

MÉMOIRE
D'HABILITATION À DIRIGER DES RECHERCHES

présenté

DEVANT L'UNIVERSITÉ DE TOULOUSE

par

JORDI INGLADA

Centre d'Études Spatiales de la Biosphère
UMR 5126 CNES, CNRS, IRD, UPS

**Contributions à l'analyse d'images d'observation de la Terre
pour la production de cartes d'occupation des sols et le suivi
des changements dans des contextes opérationnels**

Soutenu le XXX 2011 devant la commission d'examen

Composition du Jury

Président: Xxx Xxxxxx
Rapporteurs: Xxx Xxxxxx
Xxx Xxxxxx
Xxx Xxxxxx
Examineurs: Xxx Xxxxxx
Xxx Xxxxxx
Xxx Xxxxxx

Table des matières

Table des matières	1
I Synthèse des travaux	5
1 Introduction	7
1.1 Risques et catastrophes majeures	8
1.1.1 La Charte et le rôle de PM	8
1.1.2 Le développement d'outils	9
1.1.3 La déclinaison en axes de R&D	9
1.1.3.1 Le recalage d'images	9
1.1.3.2 La détection de changements	10
1.1.3.3 Les mesures de similarité	11
1.1.4 Les projets européens	11
1.2 Valorisation et promotion des images satellite à haute résolution	11
1.2.1 Spot 5 et la notion d'objet	12
1.2.1.1 La classification supervisée	12
1.2.1.2 L'extraction de primitives	12
1.2.2 Orfeo et l'interprétation de scènes	13
1.2.2.1 Le raisonnement spatial	14
1.2.2.2 Mise à jour de bases de données cartographiques	14
1.3 Conclusion	14
2 Mesures de similarité pour le recalage d'images et la détection de changements	17
2.1 Les mesures de dépendance statistique	17
2.1.1 Le coefficient de corrélation	17
2.1.2 Généralisation : interprétation probabiliste	18
2.1.3 Mesures multi-capteurs	18
2.1.3.1 Mesures utilisant les valeurs radiométriques et les probabilités	19
2.1.3.2 Mesures utilisant seulement les probabilités	19
2.2 Le recalage d'images	22
2.2.1 Modélisation du problème de mise en correspondance d'images	22

2.2.2	Modélisation de la déformation géométrique	23
2.2.3	Artefacts d'interpolation	24
2.3	La détection de changements	25
2.3.1	Comparaison de densités de probabilité	25
2.3.2	Mesures de dépendance	27
2.3.3	Classification supervisée	28
2.3.3.1	Sélection de primitives	29
2.3.3.2	Optimisation du noyau	30
2.3.3.3	Simplification de la fonction de décision	31
3	Classification pour la reconnaissance d'objets	33
3.1	Introduction	33
3.2	Classes d'objets	33
3.3	Chaîne de traitement	34
3.4	Description des imagerie	34
3.5	Résultats	35
4	Raisonnement spatial pour l'interprétation de scènes	37
4.1	Motivation	37
4.2	Region connection calculus et son extension	37
4.2.1	Le RCC-8	37
4.2.2	La reconnaissance d'objets avec le RCC-8	38
4.2.2.1	La segmentation	38
4.2.2.2	Matching de graphes	39
4.2.3	Extensions du RCC-8	40
4.3	Les relations spatiales floues	41
4.3.1	Exemple de relation spatiale : les objets alignés	41
4.3.2	Interprétation de scènes	43
5	Comparaison image/base de données	45
5.1	Introduction	45
5.2	Architecture globale	45
5.3	Éléments focaux	47
5.4	Utilisation de la théorie des croyances	47
5.5	Résultats	49
6	Perspectives de recherche	51
6.1	Les produits attendus	51
6.2	Les contraintes	52
6.3	Les solutions :	53
6.3.1	Modélisation de comportements temporels	53
6.3.2	Le multi-capteur	53
6.3.3	Les informations exogènes	53

6.3.3.1	Les modèles physiques	54
6.3.3.2	Les connaissances du domaine	54
6.4	La classification	54
6.5	La détection de changements	55
II	Curriculum Vitae	57
7	Cursus professionnel	59
7.1	Formation	59
7.2	Expérience professionnelle	59
8	Liste des publications	61
8.1	Chapitres de livre	61
8.2	Articles de revue à comité de lecture	61
8.3	Communications dans des colloques avec comité de lecture	62
9	Liste des mémoires et diplômes encadrés (2001-2011)	71
9.1	Stages	71
9.2	Thèses de doctorat	72
10	Participation à des modules d'enseignement	75
10.1	Traitement du signal radar	75
10.2	Traitement d'images radar	75
10.3	Traitement d'images pour les risques	75
10.4	Orfeo Toolbox	75
11	Projets	77
11.1	EEE-SPN (2002)	77
11.2	Robin (2005-2006)	77
11.3	GMOSS (2005-2008)	77
11.4	PREVIEW (2005-2008)	78
11.5	SAFER (2009-2012)	78
12	Autres activités liées à la recherche	81
12.1	Organisation de séminaires et colloques	81
12.2	Jurys de thèse	81
12.3	Travail éditorial	83
	Bibliographie	85
III	Annexes	89
13	Recalage d'images	91

14 Détection de changements	121
15 Reconnaissance d'objets	163
16 Raisonnement spatial	177
17 Mise à jour de cartes numériques	215
18 Traitements opérationnels	229

Première partie

Synthèse des travaux

1 Introduction



Ce mémoire synthétise les travaux de recherche que j'ai mené au CNES depuis mon arrivée en octobre 2000. Ces travaux ont été guidés par 2 axes principaux.

Le premier axe concerne l'utilisation opérationnelle des images satellite pour l'aide à la gestion de crise suite à des catastrophes majeures. On constate qu'il s'agit d'un axe applicatif – avec un ensemble de contraintes qui seront détaillées par la suite – ce qui peut sembler étrange pour des recherches menées au CNES.

Le deuxième axe concerne le travail sur l'exploitation d'images à haute et très haute résolution, soit dans un cadre de valorisation de systèmes spatiaux existants – Spot 5 – ou dans un cadre de préparation à l'utilisation de données de systèmes en développement – Orfeo/Pléiades et plus récemment Ven μ s et Sentinelle-2. Ces axes sont plus classiques dans le périmètre des compétences du CNES.

1.1 Risques et catastrophes majeures

1.1.1 La Charte et le rôle de PM

Peu avant mon arrivée au Cnes pour un post-doc, le Cnes et l'Agence spatiale européenne ont signé en avril 2000 une charte qui avait comme objectif de rendre disponible l'accès à des données satellite pour les services d'aide et de secours :

*La Charte internationale vise à offrir un système unifié d'acquisition et de livraison des données satellites dans les cas de catastrophes d'origine naturelle ou humaine par l'entremise d'utilisateurs autorisés. Chaque agence membre s'est engagée à fournir des ressources à l'appui de la Charte et contribue ainsi à atténuer les répercussions de telles catastrophes sur la vie des gens et sur la propriété.*¹

Le Cnes et l'ESA ont été ensuite rejoints par l'Agence spatiale canadienne, la NOAA (USA) et l'ISRO (Inde). La Charte compte aujourd'hui 10 membres.

Les procédures pour l'activation de la Charte par des utilisateurs autorisés ont été créées dès le début. Il en a été de même pour ce qui concerne la définition des responsabilités des différents acteurs ainsi que pour les mécanismes de demande d'acquisition des différents satellites disponibles. En revanche, rien n'était prévu pour ce qui concernait l'utilisation des images acquises. Les utilisateurs finaux étant des opérationnels de la gestion des crises, la mise à disposition d'images satellites *brutes* s'est avérée peu adaptée dès la première activation.

J'ai eu l'occasion de participer aux côtés du chef de projet de la première activation de la Charte, Francesco Sarti, lors des 2 tremblements de terre de janvier 2001 au Salvador. Dès nos premiers échanges téléphoniques avec les équipes de la Sécurité civile française envoyés sur place, nous avons pu constater l'abîme à combler entre la matrice de pixels (qui plus est, radar!) que nous avions sur nos écrans et le besoin d'information précise du pompier sur le terrain.

Pour cette activation nous avons tant bien que mal réussi à produire des cartes de changements *avant-après* mais de façon artisanale et sans beaucoup de confiance sur les résultats. Pour les activations suivantes, nous nous sommes appuyé sur les spécialistes de l'interprétation d'images du Sertit. Ceci a permis de livrer des produits de meilleure qualité, mais beaucoup de questions restaient à résoudre pour rendre ces images vraiment utiles sur le terrain :

1. Quelle est l'information utile pour l'opérationnel ? Il s'agit ici de la définition des produits cartographiques à produire. Je n'ai pas travaillé sur ce sujet.
2. Comment standardiser les procédures de façon à les rendre les moins dépendantes possible de l'interprète intervenu ou de la donnée utilisée.
3. Comment réduire les délais de mise à disposition de l'information ?

Ces questions ont motivé de développement d'outils ainsi que la définition d'axes de recherche.

1. <http://www.disasterscharter.org/>

1.1.2 Le développement d'outils

Une façon de générer des produits – cartes de dégâts – reproductibles et dans des délais réduits était d'automatiser les traitements. Ceci nous a motivé à proposer ce qui a ensuite été la *Chaîne Risques* et dont j'ai programmé une première maquette pendant mon postdoc. Le Cnes a ensuite alloué un budget qui nous a permis d'industrialiser l'outil qui a ensuite fait objet de validation et de nouveaux développements dans le cadre de projets européens.

Cette chaîne avait une architecture très simple composée de 3 blocs :

1. La mise en géométrie des images (recalage et ortho-rectification)
2. La détection de changements
3. L'interprétation des changements

L'objectif était de rendre les 2 premières étapes automatiques. La division QTIS (Qualité et traitement de l'imagerie spatiale) avait des outils qui permettaient de réaliser quasi complètement la partie géométrique. En revanche rien n'existait pour la détection de changements, même si certains croyaient qu'une simple différence d'images pouvait être utilisée pour cette détection.

Nous n'avons jamais vraiment envisagé d'automatiser l'interprétation des changements, mais nous avons développé des outils d'aide à l'interprétation, notamment dans la thèse de Tarek Habib.

1.1.3 La déclinaison en axes de R&D

La confrontation avec les besoins des utilisateurs de la Charte ainsi qu'avec les contraintes, notamment en termes de délais ont permis d'identifier les maillons manquants dans l'architecture idéale de la *Chaîne Risques*.

1.1.3.1 Le recalage d'images

Le premier manque que nous avons identifié concernait le recalage d'images. En effet, les outils maison Cnes pour les corrections géométriques (modélisation de la prise de vue, mesures de décalages locaux), bien que précis, rapides et validés, ne permettaient de couvrir que certains des besoins des applications liées aux catastrophes.

Les outils existants avaient été conçus dans un contexte mono-capteur ou, tout du moins, pour des images issues de capteurs similaires. Dans le cas de catastrophes, on ne peut pas attendre l'arrivée d'images qui soient similaires à celles que l'on retrouve dans les archives et avec lesquelles on envisage de procéder à une détection de changements.

Un autre cas non couvert par les outils disponibles à l'époque était la modélisation géométrique de la prise de vues des capteurs autres que ceux du Cnes.

Ces 2 points durs nous ont orienté vers des recherches sur le recalage sans modèle de prise de vue et la mise en correspondance d'images de modalités différentes. Le premier point n'est plus d'actualité aujourd'hui, car des bibliothèques libres de modélisation de capteurs ainsi que la généralisation de l'utilisation des modèles de fractions rationnelles rendent la modélisation géométrique relativement accessible. Nous ne nous y attarderons donc pas.

Nous détaillerons surtout l'estimation de décalages locaux pour les images multi-capteurs. Les contributions majeures ont été liées au recalage optique-radar [18] et à la gestion d'artefacts lors de l'interpolation des images [19]

1.1.3.2 La détection de changements

Même si la littérature sur les méthodes de détection de changements abrupts sur des couples d'images était relativement riche à l'époque, il fallait évaluer les approches proposées en termes de leur utilité pour les contextes opérationnels. Les critères pour le choix de méthodes étaient les suivants :

1. temps de calcul
2. robustesse aux conditions d'acquisition
3. possibilité d'utiliser des données multi-capteur
4. degré d'automatisation

Nous sommes arrivés à la conclusion que, pour les résolutions décimétriques, les approches basées sur des fenêtres locales étaient les plus appropriées et nous avons identifié les points suivants :

1. Sur les images optiques, les mesures de différence (même en utilisant les moyennes sur des voisinages locaux) sont trop sensibles aux conditions d'acquisition (angle de prise de vue, éclairages solaire). Il faut leur préférer des approches basées sur des statistiques.
2. Sur les images SAR de même incidence, le rapport de moyennes locales est assez robuste
3. Sur les images SAR acquises avec des incidences – même légèrement – différentes, il est nécessaire d'utiliser des descriptions statistiques des voisinages.

Enfin, dans le contexte multi-capteur (optique-radar) la littérature était inexistante, et les quelques essais que nous avons réalisés sont restés au stade de *prometteurs*.

Du fait que dans les premières années de la Charte très souvent les premières images disponibles étaient des acquisitions SAR, beaucoup de mes travaux ont porté là dessus. J'ai d'abord travaillé sur les images d'incidences différentes suite à l'éruption du Volcan Nyiragongo en 2002 en faisant des comparaisons entre des distributions de probabilité locales [13]. J'ai ensuite travaillé avec Grégoire Mercier sur l'extension de cette approche à des distributions quelconques avec un calcul rapide pouvant être étendu au cas multi-échelles [16]. Enfin, en collaboration avec Jean-Yves Tournet, pendant la thèse de Florent Chatelain, nous avons travaillé sur l'estimation de l'information mutuelle pour le cas particulier des lois Gamma [6, 5].

Tout ce qui a été dit jusqu'ici sur la détection de changements n'est pas directement utile pour la production de cartes d'impact, car entre 2 images entourant un événement d'intérêt il y a toujours des changements qui sont *normaux* et donc pas intéressants dans le contexte applicatif.

La séparation entre changement d'intérêt et changement non intéressant peut être envisagée de beaucoup de façons différentes (classification post-détection, détecteur par comparaison de classifications, etc.).

Nous avons seulement abordé l'approche qui consiste à faire une classification supervisée à 2 classes (changement et non changement), ce qui a été le sujet de la thèse de Tarek Habib.

1.1.3.3 Les mesures de similarité

Le recalage d'images et la détection de changements abrupts sont 2 étapes consécutives d'une chaîne de production de cartes de dégâts, par exemple. Ces 2 étapes ont beaucoup de points en commun :

- le recalage d'images peut être entendu comme la recherche d'une déformation entre 2 images permettant de les rendre les plus similaires possibles
- la détection de changements sur des images parfaitement superposables peut être posée comme la recherche de régions où le degré de similarité est inférieur à un seuil donné.

On voit dès lors que la notion de similarité est le point commun entre ces 2 procédures. Ceci a fait que le travail sur les mesures de similarité, et plus précisément celles basées sur des statistiques sur des voisinages locaux, est devenu un axe majeur des travaux de recherche qui sont synthétisés dans ce mémoire.

1.1.4 Les projets européens

Si la contribution à la Charte a été le déclencheur des travaux cités précédemment, le contexte opérationnel des activations pour des catastrophes réelles n'est pas le plus propice pour le développement et la validation des méthodes de traitement et d'analyse des images.

Le Cnes a évidemment mis en place des budgets de recherche et de développement logiciel pour la conception de nouvelles méthodes, leur validation et leur intégration dans la *Chaîne Risques*.

Nous avons aussi pu approfondir et valoriser ces travaux dans le cadre de 2 projets du Programme cadre de recherche et développement (PCRD) de l'Union Européenne. Ces 2 projets, Preview et Safer, dont le deuxième est une suite pré-opérationnelle des résultats de recherche du premier, ont permis de :

1. Continuer à développer et améliorer des méthodes de recalage automatique d'images et de détection de changements
2. Intégrer ces méthodes au sein de la *Chaîne Risques*
3. Valider les résultats par des thématiciens experts de l'interprétation d'images pour les catastrophes.

Mon rôle dans ces projets a été principalement de développer des algorithmes et de suivre l'industrialisation du logiciel.

1.2 Valorisation et promotion des images satellite à haute résolution

Un des intérêts principaux du CNES dans le développement des applications de l'imagerie de télédétection est la valorisation des capteurs développés (par le Cnes ou par ses partenaires). Lors de mon arrivée au CNES, Spot 5 était en phase finale de développement (le lancement a eu lieu en juin 2002) et le développement de Pléiades commençait.

Le développement d'outils méthodologiques dans le cadre de ces projets a guidé les axes de recherche décrits dans cette section.

1.2.1 Spot 5 et la notion d'objet

Même s'il a été lancé après Ikonos (résolution 1 m.) ou en même temps que Quickbird (résolution 60 cm.), Spot 5 et ses presque vrais 2.5 m de résolution a été un succès commercial absolu. Ceci est principalement dû à ce qu'il a été le seul capteur pouvant fournir une résolution *presque métrique* avec un champ de prise de vue comme celui des capteurs de résolution décimétrique (60 km. par rapport aux 16 km. de Quickbird).

Cette combinaison unique champ/résolution s'est avérée proche de l'optimum pour beaucoup d'applications comme l'aménagement du territoire, les risques, etc.

En revanche, les outils pour exploiter de façon efficace ces images ont dû être développés. Si la première difficulté liée à la manipulation de ces images – leur taille inédite de 24000 × 24000 pixels – a été résorbée par la loi de Moore, il y avait un autre aspect nouveau dans ces données.

Sans se rapprocher du niveau de détail de l'imagerie aérienne, on a commencé à parler d'*objets* dans les images. A cette époque, des outils du commerce comme le logiciel *eCognition*, devenu *Definiens* par la suite ont mis à la mode le terme *OBIA*, pour *Object Based Image Analysis*. Cette notion n'est pas très différente d'une segmentation (multi-échelle) suivie d'une classification hiérarchique. Le fait de faire travailler l'opérateur sur des notions de texture, compacité et forme des régions a changé la façon de travailler de beaucoup d'interprètes, mais cette approche, qui reste basée sur l'essai-erreur, ne peut être envisagée sur des gros volumes de données ou dans des contextes où les contraintes de temps de mise à disposition des données sont très fortes.

Nous avons donc lancé avec Jean-Claude Favard et Gilbert Pauc des études sur la reconnaissance d'objets sur les images Spot 5 THR (panchromatiques), notamment avec l'équipe du Professeur Georges Stamon de l'Université Paris V. En parallèle, et à l'aide de stagiaires, j'ai mené des recherches sur des approches par indexation de vignettes pour reconnaître des objets à partir d'exemples.

1.2.1.1 La classification supervisée

Afin de développer un système de reconnaissance d'objets quelconques, j'ai proposé une architecture relativement générique de classification supervisée. Au lieu de travailler par pixel, nous avons travaillé par vignette (fenêtre glissante avec un pas d'échantillonnage et une taille imposant un certain recouvrement).

A l'aide d'interprètes d'images, nous avons élaboré une base d'exemples avec une dizaine de classes (routes de différents types, ponts, bâtiments isolés, etc.) avec des spécifications très rigides en termes de taille des vignettes et localisation des objets dans la vignette.

Nous avons ensuite effectué une classification supervisée de ces vignettes. Les résultats ont donné lieu à une publication [14].

1.2.1.2 L'extraction de primitives

Qui dit classification supervisée, dit échantillon – d'apprentissage ou de validation. Sur les images Spot 1 à 4 ou Landsat, l'échantillon était le pixel et il était décrit par exemple par les

réflectances dans différentes bandes spectrales ou par des indices (NDVI, par exemple).

La classification consistait donc à demander à l'algorithme d'apprendre à reconnaître des pixels de blé, par exemple. Dans le cas de vignettes d'images panchromatiques décrivant, disons, un rond-point, ce ne sont pas les valeurs des pixels qui décrivent les classes d'objets, mais plutôt les formes et la géométrie.

Nous avons donc travaillé sur ce type de caractérisation [39].

1.2.2 Orfeo et l'interprétation de scènes

Dès 2003, j'ai commencé à travailler sur la préparation à l'utilisation des données Pléiades, dans le cadre du Programme d'accompagnement Orfeo du Cnes.

Par rapport aux précédentes missions optiques d'observation de la Terre du Cnes, ce programme préparatoire était beaucoup plus ambitieux. Il a été organisé en 2 volets :

- le volet thématique visant à recueillir les besoins des utilisateurs et à mener des études de potentialité des données, puis de validation des résultats ;
- le volet méthodologique, dont je suis devenu responsable en 2005, qui visait à mener des actions de recherche pour développer des nouvelles méthodes nécessaires pour répondre aux besoins des utilisateurs.

Le travail d'animateur du volet méthodologique m'a permis de faire un état de l'art des méthodes existantes (dont j'ai coordonné la rédaction avec des chercheurs d'une vingtaine de laboratoires de recherche en traitement des images). Cet état de l'art a ensuite permis d'alimenter les axes de recherche financés par le Cnes dans le domaine de l'extraction d'information à partir d'images à haute résolution. Les moyens du Cnes pour soutenir ce type de recherche sont les suivants :

- Les bourses de recherche (thèse et post-doc). Dans ce contexte ont été financées les thèses de Vincent Poulain et Carolina Vanegas.
- Les contrats de R&T via les Dossiers d'Axes Techniques (DAT). J'ai été responsable du DAT OT-4 de 2005 à 2009 qui a financé une moyenne de 10 contrats par an avec un budget moyen de 500 k€. J'ai aussi piloté plusieurs contrats chaque année dans ce cadre.
- Les études internes et les stages. Dans ce contexte, j'ai encadré en moyenne 2 stages par an qui ont servi à démarrer des travaux sur le raisonnement spatial (Julien Michel), la détection de changements entre images et données vecteur (Vincent Poulain), la fusion optique/SAR (Jan Wegner), la segmentation interactive d'objets (Julien Osman), la génération automatique de cartes d'occupation des sols (Christophe Lay, Malik Ciss), la détection de changements orientée objets (Éric Koun).

Le travail de recherche que j'ai mené pour le Programme Orfeo s'articule autour de 2 thématiques qui couvrent bien les besoins méthodologiques liés aux données Pléiades et Cosmo Skymed :

1. Le raisonnement spatial pour la reconnaissance d'objets complexes.
2. L'utilisation d'images à haute résolution pour la mise à jour de cartes numériques au format vecteur.

1.2.2.1 Le raisonnement spatial

Le recueil des besoins des utilisateurs thématiques Orfeo nous a permis – avec Jean-Claude Favard et Hélène de Boissezon – d’identifier un ensemble d’objets d’intérêt. La liste d’objets ainsi constituée était bien différente de ce que nous avons pu avoir pour les images moins résolues. Il s’agissait d’abord d’une liste très longue et constituée d’objets soit complexes, soit avec une variabilité intra-classe très importante.

À la différence de ce que nous avons fait pour les images Spot 5 en termes de reconnaissance d’objets, il n’était pas possible d’utiliser les approches de classification par vignette pour la reconnaissance d’objets complexes.

C’est en 2005 que j’ai passé du temps à étudier la bibliographie qui m’a permis de mettre en place un programme de recherche sur l’utilisation des techniques de raisonnement spatial pour l’interprétation de scènes de type Pléiades. Les premiers développements méthodologiques ont démarré avec le stage de Julien Michel sur l’évaluation du système RCC-8 et se sont continués par la suite par des améliorations de la procédure mise en place et l’intégration dans un système d’apprentissage supervisé [17].

Dans le cadre de la thèse d’Ahed Alboody nous avons étendu ce système de raisonnement.

La thèse de Carolina Vanegas nous a mené vers les techniques floues pour la mise en place de relations spatiales autres que celles purement topologiques du RCC-8.

1.2.2.2 Mise à jour de bases de données cartographiques

Les bases de données cartographiques numériques étant de plus en plus utilisées dans la plupart des applications thématiques de la télédétection, il a été nécessaire de se poser la question de comment les intégrer dans des approches d’analyse d’images.

L’utilisation de ces cartes numériques – qui se présentent sous la forme d’objets vectorisés (points, lignes, polygones) avec des attributs associés – se fait le plus souvent dans un contexte de mise à jour : l’image satellitaire est plus récente que la carte et sert à l’actualiser.

Si beaucoup de travaux existaient pour des résolutions plus fines (photographie aérienne) ou pour des nomenclatures bien spécifiques, nous n’avons pas trouvé d’approche générique permettant d’utiliser des types d’imagerie variés pour des nomenclatures quelconques. Le travail de thèse de Vincent Poulain – qui a suivi son stage de M2 – a consisté à démontrer qu’une approche générique pouvait être mise en place. Cette approche accepte en entrée des images optiques et SAR de résolutions allant de 50 cm. à 5 m. et peut être facilement adaptée à des types de nomenclatures – objets – différents en définissant un ensemble d’éléments focaux et leurs descripteurs associés [27].

1.3 Conclusion

Ce chapitre a présenté un survol rapide des activités de recherche menées sur les 10 dernières années. Il a été montré comment 2 cadres applicatifs déterminés par les activités du Cnes ont permis de définir des axes de recherche cohérents.

La suite du document présentera les contributions principales dans ces différents domaines

d'activité. J'éviterai de donner beaucoup de détails et renverrai le lecteur intéressé vers les publications dans les annexes.

2 Mesures de similarité pour le recalage d'images et la détection de changements

Pour les images de résolution moins fine qu'environ 5 m., le problème du recalage et celui de la détection de changements sont en quelque sorte des deux :

- le recalage, en supposant qu'il n'y a pas de changements entre les images à recaler, consiste à trouver la déformation géométrique permettant de rendre ces images les plus similaires possibles ;
- la détection de changements, en supposant que les images sont parfaitement recalées, consiste à détecter les régions où la similarité locale est faible.

Cette façon de présenter ces 2 problèmes est volontairement vague afin de rendre évident le lien entre eux. Les aspects particuliers de chacun des 2 problèmes et les contributions à leur résolution seront présentés dans les sections 2.2 et 2.3.

Dans la section suivante nous nous intéressons à l'outil de base qui sera utilisé par la suite dans les 2 cas : les mesures de similarité statistique.

2.1 Les mesures de dépendance statistique

2.1.1 Le coefficient de corrélation

Nous rappelons ici comment est calculé le coefficient de corrélation entre 2 fenêtres I et J extraites de 2 images. Les coordonnées des pixels dans les fenêtres sont notées par (x, y) :

$$\rho(I, J) = \frac{1}{N} \frac{\sum_{x,y} (I(x, y) - m_I)(J(x, y) - m_J)}{\sigma_I \sigma_J}. \quad (2.1)$$

Le coefficient de corrélation peut être caractérisé comme ceci :

- Applicable sur des images de radiométrie similaire
- Calcul rapide
- Estimation de la déformation précise
- Robuste à la présence de bruit

Cependant, son inconvénient majeur est qu'il ne peut prendre en compte que des transformations affines entre les radiométries des images comparées ($j = \alpha i + \beta$) et il ne peut donc pas être utilisé avec ces images issues de capteurs différents.

2.1.2 Généralisation : interprétation probabiliste

La formulation présentée dans l'équation 2.1 peut être approché avec un point de vue probabiliste :

$$\begin{aligned}\rho(I, J) &= \frac{1}{N} \frac{\sum_{x,y} (I(x,y) - m_I)(J(x,y) - m_J)}{\sigma_I \sigma_J} \\ &= \sum_{(i,j)} \frac{(i - m_I)(j - m_J)}{\sigma_I \sigma_J} p_{ij}\end{aligned}\quad (2.2)$$

où la somme porte sur la liste des couples radiométriques (i, j) , et p_{ij} est la valeur de l'histogramme normalisé joint (estimation de la densité de probabilité jointe, ddp, $f_{ij}(i, j)$) du couple d'images.

Ceci implique une modélisation linéaire telle que :

$$j = (i - m_I) \frac{\sigma_J}{\sigma_I} + m_J, \quad (2.3)$$

et nous évaluons sa vraisemblance par pondération de chaque couple radiométrique par p_{ij} .

On pourrait se donner d'autres modèles pour les couples radiométriques, ce qui amènerait à d'autres mesures de similarité. Par exemple, si on suppose le modèle identité, $i = j$, on obtient la norme L_n :

$$L_n(I, J) = \sum_{i,j} |i - j|^n p_{ij}, \quad (2.4)$$

Des modèles plus complexes peuvent être obtenus :

1. Moment diagonal :

$$MD(I, J) = \sum_{i,j} |i - j|(i + j - \sigma_I - \sigma_J) p_{ij}, \quad (2.5)$$

1. Cluster Shade :

$$C_{shade}(I, J) = \sum_{i,j} (i + j - \sigma_I - \sigma_J)^3 p_{ij}, \quad (2.6)$$

1. Cluster Prominence :

$$C_{pro}(I, J) = \sum_{i,j} (i + j - \sigma_I - \sigma_J)^4 p_{ij}. \quad (2.7)$$

Une étude de ces modèles peut être lue dans [4]. Ils sont très sensibles au bruit et sont rarement utilisés pour les couples d'images multi-capteurs.

2.1.3 Mesures multi-capteurs

Nous introduisons ici plusieurs mesures de similarité qui se sont montrées utiles dans le problème de recalage d'imagerie médical multi-modalités [30].

Dans la suite, les sommes sont calculées sur des valeurs de radiométrie. Nous utiliserons la moyenne conditionnelle :

$$m_{I|j} = \frac{1}{p_j} \sum_i i p_{ij}; \quad (2.8)$$

et la variance conditionnelle :

$$\sigma_{I|j}^2 = \frac{1}{p_j} \sum_i (i - m_{I|j})^2 p_{ij}. \quad (2.9)$$

Pour chacune des mesures suivantes, nous réaliserons le même type de tests de la section 2.1.1.

2.1.3.1 Mesures utilisant les valeurs radiométriques et les probabilités

Dans cette classe, nous ne prendrons pas en compte les mesures basées sur les différences de radiométries (norme L_n de la différence) [33, 8, 2], ou des mesures de texture, car leurs performances sont mauvaises.

- Écart type normalisé ou critère de Woods

Les travaux de Woods et al. d'abord sur le recalage mono-modalité [40] puis multi-modalité [41] a conduit à l'élaboration de cette mesure. Étant donnée une valeur d'intensité sur une des images, c'est à dire l'ensemble de pixels ayant cette valeur, cette mesure analyse la variabilité des valeurs des pixels homologues dans l'autre image. L'hypothèse sous-jacente est que cette variabilité (qui est en fait mesurée par la variance) sera minimum quand les images seront recalées :

$$Woods(I|J) = \sum_j \frac{\sigma_{I|j}}{m_{I|j}} p_j \quad (2.10)$$

Afin d'avoir un critère à maximiser, on utilise :

$$S_{Woods}(I|J) = 1 - \sum_j \frac{\sigma_{I|j}}{m_{I|j}} p_j \quad (2.11)$$

- Rapport de corrélation

Il s'agit ici d'une mesure bien connue des statisticiens. Sa première utilisation pour le recalage d'images a été proposée par Roche et al. [29]. Elle est définie comme ceci :

$$\eta^2(I|J) = 1 - \frac{1}{\sigma_I^2} \sum_j \sigma_{I|j}^2 p_j \quad (2.12)$$

Elle peut être interprétée de façon similaire au critère de Woods.

2.1.3.2 Mesures utilisant seulement les probabilités

Cette classe de mesures n'utilise pas les radiométries des pixels, mais seulement l'estimation de la ddp jointe. Bien entendu, cette estimation est réalisée en utilisant les pixels des images !

– Distance à l'indépendance

C'est une version normalisée du test du χ^2 :

$$\chi^2(I, J) = \sum_{i,j} \frac{(p_{ij} - p_i p_j)^2}{p_i p_j} \quad (2.13)$$

Elle mesure le degré de dépendance statistique entre les 2 images à comparer. Pour 2 variables indépendantes, la ddp jointe est identique au produit des ddp marginales. La corrélation est un test d'indépendance d'ordre 2 et celui-ci en est une généralisation.

– La famille des f-divergences

Une f-divergence [7] mesure l'espérance de la diversité du rapport de vraisemblance entre 2 distributions P et Q :

$$D_f(P, Q) = E_Q \left[f \left(\frac{dp(x)}{dq(x)} \right) \right] = \int f \left(\frac{p(x)}{q(x)} \right) q(x) dx \quad (2.14)$$

E_Q est l'espérance sur Q , $\frac{dp(x)}{dq(x)}$ est la dérivée par rapport à une densité, f est continue et convexe sur $[0, +\infty)$. Une divergence peut être interprétée comme une entropie relative. Afin de simplifier la notation, nous utiliserons : $p = p_{ij}$, $q = p_i p_j$, $\int = \sum_{i,j}$.

TABLE 2.1: Expressions de f dans les f-divergences

Mesure	$f(x)$
Distance de Kolmogorov	$\frac{1}{2} x - 1 $
Information mutuelle	$x \log x$
Divergence de Kullback	$(x - 1) \log x$
χ^2 -divergence	$\frac{1}{2}(x - 1)^2$
Distance de Hellinger	$\frac{1}{2}(\sqrt{x} - 1)^2$
Distance de Bhattacharyaa	\sqrt{x}
Distance de Toussaints	$x \frac{x-1}{x+1}$
K-divergence de Lin	$x \log \frac{2x}{1+x}$

En fonction du choix de la fonction f (voir tableau 2.1), on obtient plusieurs cas intéressants :

1. Distance de Kolmogorov :

$$V(P, Q) = \frac{1}{2} \int |p - q| \quad (2.15)$$

2. Information de Kullback ou information mutuelle :

$$K(P, Q) = \int p \log \frac{p}{q} \quad (2.16)$$

3. Divergence de Kullback :

$$K'(P, Q) = \int (q - p) (\log q - \log p) \quad (2.17)$$

4. χ^2 -divergence :

$$R(P, Q) = \frac{1}{2} \int \frac{(p - q)^2}{q} \quad (2.18)$$

5. Distance de Hellinger :

$$\mathcal{H}^2(P, Q) = \frac{1}{2} \int (\sqrt{p} - \sqrt{q})^2 \quad (2.19)$$

6. Distance de Toussaints :

$$T(P, Q) = \int p - \frac{2pq}{p+q} \quad (2.20)$$

7. K-divergence de Lin :

$$\mathcal{K}_{div}(P, Q) = \int p \log \frac{2p}{p+q} \quad (2.21)$$

Toutes ces mesures donnent des résultats très similaires [31] et ils sont aussi très proches de ceux obtenus avec la distance à l'indépendance.

– Cluster reward algorithm

Soit $H_{IJ}(k, l)$ l'histogramme joint du couple d'images et soient $H_I(k)$ et $H_J(k)$ respectivement les histogrammes marginaux et P le nombre de pixels. Nous définissons

$$I_{CRA} = \frac{\frac{\Phi}{F} - \frac{F}{P^2}}{1 - \frac{F}{P^2}}; \quad (2.22)$$

où

$$\Phi = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} H_{IJ}^2(k, l); \quad (2.23a)$$

$$F = \sqrt{h_I h_J}; \quad (2.23b)$$

$$h_I = \sum_{k=0}^{N-1} H_I^2(k); \quad (2.23c)$$

$$h_J = \sum_{k=0}^{N-1} H_J^2(k); \quad (2.23d)$$

L'indice I_{CRA} aura une valeur élevée quand l'histogramme joint a une dispersion faible. Ce manque de dispersion peut être dû à une forte corrélation (histogramme distribué le long d'une ligne) ou à cause « d'amas » de valeurs radiométriques. Dans les 2 cas, on peut prédire les valeurs de radiométrie d'une image à partir de ceux de l'autre.

Afin de comparer I_{CRA} avec les f -divergences, nous pouvons réécrire l'équation (2.22) ainsi :

$$I_{CRA} = \frac{\int p_{ij}^2 - \int p_i^2 \int p_j^2}{\sqrt{\int p_i^2 \int p_j^2 - \int p_i^2 \int p_j^2}}. \quad (2.24)$$

Si nous considérons que le dénominateur correspond à un terme de normalisation, nous pouvons nous concentrer sur le numérateur. Il contient les mêmes termes que les f -divergences, c'est à dire, un terme qui dépend de la ddp jointe et un autre qui dépend du produit des marges.

Nous pouvons donc en faire une interprétation similaire à celle des tests d'indépendance.

L'intérêt principal du CRA par rapport à la famille des $f - divergences$ réside sur le fait que le bruit d'estimation de l'histogramme a moins d'influence. Ceci permet d'estimer la mesure sur des fenêtres de taille plus faible, ce qui aura néanmoins comme conséquence de produire un pic moins franc.

2.2 Le recalage d'images

La contribution la plus importante de mes travaux dans ce domaine a été de transposer ce qui était fait dans d'autres domaines de l'imagerie, notamment médicale, au domaine de la télédétection. Il a fallu prendre en compte les spécificités de l'imagerie satellitaire (sections 2.2.1 et 2.2.2) et traiter le problème des artefacts d'interpolation liés au recalage sub-pixellique, notamment dans le cas optique-radar étudié dans la partie 2.2.3.

2.2.1 Modélisation du problème de mise en correspondance d'images

Dans cette section, nous donnons les définitions qui nous permettront de formaliser le problème de la mise en correspondance d'images. Nous commençons par définir les images *de référence* et *secondaire* :

Definition 2.2.1 *Image de référence* : image sur laquelle les autres seront mises en correspondance.

Definition 2.2.2 *Image secondaire* : image qui sera déformée géométriquement afin de la mettre en correspondance sur l'image de référence.

Deux concepts fondamentaux sont ceux de *mesure de similarité* et de *transformation géométrique* :

Definition 2.2.3 Soient I et J 2 images et c un critère de similarité, nous appelons mesure de similarité toute fonction scalaire strictement positive, S_c telle que :

$$S_c(I, J) = f(I, J, c), \quad (2.25)$$

où S_c a un maximum absolu quand les 2 images I et J sont /identiques/ au sens du critère c .

Definition 2.2.4 Une transformation géométrique T est un opérateur qui, appliqué aux coordonnées (x, y) d'un point de l'image secondaire, donne les coordonnées (u, v) de son point homologue (PH) dans l'image de référence :

$$\begin{pmatrix} u, v \end{pmatrix} = T \begin{pmatrix} x, y \end{pmatrix} \quad (2.26)$$

Enfin, nous définissons le problème de la mise en correspondance d'images :

Definition 2.2.5 *Mise en correspondance d'images* :

1. déterminer une transformation géométrique T qui maximise la similarité entre une image de référence I et le résultat de la transformation $T \circ J$:

$$\text{Arg max}_T [S_c(I, T \circ J)]; \quad (2.27)$$

2. ré-échantillonnage de J par application de T .

2.2.2 Modélisation de la déformation géométrique

La transformation géométrique de la définition 2.2.4 est utilisée pour la correction de la déformation entre 2 images à mettre en correspondance. Cette déformation, contient des informations qui sont liées à la scène observée ainsi qu'aux conditions d'acquisition. Ces déformations peuvent être classés en 3 types en fonction de leur origine physique :

1. déformation liées à l'attitude moyenne du capteur (angle d'incidence, éventuel pilotage en lacet de la plate-forme);
2. déformations liées à la parallaxe (principalement dues au relief);
3. déformations liées à l'évolution de l'attitude pendant l'acquisition (micro-vibrations présentes sur les capteurs de

Ces déformations sont caractérisées par leurs fréquence et amplitude. Elles sont présentées sur le tableau suivant :

	Amplitude	Fréquence spatiale
Attitude moyenne	Forte	Faible
Parallaxe	Moyenne	Forte et moyenne
Changement d'attitude	Faible	Faible et moyenne

En fonction du type de déformation à corriger, son modèle sera différent. Par exemple, si la seule déformation à corriger est celle introduite par l'attitude moyenne, un modèle physique de la géométrie d'acquisition, indépendant du contenu de l'image sera suffisant. Si le capteur n'est pas bien connu, cette déformation sera approchée par un modèle analytique. Quand les déformations à modéliser contiennent des hautes fréquences, les modèles analytiques ne sont pas adaptés. Dans ce cas un échantillonnage fin de la déformation, de type grille de déformation, doit être utilisé.

Les points suivants résument le problème de la modélisation des déformations :

1. Un modèle analytique est une approximation de la déformation. Il est souvent obtenu comme ceci :
 - a) Directement à partir d'une modélisation physique sans utiliser le contenu de l'image.
 - b) Par estimation des paramètres d'un modèle *a priori*. Les paramètres peuvent être estimés :
 - i. Par résolution d'équations en utilisant des PH (qui eux mêmes peuvent être obtenus manuellement ou de façon automatique).
 - ii. Par maximisation d'une mesure de similarité globale.
2. Une grille de déformation est un échantillonnage de la déformation.

Ce dernier point implique que le pas d'échantillonnage de la grille doit être suffisamment petit afin de prendre en compte les déformations de haute fréquence (théorème de Shannon). Bien entendu, si les déformations ne sont pas stationnaires – ce qui est souvent le cas pour des déformations liées au relief – l'échantillonnage peut être irrégulier.

En guise de conclusion, on peut dire que la définition 2.2.5 pose le problème de la mise en correspondance comme un problème d'optimisation. Cette optimisation peut être locale ou globale avec une mesure de similarité qui elle-même peut être locale ou globale. Le tableau 2.2 propose une synthèse du problème.

TABLE 2.2: Approches pour la mise en correspondance d'images

Modèle géométrique	Mesure de similarité	Optimisation
Physique	Aucune	Globale
Analytique avec PH	Globale	Locale
Analytique sans PH	Globale	Globale
Grille	Locale	Locale

L'approche idéale consisterait à appliquer un recalage qui soit optimisé localement du point de vue de la déformation et aussi du point de vue de la mesure de similarité. Ceci est le cas quand on utilise des grilles denses pour modéliser la déformation. Malheureusement, cette approche est la plus coûteuse en temps de calcul, ce qui fait qu'on utilise souvent, soit un sous-échantillonnage de la grille, soit une estimation de la similarité sur un nombre très restreint de points afin d'estimer un modèle analytique de déformation. Ces approximations produisent des erreurs locales qui, en fonction de la topographie locale, peuvent être de plusieurs pixels.

Même si cette précision peut être suffisante pour certaines applications (par exemple pour importer une image dans un système d'information géographique), elle ne l'est pas pour la fusion de données, la segmentation multi-canal ou la détection de changements [34]. C'est pourquoi nous nous intéressons à l'estimation utilisant des grilles denses.

Quand j'ai commencé à travailler sur ce sujet, la bibliographie disponible ne montrait pas de cas opérationnel utilisant des approches d'optimisation locale et seulement quelques auteurs, comme par exemple [23] utilisaient des mesures de similarité autres que la corrélation pour l'imagerie de télédétection. Cependant, dans le domaine de l'imagerie médicale, beaucoup de mesures de similarité avaient déjà été proposées comme des généralisations de la corrélation. Ces mesures permettent le recalage de modalités d'imagerie très différentes. Néanmoins, ces travaux n'étaient pas directement utilisables en télédétection, car les déformations modélisées en imagerie médicale sont souvent globales et utilisent des modèles paramétriques. Comme souligné plus haut, les déformations liées au relief sur les images de moyenne et haute résolution spatiale nécessitent des modèles de déformation locaux.

2.2.3 Artefacts d'interpolation

Dans l'objectif d'obtenir un recalage avec une précision sub-pixellique, la mesure de similarité est estimée sur des vignettes décalées d'une fraction de pixel. Ce décalage se réalise par interpolation d'images.

Si les images sont bien échantillonnées et que les interpolateurs sont de *bonne qualité*, cette

opération de passe sans encombre. Nous ferons ici abstraction du problème de l'échantillonnage des images – même si beaucoup de capteurs spatiaux ne respectent pas le critère de Shannon.

Étant donné que cette opération d'interpolation est coûteuse et *a priori* réalisée quelques dizaines de fois par pixel des l'image, il est habituel d'utiliser des interpolations rapides. Le coût de calcul d'une interpolation est proportionnel à la longueur (en échantillons) du filtre utilisé. La qualité de l'interpolation a une dépendance inversée par rapport à cette même longueur.

Ces interpolations approchées ne se traduisent pas seulement par des résultats bruités, mais par des artefacts avec une structure qui n'est pas spatialement aléatoire. Ceci nous a motivé à rechercher l'origine de ces artefacts et les façons de les atténuer.

Ce problème d'artefacts d'interpolation avait déjà été analysé en imagerie médicale dans le cadre limité de modèles de déformation analytiques et pour l'information mutuelle [26, 35, 21]. Nous avons généralisé ce travail en le rendant indépendant de la mesure de similarité et du modèle de déformation utilisé.

Dans [19] nous avons donné une description analytique du problème en montrant que :

1. Les artefacts étaient dus à la dépendance du degré de lissage de l'interpolateur avec le décalage sub-pixellique appliqué
2. Ces artefacts sont d'autant plus forts que l'image interpolée est bruitée
3. Un lissage préalable des images permet d'atténuer les artefacts.

2.3 La détection de changements

La détection de changements entre les images de télédétection est rendue difficile à cause de plusieurs effets, souvent présents simultanément :

1. Différence de point de vue entre les acquisitions : même dans le cas d'images parfaitement recalées, les effets d'incidence (de la lumière ou de l'onde SAR) peuvent rendre les images différentes du point de vue de la mesure physique, même en absence de tout changement.
2. Les différences d'éclairage et d'atmosphère peuvent changer aussi la mesure radiométrique

Ces 2 phénomènes, ajoutés au fait que les 2 acquisitions à comparer peuvent avoir été obtenues par des capteurs différents, font que des approches simples de mesure de variation, y compris les mesures de corrélation, s'avèrent souvent inutiles.

Nous nous sommes donc orienté vers des mesures statistiques faisant peu d'hypothèses sur les distributions des données. Nous les avons spécialisées dans le cas du radar pour aller plus loin dans la qualité de la détection.

2.3.1 Comparaison de densités de probabilité

Une des premières approches mises en place a été celle de comparer les statistiques locales dans un voisinage du pixel d'intérêt entre les 2 dates d'acquisition.

Il faut bien noter ici que nous n'estimons pas la dépendance entre les 2 dates (similarité entre les pixels), mais bien une *distance* entre les densités de probabilité.

La première contribution dans ce domaine a été d'utiliser une estimation paramétrique des densités de probabilité à l'aide du système de Pearson pour ensuite l'injecter dans la divergence de Kullback-Leibler [13]. Ceci a permis d'obtenir des résultats meilleurs que ceux obtenus par le détecteur du rapport de moyennes et celui de la distance Euclidienne entre les densités de probabilité.

L'inconvénient de cette approche était, d'un côté le coût calculatoire de l'intégration numérique dans la divergence de Kullback-Leibler, et d'un autre côté, la limitation aux lois du système de Pearson. Avec Grégoire Mercier, nous avons étendu cette approche à des lois quelconques, mais proches de la loi normale [16]. L'idée était de développer la densité de probabilité de chacune des images en série autour d'une Gaussienne, c'est le développement en série d'Edgeworth qui permet d'obtenir une expression qui ne dépend que des cumulants et qui contient très peu de coefficients si la densité n'est pas très différente de la Gaussienne.

Une fois que l'on a l'expression de la série pour les 2 variables à comparer, on peut les introduire dans l'expression de la divergence de Kullback-Leibler et on aboutit à une expression compliquée, mais qui ne dépend que des cumulants de chacune des variables.

En utilisant $\kappa_{X;i}$ pour le cumulants d'ordre i de la variable aléatoire X , la divergence de Kullback-Leibler entre les variables X et Y jusqu'à l'ordre 4 s'écrit :

$$\begin{aligned}
 KL_{\text{Edgeworth}}(X, Y) = & \frac{1}{12} \frac{\kappa_{X';3}^2}{\kappa_{X';2}^2} + \frac{1}{2} \left(\log \frac{\kappa_{Y;2}}{\kappa_{X;2}} - 1 + \frac{1}{\kappa_{Y;2}} (\kappa_{X;1} - \kappa_{Y;1} + \kappa_{X;2}^{1/2})^2 \right) \\
 & - \left(\kappa_{Y';3} \frac{a_1}{6} + \kappa_{Y';4} \frac{a_2}{24} + \kappa_{Y';3}^2 \frac{a_3}{72} \right) - \frac{1}{2} \frac{\kappa_{Y';3}^2}{36} \left(c_6 - 6 \frac{c_4}{\kappa_{X;2}} + 9 \frac{c_2}{\kappa_{Y;2}^2} \right) \\
 & - 10 \frac{\kappa_{X';3} \kappa_{Y';3} (\kappa_{X;1} - \kappa_{Y;1}) (\kappa_{X;2} - \kappa_{Y;2})}{\kappa_{Y;2}^6} \quad (2.28)
 \end{aligned}$$

où

$$\begin{aligned}
 a_1 = c_3 - 3 \frac{\alpha}{\kappa_{Y;2}}; \quad a_2 = c_4 - 6 \frac{c_2}{\kappa_{Y;2}} + \frac{3}{\kappa_{Y;2}^2}; \quad a_3 = c_6 - 15 \frac{c_4}{\kappa_{Y;2}} + 45 \frac{c_2}{\kappa_{Y;2}^2} - \frac{15}{\kappa_{Y;2}^3} \\
 c_2 = \alpha^2 + \beta^2; \quad c_3 = \alpha^3 + 3\alpha\beta^2; \quad c_4 = \alpha^4 + 6\alpha^2\beta^2 + 3\beta^4; \quad c_6 = \alpha^6 + 15\alpha^4\beta^2 + 45\alpha^2\beta^4 + 15\beta^6 \\
 \alpha = \frac{\kappa_{X;1} - \kappa_{Y;1}}{\kappa_{Y;2}}; \quad \beta = \frac{\kappa_{X;2}^{1/2}}{\kappa_{Y;2}}.
 \end{aligned}$$

Nous avons ensuite étendu cette approche à une analyse multi-résolutions, l'idée étant de mesurer cette similarité locale sur des fenêtres de tailles différentes avec un coût de calcul limité. Nous avons approché le problème en réalisant une mise à jour des moments quand un $N+1^{\text{me}}$ échantillon x_{N+1} est ajouté à un ensemble de N observations $\{x_1, x_2, \dots, x_N\}$ qui ont déjà été traitées. Pour les moments d'ordre r , on peut écrire :

$$\tilde{\mu}_{r,[N+1]} = \frac{N}{N+1} \tilde{\mu}_{r,[N]} + \frac{1}{N+1} x_{N+1}^r.$$

où $\tilde{\mu}_{r,[N]}$ (resp. $\tilde{\mu}_{r,[N+1]}$) est le moment d'ordre r estimé avec N échantillons (resp. $N+1$ échantillons). On arrive à l'expression suivante :

$$\begin{aligned}\mu_{1,[N]} &= \frac{1}{N} s_{1,[N]} \\ \mu_{r,[N]} &= \frac{1}{N} \sum_{\ell=0}^r \binom{r}{\ell} (-\mu_{1,[N]})^{r-\ell} s_{\ell,[N]},\end{aligned}\tag{2.29}$$

où la notation $s_{r,[N]} = \sum_{i=1}^N x_i^r$ a été utilisée. On peut ainsi mettre à jour le calcul des moments pour des tailles de fenêtre croissantes sans tout recalculer. Ceci permet par exemple de calculer la divergence de Kullback-Leibler sur des fenêtres de taille impaire allant de 9×9 pixels jusqu'à 51×51 pixels avec un coût de calcul proche de celui du calcul sur une fenêtre de 31×31 pixels.

2.3.2 Mesures de dépendance

Les mesures de similarité qui comparent les densités de probabilité locales ont l'avantage de n'avoir besoin que de l'estimation des marges. Par contre, elles ne sont pas capables de modéliser la dépendance entre les variables. Par exemple, elles sont insensibles à une permutation des pixels dans la fenêtre d'analyse.

Nous nous sommes donc intéressés aux mesures de dépendance statistique de même type que celles utilisées dans le recalage d'images (section 2.1). Après une petite incursion infructueuse dans le terrain des copules [24] dont l'objectif était de ne pas avoir besoin d'estimer la densité conjointe, nous nous sommes recentrés dans l'estimation de l'information mutuelle.

Les contributions principales dans ce domaine ont été faites dans la thèse de Florent Chatalein.

Nous avons commencé par évaluer l'intérêt des distributions Gamma bivariées. La première étape a consisté à mettre en place et évaluer les estimateurs des paramètres de ces lois (méthode des moments et maximum de vraisemblance). Dans un deuxième temps, ces estimateurs ont été utilisés pour mesurer l'information mutuelle entre 2 variables aléatoires suivant des lois Gamma.

L'expression de la densité d'une loi Gamma bivariée de paramètre de forme q et paramètre d'échelle P ($\Gamma(q, P)$) prend la forme suivante :

$$f_{2D}(\mathbf{x}) = \exp\left(-\frac{p_2 x_1 + p_1 x_2}{p_{12}}\right) \frac{x_1^{q-1} x_2^{q-1}}{p_{12}^q \Gamma(q)} f_q(c x_1 x_2) \mathbb{I}_{\mathbb{R}_+^2}(\mathbf{x}),$$

où $\mathbb{I}_{\mathbb{R}_+^2}(\mathbf{x})$ est l'indicatrice définie sur $[0, \infty[^2$ ($\mathbb{I}_{\mathbb{R}_+^2}(\mathbf{x}) = 1$ si $x_1 > 0, x_2 > 0$, $\mathbb{I}_{\mathbb{R}_+^2}(\mathbf{x}) = 0$ sinon), $c = \frac{p_1 p_2 - p_{12}}{p_{12}^2}$ et $f_q(z)$ est définie comme ceci :

$$f_q(z) = \sum_{k=0}^{\infty} \frac{z^k}{k! \Gamma(q+k)}.\tag{2.30}$$

En utilisant le développement mathématique présenté dans [6], on peut obtenir une approximation de l'information mutuelle entre 2 densités Gamma qui a la forme suivante ;

$$M_q \approx -\frac{1}{2} \log(1-r) + \left(q - \frac{1}{2}\right) + \log \left[\frac{\Gamma(p)}{2\sqrt{\pi}} \right] - \left(q - \frac{1}{2}\right) \frac{\Gamma'(q)}{\Gamma(q)}.\tag{2.31}$$

avec r le coefficient de corrélation.

L'intérêt de cette expression est qu'une fois le paramètre q des lois Gamma des marges estimé, le coût d'estimation de l'information mutuelle est le même que celui de la corrélation.

La limitation de cette technique est l'hypothèse de lois marginales avec le même paramètre de forme. Afin de lever cette limitation, la deuxième partie de la thèse de Florent Chatelain a consisté à développer des estimateurs pour ce qu'on a appelé les lois Gamma multivariées multicauteurs. Les résultats principaux ont été présentés dans [5] et le lecteur intéressé est invité à en consulter les détails.

2.3.3 Classification supervisée

Les techniques de détection de changements présentées plus haut ont un inconvénient majeur : elles ne sont pas capables de distinguer entre les changements *intéressants* et les autres. Même si les mesures de similarité utilisées sont robustes aux changements d'illumination, d'incidence, etc. elles ne font pas de différence entre les différents types de changements.

Dans le cas de la génération de cartes de dégâts suite à des catastrophes naturelles (inondations, tremblements de terre, etc.), il est nécessaire de distinguer les zones qui ont changé suite à l'événement catastrophique des zones qui ont eu une évolution normale (évolution de la végétation, par exemple).

Face à ce problème difficile, il nous a semblé nécessaire de placer un opérateur au sein du système de détection de changements, plutôt que de le placer après pour réaliser une sélection des changements pertinents.

L'architecture de la chaîne de changements proposée était simple. Il s'agissait de réaliser une classification à 2 classes (changement intéressant et la classe complémentaire). Nous avons abordé ce problème avec une approche pixel – les résolutions des satellites disponibles dans la Charte à l'époque ne justifiaient pas l'approche objet – et les primitives utilisées pour la caractérisation des pixels étaient toutes trouvées : les valeurs des pixels eux-mêmes et des indicateurs de changements simples (différences, ratios, corrélations).

Nous avons choisi un classifieur de type SVM à cause de sa capacité à travailler avec peu d'échantillons d'apprentissage – données par un opérateur en temps réel – et aussi parce que les SVM permettent de travailler dans des espaces de primitives à grande dimension.

Une mise en oeuvre logicielle (figure 2.1) a été mise à disposition de photo-interprètes qui l'ont validée dans le cadre du projet FP7 PREVIEW.

Cette position du problème a donné lieu à la thèse de Tarek Habib. Ce travail de thèse a eu comme objectif d'optimiser cette chaîne et de donner à l'opérateur la possibilité d'agir sur un curseur qui règle le compromis temps de calcul / qualité de la détection.

Le point de départ est donc une chaîne très simple où l'on extrait des primitives à partir des images à comparer, puis on applique une fonction de décision obtenue par classification supervisée (figure 2.2).

La figure 2.3 présente en rouge les étapes sur lesquelles des contributions ont été réalisées.

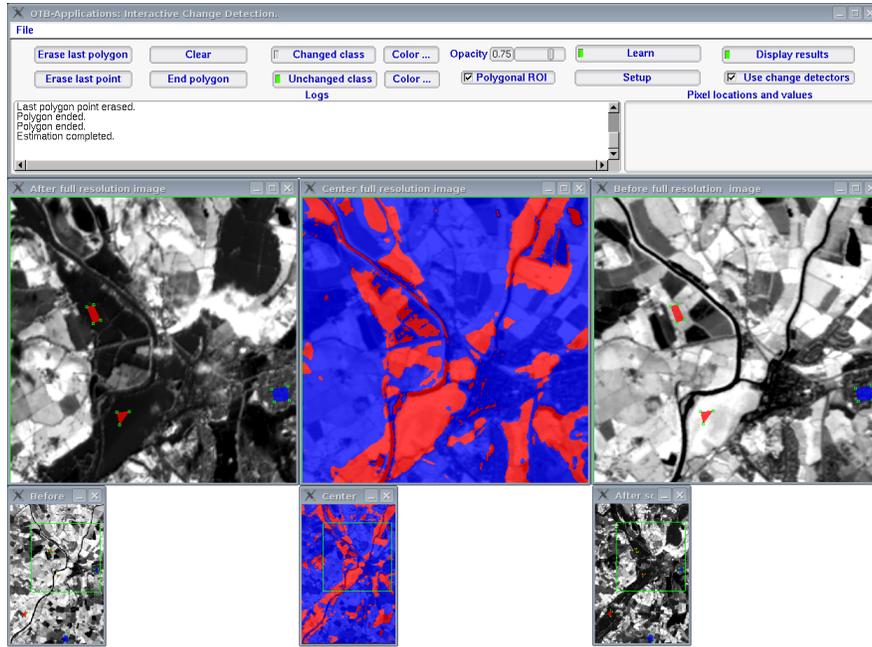


FIGURE 2.1: Logiciel de détection de changements par classification supervisée



FIGURE 2.2: Diagramme de blocs d'une chaîne de détection de changements par classification supervisée.

2.3.3.1 Sélection de primitives

Dans cette catégorie, une technique utilisant un noyau additif a été développée. Chaque composante du noyau opère sur un sous-ensemble de primitives :

$$K(X_i, X_j) = \sum_{l=1}^n K_{\gamma}(x_i^{\gamma}, x_j^{\gamma}) \quad (2.32)$$

Si un certain sous-ensemble des primitives, correspondant au sous-noyau K_{γ} , est retiré de la classification, un vecteur de test produira une erreur de classification si la condition suivant est remplie :

$$y_{test} \sum_{i=1}^m y_i \beta_i K(X_{test}, X_i) + y_{test} b \leq y_{test} \sum_{i=1}^m y_i \beta_i K_{\gamma}(x_i^{\gamma}, x_{test}^{\gamma}) \quad (2.33)$$

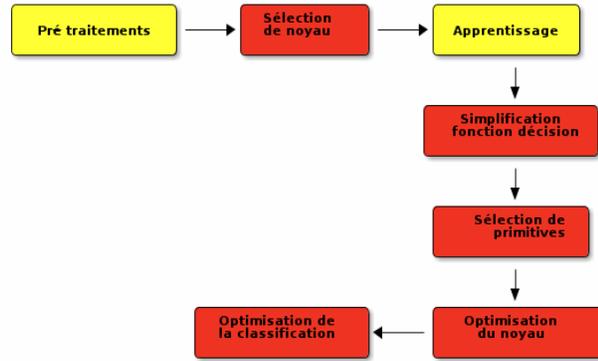


FIGURE 2.3: Diagramme de blocs de la chaîne optimisée

En utilisant cette inégalité, on peut classer les sous-ensembles de primitives en fonction du taux d'erreur produit sur un ensemble de vecteurs de test.

2.3.3.2 Optimisation du noyau

La technique mise en oeuvre ici est une optimisation de l'adéquation noyau/sous-ensemble de primitives.

En utilisant le théorème de Mercer, on peut démontrer que si $\psi(\cdot)$ est une fonction sur \mathbb{R}^p et K' est un noyau sur $\mathbb{R}^p \times \mathbb{R}^p$, alors :

$$K(x, y) = K'(\psi(x), \psi(y)) \quad (2.34)$$

est aussi un noyau. De la même façon, on peut démontrer qu'une combinaison linéaire non négative de noyaux est aussi un noyau. En combinant ces 2 propriétés, la combinaison de noyaux suivante est aussi un noyau de Mercer :

$$K(x, y) = \sum_{i=1}^M a_i K_i(\psi_i(x), \psi_i(y)) \quad (2.35)$$

Le résultat de l'équation 2.35 est particulièrement intéressant si on utilise des primitives différentes pour chaque type de noyau. Par exemple, s'il s'agit de primitives spectrales, géométriques et texturales, on peut utiliser un noyau différent pour chacun de ces groupes et composer ainsi le noyau final.

Nous avons développé une méthode pour optimiser un noyau composite en fonction des primitives disponibles en entrée du traitement.

Les sous-noyaux composant le noyau de l'équation 2.35 seront appliqués à différents groupes de primitives. La construction de ce noyau nécessite d'abord de regrouper les primitives par familles et de lister les types de noyaux élémentaires qui seront évalués.

Ensuite, une sélection des noyaux élémentaires donnant les meilleurs résultats pour chaque famille de primitives est réalisée. Pour cette sélection, de mesures de qualité sont nécessaires.

Nous avons choisi les suivantes :

- L’erreur de généralisation par *leave-one-out*.
- Le nombre d’exemples d’apprentissage mal classés.
- Une mesure originale qui combine le nombre de vecteurs support (complexité de la solution) et le nombre d’exemples mal classés : $KCost = \frac{Nb.Mis. * Nb.SV}{|w|}$.
- Une deuxième mesure originale qui intègre au $KCost$ le nombre d’itérations de l’optimisation – que nous appelons *temps* : $KTCost = \frac{Nb.Mis. * Nb.SV * time}{|w|}$.

Nous montrons ici des résultats obtenus sur des images réelles – un couple d’images SAR avant et après une éruption volcanique – sur lesquelles 3 groupes de primitives ont été calculés :

1. Différence entre les 2 dates, ratio entre les 2 dates, ratio des moyennes et ratio des médianes.
2. Corrélacion entre les 2 dates, information mutuelle entre les 2 dates, entropie et énergie (de Haralik).
3. Distance quadratique moyenne et divergence de Kullback-Leibler.

Les noyaux utilisés sont les suivants :

$$\begin{aligned} \text{Linéaire : } K(X, Y) &= \langle X, Y \rangle \\ \text{Polynomial : } K(X, Y) &= (\langle X, Y \rangle)^4 \\ \text{Gaussien : } K(X, Y) &= \exp(-0.5\|X - Y\|^2) \end{aligned}$$

Le tableau suivant montre les résultats obtenus :

Noyau	Précision
Linéaire	81.19%
Polynomial	72.64%
Gaussien	77.2%
Err. Génér.	76.9%
Err. Classif.	74.51%
KCost	86.32%
KTCost	65.22%

On observe que l’optimisation de la métrique $KCost$ permet d’obtenir un noyau additif qui fournit des résultats meilleurs que ceux obtenus par le meilleur des noyaux classiques.

2.3.3.3 Simplification de la fonction de décision

Dans le cas des noyaux non-linéaires, le coût calculatoire pour classer un vecteur est proportionnel au nombre de vecteurs support décrivant la fonction de décision.

L’objectif de cette simplification est donc d’accélérer le temps de classification.

Une première approche qui a été mise en oeuvre a consisté à classer les vecteurs support en fonction de l’erreur de classification produite quand ils sont supprimés.

Une deuxième approche plus originale a consisté à développer en série de Taylor l'expression analytique de la surface de décision :

$$f(x_{test}) = \operatorname{sgn} \left[\sum_{i=1}^m \alpha_i y_i k(x_{test}, x_i) + b \right]. \quad (2.36)$$

Pour ce faire, il faut choisir un ensemble d'origines autour desquelles le développement est réalisé. Les paramètres qui permettent de régler le compromis entre le temps de calcul et la précision de la classification sont le nombre d'origines (le nombre de séries) et l'ordre des séries.

Les résultats obtenus ont montré que l'on peut gagner un facteur 10 en temps de calcul avec une perte de 10% à 20% sur la qualité de la classification.

3 Classification pour la reconnaissance d'objets

3.1 Introduction

Notre objectif est de mettre au point un système de classification d'objets dans des images satellitaires à haute résolution (de type Spot 5). L'une des particularités de l'imagerie haute résolution est que l'information réside principalement dans la géométrie de l'image ; ainsi, un rond-point y sera mieux caractérisé par sa forme que par sa texture.

Nous avons choisi de ne prendre en compte que l'aspect géométrique des objets considérés, sans tenir compte de la radiométrie, de la texture, etc. Cela a des avantages car le système est alors indépendant de la bande spectrale ou de la saison à laquelle sont prises les images.

De plus, nous avons adopté une approche qui consiste à explorer l'image à analyser de façon séquentielle et travailler avec des vignettes de taille 100 pixels \times 100 pixels. Sur chaque vignette, nous demandons au système de déterminer la classe de l'objet centré dans la zone d'analyse. Pour ce faire, nous utilisons une procédure de classification supervisée à partir d'un apprentissage sur une base d'exemples.

Ainsi, la reconnaissance d'objets au sens où nous l'entendons consiste à effectuer une classification de l'objet présent dans une fenêtre d'analyse et ce à partir d'une caractérisation de la géométrie de l'imagette.

Ce travail a été réalisé avec l'aide de plusieurs stagiaires : Olivier Caignart, Léonard Potier et Jérôme Tagnères. Les résultats obtenus ont donné lieu à un article de revue [14].

3.2 Classes d'objets

Nous avons choisi de nous intéresser à des objets qui peuvent être intéressants dans les applications de cartographie rapide. Les classes choisies sont les suivantes :

1. **BT** bâtiments isolés ;
2. **CH** chemins ;
3. **CR** croisements ;
4. **PT** ponts ;
5. **RD** routes départementales ;
6. **RN** routes nationales ;
7. **RP** ronds-point ;

- 8. **RS** routes secondaires ;
- 9. **VF** voies ferrées ;
- 10. **ZP** zones pavillonnaires.

Par ailleurs, nous avons ajouté une 11^e classe d'objets :

- **AU** autres.

Il faut noter que la prise en compte des routes dans la liste des classes nous sert à mieux caractériser la classe de rejet, mais qu'en aucun cas nous avons envisagé de faire une détection de routes avec l'approche de reconnaissance d'objets décrite dans ce document. En effet, d'autres techniques de détection de routes beaucoup plus efficaces existent dans la littérature.

3.3 Chaîne de traitement

La chaîne de traitement comprend trois étapes :

1. L'apprentissage est réalisé à partir d'un pourcentage des imagettes ; ce pourcentage peut être fixé à n'importe quelle valeur. On détermine le vecteur caractéristique (ou vecteur des descripteurs) de chacune d'elles à partir de différents traitements (moments complexes, transformée de Fourier-Mellin, etc.), puis on fournit les vecteurs ainsi déterminés au classifieur qui va déterminer la surface de séparation optimale pour chacun des couples de classes.
2. Le test de performances est à son tour effectué, à partir des imagettes restantes. On détermine là encore les vecteurs caractéristiques, puis on les soumet au classifieur selon la stratégie « un contre un » présentée un peu plus loin.
3. L'application à une image complète consiste à essayer de reconnaître les objets dans une image de grande taille par balayage séquentiel et application du système de classification issu de l'apprentissage. Cette étape n'est pas été mise en œuvre.

Les deux premières étapes sont représentées sur la figure 3.1.

3.4 Description des imagettes

La figure 3.2 illustre l'obtention du vecteur caractéristique de dimension N pour chacune des imagettes ; voici la signification des quatre étapes apparentes :

- Étape 1 imagette d'origine ;
- Étape 2 extraction d'information ;
- Étape 3 calcul des vecteurs des différents traitements ;
- Étape 4 concaténation pour former le vecteur caractéristique.

Les études sur la sélection de primitives n'ont pas permis de réduire de façon significative le nombre de composantes du vecteur descripteur. La robustesse des SVM au phénomène de Hughes – l'augmentation de la dimensionalité de l'espace de descripteurs nécessiterait beaucoup d'échantillons – nous a permis d'obtenir des résultats satisfaisants.

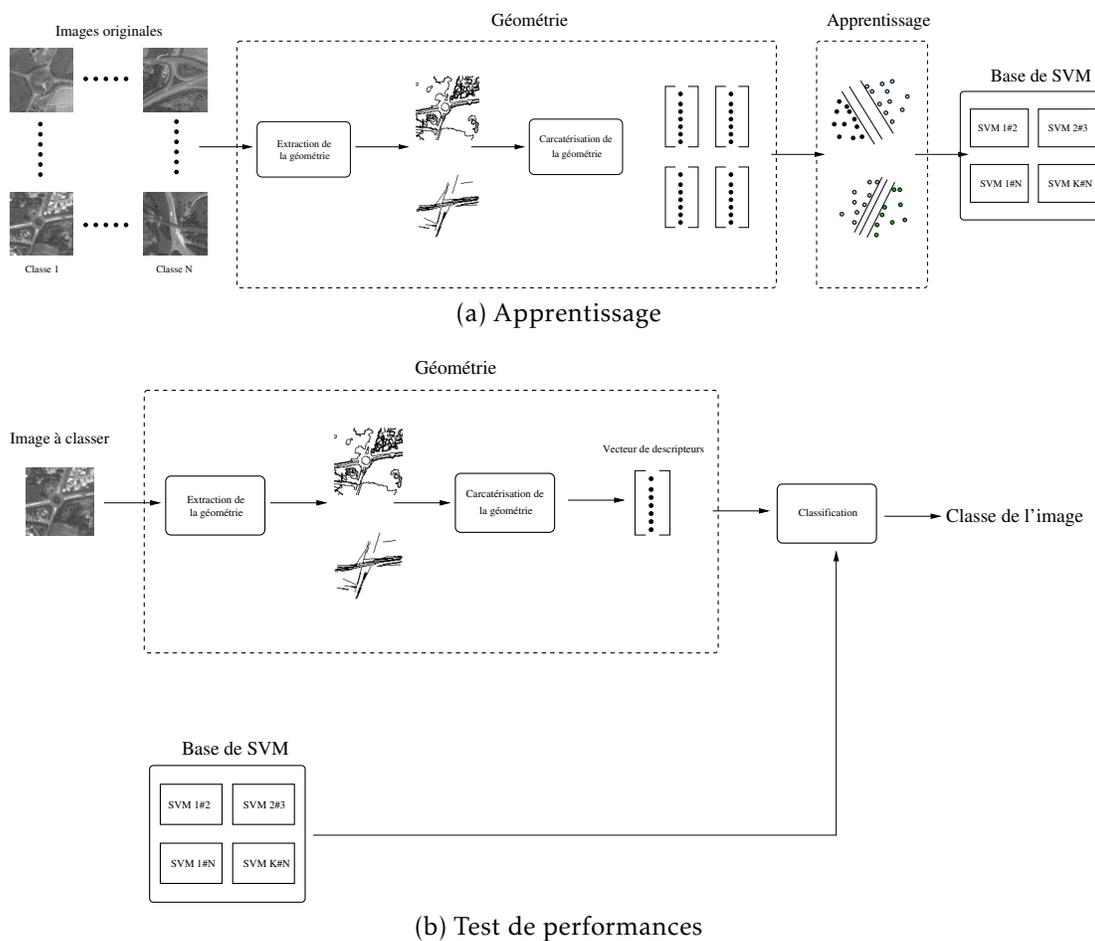


FIGURE 3.1: Les deux étapes de la chaîne de traitement.

3.5 Résultats

La matrice de confusion 3.1 montre un exemple de performances du système mis en place pour la détection de bâtiments isolés, rond-points et ponts.

TABLE 3.1: Matrice de confusion

	BT	RP	AU	PT	Pnd
BT	92.93	1.30	5.23	0.52	7.06
RP	4.01	80.80	4.91	10.26	19.19
AU	10.59	5.57	79.73	4.08	20.26
PT	2.29	8.97	2.29	86.43	13.56
Pfa	5.63	5.28	4.14	4.96	

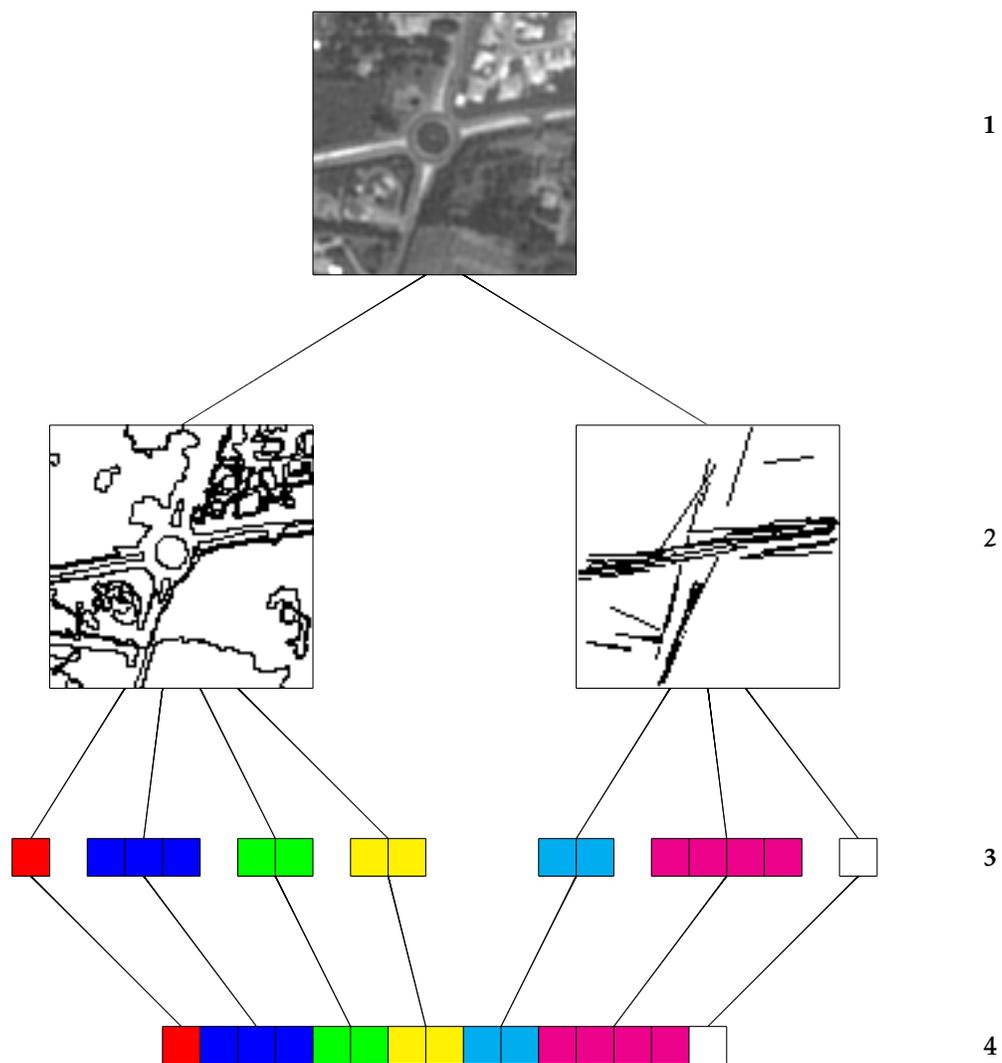


FIGURE 3.2: Obtention du vecteur caractéristique.

4 Raisonnement spatial pour l'interprétation de scènes

4.1 Motivation

L'utilisation de primitives de bas niveau comme celles utilisées dans le chapitre 3 ne permet pas de décrire de façon efficace des objets complexes ou composites. Dans les images à résolution métrique et sub-métrique, beaucoup d'objets deviennent des objets composites. Cette recherche a donc été motivée par les activités de préparation à l'utilisation des images Pléiades.

La diversité des objets d'intérêt identifiées par les utilisateurs thématiques impliqués dans le Programme préparatoire Orfeo a mis en évidence que, si on voulait échapper à une mise en oeuvre de chaînes spécifiques par type d'objet, des outils puissants de description à haut niveau d'abstraction sémantique étaient nécessaires. Nous nous sommes donc orientés vers les outils de description qualitative.

Ces outils peuvent être classés en 2 grandes familles :

1. Approches qualitatives du raisonnement métrique : on introduit l'incertitude et l'imprécision dans des descripteurs spatiaux classiques.
2. Approches topologiques : représentation qualitative de l'espace.

Nous avons commencé nos travaux par le 2ème point ci dessus, notamment sur l'utilisation du *Region Connexion Calculus* et ses extensions (stage de Julien Michel et études associées [17] et thèse d'Ahed Alboody [1]). Puis nous avons aussi abordé les approches qualitatives du raisonnement métrique via les relations spatiales floues (thèse de Carolina Vanegas [36]).

4.2 Region connection calculus et son extension

4.2.1 Le RCC-8

Le *Region Connection Calculus* (RCC) [12] est basé sur la notion de connexion entre couples de régions de l'espace. Ces connexions appartiennent à un ensemble fini de possibilités. Différents RCCs peuvent donc être obtenus en fonction du nombre de connexions considérées. Un de ces systèmes, le RCC-8, est spécialement intéressant, car il est composé d'un ensemble de relations exhaustives et mutuellement exclusives. Ces propriétés rendent le raisonnement plus simple. La figure 4.1 montre les 8 relations qui vont de la déconnexion (DC) jusqu'à l'équiva-

lence (EQ) en passant par la connexion externe (EC), le recouvrement partiel (PO), la partie propre tangentielle (TPP), la non tangentielle (NTPP) et les inverses de ces 2 dernières (TPPi et NTPPi).

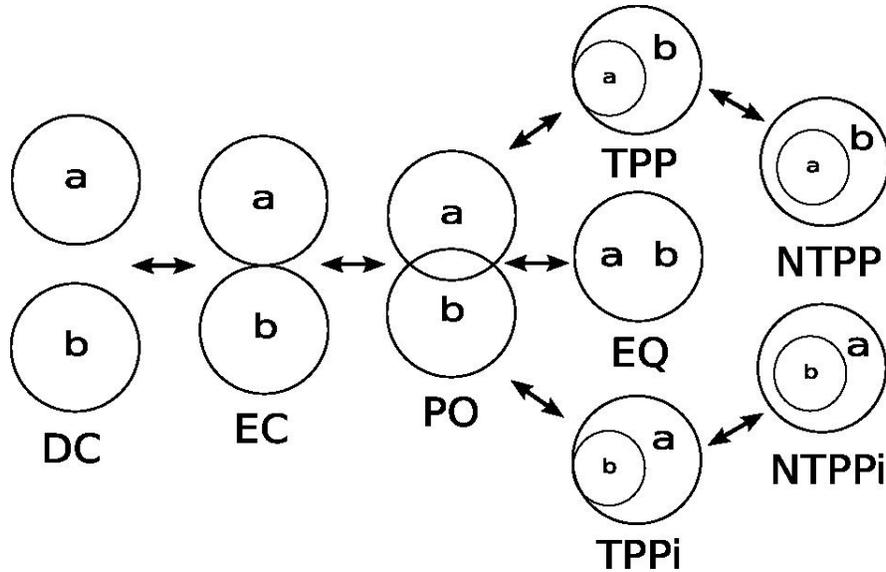


FIGURE 4.1:

Ces relations sont symétriques à l'exception de TPP et NTPP, d'où le besoin de prendre en compte leurs inverses. Dans le cas contraire, on parle de RCC-6. Si on ne prend pas en compte l'inclusion tangentielle, on parle de RCC-5.

Le raisonnement sur les relations du RCC-8 est facilité par la possibilité de déduire la relation reliant 2 régions *a* et *c* à partir de la connaissance des relations reliant *a* et *b* et *b* et *c*. La table de composition 4.1 synthétise cette information. Les cases vides indiquent qu'il est impossible d'inférer une quelconque information.

4.2.2 La reconnaissance d'objets avec le RCC-8

Après quelques travaux prospectifs sur la pertinence du RCC-8 pour décrire des objets complexes, nous avons abordé le problème de la reconnaissance d'objets.

4.2.2.1 La segmentation

Le point de départ du calcul des relations spatiales est l'ensemble de régions de l'espace à analyser. Ceci implique une première étape de segmentation d'images. Le choix d'un algorithme de segmentation n'est pas trivial, surtout si on envisage cela dans un contexte automatique.

Afin d'exploiter un maximum de relations du RCC-8, on ne peut pas utiliser des segmentations qui donnent une partition simple de l'image où chaque pixel appartient à une seule ré-

TABLE 4.1: Table de composition du RCC-8

$R_1 \setminus R_2$	DC	EC	PO	TPP	NTPP	TPPi	NTPPi	EQ
DC		DC, EC, PO, TPP, NTPP	DC, EC, PO, TPP, NTPP	DC, EC, PO, TPP, NTPP	DC, EC, PO, TPP, NTPP	DC	DC	DC
EC	DC, EC, PO, TPPi, NTPPi	DC, EC, PO, TPP, TPPi, EQ	DC, EC, PO, TPP, NTPP	EC, PO, TPP, NTPP	PO, TPP, NTPP	DC, EC	DC	EC
PO	DC, EC, PO, TPPi, NTPPi	DC, EC, PO, TPPi, NTPPi		PO, TPP, NTPP	PO, TPP, NTPP	DC, EC, PO, TPPi, NTPPi	DC, EC, PO, TPPi, NTPPi	PO
TPP	DC	DC, EC	DC, EC, PO, TPP, NTPP	TPP, NTPP	NTPP	DC, EC, PO, TPP, TPPi, EQ	DC, EC, PO, TPPi, NTPPi	TPP
NTPP	DC	DC	DC, EC, PO, TPP, NTPP	NTPP	NTPP	DC, EC, PO, TPP, NTPP		NTPP
TPPi	DC, EC, PO, TPPi, NTPPi	EC, PO, TPPi, NTPPi	PO, TPPi, NTPPi	PO, TPP, TPPi, EQ	PO, TPP, NTPP	TPPi, NTPPi	NTPPi	TPPi
NTPPi	DC, EC, PO, TPPi, NTPPi	PO, TPPi, NTPPi	PO, TPPi, NTPPi	PO, TPPi, NTPPi	PO, TPP, NTPP, TPPi, EQ	NTPPi	NTPPi	NTPPi
EQ	DC	EC	PO	TPP	NTPP	TPPi	NTPPi	EQ

gion. En effet, dans ce cas, les seules relations présentes seraient EC et DC. Nous nous sommes donc orientés vers des techniques de segmentation multi-échelles.

Le premier choix a été celui de la pyramide morphologique [22], qui a rapidement montré ses limites, notamment en termes de distorsion des régions à cause de l'élément structurant. Nous avons ensuite utilisé les profils morphologiques de Pesaresi et Benediktsson [25] qui ont donné des résultats beaucoup moins bruités.

A la fin de l'étape de segmentation, nous obtenons une suite de segmentations, en fonction des échelles. Nous avons choisi de représenter ce résultat sous forme de graphe où chaque noeud correspond à une région (toutes échelles confondues) et les arêtes représentent les relations au sens du RCC-8.

Afin de rendre la visualisation des résultats exploitable, les régions sont représentées à des niveaux différents en fonction de leur échelle d'extraction et la relation DC n'est pas représentée. Un exemple de ce type de représentation est donné dans la figure 4.2.

4.2.2.2 Matching de graphes

La représentation sous forme de graphe d'une scène permet de mettre en place de façon aisée des techniques de reconnaissance d'objets. L'idée de base consiste à disposer d'un graphe qui modélise l'objet à reconnaître, puis à faire de la mise en correspondance de (sous-)graphes avec le graphe représentant la scène dans laquelle on veut faire la reconnaissance.

Un point intéressant à souligner est que le graphe de l'objet à reconnaître peut être obtenu à partir d'exemples (imagettes contenant des objets de la classe d'intérêt), mais il peut être aussi synthétique, c'est à dire généré à partir d'une description éventuellement linguistique.

Cette mise en correspondance de graphes suppose l'existence d'une métrique pour com-

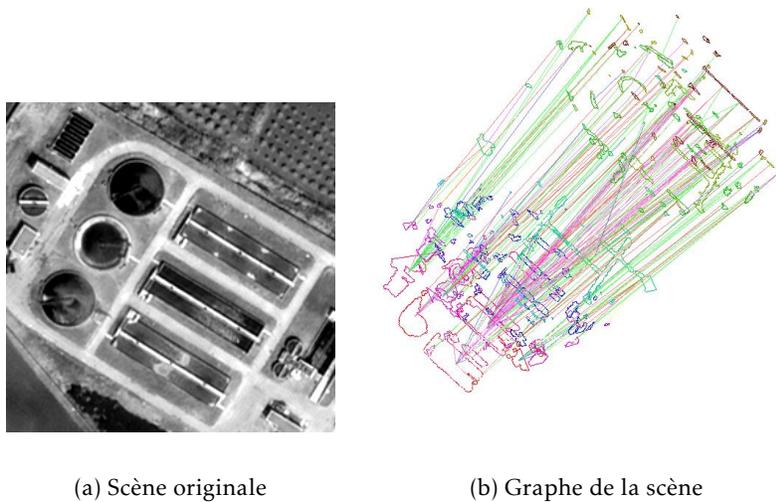


FIGURE 4.2: Exemple de graphe à 6 niveaux sur un extrait d'image Quickbird à 60 cm. de résolution.

parer les graphes. On dispose ici d'un degré de liberté très utile, car cette métrique peut se limiter à comparer la structure des graphes et les étiquettes des arêtes (relations entre les régions), mais elle peut aller beaucoup plus loin.

Nous avons mis en place une métrique qui compare aussi les régions (en termes de géométrie – forme, taille – et radiométrie). Ce type de métrique a 2 avantages :

1. Elle permet de trouver des objets avec un degré de ressemblance qui peut être choisi (privilégier la géométrie au lieu de la radiométrie, par exemple).
2. La comparaison de régions permet de façon rapide la sélection de régions candidates pour le matching sans avoir à utiliser les homéomorphismes entre les graphes dans tous les cas.

Le lecteur peut se référer à la publication [17] pour des exemples de résultats.

4.2.3 Extensions du RCC-8

Une fois que nous disposons d'une représentation de la scène en termes de graphes de régions et de relations spatiales, l'origine de ces régions n'est pas importante. Ceci ouvre la porte à l'utilisation de ce formalisme sur d'autres types de données que les images.

Dans le domaine de la télédétection, l'autre type de donnée qui peut bénéficier de ce type de représentation sont les bases de données géographiques sous forme de donnée vecteur.

Dans la communauté géomatique, l'utilisation de la topologie est fréquente pour l'analyse des données.

Nous avons souhaité utiliser ce qui avait été développé pour les images à haute résolution spatiale pour les bases de données géographiques et pour la comparaison entre ces bases et les images.

Dans la thèse d’Ahed Alboody, les travaux se sont orientés vers l’enrichissement du RCC-8. En effet, les régions issues des bases de données géographiques, sont presque exclusivement reliées par des relations DC et EC, ce qui limite l’expressivité d’une analyse à base de RCC-8. Nous avons travaillé sur la spécialisation des relations du RCC-8 afin de prendre en compte le nombre d’intersections entre les régions (lignes et points).

Ceci a permis de générer un grand nombre de relations possibles entre les régions. Ces relations ont été appliquées à la détection de changements entre images et bases de données.

Une incursion dans le domaine du flou a aussi été réalisée.

4.3 Les relations spatiales floues

Les relations de connexion modélisées par le *Region Connection Calculus*, même avec les extensions proposées dans la thèse d’Ahed Alboody, restent d’un niveau sémantique relativement peu élevé.

Lors qu’on essaie de décrire des objets complexes, il est utile de travailler à un niveau d’abstraction plus élevé qui se rapproche des descriptions linguistiques qui peuvent être élaborées par un interprète humain.

Les notions de parallélisme, d’alignement, d’entourage ou de traverse permettent de construire des descriptions comme « la maison est entre le jardin et la route », « les bateaux sont alignés le long du quai », « le jardin entoure la piscine » ou « le chemin traverse le parc ».

Cependant, quand on essaie de *mesurer* ces relations entre les objets d’une image, on ne peut pas se limiter à des réponses binaires du type « les objets sont parallèles ». Toutes les notions citées plus haut peuvent être accompagnées d’un degré de satisfaction.

Ces réflexions ont été le point de départ des travaux de recherche menés dans la thèse de Carolina Vanegas que j’ai co-encadrée avec Isabelle Bloch.

4.3.1 Exemple de relation spatiale : les objets alignés

Parmi les relations spatiales floues développées dans la thèse de Carolina Vanegas, je propose ici un zoom sur l’alignement d’objets.

On suppose que l’image a été segmentée et nous construisons un graphe d’adjacence des objets de l’image. Nous construisons aussi son graphe dual qui contiendra de l’information sur l’orientation relative des objets. Les groupes d’objets satisfaisant un critère flou sur leur alignement local seront extraits du graphe. Ces groupes sont des candidats pour les alignements globaux.

L’outil choisi pour mesurer l’orientation relative entre 2 objets, est l’histogramme d’orientations défini comme suit : Soient a et b 2 objets de l’image I , l’histogramme des angles entre a et b (de a par rapport à b) est obtenu en calculant pour chaque paire de points $p_a \in a$ et $p_b \in b$ l’angle entre le segment qui les unit et l’axe horizontal, qui est noté $\angle(p_a, p_b)$. Ces angles permettent de construire un histogramme normalisé :

$$H^a(b)(\theta) = \frac{\sum_{p_a \in a, p_b \in b | \angle(p_a, p_b) = \theta} 1}{\max_{\phi \in [0, 2\pi)} \sum_{p_a \in a, p_b \in b | \angle(p_a, p_b) = \phi} 1}. \quad (4.1)$$

Afin de déterminer si un objet a est dans une certaine direction par rapport à un objet b – par exemple « à la droite de » – nous pouvons calculer $H^a(b)$ et le comparer à un modèle d'histogramme représentant la notion recherchée.

L'histogramme d'orientations entre 2 objets par rapport à l'axe horizontal est défini comme l'histogramme d'angles entre ces 2 objets, calculé modulo π et de support π . Dans le cas où a et b sont des objets flous avec fonctions d'appartenance $\mu_a : \mathbb{I} \rightarrow [0, 1]$ et $\mu_b : \mathbb{I} \rightarrow [0, 1]$, respectivement, l'histogramme d'orientations devient :

$$O(a, b)(\theta) = \frac{\sum_{p_a, p_b \in \mathbb{I} \mid \text{mod}(\angle(p_a, p_b), \pi) = \theta} \mu_a(p_a) \wedge \mu_b(p_b)}{\max_{\phi \in [0, \pi]} \sum_{p_a, p_b \in \mathbb{I} \mid \text{mod}(\angle(p_a, p_b), \pi) = \phi} \mu_a(p_a) \wedge \mu_b(p_b)}, \quad (4.2)$$

où \wedge est une T-norme.

Afin de mesurer la similarité entre 2 histogrammes d'orientation, il faut considérer l'imprécision qui est liée à la comparaison d'angles similaires. Ceci est fait à l'aide de dilatations floues des histogrammes avec un élément structurant ν_0 approprié. On peut ensuite mesurer la similarité entre ces 2 histogrammes en prenant le maximum de leur intersection.

$$\text{sim}(O(a, b), O(c, d)) = \max_{\theta \in [0, \pi]} [D_{\nu_0}(O(a, b)) \wedge D_{\nu_0}(O(c, d))](\theta), \quad (4.3)$$

où \wedge est une T-norme et $D_{\nu_0}(\mu)(\theta) = \sup_{\tilde{\theta} \in [0, \pi]} \min(\mu(\tilde{\theta}), \nu_0(\theta - \tilde{\theta}))$ est la dilatation floue. L'extension à la mesure de similarité de plusieurs histogrammes est directe :

$$\text{sim}(O(a_0, b_0), \dots, O(a_N, b_N)) = \max_{\theta \in [0, \pi]} \bigwedge_{i=0}^N D_{\nu_0}(O(a_i, b_i))(\theta). \quad (4.4)$$

On peut ensuite définir les notions d'alignement global et alignement local. Supposons que nous disposons d'une relation de voisinage entre 2 objets $\text{Neigh}(a, b)$. On dit qu'un groupe d'objets est *globalement aligné* si tous ses membres sont connectés par Neigh et il existe un angle θ tel que chaque membre du groupe voit les autres membres dans cette direction θ ou $\theta + \pi$. Le degré d'alignement global est défini comme ceci :

Definition 4.3.1 Soit $S = \{a_0, \dots, a_N\}$, avec $N \geq 3$, un groupe d'objets dans \mathbb{I} , connectés par Neigh . Le degré d'alignement global de S est donné par :

$$\mu_{\text{ALIGN}}(S) = \text{sim}(O(a_0, S \setminus \{a_0\}), \dots, O(a_N, S \setminus \{a_N\})). \quad (4.5)$$

Un groupe S avec $\mu_{\text{ALIGN}}(S) = \beta$ est globalement aligné avec degré β .

Un groupe $S = \{a_0, \dots, a_N\}$ est localement aligné avec degré β , si pour chaque couple de couples d'objets voisins ayant un objet en commun, les orientations entre les objets de chaque couple sont similaires avec un degré β et si le groupe est connecté par la relation Neigh . Autrement dit, un groupe S avec $|S| \geq 3$ est localement aligné avec degré β s'il satisfait les relations suivantes :

$$R1 : \forall x, y, z (\text{Neigh}(x, y) \wedge \text{Neigh}(y, z)) \Rightarrow (\text{sim}(O(x, y), O(y, z)) \geq \beta)$$

$$R2 : \forall a, b \exists x_0, \dots, x_m \text{ for } m > 1 \text{ such that } x_0 = a, x_m = b \text{ and } \bigwedge_{i=0}^{m-1} \text{Neigh}(x_i, x_{i+1})$$

Avec ces définitions et l'utilisation de graphes d'adjacence que nous ne décrivons pas ici, on peut obtenir des résultats très intéressants sur des images réelles. Un exemple est donné sur la figure 4.3 où des groupes de bâtiments globalement alignés sont détectés.

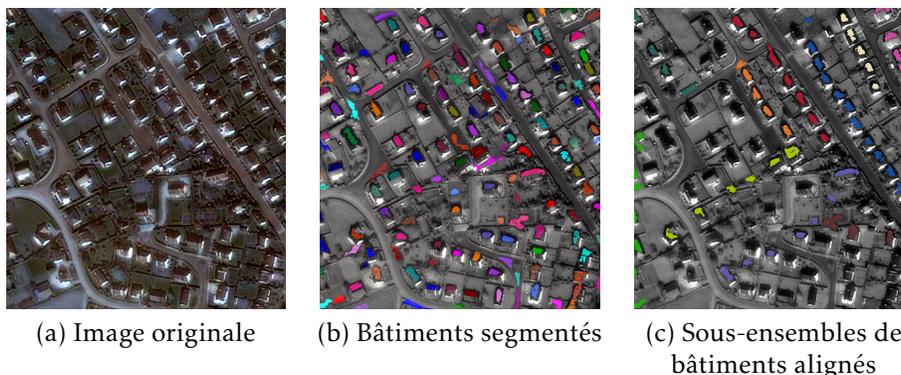


FIGURE 4.3: Détection d'objets globalement alignés avec un degré supérieur à 0.9.

4.3.2 Interprétation de scènes

Une fois qu'un ensemble de relations spatiales floues est disponible, il se pose le problème de comment les utiliser pour faire de l'interprétation de scènes. D'une façon similaire à ce qui a été présenté pour les relations du RCC-8, on peut faire de la mise en correspondance de modèles d'objet avec les graphes extraits à partir de la scène à interpréter.

Cependant, cette approche trouve vite des limites de complexité dans le cas des relations spatiales floues présentées dans cette partie du travail. En effet, étant donné qu'on veut représenter des connaissances d'un niveau sémantique élevé, il faut des outils de représentation des connaissances. Dans le cas de l'interprétation d'images à haute résolution, les concepts manipulés sont bien représentés par des description linguistiques, car ils relèvent souvent d'un travail d'interprétation humaine.

C'est pourquoi, pour réaliser l'interprétation des images en utilisant les relations spatiales floues, nous avons choisi l'approche des systèmes à base de connaissances et plus particulièrement les graphes conceptuels. Le choix de cet outil par rapport aux logiques de description a été motivé par l'aisance de la représentation graphique associée qui permet de construire facilement des modèles des concepts d'intérêt.

Afin d'utiliser les graphes conceptuels, il est nécessaire de disposer des éléments suivants :

1. Une hiérarchie de concepts : ceci correspond à une nomenclature de classification de type Corine Land Cover, par exemple. Dans notre cas, des terminologies qui ne sont pas normalisées ont été utilisées. Ceci est dû au fait qu'il n'existe pas de standard dans ce domaine pour les images à haute résolution spatiale.
2. Une hiérarchie de relations : des hiérarchies de relations spatiales disponibles dans la littérature – elles organisent les relations en *métriques* et *topologiques* et leurs sous classes

distance et directionnelles et adjacence, recouvrement, etc. – ont été étendues pour y inclure les relations floues développées ici.

3. Les graphes des concepts d'intérêt qui seront recherchés dans les images.

Le problème de recherche de concepts d'intérêt dans une image peut être abordé avec les approches suivantes, organisées par difficulté croissante :

1. Image dont les régions sont étiquetées et pour lesquelles on ne considère que des relations non floues.
2. Image étiquetée en considérant les relations floues.
3. Image non étiquetée et relations floues.

La notion d'étiquette correspond ici à l'identification du concept associé à chaque région de l'image (classification suivant la nomenclature hiérarchique).

Une fois que la hiérarchie de concepts, la hiérarchie de relations et les graphes conceptuels des objets d'intérêt sont disponibles, on peut mettre en oeuvre la recherche des objets dans des scènes complexes.

L'approche qui a été choisie pour rechercher le graphe d'un objet d'intérêt dans une image est celle des CSP (Constraint Satisfaction Problems). Un CSP [3], est défini par le triplet $\mathcal{P} = \langle \mathcal{X}, \mathcal{D}, \mathcal{C} \rangle$ où \mathcal{X} est l'ensemble de variables qui représente une caractéristique des objets du problème – les noeuds du graphe conceptuel dans notre cas –, \mathcal{D} est l'ensemble représentant les valeurs possibles des variables – les régions de l'image – et \mathcal{C} sont les contraintes du problème – les relations spatiales.

Afin d'exploiter les relations spatiales floues, nous avons utilisé les Fuzzy CSP [9].

5 Comparaison image/base de données

5.1 Introduction

Un des besoins principaux identifié avec les utilisateurs thématiques du Programme préparatoire Orfeo a été la mise à jour de bases de données cartographiques.

On peut, à juste titre, se poser la question de la nouveauté de ce besoin quand l'Institut géographique national (IGN) mène cette activité de façon routinière et à partir d'imagerie aérienne.

Le travail avec des utilisateurs nous a permis de constater que leur besoin était différent de celui qui est rempli par l'IGN. En effet, beaucoup d'utilisateurs travaillent sur des nomenclatures différentes de celles normalisées par l'IGN et aussi les données utilisées en entrée sont-elles différentes de la photographie aérienne à 25 cm de résolution.

Dans le travail de thèse de Vincent Poulain, nous avons voulu traiter le problème de la mise à jour de bases de données cartographiques à partir d'imagerie satellitaire en faisant un minimum d'hypothèses sur les données en entrée et sur les nomenclatures à utiliser.

Étant donné que ce travail a été réalisé dans le cadre du Programme Orfeo, les données image privilégiées étaient celles de type Pléiades et Cosmo Skymed, mais nous avons considérée que d'autres capteurs seraient aussi disponibles. Nous avons donc considéré des données optiques et SAR avec des résolutions allant de 60 cm jusqu'à 3 m. Ceci comprend des capteurs comme Pléiades, Quickbird, Ikonos, Spot 5, Formosat-2 pour l'optique et Cosmo Skymed, TerraSAR-X et Radarsat-2 pour le SAR.

5.2 Architecture globale

Le problème est posé de la façon suivante. On dispose d'une ou plusieurs images acquises sur une période récente. On souhaite, soit mettre à jour une base de données cartographique existante et plus ancienne que les images acquises, soit créer une nouvelle base de données (pas de mise à jour).

La figure 5.1 présente l'architecture générique mise en oeuvre. La première étape est une mise en correspondance géométrique des données disponibles. Cette mise en correspondance peut être grossière, car les traitements qui seront mis en place par la suite sont robustes à des légers décalages.

Il faut noter que nous avons aussi travaillé sur le recalage optique/SAR à haute résolution

pendant les travaux de projet de fin d'études de Jan Wegner [37, 38]. Mais ces travaux ne seront pas abordés ici.

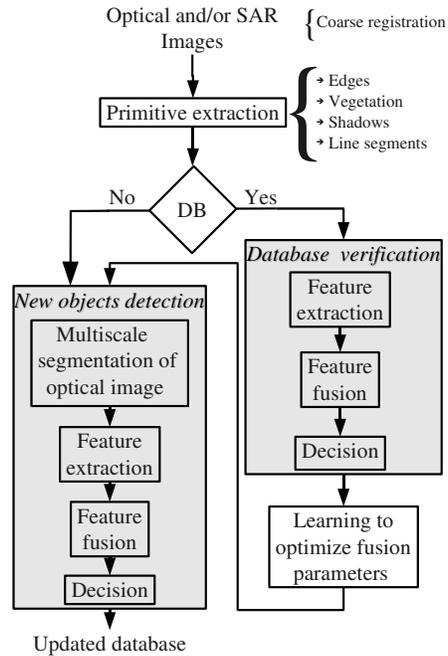


FIGURE 5.1: Architecture générique d'un chaîne de mise à jour de bases de données cartographiques.

Cette chaîne de traitement ne travaille que sur un type d'objet cartographique à la fois, mais elle peut être instanciée pour différents types d'objets avec peu de modification. Pendant la thèse de Vincent Poulain, nous avons traité 2 types d'objets : les bâtiments et les routes. Ceci a permis de valider la généralité de l'approche, car ce sont 2 types d'objets très différents en termes de propriétés radiométriques et géométriques.

La première étape des traitements consiste à extraire un ensemble de descripteurs qui aideront à décrire les objets. Nous nous sommes appuyés ici sur la multitude de travaux existants dans la littérature et leurs mises en œuvre disponibles notamment dans l'Orfeo Toolbox.

Ensuite, 2 cas sont possibles : une base de données existe et doit être mise à jour ou il n'y a pas de base de données et elle doit être créée. Dans le premier cas, on procède d'abord à vérifier l'existence, dans les images, des objets de la base de données. Dans le cas où il n'y a pas de base de données, on procède directement à la détection de nouveaux objets. Cette étape est aussi réalisée dans le cas où une base de données existe, avec la différence que, suite à la vérification de la base de données, on peut ajuster les paramètres des algorithmes d'extraction de nouveaux objets par apprentissage sur la base de données.

Les étapes de vérification et de détection de nouveaux objets sont en réalité très similaires. La seule différence réside dans le fait que, dans la détection de nouveaux objets, on doit générer des hypothèses d'objets, c'est à dire des objets qui sont potentiellement présents dans l'image. Ces hypothèses d'objets seront ensuite traitées comme s'il s'agissait d'une vérification d'objets

de la base de données.

La façon la plus simple de générer des hypothèses d'objets est de réaliser une segmentation des images et considérer les régions ainsi obtenues comme des candidats. Afin de ne pas dépendre de la qualité de la segmentation (sur- ou sous-segmentation), des segmentations à plusieurs échelles sont utilisées.

Jusqu'ici, tout est générique et il n'y a pas de brique dans cette chaîne qui soit spécifique à un type d'objet.

5.3 Éléments focaux

La façon d'introduire une description des objets d'intérêt dans la chaîne passe par une description d'un niveau sémantique assez élevé. Nous avons décidé d'utiliser une description qui s'appuie sur les descripteurs précédemment extraits.

Par exemple, un bâtiment peut être décrit par « région avec des frontières linéaires sans végétation, projetant une ombre et ayant un angle droit entre le sol et les murs ». Cette description peut être combinée avec des descripteurs image comme les indices de végétation, les détecteurs de lignes sur les images SAR, les détecteurs d'ombres, etc. pour donner la représentation graphique de la figure 5.2, qu'on appelle éléments focaux.

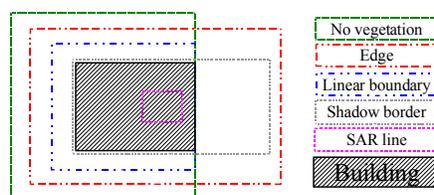


FIGURE 5.2: Exemple d'éléments focaux pour l'objet "bâtiment".

On constate donc la simplicité de la création de modèles d'objets. Cependant, une description aussi simple, ne permet pas de détecter des objets avec une précision suffisante. Ceci est dû au fait que la description linguistique, comme nous l'avons montré dans la section 4.3, est souvent entachée d'imprécision. Par exemple, dans le cas des bâtiments, dire qu'ils ne contiennent pas de végétation n'est pas toujours juste.

Il a fallu donc se doter d'outils mathématiques pour représenter cette imprécision, mais aussi l'incertitude.

5.4 Utilisation de la théorie des croyances

Nous avons choisi d'utiliser la théorie des croyances [32], ou modèle de Dempster-Shafer. Dans ce cadre, on attribue une certaine *croyance* aux hypothèses possibles. L'ensemble de toutes les hypothèses possibles est appelé *cadre de discernement*, Θ . On appelle $\mathbb{P}(\Theta)$ l'ensemble de tous les sous-ensembles de Θ . Ces sous-ensembles sont appelés *propositions*.

La quantité de *preuve* qu'une source d'information apporte à une proposition est représentée par une *fonction de masse*, FM. Une FM m satisfait les propriétés suivantes :

$$m : \mathbb{P}(\Theta) \rightarrow [0, 1], \sum_{A_i \subseteq \Theta} m(A_i) = 1, m(\emptyset) = 0. \quad (5.1)$$

Les sous-ensembles auxquels une source apporte de la croyance sont les *ensembles focaux* de la source.

Dans notre cas, chaque ensemble focal correspond à chaque hypothèse pour laquelle un descripteur apporte de la croyance. Ainsi, par exemple, pour le descripteur *végétation*, nous définissons 3 ensembles focaux : *présence de végétation*, *absence de végétation* et *incertitude*.

Il faut ensuite définir les fonctions de masse pour chaque descripteur. Pour la prise de décision sur, par exemple, l'existence d'un bâtiment, sera déterminé par fusion d'informations concernant tous les descripteurs associés. De façon concrète, cette fusion combine les fonctions de masse à l'aide de la règle orthogonale de Dempster-Shafer :

$$m_{12}(P) = m_1 \oplus m_2(P) = \frac{1}{1 - \kappa} \sum_{A \cap B = P} m_1(A)m_2(B) \quad (5.2)$$

avec $\kappa = \sum_{A \cap B = \emptyset} m_1(A)m_2(B)$ représentant le degré de conflit entre les sources. À partir de cette fonction de masse associée à l'hypothèse qui nous intéresse (bâtiment, par exemple), on peut calculer la croyance (notée *Bel* pour *belief*) et la plausibilité qui constituent respectivement les bornes supérieure et inférieure d'un intervalle de probabilité :

$$\text{Bel}(P) = \sum_{A \subseteq P} m(A). \quad (5.3)$$

$$\text{Pl}(P) = \sum_{A | A \cap P \neq \emptyset} m(A). \quad (5.4)$$

Nous attribuons à chaque objet un score qui est la moyenne de ces 2 grandeurs, même si autres choix seraient possibles. La comparaison de ce score à un seuil permet la prise de décision.

La représentation l'imprécision et l'incertitude liée aux descripteurs et aux éléments focaux est faite à l'aide des fonctions de masse. Nous avons besoin, pour chaque élément focal, de 3 fonctions de masse : celle décrivant l'hypothèse positive (ex. présence de végétation), celle décrivant l'hypothèse négative (ex. absence de végétation) et celle représentant l'incertitude. Le support de ces fonctions est la valeur numérique du descripteur associé (ex. indice de végétation).

La figure 5.3 montre un modèle de fonctions de masse pour un descripteur générique.

Ces fonctions sont complètement déterminées par le vecteur de paramètres (a, b, c, d) . Pour chaque descripteur, ces paramètres ont été optimisés en utilisant des données réelles et une fonction de coût combinant la croyance et la plausibilité.

Dans le cas de la mise à jour d'une base de données, la vérification se fait en utilisant les fonctions de masse ainsi optimisées. La détection de nouveaux objets peut bénéficier d'une nouvelle optimisation des paramètres en utilisant les objets vérifiés dans la base de données.

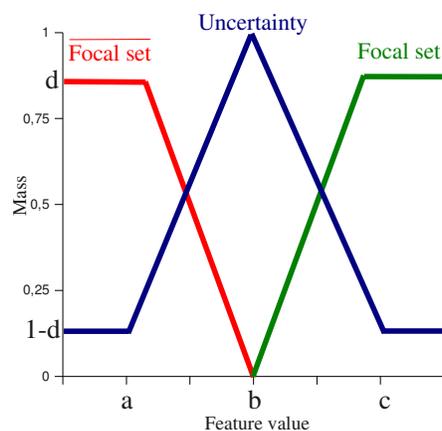


FIGURE 5.3: Modèle de fonctions de masse.

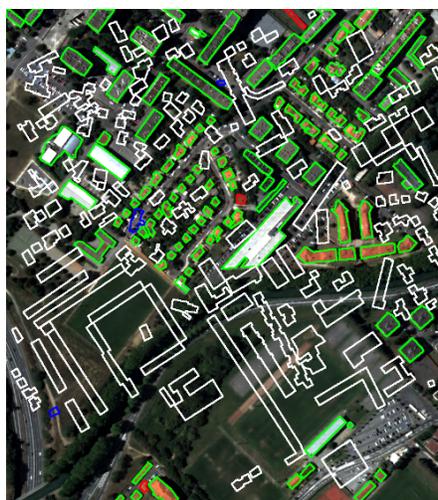


FIGURE 5.4: Exemple de vérification d'une base de données de bâtiments.

5.5 Résultats

Nous montrons ici un exemple de résultats obtenus sur un jeu de données composé d'une image de type Pléiades (multispectral 4 bandes à 70 cm.), d'une image de type Cosmo Skymed (SAR bande X à 1 m.) et d'une base de données de bâtiments où ces derniers sont représentés par leur emprise au sol.

La figure 5.4 montre les résultats obtenus en vérification. On a introduit des faux bâtiments afin de mesurer les faux positifs et les vrais négatifs. Le code de couleurs est le suivant :

- en vert : les vrais positifs, c'est à dire, les vrais bâtiments détectés correctement ;
- en blanc : les vrais négatifs, c'est à dire, les faux bâtiments correctement non détectés ;
- en rouge : les faux négatifs, c'est à dire, les vrais bâtiments non détectés ;
- en bleu : les faux positifs, c'est à dire, les faux bâtiments détectés à tort.

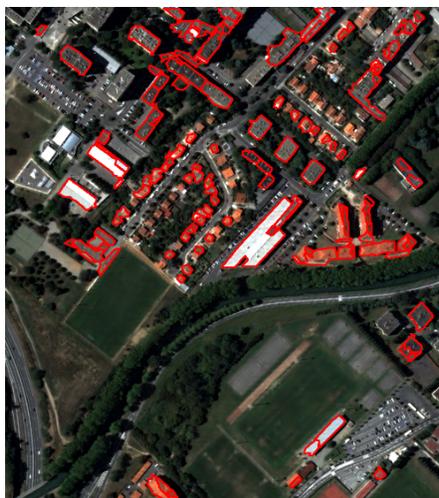


FIGURE 5.5: Exemple de détection de bâtiments.

La figure 5.5 montre un résultat de détection de bâtiments sans utilisation de base de données.

6 Perspectives de recherche

Comme il a pu être constaté tout au long des chapitres précédents, mes travaux de recherche ont été guidés par les besoins du Cnes en ce qui concerne le traitement des images de ses capteurs, que ce soit en valorisation quand ils existaient déjà, ou que ce soit en préparation à l'utilisation.

Dès 2008, l'équipe du projet *Venus* nous a sollicité pour des supports ponctuels sur le traitement de séries multi-temporelles. Nous avons décidé de commencer un travail de fond sur le traitement de ce type de données, car, peu après *Venus*, le programme des *Sentinelles* de l'ESA doit être opérationnel.

La caractéristique particulière de ce type de données est d'avoir un échantillonnage temporel très fin – allant jusqu'à une acquisition tous les 2 ou 3 jours – tout en gardant une résolution spatiale élevée (de l'ordre de 10 m.) et une résolution spectrale très riche en optique (une dizaine de bandes spectrales sur le visible et l'infra-rouge).

L'arrivée des données fournies par le programme des *Sentinelles* permettra le développement de services et d'applications qui ne peuvent être mis en place que grâce à une revisite courte et systématique de la surface de la Terre.

Le défi principal posé par ces données est le passage à l'échelle des techniques et des services actuels. Ce passage à l'échelle devra s'opérer sur les aspects suivants :

1. volume de données à gérer pour un site donnée ;
2. emprise et diversité spatiale – géographique – des zones traitées ;
3. assurance qualité des produits livrés.

6.1 Les produits attendus

Étant donné le contexte de passage à l'échelle cité ci-dessus, un certain niveau de standardisation peut être nécessaire. Trois familles de produits permettent de couvrir un grand nombre de besoins :

1. Cartes d'occupation des sols
2. Cartes de variables climatiques essentielles (ECVs, GCOS Essential Climate Variables Matrix)
3. Produits d'alerte aux changements

Les cartes d'occupation des sols devront suivre des nomenclatures normalisées et qui vont au delà de ce qui est utilisé à l'heure actuelle. La norme ISO/DIS 19144-2 *Geographic information - Classification systems – Part 2: Land Cover Meta Language (LCML)* semble être un candidat approprié pour la mise en oeuvre de systèmes de classification opérationnels.

Pour ce qui concerne les ECVs, la difficulté majeure réside dans le choix des modèles et des algorithmes qui permettront une estimation robuste et avec la fréquence de mise à jour adéquate.

Enfin, les produits d'alerte aux changements n'ont pas beaucoup été mis en avant, mais ils seront intéressants à plusieurs titres :

- ils peuvent être générés de façon plus simple que les cartes d'occupation des sols – le problème étant plus simple – et peuvent apporter presque la même quantité d'information qu'une carte d'OS si l'OS précédente est connue.
- ils peuvent être obtenus avec des données moins résolues spatialement et spectralement, ce qui les rend intéressants en termes d'exigence de données [28] ; ils pourront donc être générés de façon plus fréquente

En revanche, les algorithmes permettant d'exploiter ces potentialités n'existent pas ou sont moins mûrs que les techniques de classification ou les méthodes d'estimation de ECVs.

6.2 Les contraintes

La disponibilité des données (images) et la définition de produits attendus ne doit pas faire oublier les contraintes auxquelles un système opérationnel va être confronté.

- Il faudra des produits validés à échelle régionale, voire globale, ce qui demandera, au moins dans un premier temps, de travailler sur des nomenclatures simplifiées.
- Le manque de données de référence, vérité terrain, etc. – nécessaires pour la mise au point de méthodes, les apprentissages et les validations – demandera de travailler presque exclusivement avec des approches non-supervisées. Ces approches devront tirer un maximum d'informations sur les connaissances *a priori* disponibles sur les phénomènes et les processus observés. Ces connaissances peuvent être issues de la modélisation physique, mais aussi de connaissances qualitatives et des règles de décision.
- Même si la capacité théorique d'acquisition d'images sera importante, il ne faut pas négliger les limitations d'accès à la ressource (couverture nuageuse, temps de mise à disposition de la donnée, pannes éventuelles). Ceci est un élément crucial à prendre en compte dans un système opérationnel.
- En rapport avec le point précédent, la cadence d'acquisition des données fera émerger des besoins en termes d'accès quasi-temps-réel à l'information. Dans ce cadre, les approches simples et rapides seront à privilégier. Il faudra aussi concevoir des traitements parallélisables pour bénéficier des améliorations des capacités des machines de calcul (multi-processeurs, GPUs, etc.)

6.3 Les solutions :

Pour répondre à l'attente des utilisateurs en termes de produits, de leur qualité et de leur fréquence de mise à jour, il y a principalement 2 pistes à poursuivre : l'utilisation d'un maximum de capteurs disponibles et l'utilisation efficace d'informations exogènes aux données de télédétection.

6.3.1 Modélisation de comportements temporels

La dimension temporelle doit être traitée de façon différente à celle de la dimension spatiale des images. Le suivi des dynamiques temporelles est souvent la seule façon d'accéder à certaines informations. Les techniques de description et manipulation de séries temporelles doivent être capables de s'affranchir d'abord des problèmes d'échantillonnage irrégulier : même si le satellite a un cycle régulier, les aléas de la couverture nuageuse rendent des images inutilisables. Un autre problème qu'il faut prendre en compte est celui des comportements temporels décalés ou distordus. C'est dans ce cadre que nous avons proposé la thèse de François Petitjean.

6.3.2 Le multi-capteur

Les algorithmes capables d'ingérer des données issues de capteurs différents permettront d'augmenter la fréquence de mise à jour des produits. Ils doivent aussi permettre d'améliorer la qualité de certains de ces produits ou même d'accéder à des grandeurs autrement inaccessibles, mais nous n'abordons pas ces aspects ici. En effet, dans un cadre opérationnel, il faut envisager les traitements multi-capteurs comme un mode dégradé quand la donnée nominale n'est pas disponible, et non pas comme une exigence pour la livraison d'un produit.

Le terme *multi-capteur* est vague. Il peut être décliné à, au moins, 3 niveaux.

1. Capteurs de même classe : Ven μ s, Sentinelle-2, LDCM. On peut dans ces cas utiliser, avec peu d'adaptation, les mêmes chaînes de traitement pour toutes les données.
2. Capteurs avec des résolutions différentes. À l'horizon sur lequel nous nous plaçons, les exemples sont Sentinelle-2 avec Proba-V ou Sentinelle-3. On utilise ici la complémentarité entre haute résolution spatiale/spectrale et très haute résolution temporelle pour assurer une mise à jour très fréquente des informations. Des techniques spécifiques pour la fusion multi-résolution doivent être développées.
3. Optique/SAR : typiquement Sentinelle-1 et Sentinelle-2. On essaiera ici de bénéficier de l'aspect tous temps du SAR pour assurer une production systématique d'informations. Les propriétés physiques des 2 types de capteurs étant complètement différentes, un niveau d'indirection devra être ajouté aux algorithmes (passage par des grandeurs intermédiaires comme les classes ou les changements).

6.3.3 Les informations exogènes

L'utilisation de données exogènes dans la génération de produits à base d'imagerie satellite est classique (MNT, carte de référence, donnée météo, etc.). Le terme information exogène va

au delà de cette notion de donnée ancillaire en ajoutant 2 nouvelles composantes : les modèles physiques et les connaissances du domaine.

6.3.3.1 Les modèles physiques

Leur utilisation pour l'assimilation des données de télédétection ne peut être envisagée que pour l'estimation de variables physiques. Souvent, les modèles ne peuvent pas être envisagés dans des contextes opérationnels à cause de leur coût de calcul ou de leur besoin de paramètres non disponibles.

Nous envisageons ici leur utilisation comme outil de sélection de traitements. En effet, des modèles simples – spatialisables – peuvent être utilisés pour décider quels sont les traitements à mettre en place par la suite. Un exemple de ce type d'approche consiste à se servir de modèles de croissance de la végétation pour réaliser un premier niveau de classification et pouvoir ensuite utiliser des classifieurs spécialisés sur certaines familles de classes.

Les modèles à prendre peuvent être des types suivants :

- TSVA : transfert sol-végétation-atmosphère ;
- croissance/fonctionnement de la végétation ;
- réflectance (feuilles, sol) ;
- transfert radiatif ;
- capteur : bandes spectrales, fonction de transfert de modulation, etc.

Nous avons commencé à travailler sur ces aspects lors du séjour de Germain Forestier au CNES [11, 10]. Une application de cette approche aux séries multi-temporelles a été développée pour comparer les performances de classification entre Formosat-2 et les futurs *Venùs* et Sentinelle-2 [15].

6.3.3.2 Les connaissances du domaine

Il s'agit ici d'utiliser des informations décrivant les phénomènes et les processus d'intérêt, mais sans passer par une modélisation physique ou mathématique. Sans rentrer, en tout cas dans un premier temps, dans les domaines de la modélisation qualitative ou de la *soft data assimilation*, on peut envisager d'utiliser des terminologies et des règles décrivant, par exemple, les pratiques agricoles.

Comme dans le cas des modèles physiques, ce type de méthode est envisagée comme étape préalable à la classification, mais elle peut être aussi utilisée pour valider/invalidier des résultats de classification.

6.4 La classification

Les méthodes de classification, notamment celles qui sont développées au CESBIO depuis des années sont robustes et performantes. Des nouveaux développements sont en cours et visent principalement à mettre en oeuvre des techniques de reconnaissance de classes (passage entre l'étiquette non thématique issue d'une classification non-supervisée vers l'identification du nom thématique de la classe).

Nous nous intéressons ici à la stratégie à mettre en oeuvre pour utiliser ces méthodes dans un environnement opérationnel, dont les contraintes ont été décrites plus haut. Cette stratégie est composée des modules suivants :

1. Choix de la donnée à fournir à l'algorithme de classification : s'agit-il de réflectances, d'indices, de variables estimées ?
2. Choix de la fenêtre temporelle à utiliser pour la génération de la classification : parmi les images disponibles à un instant donné, quelles sont celles qui peuvent apporter le plus d'information ? Comment aborder le compromis entre quantité d'information et volume de données à traiter ?
3. Comment pondérer ou ajuster les résultats d'une classification en utilisant des données de référence ?
4. Comment décomposer le problème de génération d'une carte d'occupation des sols en sous-problèmes plus facilement traitables par les algorithmes de classification ?

6.5 La détection de changements

Pour certaines applications, il peut être très utile d'avoir un produit qui, même sans décrire les classes d'occupation des sols, met en évidence les évolutions des surfaces. Ces évolutions peuvent être analysées au niveau des réflectances observées, au niveau d'indices dérivées ou par le biais d'autres grandeurs d'intérêt.

Ces produits d'analyse d'évolution des surfaces peuvent être contraints par une classification existante. Par exemple, si l'on s'intéresse à l'étude du travail du sol, la détection de changements peut être limitée à la classe sol nu issue d'une classification.

Inversement, un produit d'alerte aux changements peut être fourni en entrée d'une classification pour discriminer 2 classes très similaires, mais avec des évolutions temporelles légèrement différentes.

Les travaux existants en détection de changements sont souvent focalisés sur les problèmes à 2 dates avec un changement abrupt entre les 2. Il faudra ici développer des techniques pouvant profiter d'un grand nombre d'images et éventuellement les étendre au cas multi-capteur (notamment multi-échelles).

Deuxième partie

Curriculum Vitae

7 Coursus professionnel

7.1 Formation

- 9/2000. Doctorat en Traitement du Signal et Télécommunications de l'Université de Rennes-1 : *Étude des signatures radar à la surface de l'océan de la topographie sous-marine.*
- 9/1997. DEA Signal Télécommunications Image et Radar de l'Université de Rennes 1.
- 9/1997. Ingénieur Télécom, ENST Bretagne. Spécialité Signal et Images.
- 9/1997 Ingénieur Télécom, ETSET Barcelone, Université Polytechnique de Catalogne. Spécialité Communications.

7.2 Expérience professionnelle

- Depuis 3/2010 : Chercheur au Centre d'études spatiales de la biosphère, CESBIO
- 12/2001-3/2010 : Ingénieur au Centre national d'études spatiales, CNES.
 - Animateur du Volet méthodologique du *Programme d'accompagnement ORFEO.*
 - Chef du projet ORFEO Toolbox.
 - Responsable du Dossier d'axe technique *OT4 : Méthodes d'extraction d'information des images.*
 - Chargé de la R&D en extraction d'information des images de télédétection pour des applications opérationnelles et multitemporelles :
 - détection de changements ;
 - reconnaissance d'objets ;
 - mise en correspondance d'images multi-capteurs ;
 - mesures de similarité.
 - Chef de projet pour la Charte Internationale Espace et Catastrophes Majeures.
- 10/2000 - 12/2001 : Chercheur sous contrat au CNES (Centre de Toulouse). Sujet d'étude : Développement de techniques de fusion d'imagerie spatiale (optique et radar) dans le cadre d'études multitemporelles appliquées aux risques naturels.

8 Liste des publications

8.1 Chapitres de livre

1. Andreas Wimmer, Iris Lingenfelder, Charles Beumier, **Jordi Inglada**, Simon J. Caseley : Feature Recognition Techniques. Dans B. Jasani, M. Pesaresi, S. Schneiderbauer and G. Zeug (Eds.), Remote Sensing from Space – Supporting International Peace and Security. Springer, 2009. Pages 105-118.
2. Ahed Alboody, Florence Sèdes, **Jordi Inglada** : Enriching The Qualitative Spatial Reasoning System RCC8. Dans : Qualitative Spatio-Temporal Representation and Reasoning : Trends and Future Directions. Shyamanta M. Hazarika (Eds.), Information Science Reference, août 2010.

8.2 Articles de revue à comité de lecture

1. **Jordi Inglada**, René Garello : On Rewriting the Imaging Mechanism of Underwater Bottom Topography by Synthetic Aperture Radar as a Volterra Series Expansion. IEEE Journal of Oceanic Engineering, vol. 27, no. 3, pp. 665-674, July 2002.
2. **Jordi Inglada**, Jean-Marc Le Caillec, René Garello : Inversion of Imaging Mechanisms by Regularization of Inverse Volterra Models. Signal Processing, vol. 84, no. 6, p. 1021–1034, 2004.
3. **Jordi Inglada**, Alain Giros : On the possibility of automatic multi-sensor image registration. IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 10, p. 2104–2120, October 2004.
4. **Jordi Inglada**, Jean-Claude Souyris, Caroline Henry, Céline Tison : Incoherent SAR polarimetric analysis over point targets. IEEE Geoscience and Remote Sensing Letters, vol. 3, no. 2, p. 246–249, April 2006.
5. **Jordi Inglada**, Vincent Muron, Damien Pichard, Thomas Feuvrier : Analysis of Artifacts in Subpixel Remote Sensing Image Registration, IEEE Transactions on Geoscience and Remote Sensing, vol. 45, no. 1, p. 254–264, January 2007.
6. **Jordi Inglada**, Grégoire Mercier : A Statistical Similarity Measure for Change Detection in Multitemporal SAR Images and its Extension to Multiscale Change Analysis, IEEE

- Transactions on Geoscience and Remote Sensing, Volume : 45, Issue : 5, Part 2 May 2007, pages 1432-1445.
7. Florent Chatelain, Jean-Yves Tournet, **Jordi Inglada**, André Ferrari : Bivariate Gamma Distributions for Image Registration and Change Detection, IEEE Transactions on Image Processing, Volume 16, Issue 7, July 2007, pages 1796-1806.
 8. **Jordi Inglada** : Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. ISPRS Journal of Photogrammetry and Remote Sensing. ISPRS Journal of Photogrammetry and Remote Sensing Volume 62, Issue 3, August 2007, Pages 236-248.
 9. Florent Chatelain, Jean-Yves Tournet, **Jordi Inglada**, Change Detection in Multisensor SAR Images Using Multivariate Gamma Distributions, IEEE Transactions on Image Processing. Volume : 17, Issue : 3, March 2008.
 10. **Jordi Inglada**, Julien Michel : Qualitative Spatial Reasoning for High-Resolution Remote Sensing Image Analysis. IEEE Transactions on Geoscience and Remote Sensing, Volume 47, Issue 2, February 2009, Pages 599-612.
 11. Tarek Habib, **Jordi Inglada**, Grégoire Mercier, Jocelyn Chanussot, Support Vector Reduction in SVM Algorithm for Abrupt Change Detection in Remote Sensing. IEEE Geoscience and Remote Sensing Letters, Volume 6, Issue 3, July 2009, Pages 606-610.
 12. G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco, C. Thiel, **J. Inglada**, E. Christophe, J. Chanussot and P. Gamba : Decision Fusion for the Classification of Hyperspectral Data : Outcome of the 2008 GRS-S Data Fusion Contest. IEEE Transactions on Geoscience and Remote Sensing, Volume 47, Issue 11, November 2009, Pages :3857-3865.
 13. E. Christophe, J. Michel, and **J. Inglada** : Remote sensing processing : from multicore to GPU, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. Accepted. 2010.
 14. V. Poulain, **J. Inglada**, M. Spigai, J-Y. Tournet and P. Marthon : High resolution optical and SAR image fusion for building database updating. IEEE Transactions on Geoscience and Remote Sensing, Accepted, 2011.
 15. María Carolina Vanegas, Isabelle Bloch, **Jordi Inglada** : Alignment and parallelism for the description of objects in high resolution remote sensing images, Computer Vision and Image Understanding. Soumis.
 16. François Petitjean, **Jordi Inglada**, Pierre Gançarski : Satellite Image Time Series Analysis under Time Warping, IEEE Transactions on Geoscience and Remote Sensing. Soumis.

8.3 Communications dans des colloques avec comité de lecture

1. **Jordi Inglada**, René Garello, Alfred Ramamonjariisoa, Hamed Hahhi, Martine Deveaux : Comparison of two models for the imaging mechanism of underwater bottom topography by synthetic aperture radar. Proceedings of the International Geoscience And Remote Sensing Symposium 1998. Seattle, Washington, USA.

2. **Jordi Inglada**, René Garello : Volterra Model for the Imaging Mechanism of Ocean Surface Currents by Synthetic Aperture Radar. Proceedings of the International Geoscience And Remote Sensing Symposium 1999. Hamburg, Germany.
3. **Jordi Inglada**, René Garello : Depth Estimation and 3D Topography Reconstruction from SAR Images Showing Underwater Bottom Topography Signatures. Proceedings of the International Geoscience And Remote Sensing Symposium 1999. Hamburg, Germany.
4. **Jordi Inglada**, René Garello : Volterra Model for the Study of Wave-Current Interaction as Described by the Action Balance Equation Proceedings of the 9th (1999) International Offshore and Polar Engineering Conference and Exhibition, ISOPE-99. Brest, France.
5. **Jordi Inglada**, René Garello : Underwater Bottom Topography Estimation from SAR Images by Regularization of the Inverse Imaging Mechanism. Proceedings of the International Geoscience And Remote Sensing Symposium 2000. Honolulu, HI, USA.
6. **Jordi Inglada**, René Garello : Nonlinear Model for the Imaging Mechanism of Ocean Surface Currents by SAR Using a Quadratic Source Function. Proceedings of the International Geoscience And Remote Sensing Symposium 2000. Honolulu, HI, USA.
7. **Jordi Inglada**, René Garello : Model for the Imaging Mechanism of Ocean Surface Currents by SAR. Application to Underwater Bottom Topography Estimation. Proceedings of the Second International Symposium on Physics in Signal and Image Processing. Marseille, France. 2001.
8. Luc Pigeon, **Jordi Inglada**, Basel Solaiman : Genetic algorithms for multi-agent fusion system learning. Proceedings of the SPIE 15th Annual International Symposium on Aerospace/Defense Sensing, Simulation and Controls, AeroSense, 16-20 April 2001, Orlando, FL, USA.
9. **Jordi Inglada**, Frédéric Adragna : Automatic Multi-Sensor Image Registration Algorithm by Edge Matching Using Genetic Algorithms. Proceedings of the International Geoscience And Remote Sensing Symposium 2001. Sydney, Australia.
10. **Jordi Inglada**, Francesco Sarti, Philippe Durand, Frédéric Adragna, Jean-Luc Bessis : The emergency of El Salvador in the frame of the International Charter Space and Major Disasters, X Simposio Brasileiro do Sensoramento Remoto, Foz do Iguazu, Brasil, 2001.
11. Francesco Sarti, **Jordi Inglada**, Robert Landry, Thierry Pultz : Risk management using remote sensing data : moving from scientific to operational applications, X Simposio Brasileiro do Sensoramento Remoto, Foz do Iguazu, Brasil, 2001.
12. **Jordi Inglada**, Frédéric Adragna : Blind Source Separation Applied to Multitemporal Series of Differential SAR Interferograms. Proceedings of the International Geoscience And Remote Sensing Symposium 2002. Toronto, Canada.
13. **Jordi Inglada** : Similarity Measures for Multisensor Remote Sensing Images. Proceedings of the International Geoscience And Remote Sensing Symposium 2002. Toronto, Canada.
14. Frédéric Adragna, Mihai Datcu, Irena Hajsek, **Jordi Inglada**, Kostas Papathanassiou, David Petit, Marco Quartulli, Jean-Claude Souyris : Information extraction from high

- resolution SAR data. Proceedings of European Conference on Signal Processing, EUSIPCO'02, September 2002, Toulouse, France.
15. **Jordi Inglada** : Change Detection on SAR Images by Using a Parametric Estimation of the Kullback-Leibler Divergence. Proceedings of the International Geoscience And Remote Sensing Symposium 2003. Toulouse, France.
 16. Francesco Sarti, **Jordi Inglada**, Jean-Luc Bessis : The two Emergencies of El Salvador in the Frame of the International Charter Espace and Major Disasters. Proceedings of the International Geoscience And Remote Sensing Symposium 2003. Toulouse, France.
 17. Alain Arnaud, Javier Duro, **Jordi Inglada**, Michael Eineder, Nico Adam and Ramon Hansen : ASAR/ERS interferometry over stable points. Proceedings of the International Geoscience And Remote Sensing Symposium 2003. Toulouse, France.
 18. **Jordi Inglada**, Jean-Claude Favard, Hervé Yesou, Stephen Clandillon, Claude Bestault : Lava flow Mapping during the Nyiragongo January 2002 eruption over the City of Goma in the Frame of the International Charter Espace and Major Disasters. Proceedings of the International Geoscience And Remote Sensing Symposium 2003. Toulouse, France.
 19. Hervé Yesou, Stephen Clandillon, Bernard Allenbach, Claude Bestault, Paul de Fraipont, **Jordi Inglada**, Jean-Claude Favard : A constellation of Advantages with SPOT SWIR and VHR SPOT 5 data for Flood Extent Mapping during the September 2002 Gard Event (France). Proceedings of the International Geoscience And Remote Sensing Symposium 2003. Toulouse, France.
 20. **Jordi Inglada**, Alain Giros : On the real capabilities of remote sensing for disaster management - Feedback from real cases. Proceedings of the International Geoscience And Remote Sensing Symposium 2004. Anchorage, Alaska, USA.
 21. **Jordi Inglada**, Alain Giros : Automatic man made object recognition in high resolution remote sensing images. Proceedings of the International Geoscience And Remote Sensing Symposium 2004. Anchorage, Alaska, USA.
 22. H el ene Vadon and **Jordi Inglada** : CNES research and development, and available software, in the framework of space images based risk and disaster management. Proceedings of the First International Symposium on Geo-information for Disaster Management' (Gi4DM), Delft, The Netherlands, March 21-23, 2005
 23. **Jordi Inglada** : Use of pre-conscious vision and geometric characterizations for automatic man-made object recognition. Proceedings of the International Geoscience And Remote Sensing Symposium 2005. Seoul, Korea.
 24. **Jordi Inglada**, H el ene Vadon : Fine registration of Spot5 and Envisat/ASAR images and ortho-image production : a fully automatic approach. Proceedings of the International Geoscience And Remote Sensing Symposium 2005. Seoul, Korea.
 25. Florent Chatelain, Jean-Yves Tournet, **Jordi Inglada**, Andr e Ferrari : Parameter Estimation for Multivariate Gamma Distributions. European Signal Processing Conference - EUSIPCO. Firenze (Italy), 4-8 April 2006.

26. **Jordi Inglada**, Vincent Muron, Damien Pichard, Thomas Feuvrier : Analysis of Artefacts in Sub-Pixel Remote Sensing Image Registration. ISPRS Commission 1 Symposium. Marne-la-Vallée (France), 3-6 July 2006.
27. David Petit, Guillaume Oller, **Jordi Inglada** : On the Use of SAR and Optical Images Combination. ISPRS Commission 1 Symposium. Marne-la-Vallée (France), 3-6 July 2006.
28. **Jordi Inglada**, Grégoire Mercier : The Multiscale Change Profile : a Statistical Similarity Measure for Change Detection in Multitemporal SAR Images. IEEE International Geoscience and Remote Sensing Symposium - IGARSS, Denver - Colorado (USA), 31 July - 4 August 2006.
29. Grégoire Mercier, Stéphane Derrode, Wojciech Pieczynski, Jean-Marie Nicolas, Anne Joannic-Chardin, **Jordi Inglada** : Copula-based Stochastic Kernels for Abrupt Change Detection. IEEE International Geoscience and Remote Sensing Symposium - IGARSS, Denver - Colorado (USA), 31 July - 4 August 2006.
30. Vito Alberga, Mahamadou Idrissa, Vinciane Lacroix, **Jordi Inglada** : Performance Estimation of Similarity Measures of Multi-Sensor Images for Change Detection Applications. International Workshop on the Analysis of Multi-temporal Remote Sensing Images, 2007. MultiTemp 2007. 18-20 July 2007.
31. Vito Alberga, Mahamadou Idrissa, Vinciane Lacroix, **Jordi Inglada** : Comparison of similarity measures of multi-sensor images for change detection applications. IEEE International Geoscience and Remote Sensing Symposium - IGARSS, Barcelona, Spain, 23-27 July 2007.
32. **Jordi Inglada**, Julien Michel : Spatial Reasoning and Multiscale Segmentation for Object Recognition in HR Optical Remote Sensing Images. IEEE International Geoscience and Remote Sensing Symposium - IGARSS, Barcelona, Spain, 23-27 July 2007.
33. Tarek Habib, Jocelyn Chanussot, **Jordi Inglada**, Grégoire Mercier : Abrupt change detection on multitemporal remote sensing images. A statistical overview of methodologies applied on real cases. IEEE International Geoscience and Remote Sensing Symposium - IGARSS, Barcelona, Spain, 23-27 July 2007.
34. Selma Cherchali, **Jordi Inglada** : Rapid mapping Service in risk and disaster management in the frame of PREVIEW project. 2007 CIG/ISPRS Joint Conference (May 22 - 25, 2007, Toronto, Ontario, Canada).
35. Emmanuel Christophe, **Jordi Inglada** : Robust road extraction for high resolution satellite images. IEEE International Conference on Image Processing. San Antonio Texas September 16-19, 2007.
36. Jan Dirk Wegner, **Jordi Inglada**, Céline Tison : Image analysis of fused SAR and optical images deploying open source software library OTB. ISPRS Hannover Workshop 2007. High-Resolution Earth Imaging for Geospatial Information.
37. Emmanuel Christophe, **Jordi Inglada**, Alain Giros : ORFEO Toolbox : a complete solution for mapping from high resolution satellite images, XXI Congress, The International Society for Photogrammetry and Remote Sensing, ISPRS'08. Beijing, China, July 2008.

38. **Jordi Inglada**, Emmanuel Christophe, H el ene de Boissezon : From satellite images to operational applications : generic tools for specific user needs. Space Appli 2008. Toulouse, France April 2008.
39. **Jordi Inglada**, Emmanuel Christophe, Sylvie Marzocchi-Polizzi, H el ene de Boissezon, Beno t Boissin. ORFEO Toolbox : an open source library of image processing algorithms for SAR and optical high resolution images. ESA-EUSC 2008 : Image Information Mining : pursuing automation of geospatial intelligence for environment and security. Torrejon de Ardoz, Spain, February 2008.
40. **Jordi Inglada**, Julien Michel, Thomas Feuvrier : A generic framework for disparity map estimation between multi-sensor remote sensing images. International Geoscience and Remote Sensing Symposium, IGARSS 2008.
41. Julien Michel, **Jordi Inglada** : Multi-scale segmentation and optimized computation of spatial reasoning graphs for object detection in remote sensing images. International Geoscience and Remote Sensing Symposium, IGARSS 2008.
42. Vincent Poulain, **Jordi Inglada**, Marc Spigai : High resolution remote sensing image analysis with exogenous data : a generic framework. International Geoscience and Remote Sensing Symposium, IGARSS 2008.
43. Tarek Habib, **Jordi Inglada**, Gr egoire Mercier, Jocelyn Chanussot : Speeding up support vector machine (SVM) image classification by a kernel series expansion. International Conference on Image Processing, ICIP 2008.
44. Tarek Habib, **Jordi Inglada**, Gr egoire Mercier, Jocelyn Chanussot : On the use of a new additive kernel for change detection using SVM. International Geoscience and Remote Sensing Symposium, IGARSS 2008.
45. Gr egoire Mercier, **Jordi Inglada** : Change detection with misregistration errors. International Geoscience and Remote Sensing Symposium, IGARSS 2008.
46. Tarek Habib, **Jordi Inglada**, Gr egoire Mercier, Jocelyn Chanussot : Assessment of feature selection techniques for support vector machine classification of satellite imagery. International Geoscience and Remote Sensing Symposium, IGARSS 2008.
47. Ahed Alboody, **Jordi Inglada**, Florence S edes : Post-classification and spatial reasoning : a new approach to change detection for updating GIS databasis. Information and Communication Technologies : From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on.
48. Ahed Alboody, **Jordi Inglada**, Florence S edes : Enriching the Spatial Reasoning System RCC8. ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM GIS 2008.
49. Ahed Alboody, **Jordi Inglada**, Florence S edes : Analysis of satellite imagery for updating spatio-temporal databases : Application to GIS and CLC. International Conference on Geographic Information Systems, ICGIS 2008.
50. Jan Wegner, C eline Tison, **Jordi Inglada** : Automatic Fusion of SAR and Optical Imagery based on Line Features. 7th European Conference on Synthetic Aperture Radar, EUSAR 2008.

51. Yves Auda, Danièle Ducrot, Jean-Philippe Gastellu, Olivier Hagolle, Philippe Maison-grande, **Jordi Inglada** : Enjeux et perspectives de la télédétection multitemporelle à Très Haute Résolution Spatiale. Journées Scientifiques de l'AUF 2008.
52. S. Aksoy, B. Ozdemir, S. Eckert, F. Kayitakire, M. Pesaresi, O. Aytekin, C. C. Borel, J. Cech, E. Christophe, S. Duzgun, A. Erener, K. Ertugay, E. Hussain, **J. Inglada**, S. Lefevre, O. Ok, D. K. San, R. Sara, J. Shan, J. Soman, I. Ulusoy, and R. Witz : Performance evaluation of building detection and digital surface model extraction algorithms : Outcomes of the PRRS 2008 Algorithm Performance Contest, in IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS 2008), Tampa, FL, USA, dec 2008.
53. A. Alboody, **J. Inglada** and F. Sèdes : Enriching the Spatial Reasoning System RCC8, In 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS 2008), The SIGSPATIAL Special, Vol. 1, n. 1, PP.14-20, March 2009.
54. A. Alboody, F. Sèdes and **J. Inglada** : Multi-Level Topological Relations of the Spatial Reasoning System RCC-8, In Proceedings of The First International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2009), IARIA Conferences - GlobeNet2009, IEEE Computer Society, PP. 13-21, March 2009.
55. A. Alboody, F. Sèdes and **J. Inglada** : Fuzzy Intersection and Difference Model for Topological Relations, In International Fuzzy Systems Association World Congress and Conference of the European Society for Fuzzy Logic and Technology (IFSA- EUSFLAT 2009), Calouste Gulbenkian Foundation, Lisbon, Portugal, 20-24 July 2009. Vol. IFSA-EUSFLAT 2009 Proceedings, IFSA-EUSFLAT 2009 Proceedings, PP. 1079-1084, 2009.
56. J. Osman, **J. Inglada** and E. Christophe : Interactive object segmentation in high resolution satellite images, IEEE International Geoscience and Remote Sensing Symposium, IGARSS'09, Cape Town, South Africa, jul 2009.
57. **J. Inglada** and E. Christophe : The Orfeo toolbox remote sensing image processing software, IEEE International Geoscience and Remote Sensing Symposium, IGARSS'09, Cape Town, South Africa, jul 2009.
58. E. Christophe and **J. Inglada** : Object counting in high resolution remote sensing images with OTB, IEEE International Geoscience and Remote Sensing Symposium, IGARSS'09, Cape Town, South Africa, jul 2009.
59. V. Poulain, **J. Inglada**, M. Spigai, J-Y. Tourneret, P. Marthon : High resolution remote sensing image analysis with exogenous data : a generic framework, IEEE International Geoscience and Remote Sensing Symposium, IGARSS'09, Cape Town, South Africa, jul 2009.
60. Otmane Lahlou, Julien Michel, Damien Pichard, **Jordi Inglada** : Assessment of interest points detection algorithms in OTB, , IEEE International Geoscience and Remote Sensing Symposium, IGARSS'09, Cape Town, South Africa, jul 2009.
61. Julien Michel, Thomas Feuvrier, **Jordi Inglada** : Reference algorithm implementations in OTB : Textbook cases, IEEE International Geoscience and Remote Sensing Symposium, IGARSS'09, Cape Town, South Africa, jul 2009.

62. Jean-Yves Tournet, Vincent Poulain, Marie Chabert, **Jordi Inglada** : Similarity measure between vector data bases and optical images for change detection, IEEE International Geoscience and Remote Sensing Symposium, IGARSS'09, Cape Town, South Africa, jul 2009.
63. Julien Michel, Cyrille Valladeau, **Jordi Inglada** : Focus pre-processing chain for object detection in high resolution remote sensing images, IEEE International Geoscience and Remote Sensing Symposium, IGARSS'09, Cape Town, South Africa, jul 2009.
64. Stéphane May, **Jordi Inglada** : Urban area detection and segmentation using OTB, IEEE International Geoscience and Remote Sensing Symposium, IGARSS'09, Cape Town, South Africa, jul 2009.
65. Maria Carolina Vanegas, Isabelle Bloch, **Jordi Inglada** : Fuzzy spatial relations for high resolution remote sensing image analysis : The case of « to go across », IEEE International Geoscience and Remote Sensing Symposium, IGARSS'09, Cape Town, South Africa, jul 2009.
66. Germain Forestier, Cédric Wemmert, Pierre Gancarski, **Jordi Inglada** : Mining Multiple Satellite Sensor Data Using Collaborative Clustering. Proceedings of the 2009 IEEE International Conference on Data Mining, Pages : 501-506, 2009.
67. Germain Forestier, **Jordi Inglada**, Cédric Wemmert, Pierre Gancarski : Mining spectral libraries to study sensors' discrimination ability. SPIE Europe Remote Sensing, Vol. 7478, 9 pages, Berlin, Germany, September 2009
68. E. Christophe and **J. Inglada** : Orfeo Toolbox : from satellite images to geographic information, Free and Open Source Software for Geospatial, FOSS4G 2009, Sydney, Australia, oct 2009.
69. A. Alboody, F. Sèdes and **J. Inglada** : Modeling Topological Relations between Uncertain Spatial Regions in Geo-spatial Databases : Uncertain Intersection and Difference Topological Model, In The Second International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA2010), Menuires, The Three Valleys, France, IEEE Computer Society, PP. 7-15, 11-16 Avril 2010.
70. E. Christophe, **J. Inglada**, and J. Maudlin : Crowd-sourcing satellite image analysis, IEEE International Geoscience and Remote Sensing Symposium, IGARSS'10, Honolulu, HI, USA, jul 2010.
71. E. Christophe, M. Grizonnet, J. Michel, **J. Inglada**, and T. Dhar : Bringing the raster processing algorithms of the Orfeo Toolbox Monteverdi in QGIS, Free and Open Source Software for Geospatial, FOSS4G 2010, Barcelona, Spain, sep 2010.
72. **Jordi Inglada**, Sébastien Garrigues : Land-cover maps from partially cloudy multi-temporal image series : optimal temporal sampling and cloud removal. IEEE International Geoscience and Remote Sensing Symposium, IGARSS'10, Honolulu, HI, USA, jul 2010.
73. Vincent Poulain, **Jordi Inglada**, Marc Spigai, Jean-Yves Tournet, Philippe Marthon : High resolution optical and SAR image fusion for road database updating. IEEE International Geoscience and Remote Sensing Symposium, IGARSS'10, Honolulu, HI, USA, jul 2010.

74. Julien Michel, Julien Malik, **Jordi Inglada** : Lazy yet efficient land-cover map generation for HR optical images. IEEE International Geoscience and Remote Sensing Symposium, IGARSS'10, Honolulu, HI, USA, jul 2010.
75. **Jordi Inglada**, Julien Michel : Using approximation and randomness to speed-up intensive linear filtering. IEEE International Geoscience and Remote Sensing Symposium, IGARSS'10, Honolulu, HI, USA, jul 2010.
76. María Carolina Vanegas, Isabelle Bloch, **Jordi Inglada** : Detection of aligned objects for high resolution image understanding. IEEE International Geoscience and Remote Sensing Symposium, IGARSS'10, Honolulu, HI, USA, jul 2010.
77. Marie Chabert, Jean-Yves Tournet, Vincent Poulain, **Jordi Inglada** : Logistic regression for detecting changes between databases and remote sensing images. IEEE International Geoscience and Remote Sensing Symposium, IGARSS'10, Honolulu, HI, USA, jul
78. **J. Inglada**, O. Hagolle, G. Dedieu : Assessment of the land cover classification accuracy of Venus image time series with respect to Formosat-2. RAQRS 2010, Valencia, Espagne.
79. G. Dedieu, O. Hagolle, M. Huc, M. Claverie, D. Courault, V. Debaecker, J.-F. Dejoux, V. Demarez, B. Duchemin, D. Ducrot, P. Ferrier, **J. Inglada**, A. Karnieli, L. Kergoat, C. Marais-Sicre, B. Mougenot, E. Mougin, V. Rivalland, Y. Yaniv : The VENUS mission and the benefits of Earth observation with high spatial and temporal resolutions. RAQRS 2010, Valencia, Espagne.
80. **J. Inglada**, J. F. Dejoux, C. Marais-Sicre, M. Huc, O. Hagolle, F. Baup, G. Dedieu, D. Ducrot : Use of dense time series of high resolution optical and radar images for change detection and land use classification. RAQRS 2010, Valencia, Espagne.
81. Julien Michel, **Jordi Inglada**, Julien Malik : Object based and geo-spatial image analysis : a semi-automatic pre-operational system. SPIE Remote Sensing Europe, Toulouse, September 2010.

9 Liste des mémoires et diplômes encadrés (2001-2011)

9.1 Stages

1. Peggy Malgat (Ingénieur) 2001. Estimation de décalages entre images SAR pour la mesure de déformations. École Nationale Supérieure des Télécommunications de Bretagne. Brest, France.
2. Jocelyn Briand (Ingénieur) 2002. Analyse temps-fréquence de données SAR à haute résolution. École Nationale Supérieure d'Électrotechnique, d'Électronique, d'Informatique, d'Hydraulique et des Télécommunications / Institut National Polytechnique de Toulouse. France.
3. Julien Pellier (Ingénieur) 2003. Mesures de similarité multi-capteurs pour les images de télédétection. École Nationale Supérieure de l'Électronique et de ses Applications. Cergy-Pontoise. France.
4. Olivier Caignart (DEA) 2003. Reconnaissance d'objets sur les images de télédétection à haute résolution. École Nationale Supérieure de Physique de Strasbourg / Université Louis Pasteur de Strasbourg. France.
5. Léonard Potier (Ingénieur) 2004. Caractérisations géométriques de haut niveau pour la reconnaissance d'objets. École Nationale Supérieure d'Ingénieurs Électriciens de Grenoble / Institut National Polytechnique de Grenoble. France.
6. François Bignalet-Cazalet (Ingénieur) 2004. Mise en correspondance d'images de télédétection sans modèle de prise de vue. École Nationale Supérieure de l'Aéronautique et de l'Espace. Toulouse. France.
7. Yann Le Meur (DEA) 2004. Segmentation de séries d'images multi-temporelles. École Nationale Supérieure d'Ingénieurs Électriciens de Grenoble / Institut National Polytechnique de Grenoble. France.
8. Aurélien Allan-Grandvalet (Ingénieur) 2005. Mise en correspondance d'images de télédétection sans modèle de prise de vue. École Supérieure de Chimie Physique Électronique de Lyon. France.
9. Jérôme Tagnères (Master Professionnel), 2005. Reconnaissance d'objets dans les images à haute résolution. Université Paul Sabatier - Toulouse 3.

10. Julien Michel (Master Recherche), 2006. Raisonement spatial pour l'analyse d'images à haute résolution. École Nationale Supérieure des Télécommunications de Bretagne. Brest, France.
11. Vincent Poulain (Master Recherche), 2007. Détection de changements entre images et bases de données vectorielles. Institut National Polytechnique de Toulouse. Toulouse, France.
12. Miarintsoa Ramanantsimiavona (Ingénieur), 2007. Intégration de modèles géométrique de capteurs satellitaires dans l'ORFEO Toolbox. ENSEEIHT. Toulouse, France.
13. Jan Dirk Wegner (MSc), 2007. Fusion d'images optiques et radar à haute résolution spatiale. Université de Hanovre, Allemagne.
14. Julien Osman (Ingénieur), 2008. Segmentation supervisée d'objets dans les images HR. ENSIEG. Grenoble. France.
15. Christophe Lay (Ingénieur), 2008. Génération automatique de cartes d'occupation des sols par classification SVM. ESTIA. Bayonne. France.
16. Eric Koun (Ingénieur), 2009. Détection de changements orientée objets. ENSIEG, Grenoble. France.
17. Malik Ciss (Ingénieur), 2009. Classification de séries d'images multi-temporelles. ENSEEIHT. Toulouse, France.
18. Benoît Beguet (Master Recherche), 2010. Détection, caractérisation et suivi de changements d'état des sols résultants de pratiques agricoles. Institut EGID. Bordeaux. France.

9.2 Thèses de doctorat

1. Amandine Robin : *Détection de changements et classification sous-pixelliques en imagerie satellitaire*. Dirigée par Sylvie Le Hégarat-Masclé et Lionel Moisan. Université Paris Des-cartes. Soutenue le 21/05/2007. **Co-encadrement (20%)**.
2. Florent Chatelain : *Lois Gamma multivariées pour le traitement d'images radar*. Dirigée par Jean-Yves Tournet. Institut National Polytechnique de Toulouse. Soutenue le 25/10/2007. **Co-encadrement (25%)**.
3. Tarek Habib : *Mesures de similarités pour la détection de changements abrupts en imagerie satellitaire multicapteurs*. Dirigée par Jocelyn Chanussot et Grégoire Mercier. Institut National Polytechnique de Grenoble. Soutenue le 2/12/2008. **Co-encadrement (50%)**.
4. Vincent Poulain : *Fusion d'images optique et radar à haute résolution pour la mise à jour de bases de données cartographiques*. Dirigée par Philippe Marthon et Jean-Yves Tournet. Institut National Polytechnique de Toulouse. Soutenue le 22/10/2010. **Co-encadrement (70%)**.
5. María Carolina Vanegas Orozco : *Spatial relations and spatial reasoning for the interpretation of Earth observation images using a structural model*. Dirigée par Isabelle Bloch. Télécom ParisTech. Soutenue le 13/01/2011. **Co-encadrement (30%)**.

6. Ahed Alboody : *Analyse et Interprétation d'images de télédétection et intégration dans des Systèmes d'Information Géographique*. Dirigée par Florence Sèdes. Université Paul Sabatier - Toulouse 3. Soutenance prévue en 2011. **Co-encadrement (50%)**.
7. François Petitjean : *Classification non supervisée de séries temporelles d'images satellites hétérogènes*. Dirigée par Pierre Gançarski. Université de Strasbourg. Soutenance prévue en 2012. **Co-encadrement (50%)**.

10 Participation à des modules d'enseignement

10.1 Traitement du signal radar

De 2001 à 2003 j'ai enseigné à l'ENST Bretagne un module de 12 heures sur le traitement du signal radar. Les aspects abordés étaient :

- le bilan de liaison
- les formes d'onde
- les radars de détection
- les radars Doppler
- les radars imageurs

10.2 Traitement d'images radar

Depuis 2002 j'enseigne un module de 3 h sur le traitement des images SAR en 3ème année à Supaéro et en 2ème année à l'ENSICA. Depuis 2009, ce module est aussi enseigné à l'ENSEEIHT en 3ème année. Il présente les sujets suivants :

- la modélisation et le filtrage du speckle
- la détection (cibles, contours)
- la segmentation et la classification

10.3 Traitement d'images pour les risques

Suite à mes activités de recherche et développement sur le traitement d'images pour les catastrophes naturelles, j'ai été sollicité à plusieurs reprises à réaliser des formations, typiquement sur une journée, sur ce sujet.

J'ai eu l'occasion de donner ces formations au GDTA à 2 reprises en 2003 et 2004.

Plus récemment, une formation pratique (sur le logiciel Monteverdi) a été mise en place à Télécom Paristech dans le cadre des formations ATHENS.

10.4 Orfeo Toolbox

L'intérêt suscité par l'Orfeo Toolbox a créé une demande de formation pour des utilisateurs. J'ai mis en place un programme de formation sur une durée de 3 à 4 jours qui comprend :

- Pré-requis de programmation en C++
- Concepts et approche des traitements dans OTB
- Géométrie des images
 - Modèles de capteur
 - Recalage d'images
 - Projections cartographiques
 - Ortho-rectification
- Radiométrie des images
 - Étalonnage absolu
 - Corrections atmosphériques
- Extraction de primitives
 - Indices radiométriques
 - Statistiques, textures
 - Moments géométriques
 - Morphologie mathématique
- Segmentation d'images
 - Ligne de partage des eaux
 - Croissance de régions
 - Mean-shift
- Classification d'images
 - K-moyennes
 - SVM
 - Classification orientée objets

Cette formation a été dispensée à la Maison de la Télédétection, au CESBIO, à Télécom ParisTech et à l'École Royale Militaire de Bruxelles. Elle a ensuite été sous-traitée à la société C-S, titulaire du contrat de développement industriel de l'OTB.

11 Projets

Mes travaux de recherche ont eu des liens avec des projets en partenariat avec d'autres organismes. J'en liste ici les principaux.

11.1 EEE-SPN (2002)

Le projet EEE-SPN (Exploitation d'ERS et ENVISAT en utilisant la technique des points stables) a été financé par l'ESA avec l'objectif de développer des techniques interférométriques permettant d'utiliser de façon conjointe les données des 2 générations de satellites radar européens.

Le consortium de partenaires était composé de : Altamira Information, CNES, DLR et l'Université de Delft. Les travaux de ce projet ont donné lieu aux travaux publiés dans [20].

11.2 Robin (2005-2006)

Ce projet a eu comme objectifs de produire des banques d'images, des vérités terrain et des comparaisons d'algorithmes pour la reconnaissance d'objets dans les images. Les approches analysées ont été :

- la détection de classes d'objets ;
- la détection générique d'objets ;
- la reconnaissance d'objets ;
- la catégorisation d'images.

Le projet a été financé par les Ministères de la Défense et de la Recherche dans le cadre *Technovision*. Le consortium de partenaires était composé de : Bertin technologies, CNES, Cybernetix, DGA, EADS, INRIA, ONERA, MBDA, SAGEM et THALES.

Le projet a duré 2 ans. J'ai participé à la définition des spécifications d'une base d'images d'objets cartographiques extraits de scènes Spot 5 et à la définition des épreuves d'évaluation d'algorithmes.

11.3 GMOSS (2005-2008)

Global Monitoring for Security and Stability était une réseau d'excellence de la ligne aéronautique et espace du 6ème PCRD.

L'objectif de ce réseau a été de faire travailler ensemble les acteurs de la recherche dans le domaine de la sécurité civile afin d'acquérir et de développer les connaissances et l'expérience nécessaires à développer des capacités d'observation globale en utilisant l'imagerie satellitaire. Les technologies et la recherche résidant dans le périmètre de GMOSS étaient :

1. Les méthodes, algorithmes et logiciels génériques pour l'interprétation automatique et la visualisation d'images, y compris l'extraction de primitives, la reconnaissance d'objets et la détection de changements.
2. Les technologies nécessaires à fournir :
 - a) la surveillance et le contrôle de traités internationaux sur les contrôles et la prolifération d'armes de destruction massive ;
 - b) l'estimation de la dynamique des populations à grande échelle ;
 - c) la surveillance d'infrastructures et de frontières ;
 - d) l'analyse de menaces à la sécurité des biens et des personnes et les besoins d'échanges d'information entre décideurs pour la gestion de crise.

Le réseau était composé de 22 partenaires dont l'UNOSAT, le DLR, l'EUSC, le CEA, le CNES, l'École Royale Militaire de Bruxelles ou les sociétés privées comme QinetiQ ou Definiens.

J'ai participé aux activités de recherche dans le domaine des outils et algorithmes génériques, et plus particulièrement aux tâches sur l'extraction de primitives et la détection de changements.

11.4 PREVIEW (2005-2008)

PREVIEW, *Prevention Information and Early Warning*, était un projet financé par l'Union Européenne dans le cadre du 6ème PCRD. Son objectif était de développer, au niveau européen, des nouveaux services de géo-information pour les risques naturels et industriels.

Le projet comptait avec 58 partenaires de 15 pays impliquant des chercheurs, des agences spatiales, des industriels et des utilisateurs comme les protections civiles.

Beaucoup d'activités de recherche sur la détection de changement, le recalage d'images ainsi que la validation de la *Chaîne Risques* ont réalisées dans le cadre de PREVIEW.

11.5 SAFER (2009-2012)

SAFER (Services and Applications For Emergency Response) est un projet européen financé dans le cadre de GMES. Il sert à préparer la mise en oeuvre opérationnelle du service GMES *Emergency Response*.

L'objectif principal de SAFER est de fournir des capacités de cartographie rapide en réponse à des événements catastrophiques.

Le consortium est constitué de 54 partenaires de 16 pays. Il constitue une suite logique au projet PREVIEW.

Dans ma dernière année passée à SI/AP j'ai participé aux premier mois de ce projet sur le même type de tâches qui m'ont occupé dans le cadre de PREVIEW.

12 Autres activités liées à la recherche

12.1 Organisation de séminaires et colloques

- De la séparation de sources à l’analyse en composantes indépendantes. Centres de Compétences Techniques CNES. 28 Juin 2001. Toulouse.
 - Présentation : Analyse multi-dimensionnelle et ACI.
- Mesures de similarités en traitement des images. Centres de Compétences Techniques CNES. 18 avril 2003. Toulouse.
 - Présentation : Mesures de similarités multi-capteurs.
- Session *Change Detection Techniques* du *IEEE International Geoscience and Remote Sensing Symposium*, Juin 2003.
- Non-gaussienneté, non-stationnarité et non-linéarité. Centres de Compétences Techniques CNES. 21 juin 2004. Toulouse.
 - Présentation : Non-gaussienneté en traitement du signal.
- Session *Open Source Initiatives for Remote Sensing* du *IEEE International Geoscience and Remote Sensing Symposium*, Juillet 2009.
- Session *Change Detection and Multitemporal Image Analysis* du *IEEE International Geoscience and Remote Sensing Symposium*, Juillet 2010.
- Tutoriel *Pragmatic Remote Sensing - A hands-on approach* dans le cadre du *IEEE International Geoscience and Remote Sensing Symposium*, Juillet 2010.
- Session *Change Detection and Multitemporal Image Analysis* du *IEEE International Geoscience and Remote Sensing Symposium*, Juillet 2011.

12.2 Jurys de thèse

1. Virginie Amberg : *Analyse de scènes péri-urbaines à partir d’images radar haute résolution – Application à l’extraction semi-automatique du réseau routier*. Dirigée par Philippe Marthon. Institut National Polytechnique de Toulouse. 10/11/2005.
2. Vincent Martin : *Contribution des filtres LPTV et des techniques d’interpolation au tatouage numérique*. Dirigée par Bernard Lacaze et Marie Chabert. Institut National Polytechnique de Toulouse. 28/11/2006.
3. Olivier D’hondt : *Analyse spatiale de texture non stationnaire dans les images SAR*. Dirigée par Éric Pottier. Université de Rennes 1. 0/02/2006.

4. Amandine Robin : *Détection de changements et classification sous-pixelles en imagerie satellitaire*. Dirigée par Sylvie Le Hégarat-Masclé et Lionel Moisan. Université Paris Descartes. 21/05/2007.
5. Florent Chatelain : *Lois Gamma multivariées pour le traitement d'images radar*. Dirigée par Jean-Yves Tourenet. Institut National Polytechnique de Toulouse. 25/10/2007.
6. Mathieu Fauvel : *Spectral and spatial methods for the classification of urban remote sensing data*. Dirigée par Jocelyn Chanussot et Jon Atli Benediktsson. Institut National Polytechnique de Grenoble. Novembre 2007.
7. Jérémie Jakubowicz : *La recherche des alignements dans les images digitales et ses applications à l'imagerie satellitaire*. Dirigée par Jean-Michel Morel. École Normale Supérieure de Cachan. 30/11/2007.
8. Avik Bhattacharya : *Indexing of Satellite Images Using Structural Information*. Dirigée par Michel Roux et Josiane Zerubia. Télécom ParisTech. 14/12/2007.
9. Mihai Costache : *Support Vector Machines et méthodes bayésiennes pour l'apprentissage sémantique fondé sur des catégories : recherche dans les bases de données d'imagerie satellitaire*. Dirigée par Henri Maître et Mihai Datcu. Télécom ParisTech. 12/09/2008
10. Anne-Lise Chesnel : *Quantification de dégâts sur le bâti liés aux catastrophes majeures par images satellite multimodales très haute résolution*. Dirigée par Lucien Wald. Mines ParisTech. 15/09/2008.
11. Alexandre Fournier : *Détection de cibles par une analyse des perturbations de la texture*. École Nationale Supérieure de l'Aéronautique et de l'Espace, ISAE, 31/10/2008.
12. Tarek Habib : *Mesures de similarités pour la détection de changements abrupts en imagerie satellitaire multiscapteurs*. Dirigée par Jocelyn Chanussot et Grégoire Mercier. Institut National Polytechnique de Grenoble. 2/12/2008.
13. Marie Lauginie Liéno : *Apprentissage automatique des classes d'occupation du sol et représentation en mots visuels des images satellitaires*. Dirigée par Henri Maître et Mihai Datcu. Télécom ParisTech. 2/03/2009.
14. Julien Radoux : *Updating land cover maps by GIS-driven analysis of very high resolution satellite images*. Dirigée par Pierre Defourny. Université Catholique de Louvain. 13/01/2010.
15. Aymen El Ghoul : *Phase fields for network extraction from images*. Dirigée par Ian Jermyn et Josiane Zerubia. Université de Nice - Sophia Antipolis. 12/09/2010
16. Vincent Poulain : *Fusion d'images optique et radar à haute résolution pour la mise à jour de bases de données cartographiques*. Dirigée par Philippe Marthon et Jean-Yves Tourenet. Institut National Polytechnique de Toulouse. 22/10/2010.
17. Hélène Sportouche : *Extraction et reconstruction des bâtiments en milieu urbain à partir d'images optiques et radar à haute résolution*. Dirigée par Florence Tupin. Télécom ParisTech. 10/12/2010
18. María Carolina Vanegas Orozco : *Spatial relations and spatial reasoning for the interpretation of Earth observation images using a structural model*. Dirigée par Isabelle Bloch. Télécom ParisTech. 13/01/2011.

12.3 Travail éditorial

- Éditeur associé de IEEE Transactions on Geoscience and Remote Sensing depuis février 2008.
- Relecteur pour IEEE Transactions on Geoscience and Remote Sensing.
- Relecteur pour IEEE Geoscience and Remote Sensing Letters.
- Relecteur pour IEEE Transactions on Image Processing.
- Relecteur pour the International Journal of Remote Sensing.
- Relecteur pour IEE Proceedings - Vision, Image and Signal Processing.
- Relecteur pour IEE Electronics Letters.
- Relecteur pour Remote Sensing of Environment.
- Membre du *Technical Program Committee* du *IEEE International Geoscience and Remote Sensing Symposium 2003*.
- Membre du comité de programme de SPIE Remote Sensing Europe 2010

Bibliographie

- [1] Ahed Alboody, Florence Sèdes, and Jordi Inglada. Enriching the qualitative spatial reasoning system rcc8. In Shyamanta M. Hazarika, editor, *Qualitative Spatio-Temporal Representation and Reasoning : Trends and Future Directions*. Information Science Reference, 2010.
- [2] A. Apicella, J.S. Kippenhan, and J.H. Nagel. Fast multi-modality image matching. In *SPIE, Medical Imaging III : Image Processing*, volume 1092, pages 252–263, 1989.
- [3] C. Bessière. Constraint propagation. In *Foundations of Artificial Intelligence*, volume 2, pages 29–83. Elsevier, 2006.
- [4] M. Bro-Nielsen. *Medical Image Registration and Surgery Simulation*. PhD thesis, Technical University of Denmark, aug 1996.
- [5] Florent Chatelain, Jean-Yves Tournet, and Jordi Inglada. Change detection in multi-sensor sar images using multivariate gamma distributions. *IEEE Transactions on Image Processing*, 17(3), 2008.
- [6] Florent Chatelain, Jean-Yves Tournet, Jordi Inglada, and André Ferrari. Bivariate gamma distributions for image registration and change detection. *IEEE Transactions on Image Processing*, 16(7) :1796–1806, 2007.
- [7] I. Csizar. Information type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2 :299–318, 1967.
- [8] E. de Castro and C. Morandi. Registration of translated and rotated images using finite Fourier transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5) :700–703, 1987.
- [9] Didier Dubois and Henri Prade. Possibility theory, probability theory and multiple-valued logics : A clarification. *Annals of Mathematics and Artificial Intelligence*, 32(1-4) :35–66, 2001.
- [10] Germain Forestier, Jordi Inglada, Cédric Wemmert, and Pierre Gancarski. Mining spectral libraries to study sensors’ discrimination ability. In *SPIE Europe Remote Sensing*, volume 7478, Berlin, Germany, September 2009.

- [11] Germain Forestier, Cédric Wemmert, Pierre Gancarski, and Jordi Inglada. Mining multiple satellite sensor data using collaborative clustering. In *Proceedings of the 2009 IEEE International Conference on Data Mining*, pages 501–506, 2009.
- [12] N. M. Gotts, J. M. Gooday, and A. G. Cohn. A connection based approach to common-sense topological description and reasoning. *Topology for philosophers*, 79(1) :51–75, 1996.
- [13] J. Inglada. Change detection on SAR images by using a parametric estimation of the Kullback-Leibler divergence. In *Proc of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Toulouse, France, July, 21–25 2003.
- [14] J. Inglada. Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(3) :236–248, 2007.
- [15] J. Inglada, O. Hagolle, and G. Dedieu. Assessment of the land cover classification accuracy of venus image time series with respect to formosat-2. In *Proceedings of Recent Advances in Quantitative Remote Sensing*, Valencia, Spain, 2010.
- [16] J. Inglada and G. Mercier. A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5) :1432, 2007.
- [17] J. Inglada and J. Michel. Qualitative spatial reasoning for high-resolution remote sensing image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2) :599–612, February 2009.
- [18] Jordi Inglada and Alain Giros. On the possibility of automatic multi-sensor image registration. *IEEE Trans. Geosci. Remote Sensing*, 42(10), October 2004.
- [19] Jordi Inglada, Vincent Muron, Damien Pichard, and Thomas Feuvrier. Analysis of artifacts in sub-pixel remote sensing image registration. In *ISPRS Commission I Symposium.*, Marne-la-Vallée. France, June 2006.
- [20] Jordi Inglada, Jean-Claude Souyris, Caroline Henry, and Céline Tison. Incoherent sar polarimetric analysis over point targets. *IEEE Geoscience and Remote Sensing Letters*, 3(2) :246–249, April 2006.
- [21] Jim Xiuquan Ji, Hao Pan, and Zhi-Pei Lang. Further Analysis of Interpolation Effects in Mutual Information-Based Image Registration. *IEEE Trans. Med. Imaging*, 22(9) :1131–1119, 2003.
- [22] Florence Laporterie. *Représentations hiérarchiques d’images avec des pyramides morphologiques. application à l’analyse et à la fusion spatio-temporelle de données en observation de la terre.* Thèse de doctorat, École Nationale Supérieure de l’Aéronautique et de l’Espace, Toulouse, 2002.
- [23] J. Le Moigne, A. Cole-Rhodes, R. Eastman, K. Johnson, J. Morissette, N.S. Netanyahu, H.S. Stone, and I. Zavorin. Earth Science Imagery Registration. In *International Geoscience and Remote Sensing Symposium, IGARSS 2003, CD-ROM*, 2003.

- [24] Grégoire Mercier, Stéphane Derrode, Wojciech Pieczynski, Jean-Marie Nicolas, Anne Joannic-Chardin, and Jordi Inglada. Copula-based stochastic kernels for abrupt change detection. In *IEEE International Geoscience and Remote Sensing Symposium - IGARSS*, Denver - Colorado (USA), July 2006.
- [25] Martino Pesaresi and Jon Atli Benediktsson. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2) :309–320, february 2001.
- [26] Josien P. W. Pluim, J. B. Antoine Maintz, and Max A. Viergever. Interpolation Artefacts in Mutual Information-Based Image Registration. *Computer Vision and Image Understanding*, (77) :211–232, 2000.
- [27] V. Poulain, J. Inglada, M. Spigai, J-Y. Tourneret, and P. Marthon. High resolution optical and sar image fusion for building database updating. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2) :599–612, February 2011.
- [28] Amandine Robin. *Détection de changements et classification sous-pixeliques en imagerie satellitaire*. Thèse de doctorat, Université Paris Descartes, Paris, 2007.
- [29] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *MICCAI'98, Cambridge, Massachusetts*, pages 1115–1124, October 1998.
- [30] D. Sarrut. *Recalage multimodal et plate-forme d'imagerie médicale à accès distant*. PhD thesis, Université Lumière Lyon 2, January 2000.
- [31] D. Sarrut and S. Miguët. Similarity measures form image registration. In *European Workshop on Content-Based Multimedia Indexing, Toulouse, France.*, pages 263–270, October 1999.
- [32] Philippe Smets. What is dempster-shafer's model? In *Advances in the Dempster-Shafer theory of evidence*, pages 5–34. John Wiley & Sons, Inc., New York, NY, USA, 1994.
- [33] P. Thévenaz, U.E. Ruttimann, and M. Unser. Iterative multi-scale registration without landmarks. *IEEE Transactions on Image Processing*, 7(1) :27–41, 1998.
- [34] J.R.G. Townshend, C.O. Justice, C. Gurney, and J. McManus. The impact of misregistration on change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 30(5) :1054–1060, sept 1992.
- [35] Jeffrey Tsao. Interpolation Artifacts in Multimodality Image Registration Based on Maximization of Mutual Information. *IEEE Trans. Med. Imaging*, 22(7) :854–864, 2003.
- [36] María Carolina Vanegas, Isabelle Bloch, and Jordi Inglada. Alignment and parallelism for the description of objects in high resolution remote sensing images. *Computer Vision and Image Understanding. Under review.*, 2011.
- [37] Jan Dirk Wegner, Jordi Inglada, and Céline Tison. Image analysis of fused sar and optical images deploying open source software library otb. In *ISPRS Hannover Workshop. High-Resolution Earth Imaging for Geospatial Information*, 2007.

- [38] Jan Dirk Wegner, Céline Tison, and Jordi Inglada. Automatic fusion of sar and optical imagery based on line features. In *7th European Conference on Synthetic Aperture Radar, EUSAR*, 2008.
- [39] Andreas Wimmer, Iris Lingenfelder, Charles Beumier, Jordi Inglada, and Simon J. Caseley. Feature recognition techniques. In S. Schneiderbauer B. Jasani, M. Pesaresi and G. Zeug, editors, *Remote Sensing from Space – Supporting International Peace and Security*, chapter 8, pages 105–118. Springer, 2009.
- [40] R.P. Woods, S.R. Cherry, and J.C. Mazziota. Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography*, 16(4) :620–633, 1992.
- [41] R.P. Woods, J.C. Mazziota, and S.R. Cherry. MRI-PET registration with automated algorithm. *Journal of Computer Assisted Tomography*, 17(4) :536–546, 1993.

Troisième partie

Annexes

13 Recalage d'images

On the Possibility of Automatic Multisensor Image Registration

Jordi Inglada and Alain Giros

Abstract—Multisensor image registration is needed in a large number of applications of remote sensing imagery. The accuracy achieved with usual methods (manual control points extraction, estimation of an analytical deformation model) is not satisfactory for many applications where a subpixel accuracy for each pixel of the image is needed (change detection or image fusion, for instance). Unfortunately, there are few works in the literature about the fine registration of multisensor images and even less about the extension of approaches similar to those based on fine correlation for the case of monomodal imagery. In this paper, we analyze the problem of the automatic multisensor image registration and we introduce similarity measures which can replace the correlation coefficient in a deformation map estimation scheme. We show an example where the deformation map between a radar image and an optical one is fully automatically estimated.

Index Terms—Image registration, multisensor, similarity measures.

I. INTRODUCTION

THE PROBLEM we want to deal with is the one of the automatic fine registration of images acquired with different sensors. By different sensors, we mean sensors that produce images with different radiometric properties, i.e., sensors which measure different physical magnitudes: optical sensors operating in different spectral bands, radar and optical sensors, etc.

For this kind of image pairs, the classical approach of fine correlation [1], [2], cannot always be used to provide the required accuracy, since this similarity measure (the correlation coefficient) can only measure similarities up to an affine transformation of the radiometries.

There are two main questions which can be asked about what we want to do.

- 1) Can we define what the similarity is between, for instance, a radar and an optical image?
- 2) What does *fine registration* mean in the case where the geometric distortions are so big and the source of information can be located in different places (e.g., the same edge can be produced by the edge of the roof of a building in an optical image and by the wall-ground bounce in a radar image)?

We can answer by saying that the images of the same object obtained by different sensors are two different representations of the same reality. For the same spatial location, we have two different measures. Both items of information come from the same source, and thus, they have a lot of common information.

This relationship may not be perfect, but it can be evaluated in a relative way: different geometrical distortions are compared, and the one leading to the strongest link between the two measures is kept.

The paper is organized as follows. Section II is a review of the existing remote sensing image registration literature. In Section III, we introduce a theoretical approach to image registration. The problem of modeling image deformations¹ is analyzed in Section IV. In Section V, we evaluate a set of similarity measures that can be used for the multisensor image registration problem, and we use one of them in Section VI in order to estimate the deformations between a radar image and an optical image of the same scene. Finally, in Section VII, we propose the use of deformation maps for the estimation of topography using radar and optical acquisitions.

II. REVIEW OF EXISTING WORK

In this section, we will review the works published in the literature about the automatic multisensor image registration. The literature about the subject is rather limited in the field of remote sensing compared to what has been published in the fields of medical imaging and computer vision. An interesting survey of image registration techniques can be found in [3]. This survey poses the problem of image registration using the concepts of similarity measure, geometric transformations, and feature space. We will take a similar approach in Section III.

The approach taken in most of the works consists in automatically extracting homologous points (HPs) in both images and using them to estimate a parametric analytical deformation model.

Ton and Jain [4] were among the pioneers in the research of automatic algorithms to emulate photointerpreter-based registration. They proposed an algorithm for HP selection allowing for the estimation of rotational and translational transformations on Landsat images. Their approach is similar to the one by Li *et al.* [5], who used the salient points of active contours as HPs. The main problem of this approach is the heavy computation needed for the implementation of active contours.

Several authors, as for instance Cracknell and Paithoonwatanakij [6], combine an orbital model and a HP search using correlation and heuristic planning. The idea consists in searching for couples of HP with a good correlation. The set of HPs selected is used to solve a least squares estimation of a

Manuscript received October 6, 2003; revised July 6, 2004.
The authors are with the Centre National d'Études Spatiales, DCT/SI/AP, F-31401 Toulouse Cedex 9, France (e-mail: jordi.inglada@cnes.fr).
Digital Object Identifier 10.1109/TGRS.2004.835294

¹We prefer the term *deformation* to the classical *disparity* because the latter is related to stereo-optical vision. In the case of optical-radar image pairs, the geometric relative deformations due to topography (depth) do not correspond to that kind of model. We will use the term *deformation* as a generalization of *disparity*.

parametric model. They claim to have subpixel accuracy on the National Oceanic and Atmospheric Administration's Advanced Very High Resolution Radiometer images.

Within the family of methods using automatic correlation, one can highlight the work of Foroosh *et al.* [7], where a closed-form expression for subpixel shift estimation is given. However, this approach can only be applied to translations. It is also interesting to point out the work of Stone *et al.* [8] where an algorithm for subpixel shift registration based on Fourier transforms is presented. This algorithm has the advantage of being very fast and robust to aliasing.

These approaches are not really multisensor in the sense that they use a similarity measure which is not. A way to transform these approaches into multisensor ones is by comparing extracted image primitives. For example, Inglada and Adragna [9] use a simple edge detection and a genetic algorithm in order to find the best set HP in the case of a Système Pour l'Observation de la Terre (SPOT)–European Remote Sensing Satellite (ERS) registration. The control points of the master image are randomly taken amongst the extracted edges.

Thépaut *et al.* [10] do a first geometrical correction using orbit information, and then a residual translation compensation using the correlation between the edges extracted from both, ERS synthetic aperture radar (SAR), and SPOT images. Other approaches exist for edge matching, as for instance the one of Wu and Maître [11], where a multiresolution analysis is used together with a hypothesis testing.

One of the first works on feature-based image registration was proposed by Ventura *et al.* [12]. They even applied it to the problem of image to map registration. The approach was also finding HP by matching extracted features.

Dai and Khorram [13] use a feature-based approach: they extract closed edges that are characterized using invariant moments. Then, the extracted areas are matched using their characterization. Finally, the centers of gravity of each area are used as HPs for the estimation of an affine transformation. They apply the approach to Landsat images, and they obtain an accuracy better than one pixel, which is similar to the accuracy obtained with manual registration.

Djamdjani *et al.* [14] propose a multiresolution approach, where the discrete wavelet transform is used. The automatic extraction of HPs is done by comparing thresholded wavelet coefficients.

All these approaches try to extract HP in order to compute an analytical deformation model. On the other hand, when working with images acquired with the same (type of) sensor, one can use a very effective approach. Since a correlation coefficient measure is robust and fast for similar images, one can afford to apply it in every pixel of one image in order to search for the corresponding HP in the other image. One can, thus, build a deformation grid (a sampling of the deformation map). If the sampling step of this grid is short enough, the interpolation using an analytical model is not needed, and high-frequency deformations can be estimated. The obtained grid can be used as a resampling grid and, thus, obtain the registered images.

No doubt, this approach, combined with image interpolation techniques (in order to estimate subpixel deformations) and

multiresolution strategies, allows for obtaining the best performances in terms of deformation estimation and, hence, for the automatic image registration.

Unfortunately, in the multisensor case, the correlation coefficient cannot be used. This will be justified in Section V-B. We will, thus, try to find similarity measures that can be applied in the multisensor case with the same approach as the correlation coefficient.

III. MODEL FOR THE IMAGE REGISTRATION PROBLEM

In this section, we give several definitions that allow for the formalization of the image registration problem. First of all, we define the master image and the slave image.

Definition 1 (Master Image): Image to which other images will be registered. Its geometry is considered as the reference.

Definition 2 (Slave Image): Image to be geometrically transformed in order to be registered to the master image.

Two main concepts are the one of *similarity measure* and the one of *geometric transformation*.

Definition 3: Let I and J be two images, and let c be a similarity criterion. We call a similarity measure any scalar strictly positive function

$$S_c(I, J) = f(I, J, c). \quad (1)$$

S_c has an absolute maximum when the two images I and J are *identical* in the sense of criterion c .

Definition 4: A geometric transformation T is an operator that, applied to the coordinates (x, y) of a point in the slave image, gives the coordinates (u, v) of its HP in the master image

$$\begin{pmatrix} u \\ v \end{pmatrix} = T \begin{pmatrix} x \\ y \end{pmatrix}. \quad (2)$$

Finally, we introduce a definition for the image registration problem.

Definition 5 (Registration Problem):

- 1) Determine a geometric transformation T that maximizes the similarity between a master image I and the result of the transformation $T \circ J$

$$\text{Arg max}_T (S_c(I, T \circ J)). \quad (3)$$

- 2) Resampling of J by applying T .

We must note that Le Moigne *et al.* have proposed in a recent contribution [15] a modular approach for registration that allows the analysis of different similarity measures and different optimization strategies. The presented results, which are still preliminary, are very promising. The multisensor case has been dealt with, but only for optical images (Ikonos and Landsat Enhanced Thematic Mapper Plus). The case of very different images (e.g., optical and radar) has not been explored.

IV. GEOMETRIC DEFORMATION MODELING

The geometric transformation of Definition 4 is used for the correction of the existing deformation between the two images to be registered. This deformation contains information, which

TABLE I
CHARACTERIZATION OF THE GEOMETRIC DEFORMATION SOURCES

	Intensity	Spatial Frequency
Mean Attitude	Strong	Low
Stereo	Medium	High and Medium
Attitude evolution	Low	Low to Medium

is linked to the observed scene and the acquisition conditions. The deformations can be classified into the following three classes, depending on their physical source:

- 1) deformations linked to the mean attitude of the sensor (incidence angle, presence or absence of yaw steering, etc.);
- 2) deformations linked to a stereo vision (mainly due to the topography);
- 3) deformations linked to attitude evolution during the acquisition (vibrations that are mainly present in pushbroom sensors).

These deformations are characterized by their spatial frequencies and intensities, which are summarized in Table I.

Depending on the type of deformation to be corrected, its model will be different. For example, if the only deformation to be corrected is the one introduced by the mean attitude, a physical model for the acquisition geometry (independent of the image contents) will be enough. If the sensor is not well known, this deformation can be approximated by a simple analytical model. When the deformations to be modeled are high frequency, analytical (parametric) models are not suitable for a fine registration. In this case, one has to use a fine sampling of the deformation, which means the use of deformation grids. These grids give, for a set of pixels of the master image, their location in the slave image.

The following points summarize the problem of the deformation modeling.

- 1) An analytical model is just an approximation of the deformation. It is often obtained as follows:
 - a) directly from a physical model without using any image content information;
 - b) by estimation of the parameters of an *a priori* model (polynomial, affine, etc.). These parameters can be estimated
 - i) either by solving the equations obtained by taking HP. The HP can be manually or automatically extracted;
 - ii) or by maximization of a global similarity measure.
- 2) A deformation grid is a sampling of the deformation map.

The last point implies that the sampling period of the grid must be short enough in order to account for high-frequency deformations (Shannon theorem). Of course, if the deformations are nonstationary (it is usually the case of topographic deformations), the sampling can be irregular.

As a conclusion, we can say that Definition 5 poses the registration problem as an optimization problem. This optimization can be either global or local with a similarity measure, which can also be either local or global. All this is synthesized in Table II.

TABLE II
APPROACHES TO IMAGE REGISTRATION

Geometric model	Similarity measure	Optimization of the deformation
Physical model	None	Global
Analytical model with a priori HP	Local	Global
Analytical model without a priori HP	Global	Global
Grid	Local	Local

The ideal approach would consist in a registration that is locally optimized, both in similarity and deformation, in order to have the best registration quality. This is the case when deformation grids with dense sampling are used. Unfortunately, this case is the most computationally heavy, and one often uses either a low sampling rate of the grid or the evaluation of the similarity in a small set of pixels for the estimation of an analytical model.² Both of these choices lead to local registration errors, which, depending on the topography, can amount to several pixels.

Even if this registration accuracy can be enough in many applications, (orthoregistration, import into a geographic information system, etc.), it is not acceptable in the case of data fusion, multichannel segmentation, or change detection [16]. This is why we will focus on the problem of deformation estimation using dense grids.

None of the references presented in Section II uses the local optimization approach. We can also note that in the multisensor case only few authors [15] have used any similarity measure other than the correlation coefficient. However, in the medical imaging field, as we will see in Section V, a lot of similarity measures have been proposed as a generalization of the correlation coefficient. These measures enable the registration of very different imagery modalities. Nevertheless, these works are not directly usable in our problem, since the geometric deformations present in medical images can be easily represented by global analytical models. Indeed, often a rigid model (rotation, translation, scale) or slightly elastic (affine plus a $a \cdot x \cdot y$ term) is enough, since: 1) the sensors are stable; 2) the stereo effect is small; 3) and only the point of view changes. As we have noted above, deformations due to topography can locally have high frequencies for medium- and high-resolution sensors (30 m and better), thus our need for a fine modeling. We also point out that the problem of hidden faces is beyond the scope of this paper.

V. SIMILARITY MEASURES

The fine modeling of the geometric deformation we are looking for needs for the estimation of the coordinates of nearly every pixel in the master image inside the slave image. In the classical monosensor case where we use the correlation coefficient, we proceed as follows.

²However, computation time is not nowadays a real issue. As an illustration, we can give the example of the PAN+XS fusion procedure developed at the Centre National d'Études Spatiales (CNES) for the SPOT 5 ground segment where the subpixel registration between the Panchromatic band and the multispectral channels is done on $24\,000 \times 24\,000$ images in less than 1 h on a Sun Sparc Ultra-4 workstation.

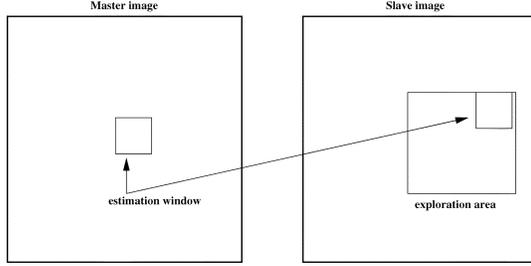


Fig. 1. Estimation of the correlation surface.

The geometric deformation is modeled by local rigid displacements. One wants to estimate the coordinates of each pixel of the master image inside the slave image. This can be represented by a displacement vector associated to every pixel of the master image. Each of the two components (lines and columns) of this vector field will be called deformation grid.

We use a small window taken in the master image, and we test the similarity for every possible shift within an exploration area inside the slave image (Fig. 1).

That means that, for each position, we compute the correlation coefficient. The result is a correlation surface whose maximum gives the most likely local shift between both images

$$\rho_{I,J}(\Delta x, \Delta y) = \frac{1}{N} \frac{\sum_{x,y} (I(x,y) - m_I)(J(x + \Delta x, y + \Delta y) - m_J)}{\sigma_I \sigma_J}. \quad (4)$$

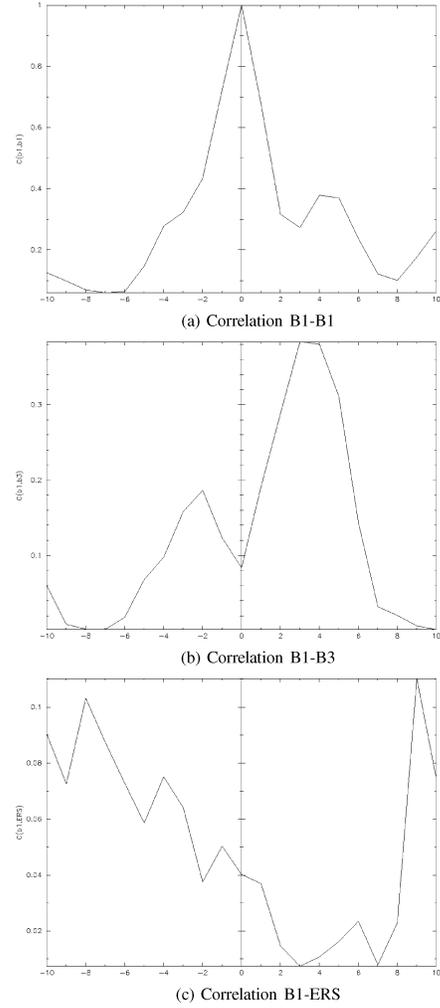
In this expression, N is the number of pixels of the analysis window; m_I and m_J are the estimated mean values inside the analysis window of, respectively, image I and image J ; and σ_I and σ_J are their standard deviations.

Quality criteria can be applied to the estimated maximum in order to give a confidence factor to the estimated shift: width of the peak, maximum value, etc. Subpixel shifts can be measured by applying fractional shifts to the sliding window. This can be done by image interpolation.

The interesting parameters of the procedure are the following.

- *The size of the exploration area:* it determines the computational load of the algorithm (we want to reduce it), but it has to be large enough in order to cope with large deformations.
- *The size of the sliding window:* the robustness of the correlation coefficient estimation increases with the window size, but the hypothesis of local rigid shifts may not be valid for large windows.

The correlation coefficient cannot be used with original gray-level images in the multisensor case. It could be used on extracted features (edges, etc.), but the feature extraction can introduce localization errors. Also, when the images come from sensors using very different modalities, it can be difficult to find similar features in both images. In this case, one can try to find the similarity at the pixel level, but with other similarity measures and apply the same approach as we have just described.


 Fig. 2. Measure of $\rho(\Delta x)$ for three different pairs of images.

The concept of similarity measure has been presented in Definition 3. The difficulty of the procedure lies in finding the function f , which properly represents the criterion c . We also need that f be easily and robustly estimated with small windows. We extend here what we proposed in [17].

A. Correlation Coefficient

We remind here the computation of the correlation coefficient between two image windows I and J . The coordinates of the pixels inside the windows are represented by (x, y)

$$\rho(I, J) = \frac{1}{N} \frac{\sum_{x,y} (I(x,y) - m_I)(J(x,y) - m_J)}{\sigma_I \sigma_J}. \quad (5)$$

In order to qualitatively characterize the different similarity measures we propose the following experiment. We take two

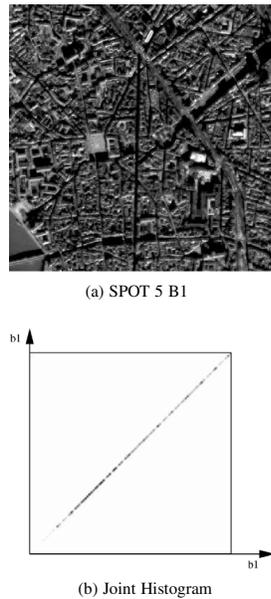


Fig. 3. Joint histogram of an image with itself.

images that are perfectly registered, and we extract a small window of size $N \times M$ from each of the images (this size is set to 101×101 for this experiment). For the master image, the window will be centered on coordinates (x_0, y_0) (the center of the image), and for the slave image, it will be centered on coordinates $(x_0 + \Delta x, y_0)$. With different values of Δx (from -10 pixels to 10 pixels in our experiments), we obtain an estimate of $\rho(I, J)$ as a function of Δx , which we write as $\rho(\Delta x)$ for short. The obtained curve should have a maximum for $\Delta x = 0$, since the images are perfectly registered. We would also like to have an absolute maximum with a high value and with a sharp peak, in order to have a good precision for the shift estimate.

In the following, we will make this experiment with different image pairs and different similarity measures. Fig. 2 shows the results obtained when the correlation coefficient is applied to [Fig. 2(a)] one extract of the B1 channel of a SPOT 5 image with itself, [Fig. 2(b)] an extract of channel B1 with the extract of channel B3, and [Fig. 2(c)] the extract of channel B1 with an extract of an ERS-2 SAR image. The images are presented in Figs. 3–5.

We can see that the correlation coefficient has a good behavior for the first pair, but its performances are bad when the images radiometries are different. The correlation coefficient can be characterized as follows:

- well-known algorithm;
- fits the registration needs when using radiometrically similar images;
- simple and fast computation;
- high precision in the estimation of the deformation;
- robust to noise.



Fig. 4. Joint histogram of two channels (B1-B3) of the same SPOT 5 image.

However, its main disadvantage is that it can only take into account affine transformations between radiometries ($j = \alpha i + \beta$), so it cannot be used in the general multisensor case.

B. Generalization: Probabilistic Interpretation

The correlation coefficient formulation [see (5)] can be revisited with a probabilistic interpretation

$$\begin{aligned} \rho(I, J) &= \frac{1}{N} \frac{\sum_{x,y} (I(x,y) - m_I)(J(x,y) - m_J)}{\sigma_I \sigma_J} \\ &= \sum_{(i,j)} \frac{(i - m_I)(j - m_J)}{\sigma_I \sigma_J} p_{ij} \end{aligned} \quad (6)$$

where the sum is taken over the list of radiometry pairs (i, j) , and p_{ij} is the value of the joint normalized histogram (estimation of the joint probability density function (pdf) $f_{ij}(i, j)$) of the pair of images.

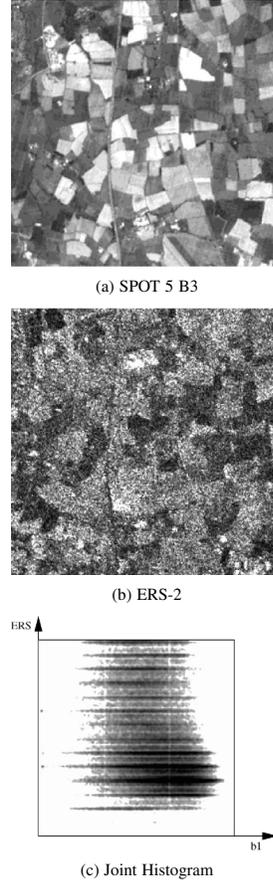


Fig. 5. Joint histogram of a SPOT 5 B3 image and a ERS-2 image.

That means that we are assuming a linear model

$$j = (i - m_I) \frac{\sigma_J}{\sigma_I} + m_J \quad (7)$$

and we evaluate its likelihood by weighting with the probability of each radiometry couple p_{ij} .

One could assume different models for the radiometry pairs leading to different measures as, for instance, the identity model $i = j$, which leads to the L_n norm

$$L_n(I, J) = \sum_{i,j} |i - j|^n p_{ij} \quad (8)$$

or more complex models based on textural approaches, as follows:

Diagonal moment:

$$MD(I, J) = \sum_{i,j} |i - j| (i + j - \sigma_I - \sigma_J) p_{ij}. \quad (9)$$

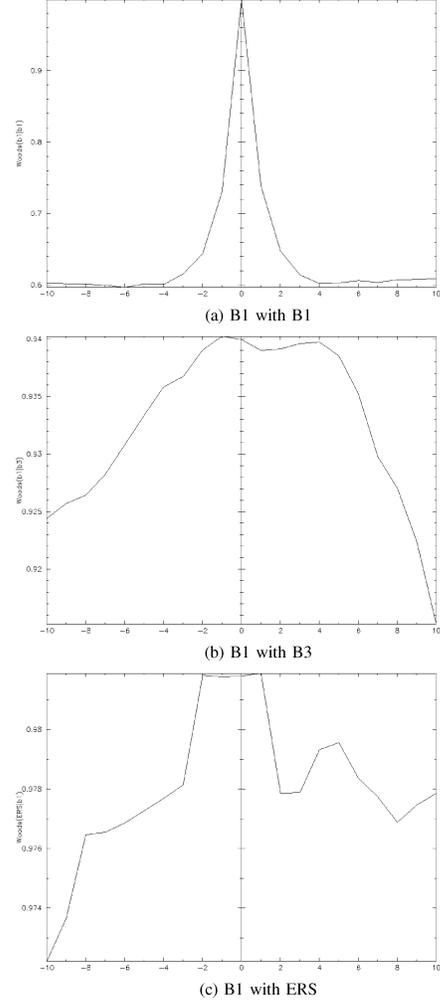


Fig. 6. Image shift experiment: Woods criterion.

Cluster Shade:

$$C_{\text{shade}}(I, J) = \sum_{i,j} (i + j - \sigma_I - \sigma_J)^3 p_{ij}. \quad (10)$$

Cluster Prominence:

$$C_{\text{pro}}(I, J) = \sum_{i,j} (i + j - \sigma_I - \sigma_J)^4 p_{ij}. \quad (11)$$

An assessment of these measures for image registration can be found in [18]. They are very sensitive to noise and are not useful for the comparison of gray levels of multisensor image pairs.

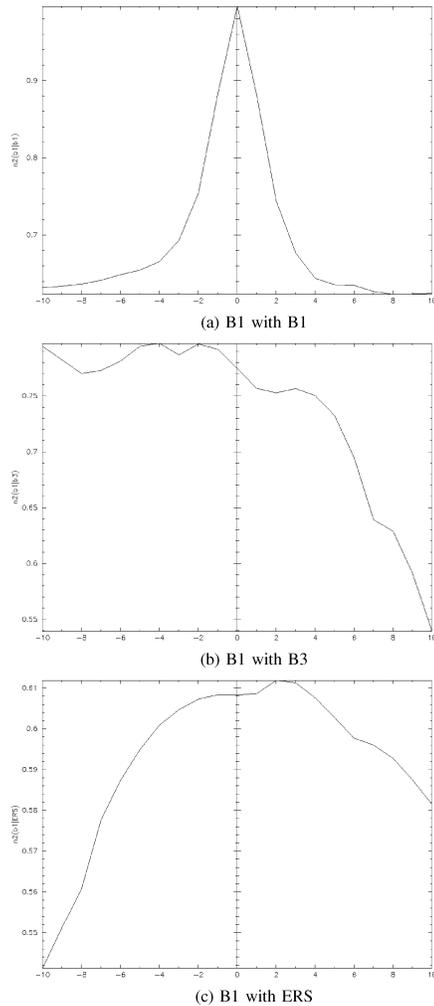


Fig. 7. Image shift experiment: Correlation ratio.

C. Estimation of Similarity Measures and p_{ij}

In the expression of the correlation coefficient the term p_{ij} is an estimation of the joint pdf of the radiometries of the images we are comparing. It can be seen as a link (transfer function) between both radiometries.

We show here several examples of the estimation of the joint histogram. In Figs. 3–5 are shown, respectively, the joint histograms of one image with itself (B1-B1), two different channels of the same SPOT 5 image (B1-B3), and a SPOT 5 B3-ERS-2 pair.

As expected, the joint histogram of an image with itself is a straight line with slope 1. It shows the full correlation between the two images: the identity transfer function. This kind of situation is well dealt with by the correlation coefficient.

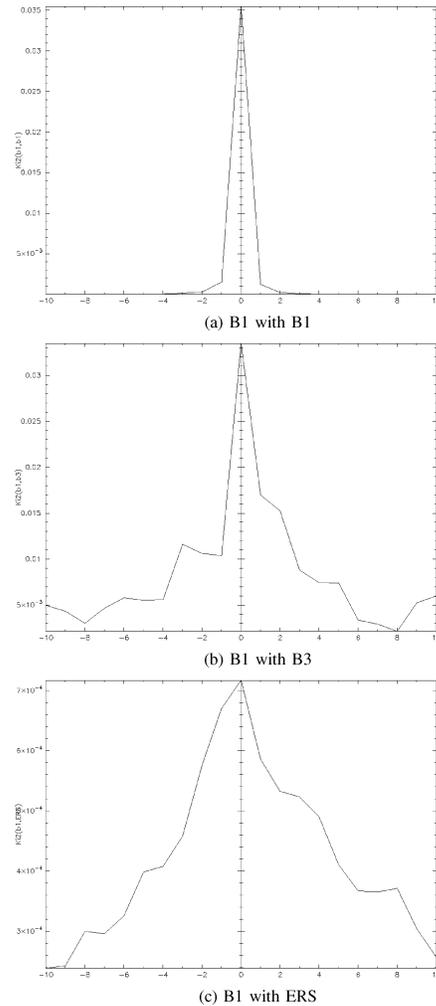


Fig. 8. Image shift experiment: Distance to independence.

TABLE III
EXPRESSIONS OF FUNCTION f IN THE f -DIVERGENCE FAMILY

Measure	$f(x)$
Kolmogorov distance	$\frac{1}{2} x-1 $
Mutual information	$x \log x$
Kullback divergence	$(x-1) \log x$
χ^2 -divergence	$\frac{1}{2}(x-1)^2$
Hellinger distance	$\frac{1}{2}(\sqrt{x}-1)^2$
Toussaints distance	$\frac{x-1}{x+1}$
Lin K-divergence	$x \log \frac{2x}{1+x}$

The B1-B3 case (Fig. 4) shows two nearly linear tendencies that are mixed up. This case cannot be dealt with by the correlation coefficient.

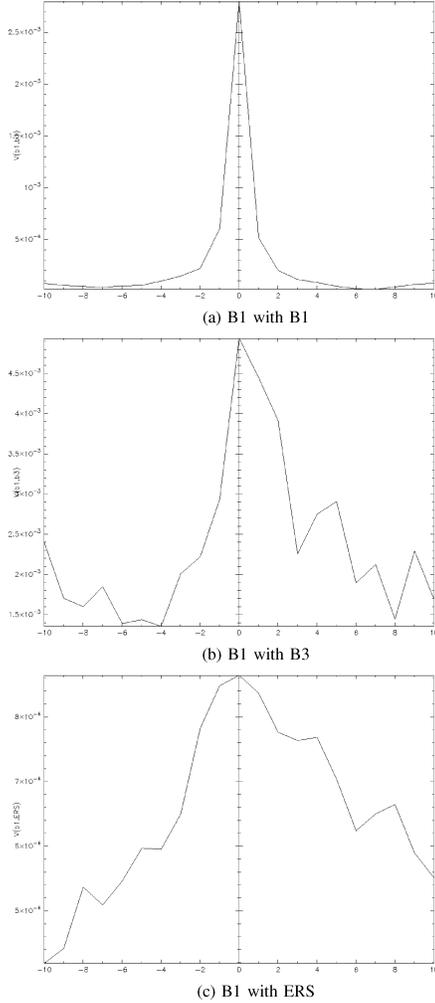


Fig. 9. Image shift experiment: Kolmogorov distance.

Finally, Fig. 5 shows the impossibility of finding any correlation link between two sensors, which are as different as an optical and a radar one.

1) *Computation Time*: The main difference between the two expressions of the correlation coefficient given by (5) and (6) is the estimation of the joint pdf needed in the second expression. This estimation is usually done by computing the joint histogram. The joint histogram can be computed with different methods, but their discussion is beyond the scope of this paper. However, it is important to note that the method chosen for histogram computation may induce significant changes in the computation cost of the similarity measure. As an example, with our implementation (counting with optimization of the number of classes), there is an increase factor of 4 in computation time between (5) and (6).

The multisensor measures that will be introduced in the next section need the estimation of the joint histogram. Hence, their computation time is comparable to the one of (6). The differences of computation complexity between these measures are negligible, since the longest part of the algorithm is taken by the joint histogram estimation.

D. Multisensor Measures

We introduce here several similarity measures that have been proved useful in the problem of multimodality medical image registration [19].

In the following, the sums will be computed over radiometry values. We will use the conditional mean

$$m_{I|j} = \frac{1}{p_j} \sum_i i p_{ij} \quad (12)$$

and the conditional variance

$$\sigma_{I|j}^2 = \frac{1}{p_j} \sum_i (i - m_{I|j})^2 p_{ij}. \quad (13)$$

For each of the following measures, we will perform the experiment described in Section V-A.

1) *Measures Using the Radiometry Values and the Probabilities*: Within this class, we will not take into account the measures that are based on the differences of radiometries (L_n norm of the difference) [20]–[22] or textural measures, since they give low-quality results.

a) *Normalized standard deviation or Woods criterion*: The work by Woods *et al.* first on monomodal registration [23] and then on multimodal registration [24] lead to the elaboration of this similarity measure. Given the intensity on one image, i.e., the set of pixels having this value, this measure takes into account the variability of the intensities of the homologous pixels in the other image. The underlying hypothesis is that this variability (which is actually a measure of variance) will be minimum when the images are registered

$$\text{Woods}(I|J) = \sum_j \frac{\sigma_{I|j}}{m_{I|j}} p_j. \quad (14)$$

In order to have a criterion which has to be maximized, we will use

$$S_{\text{Woods}}(I|J) = 1 - \sum_j \frac{\sigma_{I|j}}{m_{I|j}} p_j. \quad (15)$$

The results on our three test image pairs are shown in Fig. 6. We see that for the monosensor case, the results are similar to those of the correlation coefficient. For the two multisensor examples, we obtain high values near the zero-shift, but the location of these maxima is not accurate.

b) *Correlation ratio*: This is a very well known measure in statistics. It has been first proposed in the framework of image registration by Roche *et al.* [25]. It is defined as follows:

$$\eta^2(I|J) = 1 - \frac{1}{\sigma_I^2} \sum_j \sigma_{I|j}^2 p_j. \quad (16)$$

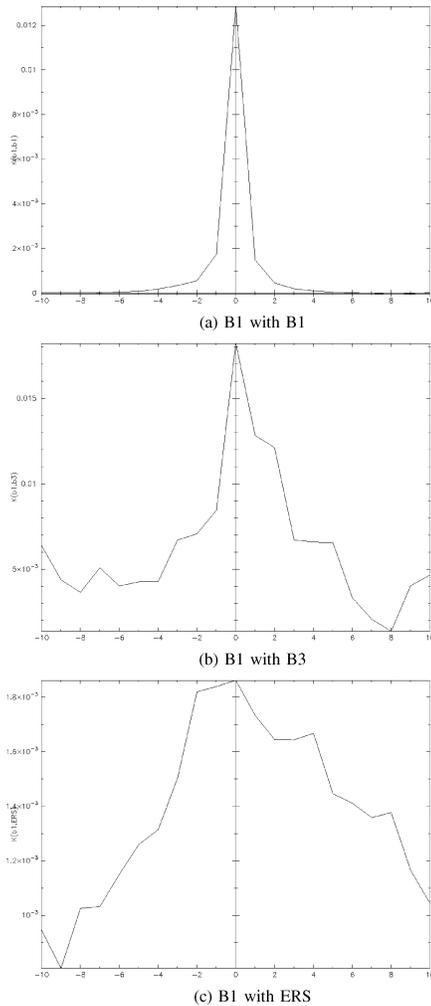


Fig. 10. Image shift experiment: Mutual information.

Its interpretation is similar to the one of the Woods criterion. The results are shown in Fig. 7, and they are worse than those of the Woods criterion.

2) *Measures Using Only the Probabilities*: This class of measures does not directly use the radiometries of the pixels, but only the estimation of the joint pdf. Of course, the pixel pairs are used for the estimation of this probability.

a) *Distance to independence*: It is a normalized version of the χ^2 test

$$\chi^2(I, J) = \sum_{i,j} \frac{(p_{ij} - p_i p_j)^2}{p_i p_j}. \quad (17)$$

It measures the degree of statistical dependence between both images, since for two independent random variables, the joint

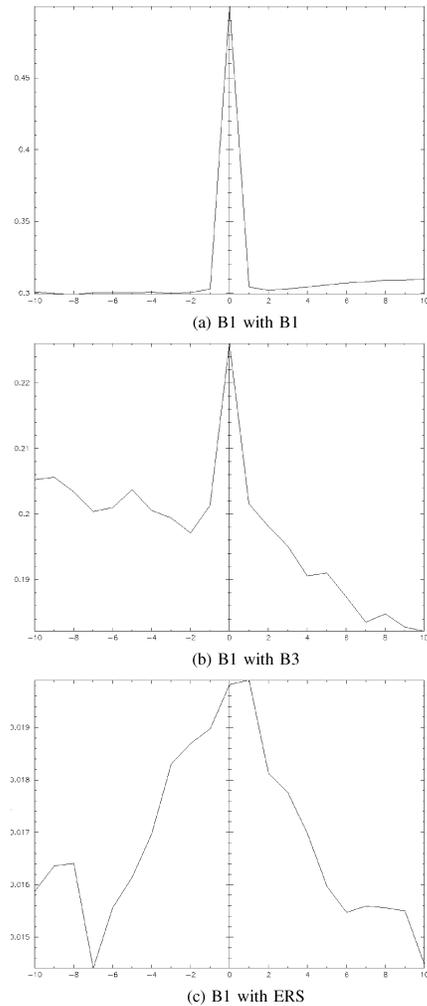


Fig. 11. Image shift experiment: CRA.

pdf is equal to the product of the marginals. The correlation coefficient is a test of independence of order 2, and this one is the generalization to any order. The results are shown in Fig. 8. In this case, for the three pairs, we obtain an absolute maximum for the zero-shift case, which is sharp enough for a robust automatic detection.

b) *f-divergence family*: An *f-divergence* [26] measures the expectation of the diversity of the likelihood ratio between two distributions P and Q

$$D_f(P, Q) = E_Q \left[f \left(\frac{dp(x)}{dq(x)} \right) \right] = \int f \left(\frac{p(x)}{q(x)} \right) q(x) dx. \quad (18)$$

E_Q is the expectation with respect to Q ; $dp(x)/dq(x)$ is the derivative with respect to a density; and f is continuous and

TABLE IV
DEFORMATION GRID ESTIMATION MEAN SQUARE ERROR

Window size	Mutual information, B1-B3. MSE in <i>pixel</i> ²					
	$T = \infty$	$T = 4000$	$T = 2000$	$T = 1000$	$T = 500$	$T = 200$
20×20	5.4	7.9	8.8	8.8	8.3	10.5
40×40	0.5	0.9	1.1	2.4	2.3	3.6
60×60	0.1	0.2	0.2	0.1	0.3	3.2
80×80	0.0	0.0	0.0	0.1	0.5	3.9
100×100	0.0	0.0	0.0	0.1	0.7	5.5

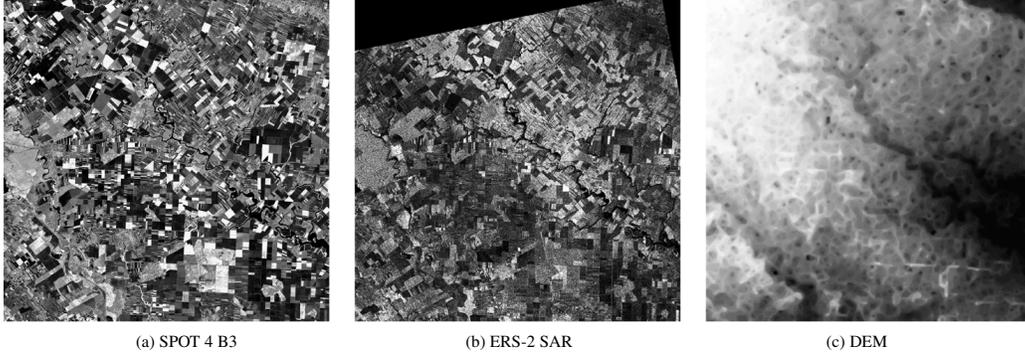


Fig. 12. Images and DEM for the test area.

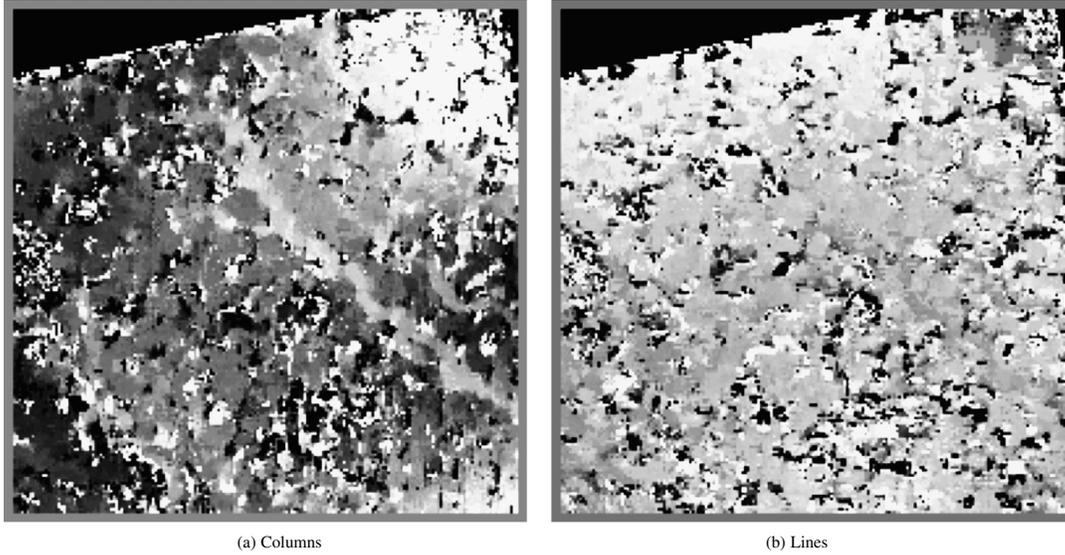


Fig. 13. Deformation grid. CRA, estimation window is 51×51 pixels, and the sampling rate is ten pixels.

convex on $[0, +\infty)$. A divergence can be seen as a relative entropy. In order to simplify the notation, we will use $p = p_{ij}$, $q = q_{ij}$, and $f = \sum_{i,j}$.

Depending on the choice of f (see Table III), we can obtain several interesting cases, as follows.

1) *Kolmogorov distance*:

$$V(P, Q) = \frac{1}{2} \int |p - q|. \quad (19)$$

2) *Kullback information or mutual information*:

$$K(P, Q) = \int p \log \frac{p}{q}. \quad (20)$$

3) *Kullback divergence*:

$$K'(P, Q) = \int (q - p)(\log q - \log p). \quad (21)$$

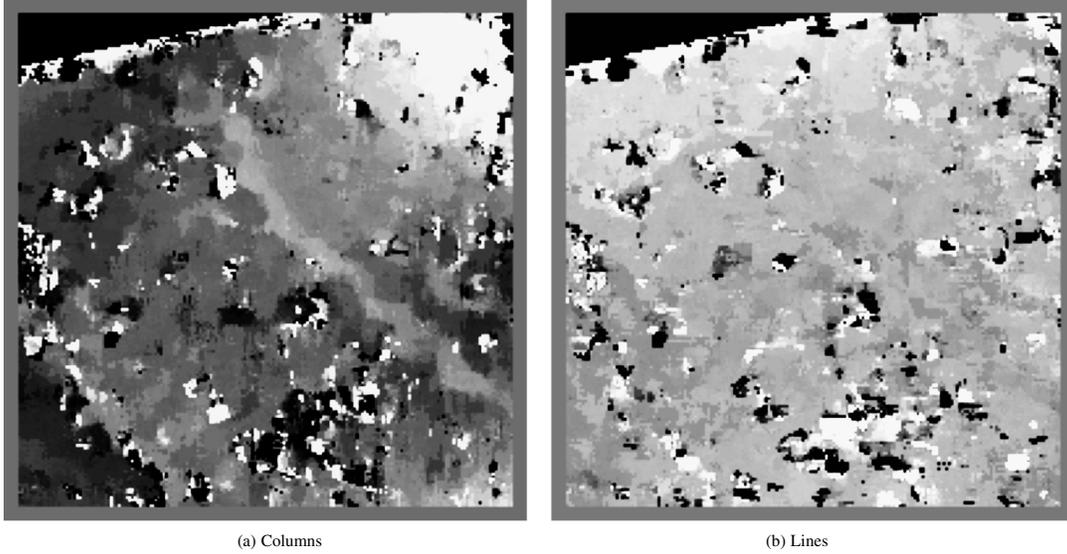


Fig. 14. Deformation grid. CRA, estimation window is 81×81 pixels, and the sampling rate is ten pixels.

4) χ^2 -divergence:

$$R(P, Q) = \frac{1}{2} \int \frac{(p - q)^2}{q}. \quad (22)$$

5) Hellinger distance:

$$\mathcal{H}^2(P, Q) = \frac{1}{2} \int (\sqrt{p} - \sqrt{q})^2. \quad (23)$$

6) Toussaints distance:

$$T(P, Q) = \int p - \frac{2pq}{p+q}. \quad (24)$$

7) Lin K -divergence:

$$K_{\text{div}}(P, Q) = \int p \log \frac{2p}{p+q}. \quad (25)$$

All these measures give very similar results [27]. We study two of them.

1) Kolmogorov distance:

$$V(P, Q) = \frac{1}{2} \int |p_{ij} - p_i p_j|. \quad (26)$$

It can be seen as a L_1 norm version of the χ^2 criterion. The results are shown in Fig. 9.

2) Mutual information:

$$K(P, Q) = \int p_{ij} \log \frac{p_{ij}}{p_i p_j}. \quad (27)$$

The results are shown in Fig. 10.

Both measures give satisfactory results, which are very similar to the ones obtained with the distance to the independence.

c) Cluster reward algorithm: Let $H_{IJ}(k, l)$ be the joint histogram of the pair of images, and let $H_I(k)$ and $H_J(k)$, respectively, be the marginal histograms and P the number of pixels. We define

$$I_{\text{CRA}} = \frac{\frac{\Phi}{F} - \frac{F}{P^2}}{1 - \frac{F}{P^2}} \quad (28)$$

where

$$\Phi = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} H_{IJ}^2(k, l) \quad (29a)$$

$$F = \sqrt{h_I h_J} \quad (29b)$$

$$h_I = \sum_{k=0}^{N-1} H_I^2(k) \quad (29c)$$

$$h_J = \sum_{k=0}^{N-1} H_J^2(k). \quad (29d)$$

The I_{CRA} index will have a high value when the joint histogram has little dispersion. This lack of dispersion can be due to a correlation (histogram distributed along a line) or to the clustering of radiometry values within the histogram. In both cases, one can predict the values of one image from the values of the other.

In order to compare I_{CRA} with the f -divergences, we can rewrite (28) as

$$I_{\text{CRA}} = \frac{\int p_{ij}^2 - \int p_i^2 \int p_j^2}{\sqrt{\int p_i^2 \int p_j^2 - \int p_i^2 \int p_j^2}}. \quad (30)$$

If we consider the denominator as a normalization term, we can focus only in the numerator. This numerator contains the same terms as the f -divergences, i.e., a term that depends on the joint pdf and a term that depends on the product of the marginals.

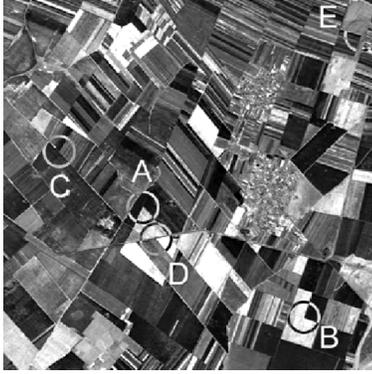


Fig. 15. Region of interest and test points for the validation.

TABLE V
RESULTS OF THE FIRST VALIDATION METHOD. SEE TEXT AND
FIGS. 15–20 FOR THE DETAILS

Point		ERS	SPOT	Deformation		
				Manual	Computed	Error
A (fig. 16)	Column	153.00	150.20	2.80	2.56	0.24
	Line	226.50	225.12	1.38	1.50	-0.12
B (fig. 17)	Column	326.56	323.22	3.34	3.32	0.02
	Line	345.00	343.44	1.56	1.44	0.12
C (fig. 18)	Column	62.00	60.11	1.89	1.48	0.41
	Line	168.33	165.89	2.44	2.31	0.13
D (fig. 19)	Column	166.56	163.44	3.12	2.56	0.56
	Line	259.78	258.00	1.78	1.57	0.21
E (fig. 20)	Column	390.20	385.10	5.10	4.57	0.53
	Line	42.50	40.75	1.75	1.37	0.38

We can, thus, make an interpretation that is similar to the independence tests. As expected, the obtained results (Fig. 11) are similar to those obtained with the f -divergences.

The main interest of the CRA with respect to the f -divergence family is that the joint histogram noise due to estimation has less influence in the similarity measure. This allows us to use smaller estimation windows. The only drawback in doing this is that the peak of the measure will be less sharp.

E. Characterization of Similarity Measures

Similarity measures can be characterized using the following criteria:

- the *geometric resolution*: maximum frequency of the deformations;
- ability to deal with images acquired from different sensors.

The first point is directly linked to the number of pixels needed to obtain a good estimate of the measure. Indeed, since when estimating deformation grids we assume that the deformation can be decomposed on local shifts, that means that there should be no deformation inside the estimation window. However, if we need large windows for the estimation, the likelihood for our assumption to be true can be very low.

In order to illustrate this problem, we have made the following experiment. We have taken the pair B1-B3 of a SPOT 5 image, and we have applied a sinusoidal deformation, with period T pixels to the slave (B3) image in the horizontal direction. We

have made the estimation of the deformation using the mutual information, and we have computed the mean square error between the estimated and the real deformation grids

$$E = \sum_i \sum_j \left(G(i, j) - \hat{G}(i, j) \right)^2$$

where $G(i, j)$ is the real sinusoidal deformation, and $\hat{G}(i, j)$ is the estimated deformation grid. The experiment has been made for different values of the period T and the estimation window size. The results, which are measured in square pixels, are shown in Table IV. We observe the following.

- For medium to long periods, the quality of the estimation increases with the window size.
- For short periods ($T = 500$ and $T = 200$), the increase of the window size produces a decrease of the performances. This is due to the fact that geometrical deformations are strong inside the estimation window.

F. Behavior in Presence of Noise

One could also analyze how the different similarity measures behave when noise is present in the data. The presence of additive noise in the data produces a dispersion of the joint histogram of the images. For the f -divergence family, since the estimation windows used in the experiments contain a high number of samples, noisy data does not cause noisy estimations of the similarity, but rather wider similarity peaks. However, the location of the similarity optimum is not affected by noise. This behavior is confirmed by the SPOT-ERS couple, since the radar image contains a strong multiplicative noise.

VI. GRID ESTIMATION: A REAL CASE

The similarity measures introduced in the previous sections have been tested in a simple framework of integer shifts in one dimension and with low frequency (Table IV). In order to further test the performances of these methods, we will apply them in the same way as we would do using the correlation coefficient, i.e., estimating a deformation grid between a radar and an optical image.

Our dataset consists in the following pair (a region of a size of 2000×2000 pixels is used for our tests):

- the B3 channel of a SPOT 4 image (20-m pixel resolution) acquired on June 24, 2001 over the East of the Bucharest area [Fig. 12(a)];
- a ERS-2 SAR three-looks intensity image (12.5-m pixel size and approximately 20-m pixel resolution) acquired on May 10, 2001 over the same area [Fig. 12(b)].

Both images were orthorectified: for the SPOT 4 image, a digital elevation model (DEM) [Fig. 12(c)] with an altimetric precision better than 10 m and a planimetric precision around 10 m has been used, together with the acquisition model (orbits, attitude) for the satellite; for the ERS-2 image, no DEM was used, but a constant altitude and homologous points manually taken on the SPOT 4 image were used in the orthorectification process. Globally, the images show a good superposition, but local errors exist, which can amount to several pixels due to the simple geometric modeling of the deformation of the radar image. We have discussed these problems in Section IV.

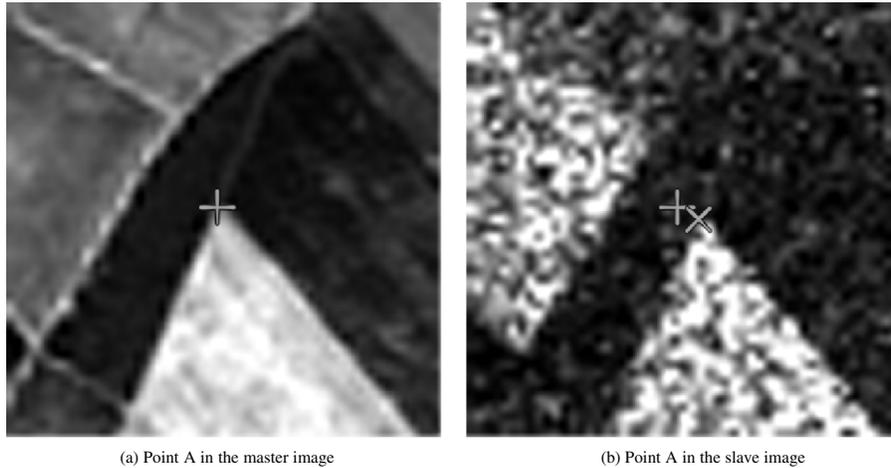


Fig. 16. Location of point A for a null displacement (+) and for the measured displacement (x).

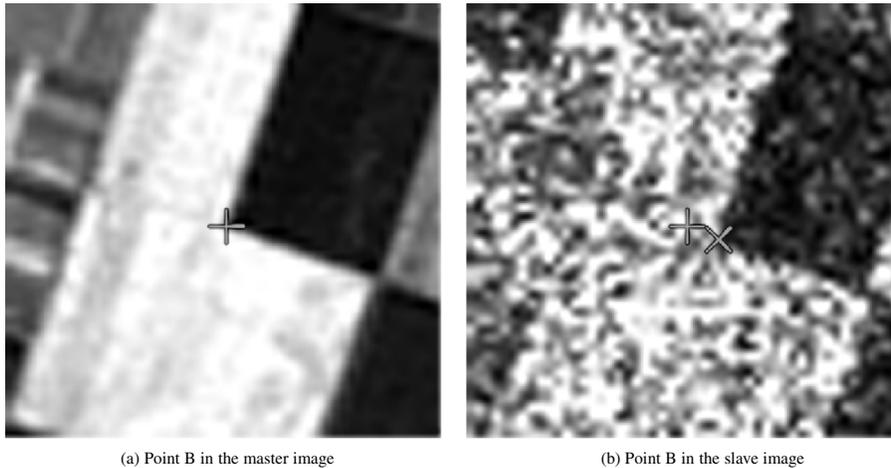


Fig. 17. Location of point B for a null displacement (+) and for the measured displacement (x).

If we analyze the DEM, we see that a gentle slope descending from northwest to southeast exists and that abrupt topography features appear in the northeast and the southwest. The shape of the river can also be identified in the DEM.

We will estimate the residual deformation between the two images with the cluster reward algorithm (CRA). We will estimate the local shifts (lines and columns) every ten pixels, thus building the deformation grids. Using noninteger shifts for the estimation window, we can estimate the deformation with subpixel accuracy. The noninteger shifts are applied by interpolating the slave image using a *sinc* function weighted by a Gaussian function whose variance is chosen so as to keep 90% of the energy in a filter with a length of 13 samples.

First of all, we will use an estimation window of size 51×51 pixels and an exploration area of ± 4 pixels around each pixel. The map of the measured deformations is shown in Fig. 13.

Applying quality criteria to the peak of the measure (value and shape), we find that 70% of the measured points are considered as valid and that 24% of the remaining pixels are not valid because the peak is in the limit of the exploration area.

Analyzing these grids we can draw the following conclusions.

- The grids are noisy. This is due to the size of the estimation window.
- The main deformations are measured in the column directions, i.e., the direction for which there is a stereoscopic effect between the two acquisitions (ERS-2 and SPOT 4 have polar orbits).
- The measured deformations are strongly correlated to the topography, mainly near the river area and at the two abrupt changes in the northeast and the southwest areas. This shows the limitations of the analytical models for the topographic deformations.

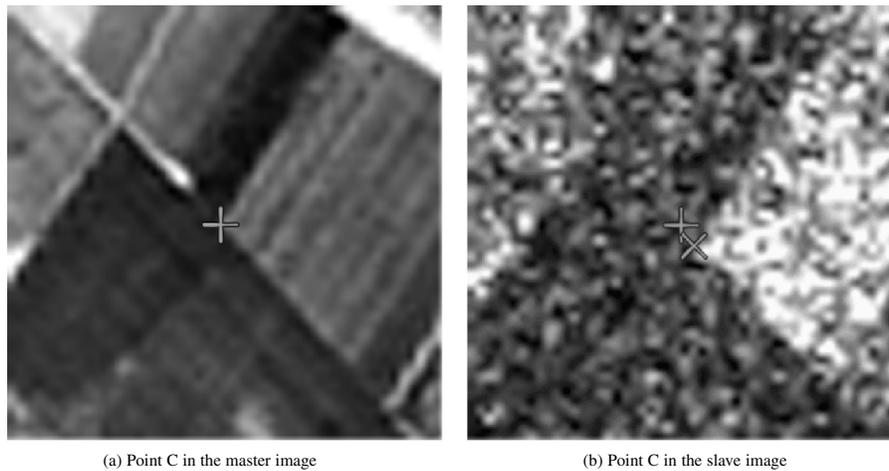


Fig. 18. Location of point C for a null displacement (+) and for the measured displacement (x).

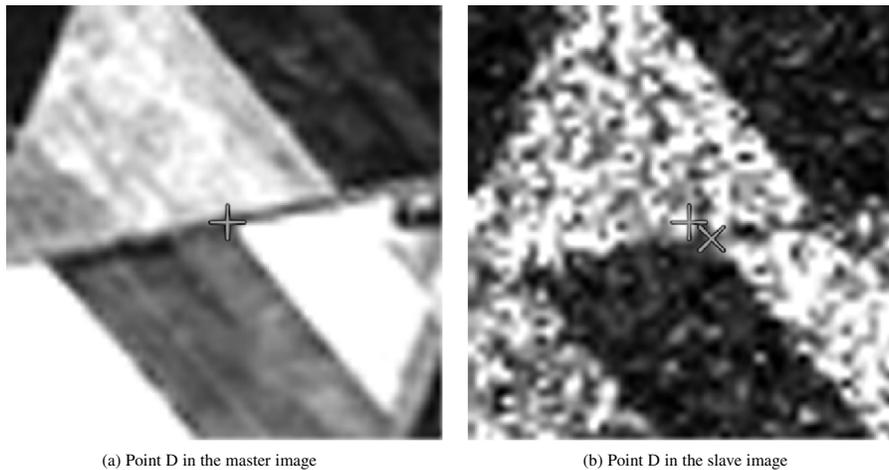


Fig. 19. Location of point D for a null displacement (+) and for the measured displacement (x).

- Low-frequency gradients of small amplitude appear on both grids. This may be due to a bad choice of the degree of the polynomial modeling.

In order to validate the assumption that links the estimation window size to the noisy aspect of the grid, we perform a new test with a window size of 81×81 pixels. We also increase the exploration area to ± 8 pixels in order to increase the number of valid points. We obtain the following results: 78% of valid points; 21% of the points have the peak in the limit of the exploration area. The deformation grids are shown in Fig. 14. We observe a decrease of the noise.

A. Validation of the Results

In order to validate the displacements obtained in the deformation grids, we have used three different approaches.

- 1) We select a set of points in the master (SPOT 4) image, and we manually find the HP in the slave (ERS-2) image and compare the obtained displacements with the ones measured in the grids. This method compares the result of the automatic processing with a manual HP selection.
- 2) For the same set of points of the master image, we point out the two points in the slave image: the point with the same coordinates (zero-shift) and the point that results in applying the shift measured in the grid. This method shows the relative improvement in registration with respect to the original image pair.
- 3) We resample the slave image using the measured grids and visually check the quality of the registration before and after resampling.

The set of test points for the first and second validation methods are shown in Fig. 15. The results of method 1 are

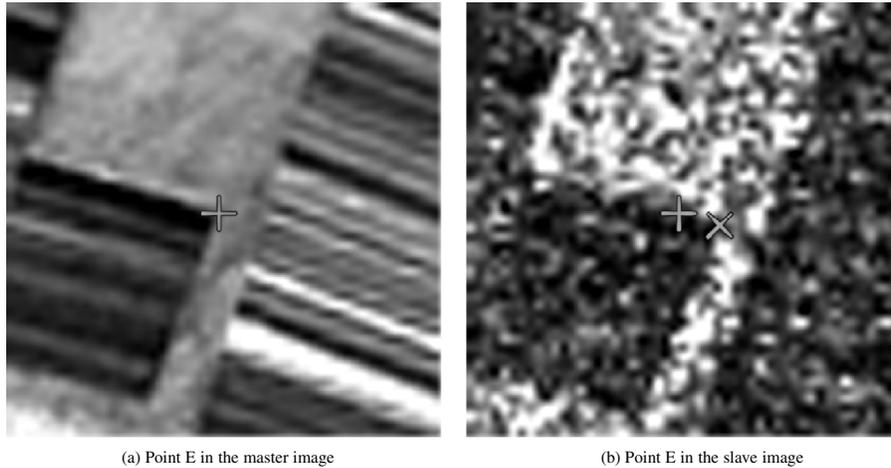


Fig. 20. Location of point E for a null displacement (+) and for the measured displacement (×).

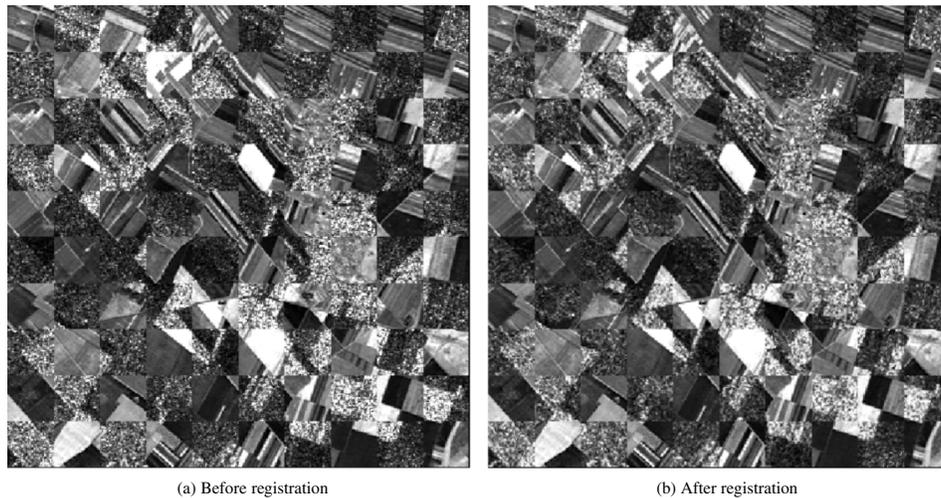


Fig. 21. Checkerboard visualization of the registration.

shown in Table V where, for each test point, we give the following:

- its coordinates (line and column) in both (ERS and SPOT) images as obtained by manual point selection;
- the deformation as a result of the difference of coordinates above;
- the deformation computed by the automatic algorithm;
- and the deformation error, i.e., the difference between manual and automatic approaches.

We see that the automatic processing yields results that are very close to manual operation. Indeed, the Euclidean distance between manual and computed deformations is always less than 0.65 pixels. One has to bear in mind that the manual measure is not perfect, so this can be considered a very good result.

The results of the second method are shown on Figs. 16–20. One can see that for every point, the measured displacement gives a better relative position of the studied point than the null displacement.

Fig. 21 shows the results of the third validation method. One can see that the resampling of the radar image using the estimated deformation grids [Fig. 21(b)] leads to a better registration of the images [Fig. 21(a)].

VII. DEM ESTIMATION

Besides the image registration approach, the results of the previous section allow for the estimation of topography using an optical-radar image pair. In this section, we show how this could be done using a single satellite. This procedure is given here as

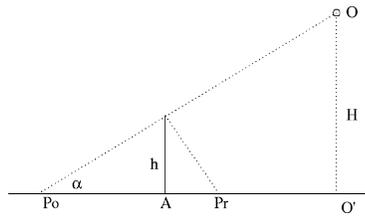
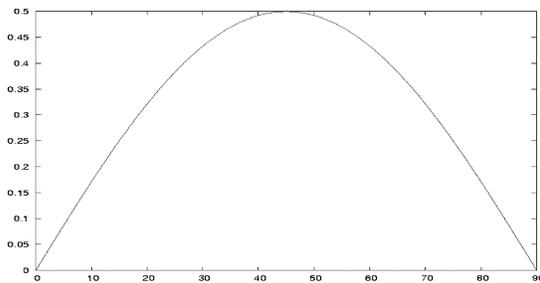


Fig. 22. DEM estimation from a single platform.

Fig. 23. Factor $\sin 2\alpha/2$ for α in degrees.

an illustration and should not be considered as a substitution of parametric sensor models when they are available.

Fig. 22 shows a geometric model for the problem. Both sensors (optical and radar) are onboard of the same satellite at point O whose vertical projection to the ground is point O' , and H is the height of the orbit. Both image acquisitions are simultaneous and the incidence angle is α . We want to measure the height h of point A . We can locally make the assumption of flat earth.

In the optical image, the point A is actually seen at point Po , while it is projected into point Pr in the radar image, if we assume a flat wavefront. If we call $\Delta x = \overline{APo} + \overline{APr}$ the measured shift of the point between both images, we obtain

$$h = \Delta x \frac{\tan \alpha}{\tan^2 \alpha + 1} = \Delta x \frac{\sin 2\alpha}{2}. \quad (31)$$

That means that the height of the point is proportional to the measured shift. It is interesting to note that it does not depend neither on any baseline between the two instruments nor on their altitude. Fig. 23 shows the dependency of the proportionality factor with the incidence angle. Typically, for optical sensors $\alpha > 60^\circ$ and for radar acquisitions, $\alpha < 70^\circ$. That means that the proportionality factor will be between 0.3–0.45.

In the case where the acquisitions are made with different incidence angles α_1 and α_2 (different platforms, for example), (31) becomes

$$h = \Delta x \frac{\tan \alpha_1}{\tan \alpha_1 \tan \alpha_2 + 1}. \quad (32)$$

VIII. CONCLUSION

We have formalized the problem of remote sensing image registration. We also have described which are the limitations of

the correlation coefficient in the multisensor case, and we have shown how other similarity measures can be used to extend the correlation coefficient approach to the multisensor problem.

These measures have been widely used in the medical imaging field, but they have always been used together with analytical models and global similarity research. We have shown that these approaches are not enough in the case of remote sensing.

A simple case study with an optical-radar pair has allowed us to show that it is possible to automatically measure high-frequency deformations between multisensor images by using a local similarity estimation leading to the estimate of deformation grids of subpixel accuracy. This result opens the possibility of automatic optical-radar DEM estimation from a single satellite.

ACKNOWLEDGMENT

The authors would like to thank their colleagues at CNES/QTIS, for all the fruitful discussions about the fine-image registration problem.

REFERENCES

- [1] C. Kuglin and D. Hines, "The phase correlation image alignment method," in *Proc. IEEE Conf. Cybernetics Soc.*, 1975, pp. 163–165.
- [2] V. N. Dvorchenko, "Bounds on (deterministic) correlation functions with applications to registration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, pp. 206–213, Feb. 1983.
- [3] L. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, 1992.
- [4] J. Ton and A. K. Jain, "Registering Landsat images by point matching," *IEEE Trans. Geosci. Remote Sensing*, vol. 27, pp. 642–651, Sept. 1989.
- [5] H. Li, B. Manjunath, and S. K. Mitra, "A contour-based approach to multisensor image registration," *IEEE Trans. Image Processing*, vol. 4, pp. 320–334, Mar. 1995.
- [6] A. Cracknell and K. Paithoonwattanakit, "Fast automatic registration for satellite imagery by heuristic planning," *Proc. Inst. Elect. Eng.*, vol. 136, no. 5, pp. 221–225, 1989.
- [7] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase-correlation to sub-pixel registration," *IEEE Trans. Image Processing*, vol. 11, pp. 188–200, Mar. 2002.
- [8] H. Stone, M. T. Orchard, E.-C. Cheng, and S. A. Martucci, "A fast direct fourier-based algorithm for subpixel registration of images," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 2235–2243, Oct. 2001.
- [9] J. Inglada and F. Adragna, "Automatic multi-sensor image registration by edge matching using genetic algorithms," in *Proc. IGARSS*, 2001.
- [10] O. Thépaut, K. Kpalma, and J. Ronsin, "ERS SAR and SPOT images automatic registration in a multichannel consensual segmentation scheme," in *Proc. IGARSS*, 1998.
- [11] Y. Wu and H. Maître, "Registration of a SPOT image and a SAR image using multiresolution representation of a coastline," in *Proc. ICPR*, 1990, pp. 913–917.
- [12] A. Ventura, A. Rampini, and R. Schettini, "Image registration by recognition of corresponding structures," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 305–314, May 1990.
- [13] X. Dai and S. Khorram, "An feature-based image registration algorithm using improved chain-code representation combined with invariant moments," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 2351–2362, Sept. 1999.
- [14] J. Djamdji, A. Bijaoui, and R. Maniere, "Geometrical registration of images: The multiresolution approach," *Photogramm. Eng. Remote Sens. J.*, vol. 59, no. 5, pp. 645–653, May 1993.
- [15] J. Le Moigne, A. Cole-Rhodes, R. Eastman, K. Johnson, J. Morissette, N. Netanyahu, H. Stone, and I. Zavorin, "Earth science imagery registration," in *Proc. IGARSS*, 2003.

- [16] J. Townshend, C. Justice, C. Gurney, and J. McManus, "The impact of misregistration on change detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 30, pp. 1054–1060, Sept. 1992.
- [17] J. Inglada, "Similarity measures for multisensor remote sensing images," in *Proc. IGARSS*, 2002.
- [18] M. Bro-Nielsen, "Medical image registration and surgery simulation," Ph.D. dissertation, Tech.Univ. Denmark, Lyngby, Denmark, 1996.
- [19] D. Sarrut, "Recalage multimodal et plate-forme d'imagerie médicale à accès distant," Ph.D. dissertation, Univ. Lumière Lyon 2, Lyon, France, 2000.
- [20] P. Thévenaz, U. Ruttimann, and M. Unser, "Iterative multi-scale registration without landmarks," *IEEE Trans. Image Processing*, vol. 7, pp. 27–41, Jan. 1998.
- [21] E. de Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 700–703, May 1987.
- [22] A. Apicella, J. Kippenhan, and J. Nagel, "Fast multi-modality image matching," *Proc. SPIE*, vol. 1092, pp. 252–263, 1989.
- [23] R. Woods, S. Cherry, and J. Mazziota, "Rapid automated algorithm for aligning and reslicing PET images," *J. Comput. Assist. Tomogr.*, vol. 16, no. 4, pp. 620–633, 1992.
- [24] R. Woods, J. Mazziota, and S. Cherry, "MRI-PET registration with automated algorithm," *J. Comput. Assist. Tomogr.*, vol. 17, no. 4, pp. 536–546, 1993.
- [25] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," in *Proc. MICCAI*, Cambridge, MA, Oct. 1998, pp. 1115–1124.
- [26] I. Csiszar, "Information type measures of difference of probability distributions and indirect observations," *Studia Sci. Math. Hungar.*, vol. 2, pp. 299–318, 1967.
- [27] D. Sarrut and S. Miguet, "Similarity measures form image registration," in *Proc. Eur. Workshop on Content-Based Multimedia Indexing*, Toulouse, France, Oct. 1999, pp. 263–270.



Jordi Inglada received the telecommunications engineer degree from both Universitat Politècnica de Catalunya, Catalunya, Spain, and the École Nationale Supérieure des Télécommunications de Bretagne, Brest, France, in 1997, and the Ph.D. degree in signal processing and telecommunications from Université de Rennes 1, Rennes, France, in 2000.

He has been since been with the Centre National d'Études Spatiales, Toulouse, France, working in the field of remote sensing image processing. He is in charge for the development of image processing algorithms for the operational exploitation of earth observation images, mainly in the fields of image registration, change detection, and object recognition.



Alain Giros received the physics engineer degree from École Nationale Supérieure de Physique de Marseille, Marseille, France, in 1982, with a specialization in optics and image processing.

He has since been with the Centre National d'Études Spatiales (CNES), Toulouse, France. He began his carrier in the SPOT 1 project team where he was responsible of the whole image processing software. Then, he was responsible for the development of several image processing systems, including DALI (the worldwide catalog of SPOT images), several image quality assessment systems for SPOT, and the Vegetation User Services Facility. Interested in automated image registration since the mid-1990s, he has been promoting this research subject at CNES. He is now in charge of the development of image processing techniques mainly in the field of image time series processing, automated registration, and image information mining.

Analysis of Artifacts in Subpixel Remote Sensing Image Registration

Jordi Inglada, Vincent Muron, Damien Pichard, and Thomas Feuvrier

Abstract—Subpixel accuracy image registration is needed for applications such as digital elevation model extraction, change detection, pan-sharpening, and data fusion. In order to achieve this accuracy, the deformation between the two images to be registered is usually modeled by a displacement vector field which can be estimated by measuring rigid local shifts for each pixel in the image. In order to measure subpixel shifts, one uses image resampling. Sampling theory says that, if a continuous signal has been sampled according to the Nyquist criterion, a perfect continuous reconstruction can be obtained from the sampled version. Therefore, a shifted version of a sampled signal can be obtained by interpolation and resampling with a shifted origin. Since only a sampled version of the shifted signal is needed, the reconstruction needs only to be performed for the new positions of the samples, so the whole procedure comes to computing the value of the signal for the new sample positions. In the case of image registration, the similarity between the reference image and the shifted versions of the image to be registered is measured, assuming that the maximum of similarity determines the most likely shift. The image interpolation step is thus performed a high number of times during the similarity optimization procedure. In order to reduce the computation cost, approximate interpolations are performed. Approximate interpolators will introduce errors in the resampled image which may induce errors in the similarity measure and therefore produce errors in the estimated shifts. In this paper, it is shown that the interpolation has a smoothing effect which depends of the applied shift. This means that, in the case of noisy images, the interpolation has a denoising effect, and therefore, it increases the quality of the similarity estimation. Since this blurring is not the same for every shift, the similarity may be low for a null shift (no blurring) and higher for shifts close to half a pixel (strong blurring). This paper presents an analysis of the behavior of the different interpolators and their effects on the similarity measures. This analysis will be done for the two similarity measures: the correlation coefficient and the mutual information. Finally, a strategy to attenuate the interpolation artifacts is proposed.

Index Terms—Disparity map estimation, image registration, interpolation artifacts, similarity measures.

I. INTRODUCTION

SUBPIXEL accuracy image registration is needed for applications such as change detection [1], pan-sharpening [2], and data fusion [3]. In order to achieve this accuracy, the deformation between the two images to be registered is usually modeled by a displacement vector field which can be estimated by measuring rigid local shifts for each pixel in the image.

Manuscript received March 15, 2006; revised May 23, 2006. This work was supported by the Centre National d'Études Spatiales under Contract 4500013574/DCT094.

J. Inglada is with the Image Analysis Office, Centre National d'Études Spatiales, 31000 Toulouse, France.

V. Muron, D. Pichard, and T. Feuvrier are with the Space Division, Communications and Systèmes, 31000 Toulouse, France.

Digital Object Identifier 10.1109/TGRS.2006.882262

In order to measure subpixel shifts, image resampling is used. Sampling theory says that, if a continuous signal has been sampled according to the Nyquist criterion, a perfect continuous reconstruction can be obtained from the sampled version. Therefore, a shifted version of a sampled signal can be obtained by interpolation and resampling with a shifted origin.

Since only a sampled version of the shifted signal is needed, the reconstruction needs only to be performed for the new positions of the samples, so the whole procedure comes to computing the value of the signal for the new sample positions.

In the case of image registration, the similarity between the reference image and the shifted versions of the image to be registered is measured, assuming that the maximum of similarity determines the most likely shift. The image interpolation step is thus performed a high number of times during the similarity optimization procedure. In order to reduce the computation cost, approximate interpolations are performed. Indeed, the ideal interpolator is a sinus cardinal function, and therefore, an infinite number of samples are needed for the computation of any new sample. Several approaches exist for reducing the computation time. For instance, a truncated sinc interpolator needs only a few samples. Other interpolators like the linear one can also be used.

Approximate interpolators will introduce errors in the resampled image which may induce errors in the similarity measure and therefore produce errors in the estimated shifts.

The problem of interpolation artifacts in image registration has been studied in the case of mutual-information-based registration in the context of medical images [5]–[7]. To our knowledge, there is no equivalent work in the field of remote sensing image registration. In this case, simple parametric geometrical transformations are not good candidates for deformation modeling. Also, similarity measures other than mutual information should be studied and characterized with respect to the interpolation artifacts.

Finally, the explanations given in the literature for the artifacts observed in registration functions do not seem to be satisfactory for all similarity measures and image modalities. Therefore, a theoretical modeling of the artifacts is proposed here. In this paper, we focus on the characterization of the artifacts for different similarity measures and interpolators, and we propose a theoretical explanation of the origin of the artifacts. Some guidelines and recommendations in order to attenuate these artifacts are also given.

The paper is organized as follows. In Section II, we pose the problem of subpixel image registration. In Section III, we show the effects of the artifacts on disparity maps with empirical tests. Section IV gives a theoretical explanation for the origin of the artifacts and checks its validity on test data. Section V

TABLE I
 APPROACHES TO IMAGE REGISTRATION

Geometric model	Similarity measure	Optimization of the deformation
Physical model	None	Global
Analytical model with a priori HP	Local	Global
Analytical model without a priori HP	Global	Global
Grid	Local	Local

proposes strategies to attenuate the effects of the artifacts; and Section VI concludes the paper.

II. DISPARITY-MAP ESTIMATION

In this section, we recall the principle of disparity-map estimation used in order to achieve subpixel accuracy.

A. Problem Position

The problem of disparity-map estimation has been described in detail in [8]. We recall hereafter the main concepts involved with it.

The estimation of a disparity map between a reference image I and secondary image J can be posed as the following optimization problem:

$$\text{Arg max}_T (S_c(I, T \circ J)) \quad (1)$$

where T is a geometric transformation and S_c is a similarity measure: a scalar strictly positive function which has an absolute maximum when the two images I and J are identical in the sense of the criterion c .

In the case of image registration, the inverse of transformation T is needed for image resampling.

Depending on the type of deformation to be corrected, the model used for T will be different. For example, if the only deformation to be corrected is the one introduced by the mean attitude, a physical model for the acquisition geometry (independent of the image contents) will be enough. If the sensor is not well known, this deformation can be approximated by a simple analytical model. When the deformations to be modeled have high frequencies, analytical (parametric) models are not suitable for a fine registration. In this case, one has to use a fine sampling of the deformation, that means the use of deformation grids. These grids give, for a set of pixels of the reference image, their location in the secondary image.

We can thus conclude that the optimization of the similarity can be either global or local with a similarity measure, which can also be either local or global. All this is synthesized in Table I.

The ideal approach would consist of a registration which is locally optimized, both in similarity and deformation, in order to have the best registration quality. This is the case when deformation grids with dense sampling are used. Unfortunately, this case is the most computationally expensive, and one often uses either a low sampling rate of the grid or the evaluation of the

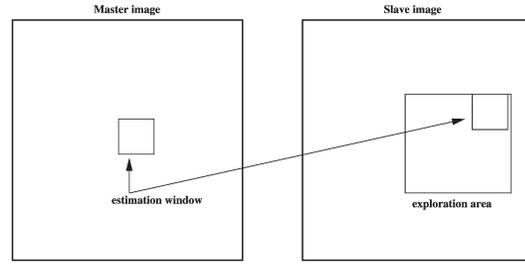


Fig. 1. Estimation of the similarity surface.

 TABLE II
 EXPRESSIONS OF FUNCTION f IN THE f -DIVERGENCE FAMILY

Measure	$f(x)$
Kolmogorov distance	$\frac{1}{2} x-1 $
Mutual information	$x \log x$
Kullback divergence	$(x-1) \log x$
χ^2 -divergence	$\frac{1}{2}(x-1)^2$
Hellinger distance	$\frac{1}{2}(\sqrt{x}-1)^2$
Toussaints distance	$x \frac{x-1}{x+1}$
Lin K-divergence	$x \log \frac{2x}{1+x}$

similarity in a small set of pixels for the estimation of an analytical model. Both of these choices lead to local registration errors which, depending on the topography, can amount several pixels.

Even if this registration accuracy can be enough in many applications (orthoregistration, import into a GIS, etc.), it may not be acceptable in the case of data fusion, multichannel segmentation, or change detection [1]. This is why we will focus on the problem of deformation estimation using dense grids.

As we have noted above, deformations due to topography can locally have high frequencies for medium- and high-resolution sensors (30 m and better), thus our need for fine modeling.

B. Estimation Procedure

The geometric deformation is modeled by local rigid displacements [8]. One wants to estimate the coordinates of each pixel of the reference image inside the secondary image. This can be represented by a displacement vector associated with every pixel of the reference image. Each of the two components (lines and columns) of this vector field will be called deformation grid.

We use a small window taken in the reference image, and we test the similarity for every possible shift within an exploration area inside the secondary image (Fig. 1). That means that, for each position, we compute the similarity measure. The result is a similarity surface whose maximum gives the most likely local shift between both images.

Quality criteria can be applied to the estimated maximum in order to give a confidence factor to the estimated shift: width of the peak, maximum value, etc. Subpixel shifts can be measured by applying fractional shifts to the sliding window. This is done by image interpolation.

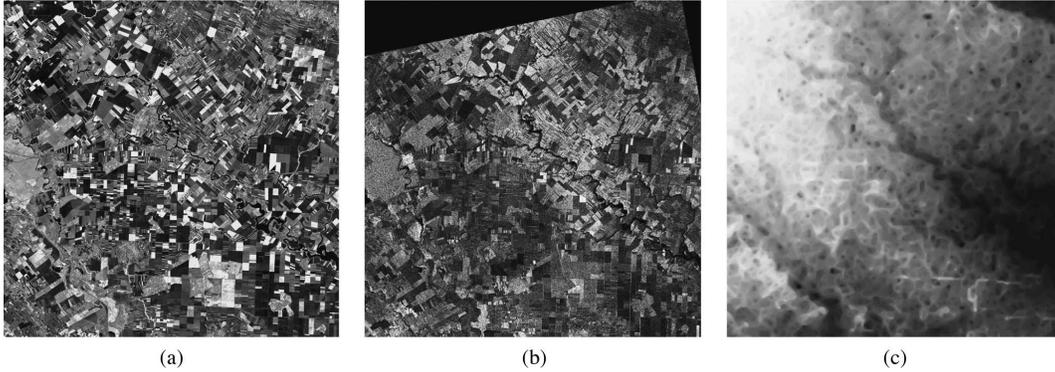


Fig. 2. Images and DEM for the test area. (a) Spot 4 B3. (b) ERS-2 SAR. (c) DEM.

The interesting parameters of the procedure are the following.

- 1) The size of the exploration area: it determines the computational load of the algorithm (we want to reduce it), but it has to be large enough in order to cope up with large deformations.
- 2) The size of the sliding window: the robustness of the similarity measure estimation increases with the window size, but the hypothesis of local rigid shifts may not be valid for large windows.

C. Similarity Measures

In this paper, we will only study two similarity measures: the correlation coefficient and the mutual information. A detailed discussion on similarity measures for image registration can be found in [8].

The correlation coefficient is the most used similarity measure for image registration. We remind here its computation for two image windows I and J . The coordinates of the pixels inside the windows are represented by (x, y)

$$\rho(I, J) = \frac{1}{N} \frac{\sum_{x,y} (I(x, y) - m_I)(J(x, y) - m_J)}{\sigma_I \sigma_J}. \quad (2)$$

In this expression, N is the number of pixels of the analysis window, m_I and m_J are the estimated mean values inside the analysis window of, respectively, images I and J , and σ_I and σ_J are their standard deviations.

The mutual-information measure between two images can be understood as the amount of information we have from one image when we know the other. This is the degree of dependence between the two images. Several approaches exist for the computation of the mutual information. We prefer to introduce it as a member of the f -divergence family.

An f -divergence [9] measures the expectation of the diversity of the likelihood ratio between two distributions P and Q

$$D_f(P, Q) = E_Q \left[f \left(\frac{dp(x)}{dq(x)} \right) \right] = \int f \left(\frac{p(x)}{q(x)} \right) q(x) dx. \quad (3)$$

E_Q is the expectation with respect to Q , $dp(x)/dq(x)$ is the derivative with respect to a density, and f is continuous and convex on $[0, +\infty)$. A divergence can be seen as a relative entropy. Depending on the choice of f , different measures can be obtained. Table II shows some interesting cases.

The mutual information between images I and J is the particular case where $f(x) = x \log x$, $p(x) = p_{IJ}(i, j)$ (the joint probability density function of the two images) and $q(x) = p_I(i)p_J(j)$ (the product of probability density functions of images I and J). It can therefore be computed as follows:

$$MI(I, J) = \sum_{ij} p_{IJ}(i, j) \log \frac{p_{IJ}(i, j)}{p_I(i)p_J(j)}. \quad (4)$$

The sum is computed over every couple of pixel values (i, j) .

III. ASSESSMENT OF THE ARTIFACTS

In this section, we introduce the problem of subpixel shift artifacts by analyzing the results obtained in a real case.

Our data set is consist of the following pair (a region of 2000×2000 pixels is used for our tests):

- 1) B3 channel of a SPOT 4 image (20-m pixel resolution) acquired on June 24, 2001, over the east of the Bucharest area [Fig. 2(a)];
- 2) European Remonte Sensing 2 (ERS-2) Satellite synthetic aperture radar three-look intensity image (12.5-m pixel size and approximately 20-m pixel resolution) acquired on May 10, 2001, over the same area [Fig. 2(b)].

Both images were orthorectified: For the SPOT 4 image, a digital elevation model (DEM) [Fig. 2(c)] with an altimetric accuracy better than 10 m and a planimetric accuracy around 10 m has been used together with the acquisition model (orbits, attitude) for the satellite; for the ERS-2 image, no DEM was used, but a constant altitude and homologous points manually taken on the SPOT 4 image were used in the orthorectification process. Globally, the images show a good superposition, but local errors exist, which can amount several pixels due to the simple geometric modeling of the deformation of the radar image.

If we analyze the DEM, we see that a gentle slope descending from northwest and southeast exists and that abrupt topography

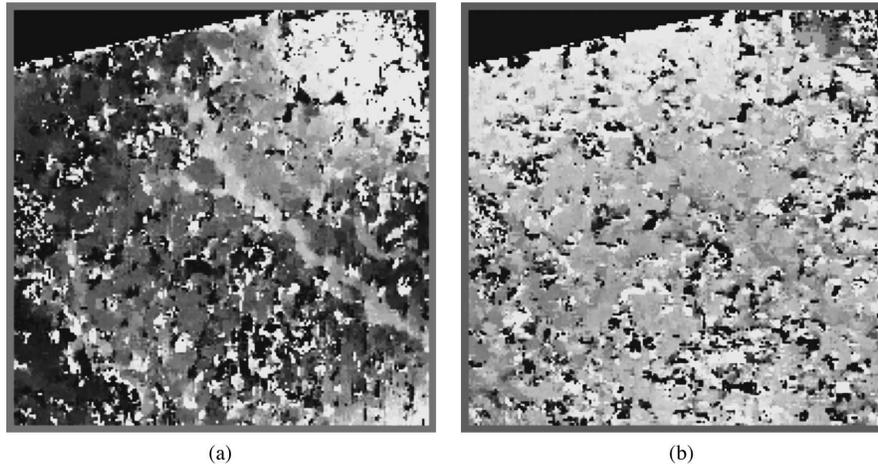


Fig. 3. Deformation grid. Mutual information: estimation window is 51×51 pixels, and sampling rate is five pixels. (a) Horizontal. (b) Vertical.

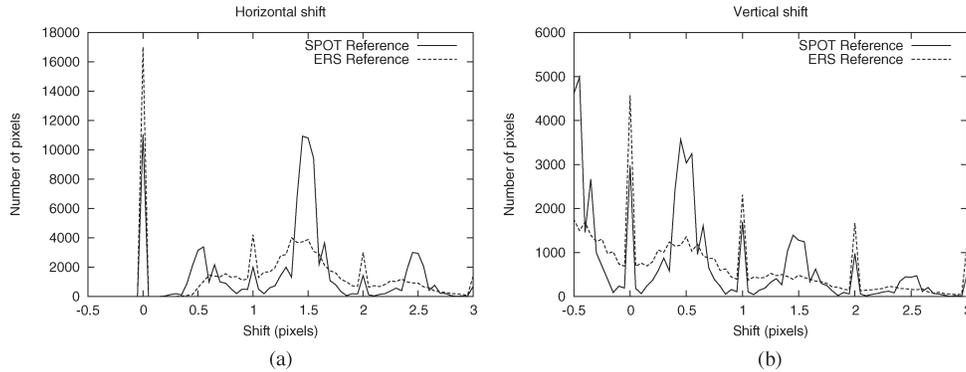


Fig. 4. Histograms of the estimated subpixel shifts [(a) horizontal and (b) vertical] with inversion of the reference and the secondary images.

features appear in the northeast and southwest. The shape of the river can also be identified in the DEM.

Fig. 3 presents the horizontal and vertical components of the displacement vector field obtained using the mutual information similarity measure and the procedure described in Section II-B. One observes a good correlation between the horizontal component and the topography shown in Fig. 2(c). As expected, the vertical (satellite along-track) direction does not show any particular structure. When this displacement vector field is used for the registration of the images, a good superposition is achieved. The detailed analysis of the procedure was carried out in [8].

If we analyze the distribution of the estimated shifts by computing their histograms, we observe the following behavior (Fig. 4). When the SPOT image is used as the reference, a high number of estimated shifts are multiples of 0.5 pixels; if the ERS image is used as the reference, this effect is attenuated, and the shifts present a more uniform distribution.

Since the similarity measure is the same for both cases and so is the optimization procedure, one can conclude that the subpixel shifts artifacts appear when the ERS image is

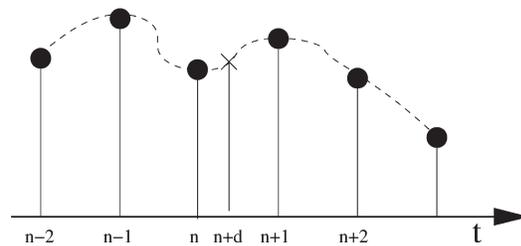


Fig. 5. Interpolation of a sampled signal.

interpolated during the similarity optimization. The following sections will study this effect in detail, and a theoretical model for the origin of the artifacts will be presented.

IV. ORIGIN OF THE ARTIFACTS

The problem of interpolation artifacts in the similarity surfaces has been studied for the case of mutual-information-based medical image registration [5]–[7]. Plum *et al.* [5] show that

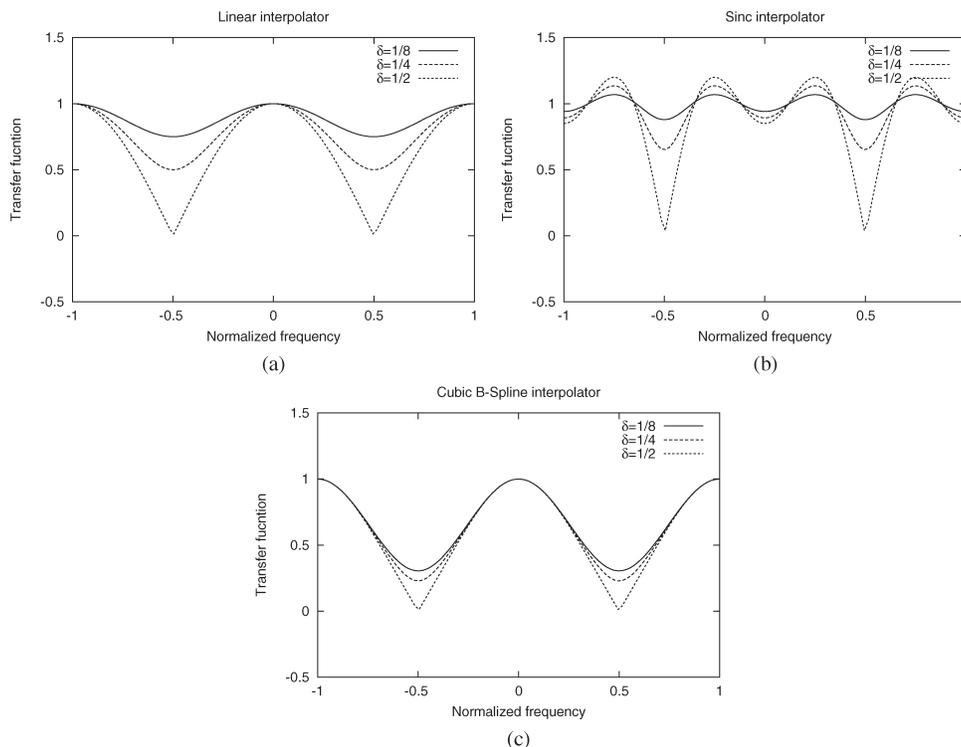


Fig. 6. Evolution of the blurring effect of the interpolators as a function of the shift. (a) Linear. (b) Truncated sinc. (c) Cubic B-spline.

entropy-based registration measures, when plotted as a function of the geometric transformation, show sudden changes. They show that this registration function depends on the interpolation method in the cases of similar sampling step for the two images to be registered and propose to resample one of the images with a slightly different sampling step. This solution can be applied in the case of global image registration where one looks for a parametric transformation (i.e., affine transformation). In this case, the resampling introduces a scaling effect which can be compensated by the geometric transformation. In the case of disparity-map estimation, the sampling step of the images has to be the same, in order to approximate the deformations by local rigid shifts.

Tsao [6] also analyzes the behavior of mutual-information surfaces for four different interpolators (nearest neighbor, linear, cubic, and Hamming-windowed sinc). He shows the influence of the number of bins used for the estimation of the histograms needed for the computation of the mutual-information measure. He proposes to blur the histograms and to introduce some jitter in the sampling step. The jitter will produce an effect which is analogous to the resampling proposed by Plum *et al.* [5]. One has to note that the blurring of the histograms can only be applied to histogram-based similarity measures. Also, the jitter of the sampling step will introduce local errors which cannot be accepted in subpixel registration.

Ji *et al.* [7] also analyze these effects and propose other strategies for reducing the artifacts: image oversampling and

intensity clustering. The image oversampling produces an effect similar to blurring, and the intensity clustering produces smoother histograms. Once again, this second solution can only be applied to histogram-based similarity measures.

Our theoretical model will show that these interpolation artifacts are not specific to the mutual-information measure, and that, therefore, their origin does not reside in the histogram estimation.

A. Theoretical Model

In this section, we show that the origin of the observed artifacts is the interpolation procedure used for the subpixel registration. In this procedure, we resample the local image patches in order to measure the similarities for different shifted positions. The resampling is performed by image interpolation. Fig. 5 illustrates the procedure. In order to obtain a shift of $\delta < 1$ pixels, we have to estimate the image gray levels at positions which lay between the samples of the image. The image to be resampled $x[n]$ is considered to be the sampled version of an ideal continuous image $x(t)$

$$x[n] = x(nT)$$

where T is the sampling step. The shifted image $y[n]$ will be obtained by sampling the same original image $x(t)$ with a shifted sampling grid. Assuming that $x(t)$ was correctly sampled (with respect to the Shannon criterion), we can retrieve

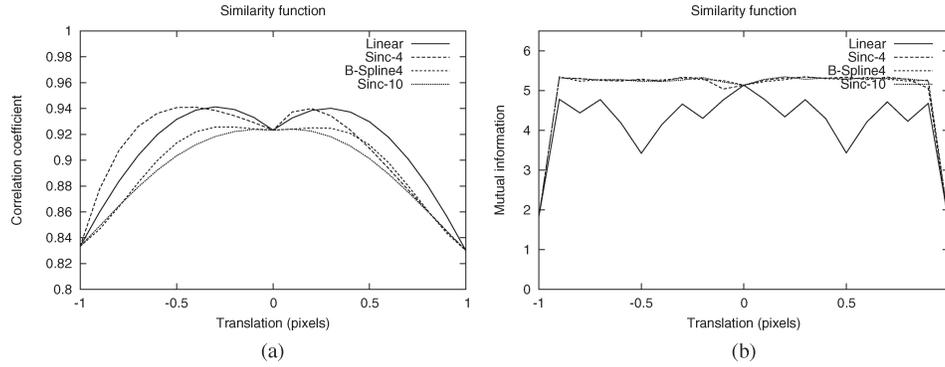


Fig. 7. Comparison of interpolators.

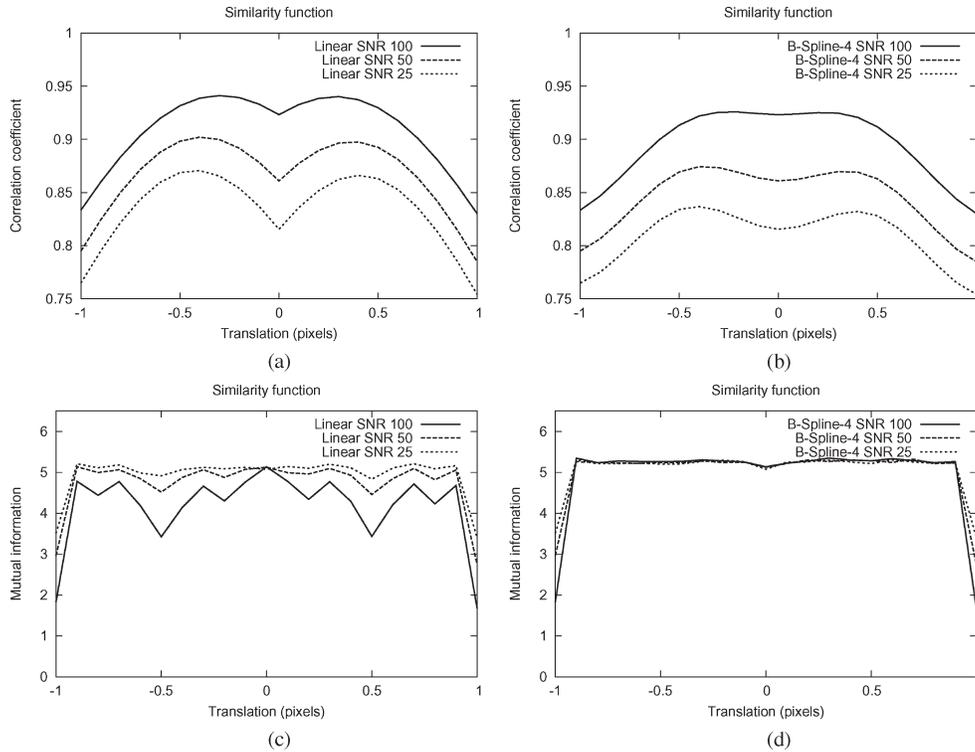


Fig. 8. Influence of noise level on the registration functions.

$x(t)$ from $x[n]$ by ideal interpolation, i.e., by using a sinc interpolator.

The sinc interpolator has an infinite impulse response. Therefore, approximate interpolators will be used. In order to increase the computation speed, we want to use interpolation filters with a low number of samples.

For a linear interpolator, the interpolated image $y(t)$ for a shift δ will take the following expression:

$$y(t) = (1 - \delta)x(t - \delta) + \delta x(t + 1 - \delta) \quad (5)$$

and its Fourier transform is

$$Y(f) = X(f) \left[(1 - \delta)e^{-j2\pi\delta f} + \delta e^{-j2\pi(\delta-1)f} \right]. \quad (6)$$

We see that the interpolated signal $y(t)$ is not exactly equal to the original signal $x(t)$ due to the fact that we are not using an ideal interpolator. Instead, we obtain a low-pass filtered version of the original signal. It is interesting to note that the blurring of

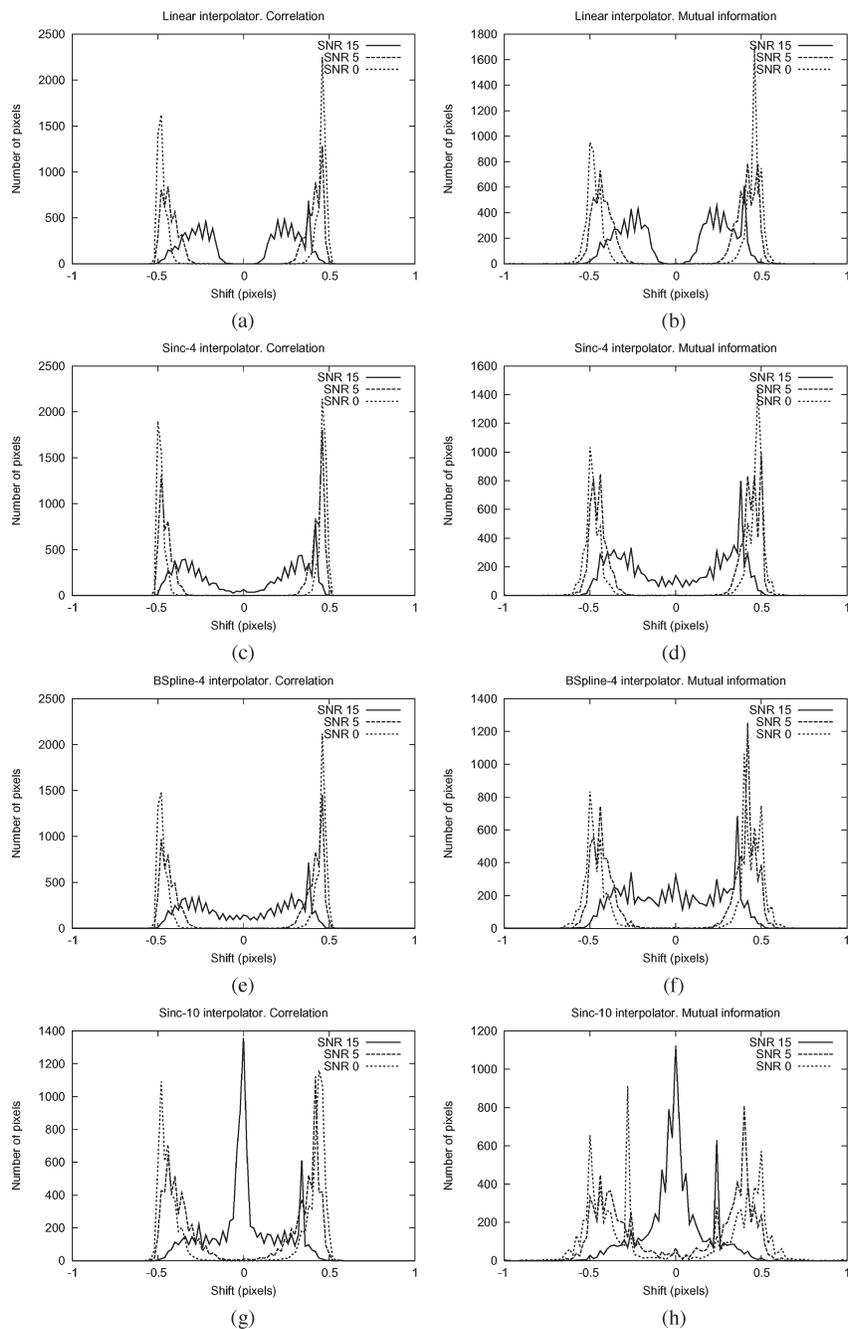


Fig. 9. Influence of noise level on the estimated shifts.

the image introduced by the interpolation depends on the shift. Fig. 6(a) shows that the blurring effect increases when the shift comes close to half a pixel ($\delta = 1/2$).

This means that, in the case of noisy images, the interpolation has a denoising effect, and therefore, it increases the quality of the similarity estimation. Since this blurring is not the same for

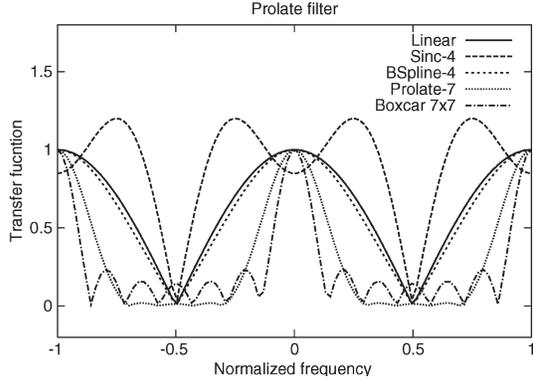


Fig. 10. Frequency response of the smoothing prolate filter compared to different interpolators for $\delta = 1/2$. The 7×7 boxcar filter is also shown.

every shift, the similarity surface may show low values for a null shift (no blurring) and higher values for shifts close to half a pixel (strong blurring). It is important to note that the artifacts do not come from the blurring effect itself, but rather from the difference of blurring effect for different shift values.

It is now interesting to analyze this effect for other interpolators. For the case of an interpolator $c(t)$ truncated to four samples, the Fourier transform of the interpolated signal takes the following expression:

$$Y(f) = X(f) \left[c(\delta+1)e^{-j2\pi(\delta+1)f} + c(\delta)e^{-j2\pi\delta f} + c(1-\delta)e^{-j2\pi(1-\delta)f} + c(2-\delta)e^{-j2\pi(2-\delta)f} \right]. \quad (7)$$

For the case of a sinc interpolator, $c(t) = \sin(\pi t)/\pi t$. The frequency response of the interpolator as a function of the shift δ is shown in Fig. 6(b). We see that the blurring effect is still dependent on the shift, but also that, for such a short filter, the continuous frequency is also filtered. This can produce effects which are worse than the linear interpolator. We will see this in the following sections.

Finally, we analyze the case of a cubic B-spline interpolator [10]. In this case, the filter coefficients take the following expression:

$$c(t) = \begin{cases} \frac{2}{3} - \frac{1}{2}|x|^2(2 - |x|), & 0 \leq |x| < 1 \\ \frac{1}{6}(2 - |x|)^3, & 1 \leq |x| < 2 \\ 0, & |x| > 2. \end{cases} \quad (8)$$

Fig. 6(c) shows the frequency response of the four-sample cubic B-spline interpolator. We can see that the blurring effect remains nearly the same for all shifts.

B. Sensitivity Analysis

We analyze here the behavior of the different interpolators and their effects on the similarity functions. The similarity function is defined as the value of the similarity measure as a function of the shift. Without loss of generality, we will apply the shifts in only one direction. In this case, the similarity function can be plotted as a one-dimensional (1-D) function.

This analysis will be done for the two similarity measures, the correlation coefficient and the mutual information.

The data used for these experiments are SPOT 4 images which are compared to a noisy version of itself. This allows us to ensure that the images are perfectly coregistered. Additive white Gaussian noise has been added to the secondary image with an SNR of 100 dB.

Fig. 7 shows the similarity functions for four different interpolators, the three studied in Section IV-A, plus a sinus cardinal of length equal to ten samples, which is a better approximation of the ideal one. Even if a similarity function shows only the behavior for a particular pixel, the examples presented here can be considered as examples of what occurs for every pixel in the image. For the case of the correlation coefficient [Fig. 7(a)], we obtain a behavior which could be predicted from the theory presented above. The linear and the sinc-4 interpolators have strong maxima close to the half-pixel shifts. We can observe that these effects are much weaker for the B-spline interpolator and that they are nearly inexistent for the sinc-10 one.

It is worth to notice that the erroneous maxima are not exactly located on the half-pixel shifts and that they are not symmetrical with respect to the null translation. This is caused by the fact that we are measuring the similarity between an image and its noisy shifted blurred version with a degree of blurring which depends on the shift. The blurring is useful for denoising and thus for increasing the similarity. On the other hand, the shift decreases the similarity because the homologous pixels are further away. Therefore, the combination of these two effects may produce a similarity maximum whose location depends on the local content of the image.

This is the case for the mutual-information plots shown on Fig. 7(b). As discussed in [8], mutual-information peaks have a higher slope than the correlation coefficient ones. That means that the effect of erroneous peaks will only appear for interpolators whose behavior is very sensitive to the shifts. Also, one could expect that the erroneous maxima will appear near the null shift. This is what can be observed in the plots. For the linear interpolator, the peaks appear for about one third of a pixel. We can also observe that, since mutual information is able to measure the dependence in the presence of noise [11], the global maximum is located at zero, even if its value is not much higher than the secondary maxima. For the case of interpolators with a more stable smoothing, one can see that there is no clear peak, meaning that the smoothing effect produces a high value of mutual information even for shifts larger than half a pixel. Also, the flatness of the measure makes it unusable in an optimization problem. Of course, the mutual-information value is low for integer pixel shifts, since no interpolation is applied in this case.

We can also analyze the influence of the noise level on the similarity functions. Fig. 8 shows the behavior of the linear and the B-spline interpolators for the correlation coefficient and the mutual information for different SNR of the secondary image.

In the case of the correlation coefficient, we observe that the relative value of the erroneous peaks with respect to the value at zero is higher for a lower SNR. Of course, the absolute value of the peaks is lower. We also see that the B-spline interpolator is more robust to high noise levels. In the case of mutual information, we observe that the similarity function becomes

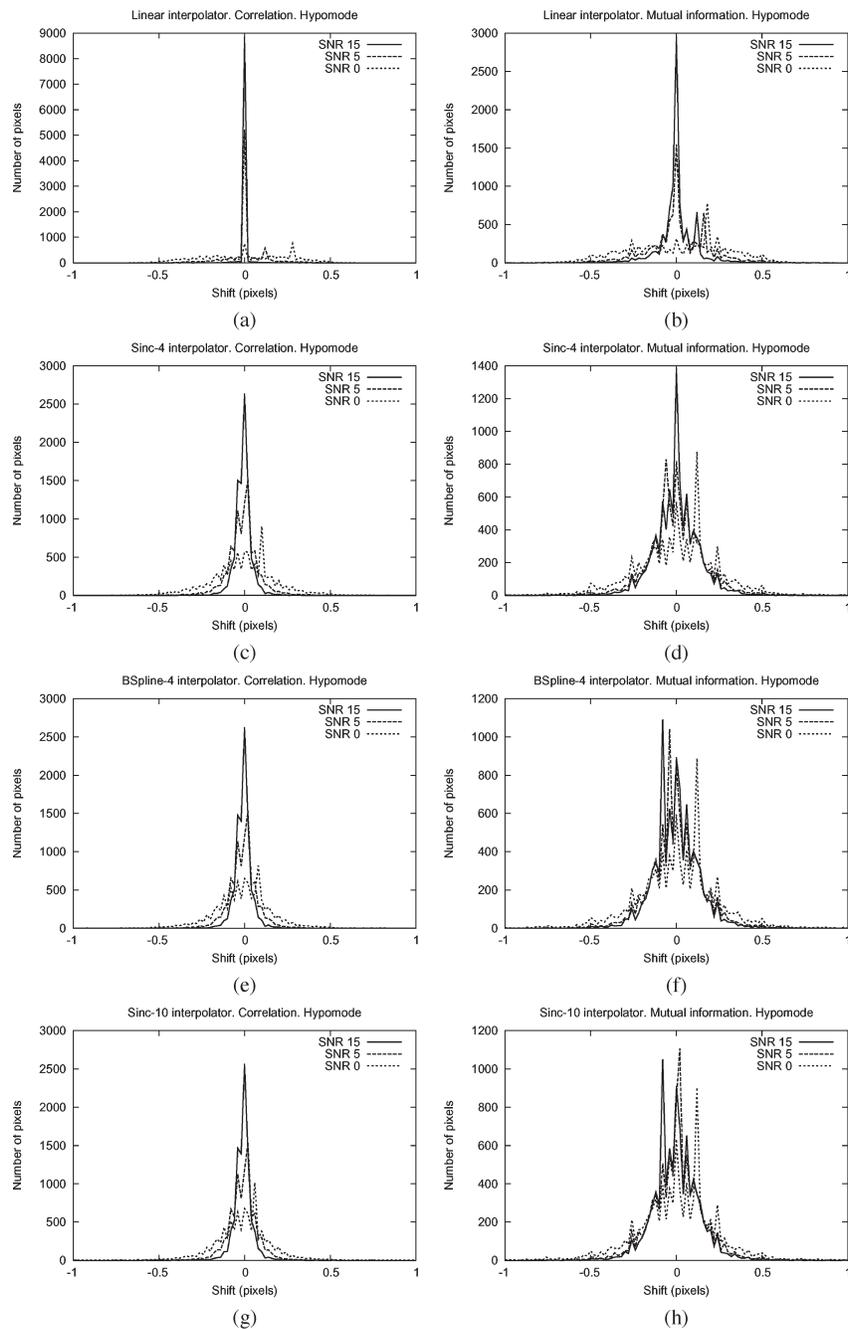


Fig. 11. Influence of low-pass filtering on the estimated pixel shifts for different noise levels.

flat when the noise increases [Fig. 8(c)]. This is coherent with what we saw in Fig. 7(b) for the most stable interpolators and is also coherent with Fig. 8(d).

Since the registration functions of Figs. 7 and 8 show only the behavior of a selected pixel of the image, it is difficult to infer the global quality of the registration from them. In order

to study the global quality, we will analyze the histograms of the estimated shifts. We will study the different combinations of interpolators (linear, sinc-4, cubic B-spline, and sinc-10), similarity measures (correlation coefficient and mutual information), and noise level. The results are shown in Fig. 9.

As for the previous simulations, 1-D shifts have been applied. In terms of noise influence, one observes that, the higher the SNR, the lower the number of shifts at a multiple of 0.5 pixels. We also observe that, when the SNR increases, the peaks move close to the null shift.

If we compare the interpolators for a given SNR, say 15 dB, we see that the better the interpolator (linear is worst, then sinc-4, B-spline, and sinc-10 is the best), the higher the number of pixels for which the estimated shift is close to zero, which is the expected value. This is true for both similarity measures.

V. ATTENUATION OF THE ARTIFACTS

As it has been stated above, we are interested in using short interpolating filters, since the interpolation is performed a high number of times during the similarity optimization procedure. As we have shown above, the interpolation artifacts are produced by the blurring effects of the interpolators. More precisely, the origin of the artifacts is not the blurring effect itself but rather the difference of blurring intensity as a function of the applied shift. We have shown, for instance, that, even if the B-spline interpolator has a stronger blurring effect than the linear interpolator, since its blurring remains nearly constant for all shifts, it has better performances for the disparity-map estimation.

The strategy proposed here for reducing the interpolation artifacts is a very simple one. Since the interpolator is going to introduce a blurring effect, we can smooth the secondary image with a filter whose transfer function is identical (in modulus) to the maximum blurring effect of the interpolator. This can be done in a preprocessing step.

However, when observing Fig. 6, we see that the evolution of the blurring effect may not be only related to the highest frequencies of the signal, and therefore, selecting the transfer function of the preprocessing filter could be tricky.

For instance, choosing a simple boxcar filter for preprocessing can produce artifacts introduced by the secondary lobes of the filter. These lobes come from the windowing used for the truncation of the filter's impulse response.

In order to study the improvement of the subpixel shift estimation for the different interpolators, we choose to use the same smoothing filter for all of them. In order to reduce the secondary lobes of the smoothing filter and to assure a short impulse response, we propose to use a prolate function [12], [13]. The prolate filter is one class of the nonrecursive finite impulse response filters. It is superior to other filters in this class in that it has a maximum energy concentration in the frequency passband and minimum ringing in the time domain.

A prolate filter with seven samples is shown in Fig. 10 and is compared to the maximum smoothing for several interpolators. The frequency response of the 7×7 boxcar filter, with its secondary lobes, is also shown.

Fig. 11 shows the same kind of analysis as Fig. 9, but with the use of the prolate filter as a preprocessing step for the secondary

image. The first remark we can make is that the peaks at multiples of 0.5 pixels have vanished for both similarity measures and for all interpolators. The behavior of the correlation is always better than that of the mutual information.

We also see that, for the high SNR values, the best results are obtained for the linear interpolator. However, for low SNR values, the better the interpolator, the better are the estimated shifts. We can also see that, for the mutual-information case, good results are only obtained with the linear and the sinc-4 interpolators, and for the highest SNR value.

VI. CONCLUSION

This paper has presented the problem of interpolation-induced artifacts in the procedure of disparity-map estimation used for subpixel image registration. The problem has been introduced with a real case, where the presence of wrongly estimated shifts when a radar image is interpolated have been shown.

A theoretical explanation of the origin of the artifacts has been given, and it demonstrated that the blurring effect of the interpolator, which is dependent on the applied shift, is responsible for the errors observed in the registration functions.

Several interpolators have been compared under different SNR conditions. Finally, it has been shown that a preprocessing step which smoothes the secondary interpolated image can solve the problem. However, attention has to be paid to the choice of the smoothing filter. Indeed, simple filters, as the boxcar one, have to be avoided since they present secondary lobes for the frequencies where the interpolation artifacts occur.

Even if the solution presented here allows for an improvement of the estimation of subpixel-accuracy disparity maps, more work has to be done in order to:

- 1) analyze the effects of the interpolation for other image modalities, as infrared data, for example;
- 2) study the influence of image resolution and type of landscape in the quality of the estimated shifts;
- 3) propose edge-preserving smoothing filters which, combined with different interpolators, could help to attenuate the interpolation artifacts at the same time that they preserve high-frequency content which can be useful for a precise disparity estimation.

REFERENCES

- [1] J. Townshend, C. Justice, C. Gurney, and J. McManus, "The impact of misregistration on change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 1054–1060, Sep. 1992.
- [2] G. Cliche, F. Bonn, and P. Teillet, "Integration of the SPOT Pan channel into its spectral mode for image sharpness enhancement," *Photogramm. Eng. Remote Sens.*, vol. 51, pp. 311–316, Mar. 1985.
- [3] P. Chavez, S. Sides, and J. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: TM & SPOT pan," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, 1991.
- [4] T. Toutin, "Multi-source data fusion with an integrated and unified geometric modeling," *EARSel Adv. Remote Sens.*, vol. 4, no. 2, pp. 118–129, 1995.
- [5] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Interpolation artefacts in mutual information-based image registration," *Comput. Vis. Image Underst.*, vol. 77, no. 2, pp. 211–232, Feb. 2000.
- [6] J. Tsao, "Interpolation artifacts in multimodality image registration based on maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 22, no. 7, pp. 854–864, Jul. 2003.
- [7] J. X. Ji, H. Pan, and Z.-P. Lang, "Further analysis of interpolation effects in mutual information-based image registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1119–1131, Sep. 2003.

- [8] J. Inglada and A. Giros, "On the possibility of automatic multisensor image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2104–2120, Oct. 2004.
- [9] I. Csiszar, "Information type measures of difference of probability distributions and indirect observations," *Studia Sci. Math. Hungar.*, no. 2, pp. 299–318, 1967.
- [10] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: Part I—Theory," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 821–832, Feb. 1993.
- [11] J. Inglada, "Similarity measures for multisensor remote sensing images," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Toronto, ON, Canada, Jun. 24–28, 2002, vol. 1, pp. 104–106.
- [12] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case," *Bell Syst. Tech. J.*, vol. 57, no. 5, pp. 1371–1429, May 1978.
- [13] J. D. Mathews, J. K. Breakall, and G. K. Karawas, "The discrete prolate spheroidal filter as a digital signal processing tool," *IEEE Trans. Acoust., Speech Signal Process.*, vol. ASSP-33, no. 6, pp. 1471–1478, Dec. 1985.



Jordi Inglada received the Telecommunications Engineer degree from both the Universitat Politècnica de Catalunya, Barcelona, Spain, and the École Nationale Supérieure des Télécommunications de Bretagne, Brest, France, in 1997 and the Ph.D. degree in signal processing and telecommunications from the Université de Rennes 1, Rennes, France, in 2000.

He has been working at the Centre National d'Études Spatiales, Toulouse, France, in the field of remote sensing image processing. He is in charge for

the development of image processing algorithms for the operational exploitation of Earth observation images, mainly in the fields of image registration, change detection, and object recognition. He hopes that one day society will be able to fully benefit from the amazing amount of information contained in remote sensing images.



Vincent Muron received the D.E.S.S. degree in remote sensing and digital image processing from Université Paul Sabatier, Toulouse, France, after a degree in physics, in 1998 and the M.Sc. degree in signal and image processing from the École Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique d'Hydraulique et des Télécommunications, Toulouse, in 2000.

Since 2000, he has been working with the Image Processing Office, Communication et Systèmes, Toulouse. His main interests include image registration and image segmentation and classification.



Damien Pichard received the Engineer degree from the École Spéciale de Mécanique et d'Electricité, Paris, France, in 1997 and the Master degree in image and signal processing from the École Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique d'Hydraulique et des Télécommunications, Toulouse, France, in 1998.

Since 2000, he has been working with Communication et Systèmes, Toulouse, as a member of the Image Processing Office in the Space Division. He has worked on image compression, image segmentation and classification, and image registration and interpolation. He is currently Responsible of the MEDICIS software development and distribution. This software computes the disparity between pairs of images. This tool is specified by the Centre National d'Études Spatiales (CNES), and it is used in the SPOT image quality processing chain by the CNES.



Thomas Feuvrier received the D.E.S.S. degree in remote sensing and digital image processing from Université Paul Sabatier, Toulouse, France, in 1998.

Since 2000, he has been working with Communication et Systèmes, Toulouse, in the field of image processing for space, industrial, and defense applications (image quality assessment, similarity measures, and image registration). At present, he takes an important role in the development of the ORFEO Toolbox, the CNES open source software library for high-resolution remote sensing image processing.

14 Détection de changements

Bivariate Gamma Distributions for Image Registration and Change Detection

Florent Chatelain, Jean-Yves Tourneret, *Member, IEEE*, Jordi Inglada, and André Ferrari

Abstract—This paper evaluates the potential interest of using bivariate gamma distributions for image registration and change detection. The first part of this paper studies estimators for the parameters of bivariate gamma distributions based on the maximum likelihood principle and the method of moments. The performance of both methods are compared in terms of estimated mean square errors and theoretical asymptotic variances. The mutual information is a classical similarity measure which can be used for image registration or change detection. The second part of the paper studies some properties of the mutual information for bivariate Gamma distributions. Image registration and change detection techniques based on bivariate gamma distributions are finally investigated. Simulation results conducted on synthetic and real data are very encouraging. Bivariate gamma distributions are good candidates allowing us to develop new image registration algorithms and new change detectors.

Index Terms—Correlation coefficient, image change detection, image registration, maximum likelihood, multivariate gamma distributions, mutual information.

I. INTRODUCTION

THE univariate gamma distribution is uniquely defined in many statistical textbooks. However, extensions defining multivariate gamma distributions (MGDs) are more controversial. For instance, a full chapter of [1] is devoted to this problem (see also references therein). Most journal authors assume that a vector $\mathbf{x} = (x_1, \dots, x_d)^T$ is distributed according to an MGD if the marginal distributions of x_i are univariate gamma distributions. However, the family of distributions satisfying this condition is very large. In order to reduce the size of the family of MGDs, S. Bar Lev, and P. Bernardoff recently defined MGDs by the form of their moment generating function or Laplace transforms [2], [3]. The main contribution of this paper is to evaluate these distributions as candidates for image registration and change detection.

Given two remote sensing images of the same scene I , the reference, and J , the secondary image, the registration problem can be defined as follows: determine a geometric transformation

T which maximizes the correlation coefficient between image I and the result of the transformation $T \circ J$. A fine modeling of the geometric deformation is required for the estimation of the coordinates of every pixel of the reference image inside the secondary image. The geometric deformation is modeled by local rigid displacements [4]. The key element of the image registration problem is the estimation of the correlation coefficient between the images. This is usually done with an estimation window in the neighborhood of each pixel. In order to estimate the local rigid displacements with a good geometric resolution, one needs the smallest estimation window. However, this leads to estimations which may not be robust enough. In order to perform high-quality estimations with a small number of samples, we propose to introduce *a priori* knowledge about the image statistics. In the case of power radar images, it is well known that the marginal distributions of pixels are gamma distributions [5]. Therefore, MGDs seem good candidates for the robust estimation of the correlation coefficient between radar images.

The change detection problem can be defined as follows. Consider two co-registered synthetic aperture radar (SAR) intensity images I and J acquired at two different dates t_I and t_J . Our objective is to produce a map representing the changes occurred in the scene between time t_I and time t_J . The final goal of a change detection analysis is to produce a binary map corresponding to the two classes: *change* and *no change*. The problem can be decomposed into two steps: 1) generation of a change image and 2) thresholding of the change image in order to produce the binary change map. The overall detection performance will depend on both, the quality of the change image and the quality of the thresholding. In this work, we choose to concentrate on the first step of the procedure, that is, the generation of an indicator of change for each pixel in the image. The change indicator can be obtained by computing the local correlation between both images, for each pixel position. For interesting approaches in the field of unsupervised change image thresholding, the reader can refer to the works of Bruzzone and Fernández Prieto [6], [7], Bruzzone and Serpico [8], and Bazi *et al.* [9]. The change indicator can also be useful by itself. Indeed, the end user of a change map often wants, not only the binary information given after thresholding, but also an indicator of the change amplitude. In order to evaluate the quality of a change image independently of the choice of the thresholding algorithm, one can study the evolution of the probability of detection as a function of the probability of false alarm, when a sequence of constant thresholds is used for the whole image. As in the image registration problem, a small estimation window is required in order to obtain a high-resolution detector, that is, a detector being able to identify changes

Manuscript received March 30, 2006; revised February 4, 2007. This work was supported in part by CNES under contract 2005/2561 and in part by the CNRS under MathSTIC Action 80/0244. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jacques Blanc-Talon.

F. Chatelain and J.-Y. Tourneret are with IRIT/ENSEEIH/TéSA, 31071 Toulouse cedex 7, France (e-mail: florent.chatelain@enseeiht.fr; jean-yves.tourneret@enseeiht.fr).

J. Inglada is with CNES, 31401 Toulouse cedex 9, France (e-mail: jordi.inglada@cnes.fr).

A. Ferrari is with LUAN, Université de Nice-Sophia-Antipolis, 06108 Nice cedex 2, France (e-mail: ferrari@unice.fr).

Digital Object Identifier 10.1109/TIP.2007.896651

with a small spatial extent. Again, the introduction of *a priori* knowledge through MGDs may improve the estimation accuracy when a small number of samples is used.

This paper is organized as follows. Section II recalls some important results on MGDs. Section III studies estimators of the unknown parameters of a bivariate gamma distribution (BGD). These estimators are based on the classical maximum likelihood method and method of moments. Section IV studies interesting properties of the mutual information for BGDs. The application to image registration and change detection is discussed in Section V. Conclusions are finally reported in Section VI.

II. MULTIVARIATE GAMMA DISTRIBUTIONS

A. Definitions

A polynomial $P(\mathbf{z})$ with respect to $\mathbf{z} = (z_1, \dots, z_d)$ is *affine* if the one variable polynomial $z_j \mapsto P(\mathbf{z})$ can be written $Az_j + B$ (for any $j = 1, \dots, d$), where A and B are polynomials with respect to the z_i s with $i \neq j$. A random vector $\mathbf{x} = (x_1, \dots, x_d)^T$ is distributed according to an MGD on \mathbb{R}_+^d with shape parameter q and scale parameter P (denoted as $\mathbf{x} \sim \mathcal{G}(q, P)$) if its moment generating function (also called Laplace transform) is defined as follows [3]:

$$\psi_{\mathcal{G}(q,P)}(\mathbf{z}) = \mathbb{E} \left(e^{-\sum_{i=1}^d x_i z_i} \right) = [P(\mathbf{z})]^{-q} \quad (1)$$

where $q \geq 0$ and P is an affine polynomial. It is important to note the following points.

- The affine polynomial P has to satisfy appropriate conditions including $P(0) = 1$. In the general case, determining necessary and sufficient conditions on the pair (q, P) such that $\mathcal{G}(q, P)$ exist is a difficult problem. The reader is invited to look at [3] for more details.
- By setting $z_j = 0$ for $j \neq i$ in (1), we obtain the Laplace transform of x_i , which is clearly a gamma distribution with shape parameter q and scale parameter p_i , where p_i is the coefficient of z_i in P .

A BGD corresponds to the particular case $d = 2$ and is defined by its moment generating function

$$\psi(z_1, z_2) = (1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2)^{-q} \quad (2)$$

with the following conditions:

$$p_1 > 0, p_2 > 0, p_{12} > 0, p_1 p_2 - p_{12} \geq 0. \quad (3)$$

In the bidimensional case, (3) are necessary and sufficient conditions for (2) to be the moment generating function of a probability distribution defined on $[0, \infty]^2$. Note again that (2) implies that the marginal distributions of x_1 and x_2 are ‘‘gamma distributions’’ (denoted as $x_1 \sim \mathcal{G}(q, p_1)$ and $x_2 \sim \mathcal{G}(q, p_2)$) with the following densities:

$$f_{1D}(x_i) = \frac{x_i^{q-1}}{\Gamma(q)p_i^q} \exp\left(-\frac{x_i}{p_i}\right) \mathbb{1}_{\mathbb{R}_+}(x_i)$$

where $\mathbb{1}_{\mathbb{R}_+}(x_i)$ is the indicator function defined on $[0, \infty]$ ($\mathbb{1}_{\mathbb{R}_+}(x_i) = 1$ if $x_i \geq 0$, $\mathbb{1}_{\mathbb{R}_+}(x_i) = 0$ else), for $i \in \{1, 2\}$. Here, $\Gamma(\cdot)$ is the usual gamma function defined in [10, p. 255].

B. Bivariate Gamma pdf

Obtaining tractable expressions for the probability density function (pdf) of a MGD defined by (1) is a challenging problem. However, in the bivariate case, the problem is much simpler. Straightforward computations allow to obtain the following density (see [1, p. 436] for a similar result)

$$f_{2D}(\mathbf{x}) = \exp\left(-\frac{p_2 x_1 + p_1 x_2}{p_{12}}\right) \frac{x_1^{q-1} x_2^{q-1}}{p_{12}^q \Gamma(q)} f_q(c x_1 x_2) \mathbb{1}_{\mathbb{R}_+^2}(\mathbf{x})$$

where $c = (p_1 p_2 - p_{12})/p_{12}^2$ and $f_q(z)$ is defined as follows:

$$f_q(z) = \sum_{k=0}^{\infty} \frac{z^k}{k! \Gamma(q+k)}. \quad (4)$$

Note that $f_q(z)$ is related to the confluent hypergeometric function (see [1, p. 462]).

C. BGD Moments

The Taylor series expansion of the Laplace transform ψ can be written

$$\psi(z_1, z_2) = \sum_{k,l \geq 0} \frac{(-1)^{k+l}}{k! l!} \mathbb{E} [x_1^k x_2^l] z_1^k z_2^l. \quad (5)$$

The moments of a BGD can be obtained by differentiating (5). For instance, the mean and variance of x_i (denoted $\mathbb{E}[x_i]$ and $\text{var}(x_i)$ respectively) can be expressed as follows:

$$\mathbb{E}[x_i] = q p_i, \quad \text{var}(x_i) = q p_i^2 \quad (6)$$

for $i = 1, 2$. Similarly, the covariance $\text{cov}(x_1, x_2)$ and correlation coefficient $r(x_1, x_2)$ of a BGD can be easily computed

$$\text{cov}(x_1, x_2) = \mathbb{E}[x_1 x_2] - \mathbb{E}[x_1] \mathbb{E}[x_2] = q(p_1 p_2 - p_{12}) \quad (7)$$

$$r(x_1, x_2) = \frac{\text{cov}(x_1, x_2)}{\sqrt{\text{var}(x_1)} \sqrt{\text{var}(x_2)}} = \frac{p_1 p_2 - p_{12}}{p_1 p_2}. \quad (8)$$

It is important to note that for a known value of q , a BGD is fully characterized by $\boldsymbol{\theta} = (\mathbb{E}[x_1], \mathbb{E}[x_2], r(x_1, x_2))$ which will be denoted $\boldsymbol{\theta} = (m_1, m_2, r)$ in the remaining of the paper. Indeed, $\boldsymbol{\theta}$ and (p_1, p_2, p_{12}) are obviously related by a one-to-one transformation. Note also that the conditions (3) ensure that the covariance and correlation coefficient of the couple (x_1, x_2) are both positive.

More computations allow to obtain a general formula for the moments $\mathbb{E}[x_1^m x_2^n]$, for $(m, n) \in \mathbb{N}^2$, of a BGD

$$\mathbb{E}[x_1^m x_2^n] = m_1^m m_2^n \frac{(q)_m (q)_n}{q^m q^n} \sum_{k=0}^{\min(m,n)} \frac{(-m)_k (-n)_k}{(q)_k} \frac{r^k}{k!} \quad (9)$$

where $(a)_k$ is the Pochhammer symbol defined by $(a)_0 = 1$ and

$$(a)_{k+1} = (a+k)(a)_k = a(a+1) \dots (a+k)$$

for any integer k (see [10, p. 256]). The mutual information of a BGD is related to the moments of $\sqrt{x_1 x_2}$ and $\log(x_j)$ for $j = 1, 2$. Straightforward computations detailed in Appendices I and II yield the following results:

$$\mathbb{E}[\log(x_j)] = \psi(q) + \log\left(\frac{m_j}{q}\right) \quad (10)$$

$$\mathbb{E}[\sqrt{x_1 x_2}] = \frac{\sqrt{m_1 m_2} \Gamma(q + \frac{1}{2})^2}{q \Gamma(q)^2} {}_2F_1\left(\frac{-1}{2}, \frac{-1}{2}; q; r\right) \quad (11)$$

where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function and ${}_2F_1$ is the Gauss's hypergeometric function (see [10, pp. 555–566]).

III. PARAMETER ESTIMATION

This section addresses the problem of estimating the unknown parameter vector $\boldsymbol{\theta}$ from n independent vectors $\mathbf{x} = (\mathbf{x}^1, \dots, \mathbf{x}^n)$, where $\mathbf{x}^i = (x_1^i, x_2^i)$ is distributed according to a BGD with parameter vector $\boldsymbol{\theta}$. Note that the parameter q is assumed to be known here, as in most practical applications. However, this assumption could be relaxed.

A. Maximum Likelihood Method

1) *Principles*: The maximum likelihood (ML) method can be applied in the bivariate case ($d = 2$) since a closed-form expression of the density is available.¹ In this particular case, after removing the terms which do not depend on $\boldsymbol{\theta}$, the log-likelihood function can be written as follows:

$$l(\mathbf{x}; \boldsymbol{\theta}) = -nq \log(m_1 m_2) - \sum_{j=1}^2 \frac{nq \bar{x}_j}{m_j (1-r)} - nq \log(1-r) + \sum_{i=1}^n \log f_q(c x_1^i x_2^i) \quad (12)$$

where $c = r q^2 [m_1 m_2 (1-r)^2]^{-1}$, and $\bar{x}_j = (1/n) \sum_{i=1}^n x_j^i$ is the sample mean of x_j for $j = 1, 2$. By differentiating the log-likelihood with respect to m_1, m_2 and r , and by noting that $f'_q(z) = f_{q+1}(z)$, the following set of equations is obtained:

$$\frac{nq \bar{x}_i}{1-r} - nq m_i - \frac{r}{(1-r)^2} q^2 m_i \Delta = 0, \quad i \in \{1, 2\} \quad (13)$$

$$\frac{nq \bar{x}_1}{(1-r)m_1} + \frac{nq \bar{x}_2}{(1-r)m_2} - nq - \frac{1+r}{(1-r)^2} q^2 \Delta = 0 \quad (14)$$

where

$$\Delta = \frac{1}{m_1 m_2} \left(\sum_{i=1}^n x_1^i x_2^i \frac{f_{q+1}(c x_1^i x_2^i)}{f_q(c x_1^i x_2^i)} \right). \quad (15)$$

The maximum likelihood estimators (MLEs) of m_1 and m_2 are then easily obtained from these equations

$$\hat{m}_{1\text{ML}} = \bar{x}_1, \quad \hat{m}_{2\text{ML}} = \bar{x}_2. \quad (16)$$

¹The problem is much more complicated in the general case where $d > 2$ since there is no tractable expression for the MGD density. In this case, the coefficients of P can be estimated by maximizing an appropriate composite likelihood criterion such as the *pairwise log-likelihood*. The reader is invited to consult [11] for more details.

After replacing m_1 and m_2 by their MLEs in (14), we can easily show that the MLE of r is obtained by computing the root $r \in [0, 1]$ of the following function:

$$g(r) = r - 1 + \frac{q}{n \bar{x}_1 \bar{x}_2} \left(\sum_{i=1}^n x_1^i x_2^i \frac{f_{q+1}(\hat{c} x_1^i x_2^i)}{f_q(\hat{c} x_1^i x_2^i)} \right) = 0 \quad (17)$$

where

$$\hat{c} = \frac{r}{(1-r)^2} \frac{q^2}{\bar{x}_1 \bar{x}_2}.$$

This is achieved by using a Newton–Raphson procedure initialized by the standard correlation coefficient estimator [defined in (25)]. The convergence of the Newton–Raphson procedure is generally obtained after few iterations.

2) *Performance*: The asymptotic properties of the ML estimators $\hat{m}_{1\text{ML}}$ and $\hat{m}_{2\text{ML}}$ can be easily derived from the moments of the univariate gamma distributions $\mathcal{G}(q, p_1)$ and $\mathcal{G}(q, p_2)$. These estimators are obviously unbiased, convergent and efficient. However, the performance of \hat{r}_{ML} is more difficult to study. Of course, the MLE is known to be asymptotically unbiased and asymptotically efficient, under mild regularity conditions. Thus, the mean square error (MSE) of the estimates can be approximated for large data records by the Cramér–Rao lower bound (CRLB). For unbiased estimators, the CRLBs of the unknown parameters m_1, m_2 and r can be computed by inverting the Fisher information matrix, whose elements are defined by

$$[I(\boldsymbol{\theta})]_{ij} = -\mathbb{E} \left[\frac{\partial^2 \log f_{2D}(\mathbf{x})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} \right] \quad (18)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)^T = (m_1, m_2, r)^T$. However this computation is difficult because of the term $\log f_q$ appearing in the log-likelihood. In such situation, it is very usual to approximate the expectations by using Monte Carlo methods. More specifically, this approach consists of approximating the elements of the Fisher information matrix $I(\boldsymbol{\theta})$ as follows:

$$[I(\boldsymbol{\theta})]_{ij} \simeq -\frac{1}{N} \sum_{k=1}^N \frac{\partial^2 \log f_{2D}(\mathbf{x}_k)}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} \quad (19)$$

where \mathbf{x}_k is distributed according to the BGD of density f_{2D} and N is the number of Monte Carlo runs.

B. Method of Moments

1) *Principles*: This section briefly recalls the principle of the method of moments. Consider a function $\mathbf{h}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^L$ and the statistic s_n of size L defined as

$$\mathbf{s}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{h}(\mathbf{x}^i) \quad (20)$$

where $\mathbf{h}(\cdot)$ is usually chosen such that s_n is composed of empirical moments. Denote as

$$\mathbf{f}(\boldsymbol{\theta}) = \mathbb{E}[s_n] = \mathbb{E}[\mathbf{h}(\mathbf{x}^1)]. \quad (21)$$

The moment estimator of $\boldsymbol{\theta}$ is constructed as follows:

$$\hat{\boldsymbol{\theta}}_{\text{Mo}} = \mathbf{g}(\mathbf{s}_n) \quad (22)$$

where $\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}$. By considering the function

$$\mathbf{h}(\mathbf{x}) = (x_1, x_2, x_1^2, x_2^2, x_1 x_2)$$

the following result is obtained:

$$\mathbf{f}(\boldsymbol{\theta}) = [m_1, m_2, m_1^2(1+q^{-1}), m_2^2(1+q^{-1}), m_1 m_2(1+rq^{-1})]. \quad (23)$$

The unknown parameters (m_1, m_2, r) can then be expressed as functions of $\mathbf{f}(\boldsymbol{\theta}) = (f_1, f_2, f_3, f_4, f_5)$. For instance, the following relations are obtained:

$$m_1 = f_1, \quad m_2 = f_2, \quad r = \frac{f_5 - f_1 f_2}{\sqrt{(f_3 - f_1^2)(f_4 - f_2^2)}} \quad (24)$$

yielding the standard estimators

$$\begin{aligned} \widehat{m}_{1M_0} &= \bar{x}_1, & \widehat{m}_{2M_0} &= \bar{x}_2, \\ \widehat{r}_{M_0} &= \frac{\sum_{i=1}^n (x_1^i - \bar{x}_1)(x_2^i - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (x_1^i - \bar{x}_1)^2} \sqrt{\sum_{i=1}^n (x_2^i - \bar{x}_2)^2}}. \end{aligned} \quad (25)$$

2) *Performance*: The asymptotic performance of the estimator $\widehat{\boldsymbol{\theta}}_{M_0}$ can be derived by imitating the results of [12] derived in the context of time series analysis. A key point of these proofs is the assumption $\mathbf{s}_n \xrightarrow{a.s.} \mathbf{s} = \mathbf{f}(\boldsymbol{\theta})$ which is verified herein by applying the strong law of large numbers to (20). As a result, the asymptotic MSE of $\widehat{\boldsymbol{\theta}}_{M_0}$ can be derived

$$\lim_{n \rightarrow \infty} nE \left[(\widehat{\boldsymbol{\theta}}_{M_0} - \boldsymbol{\theta})^2 \right] = \mathbf{G}(\boldsymbol{\theta}) \boldsymbol{\Sigma}(\boldsymbol{\theta}) \mathbf{G}(\boldsymbol{\theta})^t \quad (26)$$

where $\mathbf{G}(\boldsymbol{\theta})$ is the Jacobian matrix of the vector $\mathbf{g}(\cdot)$ at point $\mathbf{s} = \mathbf{f}(\boldsymbol{\theta})$ and

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} nE \left[(\mathbf{s}_n - \mathbf{s})(\mathbf{s}_n - \mathbf{s})^T \right]. \quad (27)$$

In the previous example, according to (24), $g : \mathbb{R}^5 \rightarrow \mathbb{R}^3$ is defined as follows:

$$\mathbf{g}(\mathbf{x}) = \left(x_1, x_2, \frac{x_5 - x_1 x_2}{\sqrt{(x_3 - x_1^2)(x_4 - x_2^2)}} \right). \quad (28)$$

The partial derivatives of g_1 and g_2 with respect to x_i , $i = 1, \dots, 5$ are trivial. By denoting $\gamma = \sqrt{(x_3 - x_1^2)(x_4 - x_2^2)}$, those of g_3 can be expressed as

$$\begin{aligned} \frac{\partial g_3}{\partial x_1} &= -\frac{x_2}{\gamma} + \frac{x_1(x_4 - x_2^2)(x_5 - x_1 x_2)}{\gamma^3} \\ \frac{\partial g_3}{\partial x_2} &= -\frac{x_1}{\gamma} + \frac{x_2(x_3 - x_1^2)(x_5 - x_1 x_2)}{\gamma^3} \\ \frac{\partial g_3}{\partial x_3} &= \frac{(x_1 x_2 - x_5)(x_4 - x_2^2)}{2\gamma^3} \\ \frac{\partial g_3}{\partial x_4} &= \frac{(x_1 x_2 - x_5)(x_3 - x_1^2)}{2\gamma^3} \\ \frac{\partial g_3}{\partial x_5} &= \frac{1}{\gamma}. \end{aligned} \quad (29)$$

The elements of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ can be computed from the moments of $\mathbf{h}(\mathbf{x})$ which are obtained by differentiating the Laplace trans-

form (2). The asymptotic MSEs (26) are then computed by using (9).

IV. MUTUAL INFORMATION FOR BGDs

Some limitations of the standard estimated correlation coefficient can be alleviated by using other similarity measures [4]. These similarity measures include the well-known mutual information. The mutual information of a BGD of shape parameter q and scale parameter $\mathbf{P} = (p_1, p_2, p_{12})$ can be defined as follows:

$$M_q(p_1, p_2, p_{12}) = \int_{\mathbb{R}^2} f_{2D}(\mathbf{x}) \log \left[\frac{f_{2D}(\mathbf{x})}{f(x_1, \cdot) f(\cdot, x_2)} \right] d\mathbf{x} \quad (30)$$

where $f(x_1, \cdot)$ and $f(\cdot, x_2)$ are the marginal densities of the vector $\mathbf{x} = (x_1, x_2)$ and $f_{2D}(\mathbf{x})$ is its joint pdf. This section shows that the mutual information of BGDs is related to the correlation coefficient r by a one-to-one transformation. Interesting approximations of this mutual information for $r \rightarrow 0$ and $r \rightarrow 1$ are also derived.

A. Numerical Evaluation of the Mutual Information

By replacing the densities $f(x_1, \cdot)$, $f(x_2, \cdot)$ and $f_{2D}(\mathbf{x})$ by their analytical expressions, the following results can be obtained:

$$M_q(p_1, p_2, p_{12}) = q \log \left(\frac{p_1 p_2}{p_{12}} \right) - c \left(E[x_1] \frac{p_{12}}{p_1} + E[x_2] \frac{p_{12}}{p_2} \right) + E \{ \log [\Gamma(q) f_q(cx_1 x_2)] \}. \quad (31)$$

The first terms of $M_q(p_1, p_2, p_{12})$ can be easily expressed as a function of $\boldsymbol{\theta}$ by using the mean of a univariate gamma distribution given in (6). The mutual information $M_q(p_1, p_2, p_{12})$ can then be expressed as follows:

$$M_q(p_1, p_2, p_{12}) = q \log(1 - r) - \frac{2qr}{1 - r} + E \{ \log [\Gamma(q) f_q(cx_1 x_2)] \}. \quad (32)$$

However, a simple closed-form expression for $A = E \{ \log [\Gamma(q) f_q(cx_1 x_2)] \}$ cannot be obtained, requiring to use a numerical procedure for its computation.

The numerical evaluation of A can be significantly simplified by noting that (x_1, x_2) and $(\alpha x_1, \beta x_2)$ have the same mutual information for any $(\alpha, \beta) \in \mathbb{R}^2$. Indeed, this property implies the following result:

$$M_q(p_1, p_2, p_{12}) = M_q \left(1, 1, \frac{p_{12}}{p_1 p_2} \right) = M_q(1, 1, 1 - r) \quad (33)$$

where $1 - r = p_{12}/p_1 p_2 \in [0, 1]$. As a consequence, $A = E \{ \log [\Gamma(q) f_q(cx_1 x_2)] \}$ can be computed by replacing (p_1, p_2, p_{12}) by $(1, 1, 1 - r)$, where $1 - r \in [0, 1]$. This expectation can be precomputed for all possible values of q and for $1 - r \in [0, 1]$, simplifying the numerical evaluation of M_q .

Moreover, it is interesting to note that (33) shows that the mutual information M_q and the correlation r are related by a one-to-one transformation. Consequently, M_q and r should provide similar performance for image registration and change de-

tection. The advantage of using the mutual information will be discussed later.

B. Approximations of the Mutual Information

The numerical evaluation of A can be avoided for values of r closed to 0 and 1 by using approximations. Indeed, the following results can be obtained:

1) $r \rightarrow 0$: The second-order Taylor expansion of $f_q(z)$ around $z = 0$ can be written

$$f_q(z) = 1 + \frac{z}{q\Gamma(q)} + \frac{z^2}{2q(q+1)\Gamma(q)} + o(z^2) \quad (34)$$

where $o(z^2)/z^2$ tends to 0, as $z \rightarrow 0$. As a consequence, A can be approximated as follows:

$$A \approx \frac{cE[x_1x_2]}{q} - \frac{c^2E[(x_1x_2)^2]}{2q} \left(\frac{1}{q(q+1)} \right). \quad (35)$$

By using (9), the mutual information M_q can be finally approximated as follows:

$$M_q \approx \frac{r^2}{2}. \quad (36)$$

2) $r \rightarrow 1$: The Taylor expansion of $f_q(z)$ around ∞ can be written

$$f_q(z) = \frac{\exp(2\sqrt{z})}{\sqrt{4\pi z^{q-1/2}}} (1 + o(1)) \quad (37)$$

where $o(1)$ tends to 0, as $z \rightarrow \infty$. As a consequence, A can be approximated as follows:

$$\begin{aligned} A &\approx E \left\{ \log \left[\frac{\exp(2\sqrt{cx_1x_2})}{\sqrt{4\pi(cx_1x_2)^{q-1/2}}} \right] \right\} \\ &\approx 2\sqrt{c}E[\sqrt{x_1x_2}] - \left(\frac{q}{2} - 1 \right) (\log c + E[\log x_1 + \log x_2]) \\ &\quad + \log \left[\frac{\Gamma(q)}{2\sqrt{\pi}} \right]. \end{aligned} \quad (38)$$

After replacing the means of $\sqrt{x_1x_2}$, $\log(x_1)$ and $\log(x_2)$ derived in Appendices I and II, the following result can be obtained:

$$M_q \approx -\frac{1}{2} \log(1-r) + \left(q - \frac{1}{2} \right) + \log \left[\frac{\Gamma(p)}{2\sqrt{\pi}} \right] - \left(q - \frac{1}{2} \right) \psi(q). \quad (39)$$

Fig. 1 shows that the mutual information M_q can be accurately approximated by (36) and (39) for $r < 0.5$ and $r > 0.9$. This figure has been obtained with the parameters $p_1 = 1$ and $p_2 = 1$ without loss of generality (see discussion at the beginning of this section).

V. APPLICATION TO IMAGE REGISTRATION AND CHANGE DETECTION

This section explains carefully how BGDs can be used for image registration and change detection. Theoretical results are illustrated by many simulations conducted with synthetic and real data.

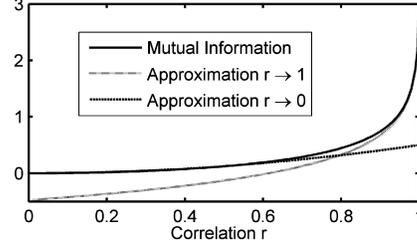


Fig. 1. Mutual information and its approximations for $r \rightarrow 0$ and $r \rightarrow 1$.

A. Synthetic Data

1) *Generation*: The generation of a vector $\mathbf{x} = (x_1, x_2)^T$ distributed according to a BGD has been performed as follows.

- Simulate $2q$ independent multivariate Gaussian vectors of \mathbb{R}^2 denoted as z^1, \dots, z^{2q} with means $(0,0)$ and the following 2×2 covariance matrix:

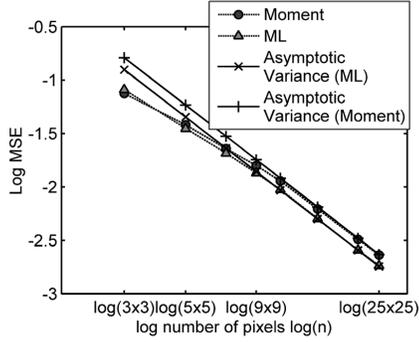
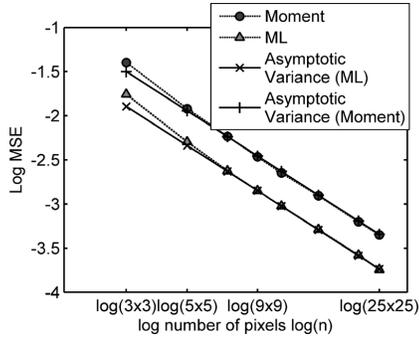
$$C = (c_{i,j})_{1 \leq i,j \leq 2} = \left(r^{\frac{|i-j|}{2}} \right)_{1 \leq i,j \leq 2}.$$

- Compute the k th component of $\mathbf{x} = (x_1, x_2)^T$ as $x_k = (m_k/2q) \sum_{1 \leq i \leq 2q} (z_k^i)^2$, where z_k^i is the k th component of z^i .

By computing the Laplace transform of \mathbf{x} , it can be shown that the two previous steps allow to generate random vectors $\mathbf{x} = (x_1, x_2)^T$ distributed according to a BGD whose marginal distributions are univariate gamma distributions $\mathcal{G}(q, m_1/q)$ and $\mathcal{G}(q, m_2/q)$. Moreover, the correlation coefficient of $\mathbf{x} = (x_1, x_2)^T$ is equal to r (the reader is invited to consult Appendix III for more details).

2) *Estimation Performance*: The first simulations compare the performance of the method of moments with the ML method as a function of n . Note that the possible values of n are $n = (2p+1) \times (2p+1)$, where $p \in \mathbb{N}$ (more precisely $3 \times 3 = 9$, $5 \times 5 = 25$, ..., $25 \times 25 = 625$). These values are appropriate for the image registration and change detection problems, as explained in the next sections. The number of Monte Carlo runs is 1000 for all figures presented in this section. The other parameters for this first example are $m_1 = 400$, $m_2 = 800$ and $q = 1$ (1-Look images). Figs. 2 and 3 show the MSEs of the estimated correlation coefficient for two different correlation structures ($r = 0.2$ and $r = 0.8$). The circle curves correspond to the estimator of moments whereas the triangle curves correspond to the MLE. These figures show the interest of the ML method, which is much more efficient for this problem than the method of moments. The figures also show that the difference between the two methods is more significant for large values of the correlation coefficient r . Note that the theoretical asymptotic MSEs of both estimators determined in (18) and (26) are also displayed in Figs. 2 and 3 (continuous lines). The theoretical MSEs are clearly in good agreement with the estimated MSEs, even for small values of n . This is particularly true for large values of r .

3) *Detection Performance*: We consider synthetic vectors $\mathbf{x} = (x_1, x_2)^T$ (coming from 128×128 synthetic images) distributed according to BGDs with $r = 0.3$ and $r = 0.65$ modeling the presence and absence of changes, respectively.


 Fig. 2. Log MSEs versus $\log(n)$ for parameter $r(r = 0.2)$.

 Fig. 3. Log MSEs versus $\log(n)$ for parameter $r(r = 0.8)$.

The correlation coefficient r of each bivariate vector $\mathbf{x}^{(i,j)} = (x_1^{(i,j)}, x_2^{(i,j)})^T$ (for $i, j = 1 \dots 128$) is estimated from vectors belonging to windows of size $n = (2p+1) \times (2p+1)$ centered around the pixel of coordinates (i, j) in the two analyzed images. The following binary hypothesis test is then considered:

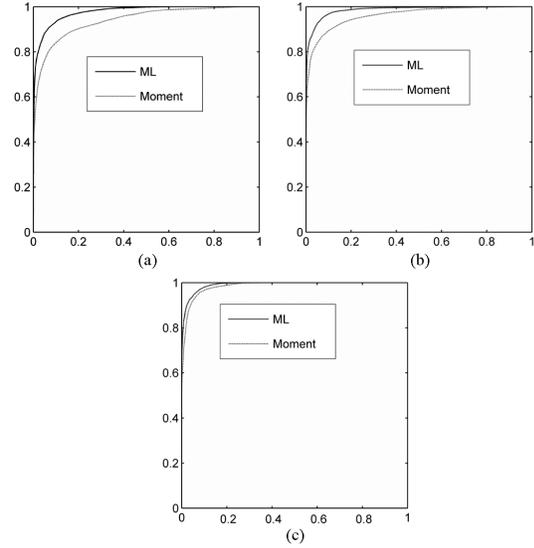
$$\begin{aligned} H_0 \quad (\text{absence of change}) : \quad & \hat{r} > \lambda \\ H_1 \quad (\text{presence of change}) : \quad & \hat{r} < \lambda \end{aligned} \quad (40)$$

where λ is a threshold depending on the probability of false alarm and \hat{r} is an estimator of the correlation coefficient (obtained from the method of moments or the maximum likelihood principle). The performance of the change detection strategy (40) can be defined by the two following probabilities [13, p. 34]

$$\begin{aligned} P_D &= P[\text{accepting } \mathcal{H}_1 | \mathcal{H}_1 \text{ is true}] \\ &= P[\hat{r} < \lambda | \mathcal{H}_1 \text{ is true}] = \int_{-\infty}^{\lambda} p_1(u) du \end{aligned} \quad (41)$$

$$\begin{aligned} P_{FA} &= P[\text{accepting } \mathcal{H}_1 | \mathcal{H}_0 \text{ is true}] \\ &= P[\hat{r} > \lambda | \mathcal{H}_0 \text{ is true}] = \int_{\lambda}^{\infty} p_0(u) du \end{aligned} \quad (42)$$

where $p_0(u)$ and $p_1(u)$ are the pdfs of \hat{r} under hypotheses H_0 and H_1 , respectively. Thus, for each value of λ , there exists


 Fig. 4. ROCs for synthetic data for different window sizes. (a) $n = 9 \times 9$; (b) $n = 15 \times 15$; (c) $n = 21 \times 21$.

a pair (P_{FA}, P_D) . The curves of P_D as a function of P_{FA} are called receiver operating characteristics (ROCs) [13, p. 38].

The ROCs for the change detection problem (40) are depicted in Fig. 4(a)–(c) for three different window sizes corresponding to $n = 2p + 1 \in \{9, 15, 21\}$. The ML estimator clearly outperforms the moment estimator for these examples. However, it is interesting to note that the two estimators have similar performances for large window sizes.

B. Application to Image Registration

This section studies an image registration technique based on BGDs. More precisely, consider two images whose pixels are denoted $\{x_1^1, \dots, x_1^n\}$ and $\{x_2^1, \dots, x_2^n\}$. Given the left image x_1 , we propose the following basic three-step image registration algorithm.

- Step 1: Determine the search area in the right image x_2 . Here, we use images that have been previously registered by a human operator using appropriate interactive software, a digital elevation model and geometrical sensor models. The use of registered images allows us to validate the results, since the expected shift between the images is equal to 0. For this experiment and without loss of generality, the search area is reduced to a line (composed of ten pixels before and ten pixels after the pixel of interest).
- Step 2: For each pixel x_2^j in the search area, estimate a similarity measure (correlation coefficient or mutual information) between x_1^i and x_2^j .
- Step 3: Select the pixel providing the largest similarity.

This three-step procedure has been applied to a couple of Radarsat 1-Look images acquired before and after the eruption of the Nyiragongo volcano which occurred in January 2002. The Radarsat images are depicted in Fig. 5(a) (before eruption)

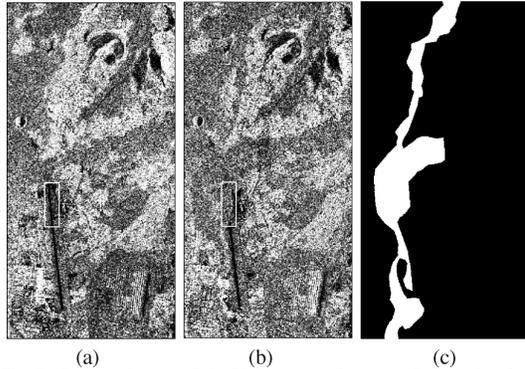


Fig. 5. Radarsat images of the Nyiragongo volcano. (a) Before; (b) after; (c) mask.

and (b) (after eruption). Note that some changes due to the eruption can be clearly seen on the landing track for example. Fig. 5(c) indicates the pixels of the image which have been affected by the eruption (white pixels). This reference map was obtained by photo-interpreters—who used the same SAR images we are using—and ground truth elaborated by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) Humanitarian Information Center (HIC) on January 27, 2002, that is, a few days after the eruption. This reference map was afterwards validated by a terrain mission. The types of change covered are: presence of a lava flow over old existing lava flows, damaged buildings (areas with different types of habitat). The area of study does not include forest or areas of dense vegetation (see <http://www.users.skynet.be/technaphot/webgamma/index.htm> for some examples of damages).

Fig. 6(a)–(c) shows an average of the estimated correlation coefficients with errorbars corresponding to mean \pm standard deviation/10. These estimates have been computed for all black pixels which have not been affected by the eruption for different window sizes. More precisely, for every black pixel x_1^i of the left figure, we consider a window of size $n = (2p+1) \times (2p+1)$ centered around x_1^i . The same window is also considered in the right picture around pixel x_2^j . The correlation coefficient between the two pixels x_1^i and x_2^j is estimated by using the $n = (2p+1) \times (2p+1)$ couples of pixels located in the left and right windows. This operation is repeated for different central pixels x_2^j belonging to the search area (i.e., the 21 pixels of $[x_2^{i+\tau}]_{-10 \leq \tau \leq 10} = [x_2^{i-10}, \dots, x_2^{i-1}, x_2^i, x_2^{i+1}, \dots, x_2^{i+10}]$), where τ is the shift between the right and left windows. The results are averaged over all black pixels displayed in the mask Fig. 5(c). The estimated correlation coefficient is maximum when $j = i$, or equivalently $\tau = 0$, i.e., when the left and right windows are centered at the same location. This result indicates that the correlation coefficient can be efficiently used for image registration. Moreover, it is interesting to study how the estimator selectivity (which can be defined as the relative amplitude of the peak compared to that of the plateau) varies from one estimator to another and depends on the window size. In particular, the ML estimator provides a slightly better selectivity than the estimator of moments. Note that the errorbars are very similar for the two estimators. Even if the different

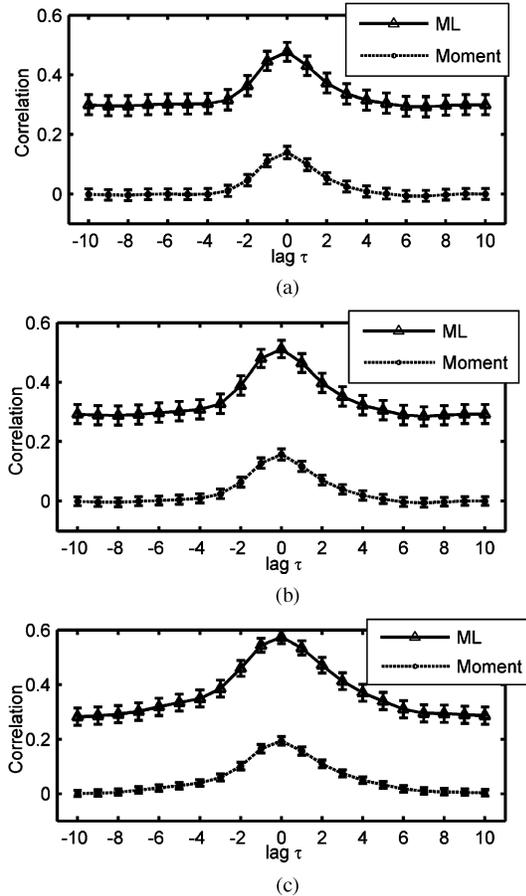


Fig. 6. Averaged correlation coefficient estimates versus τ for black pixels with errorbars (ML: maximum likelihood estimator, moment: moment estimator) for Nyiragongo images for several window sizes. (a) Window size $n = 7 \times 7$; (b) window size $n = 9 \times 9$; (c) window size $n = 15 \times 15$.

methods provide similar results for image registration, it is important to note that the proposed framework allows one to define an interesting joint distribution for the vector (x_1^i, x_2^j) . This distribution might be used for other tasks as, for instance, joint image segmentation and classification of both data sets.

The same operation is conducted on a rectangular region composed of white pixels of the mask Fig. 5(c) (which have been affected by the eruption) depicted in white in Fig. 5(a) and (b). The results presented in Fig. 7 clearly show that the estimated correlation coefficient is much smaller when computed on a region affected by the eruption (and also that there is no peak which might be used for registration). This result is interesting and can be used for detecting changes between the two images, as illustrated in the next section.

C. Application to Change Detection

This section considers two 1-Look images acquired at different dates around Gloucester (U.K.) before and during a flood

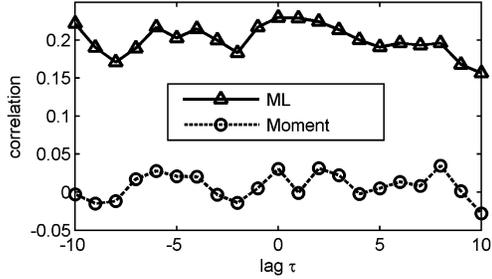


Fig. 7. Averaged correlation coefficient estimates versus τ for white pixels belonging to the Nyiragongo images square region (ML: maximum likelihood estimator; moment: moment estimator).

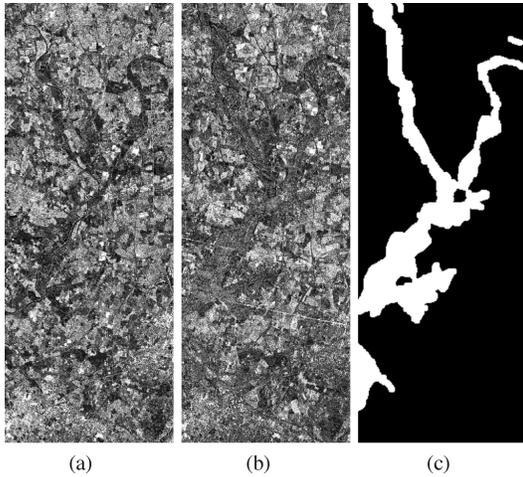


Fig. 8. Radarsat images of Gloucester before and after flood. (a) Before; (b) after; (c) mask.

(on September 9, 2000 and October 21, 2000 respectively). The images as well as a mask indicating the pixels affected by the flood are depicted in Fig. 8(a)–(c). The reference map Fig. 8(c) was obtained by photo-interpretors—who used the same SAR images we are using—and a reference map built from Landsat and SPOT data acquired one day after the radar image.

This section compares the performance of the following change detectors:

- the ratio edge detector which has been intensively used for SAR images [14], [15];
- the correlation change detector, where $\hat{\tau}$ in (40) has been estimated with the moment estimator (referred to as “*Correlation Moment*”);
- the correlation change detector, where $\hat{\tau}$ in (40) has been estimated with the ML method for BGDs (referred to as “*Correlation ML*”).

The ROCs for this change detection problem are shown in Fig. 9(a)–(c) for different window sizes n . The correlation ML detector clearly provides the best results.

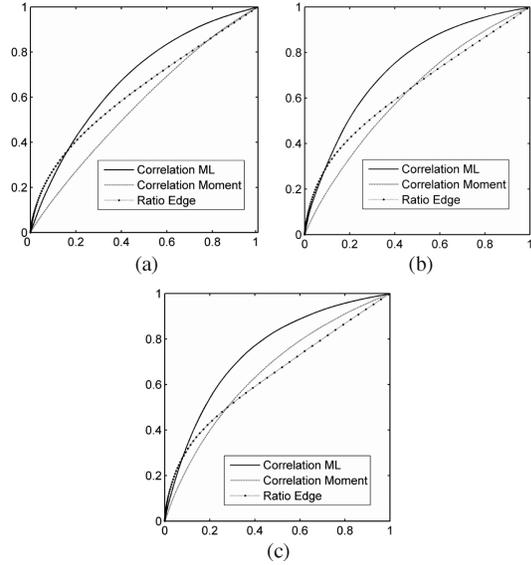


Fig. 9. ROCs for Gloucester images for different window sizes. (a) $n = 9 \times 9$; (b) $n = 15 \times 15$; (c) $n = 21 \times 21$.

The last experiments illustrate the advantage of using the mutual information for change detection. Consider the following change detector based on the mutual information:

$$\begin{aligned} H_0 \text{ (absence of change)} &: M_q(1, 1, \hat{\tau}) > \lambda_{\text{PFA}} \\ H_1 \text{ (presence of change)} &: M_q(1, 1, \hat{\tau}) < \lambda_{\text{PFA}} \end{aligned} \quad (43)$$

where $M_q(1, 1, \hat{\tau})$ is the estimated mutual information obtained by numerical integration of (32). The ROCs obtained with the detectors (40) and (43) are identical, reflecting the one-to-one transformation between the parameters $\hat{\tau}$ and $M_q(1, 1, \hat{\tau})$. However, the advantage of using the mutual information for change detection is highlighted in Fig. 10, which shows the average probability of error $P_e = (1/2)(P_{\text{ND}} + P_{\text{FA}})$ (where $P_{\text{ND}} = 1 - P_{\text{D}}$ is the probability of nondetection) as a function of the threshold λ for the change detectors (40) and (43). For a practical application, it is important to choose a threshold λ_{PFA} for these change detection problems. This choice can be governed by the value of the probability of error P_e . Assume that we are interested in having a probability of error satisfying $P_e \leq 0.39$. Fig. 10 indicates that there are clearly more values of the threshold λ_{PFA} satisfying this condition for the curve “Mutual information” than for the curve “Correlation ML.” This remains true whatever the value of the maximum probability of error P_e . Consequently, the threshold is easier to be adjusted with the detector based on the mutual information (43) than the detector based on the correlation coefficient (40).

VI. CONCLUSION

This paper studied the performance of image registration and change detection techniques based on bivariate gamma distri-

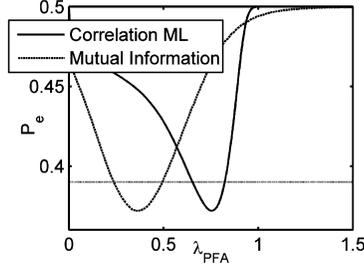


Fig. 10. Average probability of error $P_e = 1/2(P_{FA} + P_{ND})$ versus threshold λ for Gloucester images for an estimation window of size $n = 9 \times 9$.

butions. Both methods required to estimate the correlation coefficient between two images. Estimators based on the method of moments and on the maximum likelihood principle were studied. The asymptotic performance of both estimators was derived. The application to image registration and change detection was finally investigated.

The results showed the interest of using prior information about the data. On the other hand, the method presented here should not be used for more general cases where the BGD model does not hold. For these cases, the use of more general models as, for instance, copulas [16] or bivariate versions of the Pearson system [1, pp. 6–9], should be studied.

APPENDIX I

$E[\log U]$ WHERE $U \sim \mathcal{G}(a, b^{-1})$

The moment of $\log U$ can be determined by the simple change of variable $V = bU$

$$\begin{aligned} E[\log U] &= \int_0^\infty \log(u) \frac{b^a}{\Gamma(a)} e^{-bu} u^{a-1} du \\ &= \frac{1}{\Gamma(a)} \int_0^\infty \log\left(\frac{v}{b}\right) e^{-v} v^{a-1} dv \\ &= \frac{1}{\Gamma(a)} [\Gamma'(a) - \log(b)\Gamma(a)] \\ &= \psi(a) - \log(b). \end{aligned} \quad (44)$$

APPENDIX II

$E[\sqrt{x_1 x_2}]$ AND ITS APPROXIMATION FOR $r \rightarrow 1$ WHERE $(x_1, x_2) \sim \Gamma(q, \mathbf{P})$

A. $E[\sqrt{x_1 x_2}]$ Computation

The moment of the random variable $\sqrt{x_1 x_2}$ is derived from the probability density function f_{2D} of the bivariate vector $\mathbf{x} = (x_1, x_2)^T$

$$\begin{aligned} E[\sqrt{x_1 x_2}] &= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \sqrt{x_1 x_2} f_{2D}(x_1, x_2) dx_1 dx_2, \\ &= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \exp\left(-\frac{p_2 x_1 + p_1 x_2}{p_{12}}\right) \frac{x_1^{q-1/2} x_2^{q-1/2}}{p_{12}^q \Gamma(q)} \\ &\quad \times f_q(c; x_1, x_2) \mathbb{1}_{\mathbb{R}_+^2}(\mathbf{x}). \end{aligned} \quad \text{leads to}$$

The definition of f_q given in (4) yields

$$\begin{aligned} E[\sqrt{x_1 x_2}] &= \frac{1}{p_{12}^q \Gamma(q)} \sum_{k \geq 0} \frac{c^k}{k! \Gamma(q+k)} \\ &\quad \times \int_{\mathbb{R}_+} \exp\left(-\frac{p_2}{p_{12}} x_1\right) x_1^{q+k-1/2} dx_1 \\ &\quad \times \int_{\mathbb{R}_+} \exp\left(-\frac{p_1}{p_{12}} x_2\right) x_2^{q+k-1/2} dx_2 \\ &= \frac{1}{p_{12}^q \Gamma(q)} \sum_{k \geq 0} \frac{c^k}{k! \Gamma(q+k)} \left(\frac{p_{12}}{p_2}\right)^{q+k+1/2} \\ &\quad \times \Gamma(q+k+1/2) \left(\frac{p_{12}}{p_1}\right)^{q+k+1/2} \\ &\quad \times \Gamma(q+k+1/2) \\ &= \left(\frac{p_{12}}{p_1 p_2}\right)^{q+1} \frac{\sqrt{p_1 p_2}}{\Gamma(q)} \sum_{k \geq 0} \frac{\Gamma(q+k+1/2)^2}{k! \Gamma(q+k)} r^k \\ &= (1-r)^{q+1} \frac{\sqrt{m_1 m_2}}{q} \left(\frac{\Gamma(q+1/2)}{\Gamma(q)}\right)^2 \\ &\quad \times {}_2F_1\left(q + \frac{1}{2}, q + \frac{1}{2}; q; r\right) \end{aligned}$$

since

$$r = \frac{p_1 p_2 - p_{12}}{p_1 p_2} = \frac{p_{12}^2}{p_1 p_2} c.$$

Here ${}_2F_1$ is the Gauss's hypergeometric function (see [10, pp. 555–566]) defined as

$${}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}$$

and $(a)_k$ is the Pochhammer symbol presented in Section II-C (note that $(a)_k = \Gamma(a+k)/\Gamma(a)$ for any integer k and any real $a > 0$). By using the following properties of Gauss's hypergeometric functions.

- 1) The hypergeometric series ${}_2F_1(a, b; c; z)$ converges if c is not a negative integer for complex numbers z such that $|z| < 1$ or $|z| = 1$ if $\Re[c - a - b] > 0$.
- 2) ${}_2F_1(a, b; c; z) = (1-z)^{c-a-b} {}_2F_1(c-a, c-b; c; z)$ for all $|z| < 1$ (see [10, p. 559]).

The following results can be obtained

$$E[\sqrt{x_1 x_2}] = \frac{\sqrt{m_1 m_2}}{q} \frac{\Gamma(q+1/2)^2}{\Gamma(q)^2} {}_2F_1\left(\frac{-1}{2}, \frac{-1}{2}; q; r\right).$$

B. $E[\sqrt{x_1 x_2}]$ Approximation for $r \rightarrow 1$

The following identity (Gauss's hypergeometric theorem):

$${}_2F_1(a, b; c; 1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}$$

$${}_2F_1\left(\frac{-1}{2}, \frac{-1}{2}; q; 1\right) = \frac{\Gamma(q)\Gamma(q+1)}{\Gamma(q+1/2)^2}$$

and to the following first-order Taylor expansion around $z = 1$:

$${}_2F_1\left(\frac{-1}{2}, \frac{-1}{2}; q; z\right) = \frac{\Gamma(q)\Gamma(q+1)}{\Gamma(q+1/2)^2} + (z-1) \times {}_2F_1'\left(\frac{-1}{2}, \frac{-1}{2}; q; 1\right) + o(1-z)$$

where $o(1-z)/(1-z)$ tends to 0, as $z \rightarrow 1$. Using

$${}_2F_1'(a, b; c; z) = \frac{ab}{c} {}_2F_1(a+1, b+1; c+1; z) \text{ for } |z| < 1$$

the previous Taylor expansion can be written

$${}_2F_1\left(\frac{-1}{2}, \frac{-1}{2}; q; z\right) = \frac{\Gamma(q)\Gamma(q+1)}{\Gamma(q+1/2)^2} \left(1 + \frac{z-1}{4q}\right) + o(1-z).$$

Finally

$$\begin{aligned} E[\sqrt{x_1 x_2}] &= \frac{\sqrt{m_1 m_2}}{q} \frac{\Gamma(q+1/2)^2 \Gamma(q)\Gamma(q+1)}{\Gamma(q)^2 \Gamma(q+1/2)^2} \\ &\quad \times \left(1 - \frac{1-r}{4q}\right) + o(1-r) \\ &= \sqrt{m_1 m_2} \left(1 - \frac{1-r}{4q}\right) + o(1-r) \text{ for } r \rightarrow 1. \end{aligned}$$

APPENDIX III GENERATION OF SYNTHETIC DATA DISTRIBUTED ACCORDING TO BGDs

This appendix shows that the vector $\mathbf{x} = (x_1, x_2)^T$ where $x_k = (m_k/2q) \sum_{1 \leq i \leq 2q} (z_k^i)^2$ (where z_k^i is the k th component of $\mathbf{z}^i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$, with $\mathbf{C} = (c_{i,j})_{1 \leq i, j \leq 2} = (r^{|i-j|/2})_{1 \leq i, j \leq 2}$) is distributed according to a BGD whose marginals are Gamma distributions $\Gamma(q, m_1/q)$ and $\Gamma(q, m_2/q)$ and whose correlation coefficient is r . By using the independence between vectors $\mathbf{z}^1, \dots, \mathbf{z}^{2q}$, the Laplace transform of \mathbf{x} evaluated at $\mathbf{t} = (t_1, t_2)^T$ can be written

$$\begin{aligned} E[\exp(-\mathbf{t}^T \mathbf{x})] &= E\left[\prod_{k=1}^2 \exp(-t_k x_k)\right] \\ &= \prod_{i=1}^{2q} E\left[\exp\left(-\frac{(\mathbf{z}^i)^T \mathbf{S}(\mathbf{z}^i)}{2}\right)\right] \end{aligned}$$

where

$$\mathbf{S} = \begin{pmatrix} t_1 \frac{m_1}{q} & 0 \\ 0 & t_2 \frac{m_2}{q} \end{pmatrix}.$$

By using the probability density function of a bivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{C})$, the Laplace transform can be finally expressed as

$$\begin{aligned} E[\exp(-\mathbf{t}^T \mathbf{x})] &= \prod_{i=1}^{2q} \int_{\mathbb{R}_2} \frac{\exp(-\mathbf{z}^T (\mathbf{C}^{-1} + \mathbf{S}) \mathbf{z})}{(2\pi)\sqrt{\det \mathbf{C}}} d\mathbf{z} \\ &= \left[\frac{1}{\det(\mathbf{I}_2 + \mathbf{CS})}\right]^q \\ &= \left[1 + \sum_{i=1}^2 \frac{m_i}{q} t_i + (1-r) \frac{m_1 m_2}{q^2} t_1 t_2\right]^{-q} \end{aligned}$$

where \mathbf{I}_2 is the identity matrix in dimension 2. According to the definition (2), the vector \mathbf{x} is distributed according to a BGD with shape parameter q and scale parameters $p_1 = m_1/q, p_2 = m_2/q, p_{12} = (1-r)(m_1 m_2/q^2)$. Property, (8) ensures that the correlation coefficient of (x_1, x_2) is r .

ACKNOWLEDGMENT

The authors would like to thank A. Crouzil and G. Letac for fruitful discussions regarding image registration and multivariate gamma distributions, respectively.

REFERENCES

- [1] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions*, 2nd ed. New York: Wiley, 2000, vol. 1.
- [2] S. B. Lev, D. Bshouty, P. Enis, G. Letac, I. L. Lu, and D. Richards, "The diagonal natural exponential families and their classification," *J. Theoret. Probl.*, vol. 7, pp. 883–928, 1994.
- [3] P. Bernardoff, "Which multivariate Gamma distributions are infinitely divisible?," *Bernoulli*, 2006.
- [4] J. Inglada and A. Giros, "On the possibility of automatic multi-sensor image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2104–2120, Oct. 2004.
- [5] T. F. Bush and F. T. Ulaby, "Fading characteristics of panchromatic radar backscatter from selected agricultural targets," NASA STU/Recon Tech. Rep., Dec. 1973, vol. 75, p. 18460.
- [6] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.
- [7] L. Bruzzone and D. F. Prieto, "An adaptive parcel-based technique for unsupervised change detection," *Int. J. Remote Sens.*, vol. 21, no. 4, pp. 817–822, 2000.
- [8] L. Bruzzone and S. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 7, pp. 858–867, Jul. 1997.
- [9] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, Apr. 2005.
- [10] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1972.
- [11] F. Chatelain and J.-Y. Tournet, "Composite likelihood estimation for multivariate mixed poisson distributions," presented at the IEEE Workshop Statistical Signal Processing, Bordeaux, France, Jul. 2005.
- [12] B. Porat and B. Friedlander, "Performance analysis of parameter estimation algorithms based on high-order moments," *Int. J. Adapt. Control Signal Process.*, vol. 3, pp. 191–229, 1989.
- [13] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Part I*. New York: Wiley, 1968.
- [14] R. Touzi, A. Lopés, and P. Bousquet, "A statistical and geometrical edge detector for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 11, pp. 764–773, Nov. 1988.
- [15] E. J. M. Rignot and J. J. van Zyl, "Change detection techniques for ERS-1 SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 7, pp. 896–906, Jul. 1993.
- [16] R. B. Nelsen, *An Introduction to Copulas*. New York: Springer-Verlag, 1999, vol. 139.



Florent Chatelain was born in Paris, France, in 1981. He received the Eng. degree in computer sciences and applied mathematics from ENSIMAG, Grenoble, France, and the M.Sc. degree in applied mathematics from the University Joseph Fourier of Grenoble, both in June 2004. He is currently pursuing the Ph.D. degree with the Signal and Communication Group, IRIT Laboratory, Toulouse, France.

His research interests are centered around multivariate statistics applied to image processing.



Jean-Yves Tourneret (M'94) received the Ingénieur degree in electrical engineering from the Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique, Toulouse (ENSEEIH), France, and the Ph.D. degree from the National Polytechnic Institute, Toulouse, in 1992.

He is currently a Professor with the University of Toulouse (ENSEEIH). He is a member of the IRIT Laboratory (UMR 5505 of the CNRS), where his research activity is centered around estimation, detection, and classification of non-Gaussian and nonstationary processes.

Dr. Tourneret was the Program Chair of the European Conference on Signal Processing (EUSIPCO), which was held in Toulouse in 2002. He was also a member of the organizing committee for the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) held in Toulouse in 2006. He has been a member of different technical committees, including the Signal Processing Theory and Methods (SPTM) Committee of the IEEE Signal Processing Society.



Jordi Inglada received the Telecommunications Engineer degree in 1997 from both the Universitat Politècnica de Catalunya, Catalunya, Spain, and the Ecole Nationale Supérieure des Télécommunications de Bretagne, and the Ph.D. degree in signal processing and telecommunications in 2000 from Université de Rennes 1, Rennes, France.

He is with the Centre National d'Études Spatiales, French Space Agency, Toulouse, France, working in the field of remote sensing image processing. He is in charge of the development of image processing algorithms for the operational exploitation of earth observation images, mainly in the fields of image registration, change detection and object recognition.



André Ferrari was born in Senlis, France, on June 2, 1964. He received the engineering degree from École Centrale de Lyon, Lyon, France, in 1988, the M.S. degree in 1989, and the Ph.D. from University of Nice Sophia-Antipolis (UNSA), France, in 1992, all in electrical and computer engineering.

From 1993 to 2005, he was an Associate Professor at UNSA. Since 2005, he has been a Professor at the Astrophysics Laboratory, UNSA. His current research interests are in modeling and data processing in statistical optics with a particular interest

in Astrophysics.

Change Detection in Multisensor SAR Images Using Bivariate Gamma Distributions

Florent Chatelain, Jean-Yves Tourneret, and Jordi Inglada

Abstract—This paper studies a family of distributions constructed from multivariate gamma distributions to model the statistical properties of multisensor synthetic aperture radar (SAR) images. These distributions referred to as multisensor multivariate gamma distributions (MuMGDs) are potentially interesting for detecting changes in SAR images acquired by different sensors having different numbers of looks. The first part of this paper compares different estimators for the parameters of MuMGDs. These estimators are based on the maximum likelihood principle, the method of inference function for margins, and the method of moments. The second part of the paper studies change detection algorithms based on the estimated correlation coefficient of MuMGDs. Simulation results conducted on synthetic and real data illustrate the performance of these change detectors.

Index Terms—Change detection, correlation coefficient, maximum likelihood, multivariate gamma distributions.

I. INTRODUCTION

COMBINING information acquired from multiple sensors has become very popular in many signal and image processing applications. In the case of earth observation applications, there are two reasons for that. The first one is that the fusion of the data produced by different types of sensors provides a complementarity which overcomes the limitations of a specific kind of sensor. The other reason is that, often, in operational applications, the user does not have the possibility to choose the data to work with and has to use the available archive images or the first acquisition available after an event of interest. This is particularly true for monitoring applications where image registration and change detection approaches have to be implemented on different types of data [1], [2].

Both image registration and change detection techniques consists of comparing two images I , the reference, and J , the secondary image, acquired over the same landscape—scene—at two different dates. Usually, the reference image is obtained from an archive and the acquisition of the secondary image is scheduled after an abrupt change, like a natural disaster. In the case of the change detection, the goal is producing an indicator of change for each pixel of the region of interest. This indicator

of change is the result of applying locally a similarity measure to the two images. This similarity measure is usually chosen as the correlation coefficient or other statistical feature in order to deal with noisy data.

The estimation of the similarity measure is performed locally for each pixel position. Since a statistical estimation has to be performed, and only one realization of the random variable is available, the images are supposed to be locally stationary and the ergodicity assumption allows to make estimates using several neighbor pixels. This neighborhood is the so-called estimation window. In order for the stationarity assumption to hold, this estimation window has to be small. On the other hand, robust statistical estimates need a high number of samples. Therefore, the key point of the estimation of the similarity measure is to perform high quality estimates with a small number of samples. One way to do so is to introduce *a priori* knowledge about the image statistics.

In the case of power radar images, it is well known that the pixels are marginally distributed according to gamma distributions [3]. Therefore, multivariate gamma distributions (having univariate gamma margins) seem good candidates for the robust estimation of the correlation coefficient between radar images. When multirate power radar images are acquired from different sensors, the numbers of looks associated with the different images can be different. As the number of looks is the shape parameter of the gamma distribution, this leads to study multivariate gamma distributions whose margins have different shape parameters.

A family of multivariate gamma distributions has been recently defined by Bar Lev and Bernardoff [4], [5]. These distributions are defined from an appropriate moment generating function. Their margins are distributed according to univariate gamma distributions having the same shape parameter. They have recently shown interesting properties for registration and change detection in SAR images acquired by the same sensor (i.e., for images having the same number of looks) [6], [7]. This paper studies a new family of multivariate distributions whose margins are univariate gamma distributions with different shape parameters referred to as multisensor multivariate gamma distributions (MuMGDs). The application of MuBGDs to change detection in SAR images is also investigated.

This paper is organized as follows. Section II recalls important results on monosensor multivariate gamma distributions (MoMGDs). Section III defines the family of MuMGDs considered for change detection in multisensor SAR images. Section IV studies the maximum likelihood estimator (MLE), the inference function for margins (IFM) estimator, and the estimator of moments for the unknown parameters of MuMGDs. Section V presents some simulation results illustrating the performance of MuMGDs for parameter estimation and change detection on

Manuscript received November 15, 2007; revised December 9, 2007. This work was supported in part by CNES under Contract 2005/2561 and in part by the CNRS under MathSTIC Action No. 80/0244. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jacques Blanc-Talon.

F. Chatelain and J.-Y. Tourneret are with the IRIT/ENSEEIH/TéSA, 31071 Toulouse cedex 7, France (e-mail: florent.chatelain@enseeih.fr; jean-yves.tourneret@enseeih.fr).

J. Inglada is with the CNES, 31401 Toulouse cedex 9, France (e-mail: jordi.inglada@cnes.fr).

Digital Object Identifier 10.1109/TIP.2008.916047

synthetic and real SAR images. Conclusions and perspectives are finally reported in Section VI.

II. MONOSENSOR MULTIVARIATE GAMMA DISTRIBUTIONS

A. Definition

A random vector $\mathbf{X} = (X_1, \dots, X_d)^T$ is distributed according to an MoMGD on \mathbb{R}_+^d with shape parameter q and scale parameter P if its moment generating function, or Laplace transform, is defined as [5]

$$\psi_{q,P}(\mathbf{z}) = E \left(e^{-\sum_{i=1}^d X_i z_i} \right) = [P(\mathbf{z})]^{-q} \quad (1)$$

where $\mathbf{z} = (z_1, \dots, z_d)$, $q \geq 0$ and $P(\mathbf{z})$ is a so-called affine polynomial.¹ The Laplace transform of X_i is obtained by setting $z_j = 0$ for $j \neq i$ in (1). This shows that X_i is distributed according to a univariate gamma distribution with shape parameter q and scale parameter p_i , denoted as $X_i \sim \mathcal{G}(q, p_i)$. Thus, all margins of \mathbf{X} are univariate gamma distributions with the same shape parameter q .

A monosensor bivariate gamma distribution (MoBGD) corresponds to the particular case $d = 2$ and is defined by its affine polynomial

$$P(\mathbf{z}) = 1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2 \quad (2)$$

with the following conditions:

$$p_1 > 0, p_2 > 0, p_{12} > 0, p_1 p_2 - p_{12} \geq 0. \quad (3)$$

It is important to note that the conditions (3) ensure that (2) is the Laplace transform of a probability distribution defined on $[0, \infty]^2$. However, in the general case ($d > 2$), determining necessary and sufficient conditions on P and q such that (1) is the Laplace transform of a probability distribution defined on $[0, \infty]^d$ is a difficult problem (see [5] for more details). The main properties of MoBGDs have been studied in [6].

Some important results required for the present paper are recalled below.

B. Moments

The moments of an MoBGD can be obtained by differentiating the Laplace transform (2). For instance, the mean and variance of X_i (denoted as $E[X_i]$ and $\text{var}(X_i)$, respectively) can be expressed as follows:

$$E[X_i] = q p_i, \text{var}(X_i) = q p_i^2 \quad (4)$$

for $i = 1, 2$. Similarly, the covariance $\text{cov}(X_1, X_2)$ and correlation coefficient $r(X_1, X_2)$ of an MoBGD are

$$\begin{aligned} \text{cov}(X_1, X_2) &= E[X_1 X_2] - E[X_1]E[X_2] = q(p_1 p_2 - p_{12}) \\ r(X_1, X_2) &= \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var}(X_1)} \sqrt{\text{var}(X_2)}} \\ &= \frac{p_1 p_2 - p_{12}}{p_1 p_2}. \end{aligned} \quad (5)$$

¹A polynomial $P(\mathbf{z})$ where $\mathbf{z} = (z_1, \dots, z_d)$ is *affine* if the one variable polynomial $z_j \mapsto P(\mathbf{z})$ can be written $A z_j + B$ (for any $j = 1, \dots, d$), where A and B are polynomials with respect to the z_i s with $i \neq j$.

It is important to note that when $\text{cov}(X_1, X_2) = 0$ (or equivalently $p_{12} = p_1 p_2$) the Laplace transform of \mathbf{X} can be factorized as follows:

$$\begin{aligned} \psi_{q,P}(z_1, z_2) &= [1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2]^q \\ &= [1 + p_1 z_1]^q [1 + p_2 z_2]^q \end{aligned}$$

where the two factors in the right hand side are the Laplace transforms of X_1 and X_2 . As a consequence, the random variables X_1 and X_2 of an MoBGD are independent if and only if they are uncorrelated (as in the Gaussian case).

C. Probability Density Function (PDF)

The pdf of an MoBGD can be expressed as follows (see [8, p. 436] for a similar result):

$$\begin{aligned} f_{2D}(\mathbf{x}) &= \exp \left(-\frac{p_2 x_1 + p_1 x_2}{p_{12}} \right) \\ &\quad \times \frac{x_1^{q-1} x_2^{q-1}}{p_{12}^q \Gamma(q)} f_q(c x_1 x_2) \mathbb{1}_{\mathbb{R}_+^2}(\mathbf{x}) \end{aligned} \quad (6)$$

where $\mathbb{1}_{\mathbb{R}_+^2}(\mathbf{x})$ is the indicator function on $[0, \infty]^2$ ($\mathbb{1}_{\mathbb{R}_+^2}(\mathbf{x}) = 1$ if $x_1 > 0, x_2 > 0$ and $\mathbb{1}_{\mathbb{R}_+^2}(\mathbf{x}) = 0$ otherwise), $c = (p_1 p_2 - p_{12})/p_{12}^2$ and $f_q(z)$ is related to the confluent hypergeometric function [8, p. 462] defined by

$$f_q(z) = \sum_{k=0}^{\infty} \frac{z^k}{k! \Gamma(q+k)}.$$

III. MULTISENSOR GAMMA DISTRIBUTIONS

A. Definition

A random vector $\mathbf{Y} = (Y_1, \dots, Y_d)^T$ is distributed according to a MuMGD with scale parameter P and shape parameter $\mathbf{q} = (q_1, \dots, q_d)$, denoted as $\mathbf{X} \sim \mathcal{G}(\mathbf{q}, P)$, if it can be constructed as follows:

$$\begin{aligned} Y_1 &= X_1 \\ Y_i &= X_i + Z_i, \quad 2 \leq i \leq d \end{aligned} \quad (7)$$

where

- $\mathbf{X} = (X_1, \dots, X_d)^T$ is a random vector distributed according to an MoMGD on \mathbb{R}_+^d with shape parameter q_1 and scale parameter P , i.e., $\mathbf{X} \sim \mathcal{G}(q_1, P)$;
- Z_2, \dots, Z_d are independent random variables distributed according to univariate gamma distributions (with the convention $Z_i = 0$ when $q_i - q_1 = 0$) $Z_i \sim \mathcal{G}(q_i - q_1, p_i)$ with $q_i \geq q_1$;
- vector $\mathbf{Z} = (Z_2, \dots, Z_d)^T$ is independent on \mathbf{X} .

By using the independence property between \mathbf{X} and \mathbf{Z} , the Laplace transform of \mathbf{Y} can be written

$$\begin{aligned} \psi_{\mathcal{G}(\mathbf{q}, P)}(\mathbf{z}) &= E \left(e^{-\sum_{i=1}^d Y_i z_i} \right) \\ &= [P(\mathbf{z})]^{-q_1} \prod_{i=1}^d (1 - p_i z_i)^{-(q_i - q_1)}. \end{aligned} \quad (8)$$

By setting $z_j = 0$ for $j \neq i$ in (9), we observe that the random variable Y_i is distributed according to a univariate gamma distribution with scale parameter p_i and shape parameter q_i , i.e., $Y_i \sim \mathcal{G}(q_i, p_i)$. Thus, all margins of \mathbf{Y} have different shape parameters in the general case. Note that the definition above assumes that the first univariate margin Y_1 has a shape parameter q_1 smaller than all other shape parameters $q_i, i \geq 2$ without loss of generality. Note also that an MuMGD reduces to an MoMGD for $q_i = q_1, \forall i \geq 2$.

A multisensor bivariate gamma distribution (MuBGD) corresponds to the particular case $d = 2$ and is defined by its Laplace transform

$$\psi(\mathbf{z}) = \left(1 + \sum_{i=1}^2 p_i z_i + p_{12} z_1 z_2\right)^{-q_1} \times (1 - p_2 z_2)^{-(q_2 - q_1)} \quad (9)$$

with the following conditions:

$$p_1 > 0, p_2 > 0, p_{12} > 0, p_1 p_2 - p_{12} \geq 0 \text{ and } q_2 \geq q_1. \quad (10)$$

In the bidimensional case, the conditions (10) ensure that (9) is the Laplace transform of a probability distribution defined on $[0, \infty]^2$.

B. MuBGD Pdf

According to (7), a vector $\mathbf{Y} = (Y_1, Y_2)^T$ distributed according to an MuBGD (i.e., $\mathbf{Y} \sim \mathcal{G}(\mathbf{q}, P)$) is constructed from a random vector $\mathbf{X} = (X_1, X_2)^T$ distributed according to an MoBGD whose pdf is denoted as $f_{\mathbf{X}}(\mathbf{x})$ and a random variable $Z \sim \mathcal{G}(q_2 - q_1, p_2)$ independent on \mathbf{X} with pdf $f_Z(z)$. By using the independence assumption between \mathbf{X} and Z , the density of \mathbf{Y} can be expressed as

$$f_{\mathbf{Y}}(\mathbf{y}) = \int f_{\mathbf{X}}(y_1, s) f_Z(y_2 - s) ds. \quad (11)$$

Straightforward computations leads to the following expression:

$$f_{\mathbf{Y}}(\mathbf{y}) = \left(\frac{p_1 p_2}{p_{12}}\right)^{q_1} \frac{y_1^{q_1 - 1} y_2^{q_2 - 1} e^{-\left(\frac{p_2}{p_{12}} y_1 + \frac{p_1}{p_{12}} y_2\right)}}{p_1^{q_1} p_2^{q_2} \Gamma(q_2) \Gamma(q_1)} \times \Phi_3\left(q_2 - q_1; q_2; c \frac{p_{12}}{p_2} y_2, c y_1 y_2\right) \quad (12)$$

where $c = (p_1 p_2 - p_{12}) / p_{12}^2$ and where Φ_3 is the so-called Horn function. The Horn function is one of the twenty convergent confluent hypergeometric series of order two, defined as [9]

$$\Phi_3(a; b; x, y) = \sum_{m, n=0}^{\infty} \frac{(a)_m}{(b)_{m+n} m! n!} x^m y^n \quad (13)$$

where $(a)_m$ is the Pochhammer symbol such that $(a)_0 = 1$ and $(a)_{k+1} = (a+k)(a)_k$ for any positive integer k . It is interesting to note that the relation $f_q(c y_1 y_2) = \Phi_3(0; q; c(p_{12}/p_2) y_2, c y_1 y_2) / \Gamma(q)$ allows one to show that the MuBGD pdf defined in (13) reduces to the MoBGD pdf (6) for $q_1 = q_2 = q$.

C. MuBGD Moments

The moments of \mathbf{Y} can clearly be obtained from the moments of \mathbf{X} and Z . This section concentrates on MuBGDs defined by $\mathbf{Y} = (X_1, X_2 + Z)^T$, where $\mathbf{X} = (X_1, X_2)^T$ is an MoBGD

with mean (m_1, m_2) , correlation coefficient r' and shape parameter q_1 , and Z is a univariate gamma distribution with mean m_2 and shape parameter $q_2 - q_1$. Using the independence property between \mathbf{X} and Z , the following results can be obtained:

$$\begin{aligned} E[Y_1^m Y_2^n] &= \sum_{i=1}^n \binom{n}{i} E[X_1^m X_2^i] E[Z^{n-i}] \\ &= \sum_{i=1}^n \binom{n}{i} m_2^{n-i} \frac{(q_2 - q_1)_{n-i}}{(q_2 - q_1)^{n-i}} E[X_1^m X_2^i] \end{aligned} \quad (14)$$

for all $(m, n) \in \mathbb{N}^2$. The moments of an MoBGD were derived in [6]

$$\begin{aligned} E[X_1^{n_1} X_2^{n_2}] &= m_1^{n_1} m_2^{n_2} \frac{(q_1)_{n_1}}{q_1^{n_1}} \frac{(q_1)_{n_2}}{q_1^{n_2}} \\ &\times \sum_{k=0}^{\min(n_1, n_2)} \frac{(-n_1)_k (-n_2)_k (r')^k}{(q_1)_k k!} \end{aligned} \quad (15)$$

for all $(n_1, n_2) \in \mathbb{N}^2$. Expressions (14) and (15) can be used to derive analytical expressions of MuMGD moments. For instance, the first and second order moments can be written as

$$\begin{aligned} E[Y_i] &= q_i p_i, \quad \text{var}(Y_i) = q_i p_i^2, \quad i = 1, 2 \\ \text{cov}(Y_1, Y_2) &= \text{cov}(X_1, X_2) = q_1 (p_1 p_2 - p_{12}) \\ r(Y_1, Y_2) &= \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var}(Y_1)} \sqrt{\text{var}(Y_2)}} \\ &= \frac{q_1}{\sqrt{q_1 q_2}} \frac{p_1 p_2 - p_{12}}{p_1 p_2}. \end{aligned}$$

It is interesting to note that the conditions (3) ensure that the correlation coefficient satisfy the constraint $0 \leq r(Y_1, Y_2) \leq \sqrt{q_1/q_2}$. In other words, the normalized correlation coefficient defined by

$$r'(Y_1, Y_2) = \sqrt{\frac{q_2}{q_1}} r(Y_1, Y_2) = \frac{p_1 p_2 - p_{12}}{p_1 p_2}$$

is such that $0 \leq r'(Y_1, Y_2) \leq 1$. As explained in Section II-B, the random variables X_1 and X_2 are independent if and only if $p_1 p_2 - p_{12} = 0$. Since Z_2 is independent from X_1 and X_2 , a necessary and sufficient condition for the margins of an MuBGD Y_1 and Y_2 to be independent is $r'(Y_1, Y_2) = 0$. Note, finally, that for known values of the shape parameters q_1 and q_2 , an MuBGD is fully characterized by the parameter vector $\boldsymbol{\theta} = (E[Y_1], E[Y_2], r'(Y_1, Y_2))$, since $\boldsymbol{\theta}$ and (p_1, p_2, p_{12}) are related by a one-to-one transformation.

IV. PARAMETER ESTIMATION FOR MUBGDS

This section studies different methods for estimating the parameters of MuBGDs.² The following notations are used in the rest of this paper:

$$m_1 = E[Y_1], m_2 = E[Y_2], r' = r(Y_1, Y_2) \sqrt{\frac{q_2}{q_1}}$$

inducing $\boldsymbol{\theta} = (m_1, m_2, r')^T$. Note that the parameters p_1, p_2 and p_{12} can be expressed as functions of $\boldsymbol{\theta}$ as follows $p_1 =$

²The results proposed here could be used to estimate the parameters of MuMGDs by using the concept of composite likelihood. The interested reader is invited to consult [10], [11], and references therein for more details.

$(m_1/q_1), p_2 = (m_2/q_2)$ and $p_{12} = (m_1 m_2 / q_1 q_2)(1-r')$. Note also that the parameters q_1 and q_2 are assumed to be known in this paper, as in most practical applications. In the case where q_1 and q_2 are unknown, these parameters should be included in θ and estimated jointly with m_1, m_2 , and r' .³

A. Maximum Likelihood (ML) Method

1) *Principles*: The ML method can be applied to \mathbf{Y} since a closed-form expression of its pdf is available. After removing the terms which do not depend on θ , the log-likelihood function of \mathbf{Y} can be written

$$\begin{aligned} l(\mathbf{Y}; \theta) &= -nq_1 \log(1-r') - nq_1 \log m_1 - nq_2 \log m_2 \\ &\quad - n \frac{q_1}{m_1(1-r')} \bar{Y}_1 - n \frac{q_2}{m_2(1-r')} \bar{Y}_2 \\ &\quad + \sum_{i=1}^n \log \Phi_3(q_2 - q_1; q_2; dY_2^i, cY_1^i Y_2^i) \end{aligned} \quad (16)$$

where $d = (r' q_2 / m_2 (1-r'))$, $\bar{Y}_1 = (1/n) \sum_{i=1}^n Y_1^i$, $\bar{Y}_2 = (1/n) \sum_{i=1}^n Y_2^i$ are the sample means of Y_1 and Y_2 and c defined previously can be expressed as function of θ using the relation $c = (r' q_1 q_2 / m_1 m_2 (1-r')^2)$. By differentiating the log-likelihood with respect to (wrt) θ , the MLE of θ is obtained as a solution of

$$\begin{aligned} \mathbf{u}(\mathbf{Y}; \theta) &= \left(\frac{\partial l(\mathbf{Y}; \theta)}{\partial m_1}, \frac{\partial l(\mathbf{Y}; \theta)}{\partial m_2}, \frac{\partial l(\mathbf{Y}; \theta)}{\partial r'} \right)^T \\ &= \mathbf{0}^T \end{aligned}$$

where $\mathbf{u}(\mathbf{Y}; \theta) = (\partial l(\mathbf{Y}; \theta) / \partial \theta)$ is the so-called score function, or equivalently by solving

$$-nq_1 + \frac{nq_1}{m_1(1-r')} \bar{Y}_1 - r' \Delta_2 = 0 \quad (17)$$

$$-nq_2 + \frac{nq_2}{m_2(1-r')} \bar{Y}_2 - r'(\Delta_1 + \Delta_2) = 0 \quad (18)$$

$$nq_1 - \sum_{i=1}^2 \frac{nq_i}{m_i(1-r')} \bar{Y}_i + \Delta_1 + (1+r')\Delta_2 = 0 \quad (19)$$

with

$$\begin{aligned} \Delta_1 &= \frac{q_2 - q_1}{m_2(1-r')} \sum_{i=1}^n Y_2^i \\ &\quad \times \frac{\Phi_3(q_2 - q_1 + 1; q_2 + 1; dY_2^i; cY_1^i Y_2^i)}{\Phi_3(q_2 - q_1; q_2; dY_2^i; cY_1^i Y_2^i)} \\ \Delta_2 &= \frac{q_1(1-r')^{-2}}{m_1 m_2} \sum_{i=1}^n Y_1^i Y_2^i \\ &\quad \times \frac{\Phi_3(q_2 - q_1; q_2 + 1; dY_2^i; cY_1^i Y_2^i)}{\Phi_3(q_2 - q_1; q_2; dY_2^i; cY_1^i Y_2^i)}. \end{aligned}$$

³The interested reader is invited to consult [12] for a related example where the shape parameter of a mono sensor multivariate gamma distribution ($q_1 = q_2 = q$) was estimated from mixed Poisson data. This section addresses the problem of estimating the unknown parameter vector θ from n vectors $\mathbf{Y} = (Y^1, \dots, Y^n)$, where $Y^i = (Y_1^i, Y_2^i)$ is distributed according to an MuBGD with parameter vector θ

The MLE of m_2 can be obtained by summing (17)–(19) and replacing the value of $\Delta_1 + \Delta_2$ in (18)

$$\widehat{m}_{2\text{ML}} = \bar{Y}_2. \quad (20)$$

The MLEs of m_1 and r' are obtained by replacing m_2 by $\widehat{m}_{2\text{ML}}$ in (16) and by maximizing the resulting log-likelihood $l(\mathbf{Y}; (m_1, \widehat{m}_{2\text{ML}}, r'))$ wrt m_1 and r' . This last maximization is achieved by using a constrained ($m_1 > 0$ and $r' \in [0, 1]$) quasi-Newton method, since an analytical expression of the log-likelihood gradient is available.⁴ Some elements regarding the numerical evaluation of the Horn Function are detailed in Appendices I and II. It is important to note that the MLE of m_1 differs from \bar{Y}_1 in the general case.⁵ Finally, the MLE of the correlation coefficient r is deduced by functional invariance as

$$\widehat{r}_{\text{ML}} = \sqrt{\frac{q_1}{q_2}} \widehat{r}'_{\text{ML}}.$$

2) *Performance*: The properties of the ML estimator $\widehat{m}_{2\text{ML}}$ can be easily derived from the properties of the univariate gamma distribution $\mathcal{G}(q_2, p_2)$. This estimator is obviously unbiased, convergent and efficient. However, the performance of $\widehat{m}_{1\text{ML}}$ and \widehat{r}_{ML} are more difficult to study. Of course, the MLE is known to be asymptotically unbiased and asymptotically efficient, under mild regularity conditions. Thus, the mean square error (MSE) of the estimates can be approximated for large data records by the Cramer–Rao lower bound (CRLB). For unbiased estimators, the CRLB is obtained by inverting the following Fisher information matrix I

$$I(\theta) = -E \left[\frac{\partial \mathbf{u}(\mathbf{Y}; \theta)}{\partial \theta} \right].$$

Thus, the computation of I requires to determine the negative expectations of second-order derivatives of $l(\mathbf{Y}; \theta)$ wrt m_1, m_2 and r in (16). Closed-form expressions for the elements of I are difficult to obtain because of the term $\log \Phi_3$. In such situation, it is very usual to approximate the expectations by using Monte Carlo methods. This will provide interesting approximations of the ML MSEs (see simulation results of Section V).

B. Inference Function for Margins (IFM)

1) *Principles*: IFM is a two-stage estimation method whose main ideas can be found for instance in [14, Ch. 10] and are summarized below in the context of MuBGDs.

- Estimate the unknown parameters m_1 and m_2 from the marginal distributions of Y_1 and Y_2 . This estimation is conducted by maximizing the marginal likelihoods $l(Y_1; m_1)$ and $l(Y_2; m_2)$ wrt m_1 and m_2 , respectively.
- Estimate the parameter r' by maximizing the joint likelihood $l(\mathbf{Y}; \widehat{m}_{1\text{ML}}, \widehat{m}_{2\text{ML}}, r')$ wrt r' . Note that the parameters m_1 and m_2 have been replaced in the joint likelihood by their estimates resulting from the first stage of IFM.

⁴The negative log-likelihood function has a unique minimum with respect to r' in all practical cases. The reader is invited to consult [13] for discussions and simulations results.

⁵There is no closed-form expression for the MLE of m_1 contrarily to m_2 . Indeed, there is some kind of dissymmetry between Y_1 and Y_2 inherent to the proposed model (7). This dissymmetry will disappear in the method based on the inference for margins studied in section B.

The IFM procedure is often computationally simpler than the ML method which estimates all the parameters simultaneously from the joint likelihood. Indeed, a numerical optimization with several parameters is much more time-consuming compared with several optimizations with fewer parameters. The marginal distributions of an MuBGD are univariate gamma distributions with shape parameters q_i and means m_i , for $i = \{1, 2\}$. Thus, the IFM estimators of m_1, m_2, r' are obtained as a solution of

$$\mathbf{g}(\mathbf{Y}; \boldsymbol{\theta}) = \left(\frac{\partial l_1(\mathbf{Y}_1; m_1)}{\partial m_1}, \frac{\partial l_2(\mathbf{Y}_2; m_2)}{\partial m_2}, \frac{\partial l(\mathbf{Y}; \boldsymbol{\theta})}{\partial r'} \right)^T = \mathbf{0}^T$$

where l_i is the marginal log-likelihood function associated to the univariate random variable Y_i , for $i = \{1, 2\}$, and l is the joint log-likelihood defined in (16). The IFM estimators of m_1 and m_2 are classically obtained from the properties of the univariate gamma distribution

$$\widehat{m}_{1\text{IFM}} = \bar{Y}_1, \quad \widehat{m}_{2\text{IFM}} = \bar{Y}_2. \quad (21)$$

The IFM estimator of r' is obtained by replacing m_1 and m_2 by \bar{Y}_1 and \bar{Y}_2 in (16) and by minimizing the resulting log-likelihood $l(\mathbf{Y}; \bar{Y}_1, \bar{Y}_2, r')$ wrt r' . This last minimization is achieved by using a constrained quasi-Newton method (with the constraint $r' \in [0, 1]$), since an analytical expression of the log-likelihood gradient is available.

Note that the ML method presented before requires to optimize the log-likelihood $l(\mathbf{Y}; (m_1, \widehat{m}_{2\text{ML}}, r'))$ wrt m_1 and r' whereas the IFM method only requires to optimize $l(\mathbf{Y}; \bar{Y}_1, \bar{Y}_2, r')$ wrt a single variable r' . The optimization procedure is, therefore, much less time-consuming for IFM than for the ML method. Note also that the estimator of m_2 is the same for the ML and IFM methods. Finally, it is interesting to point out that the joint likelihood is the product of univariate gamma pdfs when $r' = 0$. As a consequence, the ML and IFM estimators are the same when $r' = 0$.

2) *Performance*: Asymptotic properties of the IFM estimator can be derived from the set of inferences functions $\mathbf{g}(\mathbf{Y}; \boldsymbol{\theta})$ under the usual regularity conditions for the MLE (the interested reader is invited to consult [14] for more details). In particular, the IFM estimator of $\boldsymbol{\theta}$ denoted as $\widehat{\boldsymbol{\theta}}_{\text{IFM}}$ is such that $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\text{IFM}} - \boldsymbol{\theta})$ converges in distribution to the normal distribution $\mathcal{N}(\mathbf{0}, V)$, where the asymptotic covariance matrix V is the inverse Godambe information matrix defined as

$$V = D_g^{-1} M_g D_g^{-T} \quad (22)$$

where

$$D_g = \mathbb{E}[\partial \mathbf{g}(\mathbf{Y}; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}]$$

$$M_g = \mathbb{E}[\mathbf{g}(\mathbf{Y}; \boldsymbol{\theta}) \mathbf{g}^T(\mathbf{Y}; \boldsymbol{\theta})].$$

Straightforward computations yield the following expressions for matrices D_g and M_g [15]:

$$D_g = \begin{pmatrix} \mathcal{J}_{11} & 0 & 0 \\ 0 & \mathcal{J}_{22} & 0 \\ \mathcal{I}_{13} & \mathcal{I}_{23} & \mathcal{I}_{33} \end{pmatrix}$$

$$M_g = \begin{pmatrix} \mathcal{J}_{11} & \mathcal{J}_{12} & 0 \\ \mathcal{J}_{12} & \mathcal{J}_{22} & 0 \\ 0 & 0 & \mathcal{I}_{33} \end{pmatrix}$$

where

- \mathcal{I}_{ij} are the entries of the Fisher information matrix, $I = (\mathcal{I}_{ij})_{1 \leq i, j \leq 3}$;
- \mathcal{J}_{11} and \mathcal{J}_{22} are the Fisher information associated with the margins Y_1 and Y_2 , respectively;
- $\mathcal{J}_{12} = \mathbb{E}[g_1(\mathbf{Y}; \boldsymbol{\theta}) g_2(\mathbf{Y}; \boldsymbol{\theta})] = \mathbb{E}[(\partial l_1(\mathbf{Y}_1; m_1) / \partial m_1)(\partial l_2(\mathbf{Y}_2; m_2) / \partial m_2)]$.

The terms $\mathcal{J}_{11}, \mathcal{J}_{22}, \mathcal{J}_{12}$ associated to MuBGDs are easily derived by considering the univariate log-likelihoods $l_1(\mathbf{Y}_1; m_1)$ and $l_2(\mathbf{Y}_2; m_2)$

$$\mathcal{J}_{11} = \frac{q_1}{m_1^2}, \quad \mathcal{J}_{22} = \frac{q_2}{m_2^2}$$

$$\mathcal{J}_{12} = \frac{q_1}{m_1 m_2} r'.$$

As explained in Section IV-A2, the Fisher information entries \mathcal{I}_{ij} do not have closed-form expressions. Consequently, these terms have been computed by using numerical integration (Simpson quadrature). Note that this method allows one to control the approximation error.

C. Method of Moments

The estimators of (m_1, m_2, r) derived in this paper will be compared to the standard estimators based on the method of moments

$$\widehat{m}_{1\text{Mo}} = \bar{X}_1, \quad \widehat{m}_{2\text{Mo}} = \bar{X}_2, \quad (23)$$

$$\widehat{r}_{\text{Mo}} = \frac{\sum_{i=1}^n (X_1^i - \bar{X}_1)(X_2^i - \bar{X}_2)}{\sqrt{\sum_{i=1}^n (X_1^i - \bar{X}_1)^2} \sqrt{\sum_{i=1}^n (X_2^i - \bar{X}_2)^2}}. \quad (24)$$

The asymptotic performance of the estimator $\widehat{\boldsymbol{\theta}}_{\text{Mo}} = (\widehat{m}_{1\text{Mo}}, \widehat{m}_{2\text{Mo}}, \widehat{r}_{\text{Mo}})^T$ can be derived by imitating the results of [16] derived in the context of time series analysis. More precisely, the moment estimator of $\boldsymbol{\theta}$ can be rewritten as

$$\widehat{\boldsymbol{\theta}}_{\text{Mo}} = g(\mathbf{s}_n)$$

$$= \left(s_n^1, s_n^2, \frac{s_n^5 - s_n^1 s_n^2}{\sqrt{s_n^3 - (s_n^1)^2} \sqrt{s_n^4 - (s_n^2)^2}} \right)^T$$

where

$\mathbf{s}_n = (s_n^1, \dots, s_n^5)^T = [(1/n) \sum_{i=1}^n Y_1^i, (1/n) \sum_{i=1}^n Y_2^i, (1/n) \sum_{i=1}^n (Y_1^i)^2, (1/n) \sum_{i=1}^n (Y_2^i)^2, (1/n) \sum_{i=1}^n Y_1^i Y_2^i]^T$ contains the appropriate first and second order empirical moments of $\mathbf{Y} = (Y_1, Y_2)^T$. By denoting as $\Sigma(\boldsymbol{\theta}) = n \text{cov}(\mathbf{s}_n)$ the covariance matrix of the vector $\sqrt{n} \mathbf{s}_n$ and $G(\boldsymbol{\theta})$ the jacobian of the function g defined above, it can be shown that the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\text{Mo}} - \boldsymbol{\theta})$ is $G(\boldsymbol{\theta}) \Sigma(\boldsymbol{\theta}) G(\boldsymbol{\theta})^T$ [16]. The determination of the covariance matrix $\Sigma(\boldsymbol{\theta})$ requires to know appropriate theoretical moments of $\mathbf{Y} = (Y_1, Y_2)^T$ (up to the fourth order). These moments can be determined by using the results of Section III-C. The reader is invited to consult [6] for more details regarding the asymptotic performance of the moment estimator $\widehat{\boldsymbol{\theta}}_{\text{Mo}}$ for MuBGDs.

V. SIMULATION RESULTS

Many simulations have been conducted to validate the previous theoretical results. This section presents some experiments obtained with a vector $\mathbf{Y} = (Y_1, Y_2)^T$ distributed according to an MuBGD whose Laplace transform is (9).

A. Generation of Synthetic Data

According to the definition given in Section III-C, a vector \mathbf{Y} distributed according to an MuBGD can be generated by adding a random variable Z distributed according to a univariate gamma distribution to a random vector \mathbf{X} distributed according to an MoBGD. The generation of a vector \mathbf{X} whose Laplace transform is (1) has been described in [6] and is summarized as follows:

- simulate $2q$ independent multivariate Gaussian vectors of \mathbb{R}^2 denoted as Z^1, \dots, Z^{2q} with means (0,0) and the 2×2 covariance matrix $C = (c_{i,j})_{1 \leq i,j \leq 2}$ with $c_{i,j} = r^{|i-j|/2}$;
- compute the k th component of $\mathbf{X} = (X_1, X_2)$ as $X_k = (m_k/2q) \sum_{1 \leq i \leq 2q} (Z_k^i)^2$, Z_k^i being the k th component of Z^i .

It is interesting to note that the generation of a random vector distributed according to a multivariate gamma distribution is straightforward here since $2q$ is an integer (this assumption is not a problem in practical applications since q is the number of looks of the SAR image). However, if $2q$ would not be an integer, the generation of the random vector \mathbf{X} could be achieved by using an accept-reject procedure such as the one detailed in [17, p. 51].

B. Estimation Performance

1) *ML Method and Method of Moments*: The first simulations compare the performance of the estimators based on the method of moments and the ML method as a function of the sample size n . Note that the possible values of n correspond to the numbers of pixels of squared windows of size $(2p+1) \times (2p+1)$, where $p \in \mathbb{N}$. These values are appropriate to the change detection problem. The number of Monte Carlo runs is 10 000 for all figures presented in this section. The other parameters for this example are $m_1 = 100, m_2 = 100, q_1 = 1$ (number of looks of the first image) and $q_2 = 2$ (number of looks of the second image). Fig. 1(a)–(c) shows the MSEs of the estimated normalized correlation coefficient for different values of r' ($r' = 0.2, r' = 0.5$, and $r' = 0.8$). The losange curves correspond to the estimator of moments whereas the triangle curves correspond to the MLE. Fig. 1 shows the interest of the ML method, which is much more efficient for this problem than the method of moments, particularly for large values of the correlation coefficient r' . Note that the theoretical asymptotic MSEs of both estimators are also depicted (continuous lines). They are clearly in good agreement with the estimated MSEs, even for small values of n . Finally, these figures show that “reliable” estimates of r' can be obtained for values of n larger than 9×9 , i.e., even for relatively small window sizes.

Fig. 2(a) and (b) compares the MSEs of the estimated mean m_1 obtained for the ML method and the method of moments for two values of r' ($r' = 0.8$ and $r' = 0.9$). Both estimators perform very similarly for this parameter, even if the difference is slightly more noticeable for larger values of r' . Note that the estimators of m_2 obtained for the ML and moment methods are the same. Thus, the corresponding MSEs have not been presented here for brevity.

2) *ML and IFM*: This section compares the performance of the ML and IFM estimators for the parameters r and m_1 . Fig. 3 first shows the asymptotic performance of both estimators by depicting the ratio of their asymptotic variances, referred to as

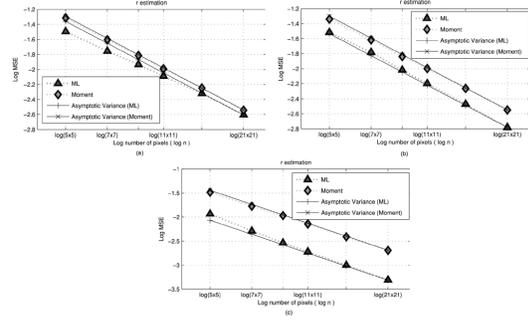


Fig. 1. Log MSEs versus $\log n$ for parameter r ($q_1 = 1, q_2 = 2, m_1 = 100$, and $m_2 = 100$).

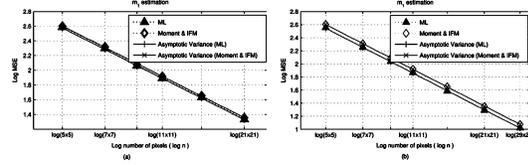


Fig. 2. Log MSEs versus $\log n$ for parameter m_1 ($q_1 = 1, q_2 = 2, m_1 = 100$, and $m_2 = 100$).

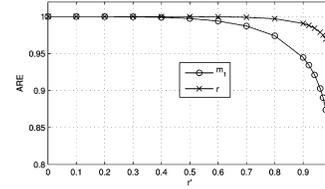


Fig. 3. ARE ($q_1 = 1, q_2 = 5, m_1 = 1$, and $m_2 = 1$).

asymptotic ratio efficiency (ARE), as a function of r . Fig. 3 shows that the ML and IFM estimators of the correlation coefficient r have very similar asymptotic variances when r' is not too close from 1. This result is confirmed in Fig. 4, which shows the MSEs of the estimated correlation coefficient obtained with the ML and IFM methods for different values of the sample size n (the parameters for this simulation are $q_1 = 1, q_2 = 2, m_1 = 100, m_2 = 100$, and $r' = 0.9$). Fig. 3 also shows that the asymptotic performance of the ML and IFM estimators for parameter m_1 differ significantly when r' approaches 1. However, this is not a major problem since the change detection algorithms proposed in this paper will be based on r only (see the next section). Based on these results, the IFM method will be preferred to the ML method since it involves much smaller computational cost.

C. Detection Performance

This section considers synthetic vectors $\mathbf{x} = (x_1, x_2)^T$ (coming from 762×292 synthetic images) distributed according to MuBGDs with $r = 0.3$ and $r = 0.7$, modeling the presence and absence of changes, respectively. The correlation

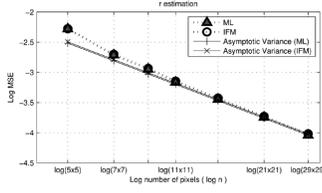


Fig. 4. Log MSEs versus $\log n$ for parameter r ($r' = 0.9$, $q_1 = 1$, $q_2 = 2$, $m_1 = 100$, and $m_2 = 100$).

coefficient r of each bivariate vector $\mathbf{x}^{(i,j)} = (x_1^{(i,j)}, x_2^{(i,j)})^T$ (for $1 \leq i \leq 762$, $1 \leq j \leq 292$) is estimated locally from pixels belonging to windows of size $n = (2p + 1) \times (2p + 1)$ centered around the pixel of coordinates (i, j) in the two analyzed images. The change detection problem in multisensor SAR images is addressed by using the following decision rule:

$$\begin{aligned} &\text{Decide } H_0 \text{ (absence of change) if } \hat{r} > \lambda \\ &\text{Decide } H_1 \text{ (presence of change) if } \hat{r} \leq \lambda \end{aligned} \quad (25)$$

where λ is a threshold depending on the probability of false alarm (PFA) and \hat{r} is an estimator of the correlation coefficient (obtained from the method of moments or the IFM method). The performance of the change detection strategy (25) can be defined by the two following probabilities [18, p. 34]:

$$\begin{aligned} P_D &= P[\text{accepting } H_1 | H_1 \text{ is true}] = P[\hat{r} < \lambda | H_1 \text{ is true}] \\ P_{FA} &= P[\text{accepting } H_1 | H_1 \text{ is true}] = P[\hat{r} < \lambda | H_1 \text{ is true}]. \end{aligned}$$

Thus, a pair (P_{FA}, P_D) can be defined for each value of λ . The curves representing P_D as a function of P_{FA} are called receiver operating characteristics (ROCs) and are classically used to assess detection performance [38, p. 38].

The ROCs for the change detection problem (25) are depicted in Fig. 5(a)–(c) for three representative values of (q_1, q_2) and two window sizes (9×9) and (21×21) . The IFM estimator clearly outperforms the moment estimator for these examples. Fig. 5(a) and (b) also shows that the detection performance seems to decrease when $q_2 - q_1$ increases, i.e., when the difference between the numbers of looks of the two images increases. In order to confirm this observation, we have derived theoretical ROCs by using the asymptotic Gaussian distribution for the estimated correlation coefficient (see Section IV-B2). In this case, by denoting r_0 and r_1 the true values of the correlation coefficient under hypotheses H_0 and H_1 , the following results can be obtained:

$$\begin{aligned} P_D &= P[\hat{r} < \lambda | H_1 \text{ is true}] = P[\hat{r} < \lambda | \hat{r} \sim \mathcal{N}(r_1, \sigma_1^2)] \\ P_{FA} &= P[\hat{r} < \lambda | H_0 \text{ is true}] = P[\hat{r} < \lambda | \hat{r} \sim \mathcal{N}(r_0, \sigma_0^2)] \end{aligned}$$

where σ_0^2 and σ_1^2 are the asymptotic variances of the estimated correlation coefficient \hat{r} under hypotheses H_0 and H_1 [calculated from the inverse Godambe information matrix defined in (22)]. By denoting as $\Phi(x)$ the cumulative distribution function of the Gaussian distribution $\mathcal{N}(0, 1)$, the following result is then classically obtained:

$$P_D = \Phi \left[\frac{r_0 - r_1}{\sigma_1} + \frac{\sigma_0}{\sigma_1} \Phi^{-1}(P_{FA}) \right]. \quad (26)$$

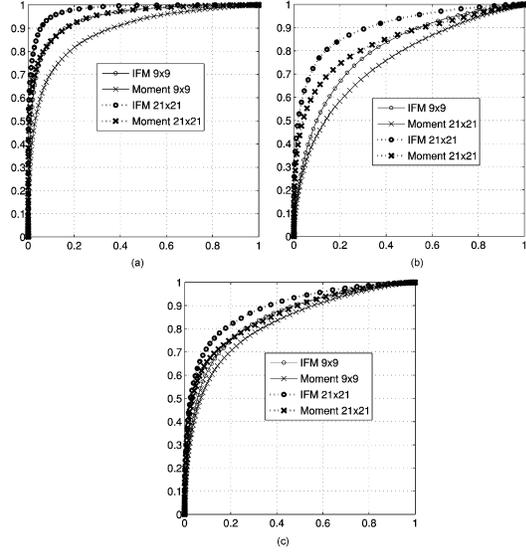


Fig. 5. ROCs for synthetic data.

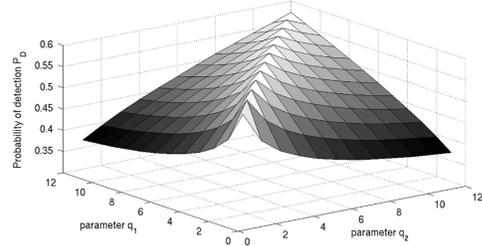


Fig. 6. P_D versus shape parameters q_1 and q_2 ($P_{FA} = 0.3$, $n = 1$).

This result provides theoretical asymptotic expressions for the ROCs associated to the detection problem (25) and allow us to analyze detection performance as functions of the MuBGD parameters. For instance, Fig. 6 shows P_D as functions of q_1 and q_2 for a given probability of false alarm $P_{FA} = 0.3$. Fig. 6 clearly confirms that the detection performance is a decreasing function of $q_2 - q_1$.

D. Change Detection in Real Images

This section first considers images acquired at different dates around Gloucester (U.K.) before and during a flood (on September 9, 2000 and October 21, 2000, respectively). The 1-look images as well as a mask indicating the pixels affected by the flood are depicted in Fig. 7(a)–(c). The reference map in Fig. 7(c) was obtained by photo-interpreters—who used the same SAR images we are using—and a reference map built from Landsat and SPOT data acquired one day after the radar image. The original 1-look images have been transformed into images with larger numbers of looks by replacing each pixel by the average of pixels belonging to a given neighborhood.

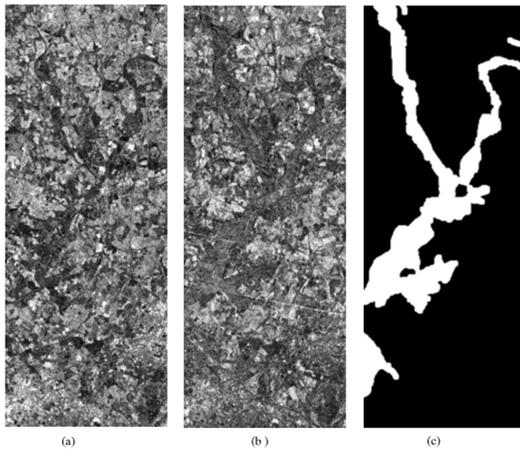


Fig. 7. ERS images of Gloucester before and after flood.

This section compares the performance of the following change detectors.

- The ratio edge detector which has been intensively used for SAR images [19], [20]. This detector mitigates the effects of the multiplicative speckle noise by computing the ratio of averages of pixel values belonging to neighborhoods of the pixels under consideration.
- The correlation change detector, where \hat{r} in (25) has been estimated with the moment estimator (referred to as “*Correlation Moment*”).
- The correlation change detector, where \hat{r} in (25) has been estimated with the IFM method for BGDs (referred to as “*Correlation IFM*”).

The ROCs for this change detection problem are shown in Fig. 8(a)–(c) for different window sizes ($n = 9 \times 9$, $n = 15 \times 15$ and $n = 21 \times 21$). The numbers of looks for the two images are $q_1 = 1$ and $q_2 = 5$. The correlation IFM detector clearly provides the best results.

The second set of experiments is related to a couple of Radarsat images acquired before and after the eruption of the Nyiragongo volcano which occurred in January 2002. The Radarsat images are depicted in Fig. 9(a) (before eruption) and (b) (after eruption). Note that some changes due to the eruption can be clearly seen on the landing track for example. Fig. 9(c) indicates the pixels of the image which have been affected by the eruption (white pixels). The ROCs for this change detection problem are shown in Fig. 10(a)–(c) for different window sizes ($n = 9 \times 9$, $n = 15 \times 15$, and $n = 21 \times 21$). The numbers of looks for the two images are $q_1 = 3$ and $q_2 = 5$. The correlation IFM detector provides better performance than the conventional correlation moment detector in all cases. The ratio edge detector also shows interesting detection performance for this example because the volcano eruption has produced significant changes in the pixel intensities. Note, however, that the proposed correlation IFM detector gives better performance for large PFAs. Even if these large PFA values are usually considered as a bad result in classical detection problems, the

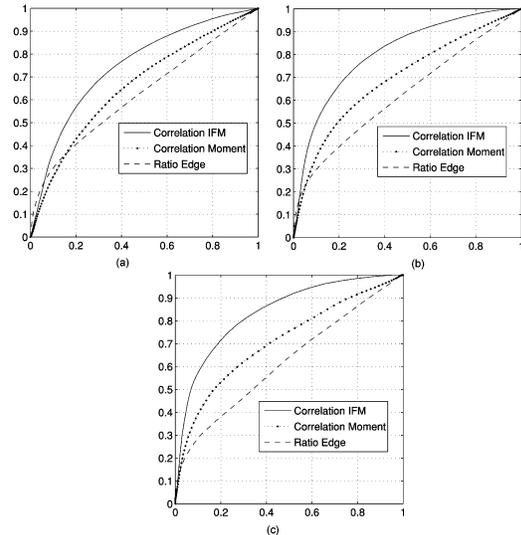


Fig. 8. ROCs for Gloucester images ($q_1 = 1$, $q_2 = 5$).

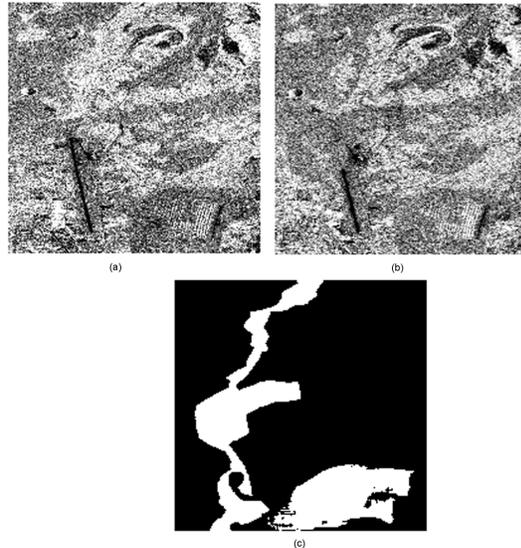
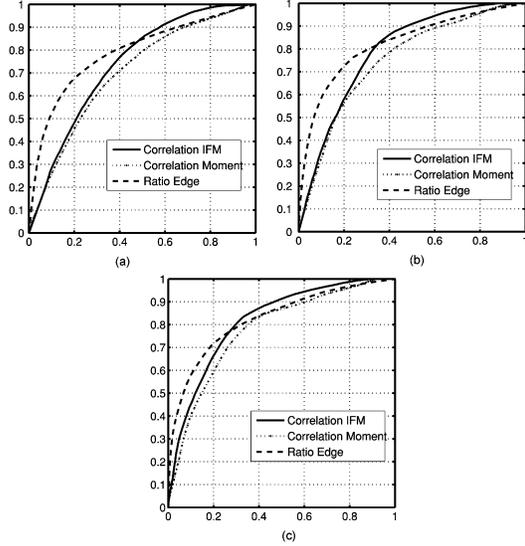


Fig. 9. Radarsat images of Nyiragongo before and after eruption.

reader has to bear in mind that when working with images, simple postprocessing strategies can dramatically improve the change detection performance. Indeed, when looking at detection maps, two types of false alarms can be observed: isolated pixels and boundary pixels. For the first type of errors, a simple median filter or a morphological opening, gives very good results. The second type of false alarm is due to the spatial extent of the estimation windows, which over-detect at the output boundaries of the change areas. This is not a main drawback in


 Fig. 10. ROCs for Nyiragongo images ($q_1 = 3, q_2 = 6$).

terms of change map production, since the change areas remain the same and only the spatial resolution of the map is affected.

VI. CONCLUSION

This paper studied a new family of multivariate gamma based distributions for multisensor SAR images referred to as MuMGDs. Estimation algorithms based on the ML method, the IFM principle and the methods of moments were studied to estimate the parameters of these distributions. In particular, the estimated correlation coefficient of MuMGDs showed interesting properties for detecting changes in radar images with different numbers of looks.

Being able to handle images with different numbers of looks is very useful, not only when the images have been acquired by different sensors, but also when both sensors have the same theoretical number of looks. Indeed, change detection algorithms require precise image co-registration which is usually achieved by image interpolation. Image interpolation and other image preprocessing steps modify locally the equivalent number of looks of the images. Therefore, even if the images have been acquired by the same sensor in the same imaging mode, differences in the equivalent number of looks can be observed. The algorithms presented in this paper could be used for detecting changes in this kind of images. Of course, in the case where the equivalent number of looks has to be estimated locally, an assessment of the influence of the estimation errors in the final MuMGD parameter estimation should be addressed. This point is currently under investigation.

APPENDIX I NUMERICAL EVALUATION OF THE HORN FUNCTION Φ_3

Some series representation in terms of special functions are useful to compute hypergeometric series of order two [21]. For the Horn function Φ_3 defined in (13), the following expansion is particularly useful:

$$\Phi_3(a; b; x, y) = \sum_{n=0}^{\infty} \frac{y^n}{(b)_n n!} {}_1F_1[a, b+n, x]$$

where ${}_1F_1$ is the confluent hypergeometric series of order one, i.e., ${}_1F_1[a, b, x] = \sum_{n=0}^{\infty} \frac{(a)_n}{(b)_n n!} x^n$. This confluent hypergeometric series ${}_1F_1[a, b, x]$ can be expressed as follows [22]:

$${}_1F_1[a, b, x] = \frac{\Gamma(b)}{\Gamma(a)} e^x x^{a-b} \sum_{i \geq 0} \frac{(b-a)_i (1-a)_i}{i! x^i} \times F_\gamma(x; i+b-a) \quad (27)$$

where $F_\gamma(x; \nu)$ is the cumulative distribution function of a univariate gamma distribution with shape parameter ν and scale parameter 1. Note that the summation in (27) is finite since $a \geq 1$ is an integer. This yields the following expression of Φ_3 :

$$\Phi_3(a; b; x, y) = \frac{\Gamma(b)}{\Gamma(a)} e^x x^{a-b} \sum_{n=0}^{\infty} \frac{(y/x)^n}{n!} \times \sum_{i \geq 0} \frac{(b+n-a)_i (1-a)_i}{i! x^i} F_\gamma(x; i+b-n-a) \quad (28)$$

where the last summation ($i \geq 0$) is finite. Equation (28) provides a numerically stable way of evaluating $\Phi_3(a; b; x, y)$ for large values of x and y . When (x, y) is close to $(0, 0)$, the definition of Φ_3 in (13) will be preferred.

APPENDIX II DERIVATIVES OF THE HORN FUNCTION Φ_3

From the series representation of the function Φ_3 defined in (13), the following results can be obtained:

$$\begin{aligned} \frac{\partial}{\partial x} \Phi_3(a; b; x, y) &= \sum_{m \geq 1, n \geq 0} \frac{(a)_m}{(b)_{m+n} (m-1)! n!} x^{m-1} y^n \\ &= \frac{\Gamma(a+1)}{\Gamma(a)} \frac{\Gamma(b)}{\Gamma(b+1)} \sum_{m, n \geq 0} \frac{(a+1)_m}{(b+1)_{m+n} m! n!} x^m y^n \\ &= \frac{a}{b} \Phi_3(a+1; b+1; x, y) \\ \frac{\partial}{\partial y} \Phi_3(a; b; x, y) &= \sum_{m \geq 0, n \geq 1} \frac{(a)_m}{(b)_{m+n} m! (n-1)!} x^m y^{n-1} \\ &= \frac{\Gamma(b)}{\Gamma(b+1)} \sum_{m, n \geq 0} \frac{(a)_m}{(b+1)_{m+n} m! n!} x^m y^n \\ &= \frac{1}{b} \Phi_3(a; b+1; x, y). \end{aligned}$$

REFERENCES

- [1] J. Inglada, "Similarity measures for multisensor remote sensing images," in *Proc. IEEE IGARSS*, Toronto, ON, Canada, Jun. 2002, pp. 104–106.
- [2] J. Inglada and A. Giros, "On the possibility of automatic multi-sensor image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2104–2120, Oct. 2004.
- [3] T. F. Bush and F. T. Ulaby, "Fading characteristics of panchromatic radar backscatter from selected agricultural targets," *IEEE Trans. Geosci. Remote Sens.*, vol. 13, no. 4, pp. 149–157, Apr. 1975.
- [4] S. B. Lev, D. Bshouty, P. Enis, G. Letac, I. L. Lu, and D. Richards, "The diagonal natural exponential families and their classification," *J. Oftheoret. Probab.*, vol. 7, no. 4, pp. 883–928, Oct. 1994.
- [5] P. Bernardoff, "Which multivariate gamma distributions are infinitely divisible?," *Bernoulli*, vol. 12, no. 1, pp. 169–189, 2006.
- [6] F. Chatelain, J.-Y. Tourneret, A. Ferrari, and J. Inglada, "Bivariate gamma distributions for image registration and change detection," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 169–1796, Jul. 2006.
- [7] F. Chatelain, J.-Y. Tourneret, J. Inglada, and A. Ferrari, "Parameter estimation for multivariate gamma distributions. Application to image registration," presented at the EUSIPCO, Florence, Italy, Sep. 2006.
- [8] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions*, 2nd ed. New York: Wiley, 2000, vol. 1.
- [9] F. O. A. Erdlyi, W. Magnus, and F. Tricomi, *Higher Transcendental Functions*. New York: Krieger, 1981, vol. 1.
- [10] B. G. Lindsay, "Composite likelihood methods," *Contemporary Math.*, vol. 50, pp. 221–239, 1988.
- [11] C. C. Heyde, *Quasi-Likelihood and Its Application. A General Approach to Optimal Parameter Estimation*. New York: Springer, 1997.
- [12] F. Chatelain and J.-Y. Tourneret, "Composite likelihood estimation for multivariate mixed Poisson distributions," in *Proc. IEEE-SP Workshop Statist. Signal Processing*, Bordeaux, France, Jul. 2005, pp. 49–54.
- [13] F. Chatelain and J.-Y. Tourneret, "Estimating the correlation coefficient of bivariate gamma distributions using the maximum likelihood principle and the inference functions for margins," Tech. Rep., Univ. Toulouse (IRIT/ENSEEIH), Toulouse, France, May 2007 [Online]. Available: <http://florent.chatelain.free.fr/Publi/ChatelainTechReport07.pdf>, available at
- [14] H. Joe, *Multivariate Models and Dependence Concepts*. London, U.K.: Chapman & Hall, May 1997, vol. 73.
- [15] H. Joe, "Asymptotic efficiency of the two-stage estimation method for copula-based models," *J. Multivar. Anal.*, vol. 94, no. 2, pp. 401–419, 2005.
- [16] B. Porat and B. Friedlander, "Performance analysis of parameter estimation algorithms based on high-order moments," *Int. J. Adapt. Control Signal Process.*, vol. 3, pp. 191–229, 1989.
- [17] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer, 2004.
- [18] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Part I*. New York: Wiley, 1968.
- [19] R. Touzi, A. Lopés, and P. Bousquet, "A statistical and geometrical edge detector for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 6, pp. 764–773, Nov. 1988.
- [20] E. J. M. Rignot and J. J. van Zyl, "Change detection techniques for ERS-1 SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 4, pp. 896–906, Apr. 1993.
- [21] F. D. Colavecchia and G. Gasaneo, "FL: A code to compute Appell's F_1 hypergeometric function," *Comput. Phys. Commun.*, vol. 157, pp. 32–38, Feb. 2004.
- [22] K. E. Muller, "Computing the confluent hypergeometric function, $M(a, b, x)$," *Numer. Math.*, vol. 90, no. 1, pp. 179–196, Nov. 2001.



Florent Chatelain received the Eng. degree in computer sciences and applied mathematics from ENSIMAG, Grenoble, France, and the M.Sc. degree in applied mathematics from the University Joseph Fourier of Grenoble, France, both in June 2004, and the Ph.D. degree in signal processing from the Signal and Communication Group, IRIT Laboratory, National Polytechnic Institute of Toulouse, France, in 2007.

He is currently a Postdoctoral Researcher at the INRIA Sophia Antipolis Méditerranée Laboratory, France, within the Ariana group. His research interests are centered around multivariate statistics applied to image processing and applications of stochastic geometry to the analysis of images.



Jean-Yves Tourneret received the Ingénieur degree in electrical engineering from the Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique, Toulouse (ENSEEIH), France, and the Ph.D. degree from the National Polytechnic Institute, Toulouse, in 1992.

He is currently a Professor with the University of Toulouse (ENSEEIH). He is a member of the IRIT Laboratory (UMR 5505 of the CNRS), where his research activity is centered around estimation, detection, and the classification of non-Gaussian and non-

stationary processes.

Dr. Tourneret was the Program Chair of the European Conference on Signal Processing (EUSIPCO), which was held in Toulouse, France, in 2002. He was also a member of the organizing committee for the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) held in Toulouse in 2006. He has been a member of different technical committees, including the Signal Processing Theory and Methods (SPTM) Committee of the IEEE Signal Processing Society.



Jordi Inglada received the Telecommunications engineer degree from the Universitat Politècnica de Catalunya, Spain, and from the École Nationale Supérieure des Télécommunications de Bretagne, France, in 1997, and the Ph.D. degree in signal processing and telecommunications from the Université de Rennes 1, France, in 2000.

He is with the Centre National d'Études Spatiales, French Space Agency, Toulouse, France, where he works in the field of remote sensing image processing. He is in charge of the development of image

processing algorithms for the operational exploitation of Earth Observation images, mainly in the fields of image registration, change detection, and object recognition.

A New Statistical Similarity Measure for Change Detection in Multitemporal SAR Images and Its Extension to Multiscale Change Analysis

Jordi Inglada and Grégoire Mercier, *Member, IEEE*

Abstract—In this paper, we present a new similarity measure for automatic change detection in multitemporal synthetic aperture radar images. This measure is based on the evolution of the local statistics of the image between two dates. The local statistics are estimated by using a cumulant-based series expansion, which approximates probability density functions in the neighborhood of each pixel in the image. The degree of evolution of the local statistics is measured using the Kullback–Leibler divergence. An analytical expression for this detector is given, allowing a simple computation which depends on the four first statistical moments of the pixels inside the analysis window only. The proposed change indicator is compared to the classical mean ratio detector and also to other model-based approaches. Tests on the simulated and real data show that our detector outperforms all the others. The fast computation of the proposed detector allows a multiscale approach in the change detection for operational use. The so-called multiscale change profile (MCP) is introduced to yield change information on a wide range of scales and to better characterize the appropriate scale. Two simple yet useful examples of applications show that the MCP allows the design of change indicators, which provide better results than a monoscale analysis.

Index Terms—Change detection, Edgeworth series expansion, Kullback–Leibler (KL) divergence, multiscale change profile (MCP), multitemporal synthetic aperture radar (SAR) images.

I. INTRODUCTION

REMOTE-SENSING imagery is a precious tool for rapid-mapping applications. In this context, one of the main uses of remote sensing is the detection of changes occurring after a natural or anthropic disaster. Since they are abrupt and seldom predictable, these events cannot be well temporally sampled—in the Shannon sense—by the polar orbit satellites, which provide the medium, high, and very high resolution imagery needed for an accurate analysis of the land cover. Therefore, rapid mapping is often produced by detecting the changes between an acquisition after the event and available archive data.

This change-detection procedure is made difficult due to the time constraints imposed by the emergency context. Indeed, the first available acquisition after the event has to be used, whatever its modality, which is more likely to be a radar image, due to weather and daylight constraints.

The kind of changes produced by the event of interest are often difficult to model. The same kind of event—a flood—can have different signatures, depending on where it happens—high-density built-up areas, agricultural areas, etc.—and on the characteristics of the sensor. Also, the changes of interest are all mixed up with normal changes, which can be the majority if the time gap between the two acquisitions is too long.

All these issues present us with a very difficult problem: detecting abrupt unmodeled transitions in a temporal series with only two dates.¹

From this position of the problem, one can make the straightforward deduction that pixelwise comparison between the two images will not be robust enough.

In the case of radar acquisitions, the standard detector is based on the ratio of local means [3]. This detector is robust to speckle noise, but it is limited to the comparison of first-order statistics. The classical model for synthetic aperture radar (SAR) intensity introduced by Ulaby *et al.* [4] assumes that the texture is a zero-mean multiplicative contribution. Therefore, changes taking place at the texture level, which preserve the mean value, will not be detected by the mean-ratio detector (MRD). One can, thus, assume a miss-detection behavior of the detectors using only the mean pixel values. This remark invites a more accurate analysis of the local statistics of the images to be compared. Bujor *et al.* [5] did a very interesting work by analyzing the interest of higher order statistics for change detection in SAR images. They concluded that the ratio of means was useful for step changes and that the second- and third-order log-cumulants were useful for progressive changes appearing in consecutive images in multitemporal series. Since higher order statistics seem to be helpful, one may want to compare the local probability density functions (pdfs) of the neighborhood of the homologous pixels of the pair of images used for the change detection.

Of course, this assumes that the pdfs are known and that there exists a robust way to compare them. The estimation of pdfs can be made with different approaches, but the straightforward histogram method should be avoided due to the need of a high number of samples for the estimation. Indeed, small analysis window sizes are required to yield high-resolution change maps. In this paper, we will present several approaches for this estimation by using only a small number of samples for the local statistics estimation, up to the fourth order.

¹In the case where a sequence of several images is to be processed, the approaches presented in [1] and [2] may be applied.

Manuscript received February 16, 2006; revised December 1, 2006.
J. Inglada is with the Centre National d'Études Spatiales, 31401 Toulouse, France.

G. Mercier is with the École Nationale Supérieure des Télécommunications de Bretagne, F-29238 Brest, France.
Digital Object Identifier 10.1109/TGRS.2007.893568

Once the pdfs are estimated, their comparison can also be performed using different criteria. Information theory shows that a good measure is the Kullback–Leibler (KL) divergence, which is also called information gain. We will use a symmetrical version of this measure and show that it is superior to the classical detector when the pdfs are correctly estimated.

Therefore, these measures will be based on the comparison of local neighborhoods where an analysis window for the computation of the local estimation of probabilities is used. The problem, which arises here, is the one of choice of the window size. Since we are facing unmodeled changes, we cannot choose the window size to fit the size of the expected changes. An inappropriate window size can produce miss- and overdetections: 1) When using a small window for a correlation analysis, no detection will be performed in a homogeneous area, which was globally changed to another homogeneous area, and 2) on the contrary, when using a larger window size, change areas have to be of larger size or strong in intensity (relative to the measure) to be detected. In these cases, it will produce a coarse-resolution change map. One way to overcome this problem is by applying a multiscale change-detection analysis.

Scale is to be understood in its geographic meaning, which is the spatial extent of the study area. It does not refer to the cartographic meaning of scale (the larger the scale, the more detailed is the information [6]; for an interesting discussion on scale issues in remote sensing, see [7]).

Image-processing techniques for multiscale analysis often use the cartographic meaning and apply low-pass filtering and possibly subsampling. For change-detection analysis, this filtering and subsampling can be justified in the case where the images are not perfectly registered [8]. In other cases, we think that it is better to use all the available information, that is, maximizing the number of available samples by using increasing window sizes. Nevertheless, pyramidal multiscale decompositions can also be useful in the case of phenomena characterizations (see, for example, [9]).

Therefore, the main point of the problem is how to choose the largest window size that robustly detects the changes, but which is small enough to preserve the resolution of the final map without misdetections.

We propose to use multiscale change profiles (MCP), which are defined as the change indicator for each pixel in the image as a function of the analyzing window size. The computation of the change detection for each window size can be very time-consuming. We present here a method for the computation of these profiles, which allows the change indicator at scale n to be computed from the value obtained at scale $n - 1$ plus a correction term which takes into account the addition of new samples only. Analytical expressions are given for three different change indicators. This paper proposes the following three main contributions:

- 1) an information-theory-based similarity measure which uses full local statistics;
- 2) the use of cumulant-based series expansions of similarity measures, which allow a robust and fast computation by using a small number of samples;
- 3) the concept of MCP and its fast implementation using recurrence evaluations.

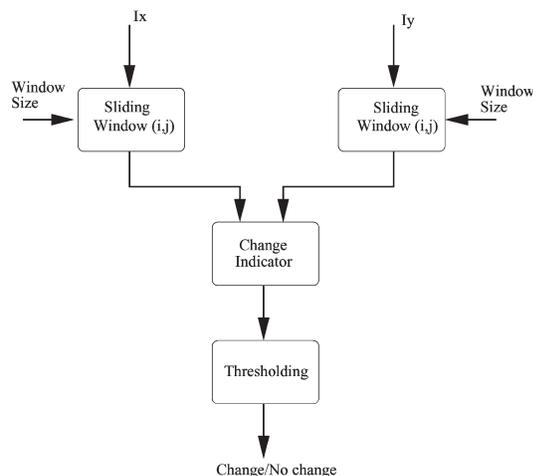


Fig. 1. Block diagram for a classical change-detection processing chain.

This paper is organized as follows. Section II presents the problem formulation; Section III introduces the measures used for the production of a change image; in Section IV, we introduce the concept of MCP and present the mathematical formulation, allowing its optimized computation; Sections V and VI present the results obtained on simulated and real data, respectively, and Section VII concludes this paper and proposes some directions for future work.

II. PROBLEM FORMULATION

Let us consider two coregistered SAR intensity images I_X and I_Y acquired at two different dates t_X and t_Y , respectively. Our objective is to produce a map representing the changes occurring in the scene between t_X and t_Y . The final goal of a change-detection analysis is to produce a binary map corresponding to the two classes: change and no change. The problem can be decomposed into two steps: the generation of a change image and the thresholding of the change image in order to produce the binary change map. Fig. 1 shows a block diagram describing a classical change-detection processing chain.

The overall performance of the detection system will depend on both the quality of the change image and the quality of the thresholding. In this paper, we choose to focus on the generation of an indicator of change for each pixel in the image. For interesting approaches in the field of unsupervised change image thresholding, the reader can refer to the works of Bruzzone and Prieto [10], [11], Bruzzone and Serpico [12], and Bazi *et al.* [13]. The reader may note that some of these approaches need a statistical modeling of the detectors' response, which is not presented here.

The change indicator can also be useful by itself. Indeed, often the end user of a change map wants not only the binary information, given after thresholding, but also an indicator of the intensity of the change and, eventually, a confidence level. In order to evaluate the quality of a change image independently of the choice of the thresholding algorithm, the evolution of

the detection probability P_{det} as a function of the false-alarm probability P_{fa} may be evaluated in the case where a set of constant thresholds is applied to the whole image. These are the so-called receiver operating characteristics (ROC), and the plots of $P_{\text{det}}(P_{\text{fa}})$ are called the ROC plots.

III. DISTANCE BETWEEN PROBABILITY DENSITIES

The main difficulty in the multitemporal analysis of SAR images is the presence of speckle noise. When moving away from interferometric configurations, the speckle is different from one image to the other, and it can induce a high number of false alarms in the change-detection procedure. Because of the multiplicative nature of speckle, the classical approach in SAR remote sensing involves using the ratio of the local means in the neighborhood of each pair of collocated pixels. The MRD is usually implemented as the following normalized quantity:

$$r_{\text{MRD}} = 1 - \min \left\{ \frac{\mu_X}{\mu_Y}, \frac{\mu_Y}{\mu_X} \right\} \quad (1)$$

where μ_X and μ_Y stand for the local mean values of the images before and after the event of interest, respectively. The logarithm of (1) may also be used. Nevertheless, this operation does not modify the performance of the detector, in terms of ROC, even if the contrast of the image of change indicator is modified. However, the logarithm is used since it modifies the initial pdf of the image of change indicator and then facilitates the development of Bayesian thresholding approaches [13].

This detector assumes that a change in the scene will appear as a modification of the local mean value of the image. If the change preserves the mean value but modifies the local texture, it will not be detected.

The change-detection algorithm proposed in this paper extends the MRD by analyzing the modification of the statistics of each pixel's neighborhood between the two acquisition dates. A pixel will be considered as having changed if its statistical distribution changes from one image to the other. In order to quantify this change, a measure, which maps the two estimated statistical distributions (one for each date at a collocated area) into a scalar change index is required. Several approaches could be taken into consideration: the mean-square error between the two distributions, the norm of a vector of moments, etc. We have chosen to use a measure derived from the information theory called the KL divergence [14].

A. KL Divergence

Let P_X and P_Y be two probability laws of the random variables X and Y . The KL divergence from Y to X , in the case where these two laws have the densities f_X and f_Y , is given by

$$K(Y|X) = \int \log \frac{f_X(x)}{f_Y(x)} f_X(x) dx. \quad (2)$$

The measure $\log(f_X(x)/f_Y(x))$ can be thought of as the information on x for the discrimination between the hypothesis \mathcal{H}_X and \mathcal{H}_Y , if hypothesis \mathcal{H}_X is associated with the pdf $f_X(x)$ and

\mathcal{H}_Y with $f_Y(x)$. Therefore, the KL divergence $K(Y|X)$ can be understood as the mean information for the discrimination between \mathcal{H}_X and \mathcal{H}_Y per observation. This divergence appears to be an appropriate tool to detect changes when we consider that changes on the ground induce different shapes on the local pdf.

Since the KL divergence can be understood as the entropy of P_X relative to P_Y , it is also called information gain. It can easily be proven that $K(Y|X) \geq 0$; $K(Y|X)$ vanishes only when the two laws are identical. $K(Y|X)$ can be used as a measure of the divergence from P_Y to P_X . This measure is not symmetric as it stands: $K(Y|X) \neq K(X|Y)$, but a symmetric version may be defined by writing

$$D(X, Y) = D(Y, X) = K(Y|X) + K(X|Y) \quad (3)$$

that will be called the KL distance.

In order to estimate the KL distance, the pdfs of the two variables to be compared have to be known. As stated in the Introduction, the processing of high-resolution change maps requires analysis windows of small size, which makes impossible the use of local histogram estimations. In the following sections, we will introduce several approaches, which allow the estimation of the pdfs by using a limited number of samples only. This requires some *a priori* information on the data, which can be introduced by using the models of local statistics.

B. Gaussian KL Detector (GKLD)

As shown in Section III, the classical detector of (1) uses first-order statistics only. Yet, second-order statistics are often used for SAR-image processing. For instance, many speckle-reduction filters [15]–[17] are based on the contrast coefficient σ_X^2/μ_X^2 , that is, the ratio between the variance and the square of the mean value. If the local statistics have to be compared up to the second order, the local random variables X and Y may be assumed to be normally distributed (i.e., of Gaussian law). Then, the pdf of P_X can be written as

$$f_X(x) = \mathcal{G}(x; \mu_X, \sigma_X) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}}. \quad (4)$$

An analogous expression holds for $f_Y(x)$.

Fig. 2(b) shows the Gaussian approximation of the probability distribution of a small region of interest (ROI) [Fig. 2(a)] extracted from a SAR image.

If this Gaussian model is used in (3), it yields the GKLD

$$r_{\text{GKLD}} = \frac{\sigma_X^4 + \sigma_Y^4 + (\mu_X - \mu_Y)^2(\sigma_X^2 + \sigma_Y^2)}{2\sigma_X^2\sigma_Y^2} - 1. \quad (5)$$

It can be seen that even in the case of identical mean values, this detector is able to underline the shading of textures, which is linked to the local variance evolution.

Nevertheless, the reader should note that the Gaussian model should not be used, since SAR-intensity values are always positive. However, this example has been given as a simple case of a parametric model, which takes into account second-order statistics. Since some Gaussianity may be introduced into

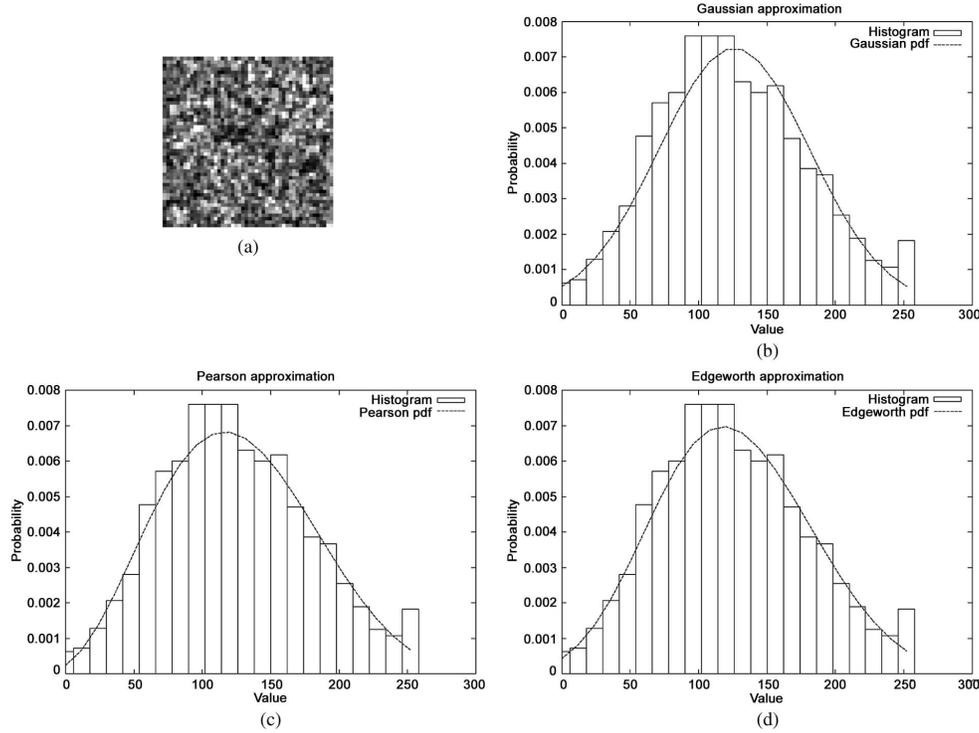


Fig. 2. Approximation of a histogram, coming from a 50×50 ROI, using three different strategies. The Pearson fitting yields a Beta distribution of the first type. (a) ROI extracted from a SAR image. (b) Gaussian fitting. (c) Pearson fitting: $\beta_1 = 2.51 \times 10^{-6}$ and $\beta_2 = 1.87$. (d) Edgeworth approximation.

the data when resampling and filtering the images during the preprocessing step, the Gaussian model may nevertheless be justified.

C. KLD Using the Pearson System

The drawback of the GKLD is that SAR-intensity statistics are not normally distributed, and the use of a bad model can induce bad performance of the detector, whatever the accuracy of the parameter estimation is. In the absence of texture, the radar intensity follows a Gamma distribution:

$$f_X(x) = \frac{1}{\Gamma(L)} \left(\frac{L}{\sigma_X} \right)^L e^{-\frac{Lx}{\sigma_X}} x^{L-1}. \quad (6)$$

The Gamma distribution is characterized by the following parameters: L is the number of looks, and σ_X is the square-root of the SAR-intensity image. $\Gamma(\cdot)$ is the Gamma function.

In the presence of texture, the local statistics can deviate from the Gamma distribution. For instance, if the texture is modeled by a Gamma distribution with a shape parameter ν , then the resulting intensity distribution follows a K -law [18]:

$$f_X(x) = \frac{2}{x} \left(\frac{L\nu x}{\mu_X} \right)^{L+\nu} \frac{1}{\Gamma(L)\Gamma(\nu)} K_{\nu-L} \left(2 \left(\frac{L\nu x}{\mu_X} \right)^{1/2} \right) \quad (7)$$

where $K(\cdot)$ is the modified Bessel function of the second kind and μ_X is the mean of X .

More generally, it is now accepted that the statistics of SAR images can be well modeled by the family of probability distributions known as the Pearson system [19]. It is composed of eight types of distributions, among which the Gaussian and the Gamma distributions may be found. The Pearson system is very easy to use since the type of distribution can be inferred from the following parameters:

$$\beta_{X;1} = \frac{\mu_{X;3}^2}{\mu_{X;2}^3} \quad \beta_{X;2} = \frac{\mu_{X;4}}{\mu_{X;2}^2}$$

where $\mu_{X;i}$ is the centered moment of order i of variable X . That means that any distribution from the Pearson system can be assessed from a given set of samples by computing the first four statistical moments. Any distribution, therefore, can be represented by a point on the $(\beta_{X;1}, \beta_{X;2})$ plane. For instance, the Gaussian distribution is located at $(\beta_{X;1}, \beta_{X;2}) = (0, 3)$, and the Gamma distributions lie on the $\beta_{X;2} = (3/2)\beta_{X;1} + 3$ line. Details about the theory of the Pearson system can be found in [20].

Fig. 2(c) shows an example of distribution estimation. The Pearson approximation fits the data better than the Gaussian one [Fig. 2(b)]. The example shown corresponds to a Beta

distribution of the first type with parameters $\beta_1 = 2.51 \times 10^{-6}$, $\beta_2 = 1.87$.

The Pearson-based KLD (PKLD) was originally introduced in [21]. It does not have a unique analytic expression, since eight different types of distribution may be held. Therefore, 64 different possibilities for the couples of pdf exist. Once the couple of pdfs is identified, the detection can be performed by numerical integration

$$r_{\text{PKLD}}(X, Y) = \int \left[\log \left(\frac{f_X(x; \beta_{X;1}, \beta_{X;2})}{f_Y(x; \beta_{Y;1}, \beta_{Y;2})} \right) \times f_X(x; \beta_{X;1}, \beta_{X;2}) + \log \left(\frac{f_Y(x; \beta_{Y;1}, \beta_{Y;2})}{f_X(x; \beta_{X;1}, \beta_{X;2})} \right) \times f_Y(x; \beta_{Y;1}, \beta_{Y;2}) \right] dx. \quad (8)$$

The correct way in proceeding to use the Pearson system is to choose a pdf using the estimated moments and, then, estimate the parameters of the distribution by maximum likelihood. While this can improve the results of the pdf estimation, the effect is not noticeable in terms of the estimation of the change indicator. This approach was not used in this paper in order to reduce the computation cost.

The reader should note that, in the case of single-look high-resolution SAR data (better than 10 m), other statistical models may be more appropriate, mainly on urban areas. Nicolas *et al.* have proposed a new model based on the log-statistics and a set of pdfs coming from the Fisher system of distributions [22], [23]. It has been applied to high-resolution SAR images on dense urban areas with promising results [24], [25].

D. Cumulant-Based KL Approximation

Instead of considering a parameterization of a given density, or set of densities, it may be of interest to describe the shape of the distribution. Such a description is based on quantitative terms that may approximate the pdf itself. The cumulants themselves do not provide such a pdf estimation directly but are necessary to describe its shape: For example, third-order (κ_3) is linked to the symmetry (i.e., skewness), while the fourth-order (κ_4) is linked to the flatness (i.e., kurtosis). The density is then estimated through a series expansion. In fact, the cumulant generating function is used for such an estimation. By definition, the cumulant generating function $\mathcal{K}_X(\cdot)$ of a random variable X is defined by

$$\mathcal{K}_X(\omega) = \ln \mathcal{M}_X(\omega) = \sum_n \kappa_{X;n} \frac{\omega^n}{n!}$$

where $\mathcal{M}_X(\cdot)$ is the moment-generating function defined by

$$\begin{aligned} \mathcal{M}_X(\omega) &= \int e^{\omega x} f_X(x) dx \\ &= \int \left(1 + \omega x + \frac{\omega^2}{2} x^2 + \dots \right) f_X(x) dx. \end{aligned}$$

For the case of the four first-order cumulants, the following expressions hold [26, p. 8]:

$$\begin{aligned} \kappa_{X;1} &= \mu_{X;1} \\ \kappa_{X;2} &= \mu_{X;2} - \mu_{X;1}^2 \\ \kappa_{X;3} &= \mu_{X;3} - 3\mu_{X;2}\mu_{X;1} + 2\mu_{X;1}^3 \\ \kappa_{X;4} &= \mu_{X;4} - 4\mu_{X;3}\mu_{X;1} - 3\mu_{X;2}^2 + 12\mu_{X;2}\mu_{X;1}^2 - 6\mu_{X;1}^4. \end{aligned} \quad (9)$$

Let us assume that the density to be approximated is not too far [27] from a Gaussian pdf (denoted as \mathcal{G}_X to underline the fact that it has the same mean and variance as X), that is, with a shape similar to the Gaussian distribution. The difference between $\mathcal{K}_X(\cdot)$ and $\mathcal{K}_{\mathcal{G}_X}(\cdot)$ can be written in terms of the difference of the cumulants $\kappa_{X;n} - \kappa_{\mathcal{G}_X;n}$. By inversion, the density may be expressed by a formal Taylor-like series

$$f_X(x) = \mathcal{G}_X(x) + c_1 \frac{d\mathcal{G}_X}{dx} + c_2 \frac{d^2\mathcal{G}_X}{dx^2} + \dots + .$$

Since a Gaussian density is used, it yields

$$f_X(x) = \sum_{r=0}^{\infty} c_r H_r(x) \mathcal{G}_X(x)$$

where $H_r(x)$ is known as the Chebyshev–Hermite polynomial of order r [27]. When choosing a Gaussian law so that its first and second cumulants agree with those of X , the number of terms of the series expansion is greatly reduced. This is the so-called Edgeworth series expansion. Its expression, when truncated to order of six, is the following:

$$\begin{aligned} f_X(x) = \left(1 + \frac{\kappa_{X';3}}{6} H_3(x) + \frac{\kappa_{X';4}}{24} H_4(x) + \frac{\kappa_{X';5}}{120} H_5(x) \right. \\ \left. + \frac{\kappa_{X';6} + 10\kappa_{X';3}^2}{720} H_6(x) \right) \mathcal{G}_X(x). \end{aligned} \quad (10)$$

It can be thought of as a model of the form $X = X_G + X'$, where X_G is a random variable with Gaussian density with the same mean and variance as X , and X' is a standardized version of X [28] with

$$X' = (X - \kappa_{X;1}) \kappa_{X;2}^{-1/2}.$$

Fig. 2(d) shows an example of such an approximation of a histogram.

The Edgeworth series expansion of the two pdfs f_X and f_Y may be introduced into the KL divergence (2). It yields an approximation of the KL divergence by Edgeworth series, truncated at a given order. In [29], such an approximation has been truncated to order of four by using the equality $(f_X/f_Y) = (f_X/\mathcal{G}_X) (\mathcal{G}_X/\mathcal{G}_Y) (\mathcal{G}_Y/f_Y)$, where \mathcal{G}_X (respectively, \mathcal{G}_Y)

is a Gaussian density of the same mean and variance as f_X (respectively, f_Y). Then

$$\begin{aligned} \text{KL}_{\text{Edgeworth}}(X, Y) &= \frac{1}{12} \frac{\kappa_{X';3}^2}{\kappa_{X;2}^2} \\ &+ \frac{1}{2} \left(\log \frac{\kappa_{Y;2}}{\kappa_{X;2}} - 1 + \frac{1}{\kappa_{Y;2}} \left(\kappa_{X;1} - \kappa_{Y;1} + \kappa_{X;2}^{1/2} \right)^2 \right) \\ &- \left(\kappa_{Y';3} \frac{a_1}{6} + \kappa_{Y';4} \frac{a_2}{24} + \kappa_{Y';3}^2 \frac{a_3}{72} \right) \\ &- \frac{1}{2} \frac{\kappa_{Y';3}^2}{36} \left(c_6 - 6 \frac{c_4}{\kappa_{X;2}} + 9 \frac{c_2}{\kappa_{Y;2}^2} \right) \\ &- 10 \frac{\kappa_{X';3} \kappa_{Y';3} (\kappa_{X;1} - \kappa_{Y;1}) (\kappa_{X;2} - \kappa_{Y;2})}{\kappa_{Y;2}^6} \end{aligned} \quad (11)$$

where

$$\begin{aligned} a_1 &= c_3 - 3 \frac{\alpha}{\kappa_{Y;2}} \\ a_2 &= c_4 - 6 \frac{c_2}{\kappa_{Y;2}} + \frac{3}{\kappa_{Y;2}^2} \\ a_3 &= c_6 - 15 \frac{c_4}{\kappa_{Y;2}} + 45 \frac{c_2}{\kappa_{Y;2}^2} - \frac{15}{\kappa_{Y;2}^3} \\ c_2 &= \alpha^2 + \beta^2 \\ c_3 &= \alpha^3 + 3\alpha\beta^2 \\ c_4 &= \alpha^4 + 6\alpha^2\beta^2 + 3\beta^4 \\ c_6 &= \alpha^6 + 15\alpha^4\beta^2 + 45\alpha^2\beta^4 + 15\beta^6 \\ \alpha &= \frac{\kappa_{X;1} - \kappa_{Y;1}}{\kappa_{Y;2}} \\ \beta &= \frac{\kappa_{X;2}^{1/2}}{\kappa_{Y;2}} \end{aligned}$$

Finally, the cumulant-based KLD (CKLD) between two observations X and Y is written as

$$r_{\text{CKLD}} = \text{KL}_{\text{Edgeworth}}(X, Y) + \text{KL}_{\text{Edgeworth}}(Y, X). \quad (12)$$

The reader should note the fact that, like for the Pearson-based detector, despite the apparent complexity of the formulas, and owing to (9), only the moments up to the order of four have to be computed.

IV. MULTISCALE-CHANGE PROFILE

Scale plays a strategic role in image analysis and more especially in change-detection applications. In Section I, it has been shown how an inappropriate scale of analysis can produce mis- or overdetections. Bovolo and Bruzzone [30] stress the fact that the scale of analysis is a key parameter for better discrimination between change and no-change areas. Such a point of view is implemented by a wavelet transform of the log-ratio estimated with a window of a user-defined size.

Instead of applying a multiscale analysis of the change image, the purpose here is to produce a set of change indicators estimated at various scales. We will call it MCP.

As stated in the Introduction, the multiscale term refers here to the size of the analyzing window. The MCP will therefore involve computing the change indicator for a pixel by using neighborhoods of increasing sizes. The so-called profile corresponds to the sequence of change measures as a function of scale. We will restrict our formulation to the case of the CKLD. Given the fact that this detector needs the estimation of the statistical moments of the samples inside the analyzing window, we are interested in finding an approach which avoids the computation from scratch of the moments at every scale.

A. Optimized Computation of the MCP

Let us consider the following problem: how to update the moments when an $(N + 1)$ th observation x_{N+1} is added to a set of N observations $\{x_1, x_2, \dots, x_N\}$ already processed? When considering raw moments of order r , the formulation comes easily as

$$\tilde{\mu}_{r,[N+1]} = \frac{N}{N+1} \tilde{\mu}_{r,[N]} + \frac{1}{N+1} x_{N+1}^r.$$

$\tilde{\mu}_{r,[N]}$ (respectively, $\tilde{\mu}_{r,[N+1]}$) stands for the raw moment of order r estimated with N samples (respectively, $N + 1$ samples). Since the analyzing window may contain textured areas, the mean value itself may be modified by the increase in the number of samples. Therefore, by using simple binomial properties, it can be shown that central moments may be characterized by

$$\begin{aligned} \mu_{1,[N]} &= \frac{1}{N} s_{1,[N]} \\ \mu_{r,[N]} &= \frac{1}{N} \sum_{\ell=0}^r \binom{r}{\ell} (-\mu_{1,[N]})^{r-\ell} s_{\ell,[N]} \end{aligned} \quad (13)$$

where the notation $s_{r,[N]} = \sum_{i=1}^N x_i^r$ has been used.

Hence, when considering a new sample x_{N+1} , each moment may be updated directly by using updates of $s_{1,[N+1]}$ and, then, $s_{r,[N+1]}$ for increasing values of order r . The Edgeworth series is also updated by transforming moments to cumulants [by using (9)] to be introduced in (10) and, then, in (11).

Fig. 3 shows an example of a pdf estimation on a homogeneous area [shown in Fig. 2(a)] when the window increases from 9×9 to 17×17 . In fact, the availability of updating the estimation of the distance between distributions from windows of any size without reprocessing the overall data is the most interesting point for multiscale change-detection purposes. This online multiscale moment estimation is the key for the operational use of the MCP concept.

For example, the computation of r_{CKLD} with windows of size ranging from 5×5 pixels to 51×51 pixels (22 different window sizes) takes only 42% additional time with respect to the computation of a single detection with a window of median size of 29×29 pixels (300 versus 210 s for a 800×400 pixel image).

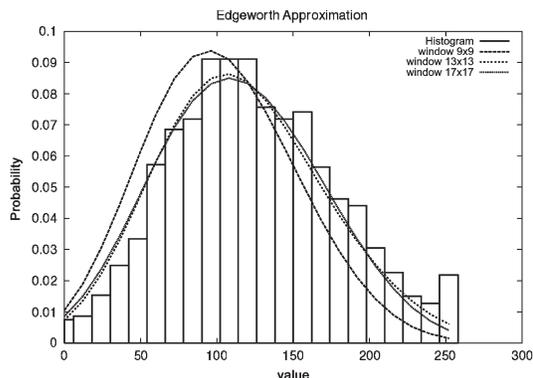


Fig. 3. Example of a pdf estimation update by increasing the sample set from a window of size 9×9 to 17×17 . The histogram has been estimated with a 17×17 window.

B. MCP Exploitation

The MCP computation produces a multichannel image (one scale per channel) whose pixels have to be transformed into scalar values in order to provide a change indicator. In order to exploit the information available at all scales, two approaches may be investigated. The first one consists in choosing the best scale for each image pixel. The second one consists in fusing the information available at all scales in order to provide a single-change value.

The development of an optimal approach for the exploitation of the MCP may be application-dependent. Indeed, multiscale fusion approaches could be tuned to a particular type of change—shape, nature, etc. In this section, two simple, yet useful, choices will be proposed, which will yield an improvement in comparison to the performance of single scale detection: 1) In order to choose the best scale, we will choose the one which produces the highest KLD value. This assumes that this scale is the one that is associated with the largest window inside a homogeneous area with respect to the classes change and no change. 2) The fusion of the multiscale information will be performed by using the principal component analysis (PCA). The first principal component of the MCP multichannel image will be considered as the change indicator. This corresponds to a linear combination of all scales which maximizes the contrast of the final image.

V. EXPERIMENTS WITH THE SIMULATED DATA

A. Data-Set Description

Simulations have been performed to better understand the behavior of the detectors relatively to a given kind of change and a given size of the change area. Since this study focuses on change detection on radar images, a speckle simulation is performed from a map of ground reflectivity. The simulated changes are applied on a small area, drawn as a circle, located in the center of the initial image.

The simulation procedure is based on the radar-image-formation mechanism. Each pixel is simulated with a given

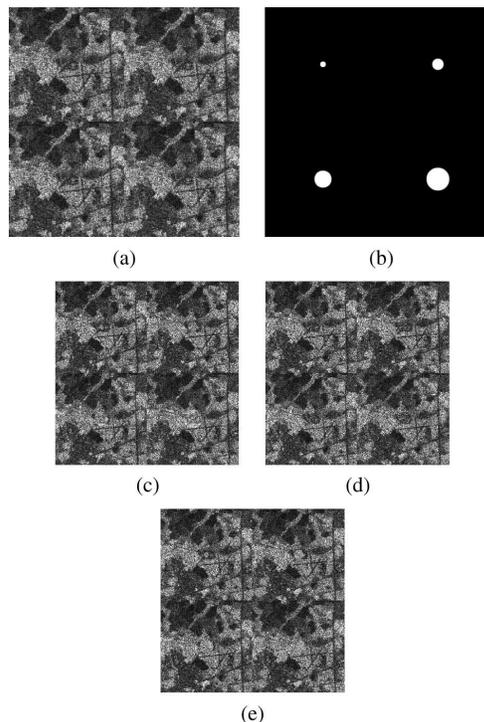


Fig. 4. Simulated data set. (a) Before. (b) Mask. (c) After offset. (d) After Gaussian. (e) After deterministic.

amplitude (coming from a SPOT near-infrared-band image, normalized to $[0, 1]$) and a thousand phases coming from independent uniform generations in $[0, 2\pi]$ to characterize elementary wave scatterers. Taking the square of the modulus of each pixel yields a one-look intensity image. A four-look intensity image is obtained by averaging and subsampling two adjacent pixels along lines and rows.

Each simulation of an change is applied to the initial image by using a change circle of a given size taken from $\{5, 10, 15, \text{ and } 20\}$. Once the speckle simulation is performed (independently from one image to another), the speckle-changed images are mosaicked on a 2×2 grid as shown on Fig. 4(b).

B. Simulation of Changes

Three kinds of change were considered.

1) *Offset Change*: The initial image is modified by applying an offset value (i.e., a shift) to the initial data [Fig. 4(c)]. This is a very simple type of change, which seldom occurs in reality, but is useful to characterize the behavior of the detectors.

2) *Gaussian Change*: The initial image is modified by applying a zero-mean Gaussian additive noise to the initial data [Fig. 4(d)]. This corresponds to a change in the state of the surface—field and vegetation. This is the main type of change that one can encounter in medium-resolution SAR images.

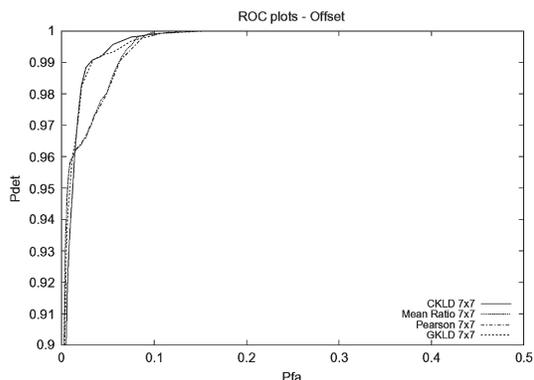


Fig. 5. ROC plot comparison of the four detectors for a simulated change consisting in an offset on reflectivity.

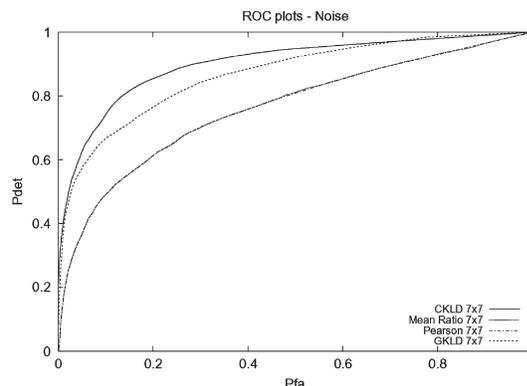


Fig. 6. ROC plot comparison of the four detectors for a simulated change consisting in a Gaussian random modification of the reflectivity.

3) *Deterministic Change*: The initial image is modified by pasting values copied from another area of the image itself [Fig. 4(e)]. This type of change can occur when there is a land-use change, anthropic activities, etc.

C. Results

1) *Monoscale Detection*: The results of the different detectors for a fixed analysis window size are analyzed.

Fig. 5 shows the ROC plots for the case where the change consists in a shift of the reflectivity value [Fig. 4(c)]. In this case, all four detectors are able to detect the changes with high accuracy. There is a slight difference in performance between the pairs CKLD–GKLD and PKLD–MRD, but it is difficult to infer the general behavior from this result. To draw a preliminary conclusion, for a simple change such as a reflectivity shift, the mean-value criterion is efficient enough for good discrimination in the changes, even on speckled images.

Fig. 6 shows the ROC plots in the case of a Gaussian change. The change is simulated by the addition of a Gaussian noise to the reflectivity (before speckle simulation). In this case, the mean value of the observed pixels remains approximately the same. It is difficult for this kind of change to be observed by a human operator. However, it is more likely to occur when the modifications affect the surface without changing its nature. In this case, even if all the detectors show bad performance in comparison to the offset case, the MRD and the PKLD are far below the GKLD and CKLD. The bad performance of the MRD is easy to understand, since the zero-mean Gaussian noise added to the reflectivity slightly changes the observed mean value. For the PKLD, it can be argued that the type of law in the Pearson system is not very different from the initial case, and the main difference is seen through the mean value, thus obtaining the same performance as the MRD. On the contrary, the GKLD assumes a simpler model than the PKLD and is able to take into account the mean and the variance modifications together. Finally, the ability of the CKLD to fit many different types of densities allows better detection for this difficult type of change.

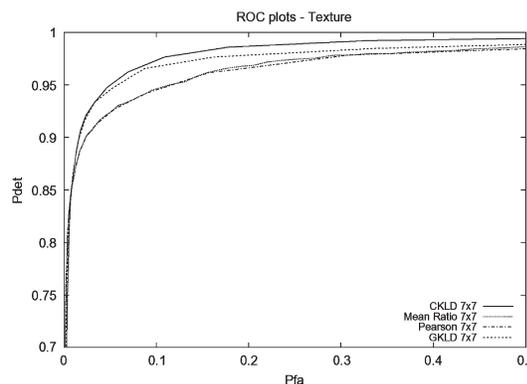


Fig. 7. ROC plot comparison of the four detectors for a simulated change consisting in a deterministic modification of the reflectivity.

The third type of change is that of a texture change, which can occur when there is a land-use change, anthropic activities, etc. In this case, as can be inferred from Fig. 7, the mean value of the regions may or may not change, and it is, therefore, interesting to analyze the shape of the density. The Pearson detector can be even worse than the MRD when the model does not fit the data, which is the case in presence of mixtures.

2) *Analysis of the MCPs*: Some collected MCPs, obtained by applying r_{CKLD} of (12) to our data set, are analyzed. Four different profiles are presented. They are extracted from a change area of the simulated data set for the case of a deterministic texture change and a radius of ten pixels. These profiles are labeled as follows: Far for the case where the analysis window is located 30 pixels from the center of the change area; Outside border for a distance of 15 pixels; Inside border for a distance of seven pixels; and Inside centered for a distance of zero pixels. Fig. 8(a) presents a diagram explaining how the profiles are extracted with respect to the change area, and Fig. 8(b) presents the profiles themselves.

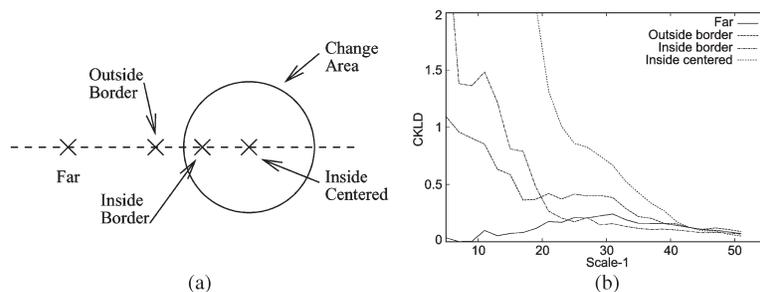


Fig. 8. Typical examples of MCPs obtained from the Edgeworth approximation of the KL distance. (a) Positions of the profiles. (b) Gaussian fitting.

The Far profile shows low values for small window sizes, and these values increase as the window size increases and it begins to include pixels from the change area. The values decrease for large window sizes, since the window stops including new change pixels while including no-change pixels present in all directions. The Outside border profile has a similar behavior, but the CKLD values are high for small scales since the pixel is nearer to the change area. The Inside border profile shows higher values for the change indicator for small window sizes. Finally, the Inside centered profile shows very high values of the detector for a large interval of window sizes. It is worth noting that the CKLD values are nearly the same for all detectors for the largest window sizes, since at this scale, all detectors include the same proportion of change and no-change pixels.

3) *MCP Exploitation*: In this section, the interest of the use of the MCP is illustrated with respect to the selection of a fixed scale of analysis (i.e., a fixed window size). The MCP allows the best scale to be selected for each pixel location in the images. Here, the maximum of the profile is used as a means to select the appropriate scale.

The maximum of the MCP and two different scales, 5×5 and 17×17 , are compared. The small window size is used in order to detect small changes, but its main drawback is that the false alarms may increase in the presence of noise. The larger window size gives a lower false-alarm rate, since the noise is averaged and, therefore, its effect is reduced. But small changes can also be averaged, and therefore, the detection probability may be lowered. Also, false alarms may be increased in the neighborhood of the change areas.

The results of the comparison are presented in Figs. 9–11. As expected, small windows were able to give high detection rates. In the case of radiometric shift, the false-alarm rates are low for a given detection probability, since the type of change is easily detected by computing the mean value over a few pixels only. However, when more complex changes occur (Figs. 10 and 11), the false-alarm rate is very high at a given detection probability. Another interesting effect can be observed in Figs. 9 and 11, where for the large window sizes, the false-alarm rate increases without an increase in the detection probability. This is due to the fact that when the window is too large for the small changes and not as large as the larger changes [see the mask in Fig. 4(b)], the new detections induce false alarms only in the neighborhood of the small changes.

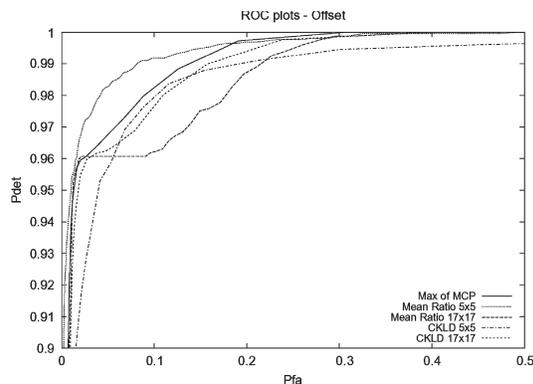


Fig. 9. ROC plot comparison between MRD—2 scales, CKLD—2 scales, and MCP—maximum of the profile, for a simulated change consisting in a reflectivity offset.

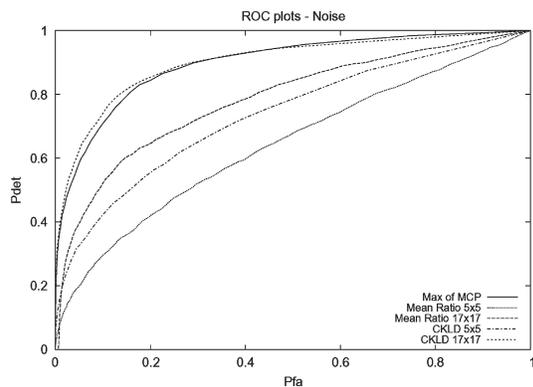


Fig. 10. ROC plot comparison between MRD—2 scales, CKLD—2 scales, and MCP—maximum of the profile, for a simulated change consisting in a Gaussian random modification of the reflectivity.

In addition, the MCP gives results which do not suffer from these drawbacks without the constraint of choosing a window size without prior information on the size of the changes in the images.

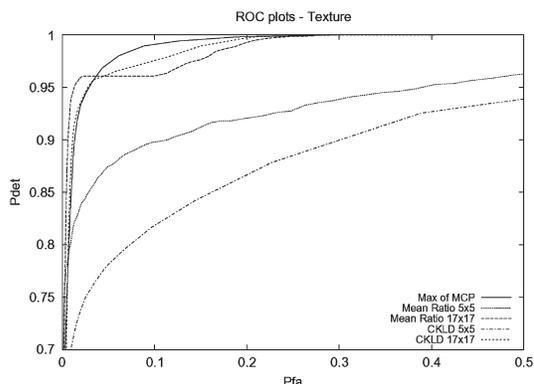


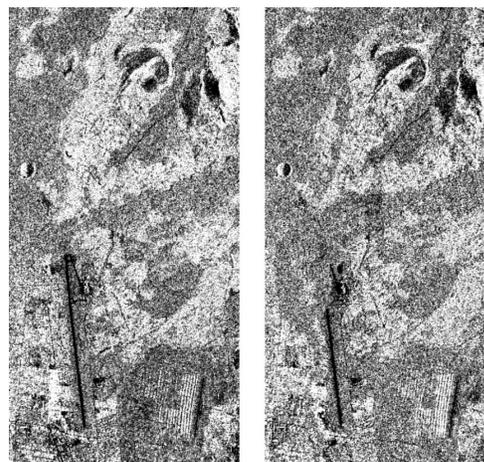
Fig. 11. ROC plot comparison between MRD—2 scales, CKLD—2 scales, and MCP—maximum of the profile, for a simulated change consisting in a deterministic modification of the reflectivity.

VI. EXPERIMENTS WITH REAL DATA

This section shows an example of applications of these algorithms to a real case. A pair of Radarsat images, acquired before and after the eruption of the Nyiragongo volcano (Democratic Republic of the Congo), which occurred in January 2002, were used. Fig. 12 shows the two images to be compared and a change map produced using the ground measures. The images have a ground resolution of 10 m and cover an area of 4×8 km. The images were orthorectified by IGN-F, the French National Geographic Institute, to a UTM35S projection, which was the same as the one used for the reference map. No filtering or calibration was applied to the data. The 16-bit to 8-bit conversion was performed using a 3σ thresholding followed by a linear intensity rescaling. It is worth noting that the image resampling applied in the orthoregistration step modifies the local statistics of the image. Indeed, the image resampling implies the local-image interpolation, which is based on approximate interpolators. A bicubic interpolation was used in this case. This type of filter has a smoothing effect, which depends on the local shift [31]. Because of these radiometric artifacts introduced during the geometric preprocessing, the theoretical models for SAR statistics may not hold locally. The area at the bottom right-hand corner of the ground-truth mask corresponds to an area where a severe misregistration exists due to the lack of a proper digital terrain model. Finally, one has to take into account the fact that the ground truth is not perfect and that all results presented in this section should be analyzed rather in a relative manner—one detector with respect to another—rather than in an absolute one—absolute value of detection probabilities.

A. Change Indicator

The comparisons between the result coming from the classical image intensity ratio and the method proposed in this paper are shown in Fig. 13. Fig. 14 gives the ROC plots using the ground truth of Fig. 12(c). It shows that the use of KL approximation by the Edgeworth series outperforms any other



(a) (b)



(c)

Fig. 12. Data and ground truth for the Nyiragongo volcanic eruption in January 2002. (a) Before. (b) After. (c) Mask.

methods such as model-based (Gaussian-based or Pearson-based) KL distance, or the ratio measure. As stated in the Introduction, a misdetection behavior of this detector can be observed, because the detector uses the mean pixel values only. It is interesting to underline the fact that the ratio criterion is not always worse than the pdf-based criteria. In fact, a density model has to fit the data in order to yield pertinent results.

For a detection probability below 0.3, it is more interesting to use the ratio criterion instead of a model-based one (by using Gaussian or Pearson assumption) in this example, even if a better change detection could have been expected by using Gaussian or Gamma laws coming from the local analysis of the two Radarsat images.

This point confirms that it is more interesting, for operational use, to consider a more flexible pdf approximation by using

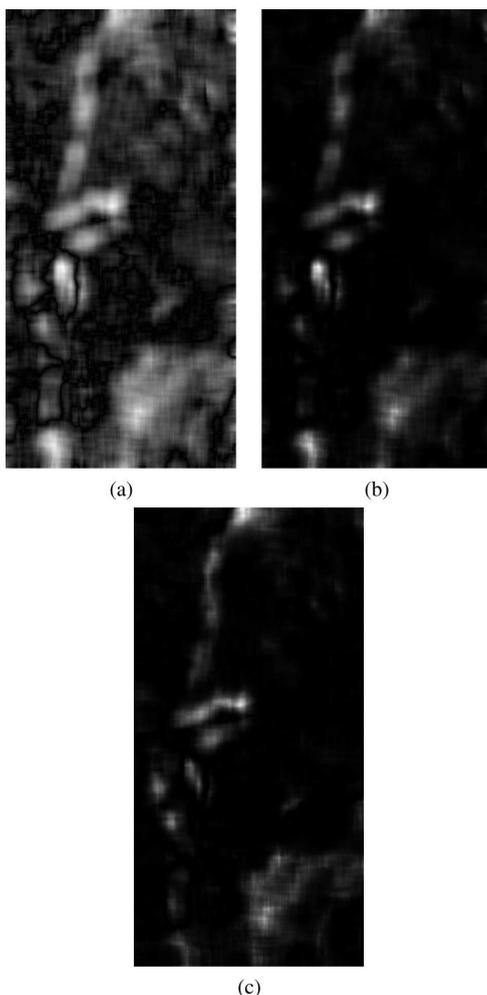


Fig. 13. Change detection. Comparison between the different change indicators using the same window size (35×35 pixels). (a) Intensity ratio. (b) Pearson KL. (c) Cumulant-based KL.

the Edgeworth series instead of a pdf parameterization. The cumulant-based approximation may give equivalent results to the Pearson-based approximation if the estimated cumulants correspond to a pdf belonging to the Pearson system of distributions, even though it may be less appropriate in the case of heavy-tailed distributions (single-look data). If cumulants of orders three and four vanish, the Edgeworth series is equivalent to a Gaussian model. If the variance of X and Y are equivalent, the Edgeworth series yields the same behavior as the ratio measure. However, when the local observations X and Y to be compared do not fit an *a priori* model, the Edgeworth series becomes a more suitable tool.

Fig. 15 draws the minimum distance of ROC curves to the point ($P_d = 1, P_{fa} = 0$). It is an interesting point of view to

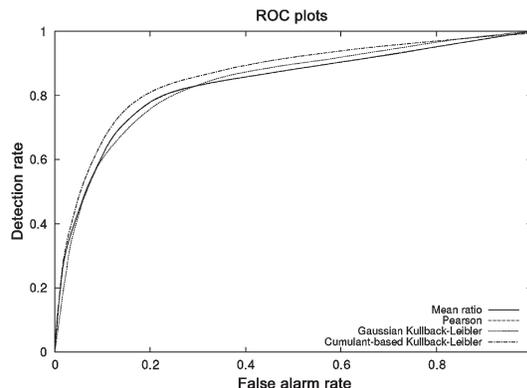


Fig. 14. ROC plots for the different detectors. The CKLD outperforms all other detectors. The Pearson-based detector gives results identical to the classical mean ratio. The Gaussian-based detector shows the worse behavior.

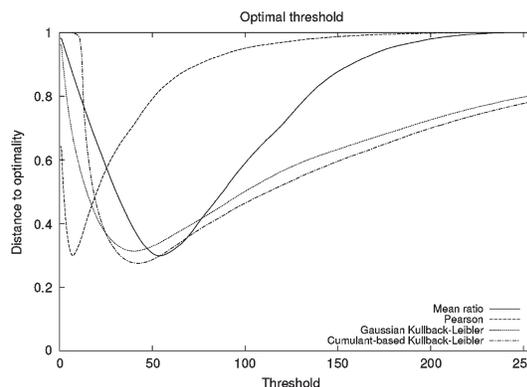


Fig. 15. Distance of ROC curves to the point ($P_d = 1, P_{fa} = 0$) for the different detectors. The Pearson detector allows trivial thresholding but is very sensitive. The cumulant-based threshold is less sensitive to threshold variations.

evaluate the threshold to be applied to obtain the best tradeoff between detection and false alarms. The best value of the threshold is to be found at the minimum of the curves.

Fig. 15 shows that this minimum is lower—and therefore, more interesting—for the Edgeworth series than for the Pearson measure or the ratio detector.

When no ground truth is available, the end-user has no *a priori* knowledge to set the value of the threshold. In this case, the Pearson measure seems to be better since a trivial value of zero could be used (i.e., pixels with values greater to zero may be considered as a change). Unfortunately, simulations and comparisons with other sets of images have shown that this trivial threshold is very sensitive to noise and fluctuations. The same observations about the sensitivity hold for the ratio measure. On the contrary, the cumulant-based measure takes its minimum for a wider range of values. Therefore, a threshold chosen *a priori* from the interval $[40, 50]$ gives an almost optimal change map for all cases.

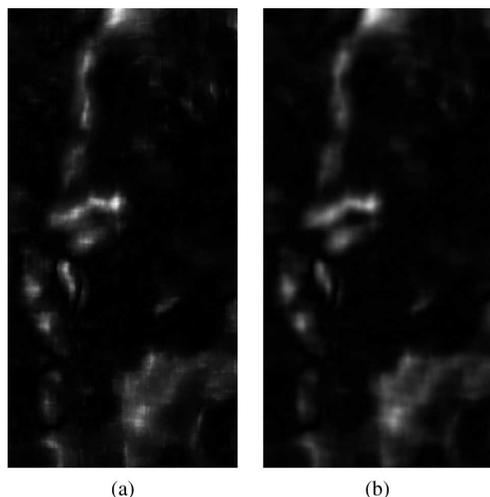


Fig. 16. Change detection results obtained with the MCP. (a) Maximum of the MCP. (b) First PC of the MCP.

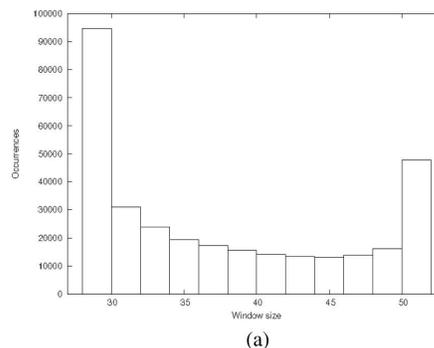
B. Multiscale Change Indicators

As stated in Section IV-B, our goal here is not to find the optimal way of exploiting the MCP but to show only the interest of the concept with simple examples. The results presented here use an MCP with window sizes ranging from 29×29 to 51×51 .

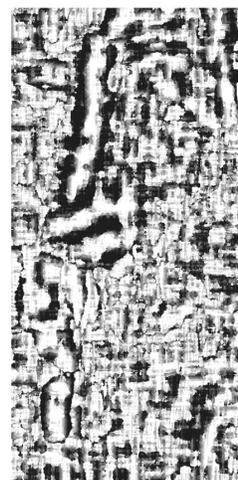
In order to select the appropriate analysis window for each pixel in the image, we will choose the maximum of the MCP. The resulting change image is shown in Fig. 16(a). Fig. 17(a) presents the histogram of the sizes of the selected analysis windows when using the maximum of the MCP. It is interesting to observe that there is a high variability of window sizes, meaning that no trivial choice exists, like for instance choosing the largest window in order to increase the number of samples. Nevertheless, two peaks may be observed in the histogram. The first maximum gives the limit of the resolution of the detector and corresponds to areas near the borders of the change and no-change classes. The second one corresponds to homogeneous areas where the window size could continue increasing. Fig. 17(b) shows the map of the selected scales. The histogram bounds of Fig. 17(a) are linearly mapped to the minimum and maximum values of the image. It is interesting to note that large windows are used inside the change and no-change areas and that small window sizes are selected near the boundaries of these areas.

The ROC plots of Fig. 18 show that this simple strategy improves the results with respect to the case where the 35×35 window was used.

As an approach to multiscale fusion, we propose here to use the first principal component of the stack of multiscale detection images. The obtained change image is presented in Fig. 16(b). The ROC plot of Fig. 18 shows that this approach also provides better performance than the monoscale detector.



(a)



(b)

Fig. 17. Analysis of the selected scales using the maximum of the MCP. (a) Histogram of the window sizes. (b) Map of selected scales.

VII. DISCUSSION AND CONCLUSION

In this paper, a new similarity measure between images has been introduced in the context of multitemporal SAR image change detection. This measure is based on the use of the cumulant-based series expansion of the local-image statistics combined with the KL divergence. The concept of MCP has been developed, and a fast and efficient implementation has been proposed. Finally, two simple approaches for the production of the change images containing multiscale information have been presented. The first one is based on the selection of the scale which gives the highest change indicator, and the second one uses the first principal component of the multiscale change image stack. The proposed similarity measure has been compared to the classical ratio of local means and also to other KLDs, which use parametric models (Gaussian- or Pearson-based). The experiments have been carried out on simulated and real data for which a reference change map was available.

The proposed original cumulant-based detector has been shown to have a more robust behavior than other detectors in terms of ROCs. The two simple yet useful schemes for the

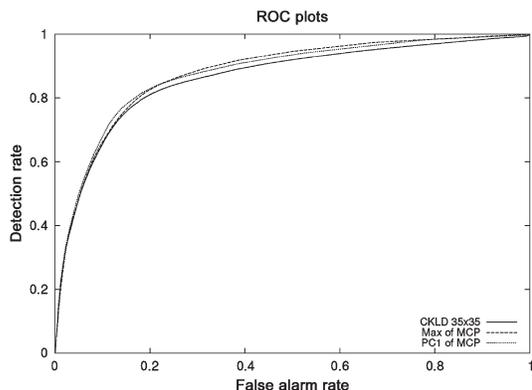


Fig. 18. ROC plots for two possibilities of MCP exploitation: The maximum and the first principal component. They outperform the CKLD for a fixed window size of 35×35 .

exploitation of the MCP provide better performance than the monoscale detector.

The main advantages of the proposed approach are the following: Our detector needs only the computation of the first four statistical moments and can deal with a great variety of pdfs and the MCP provides change information over a wide range of scales at very low computation cost.

Some improvements could be done in order to use this approach with single-look images, where the heavy-tailed distributions may need other statistical models. The use of Gamma distributions instead of Gaussian for the series expansion seems to be a good starting point.

Some questions still remain open about the use of MCPs. Indeed, it would be interesting to analyze if we could establish a classification of the profiles and thereby derive useful information, not only about the scale of the change but also about its type. This task could be carried out by visual inspection, but automatic clustering techniques, like for instance the self-organizing map [32], could be used. The parametric modeling of the profiles by projection on an orthogonal basis could be envisaged.

Another issue remaining is the automatic thresholding of the change images. Whether it is for the case of a single scale or for the case of a multiscale analysis, the statistics of the change indicators could be used in order to propose adaptive Bayesian thresholding techniques, as done in [13].

Finally, direct classification of multiscale profiles by using support vector machines seems an appropriate choice for the production of binary-change maps in the case of supervised analyses. This approach has successfully been applied to the classification of hyperspectral images [33].

All these aspects will be studied in future work.

REFERENCES

- [1] L. Bruzzone, M. Marconcini, U. Wegmuller, and A. Wiesmann, "An advanced system for the automatic classification of multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1321–1334, Jun. 2004.
- [2] P. Lombardo and T. Pellizzeri, "Maximum likelihood signal processing techniques to detect a step pattern of change in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 4, pp. 853–870, Apr. 2002.
- [3] E. J. M. Rignot and J. J. van Zyl, "Change detection techniques for ERS-1 SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 4, pp. 896–906, Jul. 1993.
- [4] F. T. Ulaby, F. Kouyate, B. Brisco, and T. H. L. Williams, "Textural information in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. GRS-24, no. 2, pp. 235–245, Mar. 1986.
- [5] F. Bujor, E. Trouvé, E. Valet, J.-M. Nicolas, and J. Rudant, "Application of log-cumulants to the detection of spatiotemporal discontinuities in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2073–2084, Oct. 2004.
- [6] N. Lam and D. Quattrochi, "On the issues of scale, resolution, and fractal analysis in the mapping sciences," *Prof. Geogr.*, vol. 44, no. 1, pp. 88–98, 1992.
- [7] D. Marceau and G. Hay, "Remote sensing contributions to the scale issue," *Can. J. Remote Sens.*, vol. 25, no. 4, pp. 357–366, 1999.
- [8] L. M. T. Carvalho, L. M. G. Fonseca, F. Murtagh, and J. Clevers, "Digital change detection with the aid of multiresolution wavelet analysis," *Int. J. Remote Sens.*, vol. 22, no. 18, pp. 3871–3876, Dec. 2001.
- [9] G. Mercier, S. Derrode, W. Pieczynski, and J. L. Caillec, "Multiscale oil slick segmentation with Markov chain model," in *Proc. IGARSS*, Toulouse, France, Jul. 21–25, 2003, pp. 3501–3503.
- [10] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.
- [11] —, "An adaptive parcel-based technique for unsupervised change detection," *Int. J. Remote Sens.*, vol. 21, no. 4, pp. 817–822, 2000.
- [12] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 858–867, Jul. 1997.
- [13] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, Apr. 2005.
- [14] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86, Mar. 1951.
- [15] J.-S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 1, pp. 165–168, Mar. 1980.
- [16] V. Frost, J. Stiles, K. Shanmugan, and J. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 2, pp. 157–165, Mar. 1982.
- [17] D. Kuan, A. A. Sawchuk, T. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, no. 2, pp. 165–177, Mar. 1985.
- [18] K. Ward, "Compound representation of high resolution sea clutter," *Electron. Lett.*, vol. 17, no. 16, pp. 561–565, 1981.
- [19] Y. Delignon, R. Garelo, and A. Hillion, "Statistical modelling of ocean SAR images," *Proc. Inst. Electr. Eng.—Radar Sonar Navig.*, vol. 144, no. 6, pp. 348–354, Dec. 1997.
- [20] N. Johnson and S. Kotz, *Distributions in Statistics: Continuous Univariate Distributions*. New York: Wiley Interscience, 1969.
- [21] J. Inglada, "Change detection on SAR images by using a parametric estimation of the Kullback-Leibler divergence," in *Proc. IGARSS*, Toulouse, France, Jul. 21–25, 2003, pp. 4104–4106.
- [22] J.-M. Nicolas, "Introduction aux statistiques de deuxième espèce: Applications des log-moments et des log-cumulants à l'analyse des lois d'images radar," *Trait. Signal*, vol. 19, no. 3, pp. 139–167, 2002.
- [23] J.-M. Nicolas, C. Tison, and F. Tupin, "Comparison of Beta1-Beta2 diagram and Logcumulant2-Logcumulant3 diagram for SAR image texture classification," in *Proc. EUSAR*, May 2004, pp. 133–136.
- [24] C. Tison, J.-M. Nicolas, and F. Tupin, "Accuracy of Fisher distributions and log-cumulant estimation in a Markovian segmentation of high resolution SAR images over dense urban areas," in *Proc. IGARSS*, Jul. 2003, vol. 3, pp. 1999–2001.
- [25] C. Tison, J. Nicolas, F. Tupin, and H. Maître, "A new statistical model of urban areas in high resolution SAR images for Markovian segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2046–2057, Oct. 2004.
- [26] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis: A Non-linear Signal Processing Framework*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [27] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics*, 5th ed. London, U.K.: Arnold, 1991.

- [28] P. McCullagh, *Tensor Methods in Statistics*. London, U.K.: Chapman & Hall, 1987.
- [29] J. Lin, N. Saito, and R. Levine. (1999). "Edgeworth approximation of the Kullback-Leibler distance towards problems in image analysis," Univ. California, Davis. Tech. Rep. [Online]. Available: <http://www.math.ucdavis.edu/~saito>
- [30] F. Bovolo and L. Bruzzone, "A wavelet-based change-detection technique for multitemporal SAR images," in *Proc. Int. Workshop Anal. Multi-Temporal Remote Sens. Images*, May 16–18, 2005, pp. 85–89.
- [31] J. Inglada, V. Muron, D. Pichard, and T. Feuvrier, "Analysis of artifacts in subpixel remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 254–264, Jan. 2007.
- [32] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.
- [33] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote-sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.



Jordi Inglada received the degree in telecommunications engineering from the Universitat Politècnica de Catalunya, Barcelona, Spain, and the École Nationale Supérieure des Télécommunications de Bretagne, Brest, France, in 1997 and the Ph.D. degree in signal processing and telecommunications from the Université de Rennes 1, Rennes, France, in 2000.

He is currently with the Centre National d'Études Spatiales, Toulouse, France, working in the field of remote-sensing image processing. He is in charge of the development of image-processing algorithms for the operational exploitation of Earth Observation images, mainly in the fields of image registration, change detection, and object recognition.



Grégoire Mercier (M'01) was born in France in the 1971. He received the degree in engineering from the Institut National des Télécommunications, Evry, France in 1993 and the Ph.D. degree from the University of Rennes, Rennes, France, in 1999.

Since 1999, he has been with the École Nationale Supérieure des Télécommunications de Bretagne, Brest, France, where he is currently an Associate Professor in the Image and Information Processing Department. His research interests are in remote-sensing image compression and segmentation, especially in hyperspectral and synthetic aperture radar. Specifically, his research is dedicated to change detection and combating pollution.

Support Vector Reduction in SVM Algorithm for Abrupt Change Detection in Remote Sensing

Tarek Habib, *Member, IEEE*, Jordi Inglada, *Associate Member, IEEE*, Grégoire Mercier, *Senior Member, IEEE*, and Jocelyn Chanussot, *Senior Member, IEEE*

Abstract—Satellite imagery classification using the support vector machine (SVM) algorithm may be a time-consuming task. This may lead to unacceptable performances for risk management applications that are very time constrained. Hence, methods for accelerating the SVM classification are mandatory. From the SVM decision function, it can be noted that the classification time is proportional to the number of support vectors (SVs) in the nonlinear case. In this letter, four different algorithms for reducing the number of SVs are proposed. The algorithms have been tested in the frame of a change detection application, which corresponds to a change-versus-no-change classification problem, based on a set of generic change criteria extracted from different combinations of remote sensing imagery.

Index Terms—Image classification, image matching, image processing, remote sensing.

I. INTRODUCTION

IN THE context of risk management and hazard assessment using satellite imagery, high-performing change detection algorithms are of great importance [1]. For this type of application, change detection problems can be viewed as a classification using two classes (change/no change). Due to its multiple advantages, the support vector machine (SVM) binary classification algorithm has been widely used for the classification of satellite imagery [2]. In this letter, the SVM algorithm is used in the context of building a generic change detection algorithm. In order to achieve the goal of genericity, a large amount of change indicators have been extracted from the images and used as input in the SVMs, yielding high-dimensionality input vectors. Despite the fact that, theoretically, the SVM algorithm can handle such large data sets with high dimensionality, in practice, the following two main problems arise: 1) The computation time required for the convergence of the optimization problem during the “learning” step of the SVM algorithm is very high, and 2) the classification time

increases in a polynomial fashion with the increase in the data dimensionality [3].

This letter focuses on the decrease of the classification time. In the SVM algorithm, the classification time is directly proportional to the number of support vectors (SVs) in the nonlinear case. Several methods for the SV reduction are proposed in Section III. The reduction of the SVs may logically result in a decrease in the classification accuracy; however, in the context of risk management and hazard assessment, a slight decrease in the classification accuracy can be tolerated for a gain in classification time.

This letter is divided into the following sections. In Section II, a brief introduction on the SVM binary classification algorithm is presented. In Section III, the proposed techniques for obtaining the reduced set of SVs are presented. The conducted experiments and the results obtained using the different methods are discussed in Section IV. Finally, the conclusions are drawn in Section V.

II. SVMs

The SVMs are state-of-the-art large margin classifiers that have gained much popularity within the image-processing community. The SVM is a learning-based binary classification algorithm that is based on the concept of structural risk minimization [4]. This construction has shown to generally outperform traditional learning machines like multilayer neural networks that are based on the concept of empirical risk minimization [5]. In this section, a brief review of the theory of this algorithm is presented; for further details, the reader is invited to review [6].

Consider a set of learning data $(x_i, y_i)_{i=1}^m$, where $x_i \in \mathbb{R}^n$ are the input feature vectors, and $y_i \in \{-1, +1\}$ are the set of corresponding labels (i.e., classes). The SVM solution finds $y_{\text{test}} = f(x_{\text{test}})$ for a new test vector x_{test} so that the probability of the error is minimal. The SVM decision for any new vector x_{test} is obtained under the hypothesis that x_{test} is issued from the same unknown probability density function that produced the learning set x_i .

If it is assumed that the two classes (i.e., the SVM is a binary classifier) can be separated by a hyperplane and that no prior information concerning the data distribution is available, then the optimal hyperplane is the one which maximizes the margin of separation between the two classes [6]. The optimal values of w and b can be obtained by solving a constrained minimization problem through the use of the Lagrange multipliers α_i . The decision function provided by the SVM can thus be put in the following form:

$$f(x_{\text{test}}) = \text{sgn} \left(\sum_{i=1}^m \alpha_i y_i K(x_{\text{test}}, x_i) + b \right). \quad (1)$$

Manuscript received September 5, 2008; revised January 19, 2009. Current version published July 4, 2009.

T. Habib is with the Centre National d'Études Spatiales, Direction du Centre de Toulouse/Système et Images/Analyse et Produits Images (DCT/SI/AP), 31401 Toulouse, France, and also with the Images and Signal Department, GIPSA-lab, 38402 Grenoble Cedex, France.

J. Inglada is with the Centre National d'Études Spatiales, Direction du Centre de Toulouse/Système et Images/Analyse et Produits Images (DCT/SI/AP), 31401 Toulouse, France.

G. Mercier is with the Laboratoire des Sciences et Technologies de l'information, Communication and Knowledge, Centre National de la Recherche Scientifique Unité Mixte de Recherche, Telecom Bretagne, Institut Telecom, 29238 Brest Cedex 3, France.

J. Chanussot is with the Images and Signal Department, GIPSA-lab, 38402 Grenoble Cedex, France.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LGRS.2009.2020306

The input feature vectors x_i having a Lagrange multiplier α_i that is not equal to zero are considered as SVs, hence the name SVMs. The kernel function $K(\cdot, \cdot)$ substitutes the scalar product $\langle \cdot, \cdot \rangle$ in order to allow the SVM to learn the nonlinear classifiers as, for example, polynomial classifiers or radial basis function (RBF) networks.

III. REDUCED SET OF SV

With the goal of providing a tradeoff between the classification accuracy and the computation time, four different approaches are proposed in this section. The objective of these approaches is the selection of a representative subset of the SVs. These approaches are intended to select the SVs according to their *importance* in order to obtain an acceptable approximation of the SVM decision boundary.

A. Set Reduction by Lagrange Multipliers

The value of the Lagrange multipliers is a valid indicator of the importance of a given vector. Using the information provided by the Lagrange multipliers, the first approach would be to keep the SVs based on the value of the Lagrange multipliers. Thus, the SVs with the highest *unbounded* Lagrange multipliers (i.e., $0 < \alpha_i < C$) will be used as the reduced set, and the rest of the SVs will be eliminated. The size of the reduced set is either specified *a priori* by a percentage of the entire set of SVs or by specifying a threshold on the value of the α_i to be retained. If the percentage is higher than the available number of unbounded SVs, the set is completed using the randomly selected bounded SVs. Since this SV reduction procedure may not respect the constraints of the SVM algorithm, namely, $\sum_{i=1}^m \alpha_i y_i = 0$, then a new learning step has to be applied. A modified classification algorithm that takes into consideration the SVM constraints is proposed as follows: 1) Perform a first learning step; 2) choose a subset of the SVs according to a maximum value of the α_i with $\alpha_i < C$; 3) reperform the learning step; and 4) classify.

The added relearning step launches the learning process only on the chosen reduced set of vectors using the same kernel as the one used in the first learning step. Intuitively, since these vectors were already considered as SVs in the first learning step, then the optimization procedure will separate them similarly to the previous learning step, and thus, the role of the second learning step will be simply to reattribute the values of the α_i in order to guarantee $\sum_{i=1}^m \alpha_i y_i = 0$. In practice, however, and due to the implementation issues of the SVM optimization, it was noted that a very limited number of vectors (basically two or three for a set of 600 SVs) are not considered as SVs anymore. It is not likely to have an impact on the overall classification performance since this quantity is negligible with respect to the total number of SVs.

B. Set Reduction by Distance to Separating Surface

The decision function as presented in (1) can be used to build a distance measure. The distance-to-hyperplane measure may be defined by

$$f_d(x) = \sum_{i=1}^m \alpha_i y_i K(x, x_i) + b.$$

The SV's relative position, with respect to the separating boundary, could be evaluated and then used as a criterion in the SV reduction. The procedures are: 1) Perform a first learning step; 2) choose a subset of the SVs according to a maximum value of $f_d(x_i)$ as in the aforementioned equation; 3) reperform the learning step; and 4) classify. As mentioned earlier, the relearned step is used in order to recompute the α_i . As implemented in this algorithm, the SVs situated far away from the separating hyperplane are considered as better suited for set reduction since they provide a better generalization of the separating hyperplane. The situation far from the hyperplane limits the eventual overlap between the two classes and hence is expected to provide a better separability between the two classes.

C. Set Reduction by Mechanical Analogy

In [7], Schlkopf and Smola present the SVM classification problem as a mechanical problem. The structure of the SVM optimization problem is closely similar to the ones that typically arise in Lagrange's formulation of mechanics. From this point of view, it is possible to give a mechanical interpretation to the optimal margin hyperplanes. Assume that each SV x_i exerts a perpendicular force of magnitude α_i and direction $y_i(w/\|w\|)$ on a solid plane sheet lying along the hyperplane. The idea is to use this mechanical analogy to reduce the number of SVs by merging the vectors using geometrical and mechanical properties.

According to this point of view, two SVs x_1 and x_2 may be replaced by a new vector x_{new} if its mean is set as the mean of the two vectors. The difficult point would be to set the value of α_{new} , which is associated to x_{new} . Considering the mechanical analogy, the problem would be to identify the direction of this force (i.e., the vector that is perpendicular to the separating boundary w cannot be computed directly in the nonlinear case). This problem could be avoided by merging the two vectors directly in the feature space (i.e., the space induced by the kernel function). Assuming that the system is stable around the origin of the feature space, the α_{new} could be obtained as follows:

$$\alpha_{new} y_{new} K(x_{new}, x_{new}) = y_1 \alpha_1 K(x_1, x_1) + y_2 \alpha_2 K(x_2, x_2).$$

Once again, in order to respect $\sum_{i=1}^m \alpha_i y_i = 0$, the difference (i.e., δ) between the computed α_{new} and the required one for stability is computed by

$$\alpha_{new} y_{new} + \delta = \alpha_1 y_1 + \alpha_2 y_2. \quad (2)$$

Once the value of δ is obtained, it is distributed uniformly on the SVs of the same class of x_{new} [according to the sign of $f(x_{new})$ in (1)], which guarantees the SVM stability constraint.

An algorithm based on this mechanical analogy was developed. In order to identify the SVs that should be merged together, a specific signature was computed for each vector, and then, the vectors with similar signatures are merged.

The proposed algorithm is decomposed in the following steps, which will be detailed further on: 1) Compute the signatures of the different SVs (s_n); 2) compute the distances between the different vectors according to their respective signature $[D(s_i, s_j)]$ by evaluating the symmetrized

Kullback–Leibler (KL) distance; and 3) define a threshold and then merge the SVs having an interdistance that is less than the defined threshold.

The signature s_n of an SV x_n is a vector of the m components (m is the number of SVs) defined, component by component, by the value of $K(x_i, x_n)$, $1 \leq i \leq m$

$$s_n = (K(x_1, x_n), K(x_2, x_n), \dots, K(x_m, x_n)). \quad (3)$$

The computation of the signature by using this procedure has two reasons: 1) The interaction between the different vectors using the kernel function is indicative of the distribution in the feature space, and 2) the kernel values are already computed during the learning phase, and hence, with proper storage, no further computation will be required. Once the signature for all the m vectors is computed, the symmetrized KL distance is used to identify the vectors to be merged. The expression of this distance was developed in [8] and is defined as follows:

$$\frac{1}{R(s_1, s_2)} = \frac{1}{D(s_1||s_2)} + \frac{1}{D(s_2||s_1)} \quad (4)$$

where $D(s_1||s_2) = \int s_1(x) \log(s_1(x)/s_2(x))dx$. The reason for using this type of distance instead of, for example, a simple quadratic distance is its interesting properties. This type of distance measurement is based on the KL divergence [9] that contains the discrimination information of the first hypothesis (i.e., represented by s_1) on the second hypothesis (i.e., represented by s_2). The addition of such a rich measurement that already contains information concerning the discrimination between the two classes can enhance the classification performance in the sense that it allows a first stage of *clustering* of the vectors.

The threshold definition procedure can be done manually or automatically according to the required number of SVs at the output of this reduction process.

This procedure has the advantage of not requiring a new learning step like the other algorithms; however, the computation of the signatures and the distances is a time-consuming task. Knowing that all these kernel evaluations have been already computed during the initial learning phase, storing these values greatly increases the speed of this algorithm. The total number of operations required for this algorithm is then $(4m + 1)$ for the KL distance computation in addition to the six operations required to evaluate the values of α_{new} and x_{new} . On the other hand, merging two vectors saves the m operations for the classification of a new test vector. Thus, if $N > (4m + 7)$, this type of reduced set SVM can be interesting, where N is the total number of test data vectors to be classified.

D. Set Reduction by Optimization Problem

Following the same procedure as that proposed for the mechanical analogy approach, another strategy may be used to evaluate α_{new} .

It takes its inspiration from the SVM optimization problem in the sense that the solution of the SVM optimization problem using the new vector is equal to the solution obtained using the original vectors x_1 and x_2 . Once the values of y_{new} and x_{new} are specified, the α_{new} of the new SV x_{new} can be computed in order to replace the two SVs α_{i_1} and α_{i_2} . α_{new} can be obtained

as follows (note that this computation can be extended to merge any number of vectors):

$$\alpha_{\text{new}} = \frac{\alpha_{i_1} + \alpha_{i_2} - \frac{1}{2}(\mathcal{T}_{i_1} + \mathcal{T}_{i_2} + \mathcal{T}_{i_1;i_2})}{1 - \frac{1}{2} \sum_{\substack{i=1 \\ i \neq i_1, i_2}}^m \alpha_i y_i y_{\text{new}} K(x_i, x_{\text{new}})}$$

where

$$\mathcal{T}_{i_1;i_2} = 2\alpha_{i_1}\alpha_{i_2}y_{i_1}y_{i_2}K(x_{i_1}, x_{i_2}), \quad \text{for } i = i_1 \text{ or } i_2$$

$$\mathcal{T}_i = \sum_{\substack{j=1 \\ j \neq i_1, i_2}}^m \alpha_j \alpha_i y_j y_i K(x_i, x_j).$$

The number of operations required to compute α_{new} from the merging of the two vectors is $(7m - 6)$; this includes the computation of the KL distance to identify the vectors to be merged. Thus, this type of reduction is interesting when $N > (7m - 6)$, where N is the number of data vectors to be classified. When three vectors are merged using this technique, the reduction is interesting when $N > (6m - 2)$; thus, the procedure becomes more interesting when the number of vectors to be merged increases.

IV. EXPERIMENTS

The proposed SV reduction algorithms were tested on three data sets. The first is the Goma data set that is composed of a pair of synthetic aperture radar images that were obtained before/after the Nyiragongo volcanic eruption in eastern Congo in 2002. The second set is composed of a pair of optical Satellite Pour l'Observation de la Terre images that were obtained before/after the earthquake that hit the Algerian city of Boumerdes.

Several features were computed, namely, difference, ratio, ratio of means, ratio of medians, correlation, mean squares, KL distance, mutual information, cardinality match, gradient difference, entropy, and energy. Along with the original images, these images were used to form the input vectors for the SVM algorithm, with each feature representing one component of the vector; this yields vectors of size 14.

For testing purposes, a soft margin SVM was used, with $C = 1$. An RBF kernel was used, with $\gamma = 0.5$ (i.e., it should be noted that the exact values of the parameters are of no interest in our case since we are only interested in obtaining information concerning the overall behavior of algorithms that are independent of these parameters). Ten different full iterations were performed on the test data. A single iteration is composed of the following steps: 1) Choose a random set of learning vectors (consisting 1% of the total number of available testing vectors); 2) perform the learning phase of the SVM algorithm in order to obtain the original SV set; 3) reduce the number of SVs by following the different algorithms (for the algorithm in Section III-A, this means 10% reduction with respect to the original set's size, whereas for the remaining three algorithms, this means increase of the threshold by 10% of the maximum distance according to the specified criteria); 4) perform the SVM classification using the new sets (four new sets following the four proposed algorithms) of SVs; and 5) restart from step 3 until the new SV set's size is only 10% of the original SV set.

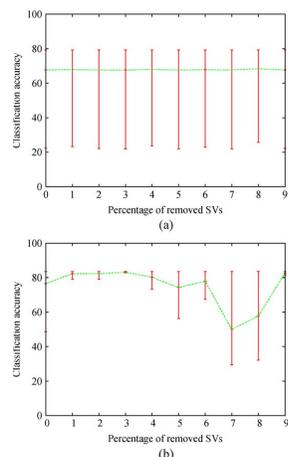


Fig. 1. Classification accuracy versus percentage (divided by ten) of the set's size reduction with respect to the original set's size using the reduced set algorithm of Section III-A applied on the different data sets. (a) Goma data set. (b) Boumerdes data set.

Note that, in the following figures, the green line is a simple connection between the mean values. This green line is intended to show the overall behavior of the different functions.

A. Set Reduction by Lagrange Multipliers

Fig. 1 shows the classification error using the reduced set as a function of the total number of SVs applied on the different testing data sets. Hence, the x -axis represents the percentage of SVs that was removed ($\times 10$). In the case shown in Fig. 1(a), the classification accuracy is almost constant with the removal of the SVs. However, due to the large variance of the classification error, following this method may not be robust to reduce the set from the Goma data.

In the case shown in Fig. 1(b), the evolution is convenient in the sense that the classification accuracy decreases with the removal of the SVs. Thus, it offers a tradeoff between the classification accuracy and the classification time (through the reduction of the number of SVs). However, when removing a large number of SVs (i.e., higher than 70%), the variance of the classification accuracy is very high, showing that the results are more and more noisy.

B. Set Reduction by Distance to Separating Surface

Fig. 2 shows the results obtained for the algorithm using the distance-to-hyperplane criteria.

From Fig. 2(a), the results show that the classification accuracy has a large variance with respect to the removal of the SVs. On the other hand, the mean values show that this technique can be used to provide a tradeoff between the classification accuracy and the classification time since the overall accuracy decreases with the decrease of the number of SVs.

When tested on the Boumerdes data set as shown in Fig. 2(b), this approach provides irregular results, where the classification accuracy seems to be independent of the number of SVs.

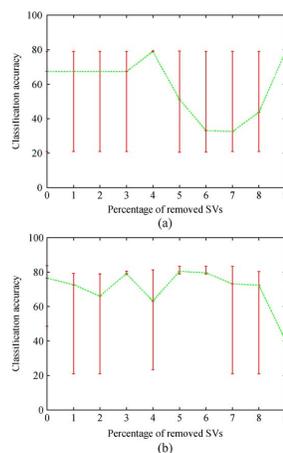


Fig. 2. Classification accuracy versus percentage (divided by ten) of the maximum distance using the reduced set algorithm defined in Section III-B. (a) Goma data set. (b) Boumerdes data set.

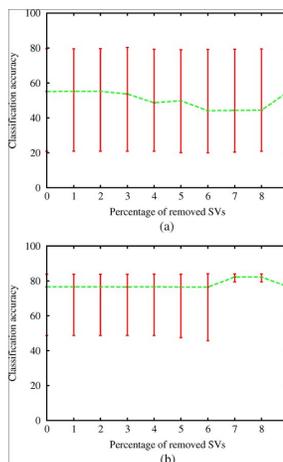


Fig. 3. Classification accuracy versus merging step using the algorithm of Section III-C. (a) Goma data set. (b) Boumerdes data set.

C. Set Reduction by Mechanical Analogy

Fig. 3 shows the results obtained using the algorithm defined in Section III-C. It can be noticed that the error's mean and variance increase with the number of merged vectors. As the number of SVs decreases, the separating surface becomes an approximation of the original surface and thus provides a lower classification accuracy. This shows that the reduced set using this technique can provide a tradeoff between the classification accuracy and the classification time through the reduction of the SVs.

Despite a large variance of the classification accuracy, the reduced set using the mechanical analogy provides adequate results when applied to the Goma data set. As can be seen in Fig. 3(a), the classification accuracy slightly decreases with the decrease of the number of SVs. This is maintained until

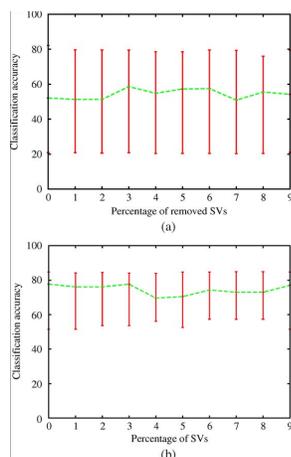


Fig. 4. Classification accuracy versus the merging step using the optimization problem algorithm. (a) Goma data set. (b) Boumerdes data set.

a very large number of SVs are removed (80%), where the classification result becomes unreliable due to the lack of representativeness of the remaining SVs.

Fig. 3(b) shows the result obtained when the algorithm is applied on the Boumerdes data set. It can be noted that the algorithm provides reasonable results in terms of the variance of the classification accuracy as well as the accuracy with respect to the number of SVs that remains constant until 70% of the SVs are removed.

D. Set Reduction by Optimization Problem

Fig. 4 shows the results obtained using the algorithm defined in Section III-D. From these results, it can be shown that the mean error increases with the decrease of the number of SVs. However, the variance value remains high independently from the number of SVs.

Fig. 4(a) shows the results of the application of the algorithm on the Goma data set. Despite a large variance of the classification accuracy, the algorithm provides good results, where the mean classification accuracy is being maintained at a relatively constant level with the removal of the SVs.

Similar to the tests on the Goma data set are the tests on the Boumerdes data set. Fig. 4(b) shows that the algorithm is capable of preserving the classification accuracy despite the removal of the SVs. Hence, the obtained approximation of the separating hyperplane is satisfactory.

After the inspection of the different obtained results, it can be observed that, in terms of classification accuracy, the mechanical-analogy-based algorithm provides the best results. The algorithm using the optimization problem is also efficient but suffers from the following two drawbacks: 1) high computational cost during the merging phase and 2) high error variance independently from the number of available SVs. The algorithms using the distance to the hyperplane and the direct filtering of the Lagrange multipliers provide irregular results, probably due to the relearn step that could change the shape of the separating hyperplane.

TABLE I
AMOUNT OF REDUCTION OPPOSED TO THE AVERAGE CLASSIFICATION TIME (IN CPU TIME) FOR THE DIFFERENT SET REDUCTION TECHNIQUES

Technique	No-Reduction	1	5	9
Lagrange Multipliers	205.5	175.6	71.3	19.3
Distance	205.5	130.6	24.7	16.3
Mechanical	205.5	157.6	137.7	87.1
Optimization	205.5	158.5	134.5	79.4

E. Classification Time

In this section, a comparison between the different methods is provided in terms of computational time. Table I shows the amount of reduction opposed to the average classification time

As can be seen from these results, despite the computational cost added for the merging schemes represented by the optimization problem or the mechanical analogy set reduction techniques, the computation time reduction as proposed in these methods remains interesting. Note, however, that the computation time required for the filtering techniques (i.e., represented by the Lagrange multiplier set reduction and the distance-to-hyperplane set reduction techniques) has a significantly less computation time, but, as shown earlier, this large decrease of the computation time comes at the expense of the classification accuracy.

V. CONCLUSION

In this letter, four different SV set reduction algorithms have been proposed. The objective of these different algorithms is to accelerate the classification process for a generic SVM-based change detection algorithm in the context of risk management in order to respect crucial operational time constraints. Since the classification time is directly proportional to the number of SVs in the nonlinear case, the adopted strategy was to obtain a reduced set of the initial SVs. It was noted that the algorithms using the mechanical analogy (Section III-C) and the optimization problem (Section III-D) show good performances. The algorithm using the mechanical analogy provides the best results since it overperforms the algorithm using the optimization problem both in required computation and in the classification accuracy.

REFERENCES

- [1] T. Habib, J. Chanussot, J. Inglada, and G. Mercier, "Abrupt change detection on multitemporal remote sensing images: A statistical overview of methodologies applied on real cases," in *Proc. IEEE IGARSS*, Barcelona, Spain, Jul. 2007, pp. 2593–2596.
- [2] N. Ghoggali and F. Melgani, "Genetic SVM approach to semisupervised multitemporal classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 212–216, Apr. 2008.
- [3] C. J. C. Burges, "Simplified support vector decision rules," in *Proc. Int. Conf. Mach. Learn.*, 1996, pp. 71–77.
- [4] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [5] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [6] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
- [7] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning)*. Cambridge, MA: MIT Press, 2001.
- [8] D. H. Johnson and S. Sinanovic, "Symmetrizing the Kullback–Leibler distance," Rice Univ., Center for Multimedia Communication, Houston, TX, 2001, Tech. Rep.
- [9] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

15 Reconnaissance d'objets



Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features

Jordi Inglada

CNES - DCT/SI/AP, 18, Avenue Edouard Belin, 31401 Toulouse Cedex 09, France

Received 6 December 2005; received in revised form 11 May 2007; accepted 15 May 2007
Available online 27 June 2007

Abstract

With the advent of Earth Observation satellite sensors producing images in the visible wavelengths with resolutions better than 5 m, it is now possible to recognize man-made objects which were not visible at lower resolutions. Because of the size and the increasing quantity of remote sensing images, tools are needed for computer aided interpretation.

In this work we present an image processing system for the detection and recognition of man-made objects in high resolution optical remote sensing images. Detection is understood here as finding a small rectangular area in the image containing an object. Recognition is the attribution of a class label. These algorithms are based on learning methods and on an example data base which contains eleven classes of objects. The examples (more than 150 for each class) have been manually extracted from SPOT 5 THR images (2.5 m resolution).

In order to build a system which is independent of the type of object to be recognized, we have used a supervised learning approach based on support vector machines. The system learns a generic model for each class of objects by using a geometric characterization of the examples in the data base.

The main novelty of this paper is the use of a high number of geometric image features which allows to characterise several classes of objects with different geometric properties using a supervised learning approach. The results show the possibility of discrimination of several classes of objects with classification rates higher than 80%.

© 2007 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

Keywords: Object recognition; Man-made objects; Support vector machines; Geometrical moments

1. Introduction

Earth observation images with resolutions better than 5 m allow the visual recognition of man-made objects of cartographic interest. Sensors with resolutions close to 1 m or better allow a better discrimination of this kind of

objects, but they usually have swath widths which are too small (less than 20 km) for mass production of maps used in town and country planning. Therefore, sensors such as SPOT 5 which combine a high resolution (2.5 m) and a wide swath (60 km) increase the interest of the use of satellite images in this application field.

Satellite data allow the possibility of making rapid mapping in the case of natural hazards, for instance. This kind of application is submitted to strong time constraints.

E-mail address: jordi.inglada@cnes.fr.

0924-2716/\$ - see front matter © 2007 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:[10.1016/j.isprsjprs.2007.05.011](https://doi.org/10.1016/j.isprsjprs.2007.05.011)

However, the exploitation images of such a large size containing a high density of information is time-consuming and also very difficult to analyze by a photo-interpreter. Indeed, more than 400 analysis screens are needed for one scene. We aim at providing photo-interpreters with tools which help them in their visual analysis task. From the user point of view, it is interesting to have a tool which is able to point out where in the whole image a particular kind of object is present, eventually giving some kind of confidence ratio.

The main goal of this paper is to investigate the possibility of recognising objects in high resolution remote sensing images using a generic approach, that is, an approach which is not object dependent.

One of the main difficulties of image processing tasks when moving from images with resolutions coarser than 10 m to metric resolution ones, is being able to deal with the high complexity of the image content. This high complexity is mainly due to the fact that, from a user point of view, the elements of interest are not only individual pixels or surfaces, but complex, structured objects.

While the land cover analysis of medium and low resolution images can be addressed with a pixel-wise approach, the specificity of high resolution images is that we are interested in objects that can only be defined by their shape and their neighborhood. Therefore, it is difficult to implement generic image processing systems which are based on the classical two-step – feature extraction plus classification – approach.

On the other hand, manipulation of high level concepts with rule-based systems seems to loose the genericity of the former approach. Indeed, for the object recognition problem, one could take the approach of manually defining a geometric template for each object and try to match it in the image. Unfortunately, the high variability inside the different classes of objects makes this approach difficult to exploit.

The main problem in our domain is the one of recognising classes of objects instead of specific objects. While the detection and recognition of objects using specific approaches – tuned to the class of object to be extracted – has been dealt with for a long time in computer vision (especially in the military), little work has been published about the detection and recognition of classes of objects in remote sensing images with generic approaches — not specific to a particular class of object.

For instance, the literature about automatic road extraction proposes methods where the roads are modeled as a network of intersections and links between the intersections Baumgartner et al. (1999), or extracted

using explicitly formulated scale-dependent models Hinz and Baumgartner (2003). The review work presented in Mena (2003) lists a high number of methods, but they are all specific to roads or networks. The object extraction from remote sensing images literature also presents a large variety of methods for building extraction, but, again, they are specific to this class of objects. Many methods use stereo image pairs Fraser et al. (2002); Gülch et al. (1998). Other methods need important *a priori* knowledge about the surrounding of a building Gerke et al. (2001).

Related works dealing with the problem of information retrieval from remote sensing image archives exist, Datcu et al. (1998, 2003), Schroeder et al. (1998, 2000), Dell'Acqua and Gamba (2001), Daschiel and Datcu (2005); and should inspire our approach. Indeed, these works show that the use of relevant features and a classification in a feature space allow the retrieval of relevant information for a high-level semantics defined by a user by means of a small set of examples.

However, the problem of information retrieval from large data bases has different constraints than the problem of object recognition in a single image. Indeed, the so-called image mining systems work with hundreds to thousands of images which need to be indexed in a very smart way in order to permit a short response time to the user which interacts in real time with the system. Since each user's query may be different, a great effort has to be made in order to build a system which is able to link the user semantics with the low-level image features. On the other hand, all the possible images are known to the system – thus all the events of the probability space –, so the use of Bayesian learning is the most suited and elegant way to proceed. As far as we know, these systems do not try to retrieve composite objects, but only classes which can be defined by radiometric properties in a pixel-wise approach.

In the object recognition problem that we mean to solve, only a few examples are available to train the system because of the high cost of example data base construction. Furthermore, each image to process is a new image to the system, so the solution of *finding the scene in the data base which best explains the observed data*, Datcu et al. (1998), is not suited to our problem.

The goal of this work is to investigate the possibility of automatically detecting and recognising objects in high resolution remote sensing images. Detection is understood as finding a small rectangular area in the image containing the object. Recognition is the attribution of a class label to the detected object. The object recognition task can be performed after a detection step. Another possible approach is to apply

the recognition algorithms to every possible position of an analysing window within the image. While this second approach is less prone to misdetections, it is computationally expensive and should be avoided in operational systems. Our solution is based in a quick pre-processing step which aims to detect the areas where there may potentially be objects of interest. This is achieved by a pre-segmentation of a low resolution version of the image. This segmentation is based on visual perception models, Thorpe et al. (2001). This is a critical step of any recognition system but the details of the pre-processing task are out of the scope of this paper. Assuming that this pre-processing step is available, we can proceed to a sequential scan of the interest areas and directly apply the recognition approach presented in this paper. A robust recognition algorithm should be invariant to scale, rotation and illumination. This will limit the use of object appearance models, Pope and Lowe (1996), Hamdan et al. (2001). Another popular approach is template matching approach, Veltkamp and Hagedoorn (1999), but the difficulty here is finding the correct templates which are able to deal with intra-class variability. In order to cope with this difficulty, we can adopt a supervised learning approach, that is, we will make the system learn from an example data base. From this example data base, a model for each class of objects can be automatically obtained. The system will therefore learn to classify image patches containing the object in the center. This constraint imposed to the learning step will allow a precise localisation of the detected objects when performing the image scanning in the detection and recognition phase. That means that for objects larger than the analysing window, multiple detections will be observed, each one for each window position.

The most popular approach for building object models from an example data base is principal component analysis, PCA, which has led to the eigenfaces approach, Turk and Pentland (1991), Turk (2001): one assumes that the set of pixels of each example in the data base is a random vector and the principal component analysis is performed in order to obtain a set of eigenvectors (the eigenfaces) which are used as a decomposition basis. Unfortunately, this approach is very constraining on the characteristics of the data base in terms of pose and illumination. In the case of data base constitution from remote sensing images, these constraints are too high. Variants of this approach have been presented as for instance Borgne et al. (2004), where Independent Component Analysis is used instead of PCA. Our experiments with these approaches have been unsuccessful. We think that the problem is that a small number of examples with a high intra-class variability does not allow

to obtain relevant principal or independent components. Indeed, this kind of analysis has given good results for problems where the intra-class variability is low and where the example data-base is built under controlled conditions (illumination, etc.).

The remaining of this paper is organised as follows. Section 2 describes the approach chosen for our system. The geometrical characterization of objects is presented in Section 3 and the feature vector classification approach is described in Section 4. Section 5 presents the performances of the system and discusses several possibilities for its optimization. Finally, Section 6 presents our conclusions.

2. The chosen approach

In order to implement a system which is generic with respect to the types of objects to be recognised, we adopt a generic feature extraction plus classification approach and choose to use a high number of geometrical descriptors in order to cope with the great diversity of possibilities for the objects of interest. We are thus facing a problem of supervised classification in a high dimensional feature space and with only several tenths of examples available per class. This limitation is a practical one in real applications due to data availability problems and cost of data base production. In order to overcome the curse of dimensionality and, at the same time, preventing to loose information with a dimensionality reduction approach, we choose to use the support vector machines (SVM).

The goal of our recognition system is to correctly label the image patches received as input. In order to do so each image patch will be characterized by a description vector which will be used in a supervised classification scheme. An operational recognition system can then be built by sequentially scanning full SPOT5 scenes.

In order to have a system which is robust to illumination, spectral band and vegetation changes, we will only use the geometry of the images. Indeed, the kind of objects we are trying to detect are very well defined by the spatial attributes of their edges and the use of textures can introduce a source of variability – illumination and seasonal effects for instance – that we should avoid. The training phase will use a subset of the example data base. Each image patch will be processed in order to obtain a description vector (Section 3) which will be fed to the learning engine (see Section 4). The description vector will only contain geometry information, but this fact is not taken into account by the learning scheme.

In order to evaluate the performances of the system, the image patches which were not used in the training

phase will be used. Each image patch will be used to compute a description vector which will be classified by the system obtained during the training phase. The performances of the system can then be evaluated by computing confusion matrices.

The example data base has been built by CNES photo-interpreters using SPOT5 THR images (panchromatic band with 2.5 m pixel sampling). We have defined ten object classes: isolated buildings (IB), paths and tracks (PT), crossroads (CR), bridges (BR), wide roads (WR), highways (HW), roundabouts (RA), narrow roads (NR), railways (RW), suburbs (SB).

While we are only interested in three of these classes (IB, RA, BR), we have included the others in order to better characterize the rejection class. We have also created a supplementary class called OT (other) in order to take into account the different possible landscapes. The road classes are also very interesting, but they can be extracted by algorithms which allow the reconstruction of the whole road network using contextual approaches, Rochery et al. (2003), Stoica et al. (2004), Baumgartner et al. (1999), Hinz and Baumgartner (2003). For a complete list of references see the review work of Mena (2003).

For each class, more than 150 examples have been entered in the data base. Each example is a 100×100 pixels image patch with the object in the center of the image. Since objects of different classes can appear in the same image patch, the object centering is important in order to lift ambiguities for the learning algorithms. This will also hopefully produce detectors which have a good localization (small footprint) so multiple detections for adjacent patches can be avoided when scanning a large image. Each image patch is thus labelled as belonging to the class of the object located in the center. Fig. 1 shows one example for each class in the data base. Additional examples have been created by applying rotations and symmetries.

The SPOT5 scenes used for the data base have been chosen over different areas in France (Toulouse, Marseille and Strasbourg) and one area in the U.S.A (San Diego, CA).

3. Geometrical characterization of objects

We aim at characterizing the objects by using geometric information. In the remote sensing field, little work exploiting the use of geometrical shape descriptors for recognition exists. Dell'Acqua and Gamba (2001) used the point diffusion technique, PDT, but this approach is only efficient in the case where the object of interest has well defined boundaries. In our case, one

object will be characterized by several sets of boundaries. The PDT is very useful for the comparison of individual closed, deterministic shapes, but it does not seem suited for our problem.

We need to be able to statistically compare sets of shapes, but also other geometric descriptions. We will use two kinds of features: region boundaries and alignments.

Alignments (that is edges and lines) are extracted using the *Gestalt* approach proposed by Desolneux et al. (2000). In this context, an event is considered meaningful if the expectation of its occurrence would be very small in a random image. One can thus consider that in a random image the direction of the gradient of a given point is uniformly distributed, and that neighbouring pixels have a very low probability of having the same gradient direction. This algorithm gives a set of straight line segments defined by the two extremity coordinates.

Region boundaries are also extracted by using the same principle but applied on the topographic map – the level-lines of the surface obtained when assuming that the pixel grey-level is an elevation value – of the image, Desolneux et al. (2001). Since we are exploiting level lines, we can obtain closed regions.

An example of the kind of extraction which can be obtained is shown in Fig. 2.

Once the geometry has been extracted, it must be coded as a description vector of a fixed length. This is achieved by the computation of a set of features. These features can be classified in two groups: low-level features and high-level features.

3.1. Low-level geometry features

Two types of low-level descriptors computed from region edges will be used: the geometric invariants and the Fourier–Mellin descriptors.

3.1.1. Geometric moments

Using the algebraic moment theory, H. Ming-Kuel obtained a family of seven invariants with respect to planar transformations called Hu invariants, Hu (1962). Those invariants can be seen as nonlinear combinations of complex geometric moments:

$$c_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + iy)^p (x - iy)^q f(x, y) dx dy, \quad (1)$$

where x and y are the coordinates of the image $f(x, y)$, i is the imaginary unit and $p+q$ is the order of c_{pq} . The geometric moments are particularly useful in the case of scale changes. Hu invariants have been very much used

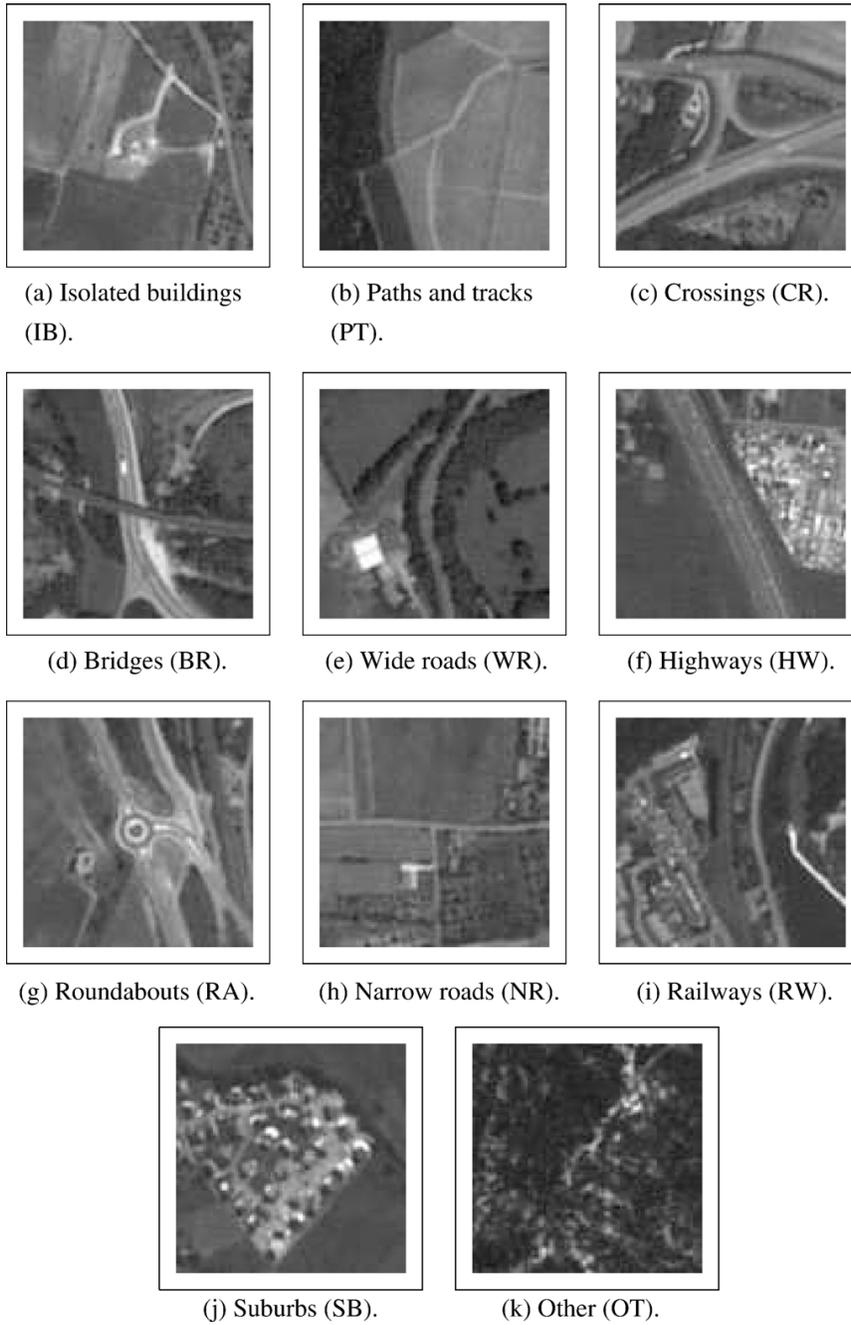


Fig. 1. Examples of images from the data base.

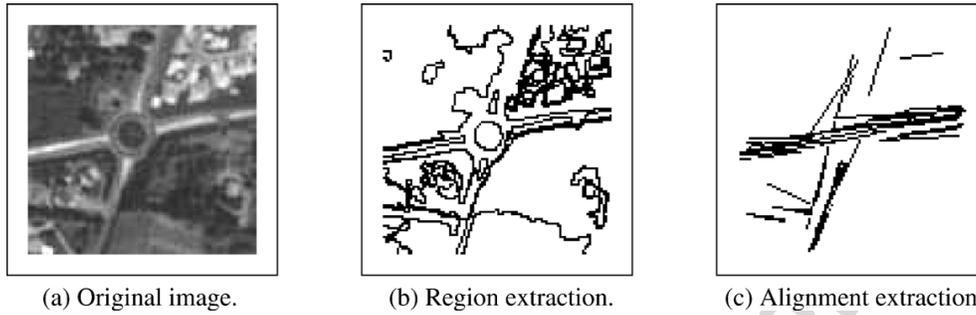


Fig. 2. Examples of geometry extraction.

in object recognition during the last 30 years, since they are invariant to rotation, scaling and translation. Flusser (2000) gives their expressions:

$$\begin{aligned} \varphi_1 &= c_{11}; \quad \varphi_2 = c_{20}c_{02}; \quad \varphi_3 = c_{30}c_{03}; \\ \varphi_4 &= c_{21}c_{12}; \quad \varphi_5 = Re(c_{30}c_{12}^3); \quad \varphi_6 = Re(c_{21}c_{12}^2); \\ \varphi_7 &= Im(c_{30}c_{12}^3). \end{aligned} \quad (2)$$

Dudani et al. (1977) have used these invariants for the recognition of aircraft silhouettes. Flusser and Suk have used them for image registration, Flusser and Suk (1994). They have been modified and improved by several authors. Flusser used these moments in order to produce a new family of descriptors of order higher than 3, Flusser (2000). These descriptors are invariant to scale and rotation. They have the following expressions:

$$\begin{aligned} \psi_1 &= c_{11} = \varphi_1; & \psi_2 &= c_{21}c_{12} = \varphi_4; \\ \psi_3 &= Re(c_{20}c_{12}^2) = \varphi_6; & \psi_4 &= Im(c_{20}c_{12}^2); \\ \psi_5 &= Re(c_{30}c_{12}^3) = \varphi_5; & \psi_6 &= Im(c_{30}c_{12}^3) = \varphi_7. \\ \psi_7 &= c_{22}; & \psi_8 &= Re(c_{31}c_{12}^2); \\ \psi_9 &= Im(c_{31}c_{12}^2); & \psi_{10} &= Re(c_{40}c_{12}^4); \\ \psi_{11} &= Im(c_{40}c_{12}^4). \end{aligned} \quad (3)$$

3.1.2. The Fourier–Mellin transform

Let $f(r)$ be a causal function ($r > 0$), the Fourier–Mellin transform, FMT, of f , if it exists, is written as:

$$\forall(k, v) \in \mathbb{Z} \times \mathbb{R}, M_f(k, v) = \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} f(r, \theta) r^{-iv} e^{-ik\theta} d\theta \frac{dr}{r}, \quad (4)$$

where k and v are respectively the angular and radial frequencies in the transform domain and θ and r respectively are the polar coordinates in the image domain.

The FMT can be seen as the Fourier transform over the group of planar similarities (rotations, translations, and dilations). It is a unique representation of the function f . It has several properties which make it very useful for grey level image analysis, Derrode and Ghorbel (2001). The most important for us will be that the modulus of the FMT coefficients is invariant under rotations and scaling.

3.2. High-level geometry features

For each region extracted from the image patches (see Fig. 2(b)) we can compute the following parameters: perimeter, surface, compacity – defined as the ratio between the surface and the square of the perimeter – and barycenter. The alignments (see Fig. 2(c)) can be described by their length, orientation, and position. They can also be used to compute the location of the intersections.

Since the number of regions and alignments is different for each image patch, we must find a way to transform these parameters into fixed-length vectors. We do so by computing histograms for each image patch. Therefore we will compute the histogram of the compacities of the closed regions, or the histogram of the lengths of the alignments. The parameters related to positions (barycenter, position of the intersections) are transformed into distances with respect to the image patch center. The parameters related to orientations are coded as relative to a principal orientation – the one having the higher weight in the histogram – in the image in order to achieve some kind of rotation invariance. Sizes are normalized with respect to their maximum in each image in order to get some kind of scale invariance.

The high-level features used by our system are:

- the entropy of the orientations of the alignments;
- the histogram of the distances of the intersections to the center of the analysis window;
- the histogram of the lengths of the alignments;

- the histogram of the distances of the barycenters of the closed regions to the center of the analysis window;
- the histogram of the compacities of the closed regions.

4. Feature vector classification

The geometric descriptors presented in the previous section allow us to build a characterization of each of the image patches stored in our data base. By selecting a subset of the data base as a learning base and the complementary subset as the test base we can build an object recognition system and measure its performances. Given the learning data base, we want to obtain a classification system which is able to label each of the images of the test base according to its likelihood of belonging to each of the defined object classes. Since we have an example data base, we can use supervised learning techniques. However, given the low number of examples available and the high dimensionality of the feature space (up to 2500 features per image), we cannot use neural networks or Bayesian learning.

Kernel based learning methods in general and the Support Vector Machines (SVM) in particular, have been introduced in the last years in learning theory for classification and regression tasks, Vapnik (1998). SVM have been successfully applied to text categorization, Joachims (1998), and face recognition, Osuna et al. (1997). Recently, they have been successfully used for the classification of hyperspectral remote-sensing images, Bruzzone and Melgani (2002).

Simply stated, the approach consists of searching for the separating surface between two classes by the determination of the subset of training samples which best describes the boundary between the 2 classes. These samples are called support vectors and completely define the classification system. In the case where the two classes are non-linearly separable, the method uses a kernel expansion in order to make projections of the feature space onto higher dimensionality spaces where the separation of the classes becomes linear. In our problem we do not use the kernel framework. Section 5.2 will show experimental evidence of the lack of interest of nonlinear SVM for our particular application.

One drawback of the SVM is that, in their classical version, they can only solve two-class problems. Some works exist in the field of multi-class SVM (see Allwein et al., 2000; Weston and Watkins, 1998, and the comparison made by Hsu and Lin, 2001), but they are not used in our system.

For problems with $N > 2$ classes, one can choose either to train N SVM (one class against all the others), or to train $N \times (N - 1)$ SVM (one class against each of the others). In

the second approach, which is the one that we use, the final decision is taken by choosing the class which is most often selected by the whole set of SVM. The classification steps consists of taking the images of the test set, computing the feature vector and then applying the set of pre-computed SVM in order to take the decision for the class.

4.1. Feature selection with linear SVM

A linear SVM uses a hyperplane in the feature space in order to choose the label of each sample depending on which side it is located in the feature space with respect to the hyperplane. This classification rule can be seen as a simple thresholding of a real value. This value is the normal distance to the hyperplane of the feature vector of the sample. As in principal component analysis, this distance can be seen as a new variable which is obtained by the linear combination of the variables (features) which are contained in the feature vectors. The weight of each feature in this linear combination is proportional to the degree of orthogonality with respect to the separating hyperplane. A feature parallel to the separating hyperplane would be useless for the classification.

This interpretation allows us to sort the features and select the most pertinent ones for the classification task. We will use this approach in Section 5.2.

5. Results and performances of the system

In order to present the performances of the system, first we show and discuss the results obtained and second, we study several possibilities of optimization of the system.

Since the construction of an exhaustive ground truth for satellite images is a very difficult task – all objects in the scene must be manually extracted and labeled – we will use image patches for the tests also. In order to simulate the behaviour of the scanning of the image by an analysis window, the examples of the class OT – other – are randomly extracted from SPOT 5 scenes. A visual analysis of these patches is then performed in order to eliminate those which contain an object of one of the classes of interest near the center. The behaviour of the system in the neighborhood of the objects will be analysed in Section 5.2.3.

5.1. Analysis of the results

In this section, we show the results obtained by the system. For the training step we will use 75% of the examples in the data base. Since we have generated new elements in the data base by applying symmetries, the total number of examples per class is 450. This gives 340

examples per class for the training and 110 examples per class for the tests.

The feature vectors have 2569 components. Features 0 to 61 are the geometric moments (complex moments, Hu and Flusser invariants), features 62 to 2125 are the Fourier–Mellin coefficients, and features 2126 to 2569 are the high-level features.

We perform two kinds of analysis. First of all, we study the two-class separability problem, that is, we compute the classification ratio for our interest classes (IB, BR, RA, OT) with respect to each of the other classes. This ratio gives us the quality of the classification given by each individual SVM. The results are shown in Fig. 3. As one can see, IB (better than 84%), RA (better than 83%) and OT (better than 87%) are very well separated of other classes. The results for BR are lower, but they are well separated from IB, RA, SB and OT (better than 80%). One can see that most of the BR misclassifications correspond to crossings and different kinds of roads. This can be explained by the fact that a bridge always contains a road and that a bridge is a crossing where the two roads are not at the same level, which is difficult to distinguish from a nadir point of view.

The second kind of analysis we perform is the one of overall system performances, that is, the classification rate for each class and how each of the classes is classified with respect to the others. The results are shown with a confusion matrix (Table 1), where each row i corresponds to the class of the test image, each column j corresponds to the class decided by the system and each cell in the table gives the percentage rate of one object of class i being classified as class j . MD stands for misclassification and FA stands for false alarm. The values are rounded to the nearest integer. The ideal system would have a diagonal confusion matrix with 100% for each term of the diagonal. One can see that this is not exactly the case for our system.

Table 1
Confusion matrix for the recognition system

	IB	PT	CR	BR	WR	HW	RA	NR	RW	SB	OT	MD
IB	87	2	0	0	0	3	1	2	0	3	2	13
PT	10	32	12	4	4	12	2	20	4	0	0	68
CR	0	6	17	20	6	13	15	11	3	6	3	83
BR	0	10	13	30	9	12	14	2	5	4	1	70
WR	3	21	6	12	12	13	2	18	8	5	0	88
HW	1	13	9	20	9	11	4	6	10	6	2	89
RA	1	5	3	1	2	9	66	1	3	9	1	34
NR	7	34	6	6	17	2	2	19	7	0	0	81
RW	0	18	7	23	17	6	2	2	20	2	6	80
SB	0	0	1	0	0	3	26	1	1	56	12	44
OT	0	2	0	0	0	0	2	0	4	36	56	44
FA	3	11	5	9	7	7	8	6	4	6	3	

$D=36.9$. $\kappa=99.330$. The class nomenclature used is: isolated buildings (IB), paths and tracks (PT), crossroads (CR), bridges (BR), wide roads (WR), highways (HW), roundabouts (RA), narrow roads (NR), railways (RW), suburbs (SB).

However, we see that two classes of interest (IB and RA) are detected at 87% and 66% respectively. We also note that most of misclassified bridges are labeled as crossroads, which is rather logical. Also, 34% of PT are classified as being NR and the different types of roads are misclassified as other types of roads, something which we could expect since most of the descriptors used are scale invariant.

In order to give a quality criterion which resumes the information of the confusion matrix we define a diagonality index, also known as Overall Accuracy, as:

$$D = \frac{\sum_i a_{ii}}{\sum_{i,j} a_{ij}} \times 100, \tag{5}$$

where a_{ij} is the element of the confusion matrix on line i and column j . This index is equal to 100 in the case of a perfect recognition system and it is equal to $\frac{100}{N}$, N being

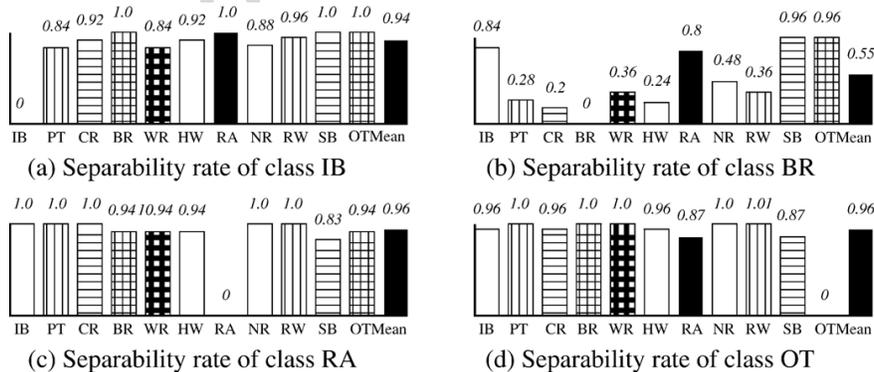


Fig. 3. Two-class separability problem.

the number of classes in the case of a system which randomly chooses the class for each object with the same probability. We will also compute the κ coefficient Congalton (1991), which is often used to measure the performances of classification systems, and which is defined by:

$$\kappa = \frac{p_0 - p_z}{1 - p_z}, \quad (6)$$

where p_0 is the mean value over the diagonal matrix elements, and p_z is the fraction of pixels that could have been accidentally classified correctly:

$$p_z = \frac{1}{N^2} \sum_{i=1}^N \left(\sum_{j=1}^N a_{ij} \sum_{j=1}^N a_{ji} \right) \quad (7)$$

For the confusion matrix of Table 1, we have $D=36.9$ and $\kappa=99.330$.

Since we are trying to reproduce a task which is usually performed by a human operator, it is interesting to compare the performances of the system with the results obtained by a photo-interpreter. In order to do so, we provided a subset of the image data base to five remote sensing experts who were not involved in the data base construction and we asked them to label each one of the image patches. The mean of the confusion matrices of each of the operators is shown in Table 2. As we can observe, the results are better than those obtained by our automatic system in terms of overall accuracy (69%). However, we obtain $\kappa=99.167$, which is lower than the value obtained by the automatic system. This may be understood as the fact that the automatic system makes systematic errors, while the set of photo-

Table 2
Confusion matrix for the human operator experiment

	IB	PT	CR	BR	WR	HW	RA	NR	RW	SB	OT	MD
IB	80	2	0	0	0	0	0	9	2	7	4	20
PT	2	43	13	8	10	2	0	19	2	2	0	67
CR	0	6	60	4	6	3	0	10	6	6	1	40
BR	0	0	16	69	4	6	0	4	0	0	1	31
WR	0	15	0	10	20	10	0	37	15	0	0	80
HW	0	4	0	0	7	82	0	0	7	0	0	18
RA	0	0	3	0	0	0	97	0	0	0	0	3
NR	2	28	12	2	2	2	0	51	0	0	0	49
RW	3	11	20	7	14	5	0	9	27	2	2	73
SB	8	0	0	0	4	0	0	0	0	87	4	13
OT	0	1	0	0	0	0	0	0	0	7	92	8
FA	1	7	6	5	5	4	0	7	3	4	0	

$D=69.0$. $\kappa=99.167$.

Table 3
Confusion matrix for the low-level geometry characterization

	IB	RA	BR	OT+	MD
IB	88.74	2.61	6.54	2.09	11.25
RA	4.90	70.29	7.35	17.45	29.70
BR	2.09	15.96	79.58	2.35	20.41
OT+	13.65	7.11	1.84	77.38	22.61
FA	6.88	8.56	5.41	7.13	

$D=79.01$. $\kappa=99.277$.

interpreters make errors which are not correlated. This may be an additional interest of an automatic system, since the results are reproducible. It is interesting to note that among the five top classes, four are common to both experiments: IB, RA, SB, OT. It is also interesting to note that we have the same trend to confuse BR and CR, and the different classes of roads between them.

Finally, we analyze the results obtained when only the three classes of interest are taken into account. All other classes are gathered under the label OT+. This setup allows us to use fewer SVM, which leads to a simpler and less error-prone voting decision system. This time we analyze the sensibility of the system to the set of descriptors used. Table 3 shows the results obtained with the low-level geometry descriptors only and Table 4 shows the results obtained with the high-level descriptors only. Table 5 shows the results obtained when the two sets of descriptors are jointly used. As one can see, the overall accuracy is much better than when the system had to learn eleven classes. We can also see that the high-level geometry descriptors have worse separability capabilities than the low-level ones, but the best results are obtained when both types of descriptors are used, achieving an overall accuracy of nearly 85%. For the three experiments the κ coefficients are very high and the combination of both approaches gives the best results.

5.2. System optimization

5.2.1. Feature selection

Reducing the computational burden of the classification system is very important in the case of operational processing of images. In this section we investigate the possibility to withdraw some of the features used in the system in order to speed up the processing. In order to identify the features which contain a little information for our problem, we apply the approach presented in Section 4.1. Another approach, as for instance performing the learning and the test steps leaving out different sets of descriptors could be envisaged, but this is very time

Table 4
Confusion matrix for the high-level geometry characterization

	IB	RA	BR	OT+	MD
IB	79.58	3.14	9.94	7.32	20.41
RA	8.84	71.49	11.30	8.35	28.50
BR	8.10	6.71	74.70	10.47	25.29
OT+	6.49	6.49	13.63	73.37	26.62
FA	7.81	5.45	10.57	9.77	

$D=74.79$. $\kappa=99.275$.

consuming if one wants to know the real pertinence of each descriptor.

Fig. 4 show the weight of each feature (direction of each separating hyperplane) for the IB *versus* BR SVM. It is interesting to note that high-level geometry seems to be more useful than low-level geometry – this is true for all SVM – and that Fourier–Mellin coefficients are less significant than geometric moments. It is also interesting to note that the weight of the Fourier–Mellin coefficients seems to follow a pseudo-periodic law. This is due to the fact that their order in the vector is a lexicographic one on the (f_θ, f_ρ) log-polar frequency plane. The results are similar for other pairs of classes, but the more significant frequencies are not the same for all SVM.

Since the SVM classification step consists of computing an euclidean distance in the feature space, this computation is much less expensive than the computation of the features themselves. That means that in order to significantly reduce the computational complexity of the recognition system one has to reduce the number of features to be computed. Only the features which are not significant for all of the SVM can be suppressed. In order to identify thus such features, if any, we compute the mean weight of the features for all the SVM of the system. The result is shown in Fig. 5. One can see that the order of significativity of the features is the same than for the individual SVM. The feature with the maximal value (number 2377, weight 0.049) is one value of the region compacity histogram, but its weight represents less than 5% of the total contribution. On the other hand, each Fourier–Mellin coefficient has a very small contribution, but their integrated contribution is of about 30%. It is also interesting to note that the computational cost of the Fourier–Mellin transform is negligible with respect to the high level geometry, since an FFT-based algorithm is used. These results lead us to the conclusion that it is not possible to identify a set of features which could be suppressed from the vector.

5.2.2. Kernel optimization

Our goal here is to choose the optimum set of parameters for the SVM in order to maximize the recog-

inition performances of the system. The SVM software used for our research, SVM-light Joachims (1999), can use different kinds of kernels (linear, RBF, polynomial and sigmoidal) and each one has its own set of parameters. The trade-off between the training error and the margin is also a parameter of the system. This means that an exhaustive search of the set of parameters of each SVM with a learn and test cycle is a very costly operation.

Using the $\xi\alpha$ -estimates of the error, the precision and the recall introduced by Joachims (2000), one can easily define the following cost function for a given SVM:

$$C = \frac{1}{100} \frac{N_{SV}^2 \cdot (MC + 1) \cdot Err}{Prec \cdot Rec}, \tag{8}$$

with N_{SV} the number of SV after the learning step, MC the number of misclassified learning examples, Err the estimated error, Prec the precision and Rec the recall. Recall is defined as the number of good detections divided by the total number of objects of the class, that is, the detection probability. Precision is defined as the number of good detections divided by the total number of detections. Its meaning is similar to the inverse of false alarm probability. So for Eq. (8), the lower the value of C, the better the generalization capabilities of the SVM. It is interesting to note that this cost function can be computed right after the learning step without the need for using a set of test examples. On the other hand, it is just a theoretical estimate, which is pessimistic. But it is sufficient for the kind of analysis that we want to perform.

Using the cost function of Eq. (8) we have been able to sort by their performances the SVM learned with different sets of parameters for several couples of object classes. Our experiments do not show any kind of trend about the type of kernel to be used. Among the ten best SVM for each couple of classes, there was always the linear kernel. The interest of using a linear kernel is twofold: first, the weight of the components of the feature vector can be directly obtained from the separating

Table 5
Confusion matrix for the high- and low-level geometry characterizations together

	IB	RA	BR	OT+	MD
IB	92.93	1.30	5.23	0.52	7.06
RA	4.01	80.80	4.91	10.26	19.19
BR	2.29	8.97	86.43	2.29	13.56
OT+	10.59	5.57	4.08	79.73	20.26
FA	5.63	5.28	4.14	4.96	

$D=84.99$. $\kappa=99.326$.

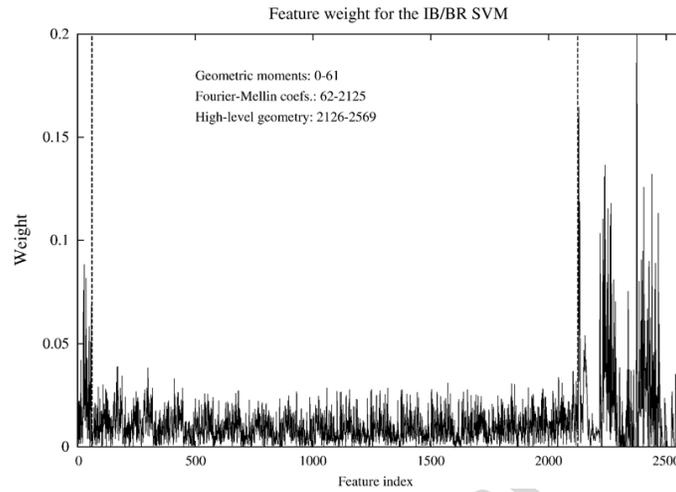


Fig. 4. Feature weight for the IB/BR SVM.

hyperplane (see Section 5.2.1); and second, there is only one parameter to set, the trade-off between the training error and the margin. These conclusions lead us to the choice of linear SVM and allow us to concentrate on the choice of the trade-off. Our experiments show that there is a small variability of the value of this parameter as a function of the couple of classes used and also as a function of the particular set of examples used for the learning step. However, this variability is small enough – in terms of cost function and also in terms of global

system performances – to choose the same value for all the SVM. In our case this parameter was chosen equal to 4.

5.2.3. Sampling step of the analyzing window

Since the object recognition in a complete scene is performed by scanning the image with the analysing window, another interesting parameter of our system is the sampling step used for this scanning. Indeed, we sequentially apply the classification system to an image

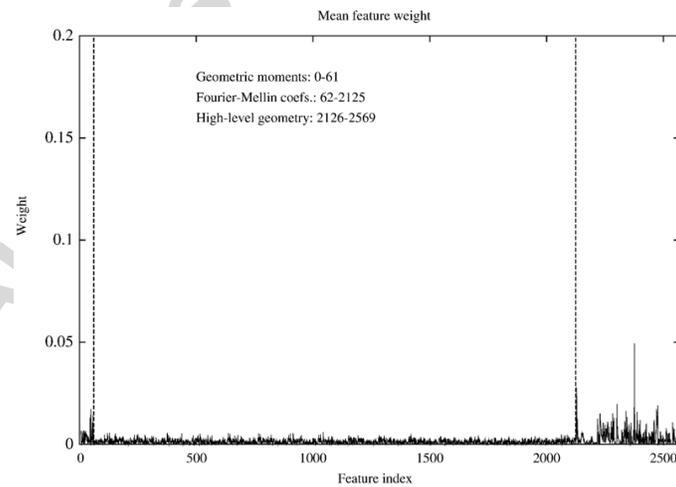


Fig. 5. Mean feature weight for the complete system.

patch extracted with a sliding window. In order to reduce the computation burden, we are interested in using a large sampling step, that is, a small overlapping between successive image patches. On the other hand, a too large sampling step can make the system miss some objects, since the system has been trained to recognise objects centered in the image patches.

So we are interested in the measure of the resolution of our system. This resolution is defined as the width of the point spread function of the system. One way of measuring this point spread function is to apply the recognition system over different kinds of objects and analyse the shape of the system output. One problem here is that the output of a recognition system is a set of labels whose order does not make any sense, so the *shape of the system output* is meaningless. What we propose to do is to measure the shape of the output of each SVM for each couple of classes and derive an

estimate (the mean, for instance) of the overall system resolution. An example of this kind of analysis is shown in Fig. 6 for the couple of classes RA/OT+.

With this kind of experiment we can estimate the maximum sampling step to five pixels which means that for an $N \times M$ pixel image, $\frac{N \times M}{5 \times 5}$ evaluations of the recognition system are needed.

6. Conclusion

We have shown that with a very simple approach it is possible to recognize complex objects on remote sensing high resolution images without explicit model construction. The system is based on the description of the geometry of the objects followed by a supervised SVM classification. We emphasize that the approach presented in this paper is completely independent of the kind of object we want to recognize – assuming that the object fits in the analysing window, this is why we can recognise roads but not road networks, and that its geometric features are pertinent – and that no *a priori* knowledge is introduced in the system. Only a set of examples for the learning step is needed. This is very important for being able to use the system in different application contexts. The system can learn to detect a new class of objects if a set of examples is available. The use of SVM makes learning possible even if only a small number of examples exist. The proposed scheme allows also the use of other descriptors. Texture or statistical grey-level descriptors could be used if they were found useful.

This system could be used in operational information extraction facilities, but it should be preceded by a focusing strategy, since the sequential scanning of entire images can be very time consuming.

Another aspect of the system which should be improved is the set of features used for the classification. We have shown that it is not possible – from an SVM point of view – to eliminate a significant number of features among those used in our system. Other feature selection methods should be tested. Also, some research in order to produce new features pertinent for high resolution remote sensing images should be conducted. Finally, further tests with additional data bases would be interesting in order to validate the applicability of the approach to other types of objects and sensors.

Acknowledgment

The author would like to thank Olivier Caignart, Leonard Potier and Jérôme Tagnères for their help in the development of the object recognition software.

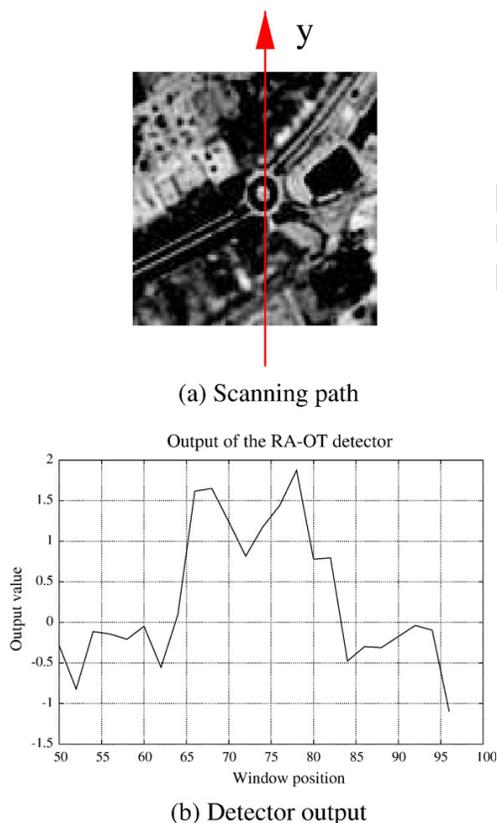


Fig. 6. Example of SVM output over an object of the class RA. The arrow indicates the scanning path of the analysis window.

References

- Allwein, E.L., Schapire, R.E., Singer, Y., 2000. Reducing multiclass to binary: a unifying approach for margin classifiers. Proceedings 17th International Conference on Machine Learning. Morgan Kaufmann, San Francisco, CA, pp. 9–16.
- Baumgartner, A., Steger, C., Mayer, H., Eckstein, W., Ebner, H., 1999. Automatic road extraction based on multi-scale, grouping, and context. *Photogrammetric Engineering & Remote Sensing* 65 (7), 777–785.
- Borgne, H.L., Guérin-Dugué, A., Antoniadis, A., 2004. Representation of images for classification with independent features. *Pattern Recognition Letters* 25 (2), 141–154.
- Bruzzone, L., Melgani, F., 2002. Support vector machines for classification of hyperspectral remote-sensing images. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, vol. 1, pp. 506–508. Jun.
- Congalton, R., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37 (1), 35–46.
- Daschiel, H., Datcu, M., 2005. Information mining in remote sensing image archives: system evaluation. *IEEE Transactions on Geoscience and Remote Sensing* 43 (1), 188–199 Jan.
- Datcu, M., Seidel, K., Walessa, M., 1998. Spatial information retrieval from remote-sensing images — Part I: information theoretical perspective. *IEEE Transactions on Geoscience and Remote Sensing* 36 (5), 1431–1445.
- Datcu, M., Daschiel, H., Pelizzari, A., Quartulli, M., Galoppo, A., Colapicchioni, A., Pastori, M., Seidel, K., Marchetti, P.G., 2003. Information mining in remote sensing image archives: system concepts. *IEEE Transactions on Geoscience and Remote Sensing* 41 (12), 2923–2936.
- Dell'Acqua, F., Gamba, P., 2001. Query-by-shape in meteorological image archives using the point diffusion technique. *IEEE Transactions on Geoscience and Remote Sensing* 39 (9), 1834–1843.
- Derrode, S., Ghorbel, F., 2001. Robust and efficient Fourier–Mellin transform approximations for gray-level image reconstruction and complete invariant description. *Computer Vision and Image Understanding* 83 (1), 57–78.
- Desolneux, A., Moisan, L., Morel, J.-M., 2000. Meaningful alignments. *International Journal of Computer Vision* 40 (1), 7–23.
- Desolneux, A., Moisan, L., Morel, J.-M., 2001. Edge detection by Helmholtz principle. *Journal of Mathematical Imaging and Vision* 14 (3), 271–284.
- Dudani, S., Breeding, K., McGhee, R., 1977. Aircraft identification by moments invariants. *IEEE Transactions on Computers* 26 (1), 39–45.
- Flusser, J., 2000. On the independence of rotation moment invariants. *Pattern Recognition* 33 (9), 1405–1410.
- Fraser, C., Baltsavias, E., Gruen, A., 2002. Processing of Ikonos imagery for submetre 3D positioning and building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing* 56 (3), 177–194.
- Gerke, M., Heipke, C., Straub, B.-M., 2001. Building extraction from aerial imagery using a generic scene model and invariant geometric moments. Proceedings of the IEEE/ISPRS joint Workshop on Remote Sensing and Data Fusion over Urban Areas, November 8–9th 2001, University of Pavia, Rome (Italy), pp. 85–89.
- Gülch, E., Müller, H., Läbe, T., Ragia, L., 1998. On the performance of semi-automatic building extraction. *International Archives of Photogrammetry and Remote Sensing* 32 (Part 3/1), 331–338.
- Hamdan, R., Heitz, F., Thoraval, L., 2001. Modèles probabilistes d'apparence: une représentation approchée de faible complexité. *Traitement du Signal* 18 (3).
- Hinz, S., Baumgartner, A., 2003. Automatic extraction of urban road networks from multi-view aerial imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (1–2), 83–98.
- Hsu, C., Lin, C., 2001. A comparison of methods for multi-class support vector machines. Technical Report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, p. 19.
- Hu, M.K., 1962. Visual pattern recognition by moment invariants. *IEEE Transactions on Information Theory* 8 (2), 179–187.
- Flusser, J., Suk, T., 1994. A moment based approach to registration of image with affine geometric distortion. *IEEE Transactions on Geoscience and Remote Sensing* 32 (2), 382–387.
- Joachims, T., 1998. Text categorization with support vector machines: learning with many relevant features. In: Nédellec, Claire, Rouveirol, Céline (Eds.), *Proceedings of the European Conference on Machine Learning*. Springer, pp. 137–142.
- Joachims, T., 1999. Making large-scale SVM learning practical. In: Schölkopf, B., Burges, C., Smola, A. (Eds.), *Advances in Kernel Methods — Support Vector Learning*. MIT Press.
- Joachims, T., 2000. Estimating the Generalization Performance of a SVM Efficiently. *Proceedings of the International Conference on Machine Learning*. Morgan Kaufman.
- Mena, J., 2003. State of the art on automatic road extraction for GIS update: a novel classification. *Pattern Recognition Letters* 24 (16), 3037–3058.
- Osuna, E., Freund, R., Girosi, F., 1997. Training support vector machines: an application to face detection. *Proceedings of IEEE Computer Vision and Pattern Recognition Conference*, pp. 130–136.
- Pope, A.A., Lowe, D.G., 1996. Learning appearance models for object recognition. *International Workshop on Object Representation for Computer Vision*, Cambridge, UK, pp. 201–219.
- Rochery, M., Jermyn, I.H., Zerubia, J., 2003. Higher order active contours and their application to the detection of line networks in satellite imagery. *Proc. IEEE Workshop Variational, Geometric and Level Set Methods in Computer Vision*, at ICCV, Nice, France.
- Schroeder, M., Rehrauer, H., Seidel, K., Datcu, M., 1998. Spatial information retrieval from remote-sensing images — Part II: Gibbs–Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing* 36 (5), 1446–1455.
- Schroeder, M., Rehrauer, H., Seidel, K., Datcu, M., 2000. Interactive learning and probabilistic retrieval in remote sensing image archives. *IEEE Transactions on Geoscience and Remote Sensing* 38 (5), 2288–2298.
- Stoica, R., Descombes, X., Zerubia, J., 2004. A Gibbs point process for road extraction from remotely sensed images. *International Journal of Computer Vision* 57 (2), 121–136.
- Thorpe, S., Delorme, A., Rullen, R.V., 2001. Spike-based strategies for rapid processing. *Neural Networks* 14 (6–7), 715–725.
- Turk, M., 2001. A random walk through eigenspace. *IEICE Transactions on Information Systems* E84-D (12), 1586–1695.
- Turk, M., Pentland, A.P., 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3 (1), 71–86.
- Vapnik, V., 1998. *Statistical Learning Theory*. John Wiley and Sons, New York.
- Veltkamp, R., Hagedoorn, M., 1999. State-of-the-art in Shape Matching. Tech. Rep. UU-CS-1999-27. Utrecht University, The Netherlands.
- Weston, J., Watkins, C., 1998. Multi-class support vector machines. Tech. Rep. CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, Egham, TW20 0EX, UK.

16 Raisonement spatial

Qualitative Spatial Reasoning for High-Resolution Remote Sensing Image Analysis

Jordi Inglada and Julien Michel

Abstract—High-resolution (HR) remote-sensing images allow us to access new kinds of information. Classical techniques for image analysis, such as pixel-based classifications or region-based segmentations, do not allow to fully exploit the richness of this kind of images. Indeed, for many applications, we are interested in complex objects which can only be identified and analyzed by studying the relationships between the elementary objects which compose them. In this paper, the use of a spatial reasoning technique called region connection calculus for the analysis of HR remote-sensing images is presented. A graph-based representation of the spatial relationships between the regions of an image is used within a graph-matching procedure in order to implement an object detection algorithm.

Index Terms—Graph theory, image analysis, image representations, spatial reasoning.

I. INTRODUCTION

BECAUSE of the diversity of sensors and the increase of their spatial resolution and repetitivity, the automatic analysis of images is a crucial asset in the remote-sensing field. It is therefore needed to design and implement new image analysis techniques which are able to perform complex processing in an efficient manner. In this context, much work has been done for the automatic information extraction from remote-sensing images aiming to provide a set of features and descriptors which are compact and parsimonious in order to feed them into learning systems. This type of features come often from low-level processing and are not able to capture the richness and complexity of high-resolution (HR) images. Indeed, the improvement of the spatial resolution makes that the objects of interest are not any more limited to several pixels, but they are represented by large areas containing subobjects. This kind of objects cannot efficiently be described by textures, edges, etc.

Recent publications present interesting advances in the recognition of particular objects as buildings [1] or urban areas [2], but this kind of approaches are not generic enough to deal with different types of objects.

It is therefore useful to use techniques which are able to deal with higher levels of abstraction for the representation and the

manipulation of the information contained in the images. In this paper, we will use spatial reasoning techniques in order to describe complex objects. We have chosen to use the well-known region connection calculus (RCC8) [3] in order to describe the relationships between the regions of an image.

The regions used for the spatial reasoning are obtained with a multiscale segmentation based on geodesic morphology [4]. The mathematical morphology approach has been selected because, in contrast to linear multiscale approaches, it allows one to select objects in terms of their size. This has been shown to be interesting in meter and submeter resolution images for detecting vehicles, buildings, etc. Our algorithm is able to produce regions superimposing across the scales.

This allows one to exploit the full extent of the RCC8 relationships set. The regions and their relationships are represented by means of an attributed relational graph (ARG) where the nodes represent the regions and the arcs represent the RCC8 relationships. The ARG can be made more complete by adding region attributes to the nodes, as for instance shape and radiometric features.

Finally, object detection and recognition can be seen as a graph-matching problem for which efficient algorithms can be implemented. In these context, a greedy search combined with a graph metric which is able to use all the information contained in the ARG has been implemented.

In this paper, the theoretical basis of the approach and examples of application to real images will be presented.

This paper is organized as follows. Section II introduces the principles of RCC. Section III presents the approach used for image segmentation. Section IV presents the graph-matching problem. Section V describes the implementation of the procedure. Section VII presents some example of results on real images. Section VIII draws some conclusions.

II. RCC

A. Spatial Reasoning

The spatial organization of the objects in images gives an important information for the recognition and interpretation tasks, particularly when the objects are located in complex environments as it is usually the case in remote-sensing HR images.

Spatial reasoning can be defined as the set of techniques for the manipulation of the description of physical objects by taking into account their mutual interactions expressed in terms of their shapes, sizes, positions, orientations, and plausible movements [5].

There are several space logics. Different theories use different primitive features: points, lines, polygons, etc. The base

Manuscript received December 7, 2007; revised March 13, 2008 and July 4, 2008. First published November 25, 2008; current version published January 28, 2009.

J. Inglada is with the Centre National d'Études Spatiales, 31401 Toulouse, France (e-mail: jordi.inglada@cnes.fr).

J. Michel is with the Communications et Systèmes, 31506 Toulouse, France (e-mail: julien.michel@c-s.fr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2008.2003435

0196-2892/\$25.00 © 2008 IEEE

spatial relationships used by each theory may also be different: inclusion, occlusion, connection, etc. Mereological—study of the parts¹ and belonging theories—, topological theories—where limit points and connectivity play a main role—, and mereotopological theories—connectivity, belonging, external connectivity—have been proposed for the last 20 years.

However, literature shows that image processing applicative works do not use the aforementioned theories, but rather more pragmatic approaches. This has the consequence of not being able to describe complex scenes, but only testing simple hypothesis on image objects. In a context of data mining, automatic interpretation, learning by example, Geographical Information Systems, etc., more powerful theories are needed.

There is a great variety of representations for storing and reasoning about spatial information. One can distinguish two main approaches: qualitative representation based on object coordinates and qualitative representations based on a high conceptual level vocabulary for the description of spatial relationships.

Qualitative languages can add very powerful functionalities to spatial information systems working with quantitative data.

We will present here two possible approaches to qualitative spatial reasoning about space:

- 1) qualitative spatial reasoning systems applied to quantitative representations of space;
- 2) qualitative theories of space and their associated reasoning systems.

The motivation for a qualitative representation of space comes from the need to perform computations allowing a computer to represent and reason with spatial concepts without relying on quantitative techniques.

del Pobil *et al.* [5] propose to define the base concepts—shape, size, relative position, etc.—but they do not provide an associated theory nor an algebra (operators).

Bennet *et al.* [6] have studied the tradeoff between expressive power, the possibility of performing calculus, and the straightforwardness of the qualitative formalisms. They have shown that a set of topological relationships can be described by a first-order language and how they can be coded in a zero-order intuitionist logic to provide an efficient algorithm. They also proposed ways of combining qualitative and quantitative information with a coherent architecture:

- 1) quantitative data structures are used inside a polygonal region database;
- 2) a qualitative relationship language is used to make high-level queries;
- 3) an intuitionist propositional logic is used to compute the inferences needed to answer user queries.

Wang and Liu [7] propose a visual knowledge network which is actually an algebraic structure on which *visual* operations (union, zoom, occlusion, etc.). Inference algorithms capable of performing generalization and specialization can be used on this type of structure.

It can therefore be concluded that many different approaches have been proposed. Cohn and Hazarika [8] made an exhaustive synthesis work.

¹Meros, Greek for part

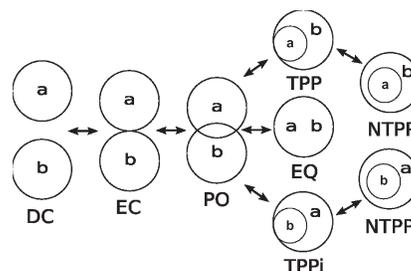


Fig. 1. Transitions between RCC8 relationships.

B. RCC8

RCC [3] is based on the notion of connection between pairs of regions in space. This connection is described by one among a set of possible relationships. One can therefore derive different RCC systems depending on the number and the nature of the connections accounted for. One of these systems, RCC8 is particularly interesting, since it is made of exhaustive and disjoint relationships. These properties simplify the reasoning techniques. As shown in Fig. 1, the eight RCC8 relationships go from disconnection, DC, to equivalence, EQ, through all intermediate possibilities by taking into account tangency. It is important to note that all these relationships are symmetrical except for TPP and NTPP which are antisymmetrical. This explains the additional relationships TPPi and NTPPi and their number (eight instead of six). If one does not take into account tangency, the RCC5 system is obtained.

After defining the RCC8 system, one can wonder how to compose relationships: knowing that region (*a*) is linked to region (*b*) by relationship R_1 and that region (*b*) is linked to region (*c*) by relationship R_2 , which is the relationship between regions (*a*) and (*c*)? Even in some cases, one can find a unique answer to this question; most of the cases do not allow this. For instance, if the two known relationships are DC, no information can be inferred about the composition [9], [10]. This is a limitation of the RCC8 system. However, the composition table of RCC8 (Table I) will help to speed up computations.

The implementation of a reasoning system based on RCC8 and more generally based on qualitative information has made the object of many research [11]–[15]. Many approaches have been proposed, some of them using fuzzy logic in order to overcome the weak nature of the composition relationships, some others trying to define an algebra for relationships. These works are usually purely theoretical and do not propose any hint for the practical implementation. Works by Jun Chen *et al.* [16] proposing modifications to RCC8 using the Voronoi regions of the objects in order to enrich the DC case, are worth noting.

III. MULTISCALE IMAGE SEGMENTATION

A. State of the Art

As mentioned in Section II, RCC deals with regions of space and how they connect and overlap each other. Therefore, the first step toward the use of this system is to obtain a suitable set of regions. The requirement for overlapping excludes a single

TABLE I
RCC8 RELATIONSHIP COMPOSITION TABLE

$R_1 \setminus R_2$	DC	EC	PO	TPP	NTPP	TPPi	NTPPi	EQ
DC	no info	DC, EC, PO, TPP, NTPP	DC, EC, PO, TPP, NTPP	DC, EC, PO, TPP, NTPP	DC, EC, PO, TPP, NTPP	DC	DC	DC
EC	DC, EC, PO, TPPi, NTPPi	DC, EC, PO, TPP, TPPi, EQ	DC, EC, PO, TPP, NTPP	EC, PO, TPP, NTPP	PO, TPP, NTPP	DC, EC	DC	EC
PO	DC, EC, PO, TPPi, NTPPi	DC, EC, PO, TPPi, NTPPi	no info	PO, TPP, NTPP	PO, TPP, NTPP	DC, EC, PO, TPPi, NTPPi	DC, EC, PO, TPPi, NTPPi	PO
TPP	DC	DC, EC	DC, EC, PO, TPP, NTPP	TPP, NTPP	NTPP	DC, EC, PO, TPP, TPPi, EQ	DC, EC, PO, TPPi, NTPPi	TPP
NTPP	DC	DC	DC, EC, PO, TPP, NTPP	NTPP	NTPP	DC, EC, PO, TPP, NTPP	no info	NTPP
TPPi	DC, EC, PO, TPPi, NTPPi	EC, PO, TPPi, NTPPi	PO, TPPi, NTPPi	PO, TPP, TPPi, EQ	PO, TPP, NTPP	TPPi, NTPPi	NTPPi	TPPi
NTPPi	DC, EC, PO, TPPi, NTPPi	PO, TPPi, NTPPi	PO, TPPi, NTPPi	PO, TPPi, NTPPi	PO, TPP, NTPP, TPPi, NTPPi, EQ	NTPPi	NTPPi	NTPPi
EQ	DC	EC	PO	TPP	NTPP	TPPi	NTPPi	EQ

layer segmentation method, where all the regions would be disconnected from each other.

Watershed segmentation allows segmentation at different level of details, but this segmentation is hierarchical, and regions at a higher level of detail are included into regions at a lower level, which is not suitable for RCC.

Laporterie [17] introduced a pyramidal decomposition of images based on mathematical morphology. She showed the interest of this method, called morphological pyramid, in the field of image analysis and fusion. Still, the morphological pyramid decomposition produces a multiresolution decomposition of the scene, which induces two types of loss of precision for the details. First, morphological filters are nonlinear, so Shannon sampling conditions not guaranteed and information is lost at each subsampling step. Second, we cannot recover precisely the extracted details at full resolution.

Pesaresi and Benediktsson [4] derived a new morphological segmentation based on the morphological profiles. They showed the interest of using a geodesic metric in the morphological operators, which make them adaptive to the structures in the image and reduces shape noise. Such operators are called morphological operators by reconstruction. They used these operators to design two functions representative of the convex and concave structures of the image with respect to the size of the structuring element. They then used these functions to design a single-layer single-scale segmentation of these structures, as well as a single-layer multiscale segmentation based on the profile of these functions for increasing structuring element sizes.

B. Proposed Segmentation Method

Neither the morphological pyramid nor the morphological profiles can fit our needs. The first is far too lossy, and the second is a single-layer method, with no possible overlapping. Still, we can take advantages of the iterative, pyramidal structure of the first one, while using the more robust concave and convex membership functions of the second one. For a given pixel, the convex membership function can be seen as the likelihood for that pixel to be part of a convex (i.e., brighter than the surrounding background) structure of the image whose size is smaller than the size of the structuring element. A similar

definition can be applied to the concave (i.e., darker than the surrounding background) membership function. The leveling function $\psi_N(f)$ is a simplification of the image f , where all convex and concave structures with a size smaller than the size of the structuring element have been obliterated.²

Definition 1 (Convex Membership Function): The convex membership function $\mu_N(f)$ is defined as follows:

$$\mu_N(f) = f - \gamma_N^*(f) \quad (1)$$

where f is the image function and γ_N^* denotes the opening by reconstruction operator, with a structuring element of size N .

Definition 2 (Concave Membership Function): The concave membership function $\eta_N(f)$ is defined as follows:

$$\eta_N(f) = \varphi_N^*(f) - f \quad (2)$$

where f is the image function and φ_N^* denotes the closing by reconstruction operator, with a structuring element of size N .

Definition 3 (Leveling Function): The leveling function $\psi_N(f)$ is defined as follows:

$$\psi_N(f) = \begin{cases} \gamma_N^* & : \mu_N(f) > \eta_N(f) \\ \varphi_N^* & : \mu_N(f) < \eta_N(f) \\ f & : \mu_N(f) = \eta_N(f). \end{cases} \quad (3)$$

Using these three functions, we propose the following segmentation scheme.

- 1) Define the range of scales $S = \{s_1, s_2, \dots, s_n\}$ for the segmentation, where n is the number of levels of analysis and s_i is the size of the structuring element used in the segmentation step i .
- 2) Define f_0 as being the image to segment.
- 3) For each scale s_i in the range S , the steps are as follows.
 - a) Segment the convex details of the image f_i using the convex membership function μ_{s_i} .
 - b) Segment the concave details of the image f_i using the concave membership function η_{s_i} .
 - c) the image f_{i+1} as being the image f_i simplified by the leveling function ψ_{s_i} .

²For more details on morphological operators, geodesic distance, and leveling, please refer to the original paper [4].

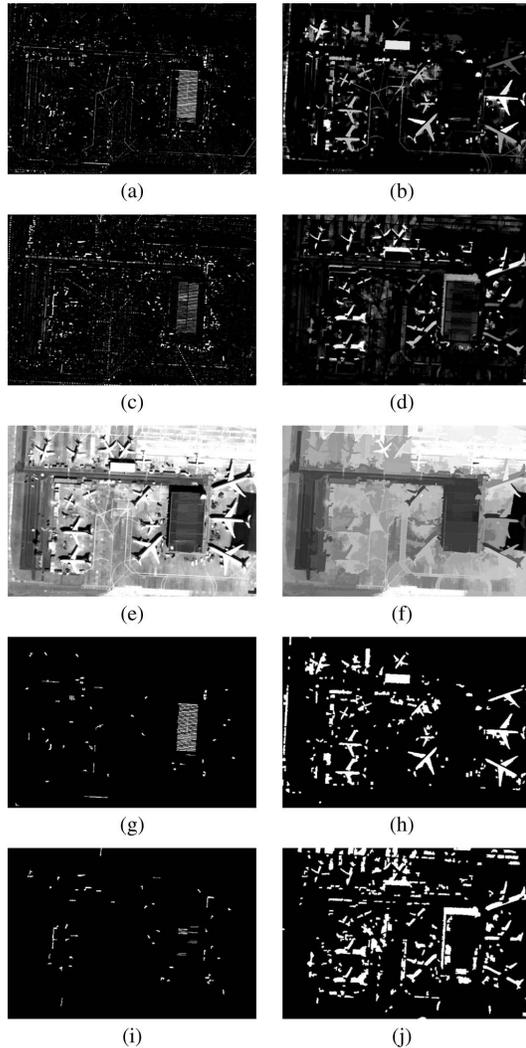


Fig. 2. Application of the multiscale geodesic segmentation algorithm to image 2. (a) Convex map (scale 2). (b) Convex map (scale 15). (c) Concave map (scale 2). (d) Concave map (scale 15). (e) Leveling map (scale 2). (f) Leveling map (scale 15). (g) Bright segmentation (scale 2). (h) Bright segmentation (scale 15). (i) Dark segmentation (scale 2). (j) Dark segmentation (scale 15).

The range of scales S can be chosen in order to select objects of a given size. Although a decision rule is proposed in [4] for the segmentation steps a) and b), we choose to use a basic thresholding technique, which seems to extract more information from the images. For a range of n segmentation scales, this segmentation method produces n segmentation layers of convex details and n segmentation layers of concave details. Fig. 2 shows the result of applying four levels of the segmentation algorithm to the image of Fig. 3. We can see the simplification effect of the leveling function, as well as the accuracy of the convex and concave membership function in



Fig. 3. Quickbird image at 60-cm resolution.

detecting brighter and darker details. The segmentation images have been postprocessed with shape regularization using binary morphology and object size filtering.

Although this segmentation approach gives satisfaction in our application, it is worthy to note that the results are far from being perfect and more research in this direction is needed.

IV. GRAPH MATCHING

Graph matching has widely been used for object recognition and image analysis. One can cite for instance the works of Buhmann *et al.* [18] where graph matching on a multiresolution hierarchy of local Gabor components is used. A similar approach is proposed in [19]. The same author extends this paper in [20] where the hierarchical graph matching allows for size and invariant object recognition. The drawback of this approach is that the model has to be built before, whereas we are looking for a representation of objects, groups of objects, etc.

In [21], each node in the graph represents a feature (e.g., curvature points) and arcs describe the distances between the features. A Hopfield binary network is used to perform sub-graph matching.

The works of Bunke [22], as for instance, constitutes a very good introduction to object recognition with graph matching. A novel concept, the mean of a set of graphs is introduced. Also, Sanfeliu *et al.* [23] give a good overview of graph-based representations and techniques for image processing and analysis.

Cyr and Kimia [24] measure the similarity between two views by a 2-D shape metric of similarity measuring the distance between the projected, segmented shapes of the 3-D object. However, the approach is limited to the recognition of the very same object and for very simple shapes.

Shams *et al.* [25] propose a comparison of graph matching and mutual information maximization for object detection restricted to the case of multidimensional Gabor wavelet features. They show that graph matching has a computational complexity which is two to three times lower than mutual information.

The originality of the approach presented here is to work on graphs which describe qualitative spatial relationships, which will allow us to represent classes of objects, rather than multiple points of view of a particular object.

A. Theoretical Background

Graphs are abstract mathematical entities which are useful for modeling and solving complex problems. They are a

generalization of the tree concept which is usually used in computer science for solving decision problems. One should note that, while the complexity and implementation of algorithms operating on trees are well known, their generalization to graphs brings new complexity issues. Graphs are nevertheless used to solve classical problems as the traveler salesman problem or the representation of complex structures as for instance neural networks. We will use the following definition.

Definition 4 (Graph): A graph G is made of a finite set of nodes V and a set E of binary relationships on $V \times V$ called arcs.

To types of graphs can be distinguished.

Definition 5 (Oriented Graph): An oriented graph G is a graph whose arcs are couples of nodes (ordered relationship).

Definition 6 (Nonoriented Graph): A nonoriented graph G is a graph whose arcs are pairs of nodes (ordered relationship).

From these definitions, one can define the node's neighborhood.

Definition 7 (Node's Neighborhood): For a nonoriented graph, a node's neighborhood is defined as the set of nodes which are linked to it by an arc.

For an oriented graph, the node's neighborhood is defined as the set of destination nodes having the given node as source.

B. Graph Similarities

Three different approaches to graph comparison can be distinguished [26].

- 1) Measure of physical properties: A set of physical characteristics which may be problem dependent are measured in the graph and are used to compare several graphs.
- 2) Occurrence of canonical structures: The comparison is made using a set of canonical structures present in the graphs.
- 3) Structural comparison: the comparison is made using a measure defined on the graph structure.

The structural approach is the one which exploits best the graph topology without being problem dependent, but is rather difficult to implement in an efficient manner.

C. Isomorphism, Distance, and Similarity Measure

In order to measure a purely structural similarity between two graphs, many measures have been proposed [27], [28]. Most of these measures are based on the research of the largest common subgraph between the two compared graphs. Sorlin and Solnon [29] defined a generic measure which unifies several approaches. It is based on an evaluation of a function which depends on the two compared graphs. This function depends on the common characteristics of the two graphs (nodes, arcs, and labels associated to both of them), and it may be tuned using two functions f and g which determine the nature of the measure. The following definitions are useful for the remainder of the presentation.

Definition 8 (Graph Match): Let G and G' be two graphs defined, respectively, over V and V' . A **graph match** is a subsample of $m \subseteq V \times V'$ containing the set of pairs (v, v') so that the nodes v and v' are matched.

Definition 9 (Common Descriptors): We call **common descriptors** of a graph match m the set $descr(G) \sqcap_m descr(G')$, where the operator \sqcap_m stands for the set of all nodes, node's labels, arcs, and arc's labels which are common to G and G' within a match m .

Definition 10 (Split Nodes): We call **Split nodes** or **splits(m)** of a match m of two graphs G and G' the set of nodes G matched to several nodes of G' and the nodes G' matched to several nodes of G .

These definitions allow us to introduce the following generic similarity measure.

Definition 11 (Similarity With Respect to a Match): The similarity between two graphs G and G' with respect to a match m is defined by

$$sim_m(G, G') = \frac{f(descr(G) \sqcap_m descr(G')) - g(splits(m))}{f(descr(G) \cup descr(G'))} \quad (4)$$

where f and g are two application dependent functions.

Definition 12 (Similarity Between Two Graphs): The similarity between two graphs G and G' is defined by

$$sim_m(G, G') = \max_{m \subseteq V \times V'} \frac{f(descr(G) \sqcap_m descr(G')) - g(splits(m))}{f(descr(G) \cup descr(G'))} \quad (5)$$

where f and g are two application dependent functions.

Depending on the choice made for f and g several measures can be obtained (see Table II).

V. IMPLEMENTATION

A. Elementary Computation

1) *Principles:* The first step for assessing the usefulness of RCC8 for image analysis, was to implement a technique for the efficient computation of the relationship between two regions of space. The first tool we will use is the nine-intersections matrix (Table III), which is a binary 3×3 matrix where 0 (empty set) and 1 (nonempty sets) represent the intersections between the exterior, the boundary, and the interior of region A with the interior, exterior, and boundary of region B.

$$\mathbf{M} = \begin{pmatrix} card(\check{A}, \check{B}) & card(\check{A}, \bar{B}) & card(\check{A}, \bar{\bar{B}}) \\ card(\bar{A}, \check{B}) & card(\bar{A}, \bar{B}) & card(\bar{A}, \bar{\bar{B}}) \\ card(\bar{\bar{A}}, \check{B}) & card(\bar{\bar{A}}, \bar{B}) & card(\bar{\bar{A}}, \bar{\bar{B}}) \end{pmatrix}$$

where $card$ is a function which is equal to 1 if the set is not empty and 0 otherwise, \check{A} is the interior of region A, \bar{A} its boundary, and $\bar{\bar{A}}$ its exterior. Among the 2^9 possible matrices, only eight are physically plausible, and they correspond to the eight RCC8 relationships. Being able to compute the nine-intersections matrix for a pair of regions of space means knowing the RCC8 relationship between them.

2) *Optimization:* This method has a main drawback: its complexity. This complexity is dependent on the image size. This is particularly annoying, since remote-sensing images are very large, while the regions of interest may be small. Therefore, using masks which have the same size of the image can

TABLE II
SEVERAL SIMILARITY MEASURES OBTAINED FROM THE GENERIC MEASURE

Name	Function f	Function g	Description
Graph isomorphism	cardinal	cardinal	All elements of graph G are found on graph G' and vice-versa
Partial sub-graph isomorphism	cardinal restricted to G	cardinal	All the elements of graph G are found on graph G'
Largest common partial sub-graph	cardinal	$g(S) = +\infty$ if $S \neq \emptyset$	Largest graph being isomorphic to partial sub-graphs of G and G'
Editing distance	cost of the elementary editing operations	$g(S) = +\infty$ if $S \neq \emptyset$	Minimum cost of transforming G into G'

TABLE III
CORRESPONDENCE BETWEEN THE RCC8 RELATIONSHIPS AND THE NINE-INTERSECTIONS MATRICES

$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$
<i>DC</i>	<i>EC</i>	<i>PO</i>	<i>TPP</i>
$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
<i>TPPi</i>	<i>NTPP</i>	<i>NTPPi</i>	<i>EQ</i>

very much slow the process. In order to reduce the computation time, three optimizations have been implemented.

The first optimization concerns the cardinal computation. Since what interests us is the fact whether this cardinal of the intersections is zero or not, we implement a lazy operator, which goes through the image and stops when it finds the first nonzero pixel.

The following optimization concentrates on the computation of the elements of the nine-intersections matrix. We have seen that only eight among the 2^9 possible matrices are physically plausible. This means that the components of the matrix are redundant. Therefore, a binary decision tree is used where only one component at each step is examined. This tree allows one to find the RCC8 relationship by examining at most four components of the matrix. The decision tree used is presented in Fig. 4.

B. Computing From Images

This implementation has first been proposed in [30]. For each region, we have a binary mask coming from the segmentation step. The exterior of the region is obtained by inverting the mask. The interior of the region is obtained by applying a morphological erosion of size 1 pixel. The boundary is obtained by subtracting the interior to the mask. Once these three regions are obtained for each one of the regions of interest, we obtain the intersections by applying a binary addition of those.

Despite of the two optimizations presented Section V-A2, the computation of the relationships from images is still very time consuming. Since, as it will be shown later on, the most common relationship between the regions of the image is disconnection, a particular optimization for this case has been implemented. We exploit the fact that most of the disconnected regions are also far away from each other. That means that their bounding boxes will also be disconnected. The bounding box computation is much faster than the morphological operations,

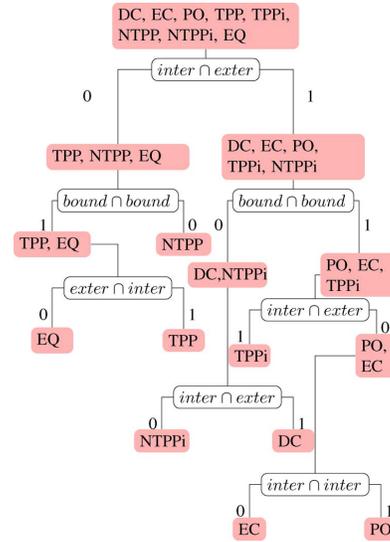


Fig. 4. Binary decision tree for the RCC8 relationships.

so this way of evaluating the DC relationship, which is the first one in the decision tree, will allow an important acceleration of the evaluation of the relationships.

C. Vector Computation

Another way of computing the RCC8 relation from the segmentation is to first vectorize the two regions into polygons. Then, we try to decide each of the four conditions in the decision tree. We track down edges configuration characterizing one of the following conditions.

- 1) Interior of A with exterior of B: We need to find one vertex of A which strictly lies out of B, or two vertices of A lying strictly inside B, and linked by an edge crossing B an even times.
- 2) Interior of B with exterior of A: This is exactly the contrary of the first point.
- 3) Interior of A with interior of B: We need to find a vertex of A strictly inside B or two vertices of A outside B and crossing B an even times.
- 4) Edge of A with edge of B: We need to find two tangent edges or a vertex of A on an edge of B, or the opposite.

This vectorization step introduces an improvement of precision and complexity. In the end, an average computation time 20 times lower than the image computation algorithm is obtained.

D. Analysis for a Set of Regions

1) *Basics of the Computation:* The application of the computation of the relationship between a pair of regions to all the pairs of regions in the image may be very time consuming. In order to reduce the number of computed relationships, we will take into account the symmetry properties of the relationships. This allows one to reduce by a factor of two the number of computed relationships. We also use the fact that the multiscale segmentation gives disconnected regions for a given scale.

2) *Optimization:* Using the composition Table I, further optimizations can be implemented. Indeed, this table allows us to unambiguously infer the relationship in 27 cases out of 64. In 16 other cases, the composition table allows one to jump the first level of the binary decision tree of Fig. 4. In other seven cases, the composition table allows one to decide the third level of the tree.

In order to fully exploit this information, before computing the relationship between a region (a) and a region (c), we will look for an intermediate region (b) for which the relationships which link it to (a) and (c) have already been computed. If such a region exists, two cases may appear.

- 1) The knowledge of the already computed relationships allows one to jump the first or the third level of the decision tree: This information is stored and the examination of the intermediate regions found goes on.
- 2) The knowledge of the already computed relationships allows one to unambiguously determine the new relationship: the computation is finished.

If at the end of the examination of the intermediate regions it was impossible to determine the needed relationship, a computation using the optimizations of the previous section is done by using all the information stored during this precomputation step.

3) *Results:* First, we segment a Quickbird 60-cm resolution panchromatic image with the multilevel segmentation technique proposed in Section III. The segmentation is performed with two different scales, which gives us four layers containing a total of 400 regions. Using the RCC8 computation based on polylines presented in Section V-C, and the group optimization presented in Section V-D2, the overall extraction process took only 50 s, including the segmentation step. Compared to the computation based on images presented Section V-B, the computation based on polylines is 20 times faster in average. The speed gain for a single computation is from 2 times faster to 1500 times faster depending on the regions' configuration. This optimization is significant regarding our previous work [3]. The optimization presented in Section V-D2 implies a gain of about 30%. This optimization helps to decide about 40% of the relationships, and can solve 11% of the cases without any computation. Table IV shows the statistics of the RCC8 relationships obtained with the Quickbird scene. Of course, these statistics will depend on the segmentation

TABLE IV
EXAMPLE OF RCC8 STATISTICS

Relationship	Proportion	Occurrences
DC	99.63 %	159408 relationships
EC	0.055 %	88 relationships
PO	0.0625 %	100 relationships
TPP	0 %	0 relationships
NTPP	0.00125 %	2 relationships
TPPi	0 %	0 relationships
NTPPi	0.00125 %	2 relationships
EQ	0.25 %	400 relationships

parameters. In particular, the number of levels of the analysis will have some effects on the less represented relationships' statistics.

VI. GRAPH COMPARISON TECHNIQUES

A. Information Representation

Once a method is available for the computation of the RCC8 relationships between the different regions of an image, a data structure is needed for their representation and manipulation. A graph structure is straightforward for our problem, since we have a set of elements—regions—which can be described by attributes (index of the image, index of the region in the image, region characteristics, etc.) and a set of relationships linking these elements (the RCC8 relationship). We have therefore decided to use a nonoriented graph, where the arcs represent a pair of relationships (the direct and the inverse relationships) without representing the DC relationship. This graph structure will allow us to use well-known algorithms for graph manipulation and matching. Figs. 5 and 6 show two examples of the RCC8 graph representation for two scenes.

B. Extraction of Connected and Biconnected Components

As shown in Fig. 6, the more complex the scene, the more dense and detailed the RCC8 graph. It is therefore interesting to have a mean of splitting the graph in a set of coherent subgraphs which correspond to objects in the scene. In order to do that, we will use to different concepts: connected graphs and biconnected graphs.

Definition 13 (Connected Graph): We call a graph G connected if for any couple of nodes (v, v') of G , one can find a path from v to v' through a set of arcs and intermediate nodes.

Definition 14 (Biconnected Graph): We call a graph G biconnected if it is connected and if the removal of a single node makes it become not connected.

We are therefore interested in partitioning a graph in a set of connected graphs of maximum size (i.e., to which one cannot add any node without making it losing its connection). The same goes for biconnection, which is a stronger condition than connection. Indeed, a connected graph can be itself split in a set of biconnected graphs. The nodes which allow the connection between the biconnected elements of a graph are called articulation points. Fig. 7 shows the difference between the two concepts. Articulation points are shown in red.

Both of this concepts are interesting for our problem. When we are looking for a particular object in an image, obtaining the

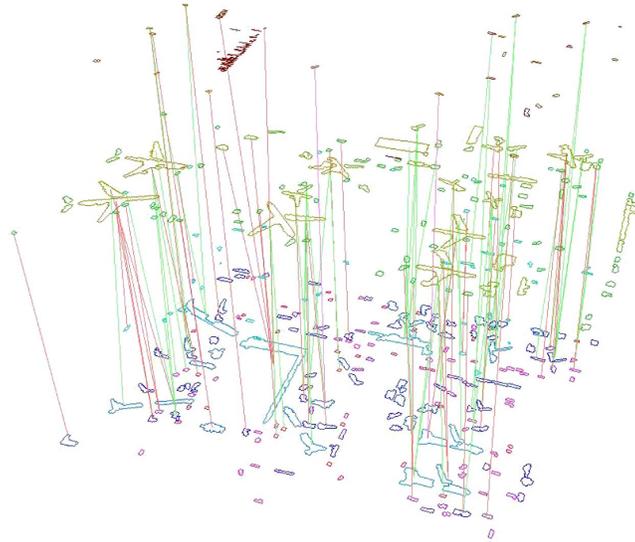
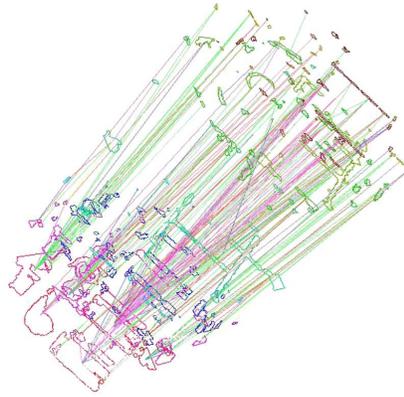


Fig. 5. Three levels graph of the airport scene.

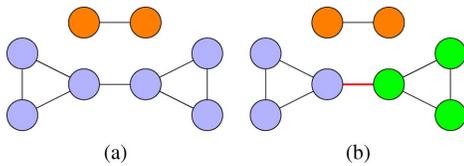


(a)



(b)

Fig. 6. Example of six levels graph of a water treatment plant 60-cm resolution image. (a) Original scene. (b) Graph of the scene.



(a)

(b)

Fig. 7. Example of connection and biconnection. (a) Connected components. (b) Biconnected components.

graph corresponding to this object is difficult, since objects in the neighboring area may be mixed up with it. It is therefore interesting to decompose the obtained graph into connected

components in order to isolate the graph of the object of interest. Fig. 8 shows an example of application.

In very large scenes, the extraction of connected components may not be enough to isolate the objects of interest. The use of biconnected components can then be used. Fig. 8 shows an example of use of biconnected components.

C. Similarity Computation Algorithm

In Section IV-C, a generic similarity measure between graphs has been presented. We have also shown that this similarity measure can be computed by computing the maximum of a

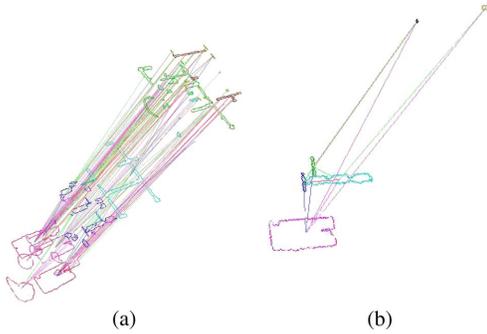


Fig. 8. Example of connected and biconnected component extracted from the graph presented Fig. 6. (a) Connected component. (b) Biconnected component.

similarity measure over the set of the sets of possible matchings between the nodes of two graphs. This is an NP-hard problem, i.e., one cannot check in polynomial time that a solution is the optimal one. This is due to the size of the solution space, which grows exponentially with the size of the problem. Therefore, any exhaustive search in the solution space has to be discarded. However, using heuristics, algorithms can be proposed to find an acceptable solution to the problem. In our case, we will use the approach proposed by Sorlin and Solnon [29]: A first step is performed using a greedy search, and the obtained result is improved by a taboo metaheuristics.

1) *Building the Search Space*: In order to implement these two algorithms, a representation of the search space under the form of a metagraph has been built. Each node of this new graph represents the association of a node of graph G with a node of graph G' of the same level of segmentation. Two nodes of the metagraph are linked with an arc if the region of graph G associated to the first node is linked to the region of graph G associated to the second node by the same relationship which links the region of graph G' associated to the first node to the region of graph G' associated to the second node. Also, this relationship has to be different from DC. Finally, labels are given to the nodes and arcs of the metagraph in order to point which node and which arc is already part of the solution. Using this structure, we can take advantage of the properties of graphs: If one considers a solution of size n , the neighboring solutions of size $n + 1$ are directly accessible. Indeed, one only needs to take into account the set of metanodes linked by an arc to the nodes which already belong to the solution. One can also make sure that a connected graph will be matched with another connected graph. Fig. 9 shows a simple example of search graph: Each node of graph 1 is matched to the nodes of graph 2 of the same level. These metanodes are then linked if the relationships which links the origin and destination nodes of graph 1 and graph 2 are of the same type.

2) *Greedy Search Heuristics*: We have seen that the computation of the similarity between two graphs is equivalent to maximizing (5).

Since the denominator of this equation is constant, one can only consider the maximum of the numerator, which we call score function. Furthermore, a potential measure will be used. The potential will account for the output arcs which are not

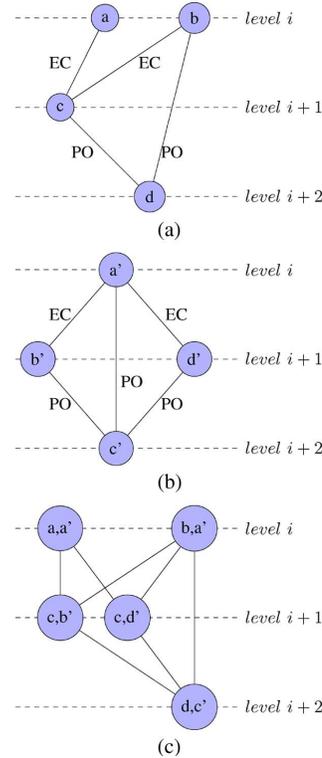


Fig. 9. Example of how to build a search space. (a) Graph 1. (b) Graph 2. (c) Search graph.

yet matched. It expresses the potential of a match to widen the search in the solution space. The principle of the greedy search is as follows.

- 1) Randomly choose a first match among those maximizing the potential and add it to the empty solution.
- 2) While there are matches which can increase the score and the maximum score is not reached, the following are performed.
 - a) Build the list of candidate matches which increase the score the most.
 - b) Among those, build the list of candidate matches maximizing the potential.
 - c) Among those, randomly choose a match and add it to the solution.
 - d) Go to step a).

This heuristic has two particularities. First of all, it is not deterministic, since, in the case of equality between matches, the choice is random. This allows one to run the algorithm several times and keep the best solution. Second, it is a gradient descent algorithm, since in each step, the steepest slope of the score function is chosen. This is why this algorithm can converge to a local minimum. Nevertheless, this is a good way to get a first guess of the solution for large complex problems.

Alignment and parallelism for the description of objects in high resolution remote sensing images

Maria Carolina Vanegas^{a,b,1,*}, Isabelle Bloch^a, Jordi Inglada^{b,c}

^a*Institut Telecom, Telecom Paristech, CNRS-LTCI UMR 5141, 46 Rue Barrault, Paris, France*

^b*Centre National d'Etudes Spatiales, 18 Avenue Edouard Belin, Toulouse, France*

^c*CESBIO UMR 5126, Toulouse, France*

Abstract

Alignment and parallelism are frequently found between objects in high resolution remote sensing images, and can be used to interpret and describe the scenes. In this work, we propose novel representations of parallelism and alignment as spatial relations. For the parallel relation we propose a fuzzy relation which is defined between a linear object and a non linear object and groups of aligned objects. For the alignment relation we propose a method for extracting the groups of aligned objects from a labeled image. The alignment relation is also defined as a fuzzy relation, and we distinguish two cases of alignment: *local* and *global*. In *local* alignment each object of the group is aligned with its neighbors, while in the *global* alignment every object of the group is aligned to all other members. To extract the *locally* aligned groups we make use of fuzzy relative position measures. The *locally* aligned groups are candidates for being *globally* aligned groups, which are determined using a graph-based approach. Illustrative examples on real images show the power of description of combining these two relations for image interpretation.

Keywords: parallelism, alignment, fuzzy spatial relations, spatial reasoning, image interpretation, high resolution remote sensing imaging

1. Introduction

Very high resolution remote sensing images (less than 1m per pixel) allow us to discriminate different objects that compose a scene, such as buildings in a urban area or airplanes in an airport. However, recognizing individual objects is not enough to determine the semantics of a complex scene and to completely interpret it. Semantical scene understanding involves the assessment of the spatial arrangement of objects. Using spatial relations does not only help us to discriminate the objects in the scene [1], but it also allows us to distinguish between different interpretations of two scenes with similar objects but different spatial arrangements [2]. Some examples of the use of spatial relations can be found in the domain of medical images to recognize different brain structures [3, 4], in image interpretation to provide linguistic scene descriptions [5], in remote sensing images to classify or mine images [2, 6], in GIS applications to monitor land use [7] and cover changes [8].

The spatial relations that are usually used to describe the spatial arrangement between image regions are topological relations as in [2, 7–9], directional relations such as “to the right of” in [2, 4, 5, 10] or distance relations [2, 4, 5, 11]. However, relations such as alignment have not been used so far in these works, although they can play an important role in image description. This is the case for instance in cartography, where it is necessary to find groups of aligned buildings for map generalization [12]. As another example, in object detection, complex semantic classes such

*Corresponding author

Email addresses: carolina.vanegas@telecom-paristech.fr (Maria Carolina Vanegas),
isabelle.bloch@telecom-paristech.fr (Isabelle Bloch), jordi.inglada@cesbio.cnes.fr (Jordi Inglada)

¹Tel: + 33 1 45 81 74 94, Fax: + 33 1 45 81 71 44

as parking areas (car parking lots, ports, truck parking lots or airports) comprise aligned groups of vehicles. Therefore, the identification of aligned groups of transport vehicles can be useful for detecting instantiations of these complex classes. Alignment is often linked to the parallel relation. For example, in object recognition the aligned groups of airplanes in an airport are parallel to the terminal buildings, or the aligned groups of boats in the ports are parallel to the deck. Another example is the determination of urban spatial patterns like groups of aligned houses parallel to other aligned groups of houses in residential areas.

In this paper, we propose representations of the parallelism and alignment relations as spatial relations between objects. Our objective is to define these relations taking into account their semantics. However, these relations are vague and imprecise when they are evaluated between objects of different sizes for the case of alignment, and between objects of different spatial extensions for the case of parallelism. Therefore we propose to define them as fuzzy spatial relations to cope with such imprecisions.

This paper is organized as follows. Section 2 reviews some of the models of alignment and parallelism in computer vision. Section 3 introduces the definitions of *local* alignment and *global* alignment. A method for extracting the *locally* aligned groups is then proposed. In this method we construct a neighborhood graph of the objects of the image, and its dual graph where we incorporate information about the relative direction of the objects, evaluated using fuzzy measures of relative position. The groups of objects satisfying the fuzzy criterion of being *locally* aligned are extracted from the dual graph. These groups are the candidates for being *globally* aligned. Section 4 discusses the issues related to the definition of the parallel relation when dealing with objects or groups of aligned objects. An original definition of parallelism is then proposed, matching the requirements. Finally, in Section 5 we apply the relations to objects in satellite images to illustrate their behavior and their usefulness in scene understanding.

2. Alignment and parallelism in computer vision

Both parallelism and alignment between low level features have been widely studied in computer vision. Some examples are parallelism between segments [13–19], alignment between groups of points [20, 21] and alignments between linear segments [13, 20] as collinearity on digital images. Most of these works are inscribed within the framework of perceptual organization. Their first objective is to find how to organize low level features, such as edge segments, into groups, such as aligned segments or parallel segments. The groups are evaluated according to their perceptually significance based on the grouping laws of the Gestalt theory [22, 23]. Their second objective is to differentiate the groupings that arise from the structure of a scene from those that arise due to accidents of view point or position [13].

The objective of perceptual organization in computer vision and our objective are different. Perceptual organization deals with how these relations take place among low level features and their meaningfulness according to the image's structure, while we are interested in defining these relations for objects in the image taking into account their semantics. Anyway, it is important to review the work that has been done on studying these relations as grouping laws between low level features in digital images, in order to understand how these relations have been modeled.

Parallelism between two linear segments is usually modeled as a relation that should satisfy three constraints. The first one is that both segments should have a small angular difference. The second one is that the perpendicular distance between the two segments should be small and the last one is that there should be a high overlap between them [14]. There exist different approaches to integrate these three constraints into a model and measure the meaningfulness of the relation. Lowe [13] was one of the first to model parallelism to perform perceptual grouping: for every couple of linear segments having a perpendicular separation d and an angular difference θ , he assigns a significant value to establish that it has not been originated by an accident of viewpoint. The significant value is used to determine the expected number of lines for a given perpendicular separation and an angular difference. In fact, this value is proportional to the prior probability of appearance and it is inversely proportional to the angular difference and the perpendicular separation. In [14] the constraint about overlapping is introduced and it is determined by the

orthogonal projection of one segment over the other and vice-versa. The meaningful segments are then obtained by applying a threshold on the measured values of the constraints. Fuzzy approaches have been proposed in [16–18, 24, 25], leading to a measure of the degree of parallelism between two linear segments, in which trapezoid type functions are used to evaluate to which degree the three constraints are satisfied. The three degrees are combined in a conjunctive way.

Parallelism between curves has been modeled as a type of symmetry [15, 16]. In [15] the authors consider that two curves are parallel if their respective orthogonal distances from the symmetry axis are almost equal. In [26] and [27] parallelism is modeled as a matching problem where two curves are parallel if there is a point-wise correspondence. In both works the parallelism is treated as a shape matching problem.

Previous works focused on parallelism between crisp linear segments. However, when dealing with objects extracted from images, it is important to consider parallelism between fuzzy linear objects. Indeed, the object extraction processes can introduce imprecision, and therefore we are not always dealing with crisp linear objects or segments. In Sec. 4 we will give some definitions about parallelism between fuzzy linear objects. The definitions of parallelism presented in this paper differ from those of the previous works in the sense that we are defining the relation considering its semantics, and we are not worried about whether or not the detected parallelism is an accident of the view point or position. We are interested in the conditions that have to be satisfied to decide whether two objects are parallel or not.

For the case of alignment, several methods to determine the alignment between points methods relying on the Hough transform [20] or the Radon transform [28] have been proposed. Other examples are the identification of aligned segments which have the same orientations as the alignment [13, 18, 20, 21, 24]. However, alignment extraction as a high level feature has been less studied. One example is the work of [29], where an algorithm to detect aligned groups of buildings in maps is presented. In this algorithm buildings with aligned barycenters are extracted, and the quality of the alignments is evaluated based on the criteria of proximity and similarity laws of Gestalt theory.

The above mentioned methods for determining alignment focused on extracting groups of objects with aligned barycenters. Consequently, it is not possible to apply them directly to determine if a group of heterogeneous objects (according to shape and size) is aligned, since neither the size nor the shape of the objects are considered. An example of this will be given in Sec. 3.1. Similarly as for parallelism, the previously proposed alignment relations have not been extended to deal with fuzzy objects.

Due to the fact that parallelism and alignment in computer vision have been studied between low-level features in the perceptual organization domain, they have been treated in an independent manner, and they are used to organize low level features but they are not combined. However, when dealing with objects and observing these relations as spatial relations, it is interesting to combine them in order to determine when two groups of aligned objects are parallel, or when a group of aligned objects is parallel to another object. These kinds of combinations give us more information about the scene and can be meaningful for the description. This point will be further developed in the following sections.

3. Alignment

Alignment can be defined as the spatial property possessed by a group of objects arranged in a straight line². As it was highlighted by the Gestalt theory, alignment should be seen as a whole [20]: if we observe each element of the group individually, then the alignment property is lost. Having to look at it as a whole makes alignment detection a difficult task, since in order to detect an aligned group of objects we have to identify its members, but to know if an object belongs to an aligned group the alignment has to be already identified. In this part we study the different approaches to model alignment, and introduce the notions of *local* and *global* alignment.

²Definition taken from ThinkMap Visual Thesaurus <http://www.visualthesaurus.com/>

3.1. Different approaches to model alignment

For the case of a group of points in \mathbb{R}^2 there are two equivalent strategies for verifying whether they are aligned. Let $\mathcal{A} = \{a_1, \dots, a_n\}$ be a group of points in \mathbb{R}^2 . \mathcal{A} is aligned if and only if:

- (i) there exists a line L that intersects all the points, or
- (ii) there exists an angle $\theta \in [0, \pi)$ such that for every pair of points a_i and a_j ($i \neq j$), a_j is located in direction θ or $\theta + \pi$ from a_i with respect to the horizontal axis.

The first strategy has been used to identify if a group of points is aligned in an image, by considering that points should fall into a strip [30], the thinner the strip the better the alignment. Extending the first definition to identify a group of aligned objects can be done by using objects' barycenters. Unfortunately this will only be appropriate for objects of similar sizes (see Fig. 1) and approximately convex. Another possibility is to search a thin strip where all the objects fall into, but the width of the strip will depend on the objects' sizes. Thus the notion of falling into a thin strip is not appropriate for objects with different sizes.

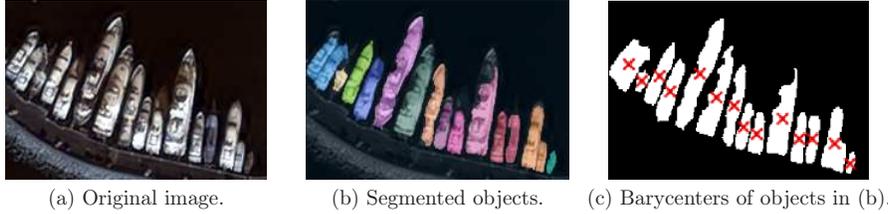


Figure 1: Example of an aligned group of objects of different sizes and with non-aligned barycenters.

The difficulty of extending the second strategy relies in determining the angle between two objects. One alternative, which is the one proposed in this work, is to use measures of relative position used in spatial reasoning. Before entering into the details of the method, we introduce the notion of orientation histogram, inspired by the angle histogram of [31]. This notion is a fundamental concept for our method.

3.2. Angle and orientation histograms

Angle histograms were introduced in [31]. They can be interpreted as a function that captures the directional position between two objects. Let us assume we have two objects A and B defined by two regions in the image space \mathcal{J} , that we also denote by A and B . The angle histogram from A to B is obtained by computing, for each pair of points $p_a \in A$ and $p_b \in B$, the angle between the segment joining them and the horizontal axis, denoted by $\angle(p_a, p_b)$ [32]. Angles are organized in a histogram normalized by the largest frequency:

$$H^A(B)(\theta) = \frac{\sum_{p_a \in A, p_b \in B | \angle(p_a, p_b) = \theta} 1}{\max_{\phi \in [0, 2\pi)} \sum_{p_a \in A, p_b \in B | \angle(p_a, p_b) = \phi} 1}. \quad (1)$$

To determine if an object A is in a given direction with respect to an object B (for example "right of"), we can compute the angle histogram $H^A(B)$ and compare it with a template for the relation "right of" by using for instance a conjunctive operator or the compatibility between the computed histogram and the template [31]. Angle histograms are also formalized for fuzzy objects. Given two fuzzy objects A and B defined through their membership functions $\mu_A : \mathcal{J} \rightarrow [0, 1]$ and $\mu_B : \mathcal{J} \rightarrow [0, 1]$ in the image space \mathcal{J} , respectively, it is possible to define the angle histogram between A and B by considering the membership of each point to the fuzzy set:

$$H^A(B)(\theta) = \frac{\sum_{p_a, p_b \in \mathcal{J} | \angle(p_a, p_b) = \theta} \mu_A(p_a) \wedge \mu_B(p_b)}{\max_{\phi \in [0, 2\pi)} \sum_{p_a, p_b \in \mathcal{J} | \angle(p_a, p_b) = \phi} \mu_A(p_a) \wedge \mu_B(p_b)}, \quad (2)$$

where \wedge denotes a t-norm and \vee a t-conorm. When μ_A and μ_B are crisp, (2) and (1) are equivalent. In the following, \wedge denotes a t-norm and \vee a t-conorm. Angle histograms have proved to be an adequate way for evaluating the directional spatial relation between two objects [31] since they take into account the shape of the regions. In addition, they are invariant to simultaneous translation, scaling and rotation of both objects. They are not symmetrical, but they satisfy: $H^A(B)(\theta) = H^B(A)(\theta + \pi)$.

Since we are interested in the orientation between two objects, we introduce the notion of orientation histogram, which is simply an angle histogram where the angles are computed modulo π and its support has a length equal to π :

$$O(A, B)(\theta) = \frac{\sum_{p_a, p_b \in \mathcal{I} | \text{mod}(\angle(p_a, p_b), \pi) = \theta} \mu_A(p_a) \wedge \mu_B(p_b)}{\max_{\phi \in [0, \pi)} \sum_{p_a, p_b \in \mathcal{I} | \text{mod}(\angle(p_a, p_b), \pi) = \phi} \mu_A(p_a) \wedge \mu_B(p_b)}. \quad (3)$$

The orientation histogram is a fuzzy subset of $[0, \pi[$ that represents the orientation between two objects, it preserves the same properties as the angle histogram and, in addition, is symmetrical.

3.2.1. Similarity degree for orientation histograms

There are several ways to define a degree of similarity between orientation histograms. One possibility is to interpret the orientation histograms as fuzzy sets and use similarity measures based on aggregation [33]. Another possibility is to interpret the orientation histogram as a function and use a similarity degree based on distance between functions. We use here a simple similarity degree based on intersection, but other similarity measures are possible. More information on distances between histograms and similarity measures between fuzzy sets can be found in [33–35].

A degree of similarity between two orientation histograms can be defined as the maximum height of the fuzzy intersection of the two orientation histograms [33]. Let $O(A, B)$ and $O(C, D)$ be two orientation histograms. The degree of similarity between them is given by:

$$\text{sim}(O(A, B), O(C, D)) = \max_{\theta \in [0, \pi)} [O(A, B)(\theta) \wedge O(C, D)(\theta)]. \quad (4)$$

In this measure, the possibility for each angle is aggregated in a conjunctive way, and we take its maximum value. In the case were the orientation histograms do not intersect then the similarity value is 0.

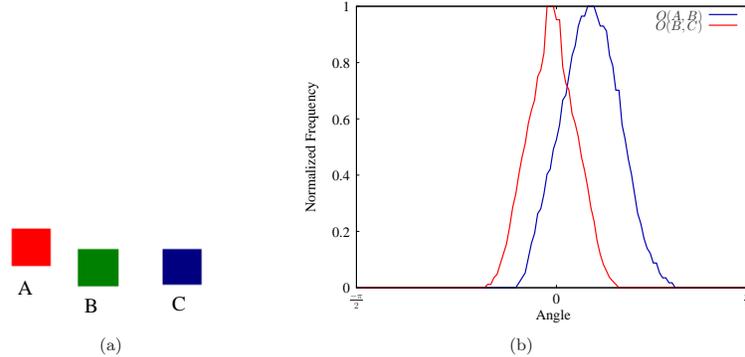


Figure 2: (a) Objects. (b) Orientation histograms of the objects in (a).

Figure 2(b) shows the two orientation histograms computed for the objects of Fig. 2(a). Both histograms have a well defined maximum, showing a strong main orientation, and these maximum values are close to each other. Therefore a high similarity value is expected. However, if the

\min operator is used as a t-norm in (4), the similarity value between the two histograms is 0.67. This non-high value is due to the comparison of the value for each angle separately, since the aggregation is done for every angle. However, in this context referring to an orientation equal to an angle θ actually represents the quantity “approximately θ ”. Therefore, in order to compare if two orientation histograms are similar, it is important to consider the imprecision that is linked to the comparison of two angles that are approximately the same. When a fuzzy morphological dilation [36] is performed on an orientation histogram using a structuring element ν_0 , then the high values of the histogram will be propagated to the similar angle values according to ν_0 . The structuring element ν_0 is designed such that $\nu_0(\theta - \hat{\theta})$ represents the degree to which $\hat{\theta}$ and θ are “approximately” equal, and in our experiments we modeled ν_0 as a trapezoid function:

$$\nu_0(\theta) = \begin{cases} 1 & \text{if } |\theta| \leq t_1 \\ \frac{t_2 - |\theta|}{t_2 - t_1} & \text{if } t_1 < |\theta| \leq t_2 \\ 0 & \text{if } |\theta| > t_2, \end{cases} \quad (5)$$

where t_1 and t_2 represent the parameters of a trapezoid function. The parameters t_1 and t_2 are related to the imprecision linked to computing angles in a discrete grid, as is the case for images. This imprecision is a function of the distance between the points for which the angle is computed: the greater the distance between the points, the more precise is the calculation. However, in our experiments, since most of the time we are dealing with objects which have a similar distance between them, we have chosen to use:

$$t_1 = \arcsin\left(\frac{1}{0.5d_{average}}\right) \quad (6)$$

$$t_2 = \arcsin\left(\frac{1}{0.25d_{average}}\right) \quad (7)$$

where $d_{average}$ is the average distance in pixels between the “neighbor” object’s barycenters. This values have been used in all experiments, and did not required any fine tuning of the parameters.

By performing a morphological dilation of $O(A, B)$ with the structuring element ν_0 we introduce the imprecision to all the angles in $O(A, B)$ (as in [37]). For this reason, the similarity should be computed using the dilated orientation histograms, and (4) becomes:

$$sim(O(A, B), O(C, D)) = \max_{\theta \in [0, \pi]} [D_{\nu_0}(O(A, B))(\theta) \wedge D_{\nu_0}(O(C, D))(\theta)], \quad (8)$$

where $D_{\nu_0}(O(X, Y))$ is the dilation of $O(X, Y)$ by a structuring element ν_0 and it is given by $D_{\nu_0}(\mu)(\theta) = \sup_{\hat{\theta} \in [0, \pi]} \min(\mu(\hat{\theta}), \nu_x(\theta - \hat{\theta}))$ [36].

Figure 3(a) shows the result of performing a morphological dilation of the orientation histograms of Fig. 2(b). The parameters of Eqs. 6 and 7 are $t_1 \approx 0.07$ and $t_2 \approx 0.12$ for the structuring element. The similarity value between the histograms obtained with Eq. (8) is 0.93 which is more consistent with the perceived orientation of the objects than the value obtained by applying Eq. (4).

When orientation histograms are not similar (see Fig.4) we obtain a zero similarity value, as desired.

This measure of similarity can be extended to compare several orientation histograms. Let $\{O(A_i, A_j)\}_{i=0, j \neq i}^N$ be a set of orientation histograms, the similarity degree between them is equal to:

$$sim(O(A_0, A_1), \dots, O(A_i, A_j), \dots, O(A_N, A_{N-1})) = \max_{\phi \in [0, \pi]} \bigwedge_{i=0, j \neq i}^N D_{\nu_0}(O(A_i, A_j))(\phi) \quad (9)$$

3.3. Definition and identification of aligned groups of objects

We consider a set of objects $\mathcal{A} = \{A_0, \dots, A_n\}$ and want to determine the subsets of aligned groups of \mathcal{A} . For the sake of clarity we will call them *globally* aligned groups. First we identify the

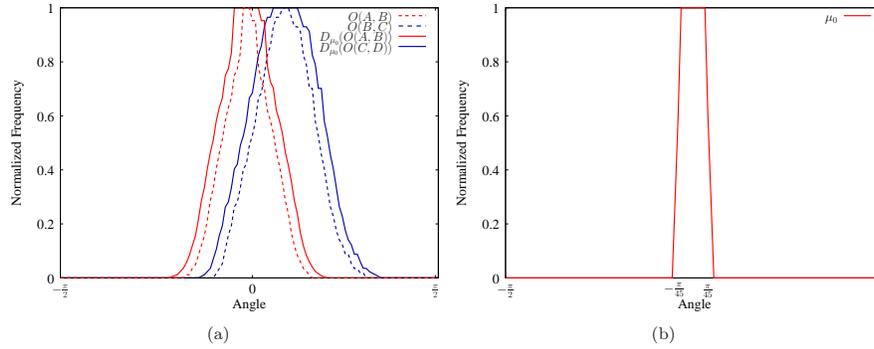


Figure 3: (a) Orientation histograms of the objects in Fig. 2(a) and dilated orientation histograms. (b) Structuring element used for the dilation.

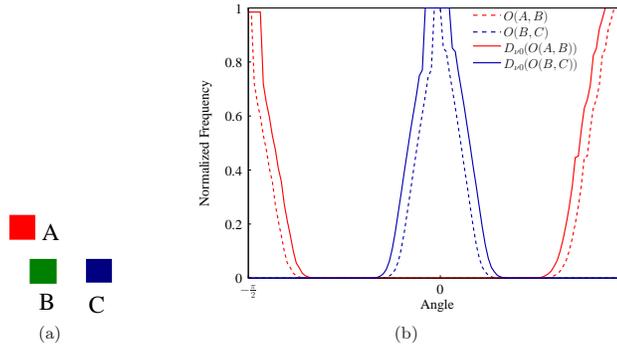


Figure 4: (a) Objects. (b) Orientation histograms of the objects (a) and dilated orientation histograms.

locally aligned groups (defined in Sec. 3.3.2), and the set of *locally* aligned groups will be denoted by \mathcal{L} . For each *locally* aligned group $\mathcal{S}_i \in \mathcal{L}$ we measure the degree of *global* alignment (defined in Sec. 3.3.1). If this degree is lower than a user acceptance value α , then elements of the group will be deleted until the degree is equal or greater than α or until the group has less than 3 elements. In the case where the degree is greater than α , the group \mathcal{S}_i will be added to the set of aligned groups \mathcal{G} . A last step of addition and fusion of the groups is done to obtain the largest possible *globally* aligned groups. This method is illustrated in Fig. 5. In the following sections we detail each step of the method.

3.3.1. Globally aligned groups

Before defining a *globally* aligned group of objects, we define two preliminary concepts. Let \mathcal{S} be a group of objects, and $A, B \in \mathcal{S}$, we define the *Neigh*(A, B) relation as being satisfied if and only if $B \cap A = \emptyset$ and $B \cap N(A) \neq \emptyset$ where $N(A)$ is defined as a neighborhood of A . One possible choice for the neighborhood of an object is $N(A) = N_d(A)$ where $N_d(A)$ is the Voronoi neighborhood constrained by a distance d . Other possible choices will be discussed in Sec. 3.3.3.

Definition 3.1. A group \mathcal{S} is called *connected* by the *Neigh* relation if for every $A, B \in \mathcal{S}$, there

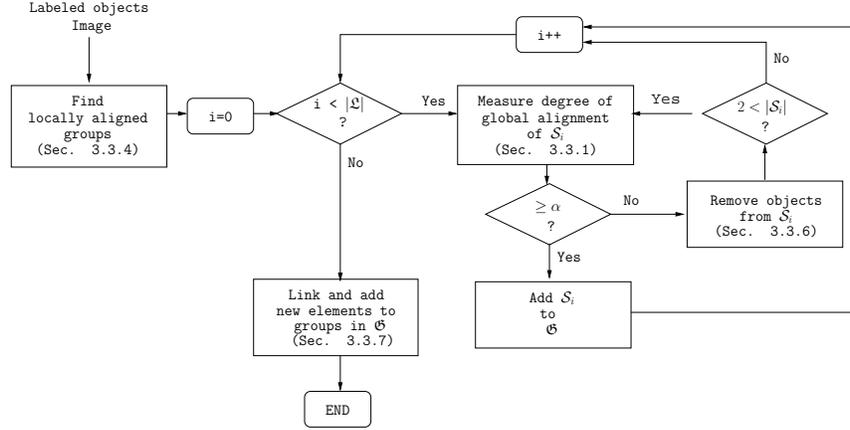


Figure 5: Proposed method for determining the aligned groups of objects.

exist C_0, \dots, C_M objects in \mathcal{S} , such that $C_0 = A$, $C_M = B$ and for every $m = 0, \dots, M - 1$, the relation $Neigh(C_m, C_{m+1})$ is satisfied.

Returning to the discussion of Sec. 3.1, the group \mathcal{S} is *globally* aligned if the following conditions are satisfied:

- (i) \mathcal{S} is connected by the *Neigh* relation,
- (ii) $|\mathcal{S}| \geq 3$, and
- (iii) there exists $\theta \in [0, \pi[$ such that for every $A, B \in \mathcal{S}$, A is able to see B in direction θ or $\theta + \pi$ with respect to the horizontal axis.

The first condition ensures that the group is not “divided”, for instance the red objects in Fig. 6(a) do not satisfy this condition and can be considered as two groups. The second condition states that an aligned group should have at least 3 elements. To verify the third condition it would be necessary to compare all the orientation histograms between any two objects of \mathcal{S} . Unfortunately, this measure is very restrictive, and a more flexible measure is to consider that all the orientation histograms of $O(A_i, \mathcal{S} \setminus \{A_i\})$ are similar for all $A_i \in \mathcal{S}$. Figure 7 shows the dilated orientation histograms $D_{\nu_0}(O(A_i, \mathcal{S} \setminus \{A_i\}))$ and the dilated orientation histograms between the objects of the group in Fig. 6(b). For each orientation histogram we used the same structuring element ν_0 . We can notice that for the dilated histograms $D_{\nu_0}(O(A_i, \mathcal{S} \setminus \{A_i\}))$ it is possible to see a tendency towards a similar angle, while for the dilated orientation histograms $D_{\nu_0}(O(A_i, A_j))$ this is not the case. This is reflected when histograms are aggregated using the t-norm of Lukasiewicz in Fig. 7(c) where the aggregation of $D_{\nu_0}(O(A_i, \mathcal{S} \setminus \{A_i\}))$ results in a function with a maximum of 0.81, while the aggregation of $D_{\nu_0}(O(A_i, A_j))$ produces a constant function equal to zero, which is not meaningful here. One should notice that when using the conjunction of the dilated orientation histograms $D_{\nu_0}(O(A_i, A_j))$, if two pairs of objects do not have a similar orientation then the whole conjunction is equal to zero. However, when using the dilated histograms $D_{\nu_0}(O(A_i, \mathcal{S} \setminus \{A_i\}))$ the dissimilarity between the orientations of two pairs of objects will not affect the whole conjunction, since it is a comparison between the orientations of the whole group with respect to its members.

Thus, it is possible to define the degree of *global* alignment as follows:

Definition 3.2. Let $\mathcal{S} = \{A_0, \dots, A_N\}$, with $N \geq 3$, be a group of objects in \mathcal{I} , connected by the *Neigh* relation. Then, the degree of global alignment of \mathcal{S} is given by:

$$\mu_{ALIG}(\mathcal{S}) = \text{sim}(O(A_0, \mathcal{S} \setminus \{A_0\}), \dots, O(A_N, \mathcal{S} \setminus \{A_N\})). \quad (10)$$

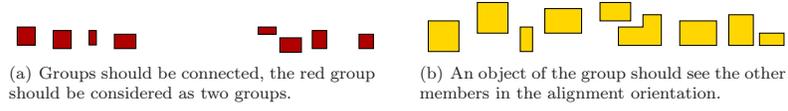


Figure 6: Considerations for an aligned group.

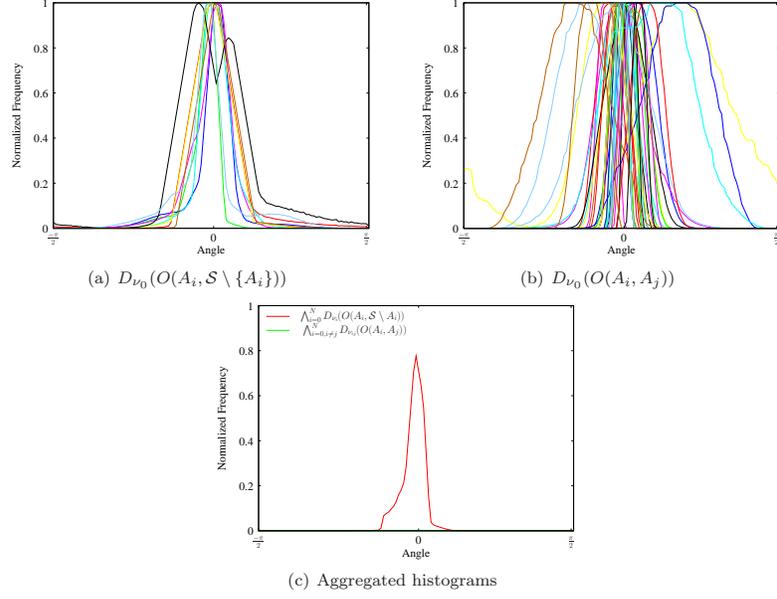


Figure 7: Dilated orientation histograms for objects of Fig. 6(b) and their aggregation using the t-norm of Lukasiewicz.

3.3.2. Locally aligned groups

We can say that a group \mathcal{S} is *locally aligned* if it satisfies:

- (i) for every $A \in \mathcal{S}$ the elements in the neighborhood $N(A)$ in \mathcal{S} are aligned, and
- (ii) it is connected by the *Neigh* relation.

For the first condition, we will only verify for simplicity that for every couple of elements $B, C \in \mathcal{S}$ belonging to $N(A)$, the orientations $O(A, B)$ and $O(A, C)$ are similar. Thus, we can define the degree of *locally alignment* as follows:

Definition 3.3. Let $\mathcal{S} = \{A_0, \dots, A_N\}$, with $N \geq 3$, be a group of objects in \mathfrak{J} , connected by the *Neigh* relation. The degree of locally alignment of \mathcal{S} is given by:

$$\mu_{LA}(\mathcal{S}) = \min_{X, Y, Z: Neigh(X, Y) \wedge Neigh(Y, Z)} sim(O(X, Y), O(Y, Z)). \quad (11)$$

We will say that a group of objects \mathcal{S} is *locally aligned* to a degree β if $\mu_{LA}(\mathcal{S}) \geq \beta$. The preceding definition can be summarized by saying that a group \mathcal{S} with $|\mathcal{S}| \geq 3$ is *locally*

aligned to a degree β if it satisfies the following relations:

$$R1 : \forall X, Y, Z \quad (Neigh(X, Y) \wedge Neigh(Y, Z)) \Rightarrow (sim(O(X, Y), O(Y, Z)) \geq \beta) \quad (12)$$

$$R2 : \forall A, B, \exists X_0, \dots, X_m \text{ such that } X_0 = A, X_m = B \text{ and } \left(\bigwedge_{i=0}^{m-1} Neigh(X_i, X_{i+1}) \right) \quad (13)$$

3.3.3. Other neighborhood choices

The choice $N_d(A)$ in Sec. 3.3.1 as the Voronoi neighborhood of A constrained by a distance d was motivated by the fact that in a group, any subgroup of aligned objects should be “consecutive” (see Fig. 8), and that successive objects should be close to each other. Nevertheless, there are other possibilities for choosing the neighborhood $N(A)$. For instance, we can consider the neighborhood of objects that are “near” A which can be defined by the fuzzy morphological dilation $D_\nu^d(A)$ of A , where ν^d is a fuzzy structuring element representing the notion of a distance “less than” d . For

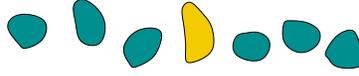


Figure 8: Illustration of a non-consecutive group. In a consecutive group there should be a succession of elements, which is not the case for the blue group, since the yellow object breaks the succession.

a fixed ν^d the neighborhood $N(A)$ is also univocally defined. This choice gives a more restrictive condition of *locally* alignment since the neighborhood contains more objects that should verify the alignment conditions, and it can also be used in the case where A and B are fuzzy objects. A fuzzy object is defined as a fuzzy set over the image space (usually \mathbb{Z}^2 for 2D digital images), and the membership function of the set represents the imprecision in the spatial extent of the object [1]. In the following of this subsection we assume that we are dealing with fuzzy objects A and B defined through their membership functions μ_A and μ_B , since all equations that will be presented are valid for crisp and fuzzy objects. When considering a fuzzy neighborhood, the *Neigh* relation becomes also fuzzy, and we denote by $\mu_{Neigh}(A, B)$ its degree of satisfaction. Similarly to the degree of adjacency in [1], the degree of μ_{Neigh} is defined as a conjunction of the degree of intersection between μ_A and $\mu_{N(B)}$, and μ_B and $\mu_{N(A)}$. Therefore, the degree of neighborhood is defined by:

$$\mu_{Neigh}(A, B) = \mu_{\neg int}(\mu_A, \mu_B) \wedge \mu_{int}(\mu_{N(A)}, \mu_B) \wedge \mu_{int}(\mu_{N(B)}, \mu_A) \quad (14)$$

where μ_{int} and $\mu_{\neg int}$ are a degree of intersection and non-intersection, respectively [1]. To have a symmetrical relation we consider the intersection between μ_A and $\mu_{N(B)}$, and between μ_B and $\mu_{N(A)}$. However, in the case where $D_\nu^d(A)$ with ν_d symmetric is used as neighborhood, $\mu_{int}(\mu_{N(A)}, \mu_B)$ and $\mu_{int}(\mu_{N(B)}, \mu_A)$ are equal, and therefore it is only necessary to verify one of the two conditions.

Consequently, the relation “connected by *Neigh*” also becomes a fuzzy relation denoted by μ_{conn} and the degree of connectedness in a group \mathcal{S} between two objects A, B in \mathcal{S} is defined as in [38] by:

$$\mu_{conn}(A, B) = \max_{p \in P_{ab}} \left[\min_{1 \leq i \leq l_p} \mu_{Neigh}(C_{i-1}^{(p)}, C_i^{(p)}) \right] \quad (15)$$

where p is a list of objects $\langle C_0^{(p)} = A, C_1^{(p)}, \dots, C_{l_p}^{(p)} = B \rangle$ in \mathcal{S} , called path, and P_{ab} is the set of all the paths from A to B in \mathcal{S} . The degree of connectedness of a group can be defined as the minimum degree of connectedness between its elements:

$$\mu_{conn}(\mathcal{S}) = \min_{A, B \in \mathcal{S}} \mu_{conn}(A, B) \quad (16)$$

When using a fuzzy neighborhood, the definitions of *globally* aligned and *locally* aligned have to be revised. The degree of *global* alignment of a group of objects $\mathcal{S} = \{A_0, \dots, A_N\}$ and with

$N \geq 3$ becomes:

$$\mu_{ALIGN}(\mathcal{S}) = \mu_{conn}(\mathcal{S}) \wedge sim(O(A_0, \mathcal{S} \setminus \{A_0\}), \dots, O(A_N, \mathcal{S} \setminus \{A_N\})) \quad (17)$$

This definition represents a conjunctive combination of the condition of being connected by the neighborhood relation and the similarity among the orientation histograms. When using a crisp neighborhood this definition is equivalent to the one given in Def. 3.1, where the condition of satisfying the relation of connection by the neighborhood is implicit in the definition.

In a similar way, it is possible to extend the definition of *locally* alignment to:

$$\mu_{LA}(\mathcal{S}) = \mu_{conn}(\mathcal{S}) \wedge \left[\min_{X,Y,Z} (sim(O(X,Y), O(Y,Z)) \wedge \mu_{conn}(\{X,Y,Z\})) \right]. \quad (18)$$

Again, the equation represents the conjunction of combining two conditions, the first one is the condition of being connected and the second one represents that objects X and Z belong to the neighborhood of Y , and that the orientation histograms $O(X,Y)$ and $O(Y,Z)$ should be similar. In the case where a crisp neighborhood is used, we obtain the same degree as in Def. 3.3.

3.3.4. Identification of locally aligned groups

In this section we explain how it is possible to extract the *locally* aligned subgroups from a group of objects. For clarity purposes we first use a crisp neighborhood, and in the next section we explain how the algorithm is extended to cope with fuzzy neighborhoods.

As discussed in Sec. 3.3.2, the notion of *local* alignment strongly depends on the notion of neighborhood, since an aligned group should be connected by the *Neigh* relation. Therefore, we propose to construct a neighborhood graph G_N to obtain the information of which objects are connected via the *Neigh* relation. In a neighborhood graph $G_N = (\mathcal{V}, E)$ the vertices represent the objects of the group, and there is an edge between two vertices if and only if the corresponding objects are neighbors. Notice that only the connected subsets of three vertices X, Y and Z in G_N which share a common vertex, for example Y , satisfy:

$$Neigh(X,Y) \wedge Neigh(Y,Z) \quad (19)$$

These connected subsets are called *triplets*. According to *R1*, only the *triplets* $\{X,Y,Z\}$ that satisfy (8):

$$sim(O(X,Y), O(Y,Z)) \geq \beta \quad (20)$$

are aligned and can belong to a group which is *locally* aligned to a degree β . *Triples* can be easily identified as the edges of the dual graph, when the dual graph is constructed in the following manner. The dual graph is denoted by $\tilde{G}_N = \{\tilde{\mathcal{V}}, \tilde{E}\}$ where each vertex \tilde{V}_i represents an edge in the graph G_N . An edge exists between two vertices \tilde{V}_i and \tilde{V}_j of \tilde{G}_N if the two corresponding edges of the graph G_N have a common vertex. If, additionally, we attribute to each edge (i,j) the similarity degree between the orientation histograms of \tilde{V}_i and \tilde{V}_j that we denote by \tilde{s}_{ij} , then it is possible to verify if relation *R1* holds for its corresponding *triplet*. Figure 9 shows an example of a neighborhood graph and its dual graph. Notice that the edges of \tilde{G}_N with a high value represent the *triples* of objects with a similar orientation histograms. For instance, in the dual graph the edge between the nodes (1 - 2) and (2 - 3) has a similarity value of 1, this edge corresponds to the objects labeled 1, 2 and 3 of Fig. 9(a). In a similar way, edges with a low value represent objects which are not aligned, for example in the dual graph the edge between the nodes (1 - 2) and (6 - 2) has a similarity value of 0.11 and corresponds to the objects labeled 1, 2 and 6, which do not form a *globally* aligned *triplet*.

Returning to the conditions of *local* alignment *R1* (12) and *R2* (13), the first one states that *triples* should be *globally* aligned, and the second one that the group should be formed by connected objects according to the *Neigh* relation. Then a group \mathcal{S} satisfies these relations if and

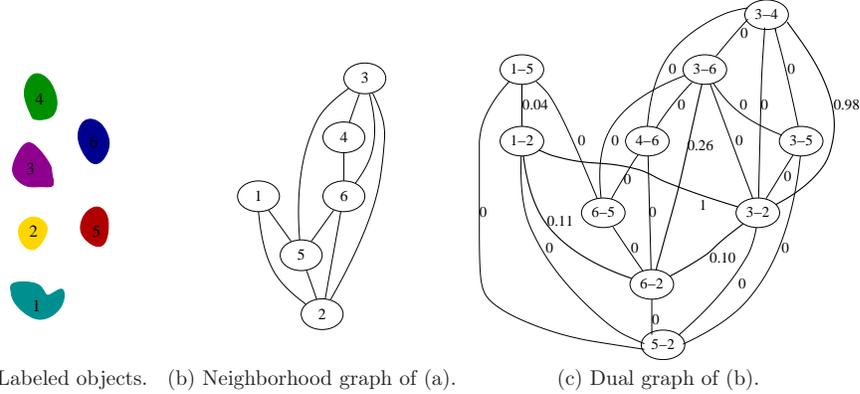


Figure 9: Neighborhood graph and dual graph of a group of objects.

only if the subset $\tilde{\mathcal{S}} \subseteq \tilde{\mathcal{V}}$ which represents the dual of \mathcal{S} satisfies the following relations:

$$\tilde{R}1 : \forall \tilde{V}_i, \tilde{V}_j \text{ Connected}(\tilde{V}_i, \tilde{V}_j) \Rightarrow (\tilde{s}_{ij} \geq \beta) \quad (21)$$

$$\begin{aligned} \tilde{R}2 : \forall \tilde{V}_i, \tilde{V}_j \exists \tilde{U}_0, \dots, \tilde{U}_K \text{ for } K \geq 1 \text{ such that } \tilde{U}_0 = \tilde{V}_i, \tilde{U}_K = \tilde{V}_j \\ \text{and } \bigwedge_{k=0}^{K-1} \text{Connected}(\tilde{U}_k, \tilde{U}_{k+1}), \end{aligned} \quad (22)$$

where $\text{Connected}(\tilde{U}, \tilde{V})$ is true if there exists an edge between \tilde{U} and \tilde{V} . Condition $\tilde{R}2$ expresses that $\tilde{\mathcal{S}}$ should be connected, since if $\tilde{\mathcal{S}}$ is not connected then \mathcal{S} is not connected. Therefore, a *locally* aligned group is a subset $\mathcal{S} \subseteq \mathcal{V}$ for which its dual set $\tilde{\mathcal{S}} \subseteq \tilde{\mathcal{V}}$ is connected in $\tilde{\mathcal{G}}$ and the value of all the edges joining the vertices within $\tilde{\mathcal{S}}$ is greater than or equal to β .

Algorithm 1 can be used to extract the $\tilde{\mathcal{S}}_i \subset \tilde{\mathcal{V}}$ corresponding to the dual sets of the *locally* aligned sets $\mathcal{S}_i \subset \mathcal{V}$. First the connected components of a graph $\tilde{\mathcal{G}}_{TH}$ are computed and stored in \mathfrak{C} . $\tilde{\mathcal{G}}_{TH}$ is a non-attributed graph containing the same vertices as $\tilde{\mathcal{G}}$ and there is an edge between two vertices if the edge in $\tilde{\mathcal{G}}$ has a degree greater than or equal to β . Then, for each component \mathcal{C}_k we obtain the minimum value of its edges in $\tilde{\mathcal{G}}$ that we call consistency degree of \mathcal{C}_k and is denoted by $\text{cons}(\mathcal{C}_k)$:

$$\text{cons}(\mathcal{C}_k) = \min\{\tilde{s}_{ij} | \tilde{V}_i, \tilde{V}_j \in \mathcal{C}_k\}$$

If $\text{cons}(\mathcal{C}_k) < \beta$ then \mathcal{C}_k does not satisfy $\tilde{R}1$, thus vertices are removed until $\text{cons}(\mathcal{C}_k) \geq \beta$. If in the process of vertex removal \mathcal{C}_k becomes disconnected, then each of the connected components of \mathcal{C}_k is treated separately. The vertices which are removed are the ones having more conflict with their neighbors in \mathcal{C}_k . We say that two connected vertices \tilde{V}_i and \tilde{V}_j are in conflict if \tilde{s}_{ij} is close to zero, that is if the corresponding orientation histograms of both vertices are not similar. The conflict of a vertex \tilde{V}_t with its neighbors in \mathcal{C}_k is measured by using what we call the degree of the vertex in \mathcal{C}_k :

$$\text{deg}(\tilde{V}_t) = \frac{\sum_{\tilde{V}_j \in \mathcal{C}_k} \tilde{s}_{tj}}{|\{(i, j) | \tilde{V}_j \in \mathcal{C}_k\}|}. \quad (23)$$

This degree represents the average edge value over all the edges connected to \tilde{V}_t . It is clear that if \tilde{V}_t is in conflict with several of its connected vertices in \mathcal{C}_k then $\text{deg}(\tilde{V}_t)$ will be close to 0, and it will be close to 1 if there is no conflict. Then the conflict of a vertex will be given by $1 - \text{deg}(\tilde{V}_t)$.

Fig. 10 shows an example where there is a conflict between the vertices of a connected component for $\beta = 0.8$. Fig. 10(b) shows the dual graph of the objects, and Fig. 10(c) shows the thresholded graph, where there is a connected component with three vertices that we denote by

C_0 . The consistency degree of C_0 is $cons(C_0) = 0.05$, which is inferior to β . The conflict of the nodes (1-2), (2-3) and (2-4) are 0.64, 0.49 and 0.08, respectively. Therefore, the nodes (1-2) and (2-3) have a conflict with their neighbors in C_0 . To reduce the conflict we remove the node (1-2) since it is the one having the higher conflict. By removing this edge the conflict is solved and the consistency degree of C_0 becomes $cons(C_0) = 0.89$, which is higher than β .

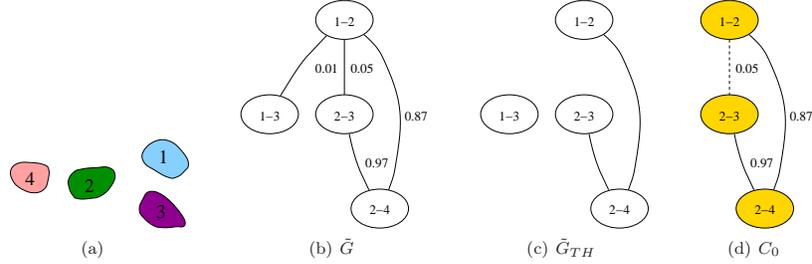


Figure 10: (b) Dual graph of objects in (a). (c) Thresholded dual graph for $\beta = 0.8$. (d) Vertices of the connected component of \tilde{G}_{TH} seen by \tilde{G} .

```

Input: Dual graph  $\tilde{G}$ ,  $\beta$ 
Output:  $\mathfrak{L}$ 
1 Create  $\tilde{G}_{TH} = (\tilde{V}, \tilde{E}_{TH})$  where  $E_{TH} = \{(i, j) \in E | \tilde{e}_{ij} \geq \beta\}$ ;
2 Find  $\mathfrak{C}$  the set of connected components of  $\tilde{G}_{TH}$ ;
3 foreach  $C_k \in \mathfrak{C}$  do
4   Let  $cons = \min\{\tilde{s}_{ij} | \tilde{V}_i, \tilde{V}_j \in C_k\}$ ;
5   while  $cons < \beta$  and  $|C_k| \geq 2$  do
6     foreach  $\tilde{V}_i \in C_k$  do
7        $deg_i = \frac{\sum_{\tilde{V}_j \in C_k} \tilde{s}_{ij}}{| \{(i, j) | \tilde{V}_j \in C_k \} |}$ ;
8     end
9     Delete from  $C_k$  the  $\tilde{V}_j$  for which  $\tilde{V}_j = \min_{\tilde{V}_i \in C_k} deg_i$ ;
10    if  $C_k$  is disconnected in  $\tilde{G}_{TH}$  then
11      Let  $\mathfrak{D} = \{\mathcal{D}_0, \dots, \mathcal{D}_L\}$  the connected components of  $C_k$ ;
12       $C_k = \mathcal{D}_0$ ;
13      for  $l = 1$  to  $L$  do
14        | Add  $\mathcal{D}_l$  to  $\mathfrak{C}$ ;
15      end
16      Update  $cons = \min\{\tilde{s}_{ij} | \tilde{V}_i, \tilde{V}_j \in C_k\}$ ;
17    end
18  end
19  if  $cons \geq \beta$  then
20    | Add  $C_k$  to  $\mathfrak{L}$ ;
21  end
22 end

```

Algorithm 1: Algorithm for finding locally aligned groups \mathfrak{L} from \tilde{G} .

3.3.5. Extension for fuzzy neighborhoods

If instead of having a crisp *Neigh* relation we have a fuzzy relation μ_{Neigh} , the procedure should be adapted for the extraction of the *locally* aligned groups.

When constructing the neighborhood graph G_N each edge should be attributed with the degree of satisfaction of μ_{Neigh} . The notion of *triplets* also becomes fuzzy, and the degree to which three vertices X, Y and Z form a *triplet* is given by the degree of connectedness of $\{X, Y, Z\}$ using Eq. (16).

The degree of connectedness is taken into account in the construction of the dual graph \tilde{G}_N , and only the *triplets* with a connectedness value greater than the user defined acceptance value β will be considered. Each edge (i, j) between the vertices \tilde{V}_i and \tilde{V}_j of the dual graph will be attributed with the degree:

$$\tilde{s}_{ij} = sim(O(X, Y), O(Y, Z)) \wedge \mu_{conn}(\{X, Y, Z\}), \quad (24)$$

where X, Y and Z build the vertices *triplet* represented by the vertices \tilde{V}_i and \tilde{V}_j .

The extraction of *locally* aligned groups is performed by applying Algorithm 1 on the dual graph. Let \mathcal{S} be a resulting group. Due to the choice of construction of the dual graph, we can guarantee that, for every pair of elements $A, B \in \mathcal{S}$ the degree $\mu_{Neigh}(A, B)$ is greater than β , since only the edges satisfying this condition were used for the construction of the dual graph. Therefore, the degree of connectedness of the group is $\mu_{conn}(\mathcal{S}) \geq \beta$. It is straightforward to see that \mathcal{S} satisfies the second condition of Eq. (18) since this condition is imposed by the choice of \tilde{s}_{ij} . Hence, the resulting groups are *locally* aligned to a degree greater than β according to (18).

3.3.6. Candidates for globally aligned groups

The *locally* aligned groups \mathfrak{L} to a degree β are the possible candidates for being *globally* aligned groups to a degree α , for $\alpha \leq \beta$. The evaluation is performed by measuring the degree of *global* alignment using Eq. (11). Usually the *locally* aligned groups are *globally* aligned. However there are cases as the one shown in Fig. 11 where a *locally* aligned group is not *globally* aligned.

To increase the degree of *global* alignment of a group \mathcal{S} we divide the group by eliminating the vertices in $\tilde{\mathcal{S}}$ with the minimum vertex degree (Eq. 23) in $\tilde{\mathcal{S}}$, we repeat this step until $\mu_{ALIGN}(\mathcal{S}) \geq \alpha$. If the degree of all vertices in $\tilde{\mathcal{S}}$ is equal to one, and $\mu_{ALIGN}(\mathcal{S}) < \alpha$ it means that a lot of imprecision was introduced for the similarity computation, and the measurement of similarity is very permissive, thus the whole process should be repeated using a ν_0 with a tighter support in (9).

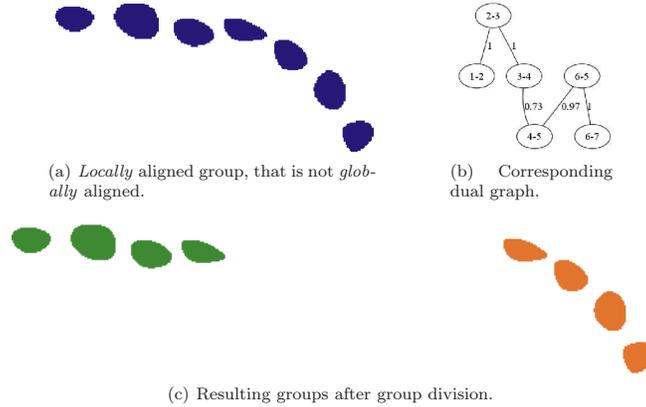


Figure 11: (a) *Locally* aligned group which is not *globally* aligned. (b) Its dual graph, where the objects are labeled from 1 to 7 (left to right). (c) Resulting groups obtained after solving the conflict.

3.3.7. Adding more elements to the group

Once the *globally* aligned groups of objects are identified it is possible to add new objects to the group or fuse two *globally* aligned groups to obtain a larger *globally* aligned group. For each group \mathcal{S}_i we perform two morphological directional dilations of the group in the directions θ and $\theta + \pi$, where θ is the orientation of the alignment (the angle which maximizes the conjunction of the orientation histograms $O(A_i, \mathcal{S} \setminus \{A_i\})$). These dilations will be denoted by $D_{\nu_\theta}(\mathcal{S}_i)$ and $D_{\nu_{\theta+\pi}}(\mathcal{S}_i)$. The directional dilation of a fuzzy set μ in a direction \vec{u}_θ is defined as [36]:

$$D_{\nu_\theta}(\mu)(x) = \sup_y \min[\mu(y), \nu_\theta(x - y)] , \quad (25)$$

where ν_θ is a fuzzy directional structuring element chosen so as to have high membership values in the direction \vec{u}_θ and its value at a point $x = (r, \alpha)$ (in polar coordinates) is a decreasing function of $|\theta - \alpha|$ modulo 2π . The fuzzy sets $D_{\nu_\theta}(\mathcal{S}_i)$ and $D_{\nu_{\theta+\pi}}(\mathcal{S}_i)$ represent the regions of space that are in direction θ and $\theta + \pi$ of \mathcal{S}_i . An object A which satisfies the *Neigh* relation with one of the members of \mathcal{S}_i and which is included in $D_{\nu_\theta}(\mathcal{S}_i)$ or $D_{\nu_{\theta+\pi}}(\mathcal{S}_i)$ with a degree greater than or equal to β (that is $\mu_{include}(A, D_{\nu_\theta}(\mathcal{S}_i) \cup D_{\nu_{\theta+\pi}}(\mathcal{S}_i)) \geq \beta$, where $\mu_{include}$ denotes a degree of inclusion [1]) is added to \mathcal{S}_i , since is in the same direction as the orientation alignment and it is connected to the group. If a whole group \mathcal{S}_j is included in $D_{\nu_\theta}(\mathcal{S}_i)$ or $D_{\nu_{\theta+\pi}}(\mathcal{S}_i)$ with a degree greater than β and one of the elements of \mathcal{S}_i is connected to one of the members \mathcal{S}_j and both groups have a similar orientation, then both groups are fused into one.



Figure 12: (a) Labeled image. (b) Locally aligned group. (c) The region seen by the group of (b) in the direction of the alignment (white = high value of visibility). (d) Group obtained after adding new elements.

Figure 12 shows an example of extracting the *locally* aligned groups of Fig. 12(a). The resulting group is *locally* aligned to a degree 0.9, and it is *globally* aligned to a degree 0.85. This group is extended to add more elements to the group resulting in a larger group with a degree of *global* alignment of 0.8.

3.4. Stability with respect to segmentation errors

One interesting feature of our approach is that it is robust to the quality of the segmentation of the objects. This property is particularly important in real applications where it is difficult to guarantee that all objects have been segmented and that the segmentation is accurate. Figure 13 shows two examples of the stability of the algorithm with respect to segmentation errors such as the absence of an object, or the merging of an undesired region to one of the objects. Figure 13(b) shows a segmentation of the houses of Fig. 13(a), which is almost correct, except for some false detections and a missing house. Two of the aligned groups of objects extracted from this segmentation are shown in Fig.13(c). Actually more groups are extracted by the algorithm but for the sake of clarity only a few groups are shown. Figure 13(d) shows another segmentation which was manually modified to introduce errors: two missing houses, and one of the houses is merged with other regions. Figure 13(e) shows two of the retrieved aligned groups, which correspond to the groups found in Fig.13(c). We can see that the blue group in Fig. 13(e) is retrieved even with the absence of two objects, and that the orange group is retrieved although the center of mass of one of its members has been displaced.

For this example we used a Voronoi neighborhood constrained by a distance d , where d was larger than the separation between the objects. However, if we had used a smaller distance,

then the algorithm would not have retrieved the purple group, since there would have been a disconnection between the upper and the lower part of the purple group of Fig. 13(e), and therefore only the lower part of the group would have been retrieved. This is an expected behavior.

In both experiments we used the same parameters for the extraction of the *locally* aligned groups, which are $\beta = 0.8$ and $d = 50$ pixels. The degrees of *global* alignment in both cases were very similar, for the purple groups the degree was $\mu_{ALIGN} = 1.0$ in the two cases, and for the orange groups it was $\mu_{ALIGN} = 1.0$ in Fig. 13(c), and $\mu_{ALIGN} = 0.94$ in Fig. 13(e). The degrees of *global* alignment remained almost the same, since the missing objects and the modification of the object are local changes which do not affect the *global* orientation of the groups. Nevertheless if one of the objects is severely modified as in Fig. 13(f) then it is not possible to retrieve the same aligned groups as before, which is again an expected behavior.

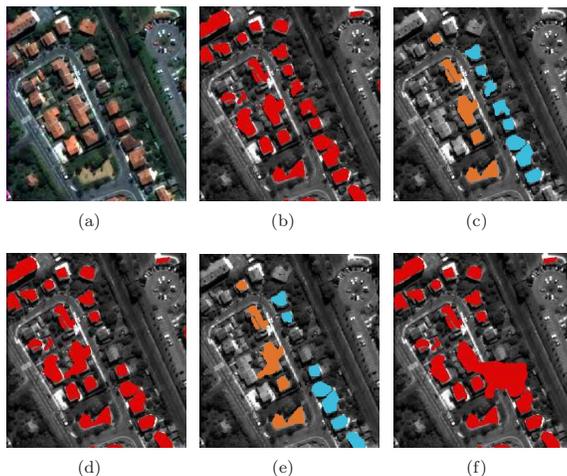


Figure 13: (a) Original image. (b) Segmented houses (red). (c) Some of the extracted *globally* aligned groups from objects of (b). (d) Segmented houses with errors of missing objects and some merged regions. (e) Some of the extracted *globally* aligned groups from objects of (d). (f) Segmented houses with errors that do not allow the recovery of the *globally* aligned groups.

3.5. Complexity analysis

In this section we deal with the cost of the basic operations of the algorithm for extracting *locally* aligned groups and *globally* aligned groups.

First, we consider the complexity of extracting *locally* aligned groups. Consider we have N objects each with at most n_o points. The complexity of the algorithm is $O(N^2)$ since most of steps of the algorithm deal with operations over the graph or its dual. It should be noticed that the step which corresponds to the construction of the orientation histograms has a complexity of $O(N^2 n_o^2)$, since at maximum there are $N(N-1)$ edges in the graph and for each edge an orientation histogram is constructed with a complexity of $O(n_o^2)$.

The complexity of finding a *globally* aligned group from a *locally* aligned group with N_A elements each having at most n_o points lies in the following steps. The first step consists in evaluating the degree of *global* alignment and division of the group in the case where it is not aligned, and this step has a complexity of $O(N_A^2 n_o^2)$. The second step consists in performing the morphological directional dilations of the group in the directions of alignment θ and $\theta + \pi$, and has a complexity of $O(N_I)$ [39], where N_I is the number of points in the image (see [39] for the implementation of the directional morphological dilation using a propagation method). And finally, the complexity

of the step of evaluating the degree of inclusion of each object not belonging to the group into the directional dilations of the group is $O((N - N_A)n_o^2)$, where N is the total number of objects. Hence, summing the three steps we obtain that the total complexity is $O(N_A^2 n_o^2 + N_I)$.

3.6. Discussion

In this part we have introduced the definitions of *globally* aligned groups and *locally* aligned groups of objects, and gave a method to extract alignments from an image of labeled objects. Both definitions are appropriate to determine alignments of objects of different sizes. Therefore, the extraction method can be used to find alignments of objects of the same type or class, for instance buildings in a urban scene, since all the objects do not necessarily have the same size. However, not all the obtained groups are meaningful for the description of the scene, since the subsets which are found only satisfy the conditions of alignment. For example Fig.14 shows two *globally* aligned subsets of airplanes extracted using the proposed algorithm, the group of Fig.14(c) is *globally* aligned but does not give any information about the arrangement of the airplanes, while the group of Fig.14(d) gives us more information about the arrangement of the airplanes. Hence it is necessary to use additional information to put the aligned groups into context, for example find whether the alignments are parallel between them or parallel to a linear structure. In the case of Fig.14 it would be interesting to find the alignment parallel to the buildings. This point will be further discussed in the next section.

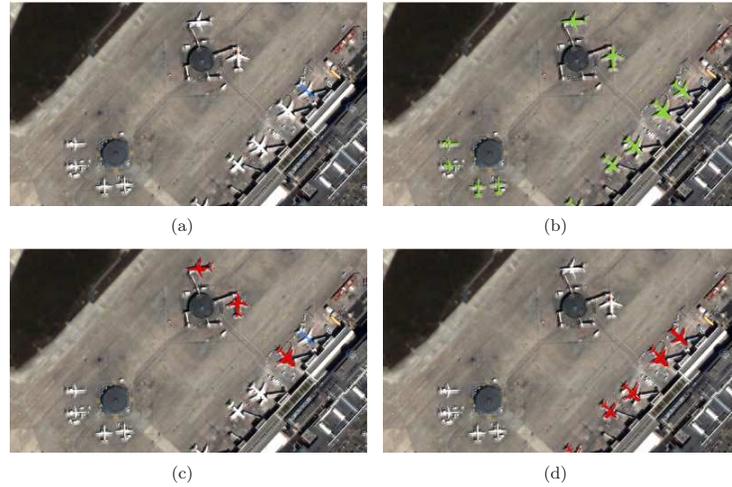


Figure 14: (a) Airport image. (b) Segmented airplanes (green). (c) Extracted *globally* aligned group in red with degree 0.97. (d) Extracted *globally* aligned group in red with degree 0.99.

The proposed method for alignment extraction is very flexible. One should notice that it is possible to incorporate more information according to the type of alignment. For instance, if we are searching for an alignment where the objects of the alignment have the same orientation as the alignment, this could be incorporated in the weight \tilde{s}_{ij} attributed to the edges of the dual graph. The new weight \tilde{s}_{ij} of the edge between the vertices \tilde{V}_i and \tilde{V}_j , which represents a *triplet* $\{X, Y, Z\}$, is given by a conjunction combining the condition that $O(X, Y)$ and $O(Y, Z)$ should be similar, and the condition that every member of the *triplet* should have a similar orientation to the orientation histograms between itself and the other members of the triplet which are connected to it. For instance, the condition for Y is expressed as:

$$\text{sim}(O(X, Y), \delta_{\theta_Y}) \wedge \text{sim}(O(Z, Y), \delta_{\theta_Y}),$$

where δ_{θ_Y} is the Dirac function at the angle θ_Y which represents the orientation of Y . For X and Z the condition is verified by observing the degree of $\text{sim}(O(X, Y), \delta_{\theta_X})$ and $\text{sim}(O(Z, Y), \delta_{\theta_Z})$, respectively. Since X and Z are connected to Y in the *triplet*, and the only histograms that involve them are $O(X, Y)$ and $O(Z, Y)$, respectively. Finally, when combining all the conditions the weight \tilde{s}_{ij} is given by:

$$\tilde{s}_{ij} = \text{sim}(O(X, Y), O(Y, Z)) \wedge [\min[\text{sim}(O(X, Y), \delta_{\theta_Y}) \wedge \text{sim}(O(Z, Y), \delta_{\theta_Y}), \text{sim}(O(X, Y), \delta_{\theta_X}), \text{sim}(O(Z, Y), \delta_{\theta_Z})]] \quad (26)$$

In this equation we use the min to ensure that the second condition is satisfied by all the members of the *triplet*.

4. Parallelism

For linear objects to be parallel, we expect a constant distance between them, or that they have the same normal vectors and the same orientation. Although classical parallelism in Euclidean geometry is a symmetric and transitive relation, these properties are subject to discussion when dealing with linear objects of finite length. When objects have different extensions as in Fig. 15(a), where B can be a house, and A a road, the symmetry becomes questionable. The statement “ B is parallel to A ” can be considered as true, since from every point on the boundary of B that faces A it is possible to see (in the normal direction to A ’s principal axis) a point of A , and the orientations of A and B are similar. On the other hand, the way we perceive “ B parallel to A ” will change depending on our position: from point d it is possible to see a point of B in the normal direction of B , while this is not possible from point c . In both cases (symmetrical and non symmetrical ones) the transitivity is lost. For example, in Fig. 15(b) and 15(c) the statements “ A is parallel to B ” and “ B is parallel to C ” hold, but “ A is not parallel to C ” since it is not possible to see C from A in the normal direction to C . This example also illustrates the interest of considering the degree of satisfaction of the relation instead of a crisp answer (yes/no). Then the relation “ B parallel to A ” will have a higher degree than “ A parallel to B ” in Fig. 15(a).

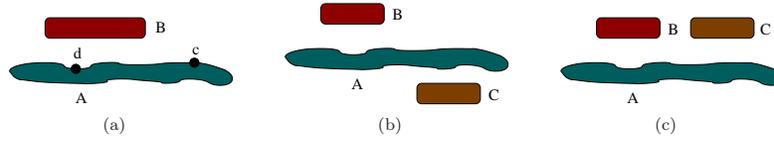


Figure 15: Examples where parallelism should preferably be considered as a matter of degree, and should not be necessarily symmetrical and transitive.

The parallel relation can also be considered between a group of objects $\mathcal{A} = \{A_i\}$ and an object B , typically when the objects in the group are *globally* aligned and B is elongated. For example a group of boats and a deck in a port. When evaluating the relation “ A is parallel to B ”, actually we are evaluating whether the whole set \mathcal{A} and the boundary of B that faces A have a similar orientation, and whether there is a large proportion of $\cup_i A_i$ that sees B in the normal direction to the group. Similar considerations can be derived when considering the relation “ B is parallel to \mathcal{A} ” or the relation between two groups of objects. All these considerations form the basis for the formal models provided in this section.

For “ A to be parallel to B ” it is only necessary that A is a linear object, while B can be a non linear object, and in this case A would be parallel to the boundary of B which is facing A . The same idea is also applicable for the parallelism between a *globally* aligned group of objects parallel to an object (see Fig. 16).

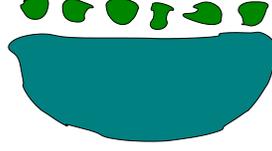


Figure 16: A group of *globally* aligned objects parallel to a non-linear object.

4.1. Parallelism with (fuzzy) linear objects

In this section we propose a definition of parallelism between a fuzzy linear object³ and a fuzzy object, including the particular case of crisp linear segments, and taking into account the above mentioned considerations. Suppose A is a linear object and B an object that it is not necessarily linear, let θ_A be the orientation of A and $\vec{u}_{\theta_A + \frac{\pi}{2}}$ be the normal unit vector to the principal axis of A . Then, according to the considerations of the previous section, the degree of satisfaction of the relation “ A is parallel to B ” depends on two conditions:

- (i) There should be a large proportion of A that sees B in the direction $\vec{u}_{\theta_A + \frac{\pi}{2}}$.
- (ii) The orientation of A and the orientation of the boundary of B that is facing A and that is seen by A in the direction $\vec{u}_{\theta_A + \frac{\pi}{2}}$ should be similar.

Both conditions deal with the notion of visibility. Let p be a point, X be a fuzzy object with membership function μ_X and \vec{u}_θ a vector with angle θ with respect to the x-axis. Then the subset of X that is seen by p in the direction \vec{u}_θ , that we denote by $X_{vis(p,\theta)}$, is equal to the intersection of X with the visual field of p , when p observes in the direction \vec{u}_θ . The visual field is represented as a morphological directional dilation in the direction θ of p . The set $X_{vis(p,\theta)}$ is a fuzzy set with membership function $\mu_{X_{vis(p,\theta)}}$, where $\mu_{X_{vis(p,\theta)}}(x)$ represents the degree to which x is in X and is seen by p in the direction \vec{u}_θ :

$$\mu_{X_{vis(p,\theta)}}(x) = \mu_X(x) \wedge D_{\nu_\theta}(p)(x), \quad (27)$$

where $D_{\nu_\theta}(p)(x)$ is the morphological directional dilation defined in Eq. (25).

Let Y be a fuzzy object with membership function μ_Y not intersecting X . We denote by $X_{vis(Y,\theta)}$ the subset of X that is seen by the points on the boundary of Y , i.e the subset of X that is seen by Y , and it is defined by:

$$\mu_{X_{vis(Y,\theta)}}(x) = \mu_X(x) \wedge D_{\nu_\theta}(\mu_Y)(x). \quad (28)$$

Fig.17 shows two objects A and B , and $B_{vis(A,\theta_A + \frac{\pi}{2})}$, where θ_A is the orientation of A . When A and B are linear segments, $A_{vis(B,\theta_A + \frac{\pi}{2})}$ can be interpreted as the projection of B onto A .

For the first condition of parallelism, we are interested in the proportion of A that sees B in the direction $\vec{u}_{\theta_A + \frac{\pi}{2}}$. This subset is equivalent to $A_{vis(B,\theta_A - \frac{\pi}{2})}$, since the degree to which a point $x \in A$ sees B in the direction $\vec{u}_{\theta_A + \frac{\pi}{2}}$ is equivalent to the degree to which the point is seen by B in the direction $\vec{u}_{\theta_A - \frac{\pi}{2}}$. Therefore, the proportion of A that sees B in the normal direction $\vec{u}_{\theta_A + \frac{\pi}{2}}$ is equal to the fuzzy hypervolume of $A_{vis(B,\theta_A - \frac{\pi}{2})}$ over the fuzzy hypervolume of A , where the fuzzy hypervolume V_n of a fuzzy set μ is given by: $V_n(\mu) = \sum_{x \in \mathcal{J}} \mu(x)$ [1]. Hence, the proportion is equal to:

$$\frac{V_n(\mu_{A_{vis(B,\theta_A - \frac{\pi}{2})}})}{V_n(\mu_A)}$$

³We consider that an object is linear if the ratio of its principal axis given by $\frac{c_{yy} + c_{xx} - \sqrt{(c_{xx} + c_{yy})^2 - 4(c_{xx}c_{yy} - c_{xy}^2)}}{c_{yy} + c_{xx} + \sqrt{(c_{xx} + c_{yy})^2 - 4(c_{xx}c_{yy} - c_{xy}^2)}}$ where $C = \begin{pmatrix} c_{xx} & c_{xy} \\ c_{yx} & c_{yy} \end{pmatrix}$ corresponds to the second moments matrix, is high [40].

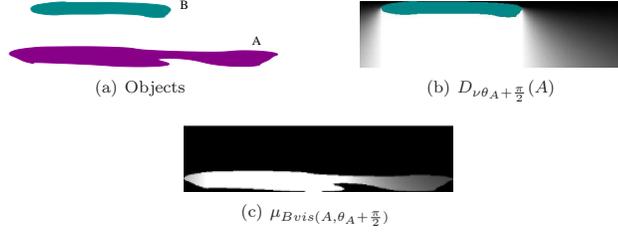


Figure 17: Illustration of the notion of visibility for the objects A and B of (a). (b) Visual field of A in the direction of $\theta_A + \frac{\pi}{2}$ (the white pixels have a high membership value of being observed by A in the direction θ_A). (c) membership function of $B_{vis(A, \theta_A + \frac{\pi}{2})}$, the white pixels are the points which have a high membership value.

For the second condition we are interested only in the subset of the boundary of B that faces A and that is seen by the boundary of A in the direction θ_A . The boundary of B which faces an object A corresponds to the points on the boundary of B that delimit the region between A and B , and are defined as the extremities of the admissible segments [41]. These are the points $b \in B$ for which there exists a point $a \in A$ such that the segment $]a, b[$ is included in $A^C \cap B^C$. In the case where A and B are fuzzy objects, we will be interested in the points which are the extremities of a segment with a high degree of admissibility [41]. Therefore, the subset of the boundary of B that faces A and that is seen by the boundary of A in the direction $\theta_A + \frac{\pi}{2}$ is a fuzzy subset where the membership of a point $x \in \mathcal{I}$ is equal to the conjunction between its membership to B , the degree of being the extremity of an admissible segment and the degree of being seen by A . We denote this subset by $\delta B_{vis(A, \theta_A + \frac{\pi}{2})}$ and its membership function by $\mu_{\delta B_{vis(A, \theta_A + \frac{\pi}{2})}}$:

$$\mu_{\delta B_{vis(A, \theta_A + \frac{\pi}{2})}}(x) = \mu_{adm}(x) \wedge \mu_{B_{vis(A, \theta_A + \frac{\pi}{2})}}(x) \quad (29)$$

where μ_{adm} represents the degree of being the extremity of an admissible segment.

Definition 4.1. The relation “ A is parallel to B ” is given by the following measure:

$$\mu_{\parallel}(A, B) = \frac{V_n(\mu_{A_{vis(B, \theta_A - \frac{\pi}{2})}})}{V_n(\mu_A)} \bigwedge \nu_0(\theta_{\delta B_{vis(A, \theta_A + \frac{\pi}{2})}} - \theta_A), \quad (30)$$

where $\nu_0(\theta)$ is the same as in Eq. (5) and it evaluates the degree to which $\theta_{\delta B_{vis(A, \theta_A + \frac{\pi}{2})}}$, the normal angle to $\delta B_{vis(A, \theta_A + \frac{\pi}{2})}$, and θ_A are “approximately” equal.

In some contexts a symmetrical relation is needed (for example in perceptual organization), and is then expressed as “ A and B are parallel”. In such cases, we verify at that least one of the sets is visible from the other in the normal direction and that the orientations of both sets are similar, leading to the following definition.

Definition 4.2. The degree of satisfaction of the symmetrical relation, “ A and B are parallel” is expressed by:

$$\mu_{\parallel S}(A, B) = \left[\frac{V_n(\mu_{A_{vis(B, \theta_A - \frac{\pi}{2})}})}{V_n(\mu_A)} \bigvee \frac{V_n(\mu_{B_{vis(A, \theta_B - \frac{\pi}{2})}})}{V_n(\mu_B)} \right] \bigwedge \nu_0(\theta_{\delta B_{vis(A, \theta_A + \frac{\pi}{2})}} - \theta_A) \bigwedge \nu_0(\theta_{\delta A_{vis(B, \theta_B + \frac{\pi}{2})}} - \theta_B) \quad (31)$$

Proposition 4.1. Both relations (Definitions 4.1 and 4.2) are invariant with respect to geometric transformations (translation, rotation, scaling).

None of the relations is transitive, as discussed previously. But we have the following partial result in the crisp case:

Proposition 4.2. *Let A, B, C be linear crisp segments, if $\mu_{\parallel}(A, B) = 1$, $\mu_{\parallel}(B, C) = 1$ and $\theta_A = \theta_B = \theta_C$, then $\mu_{\parallel}(A, C) = 1$.*

This result shows that in the crisp case we have transitivity. To have the transitivity property, it is necessary that $\theta_A = \theta_B = \theta_C$, since $\nu_0(\theta_A - \theta_B) = 1$ and $\nu_0(\theta_B - \theta_C) = 1$ do not imply $\nu_0(\theta_A - \theta_C) = 1$ due to the tolerance value t_1 of the function ν_0 (see Eq. 5). To have the transitivity without imposing the condition $\theta_A = \theta_B = \theta_C$, it is necessary that ν_0 is a linear function (i.e. $t_1 = 0$). But this is restrictive.

It is clear that both relations are reflexive. However, depending on the context we may not want to consider intersecting objects as parallel. In this case, it is necessary to combine in a conjunctive way the previous degree (Def. 4.1 or 4.2) with a degree of non-intersection between the two sets.

4.2. Parallelism with a globally aligned group of objects

When considering parallelism with a *globally* aligned group of objects, the group has a similar role as the linear object in the definitions introduced in the previous section. When defining the relation with a group there is a modification, with respect to the case of a linear object, in the way the visibility constraint is computed, this modification will be discussed in the following.

4.2.1. A group of globally aligned objects parallel to an object

Let $\mathcal{S} = \{A_0, \dots, A_N\}$ be a group of globally aligned objects, as defined in Sec. 3.3.1, and let B be another object. For \mathcal{S} to be parallel to B it is necessary that there is a large portion of \mathcal{S} that sees B , and this is computed in the same way as for the case of parallelism between a linear object and an object. For the second condition we need to create the fuzzy set $\beta_{\mathcal{S}}$ which is composed of the union of the regions between two consecutive elements of \mathcal{S} . $\beta_{\mathcal{S}}$ can be constructed using the definition that involves the convex hull presented in [41]. In Fig. 18(b) an example of the region $\beta_{\mathcal{S}}$ of a group \mathcal{S} is shown in light purple. From Fig. 18(a) we can see that the boundary of B that faces \mathcal{S} and that is visible by \mathcal{S} depends on the separation between the members of the group. However, it is desirable that the degree of parallelism of a group to an object is independent of the separation of its members, since if we add more elements to a group without changing its orientation or its extension the degree of alignment to the object should remain the same. Therefore, in order to have a degree of parallelism independent of the separation of the members of the group we should use $\beta_{\mathcal{S}}$ in the second condition of parallelism. Then, the second condition becomes that the boundary of B that faces \mathcal{S} and that is visible by $\beta_{\mathcal{S}}$ or by \mathcal{S} should have the same orientation as the orientation of the alignment of \mathcal{S} .

Finally, we define the relation as:

Definition 4.3. *The degree of satisfaction of the relation “ \mathcal{S} is parallel to B ” is given by:*

$$\mu_{\parallel}(\mathcal{S}, B) = \frac{V_n(\bigvee_i \mu_{A_i \text{vis}(B, \theta_s - \frac{\pi}{2})})}{V_n(\bigvee \mu_{A_i})} \bigwedge \nu_0(\theta_{\delta B \text{vis}(\beta_{\mathcal{S}}, \theta_s + \frac{\pi}{2})} - \theta_{\mathcal{S}}), \quad (32)$$

where θ_s is the orientation of the alignment of the group \mathcal{S} .

In this definition, the first part of the equation represents that there should be a large portion of the union of all the $A_i \in \mathcal{S}$ that see B , and the second part evaluates the degree of similarity between the orientation of the group and the orientation of the boundary of the object seen by the group in the direction $\theta_s + \frac{\pi}{2}$.

4.2.2. A linear object parallel to a globally aligned group of objects

Using the same notations as above, suppose B is a fuzzy linear object. Then for “ B is parallel to \mathcal{S} ” to be true, it is necessary that B has a similar orientation to the orientation of the alignment of \mathcal{S} , and that there is a large proportion of B that sees the group of objects or $\beta_{\mathcal{S}}$. As in Def. 4.3 it is necessary to use $\mathcal{S} \cup \beta_{\mathcal{S}}$ in order to assure that the parallel relation is independent of the separation of the element of \mathcal{S} .

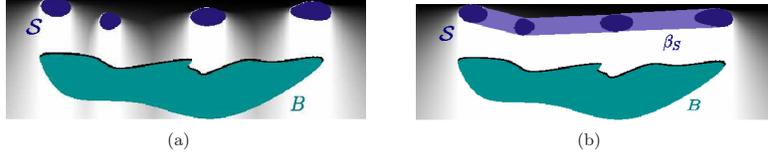


Figure 18: (a) Visibility field of the group of objects S , and the boundary of the object facing S (in black). (b) Visibility field of the region between two consecutive members of the group β_S (light purple) and the group S , and the boundary of B facing $S \cup \beta_S$ (in black).

Definition 4.4. The degree of satisfaction of the relation “ B is parallel to S ” is given by:

$$\mu_{\parallel}(B, S) = \frac{V_n(\bigvee_i \mu_{B \text{vis}(S \cup \beta_S, \theta_B - \frac{\pi}{2})})}{V_n(\bigvee \mu_B)} \bigwedge \nu_0(\theta_S - \theta_B), \quad (33)$$

where θ_B is the orientation of B .

4.2.3. Parallelism between two globally aligned groups of objects

Using the same notation as in Def. 4.4, we can define the parallelism between two *globally* aligned of fuzzy sets $\mathcal{S} = \{A_0, \dots, A_N\}$ and $\mathcal{T} = \{B_0, \dots, B_M\}$:

Definition 4.5. The degree of satisfaction of the relation “ S is parallel to T ” is given by:

$$\mu_{\parallel}(\mathcal{S}, \mathcal{T}) = \frac{V_n(\bigvee_i \mu_{A_i \text{vis}(\mathcal{T} \cup \beta_T, \theta_s - \frac{\pi}{2})})}{V_n(\bigvee \mu_{A_i})} \bigwedge \nu_0(\theta_T - \theta_S), \quad (34)$$

where β_T is the region formed between two consecutive elements of \mathcal{T} and θ_T is the orientation of alignment of the group \mathcal{T} .

4.3. Discussion

In this section we discussed the considerations that should be taken into account when modeling the parallel relation. We highlighted that the parallel relation should be modeled as a fuzzy relation represented as the conjunction of two conditions, one dealing with visibility and the other with similarity of orientation. Using the directional morphological dilation to model the visibility condition allows us to identify the region on the image where it is possible to find objects to which the object of interest or groups of objects are parallel to.

5. Illustrative examples

In this section we present two examples to illustrate the usefulness of the defined relations. The first example deals with urban patterns. By using the relations of *globally* alignment and parallelism between *globally* aligned groups we are able to determine the residential areas composed of organized houses. The method for determining the *globally* aligned groups of objects was applied on the segmented buildings of Fig. 19. The buildings were obtained by using the method described in [42]. For the extraction we used a $\beta = 0.85$ and a Voronoi neighborhood constrained by a distance of 30 pixels equivalent to approximately 21 m, since we are interested in residential area, in which houses are usually close to each other. Some of the *globally* aligned groups of houses are shown in Fig. 19(c). It is not possible to show all the aligned groups found by the algorithm since there are objects which belong to more than one group. We can see that the method obtains groups of *globally* aligned objects of different sizes which do not have an aligned barycenter. The obtained groups contain few elements due to the small neighborhood used to extract them. From the obtained *globally* aligned groups of houses we extracted the groups which

are parallel to another group or which have a group parallel to it and is close to another group to a degree greater than or equal to 0.8. The spatial relation “close to” was modeled in the same way as in [43]. The groups of houses satisfying the previous condition of parallelism are shown in Fig. 19(d). The regions containing organized groups of houses are well detected.

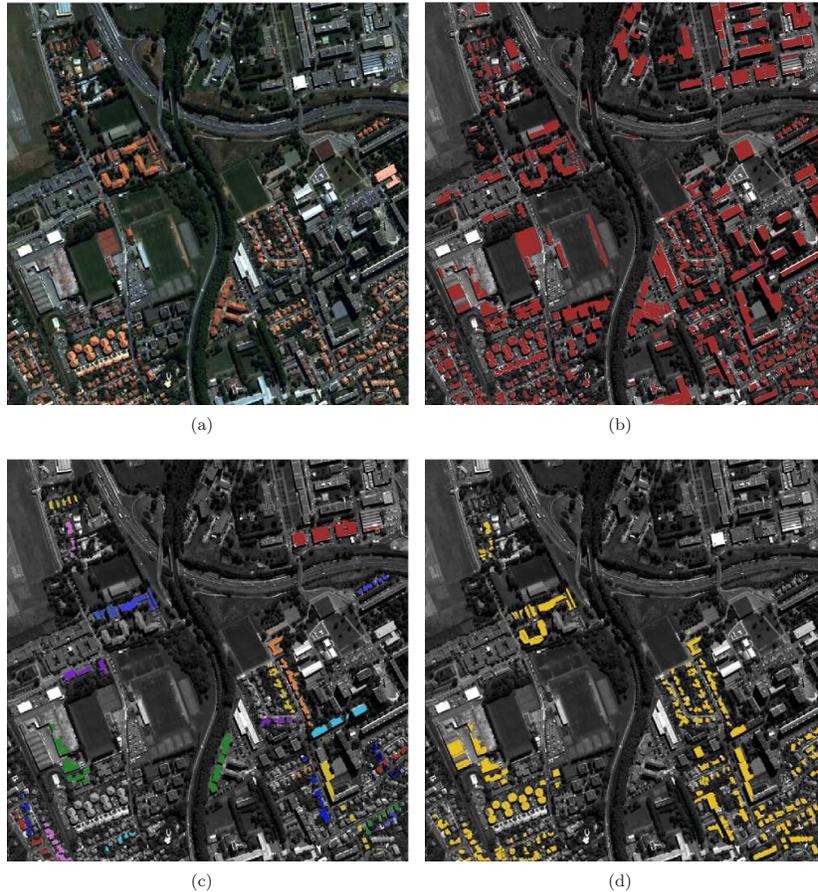


Figure 19: (a) Original image. (b) Segmented buildings. (c) Some of the *globally* aligned subsets of houses found by the algorithm with a degree of alignment greater than 0.85. (d). Clusters of houses belonging to *globally* aligned groups which are parallel and near to other groups with a degree greater than or equal to 0.8.

In this example we showed that by using the spatial relations of alignment and parallelism it is possible to recognize spatial urban patterns in the image.

The second example deals with the elimination of false detections of roads near urban areas obtained from a road extractor [44]. Figure 20(a) shows the result of a road detection algorithm. Some of the obtained roads are false detections. We know that in the residential areas, the houses are aligned forming groups which are parallel to the roads. Therefore, in this example, we are mainly interested in determining the roads which have a group of buildings parallel to them. However, since some of the roads extracted in Fig. 20(a) consist of road’s fragments, then we will be also interested in determining the roads which are parallel to a group of aligned buildings.

Considering that the groups of aligned buildings which are parallel to roads do not have to satisfy the constraint of buildings being close to each other, we extracted the groups of aligned buildings using $\beta = 0.85$ and a Voronoi neighborhood constrained by a distance of 70 pixels equivalent to approximately 49 m. Some of the obtained groups of aligned buildings are shown in Fig. 20(b). If we compare the results of Fig. 20(b) and of Fig. 19(c) we can see that we obtain longer groups, and some of the groups of Fig. 19(c) are included in Fig. 20(b). By allowing a larger distance between the members of an aligned group, we are more permissive and therefore we can obtain groups such as the green group on the bottom right part of the image, which is an aligned group made of distant objects, and does not represent a meaningful alignment for the description of the scene.

As in the previous example we made use of the “close to” relation. The resulting roads which are “close to” an aligned group of objects and which are parallel to the group or which have a group parallel to them are shown in Fig. 20(c) (we call this condition the constraint of parallelism). As we are interested only in the roads on residential areas, since the hypothesis of the constraint of parallelism is only valid these areas, Fig. 21 shows a subregion of the image where the roads that satisfy the constraint of parallelism. We note that most of the roads which have a low degree of satisfaction of this constraint are the roads which can be classified as false detections. However on the bottom of the image we see three false detections that continue to be detected with a high degree. This is due to the fact that there exist groups as the green aligned group in Fig. 19(c) for which these roads are parallel. Although there are still some false detections, we can observe that the number of false detections has been significantly reduced. Determining the roads which satisfy the constraint of parallelism can be seen as an intermediary step for a road and building extraction method. We can further think of combining the parallel and the alignment relations, with the relation “between” [41] to determine the region between two parallel groups of aligned buildings where it is possible to find a road.

Both presented examples demonstrated the need of modeling the alignment relation considering each object as a whole, rather than only its center of mass. For example the orange group in Fig. 20(b) is an aligned group with non aligned barycenters.

6. Conclusion

After having highlighted the importance of alignment and parallelism for high resolution remote sensing image interpretation, we proposed fuzzy models for determining what is the degree of satisfaction of these relations between a set of objects in an image. The proposed models take into account the semantic meaning of the relations. Two original definitions were presented for the alignment relation: *local* and *global* alignment. An algorithm for extracting *local* aligned groups of objects was proposed, based on fuzzy relative position measures, which takes into account the imprecision inherent to images and to the segmentation process. Based on the extraction of *locally* aligned groups of objects, we proposed a method for extracting *globally* aligned groups of objects using a graph based approach.

As the parallel relation and the alignment relation are frequently found together, we presented a model for determining when a linear object is parallel to an object, when a group of *globally* aligned objects is parallel to an object, and when a linear object is parallel to a group of objects. We showed how all the proposed models and algorithms can be extended to the case of fuzzy objects or when using fuzzy neighborhoods for the alignment relations.

The examples, on real objects extracted from satellite images, have shown the usefulness and power of the proposed models for scene understanding. We also highlighted how these relations can be used as intermediary steps for extracting objects in images.

Future work aims at combining the newly proposed spatial relations with state of the art spatial relations for the interpretation of complex scene on satellite images, which can be integrated in content-based image retrieval applications.

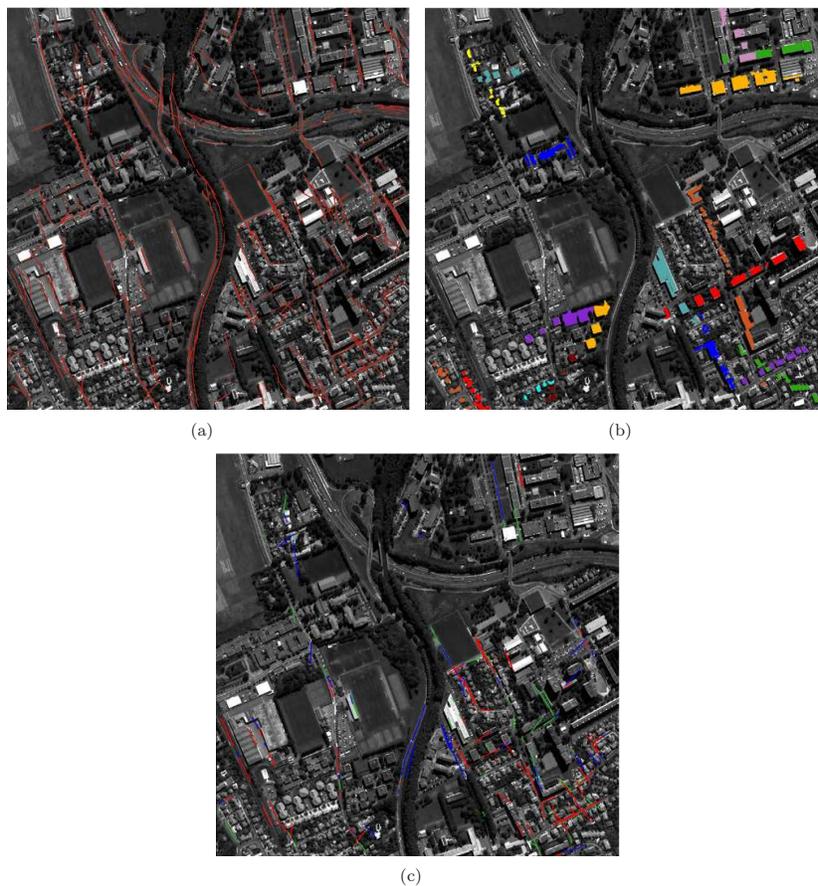


Figure 20: (a) Original roads. (b) Some of the *globally* aligned subsets of houses found by the algorithm with a degree of alignment greater than 0.85. (c) Obtained roads, after eliminating the roads which were not parallel to a group or did not have a group parallel to them. The green roads represent the roads which satisfy the constraint of parallelism with a degree between 0.3 and 0.5, the blue roads between a degree 0.5 and 0.8 and the red roads between a degree of 0.8 and 1.0.

Acknowledgements

This work was partially founded by CNES-DLR-ENST Competence Center for Image Understanding and Information Extraction. The authors would like to thank Vincent Poulain for providing the segmentations of buildings and roads.

References

- [1] I. Bloch, Fuzzy spatial relationships for image processing and interpretation: a review, *Image and Vision Computing* 23 (2005) 89–110.
- [2] S. Aksoy, C. Tusk, K. Koperski, G. Marchisio, Scene modeling and image mining with a

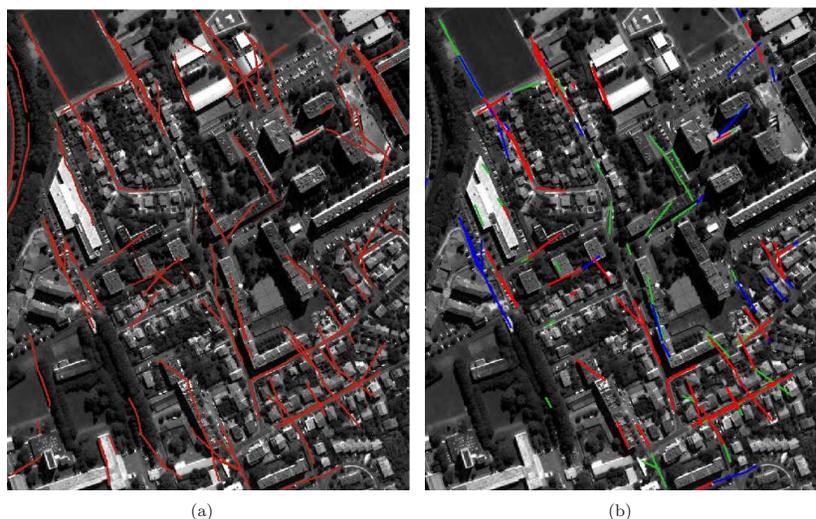


Figure 21: (a) Original roads of a subregion of Fig. 20(a). (c) Obtained roads, after eliminating the roads which were not parallel to a group or did not have a group parallel to them. The green roads represent the roads which satisfy the constraint of parallelism with a degree between 0.3 and 0.5, the blue roads between a degree of 0.5 and 0.8 and the red roads between a degree of 0.8 and 1.0.

visual grammar, in: C. Chen (Ed.), *Frontiers of Remote Sensing Information Processing*, World Scientific, 2003, pp. 35–62.

- [3] I. Bloch, T. Géraud, H. Maître, Representation and Fusion of Heterogeneous Fuzzy Information in the 3D Space for Model-Based Structural Recognition - Application to 3D Brain Imaging, *Artificial Intelligence* 148 (2003) 141–175.
- [4] O. Colliot, O. Camara, I. Bloch, Integration of Fuzzy Spatial Relations in Deformable Models - Application to Brain MRI Segmentation, *Pattern Recognition* 39 (2006) 1401–1414.
- [5] J. Keller, X. Wang, A fuzzy rule-based approach to scene description involving spatial relationships, *Computer Vision and Image Understanding* 80 (2000) 21–41.
- [6] D. Guo, H. Xiong, V. Atluri, N. Adam, Object discovery in high-resolution remote sensing images: a semantic perspective, *Knowledge and Information Systems* 19 (2009) 211–233.
- [7] J. Mota, G. Câmara, M. Escada, O. Bittencourt, L. Fonseca, L. Vinhas, Case-based reasoning for eliciting the evolution of geospatial objects, in: K. S. Hornsby, C. Claramunt, M. Denis, G. Ligozat (Eds.), *9th International Conference on Spatial Information Theory, COSIT'09*, volume 5756 of *Lecture Notes in Computer Science*, Springer, Aber Wrac'h, France, 2009, pp. 405–420.
- [8] F. L. Ber, A. Napoli, Object-based representation and classification of spatial structures and relations, in: *14th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2002*, pp. 268–275.
- [9] J. Inglada, J. Michel, Qualitative spatial reasoning for high resolution remote sensing image analysis, *Transactions on Geoscience and Remote Sensing* 47 (2009) 599–612.

- [10] A. Deruyver, Y. Hodé, L. Brun, Image interpretation with a conceptual graph: Labeling over-segmented images and detection of unexpected objects, *Artificial Intelligence* 173 (2009) 1245–1265.
- [11] J. Metzger, F. L. Ber, A. Napoli, Modeling and representing structures for analyzing spatial organization in agronomy, in: A. de Moor, W. Lex, B. Ganter (Eds.), *Conceptual Structures for Knowledge Creation and Communication*, 11th International Conference on Conceptual Structures, ICCS 2003 Dresden, Germany, volume 2746 of *Lecture Notes in Computer Science*, Springer, 2003, pp. 215–228.
- [12] S. Steiniger, R. Weibel, Relations among map objects in cartographic generalization, *Cartography and Geographic Information Science* 34 (2007) 175–197.
- [13] D. G. Lowe, Three-dimensional object recognition from single two-dimensional images, *Artificial intelligence* 31 (1987) 355–395.
- [14] R. Mohan, R. Nevatia, Using perceptual organization to extract 3D structures, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (1989) 1121–1139.
- [15] R. Mohan, R. Nevatia, Perceptual organization for scene segmentation and description, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (1992) 616–635.
- [16] H.-B. Kang, E. Walker, Perceptual grouping based on fuzzy sets, in: *IEEE International Conference on Fuzzy Systems*, 1992, pp. 651–659.
- [17] J. Rouco, M. Penas, M. Penedo, M. Ortega, C. Alonso-Montes, Artificial Intelligence Techniques and Recognition-Certainty Measure of Pairwise Line Segment Perceptual Relations Using Fuzzy Logic, in: *Iberoamerican Congress on Pattern Recognition*, volume 4756, Springer, 2007, pp. 477–486.
- [18] A. L. Ralescu, J. G. Shanahan, Perceptual organization for inferring object boundaries in an image, *Pattern Recognition* 32 (1999) 1923–1933.
- [19] S. Sarkar, K. L. Boyer, Perceptual organization in computer vision: A review and a proposal for a classificatory structure, *IEEE Transactions on Systems, Man, and Cybernetics* 23 (1993) 382–399.
- [20] A. Desolneux, L. Moisan, J. M. Morel, From Gestalt Theory to Image Analysis: A Probabilistic Approach, volume 34 of *Interdisciplinary Applied Mathematics*, Springer-Verlag, 2008.
- [21] M. Ortner, X. Descombes, J. Zerubia, Building outline extraction from digital elevation models using marked point processes, *International Journal of Computer Vision* 72 (2007) 107–132.
- [22] M. Wertheimer, Laws of organization in perceptual forms., in: W. H. Ellis (Ed.), *Source Book of Gestalt Psychology*, Routledge and Kegan Paul, London, 1938.
- [23] K. Koffka, *Principles of Gestalt Psychology*, Harcourt-Brace, New York, 1935.
- [24] H. Kang, E. Walker, Characterizing and controlling approximation in hierarchical perceptual grouping, *Fuzzy Sets and Systems* 65 (1994) 187–223.
- [25] S. Toh, Extracting viewpoint invariance relations using fuzzy sets, in: *Proceedings of the Intelligent Vehicles' 92 Symposium*, 1992, pp. 325–329.
- [26] H. Ip, W. Wong, Detecting perceptually parallel curves: Criteria and force driven optimization, *Computer Vision and Image Understanding* 68 (1997) 190–208.
- [27] D. Shen, W. Wong, H. Ip, Affine-invariant image retrieval by correspondence matching of shapes, *Image and Vision Computing* 17 (1999) 489–499.

- [28] L. Likforman-Sulem, A. Hanimyan, C. Faure, Extracting lines on handwritten documents by perceptual grouping, in: Proceedings of the Third International Conference on Document Analysis and Recognition, ICDAR 1995, volume 2, p. 774.
- [29] S. Christophe, A. Ruas, Detecting Building Alignments for Generalisation Purposes, in: Advances in Spatial Data Handling: 10th International Symposium on Spatial Data Handling, 2002, pp. 419–432.
- [30] A. Desolneux, L. Moisan, J. Morel, A grouping principle and four applications, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2003) 508–513.
- [31] K. Miyajima, A. Ralescu, Spatial organization in 2D segmented images: Representation and recognition of primitive spatial relations, Fuzzy Sets and Systems 65 (1994) 225–236.
- [32] I. Bloch, A. Ralescu, Directional relative position between objects in image processing: a comparison between fuzzy approaches, Pattern Recognition 36 (2003) 1563–1582.
- [33] R. Zwick, E. Carlstein, D. V. Budescu, Measures of similarity among fuzzy concepts: A comparative analysis, International Journal of Approximate Reasoning 1 (1987) 221–242.
- [34] I. Bloch, On fuzzy distances and their use in image processing under imprecision, Pattern Recognition 32 (1999) 1873–1895.
- [35] J. Rabin, J. Delon, Y. Gousseau, Circular earth mover’s distance for the comparison of local features, in: International Conference on Pattern Recognition, ICPR 2008, pp. 1–4.
- [36] I. Bloch, H. Maître, Fuzzy Mathematical Morphologies: A Comparative Study, Pattern Recognition 28 (1995) 1341–1387.
- [37] I. Bloch, C. Pellot, F. Sureda, A. Herment, Fuzzy Modelling and Fuzzy Mathematical Morphology applied to 3D Reconstruction of Blood Vessels by Multi-Modality Data Fusion, in: D. D. R. Yager, H. Prade (Eds.), Fuzzy Set Methods in Information Engineering: A Guided Tour of Applications, John Wiley and Sons, New-York, 1996, pp. 93–110.
- [38] A. Rosenfeld, Fuzzy digital topology, Information and Control 40 (1979) 76–87.
- [39] I. Bloch, Fuzzy Relative Position between Objects in Image Processing: a Morphological Approach, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (1999) 657–664.
- [40] M. Peura, J. Iivarinen, Efficiency of simple shape descriptors, in: Third International Workshop on Visual Form, VF 1997, pp. 443–451.
- [41] I. Bloch, O. Colliot, R. Cesar, On the Ternary Spatial Relation Between, IEEE Transactions on Systems, Man, and Cybernetics SMC-B 36 (2006) 312–327.
- [42] V. Poulain, J. Inglada, M. Spigai, J.-V. Tourneret, P. Marthon, Fusion of high resolution optical and SAR images with vector data bases for change detection, in: IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2009.
- [43] C. Hudelot, J. Atif, I. Bloch, Fuzzy spatial relation ontology for image interpretation, Fuzzy Sets Systems 159 (2008) 1929–1951.
- [44] V. Poulain, J. Inglada, M. Spigai, J.-V. Tourneret, P. Marthon, High resolution optical and sar image fusion for road database updating, in: IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2010, Hawaii.

17 Mise à jour de cartes numériques

High resolution optical and SAR image fusion for building database updating

Vincent Poulain, Jordi Inglada, *Member, IEEE*, Marc Spigai, Jean-Yves Tournet, *Senior Member, IEEE*, and Philippe Marthon

Abstract—This paper addresses the issue of cartographic database creation or updating using high resolution SAR and optical images. In cartographic applications, objects of interest are mainly buildings and roads. This paper proposes a processing chain to create or update building databases. The approach is composed of two steps. First, if a database is available, the presence of each database object is checked in the images. Then, we verify if objects coming from an image segmentation should be included in the database. In order to do those two steps, relevant features are extracted from images in the neighborhood of the considered object. The object removal/inclusion in the database is based on a score obtained by the fusion of features in the framework of Dempster-Shafer evidence theory.

Index Terms—Image analysis, change detection, data fusion, image databases, feature extraction.

I. INTRODUCTION

WITH the recent (or in the very next future) availability of high resolution (HR) optical and radar satellite sensors, such as in the ORFEO program¹, the need of multi-sensor image processing chains that are able to assist a human expert in scene interpretation is increasing.

In this work, we focus on the problem of cartography creation/update, and more precisely on built-up areas. We propose a generic image processing and interpretation chain for cartography creation/update. This chain is generic because it can process multi-sensor data (optical and SAR images are considered in this work), at various resolutions (between 70cm and 2.5m for results presented in this paper) and can take into account ancillary data (typically a digital map). HR optical images are often used in cartographic applications, thanks to their easy interpretation. However optical sensors are time and weather dependent. On the contrary, SAR sensors have an equal effectiveness at any time of the day and night. They can provide information quickly in emergency situations or in cloudy area for instance. However the interpretation of SAR images is more complex. Consequently a multi-sensor application can exploit the complementarity of sensors to provide a maximum of information on a scene.

Several scenarios are possible according to available images and data. The basic (and optimistic) case is the one where available input data are: a multispectral high resolution image (for instance Quickbird or Pleiades), a high resolution SAR image (for instance Cosmo-SkyMed or TerraSAR-X) and a vector database. The aim is then to update the vector database. However, other (less optimistic) scenarios of input data are

foreseen to be processed by the chain: for instance the use of a single optical or SAR image, or two images of same kind (Cosmo-SkyMed and TerraSAR-X for instance), or an optical image at a lower resolution (SPOT-5 for instance) or panchromatic, with or without database as prior information.

Often, existing methods are specific to one sensor in single mode. Indeed, in the field of building extraction with a single optical image, many methods have been proposed. In [1], [2] hypotheses of buildings are created by grouping primitives extracted from airborne images. Buildings are extracted in [3] from a panchromatic QuickBird image using clustering and edge detection. Methods based on segmentation of high resolution images followed by a segment classification to detect buildings are presented in [4] with multispectral images, in [5] with panchromatic images, and in [6], [7] with aerial RGB images. In [8] is presented a method based on active contours to check a digital map of buildings using a panchromatic Quickbird image. Methods based on a shape prior, using morphological operators are presented in [9]–[11]. A building detection method based on the integration of shape priors in a level-set image segmentation is proposed in [12]. Approaches described in [13], [14] use graph theory to extract buildings from optical images.

With SAR sensors, the analysis of a single image to extract buildings is a more challenging task. Some promising methods are based on marked point processes [15], [16]. However robust results are very hard to achieve. Building detection and height estimation methods are proposed in [17]–[19] using interferometric SAR data. A method for extracting building outlines using a SAR intensity image is explained in [20], and using features extracted from SAR and optical images in [21]. In dense urban environments, single SAR images are more efficiently used to extract the road network. For instance, in [22], a road network extraction is proposed based on a Hough transform and a road tracking algorithm.

Contrary to these approaches, this paper proposes a generic chain. It is able to integrate multi-sensor images and exogenous data. The goal is to exploit all the available information on a scene. The proposed chain can also evolve with the easy integration of new features.

The paper is organized as follows. Section II gives an overview of this work by describing our generic processing chain. Section III describes features used to characterize the objects of interest (buildings). A fusion method allowing one to combine all advantages of these features is studied in section IV. Finally, results are presented in section V.

¹<http://smc.cnes.fr/PLEIADES/>

II. PROCESSING CHAIN

Our processing chain is presented in Fig. 1. Inputs of the chain are high resolution images: an optical and/or a SAR image. Resolution of images should be in the range of 0.6 up to 2.5 meters for the optical images, and around 1 meter for the SAR images. A cartographic database (DB) can be available. The registration of images and DB is not part of this work. Thus images and DB are assumed to be registered. As we work at object level with buffer regions, a coarse registration (that can be performed automatically with an accuracy of several pixels) is sufficient to define appropriate buffer regions. Consequently a fine registration is not needed. The goal of this chain is to update (if available) or create a vector DB representing buildings. If a DB is available, our approach consists of two steps: first we consider each DB object and we check if the object is present in the SAR and optical images. To do so, some relevant features based on primitives are computed in the neighborhood of each object and fused to decide if the object should be kept in the DB. The second step consists of detecting buildings that are missing in the DB and including them in the DB. The proposed method is similar to the first step. However, instead of considering each DB object, we consider each region coming from a multiscale segmentation [23] of the optical image. In this work an optical image is required to extract new buildings. If no optical image is available, the detection of new buildings should be performed with approaches specific to building detection in SAR images such as [15], [16]. However, these methods are not used in this work.

The goal of the two steps presented in Fig. 1 is to provide a score for each object (coming from the DB or from the segmentation) representing its likelihood of being a building. Consequently, simple cases can be processed automatically whereas more complex cases may require the intervention of a human operator.

Moreover, a constraint of our algorithm is the absence of learning set to classify building candidates. So the classification has to be performed with prior knowledge on buildings. However, if a DB is available, the DB verification will be performed with the prior knowledge. Accepted buildings after this step will be used to optimize parameters of the building detection algorithm.

III. FEATURE EXTRACTION

A. Hypothesis generation

In the DB verification step, we consider objects coming from the DB (for buildings, these objects are polygons). As detailed in Fig. 1, objects are subject to feature computation, feature fusion and decision. For the detection of new objects (left part of the chain shown in Fig. 1), the three same steps are applied to objects (polygons). To do so, we must generate these objects, i.e., extract object hypotheses from images. In this work we only extract hypotheses from the optical images. They can be also extracted from SAR images, however it is more challenging. To generate building hypotheses, we perform a multi-scale segmentation of the optical image by using the Mean-Shift algorithm [23]. This

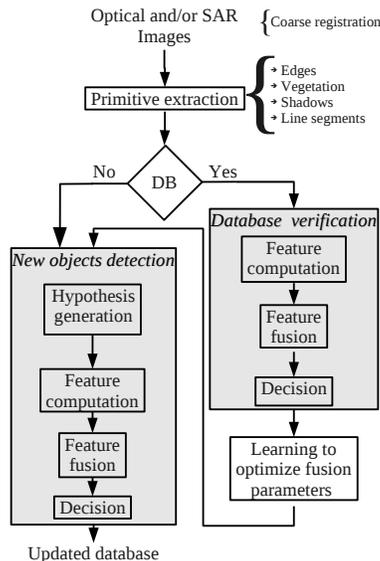


Fig. 1. Processing chain for building DB creation or updating.

algorithm requires three parameters: the spatial radius, spectral radius and minimum region size. We use various spectral radius to obtain segmentations at different scales. The resulting regions are then transformed into polygons that will be used for the detection of new objects. Note that if a DB is available, it was checked in the previous step. Consequently, building hypotheses are regions coming from the segmentation of the optical image that do not intersect polygons kept in the DB.

B. Feature computation

The goal of feature computation is to find clues in images about the presence of buildings. As the proposed approach must be generic, we have to find features common to most kinds of buildings. As represented in Fig. 2, in optical images, most buildings contrast with their surrounding, cast a shadow (as a building is higher than its surrounding), contain no vegetation and have linear walls (this feature characterizes man-made structures). In SAR images, some buildings present a contrast between the layover and the shadow area [24]. However these characteristics are not observed for all buildings. Indeed, if the sun is at its zenith, elevated objects do not cast a shadow. Moreover small houses do not have long enough walls to be detected as meaningful segments. A fusion process detailed in the next part is used to combine all available information extracted from images. The result of this fusion is used to make a decision on the relevance to include an object in the DB.

Algorithms used to extract primitives associated to this prior information are provided in Table I. Once primitives have been extracted from input images, appropriate features are computed in the vicinity of each building hypothesis providing a score associated to the presence of buildings. These features

are:

- 1) **Shadow:** This feature requires to know the direction of the sun as a prior information on the optical image. This information is used to determine which building hypothesis walls are oriented toward the sun. A shadow mask is also required. In this work we obtain the shadow mask by using an empirical thresholding of the optical image. However automatic methods such as the ones developed in [25], [26] could also be used for that purpose. We consider building hypothesis wall pixels, that are pixels of the optical image lying on the edges of input polygons. Around each building hypothesis wall not oriented toward the sun, we define a buffer region (width of several pixels depending on the image resolution) and compute the percentage of wall pixels that contain a shadow pixel in their neighborhood.
- 2) **Line segments:** Segments are extracted from the optical image using the Line Segment Detector [27]. For each building hypothesis wall (each line of input polygon), we consider extracted segments that are in their neighborhood and parallel to the wall (for our tests we allowed a tolerance of 10 degrees). We compute the percentage of building hypothesis wall pixels containing an extracted segment parallel to the wall in their neighborhood.
- 3) **Edges:** We compute the contrast between the building hypothesis and its neighborhood. The resulting score is the mean distance between building hypothesis borders and optical image nearest edges (extracted using Mean Shift [23]).
- 4) **No vegetation:** This feature requires a multispectral image. We determine the percentage of not vegetated pixels located inside the building hypothesis (a vegetation mask is obtained thanks to a thresholding of the NDVI [28] of the multispectral image).
- 5) **SAR contrast:** We define buffer regions around walls oriented toward the sensor (layover region) and behind opposite walls (shadow region). For each building hypothesis, we compute the ratio of means: score = $\log\left(\frac{m(\text{layover})}{m(\text{shadow})}\right)$.

These five features are examples that have been implemented in order to provide data to the fusion step, which is the crucial point of this work. Moreover, once the strategy of feature adding has been set up, new features can be easily integrated to the proposed generic processing chain. Note that according to the scenario, some features may not be computable. For instance if only one optical panchromatic image is available, only three features will be computed (shadow, line segments and edges). The score indicating the likelihood of being a building is computed from a fusion procedure described in the next section.

IV. FEATURE FUSION

A. Fusion framework

As explained in the previous section, several features are extracted from images. Each feature brings evidence on the presence of a building. To benefit from all information brought by features, we need to combine scores coming from each

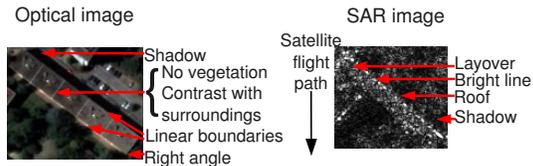


Fig. 2. Characterization of buildings in optical and SAR images

feature. The goal of the fusion is to exploit redundancy and to reduce uncertainty. Evidential reasoning can be based on three frameworks: Bayesian probability theory [30], Dempster-Shafer theory of evidence [31] and possibility theory [32]. Bayesian probability theory is a classical method for data fusion that is based on a well-developed decision-making theory. However, it requires a considerable amount of prior knowledge and cannot easily model imprecise, incomplete and not totally reliable information. Dempster-Shafer theory of evidence is a generalization of probability theory that allows us to capture the imprecise nature of evidence. The resulting decision is not very well defined since degrees of likelihood are measured by probability intervals instead of probabilities for the Bayesian framework. Possibility theory, based on fuzzy set theory [33], is also adapted to uncertain and imprecise information. It might be used as well in our application. However, with the possibility theory, several combination rules are possible, and the choice between these rules is not straightforward. The Dempster-Shafer evidence theory has a clearer and more rigorous foundation. Moreover it provides interesting byproducts, such as conflict between sources, ignorance (as the confidence is expressed through intervals), that can be used to analyze complex cases that do not fit the established model. A human operator might want to focus on those complex cases while cases with a great confidence value might will be processed automatically. In our application, features bring pieces of evidence on the probability of being a building. However most of them do not discriminate only buildings (for instance, absence of vegetation can be an evidence of presence of buildings, but also of roads). The Dempster-Shafer theory of evidence appears as the best adapted framework to represent and manage imprecision of features, and also to allow the easy integration of new features in the chain.

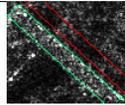
B. Evidence theory

In Dempster-Shafer framework, evidence is assigned to elements of the set of all possible propositions called *frame of discernment*, often denoted by Θ . The power set $\mathbb{P}(\Theta)$ is the set of all possible subsets of Θ . Subsets of Θ are called *propositions*. The quantity of evidence that a source assigns to a proposition is represented by a *mass function* (also called *basic probability assignment*). A mass function (MF) m satisfies the following properties

$$m : \mathbb{P}(\Theta) \rightarrow [0, 1], \sum_{A_i \subseteq \Theta} m(A_i) = 1, m(\emptyset) = 0. \quad (1)$$

Subsets that are assigned a mass by a source are called *focal sets* of the source.

TABLE I
FEATURE EXTRACTION.

Sensors	Optical				SAR	
	Panchromatic		Multispectral			
Properties	Higher than surroundings	Linear structure		Contrasts with surroundings	No vegetation	SAR contrast
Primitives	Shadow mask	Line [27]	Segment	Edges [23] and distance to edge measure [29]	Vegetation mask(NDVI [28] thresholding)	Ratio layover/shadow
Examples						

The uncertainty corresponds to the set Θ . Considering the focal set model built according to feature imprecision, masses assigned to sets included in (resp. containing) the building set will constitute the belief (resp. the plausibility) of building hypothesis. After defining focal sets for each feature, the fusion of information is performed thanks to the Dempster-Shafer orthogonal rule. The mass of a proposition P , resulting from the combination of two sources 1 and 2 is expressed as follows

$$m_{12}(P) = m_1 \oplus m_2(P) = \frac{1}{1 - \kappa} \sum_{A \cap B = P} m_1(A)m_2(B) \quad (2)$$

with $\kappa = \sum_{A \cap B = \emptyset} m_1(A)m_2(B)$.

Information about each proposition is represented by an interval, bounded by two values: the belief and the plausibility. The belief function contains all evidences attached to subsets of the proposition P

$$\text{Bel}(P) = \sum_{A \subseteq P} m(A). \quad (3)$$

The plausibility function is the sum of all the masses that intersect the set of interest P

$$\text{Pl}(P) = \sum_{A | A \cap P \neq \emptyset} m(A). \quad (4)$$

There are various ways to take a decision in the Dempster-Shafer framework. The main decision rules are the maximum of belief, the maximum of plausibility and the center of the interval whose boundaries are belief and plausibility. For our tests we have chosen the trade-off consisting of taking the mean of belief and plausibility.

C. Imprecision representation

To take into account the imprecision of our features, we build a model representing relationships between focal sets. To build the model, we consider each feature and the type of object they can discriminate. The approach is the following. A ground truth of an image is used to generate samples of various classes in an image. In dense urban environment we consider typically the following classes: building, vegetation, road, shadow, water, and heterogeneous regions (regions composed of parts of other classes). Feature values for the various classes

are used to build histograms. For instance histograms for the features *Shadow* and *No vegetation* are represented in Fig. 3. The histogram depicted in Fig. 3.(a) shows that elements of classes *Roads*, *Shadows* and *Heterogeneous objects* have a very low value, while elements of classes *Buildings* and *Vegetation* are more spread over the histogram. Consequently most buildings and vegetation can be distinguished from the other classes. Moreover the histogram depicted in Fig. 3.(b) shows that the feature *No vegetation* allows one to discriminate between vegetation and buildings. More generally, considering histograms for the five features, the analysis of feature value repartition for each class leads to following properties:

- Buildings contain no vegetation. However there are other objects that are not vegetated (like roads)
- Objects contrast with their surrounding
- Most buildings and trees project a shadow
- Most objects with linear borders are man-made structures (roads or buildings)
- Only some buildings present a contrast between the layover and the shadow area.

Based on these remarks, we build the model of relationships between focal sets displayed in Fig. 4. This figure shows relationships of partial or total inclusion between focal sets. Even if most buildings contain no vegetation, a no vegetated object is not always a building. Consequently, the set *Building* (hatched in the Fig. 4) is included in the focal set *No vegetation*. We mentioned previously that the feature *Shadow* can distinguish elevated objects (high vegetation and most buildings) from other objects. Consequently, in the prior model of Fig. 4, the only non vegetated set included in the *Shadow border* focal set is the *Building* set. Other elements of the *Shadow border* set are vegetated objects. In SAR images, according to their size and their orientation, some buildings present a contrast between layover and shadow area. Consequently the *SAR contrast* focal set is totally included in the *Building* set. The Dempster-Shafer evidence theory assigns pieces of evidence to each focal set. During the fusion step, the mass of each set is computed according to the Dempster-Shafer orthogonal rule. Finally, the decision is taken according to the value of belief and plausibility of the building set (hatched in Fig. 4). This model gathers prior information on feature imprecision. Data fitting will be represented by MFs

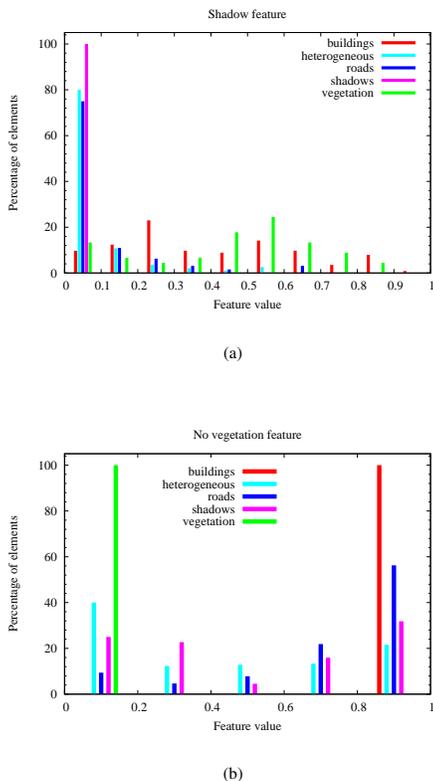


Fig. 3. Examples of histograms for features *Shadow* (a) and *No vegetation* (b).

detailed in the next part.

Note that the Bayesian probability theory would be more complicated to apply. Indeed, it would require to define the conditional probabilities of each subset for the fusion (for instance as the shadow feature cannot discriminate between buildings and trees, $P(\text{Building}|\text{Cast shadow})$ and $P(\text{Tree}|\text{Cast shadow})$ should be known, which requires a lot of prior knowledge).

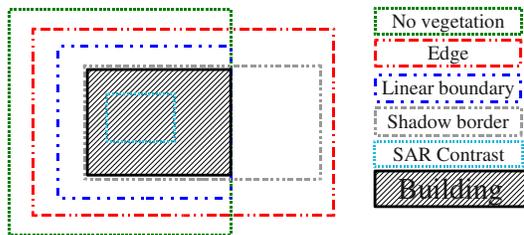


Fig. 4. Model of focal sets.

D. Representation of uncertainty

MFs have to be defined to represent uncertainty. MFs determine the quantity of evidence brought by a source to each set defined in the model represented in Fig. 4. Each feature that we have implemented brings information to three sets: the corresponding focal set, its complementary and uncertainty. To determine the quantity of evidence brought to each of these sets, we use trapezoidal mass functions which have shown simplicity and efficiency in a similar application [34]. The trapezoidal functions used in our study are represented in Fig. 5. For instance, an object with a high score for the feature *No vegetation* will lead to a high mass for the set *No vegetation* and a low mass for the uncertainty. Conversely, if the object is composed of vegetation pixels in almost half of its surface, the uncertainty will be high for this feature. MFs depend on four parameters (as represented in Fig. 5) gathered in $\theta_j = (a_j, b_j, c_j, d_j)$ where j varies from 1 to 5 and corresponds to each feature. At the beginning of the chain, MF parameters have to be set up. It can be achieved empirically, or optimized thanks to a ground truth. After the DB verification (if a DB is present), a learning set of verified buildings is available. The next subsection addresses the problem of optimizing the vector $\theta = (\theta_1, \dots, \theta_5)$ thanks to a learning set. In our tests, we have used true buildings coming from a ground truth, and false buildings (corresponding to other objects like shadows, roads, trees, and heterogeneous objects). Then MFs have been optimized using the method detailed in Section IV-E. Note that once MFs have been optimized with an image dataset, they can be used to process other images acquired in the same conditions (image type, resolution, environment, illumination, ...).

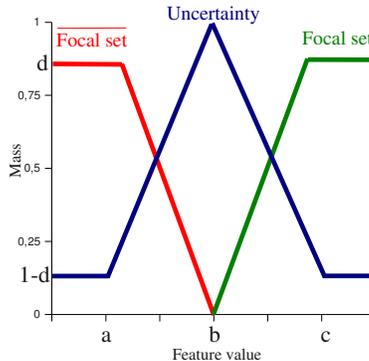


Fig. 5. Model for mass functions.

E. Parameter optimization

For each feature (shadow, line segment, edge, no vegetation and SAR contrast), three MFs are needed (for the focal set, its complementary and uncertainty), that lead to four parameters to optimize per feature (so 20 parameters in total). As described in the processing chain shown in Fig. 1, parameter optimization is conducted in the case of an available DB. After

verifying the DB, input objects are divided into two classes: buildings kept in the DB, and buildings removed from the DB. We use those samples to optimize the 20 parameters of the MF model, which is used to perform the detection of new buildings. The function that is minimized to estimate θ is defined as:

$$F(\theta) = p \sum_{i=1}^{n_0} \left(1 - \frac{\text{Bel}(B_i) + \text{Pl}(B_i)}{2} \right)^2 + (1-p) \sum_{i=n_0+1}^N \left(\frac{\text{Bel}(B_i) + \text{Pl}(B_i)}{2} \right)^2$$

where B_1, \dots, B_N represent the *Building* focal set of each object. We consider that objects numbered from 1 to n_0 are kept in the DB, while those numbered from $n_0 + 1$ to N are removed from the DB. Minimizing $F(\theta)$ consists of maximizing the mean of belief and plausibility for accepted buildings and minimizing this mean for rejected buildings, p weighting the two terms appearing in F (p is set to 0.5 in this paper reflecting the absence of knowledge for this parameter). The optimization can be achieved using a numerical optimization procedure. Results presented in this paper have been obtained thanks to the Nelder-Mead method [35].

Mass functions optimized for the feature *No vegetation* are provided in Fig. 6. The four parameters for this feature are estimated using real buildings (coming from an up to date DB) and false buildings (corresponding to other objects). Parameters are initialized using prior information: we consider that the uncertainty is maximum when half of object pixels are vegetation pixels ($b = 50$). Moreover, we have chosen $a = 0$, $c = 100$ and a reliability (parameter d) equal to 0.8 (as we have no information about these parameters). The result of the optimization proves that most of buildings contain no vegetation. Indeed, if an object contains at least 5.5% of vegetated pixels ($b = 94.5$), no evidence will be brought to the *No vegetation* focal set. The reliability of this feature has a value of 0.75. Consequently a highly vegetated building will present an uncertainty of 0.25, keeping a possibility to be considered as a building if it has other characteristics of a building (like a contrast in SAR images, line parallel to its wall,...).

Note that if some wrong buildings are kept in the DB or true buildings are rejected, the parameter estimation will be impacted. According to the application, solutions are possible to overcome this issue. For instance, a human operator can be included in the chain after the DB verification step to validate results. He will focus on uncertain and complex cases. Consequently the learning step will be performed with a correct learning set. If the chain is used in a fully automatic way, errors after the DB verification will reduce the quality of the object detection step.

V. SIMULATION RESULTS

Experiments have been performed to evaluate both the DB verification and the detection of new buildings. Two datasets are available. The first one has been acquired over

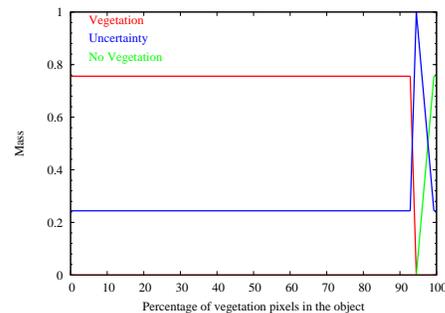


Fig. 6. Optimized mass functions for the feature *No vegetation*.

Toulouse, France, in an urban environment. More precisely, we used a Pleiades simulated image (coming from the airborne sensor PELICAN downsampled to 70cm and to 2.5m) and a TerraSAR-X satellite image at 1m resolution associated to the same area. The test area contains 111 buildings, i.e., 70982 pixels in the 70cm resolution optical image (whose size is 787x888 pixels). The second dataset has been acquired over Haiti. It consists of a 60cm QuickBird image and a TerraSAR-X image at 1m resolution. This test area contains 100 buildings, i.e., 175824 pixels in the 60cm resolution optical image (whose size is 1100x1332 pixels). Building DBs contain vector data representing the 2-D coordinates of building outlines.

Results are evaluated both at object and pixel levels. At object level, a reference building is considered as detected if more than 50% of its surface has been detected. A detected object is considered as a false alarm if more than 50% of its surface does not correspond to a building in the ground truth. Note that this threshold of 50% is a balanced value used in our tests. However, a more strict criterion might be used, for instance by considering a building as detected if at least 90% of its surface has been detected, and an object as a false alarm if more than 10% of its surface does not correspond to a building. Such a strict criterion would tolerate less imprecision in the detected building delineation. In our tests it would not impact the DB verification results because we use a ground truth to provide true buildings. A correctly accepted building is a true positive (TP). A correctly rejected object is a true negative (TN). A wrongly rejected building is a false negative (FN). A wrongly accepted object is a false positive (FP). Considering the number of TP, TN, FN and FP for each case we compute precision and recall [36]. The precision can be seen as a measure of correctness (corresponds to the probability that an accepted object is a true building), whereas the recall is a measure of completeness (corresponds to the probability that a reference building is accepted). The more those coefficients are close to 1 the best is the result. The F-measure corresponds to the harmonic mean of precision and recall. The pixel level evaluation is performed using the

false alarm rate (FAR) and the detection rate (DR). It is a classical evaluation method in this type of work. However, this evaluation is not suitable for the object level evaluation of new building detection. Indeed, FAR depends on the number of true negatives (false buildings correctly rejected). In the building detection step, the number of false buildings is huge (they are due to oversegmentation). Consequently FAR is always very low, and thus not interesting to evaluate results. The advantage of recall/precision evaluation is its independence to the number of true negatives.

For the first dataset, acquired over Toulouse, results are presented in seven cases which gather possible scenarios of our processing chain

- Case 1: a multispectral 70cm resolution image, and a 1m resolution TerraSAR-X image.
- Case 2: a multispectral 70cm resolution image.
- Case 3: a panchromatic 70cm resolution image, and a 1m resolution TerraSAR-X image.
- Case 4: a panchromatic 70cm resolution image.
- Case 5: a multispectral 2.5m resolution image, and a 1m resolution TerraSAR-X image.
- Case 6: a multispectral 2.5m resolution image.
- Case 7: a 1m resolution TerraSAR-X image.

For the second dataset, acquired over Haiti, scenarios are as follows

- Case 1: a multispectral 60cm resolution QuickBird image, and a 1m resolution TerraSAR-X image.
- Case 2: a multispectral 60cm resolution QuickBird image.
- Case 3: a panchromatic 60cm resolution QuickBird image, and a 1m resolution TerraSAR-X image.
- Case 4: a panchromatic 60cm resolution QuickBird image.
- Case 5: a 1m resolution TerraSAR-X image.

The proposed decision is taken thanks to a threshold T that can vary in the interval $[0, 1]$. An object is accepted if

$$\frac{Pl(B) + Bel(B)}{2} \geq T. \quad (5)$$

This threshold is empirical and depends on the application where a low false alarm rate or a high detection rate may be preferred. It can be adjusted by a human operator accordingly, or kept to a default value in a fully automatic use of the processing chain.

A. DB verification

To evaluate the DB verification with the dataset acquired over Toulouse, we have created a DB composed of true building outlines coming from a ground truth (111 buildings) and of false buildings (polygons) created manually (148 objects). The evaluation consists of checking if the processing chain is able to keep true buildings and remove false buildings. Results are presented in Table II for the seven cases. Our results prove that the DB verification step can be performed efficiently in the first six cases. However the problem is more challenging when only a SAR image is available. The analysis of Cases 5 and 6 highlights the interest of a 1m resolution SAR image when the optical image has a resolution of 2.5m. However,

when the optical image has a high resolution (70cm), Cases 1 and 2 show that the presence of a 1m resolution SAR image does not improve results. Fig. 7 illustrates the DB verification procedure for the first case. The threshold T mentioned in Eq. 5 was tuned empirically in order to have a balanced result, with three false positives and three false negatives. This result was obtained with a threshold value of 0.25.



Fig. 7. Result of DB verification with the Toulouse dataset for the case 1, with the following colors: TP (in green): 108 buildings correctly accepted, FN (in red): 3 buildings wrongly rejected, FP (in blue): 3 objects wrongly accepted and TN (in white): 145 objects correctly rejected.

To obtain those results, MFs have been determined thanks to a parameter optimization based on samples coming from a ground truth. Some results associated to the DB verification for Case 1 are presented in Table III. This table shows interesting objects (referred to as Case (a), ..., (f)) extracted from Fig. 7 and the corresponding masses brought by each feature. The third and fifth rows of this table correspond to the hypothesis of building outline projected respectively on the optical and SAR images. The fourth row corresponds to primitives extracted from the optical image with the following colors: edges in white, line segments in red, vegetation mask in green and shadow mask in blue. The considered cases have the following characteristics:

- Case (a) corresponds to a true building correctly kept in the database (like 107 other buildings). It is a large and high building, that contrasts with its neighborhood, projects a large shadow, with linear boundaries, contains no vegetation and presents a layover and a shadow area in the SAR image. Masses assigned to each focal set highlight the presence of all characteristics for this building. Consequently, the score corresponding to the mean of belief and plausibility for the set *Building* is high (0.87).
- Case (b) corresponds to a real building which has been

TABLE II
DB VERIFICATION RESULTS USING TOULOUSE DATASET.

Scenario	Object level evaluation							Pixel level evaluation	
	TP	TN	FN	FP	Precision	Recall	F-measure	FAR	DR
Case 1	108	145	3	3	0.973	0.973	0.973	0.00455	0.938
Case 2	108	145	3	3	0.973	0.973	0.973	0.00420	0.920
Case 3	103	140	8	8	0.928	0.928	0.928	0.00790	0.920
Case 4	99	140	12	8	0.925	0.892	0.908	0.00823	0.900
Case 5	101	140	10	8	0.927	0.910	0.918	0.0127	0.841
Case 6	97	140	14	8	0.924	0.874	0.898	0.0123	0.826
Case 7	82	129	29	19	0.812	0.739	0.774	0.0224	0.811

removed from the DB (like 2 other buildings). MFs show that it is caused by the absence of cast shadow (because of high vegetation surrounding the house), the absence of noticeable contrast in SAR image and the absence of linear boundaries.

- Case (c) corresponds to a false building accepted in the DB (like 2 other false buildings). The polygon is located in a heterogeneous region, close to edges and linear boundaries due to the road. Note that the belief value of the *Building* set is the same than for the previous case. However, this case is accepted thanks to its high plausibility value, due to the high uncertainty of most features.
- Case (d) corresponds to a true building that has been accepted with a high value of conflict. As explained previously, the conflict is an interesting byproduct of the Dempster-Shafer evidence theory. Conflict denotes an error in the model, or a problem of feature reliability. In this case, the building is high and projects a large shadow. However, in the optical image, its radiometry is very close to the radiometry of the adjacent road. Consequently, edges between the building and the road are not detected by the edge detector. The polygon is considered as located far from edges, and a conflict appears between the features “shadow” and “edges”. For this situation, a human operator should be alerted to process himself complex cases that do not correspond to the model. Indeed risks to have a wrong automatic decision in those cases is high.
- Case (e) corresponds to a small house, correctly kept in the DB. As this house is very low, its shadow is not visible. Moreover in the SAR image, dimensions of the house are too small to present a significant contrast between the layover and the shadow area. However, this house contrasts with its neighborhood, linear boundaries are present and it contains no vegetation. So even if the belief of the *Building* set is equal to 0, the plausibility is high enough to be kept in the DB.
- Case (f) corresponds to a false building correctly removed from the DB (like 144 other false buildings). The building hypothesis has been positioned in a parking lot. So it contains no vegetation however the outline is far from edges, it does not project a shadow, has no linear boundaries, and

does not present a contrast in the SAR image. Thus the plausibility of the *Building* set is low and the object has been rejected.

The evaluation of the database verification with the Haiti dataset is presented in Table IV. The DB was composed of 100 true buildings and 50 false buildings created manually. Results presented in Table IV confirm the ability of the processing chain to perform the DB verification in various environments. Fig. 8 represents accepted and rejected building hypotheses for the third case (a panchromatic 60cm resolution QuickBird image and a 1m resolution TerraSAR-X image). This figure shows that most false negative buildings are very small. Their primitives are hard to detect, even in a 60cm resolution image. False negative objects are located near linear edges. As in this scenario (Case 3) the multispectral information is not available, the presence of vegetation can not be used to reject building hypotheses located over a vegetated area.

B. Building detection

The previously checked DB was finally used for parameter optimization. To evaluate the detection of new buildings, we do not use this DB anymore. Indeed, in an operational use of this chain, we could try to detect buildings that are not in the DB. However, this number of buildings is very small after the DB verification. To have more representative results, we have considered an empty DB. In this case, the ideal goal of the chain is to detect all buildings in the images, i.e., to create the DB. In this step, we consider polygons provided by the multi-scale segmentation of the optical image. These polygons are processed similarly to DB objects, as detailed in the previous part.

Results are presented in Fig. 9 for the six first cases related to the Toulouse dataset described previously. The case where only a SAR image is available has not been considered here because there was no building extraction method from SAR images available in the free image processing library Orfeo Toolbox. The object level evaluation is represented with precision-recall curves in Fig. 9(a). The pixel level evaluation is represented in Fig. 9(b). These curves are obtained by changing the value of the threshold T defined in Eq. 5. Curves shown in Fig. 9 confirm that the availability of a SAR image slightly improves results. Moreover, Fig. 9(a) shows that contrary to the DB verification, building detection is performed

TABLE III
SIMULATION RESULTS.

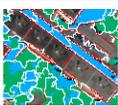
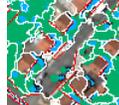
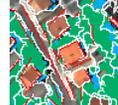
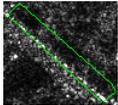
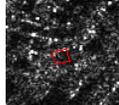
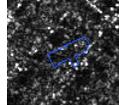
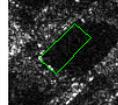
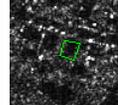
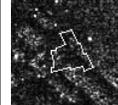
Cases		(a)	(b)	(c)	(d)	(e)	(f)
Status		true positive	false negative	false positive	true positive	true positive	true negative
Optical image							
Extracted primitives							
SAR image							
Shadow	m(Shadow)	0.74	0	0	0.67	0	0
	m(Shadow)	0	0.39	0.41	0	0.30	0.48
	m(Θ)	0.26	0.61	0.59	0.33	0.70	0.52
Linear boundary	m(Linear bnd.)	0.81	0	0.11	0.81	0.81	0
	m(Linear bnd.)	0	0.38	0	0	0	0.12
	m(Θ)	0.19	0.62	0.89	0.19	0.19	0.88
Edges	m(Edges)	0.25	0.46	0	0	0.51	0
	m(Edges)	0	0	0.03	0.81	0	0.71
	m(Θ)	0.75	0.54	0.97	0.19	0.49	0.29
No vegetation	m(No veg.)	0.71	0.55	0	0.73	0.70	0.70
	m(No veg.)	0	0	0.09	0	0	0
	m(Θ)	0.29	0.45	0.91	0.27	0.30	0.30
SAR contrast	m(SAR cont.)	0.12	0	0	0.61	0	0
	m(SAR cont.)	0	0.11	0.82	0	0.21	0.36
	m(Θ)	0.88	0.89	0.18	0.39	0.79	0.64
Conflict		0	0	0	0.13	0	0
Bel(building)		0.73	0	0	0.78	0	0
Pl(building)		1	0.38	0.53	0.90	0.70	0.13
Decision		0.87	0.19	0.27	0.84	0.35	0.067

TABLE IV
DB VERIFICATION RESULTS USING HAITI DATASET.

Scenario	Object level evaluation							Pixel level evaluation	
	TP	TN	FN	FP	Precision	Recall	F-measure	FAR	DR
Case 1	97	48	3	2	0.980	0.970	0.975	0.00738	0.967
Case 2	97	47	3	3	0.970	0.970	0.970	0.00762	0.967
Case 3	93	46	7	4	0.959	0.930	0.944	0.00668	0.954
Case 4	93	45	7	5	0.949	0.930	0.939	0.00923	0.954
Case 5	79	34	21	16	0.832	0.790	0.810	0.0220	0.824

better with a 70cm panchromatic image than with a 2.5m multispectral image. In the building detection step, the spectral information seems more important than the spatial resolution. This is due to the segmentation step, which is used to generate

building hypotheses. A good delineation of buildings is hard to achieve with the segmentation of a panchromatic image. Results show that for the first case, at pixel level, almost 70% of building pixels are detected for only 2% of false alarm. This

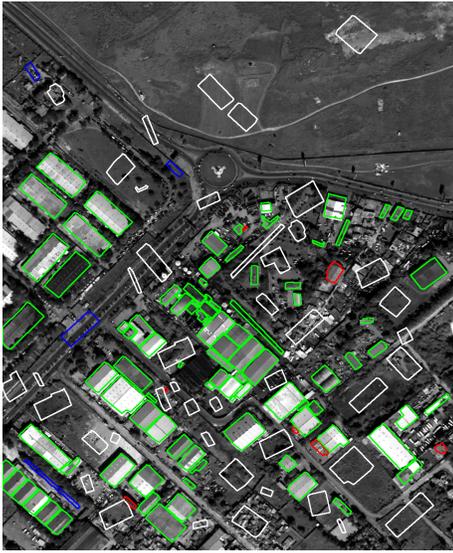
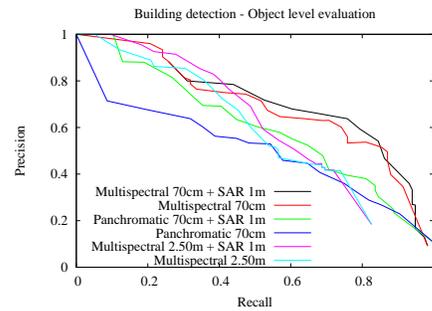


Fig. 8. Result of DB verification with the Haiti dataset for the case 3, with the following colors: TP (in green): 93 buildings correctly accepted, FN (in red): 7 buildings wrongly rejected, FP (in blue): 4 objects wrongly accepted and TN (in white): 46 objects correctly rejected.

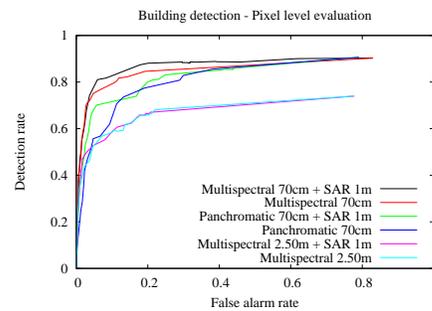
result corresponds to a precision of 72% and a recall of 62%. Detected buildings for this result are represented in Fig. 10. Results projected on the optical image (Fig. Fig. 10(a)) and on the SAR image (Fig. 10(b)) highlight that small houses are hard to detect. When buildings are correctly segmented in the optical image, results are very close to those obtained in the DB verification step. It corresponds to buildings that present a high contrast with their neighborhood. On the contrary, Fig. 10(a) shows that some large buildings whose radiometry is close to the adjacent road are not correctly detected. This is due to the problem of segmentation. Indeed, at a coarse scale, those buildings are merged with the adjacent road. However at a fine scale, they are divided into small regions, which do not verify characteristics of a building.

VI. CONCLUSION

This paper described a generic processing chain to update/create a cartographic DB with SAR and optical input images. Our results showed that the chain can process images in a large range of resolution (tested at 0.6m and 2.5m resolution for the optical image and at 1m for the SAR image). The chosen fusion framework was well adapted to the representation of feature imprecision. New features can be easily included in the proposed chain to improve the building detection. A score for each building hypothesis based on belief and plausibility gave a confidence value for each DB element. It also allowed a human expert to focus only on complex cases while simple elements were processed automatically. Future work will consist of analyzing these complex cases to extract new features able to discriminate these elements.



(a)



(b)

Fig. 9. Building detection: evaluation at object level with Precision-Recall curves (a) and at pixel level with ROC curves (b)

For instance new features could characterize the shape of segmented regions in the optical image. This will be useful to discriminate buildings and vegetation if just a panchromatic image is available. A similar processing chain is currently under investigation to perform the update of road DBs. Finally, it is interesting to mention that the proposed building processing chain was implemented using CNES ORFEO Toolbox free software <http://www.orfeo-toolbox.org>.



(a)



(b)

Fig. 10. Building detection without DB: building outlines projected on the optical image (a) and on the SAR image (b)

REFERENCES

- [1] C. A. Lin and R. Nevatia, "Building detection and description from a single intensity image," *Computer Vision and Image Understanding*, vol. 72, no. 2, pp. 101–121, Nov. 1998.
- [2] A. Katartzis and H. Sahlbi, "A stochastic framework for the identification of building rooftops using a single remote sensing image," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 259–271, Jan. 2008.
- [3] Y. Wei, Z. Zhao, and J. Song, "Urban building extraction from high-resolution satellite panchromatic image using clustering and edge detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Anchorage, Alaska, USA, Sep. 2004.
- [4] Z. J. Liu, J. Wang, and W. P. Liu, "Building extraction from high resolution imagery based on multi-scale object oriented classification and probabilistic Hough transform," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Seoul, South Korea, Jul. 2005.
- [5] L. Wei and V. Prinet, "Building detection from high-resolution satellite image using probability model," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Seoul, South Korea, Jul. 2005.
- [6] S. Muller and D. W. Zaum, "Robust building detection in aerial images," in *International Archives of Photogrammetry and Remote Sensing (IAPRS)*, Vienna, Austria, Aug. 2005.
- [7] B. Sirmacek and C. Unsalan, "Building detection from aerial images using invariant color features and shadow information," in *23rd International Symposium on Computer and Information Sciences, ISCIS '08*, Oct. 2008, pp. 1–5.
- [8] T. Bailloeuil, V. Prinet, B. Serra, P. Marthon, P. Chen, and H. Zhang, "Urban building land use change mapping from high resolution satellite imagery, active contours and Hough voting," in *Proc. 9th International Symposium on Physical Measurements and Signature in Remote Sensing (ISPMSRS)*, Beijing, China, 2005.
- [9] O. Aytekin, I. Ulusoy, E. Z. Abacioglu, and E. Gokcay, "Building detection in high resolution remotely sensed images based on morphological operators," in *4th International Conference on Recent Advances in Space Technologies, RAST '09*, Jun. 2009, pp. 376–379.
- [10] S. Lefevre, J. Weber, and D. Sheeren, "Automatic building extraction in VHR images using advanced morphological operators," in *IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas (URBAN)*, Paris, France, Apr. 2007.
- [11] H. Sportouche, F. Tupin, and L. Denise, "Building extraction and 3D reconstruction in urban areas from high-resolution optical and SAR imagery," in *Urban Remote Sensing Event*, Shanghai, China, May 2009.
- [12] K. Karantzas and N. Paragios, "Recognition-driven two-dimensional competing priors toward automatic and accurate building detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 133–144, Jan. 2009.
- [13] B. Sirmacek and C. Unsalan, "Urban-area and building detection using SIFT keypoints and graph theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1156–1167, Apr. 2009.
- [14] T. Kim and J. P. Muller, "Development of a graph-based approach for building detection," *Image Vis. Comput.*, vol. 17, no. 1, pp. 3–17, Jan. 1999.
- [15] M. Quartulli and M. Datcu, "Information extraction from high resolution SAR data for urban scene understanding," in *Proc. 2nd GRSS/ISPRS Joint workshop on data fusion and remote sensing over urban areas (URBAN 2003)*, 2003, pp. 115–119.
- [16] M. Quartulli and M. Datcu, "Stochastic geometrical modeling for built-up area understanding from a single SAR intensity image with meter resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, pp. 1996–2003, 2004.
- [17] C. Tison, F. Tupin, and H. Maitre, "Retrieval of building shapes from shadows in high resolution SAR interferometric images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Anchorage, Alaska, USA, Sep. 2004, pp. 1788–1791.
- [18] P. Gamba, B. Houshmand, and M. Saccani, "Detection and extraction of buildings from interferometric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 611–617, Jan. 2000.
- [19] A. Thiele, E. Cadario, K. Schulz, U. Thoennessen, and U. Soergel, "Building recognition from multi-aspect high-resolution InSAR data in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3583–3593, Nov. 2007.
- [20] A. J. Bennett and D. Blacknell, "The extraction of building dimensions from high resolution SAR imagery," in *Proceedings of the International Radar Conference*, Sep. 2003, pp. 182–187.
- [21] F. Tupin and M. Roux, "Detection of building outlines based on the fusion of SAR and optical features," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, no. 1, pp. 71–82, Jun. 2003.
- [22] V. Amberg, M. Spigai, M. Coulon, and P. Marthon, "Improvement of road extraction in high resolution SAR data by a context-based approach," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Seoul, South Korea, Jul. 2005.
- [23] D. Comaniciu and P. Meer, "Mean shift analysis and applications," in *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, Kerkyra, Greece, 1999.
- [24] D. Massonnet and J. C. Souyris, *Imaging with synthetic aperture radar*, chapter 3, pp. 165–167, CRC Press, Boca Raton, FL, 2008.
- [25] K. S. Gudmundsson and F. Cagatin, "Complex shadow extraction," in *International Society for Optical Engineering (SPIE)*, San Diego, CA, USA, Aug. 2008, vol. 7072, pp. 707211.1–707211.10.
- [26] H. T. Guo, Y. Zhang, J. Lu, and G. W. Jin, "Research on the building shadow extraction and elimination method," in *International Archives of Photogrammetry and Remote Sensing (IAPRS)*, Beijing, China, 2008, pp. 569–574.
- [27] R. Grompone Von Gioi, J. Jakubowicz, J. M. Morel, and G. Randall, "On straight line segment detection," *J. Math. Imaging Vis.*, vol. 32, no. 3, pp. 313–347, 2008.

- [28] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sensing of Environment*, pp. 127–150, 1979.
- [29] P. E. Danielsson, "Euclidean distance mapping," *Computer Graphics and Image Processing*, vol. 14, 1980.
- [30] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, pp. 98–105, John Wiley and Sons, 1973.
- [31] P. Smets, "What is Dempster-Shafer's model?," in *Advances in the Dempster-Shafer theory of evidence*, pp. 5–34, John Wiley & Sons, Inc., New York, USA, 1994.
- [32] D. Dubois and H. Prade, "Possibility theory, probability theory and multiple-valued logics: A clarification," *Annals of Mathematics and Artificial Intelligence*, vol. 32, no. 1-4, pp. 35–66, 2001.
- [33] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets Syst.*, vol. 100, pp. 9–34, 1999.
- [34] F. Tupin, I. Bloch, and H. Maître, "A first step toward automatic interpretation of SAR images using evidential fusion of several structure detectors," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1327–1343, May 1999.
- [35] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, Jan. 1965.
- [36] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *ICML '06: Proceedings of the 23rd international conference on Machine learning*, New York, NY, USA, 2006, pp. 233–240, ACM.



Vincent Poulain was born in Dieppe, France, in 1983. He received the Engineer degree in electrical engineering and signal processing from ENSEEIHT, Toulouse, France, in 2007 and the Ph.D. degree in image processing in 2010 from the National Polytechnic Institute of Toulouse, France. He joined Thales Group in 2010 where he is currently working in the field of remote sensing image processing.



Jordi Inglada received the Telecommunications Engineer degree from both the Universitat Politècnica de Catalunya, Barcelona, Spain, and the École Nationale Supérieure des Télécommunications de Bretagne, Brest, France, in 1997 and the Ph.D. degree in signal processing and telecommunications in 2000 from Université de Rennes 1, Rennes, France. He is currently with the Centre National d'Etudes Spatiales (French Space Agency), Toulouse, France, working in the field of remote sensing image processing at the CESBIO laboratory. He is in charge of the development of image processing algorithms for the operational exploitation of Earth observation images, mainly in the field of multitemporal image analysis for land use and cover change. Dr. Inglada is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.



Marc Spigai was born in Marseille, France, 1965. He received the Engineering degree in signal and image processing from the École Nationale Supérieure des Télécommunications de Bretagne (Telecom Bretagne), France (1989). From 1990 to 1998, he works on signal and data processing at Thales System Airborne, Elancourt France. From 1998 to 2000 he was Head of a research team on fusion of radar, vision and navigation for driving assistance at Renault, Guyancourt, France. He joined Thales Alenia Space, Toulouse, France, in 2000, where he currently works on SAR signal and image processing at the Observation Systems and Radars Unit.



Jean-Yves Tourneret (SM'08) received the ingénieur degree in electrical engineering from ENSEEIHT (Toulouse, France) in 1989 and the Ph.D. degree from the National Polytechnic Institute from Toulouse in 1992. He is currently a professor in the university of Toulouse and a member of the IRIT laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal processing with a particular interest to MCMC methods. He was the program chair of the European conference on signal processing (EUSIPCO). He was also member of the organizing committee for the international conference ICASSP'06 which was held in Toulouse (France) in 2006. He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society (2001 – 2007, 2010 – present). He is currently serving as an associate editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.



Philippe Marthon received the engineering degree and the Ph.D. degree in computer science from the Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique et d'Hydraulique de Toulouse (ENSEEIHT), Toulouse, France, in 1975 and 1978, respectively, and the Doctorat d'Etat Degree from the Institut National Polytechnique de Toulouse (INPT), in 1987. He has published over 80 scientific articles in international journals and conferences. His research interests include systemics, operations research, remote sensing image processing and computer vision. He is currently Professor at ENSEEIHT and Research Scientist at the Institut de Recherche en Informatique de Toulouse (IRIT).

18 Traitements opérationnels

Remote sensing processing: from multicore to GPU

Emmanuel Christophe, *Member, IEEE*, Julien Michel, and Jordi Inglada, *Member, IEEE*

Abstract—As the amount of data and the complexity of the processing rise, the demand for processing power in remote sensing applications is increasing. The processing speed is a critical aspect to enable a productive interaction between the human operator and the machine in order to achieve ever more complex tasks satisfactorily. Graphic processing units (GPU) are good candidates to speed up some tasks. With the recent developments, programing these devices became very simple. However, one source of complexity is on the frontier of this hardware: how to handle an image that does not have a convenient size as a power of 2, how to handle an image that is too big to fit the GPU memory? This paper presents a framework that has proven to be efficient with standard implementations of image processing algorithms and it is demonstrated that it also enables a rapid development of GPU adaptations. Several cases from the simplest to the more complex are detailed and illustrate speedups of up to 400 times.

Index Terms—GPU, CUDA, OpenCL, implementation

I. INTRODUCTION

THE amount of data acquired by imaging satellites has been growing steadily in recent years. There is a rapidly increasing number of applications that benefit from the decline in prices and the easier access to such data. With this proliferation of data, relying on humans to do most of the high level interpretation tasks is no longer possible. Some (but not all) advanced tasks need to be processed automatically. However, these tasks are more complex, thus raising the computational power required.

As highlighted in an insightful report from Berkeley [1], the increase in computational power for the coming years goes through a parallel approach. High Performance Computing (HPC) is a natural solution to provide the computational power required. There are several approaches to HPC: clusters, grids or clouds are some examples. However, we chose here to focus on desktop HPC with the use of graphics processing units (GPU). The idea is to bring the processing power as close as possible to the final user to enable a better human-algorithm interaction.

It is now possible to use GPUs to do general purpose processing. Benefiting from investment from the movie and gaming industries [2], the processing power of GPUs has increased dramatically. They have evolved in a different direction than the general purpose central processing units (CPU). They harbor hundreds of processing units that are able to work at the same time. CPUs and GPUs rely on different trade-offs

E. Christophe was with the Centre for Remote Imaging, Sensing and Processing (CRISP), National University of Singapore, Singapore at the time of writing. He is now with Google Inc. USA.

J. Michel was with Communications et Systèmes (CS), Toulouse, France. He is now with CNES, Toulouse, France

J. Inglada is with the Centre d'Etudes Spatiales de la Biosphère (CESBIO), CNES, Toulouse, France

Manuscript received...

regarding the amount of cache memory versus the number of processing units.

To benefit from these hundreds of processing units, the inherent parallelism of the algorithms needs to be exposed. Often in the literature, the focus is on the implementation of the core algorithm. However, one critical difficulty arises from boundary conditions (when the number of pixels in the image is not a convenient multiple) and also from the data size that often cannot be kept in the hardware memory in one go (thus requiring several passes to reach the final result).

When designing a library to benefit from the capabilities of GPUs, one has to think of both the final user of the program and the developer who is going to write new programs. For the former, it is important to keep him isolated from these implementation details: the program should work for any image size on any hardware. The latter will benefit from a framework to simplify the nitty-gritty mechanisms so that he can focus on performances.

The aim of this paper is to present a framework enabling an easier implementation of the GPU kernel for some parts of a global remote sensing image processing pipeline. The framework is available as open source software in the Orfeo Toolbox library [3].

The Orfeo Toolbox (OTB) is an open source library developed by CNES (the French Space Agency). It contains numerous algorithms for preprocessing as well as for information extraction from satellite images [4].

One of the main objectives of the Orfeo Toolbox (OTB) is to provide a strong and robust software architecture to facilitate the *scalability* of newly implemented algorithms and to relieve (at least partially) the researcher from such concerns. The processing model for OTB has its roots in the Insight Toolkit [5] and has been proven to be effective for remote sensing images as well as for medical images. The general architecture of OTB is described in section II-A and in section II-B we describe how it can be used to exploit the GPU processing capabilities.

In section III, several examples of implementation with increasing complexity are described and show improvement ranging from zero to 400 time faster than the comparable CPU implementation.

Section IV discusses some of the perspectives that arise from such a speedup in the way we design and we work with algorithms.

Finally, V concludes presenting some directions for further improvements.

II. PROBLEM STATEMENT

This section presents the main issues related to the processing of remote sensing images: scalability. There are two

dimensions to this problem: the size of the data and the time required for processing. A significant bottleneck for the use of new algorithms is the jump from toy image examples to real satellite images. The first issue is related to the size of the image which cannot be held into memory any more (section II-A) and the second is related to the computation time and how to benefit from multiple computation units as commonly found in CPUs and GPUs (section II-B). Even if those two problems are related, they play a part at different levels.

A. Image size scalability: streaming

The sheer size of satellite images – several gigabytes – makes processing by usual methods inapplicable on standard computers. It is not desirable or possible to load the entire image into memory before doing any processing. In this situation, it is necessary to load only part of the image and process it before saving the result to the disk and proceeding to the next part. This corresponds to the concept of on-the-flow processing.

Remote sensing processing can be seen as a *chain of events or steps that lead to a final output* [6]. Each of these steps is generally independent from the following ones and generally focuses on a particular domain. For example, the image can be radiometrically corrected to compensate for the atmospheric effects, indices (such as NDVI) computed, before an object extraction based on these indexes takes place. The typical processing chain will process the whole image for each step, returning the final result after everything is done.

For some processing chains, iterations between the different steps are required to find the correct set of parameters. Due to the variability of satellite images and the variety of the tasks that need to be performed, fully automated tasks are rare. Humans are still an important part of the loop.

In these conditions, it is valuable to be able to provide some feedback quickly for only parts of the image and reprocess this part for a different set of parameters. Better yet if only the modified steps are reprocessed and not the whole chain; this is the concept of on-demand processing.

These concepts are linked in the sense that both rely on the ability to process only one part of the data. In the case of simple algorithms, this is quite easy: the input is just split into different non-overlapping pieces that are processed one by one. But most algorithms do consider the neighborhood of each pixel. As a consequence, in most cases, the data will have to be split into partially overlapping pieces.

The objective is to obtain the same result as the original algorithm as if the processing was done in one go. Depending on the algorithm, this is unfortunately not always possible.

In the Orfeo Toolbox, the processing elements are organized in the library as filters. Filters perform operations such as reading and writing the data, but also processing, e.g. linear filtering, thresholding or classification. Writing a new application (or a new processing chain) consists of plugging a few filters together to create a processing pipeline. As highlighted above, in most cases, the whole image cannot be held in memory at once and a different processing model is required.

Figure 1 illustrates the process on a simple example. In this case, four filters are connected together:

- a reader that loads the image, or part of the image in memory from the file on disk;
- a filter which carries out a local processing that does not require access to neighboring pixels (a simple threshold for example), the processing can happen on CPU or GPU;
- a filter that requires the value of neighboring pixels to compute the value of a given pixel (a convolution filter is a typical example), the processing can happen on CPU or GPU;
- a writer to output the resulting image in memory into a file on disk, note that the file could be written in several steps.

We will illustrate on this example how it is possible to compute part of the image in the whole pipeline, occurring only a minimal computation overhead.

Once all the filters are connected together and the pipeline is created as in Fig. 1 (a), the processing is started by a call on the last filter of the pipeline, which is the writer in our example. This filter requests its input to provide the information regarding the size of the image it will produce (Fig. 1 (b)). The reader gets the information by reading the meta-data of the file and the information is propagated through the pipeline (Fig. 1 (c)). Eventually, filters can modify this information, depending on the processing they apply on their input.

Once the writer has the information regarding the size of the data it has to produce, it can compute the splitting strategy: depending on the maximum memory specification, it will select only a small area of the image to request to its input filter. In Figure 1 (d), this area is represented by the red rectangle. The writer requests this area to its input: filter 2. This filter needs the value of the neighboring pixels of each pixel to be able to process its output. For the pixels in the middle of the region, this is not a problem, however, the region needs to be expanded to accommodate the need of the pixels at the border. This extension is represented by the blue rectangle in Figure 1 (d).

Here, there are two different cases: either the value is part of the image and can be obtained, or it is outside of the image area. In the first case (the bottom line in our example), the region is simply extended and the value will be generated by the input filter. If the value is outside of the image (top, left and right of the red region in our example), a strategy is necessary to create this value. This is handled at the filter level and several strategies (boundary conditions) are available: constant value, mirror, zero flux Neumann. In section III-B, we will see how this strategy can ease the constraints on the GPU implementation.

Once the request reaches the reader, which is the first filter of our pipeline, it can generate the requested area from the file and pass it to the next filter (Fig. 1 (e)). Once the region reaches the writer, it is written on the disk and the process continues with the next tile (Fig. 1 (f)).

This process relies on the capability of each filter to determine the size of the input needed to produce the output required by the downstream filter. Some specific algorithms

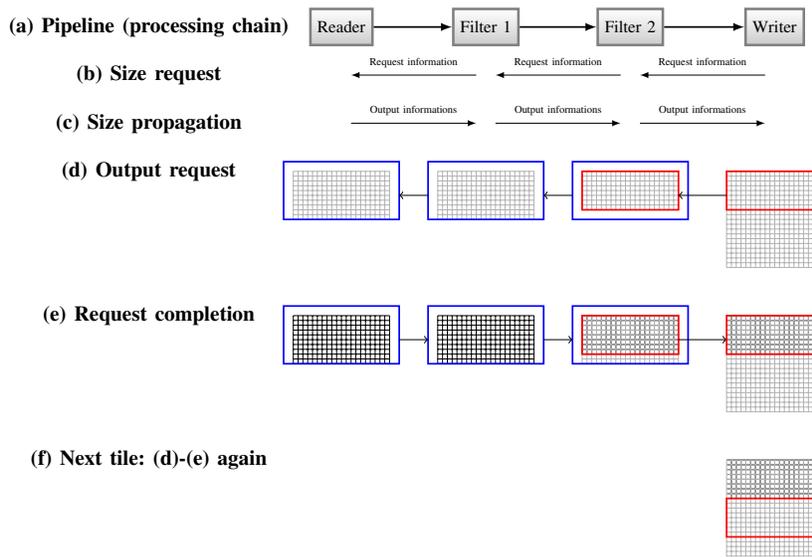


Fig. 1. Streaming model for OTB: illustration on a simple pipeline with two processing filters between a reader and a writer. One filter (filter 1) does pixel by pixel processing, while the other (filter 2) does a neighborhood processing.

cannot be rigorously implemented to work only on extracts of the image: Markov fields for example where the iterations introduce potential long range dependencies between pixels. In these situations, an approximation can usually be obtained. Note that the processing of the different tiles is independent and does not use common memory. It could ideally be coupled with distributing processing techniques (cluster or grids) where each node would process a tile, but this point is outside the scope of this paper.

The process is illustrated above in great detail, but it is worth mentioning that the user of the library does not need to understand this mechanism. Indeed, he may not even need to be aware of it. The developer of new filters, does not need to fully understand it, having a knowledge of the relevant customization points is sufficient. This design will appear to be critical in the context of the use of GPU where the memory is more limited.

B. Processing unit scalability: multithreading

1) *Short review of CPU vs GPU architecture:* Several architectures are available to enable parallel processing of data. The most common are SIMD and MIMD (using Flynn's taxonomy [7]). SIMD (Single Instruction, Multiple Data streams) where the same set of instruction is applied on several data streams is particularly suited to digital image processing. SIMD is available in CPU and enable to apply a single operation on multiple data at once. MIMD (Multiple Instruction, Multiple Data streams) corresponds to using several cores in a single die. These cores are able to execute independent instructions on different data.

Recent CPU combines several parallelization techniques to increase performances while giving the impression that they

work sequentially: branch prediction, out-of-order execution, superscalar. All these techniques increase the complexity of the CPU, limiting the number of CPUs that can be included on a single chip.

On the other hand, GPUs keep each processing unit simple but pack thousands of them on the chip. One of the critical difference is the lower amount of cache memory available. As a consequence, the GPU will work well when the level of data parallelism is high and enable masking the latency of each thread.

If we omit some of the details above and compare how CPU and GPU will process the pixels of the image: we can consider that the CPU will process each pixel sequentially but very fast while the GPU will process them slower but a whole lot of them at a time.

2) *The rise of multicore CPUs:* As the frequency of processors is reaching limits due to heating issues, advances in processing capabilities of recent CPUs are geared towards the increase in the number of cores [1], [8]. Recent CPUs are able to handle 8 to 12 threads simultaneously.

However, designing an application to benefit from these multicore architectures is not straightforward. The problem is to be able to split the data into different entities that can be processed simultaneously.

In section II-A, the main issue was limiting the size of the image to load into memory; here it is to draw on the availability of several processors. One common point between the two problems is that they rely on the possibility to process an extract of the data at a time. A major difference is that multithreading in the context of multicores will have access to shared memory between the cores. In the case of streaming, there is no shared memory between the tiles.

is almost identical in both cases which means that investing efforts into one technology will also benefit the other in case a change is required in the future. The setup of the kernel (the part of the code which prepares the data and transfers it to the processing units) is more complicated in the OpenCL case as there is some overhead due to the support for more heterogeneous hardware and for the compilation of the kernel during execution. However, this increased complexity enables more portability as OpenCL targets CPUs and GPUs alike. Concerning the availability of existing code, CUDA benefits from a head start of a few years compared to OpenCL and from many libraries heavily optimized for most common tasks (FFT, linear algebra...).

In term of performances, this difference in development stage has an impact. Earlier study [17] concludes that the OpenCL implementation is 13 to 63% slower. Figure 10 compares the performances of equivalent implementations using CUDA and OpenCL on the same hardware (OpenCL about 45% slower).

	CUDA	OpenCL
Kernel code	Simple	Simple
Kernel setup	Simple	More complicated
Portability	Low	High
Library availability	High	Low

TABLE I
COMPARISON BETWEEN CUDA AND OPENCL

GPU memory is more limited than CPU memory. As several types of memory are available special attention to which memory is used is required for further optimizations.

In the case studies described in the next section, the amount of data to be processed can be above the memory size. Another issue appears related to the size of the images. In most cases, each thread of the GPU is going to process one pixel. Due to the hardware architecture, threads need to be gathered into thread blocks, forming, for example, a group of 16×16 threads. This is fine if the image size is a multiple of 16. When processing regular images, this is unlikely, and this problem needs to be accounted for.

It can be handled at the GPU level, but it usually involves branching conditions. If possible, it is better to handle it before transferring the data to the GPU, making sure that the data size is suitable (a multiple of 16 in the example above). Thankfully, the pipeline model (Fig. 1) does just that: when the processing requires some neighborhood information (a simple case is a convolution), the pipeline is able to adapt the requested region to ensure that the necessary data are available. To comply with the GPU requirements, we just adapt this request to make sure that the region size is a multiple of 16.

In the following examples, the size of the image is handled as described in section II-A. The only modification, which fits perfectly in the pipeline structure is how the filter is computing the region required to produce the output.

III. CASE STUDIES OF GPU MIGRATION

In this section we introduce several examples of processing algorithms in order to illustrate different trade-offs in terms of

memory size and computation complexity which will allow the reader to get insight on the benefit of GPU-based approaches and when it would be most profitably applied. We have selected 3 algorithm categories which cover most of the steps of a classical remote sensing image processing chain.

Relative performances are compared between the original program on CPU using either a multithreaded implementation or otherwise, and the same program running on GPU which provides similar results. Outputs are compared to make sure that no differences other than those due to the single precision computation appears.

It is always a delicate task to compare programs using their execution time as it depends heavily on the quality of the implementation. Unlike what was done in the Intel's study [16], we do not push the implementation optimization to the maximum, but instead choose to focus on good quality implementation attainable with reasonable effort and hardware knowledge by the typical remote sensing scientist. There is probably room for improvement on the CPU side (using SIMD) as well as on the GPU side (coalescing access).

The hardware is an Intel i7-920 with 6 GB of RAM, the GPU is a Nvidia GTX-260 used purely for processing, the display being handled by another card. In terms of software, the C/C++ compiler gcc/g++ 4.4.3 is used with the version 3.0 of the CUDA toolkit for both CUDA and OpenCL simulations. Compilation option used is -O3 which turn on all the optimization available.

A. Pixel-based processing, a first naive example

The first category of algorithms refer to those that perform operations on single pixels without the need of context. This category can include any arithmetic or logical operation on pixels such as simple additions and thresholdings or more complex computations such as numerical solutions of equations where the pixel value is a parameter. This category also includes pixel based classification such as maximum likelihood, neural networks or SVM classifiers. Finally, another interesting subset of algorithms for remote sensing image processing which belong to this category are the coordinate transformations used in image orthorectification (through sensor models), map projection transforms and any analytical model-based in preparation for image resampling.

As the data transfer from the CPU to the GPU is relatively slow, the key factor in order to benefit from the GPU massively parallel architecture will be the complexity of the pixel-wise operation. In order to illustrate this, we have selected 2 classical algorithms.

1) *Algorithm description*: The first algorithm is the computation of the Normalized Difference Vegetation Index (NDVI) [18] which is a radiometric index which combines the red (N) and near infrared (NIR) reflectances in order to estimate the amount of vegetation:

$$NDVI = \frac{NIR - R}{NIR + R}. \quad (2)$$

The second algorithm is the spectral angle mapper (SAM), which computes, for each pixel of the image p with n_b bands,

```

__global__ void ndviKernel(float* pix, float* ndvi,
    int numBands, int indexRed, int indexNIR,
    int imageWidth)
{
    int x = blockIdx.x * blockDim.x + threadIdx.x;
    int y = blockIdx.y * blockDim.y + threadIdx.y;

    float nir = pix[numBands*(x + y*imageWidth)
        + indexNIR];
    float red = pix[numBands*(x + y*imageWidth)
        + indexRed];
    ndvi[x + y*imageWidth] = (nir - red)/(nir + red);
}

```

Fig. 3. Naive kernel example for NDVI: unfortunately, the amount of processing for each pixel is too low to get any gain from GPU

the spectral angle with respect to a reference pixel r . The spectral angle is defined as:

$$SA = \arccos \left(\frac{\sum_{b=1}^{n_b} r(b) \cdot p(b)}{\sqrt{\sum_{b=1}^{n_b} r(b)^2 \sum_{b=1}^{n_b} p(b)^2}} \right), \quad (3)$$

b being the spectral band, r is the reference pixel and p the current pixel. The interest of evaluating the SAM is that its computation is more costly due to the square root and the arccos function.

2) *Implementation details*: The straightforward way to implement these pixel processing algorithms is to get each thread of the GPU to process one pixel. In this case, the kernel is very simple as shown in Fig. 3.

Unfortunately, in this case, the GPU processing appears to be slower than the CPU by about 20% (Fig. 4). A quick profiling of this case shows that the kernel spends about 92% of its time for memory transfer and only 8% for the processing. This shows clearly that the NDVI computation is too simple to get any benefit by itself from the GPU architecture.

On the other hand, the computation cost for the spectral angle (Eq. 3) is higher than the NDVI. In this case, we start to see some gain from using the GPU, but it is not yet convincing (Fig. 5): the GPU is faster than the CPU using one thread, but comparable to the CPU fully using its 8 threads.

3) *Results*: Figure 4 shows the computation time for the NDVI for image size from 1000×1000 to 6000×6000 . Due to the significance of the IO operations compared to the computation, the execution time displays a large variance; average timing and the standard deviation for these timings on at least 20 runs are represented. As one can observe, the NDVI computation is too simple to be a good candidate for GPU optimization on its own. Actually, the GPU version is slower than the CPU versions. However, the multithreaded CPU implementation brings some benefit with respect to the single-threaded one but the improvement is also limited and much lower than the factor 8 expected.

Figure 5 shows the same kind of simulation for the spectral angle computation. Here, the parallel implementations are much more efficient than the single threaded one, but the spectral angle is still too simple for the GPU implementation to provide a speedup with respect to the CPU multithreaded one.

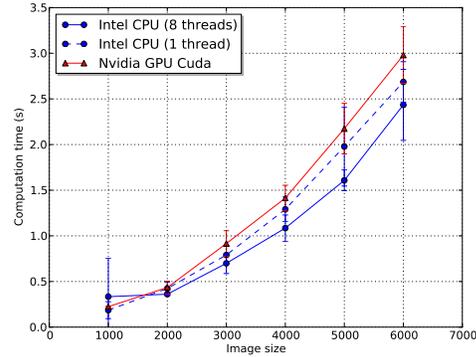


Fig. 4. Computation time for the NDVI for increasing image sizes: NDVI is not a good candidate to benefit from a GPU implementation on its own.

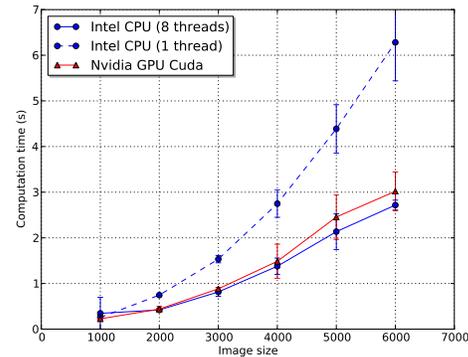


Fig. 5. Computation time for the spectral angle for image size from 1000×1000 to 6000×6000 : there is some improvement over the CPU implementation on a single thread, but it is still not sufficient to justify a GPU implementation.

B. Neighborhood-based processing

The second category of algorithms we are interested in is the ones which use a pixel and its close neighbors (on a regular grid) in order to compute the output value for a single pixel. This category includes many image processing tasks such as linear filtering by convolution, but also non-linear filtering (median, mathematical morphology), local statistics, etc.

Texture estimations using the Grey-Level Co-occurrence Matrices (GLCM) are also in this category, but they are more computationally intensive since 2 shifted neighborhoods are used, so they are very interesting candidates for GPU implementation.

1) *Algorithm description*: We choose here another particular case where the local computation is made using several pixel neighbors: fine image correlation. This technique is used for disparity map estimation between stereo image pairs [19]. Let I be the reference image and J be the secondary image,

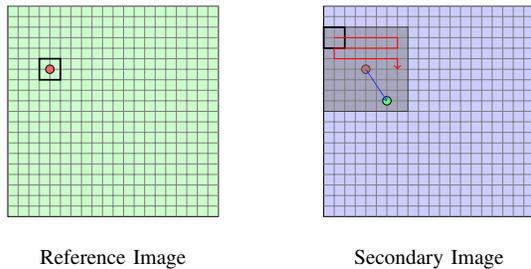


Fig. 6. Illustration of the fine correlation process

which are supposed to be roughly superimposable, one is interested in finding the local shift $(\Delta x, \Delta y)$ between small image patches which maximizes the correlation coefficient:

$$\rho_{I,J}(\Delta x, \Delta y) = \frac{\sum_{x,y} I(x,y)J(x + \Delta x, y + \Delta y)}{\sqrt{\sum_{x,y} I(x,y) \sum_{x,y} J(x + \Delta x, y + \Delta y)}} \quad (4)$$

This processing is applied for every shift in a given exploration area and for every pixel in the image (see Figure 6).

2) *Implementation details:* This particular problem of the estimation of a disparity map by fine correlation poses several issues for the implementation on GPUs and the adaptation is not as straightforward as the previous examples. On the other hand, the amount of computation per pixel is much greater than in the previous cases, so the potential gain is significant.

An adaptation of a similar algorithm to perform the auto-correlation of an image is given in [20], but the search window size was specific to their problem and involved different trade-offs.

The first question is which part of the processing should be implemented on the GPU. After profiling the CPU version of the algorithm, it appeared that 95% of the time was spent in computing the correlation for each shift. Producing the correlation map for each pixel seems to be the ideal part to be implemented.

The second question concerns the degree of flexibility required in the implementation. Of course, we do not want any limit concerning the size of the image to be processed. But given the streaming process described in II-A, it is not an issue. However, we also do not want a restriction on the size of the patch used to compute the local correlation or on the size of the exploration window.

The approach chosen for this case is that each thread will compute the correlation value corresponding to one displacement. Each block of threads will compute the correlation map for one pixel. Of course as there is no restriction in the size of the search window, the number of possible displacements for one pixel (which is computed by one thread block) can be greater than the maximum number of threads in one block (which is currently limited to 512). To go around this limitation, when the window search size is too big, the whole correlation map is computed through multiple kernel launches with different parameters.

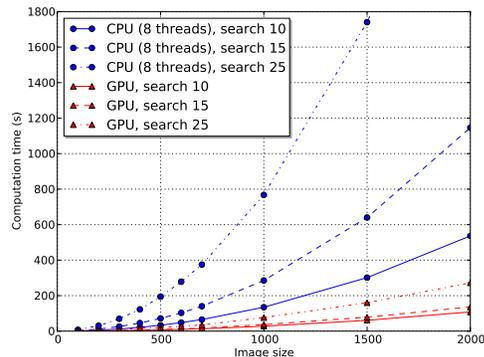


Fig. 7. Computation time of the fine correlation for different hardware for different search radius size and for a image size from 100×100 to 700×700 pixels: the GPU version is 8 to 12 times faster that the multithreaded CPU version fully using 8 threads.

There is an issue with the matching window which can extend outside of the image. One way to solve the problem is to make costly checking for each access directly in the GPU kernel. Here we avoid the issue altogether by using the strategy presented in II-A which ensure that all accesses will be valid.

The final result (Fig. 7) shows a speedup of 8 to 12 times compared to the CPU implementation for the whole fine registration process. The correlation computation is no longer the limiting factor and further improvement would require improving the interpolation process to find subpixel displacements.

3) *Results:* Figure 7 shows the computation time for the fine correlation for different image size and different radius for the search windows. A search window of radius 25 means that 2601 (51×51) different possible displacements will be explored for each pixel. In all cases, the size of the patch was fixed to 11×11 pixels. The CPU time corresponds to the multithreaded version, making full use of the processor. Here the difference between the GPU version and the CPU version is significant. For example, a processing that takes 4 min 30 s on the GPU takes more than 51 min on the CPU. It is worth noting that the correlation computation part, the only one implemented on the GPU here, used to represent 95% of the total computation time. The execution time for this part has been reduced by a factor of 20 and now represents only 50% of the total computation time. To obtain significant further improvements it is required to work on other part of the computation such as the correlation map interpolation (cf. Amdahl's law in equation (1)).

C. Irregular or non-local processing

This third class of algorithms consists of the cases where the pixels we are interested in are in irregular positions or represent only a small percentage of the pixels of the image and the computation to be performed is very demanding.

In these cases, the amount of data to be transferred to the processing unit is small and the computing cost is large.

In remote sensing image analysis, some examples of these algorithms are: irregular interpolation by thin plate splines, histogram kernel estimation, Voronoi/Delaunay triangulations and vector object processing (operations on polygons or polylines yielded by segmentation and detection algorithms).

1) *Algorithm description*: One example of such processing is point density estimation. This program estimates the point density for every point in an image for a given set of points. Several applications use this density estimation: we can mention point feature extraction such as SIFT [21], permanent scatterers in SAR images [22], etc. One example of possible output for this process is illustrated in Fig. 8.

The estimation is done using a Gaussian kernel as in Eq. (5) where $d_{p,h}$ denotes the distance between the pixel p where the density is computed and the point h of the set of points.

$$\rho_p = \frac{1}{2\pi\sigma^2} \sum_h e^{-\frac{d_{p,h}^2}{2\sigma^2}} \quad (5)$$

This equation means that the density is spread around point h following a Gaussian model.

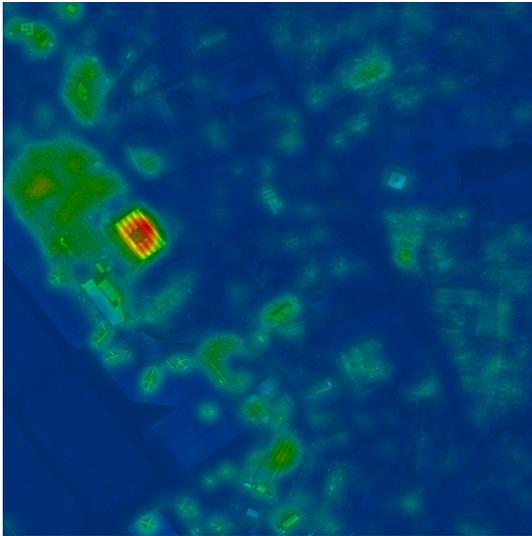


Fig. 8. Example of possible output of a point density estimation: in this example, the points correspond to SIFT detections. About 10000 points are detected on this image.

The decision to optimize this problem was taken after noticing that the CPU performances of the original version of the program were not satisfying, taking hours to provide a density map of permanent scatterers.

Another example is the learning step of SVM classifiers [23]. As mentioned in section III-A, SVM classification belongs to pixel-based processing, since class prediction for a given pixel is only a matter of a few scalar products with support vectors. Prior to SVM classification, SVM learning is

the task of identifying these support vectors into a training set and involves several iterations over a set of training examples which can be located anywhere in an image. This makes the SVM learning algorithm falls into the class of irregular of algorithms.

2) *Implementation details*: For the density estimation example, there are two different approaches to compute the final image. One is to go through the set of points and for each of them compute their impact on the whole image using an accumulator. The other approach is to go through all pixels and, for each of them, sum up the impact of all points in the point set. Depending on the relative size of the image and the point set, one or the other can be privileged in the case of CPU implementation. The first one corresponds to a *scatter* approach while the second one is a *gather* approach.

The *gather* approach is one where the computation of one output pixel is done at once by a single computation unit, gathering information from several positions from the input. The *scatter* approach is the reverse, when input are accessed once by a single computation unit and their impact is reported to the relevant output pixel which works as an accumulator.

GPUs are more suited to the *gather* approach and this is the one selected here: for each pixel, we go through the point set and compute the impact of each point to this pixel. The other advantage is that this approach is perfectly suitable for the pipeline model described in the previous sections.

This simple approach, where one thread processes one pixel, iterating over the list of points already provides an impressive gain compared to the CPU version. However, using additional features available from GPUs, additional performance gains are possible.

The first improvement is to use the constant memory of the GPU. Accesses to the constant memory are much faster when they are synchronized between the threads. Constant memory is the ideal candidate to store the point coordinates as they are frequently accessed by each thread.

Another source of improvement is to factorize some computations that are common to different threads, thus reducing the total amount of computations to be performed. When computing the distance between one point and one pixel, the computation is done for the x component and the y component. The y component will be the same for all the points of a line. By using one thread to process not only one point, but several points on the same line, this part of the computation can be done only once. A trade-off has to be made to keep the number of threads high enough to use all the computation units of the GPU and not to use too many registers (as that would reduce the number of threads that can run concurrently). In the present example, each thread processes 4 consecutive points on the same line.

In this case, as shown in Figure 10, the improvement over the GPU is impressive: 130 times faster than the multithreaded CPU version.

Regarding the SVM learning problem, the CPU implementation relies on LibSVM [24], a widely known library to perform SVM classification and regression, while the GPU version was handled by cuSVM [25], a CUDA implementation of SVM restricted to regression and two-class classification

```

__constant__ float pt_c[CHUNK_SIZE*2];

global void pointDensityKernel(float* pix,
int numPoint, int originX, int originY,
float spacingX, float spacingY,
int imageWidth, int radiusSq)
{
    int x1 = blockDim.x*blockDim.x*PIX_PER_THREAD
        + threadIdx.x;
    int x2 = x1 + blockDim.x;
    int x3 = x1 + 2*blockDim.x;
    int x4 = x1 + 3*blockDim.x;
    int y = blockDim.y*blockDim.y + threadIdx.y;

    float accum1 = 0.0f;
    float accum2 = 0.0f;
    float accum3 = 0.0f;
    float accum4 = 0.0f;
    for (int k = 0; k < numPoint; k++)
    {
        float ptX = pt_c[2 * k];
        float ptY = pt_c[2 * k + 1];
        float pixX1 = x1 * spacingX + originX;
        float pixX2 = x2 * spacingX + originX;
        float pixX3 = x3 * spacingX + originX;
        float pixX4 = x4 * spacingX + originX;
        float pixY = y * spacingY + originY;
        float disty_sq = (ptY - pixY) * (ptY - pixY);
        float distsq;
        distsq = (ptX - pixX1) * (ptX - pixX1)
            + disty_sq;
        accum1 += __expf(-distsq/radiusSq/2);

        distsq = (ptX - pixX2) * (ptX - pixX2)
            + disty_sq;
        accum2 += __expf(-distsq/radiusSq/2);

        distsq = (ptX - pixX3) * (ptX - pixX3)
            + disty_sq;
        accum3 += __expf(-distsq/radiusSq/2);

        distsq = (ptX - pixX4) * (ptX - pixX4)
            + disty_sq;
        accum4 += __expf(-distsq/radiusSq/2);
    }
    pix[x1+y*imageWidth] += accum1/NORMALIZATION;
    pix[x2+y*imageWidth] += accum2/NORMALIZATION;
    pix[x3+y*imageWidth] += accum3/NORMALIZATION;
    pix[x4+y*imageWidth] += accum4/NORMALIZATION;
}

```

Fig. 9. CUDA kernel for the point density computation: each thread computes the density value for 4 points of the same line simultaneously. The OpenCL kernel is very similar.

with a Gaussian kernel. In this case, all that is needed is to fit the LibSVM and cuSVM calls into the Orfeo Toolbox framework. In each case, the learning and classification steps are implemented in separated filters and the classification filters are given streaming capabilities. Parameters from both libraries are tuned to match results as closely as possible.

3) *Results*: Figure 10 presents the results obtained by the implementation of the point density computation. Here, the gain is so significant that a logarithm scale is used to show the CPU and the GPU time on the same plot. The speedup from the GPU version is about 130 times. For this particular case, two implementations for CUDA and OpenCL are realized. It appears that for this particular case the OpenCL implementation is about 45% slower than the CUDA implementation.

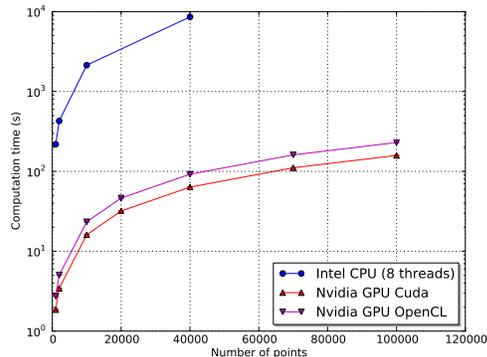


Fig. 10. Computation time of the point density for different hardware for different number of points for an image of 36 MPixels: the GPU version is about 130 times faster than the multithreaded CPU version (Note that the timing is on logarithm scale).

However, the gain compared to the CPU is still impressive, and with respect to the CUDA implementation, there is the advantage of supporting a wider range of hardware including CPUs and GPUs. Computation speed improvements of such an order of magnitude open a whole range of possibilities where interactive processing becomes convenient. Advantage of supporting a wider range of hardware including CPUs and GPUs. Such order of magnitude in computation speed improvement opens a whole range of possibilities where interactive processing becomes convenient.

Regarding the SVM learning and classification problem, both CPU and GPU filters have been used to perform change detection on a pair of registered SPOT5 images in the context of a severe flooding event. This dataset contains 2320x4320 pixels with 3 spectral bands for each date. A training mask of 62676 pixels denoting change and no-change areas was used to train the SVM. The learning step took 572 seconds with the CPU implementation, while only 1.35 seconds were necessary for the GPU implementation to converge, which is tremendously faster (more than 400 times), although we should mention that the CPU version of the learning step is mono-threaded in this case. As expected, being a simpler algorithm, the classification step shows smaller time improvements, with 489 seconds for the multithreaded CPU version and 146 seconds for the GPU one – only about 3 times faster.

IV. DISCUSSION AND PERSPECTIVES

As we have seen in the previous sections, some classes of algorithms can benefit tremendously from a GPU implementation. Typically, these algorithms can be identified as algorithms that do mostly local processing (limited distance impact) and intensive computation for each pixel. We have seen for example that the SAM is at the lower bound in terms of computation to make it valuable to implement on GPU for an image with 4 bands (fig. 5). In that particular case, the limit appears to be a few hundred operations (mainly

due to the trigonometric function). Below this limit, the time spent transferring the data from the CPU memory to the GPU overcomes any benefit in the computation speed.

Above this limit, the benefits can be very important. In some cases, they can be so important that they could change the way the human interacts with the process. When the process takes hours, or even minutes, it is not conceivable to have the user sit down in front of the computer, waiting for the results. In this case, the data will be processed from end to end and the user will exploit the final result. If there is a need to adjust some parameters, the data will be reprocessed.

One drawback of this classical scheme is that it tends to limit the interactivity between the human and the algorithm. When we reach a situation where we can process a screen size area in about a second, it becomes possible to do the processing in real time. In this situation, any modification of any parameter will trigger immediate feedback: the user is able to interact much more with the algorithm. The human can become a real part of the processing chain.

This can lead to an improvement of the classic processing chains in use, but it can also lead to the development of new paradigms. One obvious example is the application of active learning to remote sensing problems [26]. Benefiting from the major speed-up brought by the GPU implementation of SVM learning (see section III-C3), the training samples selection step could change drastically: near real-time feedback on the pertinence of selected samples and on primitive-wise confidence of the classifier becomes achievable. Other common remote sensing image processing tasks could benefit from immediate quality feedback, such as image co-registration for instance.

V. CONCLUSIONS

As demonstrated in this paper, adapting the most expensive part of a processing pipeline to benefit from the processing power of GPUs is quite simple. With a minimum investment (hardware cost is around US\$ 200, the software used here is free and open source), performance gains can attain 10 to 400 times on the critical portion of the processing.

One of the main shortcomings, which is the relatively slow computation in double precision – important for some scientific computations – has been addressed by the new Fermi architecture released by Nvidia in April 2010. We will definitely witness an increasing number of GPU implementations for remote sensing processing algorithms in the near future.

Still, benefiting from this massive speed-up requires one to carefully select those algorithms which fit well in the GPU computing architecture, identify the critical sections to optimize, and have a close look at how things are implemented.

All this complexity should remain hidden to the end-user, which is exactly what the high level of abstraction provided by the Orfeo Toolbox framework allows us to do. Further improvements can be made in that direction by proposing a mechanism to switch seamlessly from CPU to GPU versions of algorithms depending on available hardware. Another interesting perspective would be to run GPU-enabled filters on GPU blades, which gather several GPU devices on a single hardware.

The source code corresponding to the examples presented in this paper used to generate the results is available for download from the Orfeo Toolbox web site (<http://www.orfeo-toolbox.org/OTB-GPU>). Most of it will be integrated in the upcoming releases of the library.

REFERENCES

- [1] K. Asanovic, R. Bodik, B. Catanzaro, J. Gebis, P. Husbands, K. Keutzer, D. Patterson, W. Plishker, J. Shalf, S. Williams, and K. Yelick, "The landscape of parallel computing research: A view from Berkeley," University of California, Berkeley, Tech. Rep., Dec. 2006.
- [2] D. Blythe, "Rise of the graphic processor," *Proceedings of the IEEE*, vol. 96, no. 5, May 2008.
- [3] "The Orfeo Toolbox software guide," <http://www.orfeo-toolbox.org>, 2010.
- [4] E. Christophe and J. Inglada, "Open source remote sensing: Increasing the usability of cutting-edge algorithms," *IEEE Geoscience and Remote Sensing Newsletter*, Mar. 2009.
- [5] "ITK, the Insight Toolkit," <http://www.itk.org>.
- [6] J. R. Schott, *Remote sensing: the image chain approach*. Oxford University Press US, 2007.
- [7] M. J. Flynn, "Some computer organizations and their effectiveness," *IEEE Transactions on Computers*, vol. 21, no. 9, pp. 948–960, Sep. 1972.
- [8] J. L. Manferdelli, N. K. Govindaraju, and C. Crall, "Challenges and opportunities in many-code computing," *Proceedings of the IEEE*, vol. 96, no. 5, May 2008.
- [9] N. Satish, M. Harris, and M. Garland, "Designing efficient sorting algorithms for manycore GPUs," in *IEEE International Symposium on Parallel & Distributed Processing*, May 2009.
- [10] G. M. Amdahl, "Validity of the single-processor approach to achieving large scale computing capabilities," in *AFIPS Conference Proceedings*, vol. 30, Apr. 1967, pp. 483–485.
- [11] J. Setoain, M. Prieto, C. Tenllado, and F. Tirado, *High performance computing in remote sensing*. Chapman & Hall/CRC, 2007, ch. Real-time onboard hyperspectral image processing using programmable graphic hardware, pp. 411–451.
- [12] *The OpenCL Specification*, Khronos Group Std. 1.0, Dec. 2008.
- [13] D. Kirk and W.-M. W. Hwu, *Programming Massively Parallel Processors: A Hands-on Approach*. Elsevier Science & Technology, 2010.
- [14] P. Harish and P. J. Narayanan, *High Performance Computing – HIPC 2007*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2007, ch. Accelerating large graph algorithms on the GPU using CUDA.
- [15] S. Che, M. Boyer, J. Meng, D. Tarjan, J. W. Sheaffer, and K. Skadron, "A performance study of general-purpose applications on graphics processors using CUDA," *Journal of Parallel and Distributed Computing*, vol. 68, no. 10, pp. 1370–1380, Oct. 2008.
- [16] V. W. Lee, C. Kim, J. Chhugani, M. Deisher, D. Kim, A. D. Nguyen, N. Satish, M. Smelyanskiy, S. Chennupati, P. Hammarlund, R. Singhal, and P. Dubey, "Debunking the 100X GPU vs. CPU myth: an evaluation of throughput computing on CPU and GPU," pp. 451–460, Jun. 2010.
- [17] K. Karimi, N. G. Dickson, and F. Hamze, "A performance comparison of CUDA and OpenCL," *CoRR*, vol. abs/1005.2581, 2010.
- [18] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sensing of Environment*, no. 8, pp. 127–150, 1979.
- [19] J. Inglada and A. Giros, "On the possibility of automatic multi-sensor image registration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 10, Oct. 2004.
- [20] P. J. Lu, H. Oki, C. A. Frey, G. E. Chamitoff, L. Chiao, B. J. Au, M. Christiansen, A. B. Schofield, D. A. Weitz, P. J. Lu, D. A. Weitz, H. Oki, C. A. Frey, G. E. Chamitoff, L. Chiao, E. M. Fincke, D. M. Tani, P. A. Whitson, J. N. Williams, W. V. Meyer, R. J. Sicker, B. J. Au, M. Christiansen, and A. B. Schofield, "Orders-of-magnitude performance increases in GPU-accelerated correlation of images from the international space station," *Journal of Real-Time Image Processing*, 2009.
- [21] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, Feb. 2004.
- [22] A. Ferretti, C. Prati, and F. Rocca, "Permanent scatterers in SAR interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 8–20, Jan. 2001.

- [23] V. Vapnik, *Statistical learning theory*. John Wiley and Sons, NewYork, 1998.
- [24] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- [25] A. Carpenter, "cuSVM: A CUDA implementation of support vector classification and regression," <http://patternsonscreen.net/cuSVMDesc.pdf>, Jan. 2009.
- [26] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification." *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2218–2232, Jul. 2009.



Emmanuel Christophe received the Engineering degree in Telecommunications from École Nationale Supérieure des Télécommunications de Bretagne, Brest, France and the DEA in Telecommunications and image processing from University of Rennes 1 in 2003. In October 2006, he received the PhD degree from Supaero and University of Toulouse in hyperspectral image compression and image quality. He has been a visiting scholar in 2006 at Rensselaer Polytechnic Institute, Troy, NY, USA. From 2006 to 2008, he was a research engineer at CNES, the French Space Agency, focusing on information extraction for high resolution optical images. Between 2008 and 2010, he moved to Singapore at CRISP, National University of Singapore, where he was tackling new challenges for remote sensing in tropical areas. He is now with Google Inc. in California.



Julien Michel received the Telecommunications Engineer degree from the École Nationale Supérieure des Télécommunications de Bretagne, Brest, France, in 2006. From 2006 to 2010, he has been with Communications et Systèmes, Toulouse, France, where he has been working on studies and developments in the field of remote sensing image processing. He is now with the Centre National d'Études Spatiales (French Space Agency), Toulouse, France, where he is in charge of the development of image processing algorithms for the exploitation of Earth observation images, mainly in the field of Very High Resolution image analysis.



Jordi Inglada received the Telecommunications Engineer degree from both the Universitat Politècnica de Catalunya, Barcelona, Spain, and the École Nationale Supérieure des Télécommunications de Bretagne, Brest, France, in 1997 and the Ph.D. degree in signal processing and telecommunications in 2000 from Université de Rennes 1, Rennes, France. He is currently with the Centre National d'Études Spatiales (French Space Agency), Toulouse, France, working in the field of remote sensing image processing at the CESBIO laboratory. He is in charge of the development of image processing algorithms for the operational exploitation of Earth observation images, mainly in the field of multitemporal image analysis for land use and cover change. Dr. Inglada is an Associate Editor of the *IEEE Transactions on Geoscience and Remote Sensing*.