# Génération de modèles numériques de surface et détection de changements 3D à partir d'imagerie satellite stéréoscopique très haute résolution

Cyrielle Guérin

▶ **To cite this version:**

Cyrielle Guérin. Génération de modèles numériques de surface et détection de changements 3D à partir d'imagerie satellite stéréoscopique très haute résolution. Modélisation et simulation. Université René Descartes - Paris V, 2014. Français. NNT : 2014PA05S003 . tel-00953485

**HAL Id: tel-00953485**

**https://theses.hal.science/tel-00953485**

Submitted on 28 Feb 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THÈSE DE DOCTORAT DE**

**l'UNIVERSITÉ PARIS DESCARTES**

Spécialité : **Informatique**

École doctorale Informatique, Télécommunications et Électronique (Paris)

Présentée par

**Cyrielle GUERIN**

Pour obtenir le grade de

**DOCTEUR de l'UNIVERSITÉ PARIS DESCARTES**

# Génération de Modèles Numériques de Surface et Détection de Changements 3D à Partir d'Imagerie Satellite Stéréoscopique Très Haute Résolution

soutenue le 18 février 2014

devant le jury composé de :

| | |
|---|---|
| M. Pascal Monasse | Rapporteur |
| M. Laurent Polidori | Rapporteur |
| M. Andres Almansa | Examinateur |
| M. Bruno Vallet | Examinateur |
| Mme. Nicole Vincent | Examinatrice |
| M. Renaud Binet | Encadrant |
| M. Marc Pierrot-Deseilligny | Directeur de thèse |

# Remerciements

Avant tout, mes remerciements vont à Renaud Binet et Marc Pierrot-Deseilligny, initiateurs et encadrants de ce sujet de thèse qui m'a captivé pendant ces trois années. Merci à Marc de m'avoir fait découvrir MicMac et la programmation dynamique que j'utilise maintenant partout ! Merci Renaud pour ton enthousiasme, tes nombreuses idées au cours de la thèse et surtout pour m'avoir beaucoup appris dans le domaine de l'imagerie optique.

Je tiens aussi à remercier la DGA d'avoir financé ces travaux. Participer à ce programme a été très motivant.

Merci à Philippe Blanc, toujours de très bon conseil et à Bruno Vallet pour le travail très enrichissant que nous avons réalisé ensemble.

Merci à toute l'équipe du CALTECH pour m'avoir invitée dans vos murs et fait découvrir d'autres pistes de recherche ! Merci Sébastien pour toute l'aide que tu m'as apportée durant ce stage mais aussi pour m'avoir fait visiter ce petit bout des États–Unis ! J'espère que notre collaboration ne s'arrêtera pas là.

Je tiens aussi à remercier les enseignants du Master 2 de l'EGID et surtout Samia Boukir et Nesrine Chehata. Merci de m'avoir incitée à continuer dans la recherche quand je n'étais qu'étudiante en master ! J'espère Nesrine qu'on se retrouvera encore souvent pour des conférences, IGARSS a été un très bon moment.

Je n'oublie sûrement pas tout le labo TSE avec lequel j'ai travaillé pendant ces trois années. Merci bien sûr à Philippe pour m'avoir accueillie dans ce labo et m'avoir encouragée tout au long de la thèse. Un grand bravo à Adrien pour m'avoir côtoyée dans le même bureau pendant plus de 2 ans ! Tu resteras mon irremplaçable collègue de bureau et ami.

Merci à Thierry d'avoir toujours eu confiance en moi et de m'avoir soutenu tous les jours, bon

ou moins bon...Merci à Béatrice pour nos collaborations sur divers projets. Merci à Philippe, Rodolphe et Eric pour avoir accepté de relire cette thèse. Finalement merci à vous tous, ça a été un réel plaisir de travailler avec vous dans ce labo.

Merci aussi à tous ceux que j'ai rencontrée durant cette thèse. Merci à Franck pour d'inoubliables conversations et fous rires si décompressants sur la ligne C3 ! Je te souhaite le meilleur pour ta nouvelle carrière et j'espère qu'on ne se perdra pas de vue. Bien sûr, je remercie Boris, rencontré de façon si inopiné et qui s'est révélé être un ami très précieux, merci d'avoir passé ces heures à m'épauler pour mon article, tu m'as beaucoup apporté. You rock !

Merci enfin à mes parents et à mes soeurs, vous m'avez soutenue durant toutes ces années d'études. J'ai pu compter sur votre soutien inflexible lors de mes plus grandes périodes de stress. Du fond du coeur, merci. C'est à vous que je dédie cette thèse.

# Résumé

L'imagerie satellite permet aujourd'hui l'acquisition d'un nombre croissant de données dont l'augmentation des résolutions spatiale et temporelle permet de caractériser de plus en plus finement une scène et son évolution. Dans ce contexte, les méthodes de détection des changements apparus entre deux scènes sont particulièrement étudiées. Elles sont généralement basées sur les différences radiométriques entre les images. Cependant, ces méthodes s'avèrent souvent peu robustes à des changements radiométriques non pertinents tels que ceux induits par la variation des conditions d'acquisition des images.

L'objectif de cette thèse est ainsi de développer une méthode alternative, basée sur la recherche des changements d'élévation de la scène. L'élévation représente en effet une information pertinente et adaptée, notamment dans un contexte de détection des changements de type urbain (construction, destruction ou modification d'infrastructures).

Pour répondre à des besoins en analyse d'image qui nécessitent des résultats rapides et fiables, la méthode que nous proposons est une chaîne de traitements complète et automatique basée sur l'exploitation de couples d'image satellites stéréoscopiques très haute résolution permettant la génération et la comparaison de Modèles Numériques de Surface (MNS). Afin de limiter les fausses alarmes de changements dues aux erreurs liées à la génération des MNS, une étape clé de cette thèse a consisté à augmenter la précision des MNS, notamment à travers la prise en compte des zones d'occlusions et de mauvaise corrélation.

La méthode de génération des MNS à ainsi été améliorée et une technique innovante de fusion des deux MNS provenant du même couple d'images a été développée. La comparaison des MNS générés avec un MNS LiDAR montre que notre approche permet une nette augmentation de la qualité des MNS, les erreurs de corrélation sont réduites tandis que les zones d'occlusion sont

précisément localisées.

La méthode de détection des changements d'élévation est, quant à elle, basée sur une labellisation par optimisation des pixels du MNS différentiel calculé à partir des MNS produits à chaque date. Cette étape permet de mettre en évidence les vrais changements de la scène parmi le bruit résiduel des MNS.

Les résultats obtenus sur différents sites testés montrent que plus de 80% des changements de taille supérieure à 15 pixels x 15 pixels (ou 100 $m^2$ avec des images très haute résolution) sont détectés par notre méthode, avec moins de 20% d'erreurs. Nous montrons cependant que ces résultats dépendent principalement du paramètre de régularisation de la détection des changements, qui contrôle le taux de fausses alarmes par rapport au taux de bonnes détections du résultat.

# Abstract

The growing amount of satellite data, increasingly resolved spatially and temporally, represents a high potential of information allowing the accurate characterization of the evolution of an area of interest. For this reason, automatic analysis techniques such as change detection methods are widely investigated. Most of them are based on radiometric changes between remote sensed optical images. These methods are however very sensitive to a significant number of irrelevant changes such as those due to the variation of the geometrical conditions between two different acquisitions.

The objective of this work is then to develop an alternative method based on the elevation change detection. The advantage of using the elevation is that this information is particularly relevant and well adapted in a context of urban monitoring where the elements of interest correspond to buildings that can be constructed, modified or destroyed between two dates.

In order to satisfy new needs in image analysis which require quick and reliable results, our method is a complete and automatic processing flow based on the analysis of high resolution satellite stereoscopic couples and the generation of Digital Surface Models (DSM).

Stereoscopic DSMs, however, generally suffer from a high number of correlation errors leading to false alarms in the final change detection map. One of the main contribution of this work consisted in increasing the DSM accuracy, especially through a better handling of the occlusion and miss-correlation areas. For this purpose, the image matching technique has been improved and all DSMs computed from the same stereoscopic couple are then fusioned through a new approach, based on an optimization method. The comparison between our DSM with a LiDAR-based DSM indicates that our method largely improves the DSM quality, the amount of correlation errors is decreased while the occlusion areas are accurately localized.

The change detection method itself is based on the labelization of the pixels of the differential

DSM computed from the DSMs generated at each date of interest. This step, performed through another optimization process, enables to bring forward the relevant changes among the residual noise of the DSMs. The results, obtained for several experimental areas, show that more than 80% of the changes larger than 15 pixels x 15 pixels (100 m $m^2$ with high resolution images) are detected with our method, with less than 20% of false alarms. We also show that these results mainly depend on the regularization parameter which controls the balance between the amount of false alarms towards the amount of true detections in the final results.

# Table des matières

**Introduction Générale** **13**

**1 Études bibliographiques et stratégie adoptée** **17**

1.1 La détection de changements : état de l'art . . . . . 18

1.1.1 Méthodes de détection des changements radiométriques . . . . . 18

1.1.2 Méthodes alternatives : détection des changements d'élévation . . . . . 19

1.2 Méthodes de restitution 3D : état de l'art . . . . . 23

1.2.1 Mise en correspondance d'images par méthode d'optimisation . . . . . 25

1.2.2 Caractéristiques des corrélateurs . . . . . 27

1.2.3 Les stratégies de mise en correspondance . . . . . 29

1.2.4 Les outils de restitution 3D existants . . . . . 33

1.3 Description générale de la chaîne de traitements développée . . . . . 35

1.4 Capteurs utilisés et données d'entrée . . . . . 37

1.5 Conclusion . . . . . 39

**2 Outils et méthodes nécessaires à la chaîne de traitements** **41**

2.1 Co–localisation des images : utilisation d'un outil de spatio–triangulation . . . . 42

2.1.1 Contexte et objectif . . . . . 42

2.1.2 Principe de la spatio–triangulation . . . . . 43

2.2 Outil pour la résolution de problèmes de labellisation . . . . . 47

2.2.1 Trois problèmes de labellisation . . . . . 47

2.2.2 Les méthodes d'optimisation existantes . . . . . 49

2.2.3 Description de l'algorithme sélectionné . . . . . 55

2.3 Conclusion . . . . . 59

**3 Génération de Modèles Numériques de Surface sur une grille régulière à partir de couples stéréoscopiques 61**

3.1 Description de la stratégie de mise en correspondance adoptée . . . . . 62

3.1.1 Technique de mise en correspondance . . . . . 62

3.1.2 Fonction d'optimisation et régularisation . . . . . 63

3.1.3 Prise en compte des zones d'occlusion . . . . . 65

3.2 Développement d'une méthode de basculement des MNS sur une grille géographique régulière . . . . . 65

3.2.1 Interpolation des points de la grille régulière . . . . . 66

3.2.2 Conditions au calcul de l'élévation d'un point . . . . . 67

3.3 Amélioration des MNS : développement d'une méthode de fusion des MNS . . . 72

3.3.1 Erreurs des MNS asymétriques . . . . . 72

3.3.2 Méthode de fusion des MNS asymétriques . . . . . 74

3.4 Qualité et précision des MNS obtenus . . . . . 82

3.4.1 Co–localisation LiDAR et MNS . . . . . 82

3.4.2 Critères de qualité . . . . . 83

3.5 Application au cas de la multi–stéréoscopie . . . . . 87

3.6 Conclusion . . . . . 91

**4 Développement d'une méthode de détection des changements d'élévation 93**

4.1 Calcul et analyse du MNS différentiel . . . . . 94

4.2 Méthode de détection des changements . . . . . 97

4.3 Conclusion . . . . . 100

**5 Expérimentation et discussion des résultats de la méthode de détection des changements d'élévation 101**

5.1 Présentation des zones de test . . . . . 102

5.1.1 Phoenix . . . . . 102

5.1.2 Christchurch . . . . . 104

5.1.3 Région de Tohoku : analyse d'une zone catastrophée . . . . . 105

5.2 Métriques d'analyse des résultats de la détection de changements . . . . . 109

5.2.1 Variables calculées . . . . . 109

5.2.2 Métriques d'analyse . . . . . 111
5.3 Analyse de sensibilité des résultats obtenus . . . . . 112
5.3.1 Influence des paramètres de la méthode sur les résultats . . . . . 113
5.3.2 Impact de la qualité des données d'entrée . . . . . 118
5.3.3 Intérêt d'une méthode globale : comparaison avec un filtrage local . . . . 121
5.3.4 Influence du paysage de la zone d'étude . . . . . 124
5.4 Analyse des résultats sur la zone de Sendai . . . . . 127
5.5 Conclusion . . . . . 131

**6 Conclusion générale** **133**
6.1 Conclusion . . . . . 133
6.2 Perspectives . . . . . 135

**Table des figures** **139**

**Liste des tableaux** **143**

**Bibliographie** **145**

# Introduction Générale

Depuis le lancement des premiers satellites dédiés à la télédétection spatiale, leur nombre, leur résolution et leur capacité d'acquisition n'ont de cesse d'augmenter. Aujourd'hui, Pléiades 1A et 1B fournissent des images à 70 cm de résolution tandis que WorldView-3 est attendu avec une résolution inférieure à 50 cm. Ce nombre de données, d'une qualité toujours meilleure, permet d'accéder à une grande quantité d'informations pour caractériser et analyser une scène et son évolution à de très fines échelles spatiales et temporelles.

Pour traiter ces nouveaux flux de données, les méthodes permettant l'analyse de séries multi–temporelles d'images sont alors particulièrement intéressantes et en fort développement. Les applications à ces études sont très nombreuses, à commencer par l'agriculture, pour caractériser l'évolution des sols, jusqu'aux géosciences, pour l'observation des glaciers en termes de volume, d'étendue ou de vitesse [1], pour la surveillance des volcans [2] ou encore pour les mesures des déformations tectoniques [3, 4].

L'augmentation drastique de la résolution des images permet aussi aujourd'hui une observation beaucoup plus fine des milieux urbains pour une meilleure analyse géographique et démographique des villes à travers la mise à jour des plans cadastraux ou autres bases de données [5] mais aussi pour la cartographie de dégâts matériels (en terme de bâtiments ou d'infrastructures détruits) après une catastrophe majeure telle qu'un séisme, un tsunami, un glissement de terrain, un événement météorologique ou encore une guerre [6]. Dans ce contexte, les méthodes de détection des changements, automatiques ou semi–automatiques, sont particulièrement avantageuses.

Les algorithmes de détection de changements sont traditionnellement basés sur l'étude

d'images radars ou optiques. En imagerie optique, la plupart des méthodes s'intéressent aux différences de réflectances entre les données acquises à des dates $t_1$ et $t_2$ [7]. Cependant, les résultats montrent que, dans un contexte de suivi des infrastructures urbaines, de nombreuses fausses alarmes ou changements non pertinents sont générés par ces méthodes de détection 2D. Ces alarmes sont généralement dues aux différentes conditions d'acquisition des images (conditions d'éclairement ou d'atmosphère, azimut solaire, saison ou encore angles d'acquisition des images) lors des prises de vue ainsi qu'à de nombreux changements radiométriques tels que l'évolution des sols, les rénovations de toitures ou de revêtement de sol.

Le travail de thèse présenté ici s'inscrit dans cette problématique de détection de changements. Dans le but de s'affranchir des changements radiométriques, nous nous intéressons aux changements d'élévation de type urbain (directement reliés au bâti) et dans un contexte de suivi temporel ou de crise majeure nécessitant des résultats fiables et rapides. Afin de répondre au mieux à cette problématique, la technique proposée est basée entièrement sur les différences d'élévation de la scène entre les deux dates d'intérêt. En effet, cette information, dont la variation en milieu urbain est généralement liée à des changements du bâti, est robuste à la plupart des changements 2D et est donc particulièrement adaptée à cet objectif [8].
Aujourd'hui, différents systèmes d'acquisition permettent d'accéder à l'information d'élévation d'une scène : les lasers aéroportés (Airborne Laser Scaning ou ALS), le LiDAR (Light Detection And Ranging), les satellites radar ou encore l'imagerie stéréoscopique, aérienne ou spatiale, à travers la génération de Modèles Numériques de Surface (MNS). Du fait de leur très haute résolution, leur large capacité d'acquisition et la rapidité de livraison et de traitement de ces données, les satellites optiques produisant des images stéréoscopiques apparaissent comme la meilleure réponse pour la détection des changements d'élévation. De plus, le coût de ces images est relativement faible comparé aux autres systèmes d'acquisition, voire nulle lors d'une crise majeure car ces images sont alors gratuitement mises à disposition des agences cartographiques à travers la Charte Internationale Espace et Catastrophes Majeures, signée par de nombreux pays.

L'objectif de ce travail est ainsi de développer et de tester, dans un contexte urbain, une méthode de détection de changements innovante et basée sur les changements d'élévation d'une scène entre deux dates. Afin de répondre de façon adaptée à la problématique de quantification

de dégâts après une catastrophe, la technique de détection de changements développée est une chaîne de traitements complète et automatique, débutant par le recalage des données brutes et fournissant, en sortie, une carte géoréférencée des changements détectés.

Au cours de cette thèse, nous détaillerons toutes les étapes de cette chaîne de traitements, leurs objectifs et leurs résultats.

Dans une première partie, nous montrerons l'intérêt de la recherche des changements d'élévation (ou 3D) par rapport aux changements radiométriques à travers un état de l'art de ces méthodes. Nous étudierons aussi les méthodes de génération de MNS à partir d'images stéréoscopiques (chapitre 1), étape primordiale de la chaîne de traitements.

Dans une seconde partie, nous décrirons les outils existants nécessaires à notre chaîne de traitements, qui consistent en un outil de recalage des images, étape indispensable à tout traitement géométrique des images et un outil de résolution de problèmes de labellisation par programmation dynamique (chapitre 2). Cet algorithme est en effet central pour trois étapes de calcul de notre chaîne de traitements.

Puis, dans le chapitre 3, nous détaillerons la méthode complète de génération des MNS. Cette méthode, qui débute par la mise en correspondance des images stéréoscopiques présente ensuite deux innovations majeures permettant d'améliorer la précision des MNS en vue de la détection des changements.

La technique de détection de changements 3D, basée sur une classification de la carte des différences d'élévation par une méthode de labellisation, sera finalement exposée dans le chapitre 4.

Enfin, la dernière partie aura pour objectif d'illustrer l'intérêt de la chaîne de traitements développée et d'analyser les résultats obtenus dans différentes scènes urbaines, que ce soit dans un contexte de mise à jour des bases de données ou pour la caractérisation et la quantification rapide des dégâts après une catastrophe majeure (chapitre 5).

# Chapitre 1

# Études bibliographiques et stratégie adoptée

Dans le domaine de la détection de changements par imagerie satellite, de nombreuses méthodes existent mais elles sont principalement basées sur l'exploitation de la radiométrie des images tandis que très peu de méthodes utilisent l'information 3D de la scène. C'est cette approche que nous avons choisi de développer dans notre chaîne de traitements car elle se révèle particulièrement pertinente pour la détection des changements de type bâti, dans un contexte urbain.

La chaîne de traitements proposée nécessite ainsi une étape de reconstruction 3D de la scène. Cependant, les contraintes liées aux acquisitions satellites rendent le calcul de la mise en correspondance des images particulièrement complexe et de multiples techniques ont été développées afin d'améliorer leur robustesse.

Dans une première partie de ce chapitre, nous présenterons quelques unes des méthodes classiques de détection de changements radiométriques. Nous détaillerons ensuite plus particulièrement les méthodes développées pour la détection des changements d'élévation et nous montrerons leur intérêt, notamment en milieu urbain. Puis, dans une seconde partie, nous analyserons les méthodes de restitution 3D existantes et l'intérêt de certaines techniques pour notre chaîne de traitements.

Enfin, dans une dernière partie, nous détaillerons toutes les étapes de notre chaîne de traitements et les données utilisées lors des phases expérimentales.

## 1.1 La détection de changements : état de l'art

### 1.1.1 Méthodes de détection des changements radiométriques

Les méthodes traditionnelles en détection de changements sont basées sur la comparaison temporelle des radiométries qui peut être effectuée directement à l'échelle du pixel. Un large inventaire des méthodes les plus connues est fourni par Radke [7].
Les méthodes les plus simples et rapides à implémenter sont celles utilisant les différences radiométriques ou les rapports (ratios) entre les pixels correspondant en $t_1$ et $t_2$ [7, 9, 10]. Cependant, ces méthodes nécessitent souvent des seuillages ou des règles de décisions et restent très sensibles au bruit des images, aux erreurs de recalage et surtout aux conditions d'acquisition des images, différentes entre $t_1$ et $t_2$, qui génèrent de nombreuses alarmes de changements, généralement non pertinentes.

D'autres méthodes, de type statistique, sont aussi employées à l'échelle du pixel [7, 9]. Le but est de décider, pour chaque pixel, s'il correspond à un changement ou non en fonction de l'hypothèse changement ou non changement qui décrit le mieux son intensité. Ces hypothèses sont posées à partir des variations d'intensité des pixels dans la série temporelle [7].
L'analyse par composantes principales (PCA) est aussi largement utilisée [10–12]. Elle permet de réduire la dimension spectrale des images aux composantes les plus riches en information et donc où les changements les plus importants ont la plus grande probabilité d'apparaître.
Cependant, les hypothèses, nécessaires pour la mise en oeuvre de ces méthodes, proviennent souvent d'une première analyse des changements recherchés et limitent ainsi l'automatisation de ces méthodes.

Les méthodes de classification des pixels sont aussi très développées. Deux méthodes sont principalement utilisées : la post classification et la classification directe multi–date [13].
La première consiste à classer les pixels, de façon supervisée ou non, sur chaque image indépendamment puis à comparer les classes. Le problème de cette approche est l'impact important d'une erreur de classe dans l'une des images sur le résultat final. Cette méthode est ainsi plutôt utilisée pour la détection de changements à grande échelle comme pour l'occupation des sols par exemple [7, 9, 10, 12].

La classification directe multi–dates nécessite la concaténation des images, puis la classification, souvent supervisée, des pixels selon leur valeur dans toutes les bandes afin de faire apparaître les changements [11]. Enfin, les images multi–spectrales permettent aussi l'utilisation d'outils tels que les Support Vector Machine (SVM) [14]. Les valeurs spectrales de chaque pixel sont alors représentées par un vecteur dans l'espace temporel, la magnitude et la direction de ces vecteurs permettent le classement des pixels en changement ou non changement dans l'image.
Ces méthodes de classification s'avèrent souvent plus robustes aux erreurs de recalage ou au bruit des images mais elle nécessitent souvent des processus supervisés.

Aujourd'hui, avec l'augmentation de la résolution spatiale des images, l'intégration d'informations sur le voisinage des pixels devient indispensable car un objet est maintenant représenté par plusieurs pixels dans l'espace image. L'image peut alors être analysée de façon plus efficace en considérant cette information spatiale [11, 12, 15]. Ainsi, Bruzzone [16] utilise une approche basée sur les Champs de Markov Aléatoires (*Markov Random Fields*, MRF) afin de modéliser les probabilités d'appartenance d'un pixel à une classe en fonction de son voisinage. Al–Khudhairy [15] se base sur la forme et l'homogénéité des objets afin de les classer.
Ces approches, de plus en plus répandues, améliorent la robustesse des algorithmes au bruit et aux distorsions géométriques des images mais sans pour autant supprimer la sensibilité à tous les changements radiométriques non pertinents tels que les variations de réflectance des surfaces, l'évolution des sols ou des revêtements d'infrastructures.

### 1.1.2 Méthodes alternatives : détection des changements d'élévation

Dernièrement, de nouvelles approches, alternatives à la détection de changements 2D, ont émergé. Leur but est de concentrer la détection sur les changements d'élévation de la scène tout en étant plus robuste aux conditions d'acquisition des images ou aux changements non pertinents qui correspondent souvent à des changements 2D. Dans le cadre d'une détection des changements urbains de type bâti, ces méthodes apparaissent donc particulièrement appropriées. Les données d'entrée utilisées pour obtenir cette information d'élévation proviennent alors de campagnes LiDAR, de lasers aéroportés ou plus fréquemment d'imagerie stéréoscopique, spatiale ou aérienne [17, 18].
Ces techniques de détection des changements 3D peuvent être divisées en deux catégories

principales : les techniques image à base de données et les techniques image à image.

Les techniques dites image–base de données consistent généralement à comparer une base de données vecteur, contenant les empreintes de bâtiments (telles que des données cadastrales) et une image de l'élévation de tous les points de la scène. Ces techniques sont les plus utilisées en détection de changements car elles sont nécessaires pour la mise à jour automatique ou semi–automatique de bases de données déjà existantes.
La stratégie généralement adoptée dans ce cas est l'extraction des bâtiments sur la carte d'élévation et la comparaison des empreintes obtenues avec la base de données initiale afin de constater la présence ou non du bâtiment à la date de la base de données [5, 19, 20].
L'extraction des bâtiments est alors le point sensible de la méthode et peut être effectuée, selon le type de données d'entrée, par un algorithme d'extraction de structures dans un nuage de points généré par des données laser [18], par des méthodes de classification à partir de MNS et de données multispectrales [19, 21] ou bien par l'extraction des contours 2D des bâtiments sur un MNS afin de comparer les segments obtenus avec la base de données vecteur [5]. Cette étape d'extraction est particulièrement sensible car une erreur peut générer une fausse différence entre la base de données et les empreintes calculées, notamment lorsque la carte des élévations de la scène est générée à partir d'images, plus bruitées que les données laser. De même, l'étape de comparaison nécessite généralement que les données vecteur et image soient très précisément recalées. Des post–filtrages sont souvent appliqués sur les extractions de bâtiment afin de compenser ce type d'erreur, mais ces filtrages peuvent altérer la détection des changements les plus petits, tels que les modifications apportées à un bâtiment.
Enfin, l'inconvénient majeur de ces techniques réside dans le fait que les bases de données cadastrales sont très rarement disponibles et mises à jour, rendant ce type de méthode totalement inadapté pour la quantification des dégâts après une catastrophe majeure par exemple [19].

Les techniques image à image représentent alors la seule alternative lorsqu'aucune donnée externe n'est disponible ou valide. Elles consistent à calculer la différence entre deux images d'élévation acquises aux dates $t_1$ et $t_2$ afin de mettre en évidence les changements d'élévation apparus.
Cette méthode est largement utilisée pour des problématiques de géoscience telles que l'étude

d'objets géologiques comme les volcans [2] ou les glaciers [1], la mesure des mouvements topographiques ou des glissements de terrain [3, 4].
Dans ce contexte de surveillance des milieux, il est souvent nécessaire d'acquérir des séries temporelles composées de nombreuses images afin de suivre les variations sur une grande période de temps. Les données employées sont alors très souvent des données satellitaires ou aériennes qui permettent d'acquérir des scènes sur de vastes emprises terrain.
Pour des applications en géoscience, les variations recherchées sont typiquement de très basse fréquence spatiale et de faible amplitude altimétrique, visibles à grande échelle et donc facilement détectables par une simple différence entre les cartes d'élévation, générées à basse ou moyenne résolution [3, 4]. Au contraire, dans le contexte urbain dans lequel se place ce travail, les changements recherchés sont généralement de très haute fréquence spatiale et de forte amplitude, nécessitant une précision de détection qui ne peut être obtenue qu'avec des images très haute résolution. Les méthodes développées pour la basse ou moyenne résolution ne sont pas adaptées à cette problématique car il est alors nécessaire de mettre en évidence les changements recherchés parmi les nombreuses fausses alarmes générées à haute ou très haute résolution spatiale [8, 22].

Parmi les méthodes de détection image–image à très haute résolution, deux types de techniques sont appliqués : les techniques semi–automatiques, basées sur l'utilisation d'informations multi–spectrales ou contextuelles telles que la taille ou la forme des objets ou encore sur l'utilisation de zones d'apprentissage et les techniques entièrement automatiques.
Jung [23] propose une méthode semi–automatique qui utilise des MNS générés par imagerie stéréoscopique aérienne. Dans une première étape, un algorithme de *focusing* permet de rechercher les zones ayant probablement changé grâce à la comparaison des MNS de chaque date. Puis une seconde étape permet, pour chaque date, de classer ces zones d'intérêt selon les labels "bâtiment" ou "non–bâtiment" à l'aide d'arbres de décision générés à partir de zones d'apprentissage. Les deux segmentations ainsi obtenues sont alors comparées pour retrouver les changements du bâti. Avec ce type de méthode, la qualité de la détection dépend alors de chacune des segmentations et ne permet pas de retrouver les modifications apportées à un bâtiment.

Aujourd'hui, peu d'études portent sur la détection des changements d'élévation entre deux

MNS générés à partir d'images satellites stéréoscopiques très haute résolution.
En effet, malgré les avantages qu'apporte la haute résolution temporelle de ces images qui sont aussi moins onéreuses qu'une campagne d'acquisition aérienne, leur résolution reste beaucoup plus basse que celle des images aériennes (aujourd'hui WorldView–1 atteins 50 cm de résolution nadir lorsque les images aériennes ont une résolution de l'ordre du décimètre). De plus, les images aériennes sont plus contrastées que les images satellites permettant une mise en correspondance plus efficace pour la génération de MNS précis. Enfin, leur acquisition quasi nadir permet de limiter les zones d'occlusion dans les images.
Ainsi, les MNS produits à partir de couples d'images satellites souffrent de plus nombreuses erreurs de corrélation et de zones d'occlusion par rapport à ceux produits par l'aérien, erreurs se propageant souvent sous forme de fausses alarmes sur les cartes finales de détection de changements.

Afin de réduire l'impact des erreurs de MNS dans le MNS différentiel, Tian [8] propose le calcul d'une différence robuste entre les MNS des deux dates d'acquisition qui se traduit par l'affectation, pour chaque pixel $(i, j)$ de l'image des différences, de la différence minimale obtenue entre le pixel du MNS à $t_1$ en $(i, j)$ et un pixel du MNS $t_2$ appartenant à une fenêtre de taille définie et centrée sur le pixel $(i, j)$. Puis, dans une seconde étape, la carte des différences obtenue est analysée par une extraction de contours suivie par un affinage de formes par un algorithme de *box-fitting*.
Dans une autre approche, Tian [24] propose, après le calcul de la différence robuste, l'application d'un masque des zones d'ombre, basé sur l'exploitation des bandes multispectrales tel que l'a implémentée Marchant [25]. Ce masque est alors combiné à un masque regroupant les corrélations de faible coefficient obtenues lors de la génération des MNS. La différence masquée ainsi obtenue est alors segmentée puis les éléments mis en évidence sont filtrés à partir de critères de tailles et de formes décrits par Chaabouni–Chouayakh [22], afin de mieux séparer les changements appartenant au bâti des fausses erreurs.
Cette approche permet la détection de 60% à 80% des objets de plus de 100 $m^2$ et de 2 m à 3 m d'élévation différentielle, respectivement et avec 50% à 15% de fausses alarmes.
Plus récemment, Tian [26] a amélioré cette approche grâce à l'ajout d'un indicateur de changements calculé directement entre les images panchromatiques ayant permis la génération

des MNS. Cet indicateur de changements est basé sur l'analyse de l'évolution des statistiques locales d'une image entre deux dates. Le résultat de la fusion des informations obtenues par la différence robuste segmentée et par les indicateurs de changements radiométriques est ensuite filtré à partir des mêmes critères de formes et de tailles. Les résultats montrent alors de 55% à 93% des changements détectés avec 45% à 16% de fausses alarmes.
Cependant, si les résultats de ces méthodes apparaissent prometteurs, les nombreux seuils et critères nécessaires peuvent limiter les changements détectables à certaines formes et tailles.

Enfin, Reinartz [6] effectue une segmentation des images aux dates $t_1$ et $t_2$ en utilisant une méthode de classification basée sur le IR–MAD (Iteratively Reweighted Multivariate Alteration) développée par Nielsen [27]. Les résultats de segmentation aux deux dates sont alors associés avec le MNS différentiel afin de décider, pour chaque région segmentée, si elle correspond à des bâtiments détruits ou nouveaux.

En conclusion de cette étude, nous noterons que, à notre connaissance, aucune méthode totalement automatique n'existe pour la détection des changements d'élévation à partir d'imagerie satellite seule, sans ajout d'informations externes ou contextuelles. Pourtant, la capacité de réactivité des satellites face à une crise, leur agilité pour la stéréoscopie ainsi que leur résolution inférieure au mètre sont des atouts majeurs pour des problématiques de détection de changements d'élévation, même si la génération de MNS précis à partir de ces données demeure difficile.

## 1.2 Méthodes de restitution 3D : état de l'art

Classiquement, la reconstruction 3D d'une scène nécessite le calcul de la disparité entre tous les pixels homologues des deux images d'un couple stéréoscopique.
Les pixels homologues correspondent à des pixels représentant un même point terrain, repérés sur chaque image du couple stéréoscopique. La disparité correspond au décalage géométrique (en pixels) mesuré entre deux pixels homologues lorsque les angles de prise de vue des images diffèrent. Cette valeur de disparité est reliée à l'élévation de la scène au pixel considéré, elle est donc calculée pour tous les pixels d'une image par la mise en correspondance dense des images

afin d'obtenir l'élévation de tous les points de la scène.

Cependant, les contraintes liées aux acquisitions satellite rendent le calcul de la disparité particulièrement complexe car de nombreuses difficultés se présentent :

- les angles d'acquisition des images sont souvent importants (jusqu'à 30° d'incidence) générant des différences radiométriques importantes,
- ces angles d'incidence génèrent d'importantes zones d'occlusion : c'est–à–dire des zones cachées dans l'image, typiquement les zones aux abords des bâtiments. Dans ce cas il n'est pas possible de trouver des pixels homologues dans ces zones,
- des d'objets peuvent être en mouvement : la mise en correspondance des images d'un couple implique comme hypothèse que la scène soit parfaitement immobile entre les deux acquisitions, or, dans la réalité, cette hypothèse n'est pas réalisable,
- certaines zones sont trop homogènes ou possèdent une texture répétitive ou bien une structure 3D très complexe (c'est le cas pour la structure de la végétation), dans tous ces cas il est très difficile de déterminer précisément les pixels homologues,
- il existe des changements radiométriques entre les acquisitions : toutes les surfaces n'étant pas lambertiennes, c'est–à–dire que la lumière n'est pas forcément réfléchie de la même façon dans toutes les directions, une même surface peut présenter différentes radiométries selon l'angle de l'acquisition et ainsi compliquer la mise en correspondance,
- certains défauts géométriques résiduels subsistent, même après l'affinage, comme nous le montrerons dans la section 2.1.

Depuis les 30 dernières années, de nombreux algorithmes ont été développés pour la reconstruction 3D de scènes à partir d'imagerie terrestre, aérienne ou spatiale. Des inventaires de ces techniques ont été réalisés par Scharstein [28], Brown [29] puis plus récemment par Lazaros [30]. La base de données de Middlebury [1] a notamment été très utilisée pour réaliser ces inventaires. Ils permettent de comparer un grand nombre de ces techniques en termes de performance et de méthodologie. Les techniques présentées ne sont cependant pas toutes transposables aux images satellites.

1. Ensemble de jeux de données correspondant à des images stéréoscopiques accompagnées de vérités terrain mis à disposition des utilisateurs pour tester des algorithmes de mise en correspondance. Le contexte et les images sont cependant éloignés de notre problématique de reconstruction 3D par imagerie satellite en milieu urbain.

### 1.2.1 Mise en correspondance d'images par méthode d'optimisation

Pour de nombreux auteurs, la mise en correspondance d'images peut être vue comme un problème de labellisation [31],[32],[33],[34] : chaque valeur de disparité ou d'élévation (alors considérée comme un label) affectée à un pixel d'une image maître est déduite de la sélection du pixel homologue parmi plusieurs candidats dans l'image esclave.
La mise en correspondance peut ainsi être résolue à partir d'une méthode d'optimisation : l'objectif est alors de retrouver, pour chacun des pixels d'une image maître, les pixels homologues dans l'image esclave (ou dans toutes les autres images dans le cas de la multi–stéréoscopie).
La mise en correspondance peut alors être réalisée par des méthodes locales d'optimisation : la recherche de primitives homologues (pixels, segments ou régions) est alors effectuée de façon locale uniquement, ou bien à travers des méthodes globales (ou semi–globales) selon lesquelles la mise en correspondance est résolue en recherchant la solution de disparité sur l'image entière (ou des subdivisions de l'image).
Ainsi, les méthodes de mise en correspondance peuvent être basées sur les algorithmes d'optimisation que nous verrons plus en détails dans le chapitre 2, section 2.2.2.

#### Méthodes locales de mise en correspondance

Les méthodes locales peuvent être basées sur la mise en correspondance des segments ou des contours de l'image, tels que les contours de bâtiments. Plusieurs auteurs utilisent ces contours pour aider à la mise en correspondance radiométrique des images.
La méthode proposée par Baillard [35] recherche les points de contours des bâtiments sur des images aériennes afin d'apporter une contrainte géométrique à la corrélation radiométrique des pixels, effectuée dans une seconde étape.
Zhang [36] propose, sur des images satellites, une méthode hybride basée sur la combinaison de la mise en correspondance dense des pixels et celle des contours de l'image. Cette technique permet d'imposer une forte contrainte de régularisation entre les contours détectés afin de garantir une surface régulière tout en préservant les discontinuités au niveau des contours.
Ces techniques montrent des résultats particulièrement intéressants, notamment pour les zones homogènes mais elles peuvent être sensibles à la détection de contours. Les méthodes basées

sur les contours restent cependant peu utilisées à cause des erreurs produites par la détection de contours et la difficulté de retrouver les plus petites structures.

Les méthodes les plus populaires de mise en correspondance sont généralement basées sur l'intensité des pixels des images et la mise en correspondance dense de tous les pixels de l'image, souvent à travers la méthode dite de *block matching*. Cette méthode est basée sur la corrélation d'une fenêtre de pixels définie sur l'image maître avec une fenêtre glissante sur l'image esclave et le calcul d'un score de corrélation. La disparité est ensuite évaluée à partir du pixel ayant engendré le meilleur score de corrélation (technique du *winner take all*) [37].
La zone de recherche est alors souvent contrainte par la ligne épipolaire (ou pseudo–épipolaire dans le cas des images satellites). Ce type de méthode, très rapide, est cependant particulièrement sensible au bruit des images, aux différences radiométriques ou encore aux défauts d'orientation résiduels des images lorsque la mise en correspondance est réalisée le long de l'épipolaire seulement.
Récemment, d'autres techniques ont émergé, basées sur la mise en correspondance de zones segmentées d'une image [38] et l'assignement d'une valeur de disparité à ces zones par méthode d'optimisation. Ces techniques sont cependant sensibles à toutes les erreurs de segmentation qui sont alors susceptibles de provoquer des erreurs de reconstruction 3D.

#### Méthodes globales ou semi–globales

Les méthodes de reconstruction 3D les plus populaires et les plus efficaces sont aujourd'hui les méthodes basées sur la mise en correspondance à travers une optimisation sur l'image. Ces méthodes fonctionnent par le calcul d'un coût de mise en correspondance lié au score de corrélation obtenu entre les fenêtres de pixels candidates sur chaque image (*block matching*), auquel s'ajoute un terme de régularisation. Nous décrirons en détail cette optimisation dans le chapitre 3, section 3.1.
Les différences entre les nombreuses méthodes basées sur une optimisation sont liées à la mesure de similarité, à la contrainte de régularisation et à la méthode d'optimisation employée.
Comme nous le verrons dans le chapitre 2, section 2.2.2, de nombreuses méthodes d'optimisation ont été développées et appliquées au traitement d'images et plus particulièrement à la mise en

correspondance d'images.

On trouve notamment les méthodes globales telles que le *belief propagation* [39], [33], ou les *graph-cuts* [40], [41], l'objectif est alors de retrouver, pour tous les pixels de l'image simultanément, une solution qui minimise une fonctionnelle d'énergie globale sur l'image.

Les méthodes dites semi–globales [42] permettent de calculer une solution non plus sur toute l'image simultanément mais sur des séquences de pixels de l'image traitées indépendamment [43]. Ces séquences peuvent être formées par les lignes [34] de l'image ou bien par les lignes, colonnes et toutes les directions intermédiaires de l'image, selon la méthode de Hirschmüller [31] ou Pierrot-Deseilligny [32] ou encore par un graphe sélectionné dans l'image comme le propose Veksler [44]. Les résultats obtenus sur chaque séquence de l'image sont ensuite fusionnés afin d'obtenir la solution semi–globale.

Nous verrons dans la section 2.2.3 que la méthode de reconstruction que nous avons sélectionnée est basée sur ce type d'optimisation semi–globale et résolue par programmation dynamique à l'image des méthodes proposées par Hirschmüller et Pierrot-Deseilligny [31], [32].

### 1.2.2 Caractéristiques des corrélateurs

Quelle que soit la méthode d'optimisation choisie, la mesure de similarité employée pour calculer la correspondance entre les pixels ou entre les fenêtres de pixels de chaque image (*matching cost*), est un point important pour sélectionner les pixels homologues.

Une étude des différentes mesures de la littérature a été réalisée par Scharstein et Szeliski [28], puis Hirschmüller [45].

Les mesures les plus utilisées sont :

- la somme des différences radiométriques au carré (*Sum of Absolute Differences* ou SAD),
- la somme des différences radiométriques absolues (*Sum of Squared Differences* ou SSD),
- le *normalized cross correlation coefficient* (ou NCC).

De nombreuses autres mesures de similarité sont proposées dans la littérature. Birchfield [46] propose une mesure de ressemblance robuste à l'échantillonnage pixellique des images : au lieu de comparer deux fenêtres de pixels décalées par un nombre entier de pixels, il compare chaque pixel de l'image maître avec une interpolation linéaire d'une fenêtre de pixels sur l'autre

image.

Afin d'être le plus robuste possible aux différences radiométriques provoquées par la variation d'angle de prise de vue, des méthodes dites "non-paramétriques" sont aussi très utilisées. Elles consistent à filtrer les données avant la corrélation afin de réduire l'impact de ces différences [37]. Le filtrage de *rank*, par exemple, recherche dans une fenêtre centrée sur un pixel, le nombre de pixels dont la valeur est en dessous de celle du pixel considéré puis remplace la valeur du pixel central par cette valeur. La correlation (réalisée avec la mesure SAD) va alors dépendre de l'agencement des valeurs et non plus de l'intensité. Cependant, ce filtrage engendre une perte d'information spatiale qui réduit le pouvoir discriminant de la méthode [29].

Une variation à ce filtrage, nommé le filtre census, préserve la distribution spatiale de la fenêtre en encodant le résultat du filtre de *rank* pour chaque pixel dans une chaîne de bytes [29].

D'autres filtrages de la fenêtre de mise en correspondance sont utilisés tels que le *Laplacian of Gaussian* ou le filtrage moyen [45]. Tous ces filtres tentent d'améliorer la robustesse aux conditions d'acquisition et aux erreurs de corrélation des images car ils diminuent la sensibilité aux *outliers* des images. Ils sont majoritairement suivis d'une corrélation par SAD.

Hirschmüller [31] propose une mesure de similarité basée sur la notion d'information mutuelle, réputée moins sensible au recalage et aux conditions d'illumination des images. L'information mutuelle entre deux fenêtres de pixels est calculée en additionnant l'entropie des histogrammes de chaque image puis en soustrayant l'entropie de l'histogramme joint.

Hirschmüller et Szeliski [45] ont comparé différents coûts de corrélation et leur efficacité pour la mise en correspondance dense des images. Ils montrent d'abord que la performance d'une fonction de coût dépend avant tout de la méthode d'optimisation employée indiquant que la fonction de coût a finalement un impact modéré sur le résultat final.

Cependant, ils observent que le filtrage de *rank* et l'information mutuelle semblent plus efficaces au niveau des discontinuités car ils permettent notamment une meilleure robustesse à l'effet d'adhérence qui conduit généralement à la dilatation des objets de la scène. Ce problème, inhérent aux méthodes de *block matching*, apparaît lorsqu'il y a un saut brutal de disparité dans l'une des images qui ne peut être parfaitement localisé à l'intérieur de la fenêtre de corrélation. Finalement, bien que plus performants, le filtre de rank ou l'information mutuelle nécessitent un temps de calcul 1,5 à 5 fois plus important que le temps requis pour une mesure de type SAD. Cette dernière est la mesure implémentée dans l'outil de mise en correspondance que

nous utilisons.

### 1.2.3 Les stratégies de mise en correspondance

#### L'approche multi–résolution

Aujourd'hui, la plupart des auteurs utilisent une approche *coarse to fine* ou multi–échelle pour la génération des MNS. Cette méthode est basée sur la corrélation réalisée sur les images ré–échantillonnées de l'échelle la plus grande à la plus petite. Le résultat obtenu pour une certaine échelle sert alors d'initialisation pour la corrélation à échelle suivante. Le grand avantage de cette technique est un important gain de temps [31], [32].

Cependant, Szintsev [40] montre que de tels algorithmes ont tendance à flouter les bords de bâtiments car lors de l'agrégation des pixels d'une fenêtre de corrélation, des pixels appartenant à deux surfaces différentes sont mélangés aux échelles les plus grandes et les discontinuités sont plus difficilement retrouvées. Des erreurs peuvent alors se propager à toute la pyramide d'images.

Certains auteurs proposent des fenêtres de corrélation adaptatives au niveau des zones de fortes discontinuités [47]. Sizintsev [40] et Pierrot-Deseilligny [32] proposent l'équivalent d'un filtrage morphologique à chaque niveau de la pyramide d'images en utilisant la fenêtre de corrélation comme élément structurant afin de simuler une fenêtre adaptative pouvant se déplacer autour de la discontinuité qui a été calculée au niveau supérieur de la pyramide, augmentant la précision au niveau de ces zones de discontinuités. Cet algorithme est décrit par Pénard et al.[48] et est implémenté dans l'outil de mise en correspondance que nous avons sélectionné.

#### La prise en compte des occlusions

L'un des points les plus sensibles de la mise en correspondance réside dans la présence des zones d'occlusion sur les images stéréoscopiques. Elles correspondent à des zones observées sur une des deux images seulement et pour lesquelles il est impossible de retrouver les pixels homologues (ces zones sont souvent appelées demi–occlusion ou *half–occlusion* dans la littérature car l'occlusion n'est présente que sur une des deux images). Ces zones se situent aux abords

des bâtiments et ont une taille variable en fonction de l'angle d'incidence de l'acquisition de l'image. Or, pour la stéréoscopie, les angles d'incidence peuvent être très importants (jusqu'à 30° pour les images WorldView) créant des zones d'occlusion de taille importante dans l'image. Ces dernières années, de nombreux auteurs se sont intéressés à ce sujet. Brown [29] définit trois classes d'algorithmes prenant en compte les occlusions : ceux qui détectent les occlusions (généralement après le calcul de la disparité), ceux qui sont plus robustes aux zones d'occlusion grâce aux fonctions de coût implémentées et enfin ceux qui modélisent les zones d'occlusion lors de la mise en correspondance.

Les méthodes les plus simples pour la détection des zones d'occlusion sont basées sur la détection des discontinuités dans l'image. Fua [49] et Qingxiong [33] calculent deux cartes de disparité en inversant les rôles maître et esclave des images à chaque fois. Les disparités dont les valeurs sont trop éloignées entre elles pour le même point sont alors considérées comme des occlusions. Cette méthode dite de *back matching* est aujourd'hui implémentée dans de nombreux algorithmes [29]. Elle est cependant longue en temps de calcul et peut générer de très nombreuses zones dites d'occlusion qui sont en réalité dues à des différences d'illumination, des zones homogènes ou plus généralement des zones de mauvaise corrélation sur l'une des deux cartes de disparité.

Egnal [50] compare cinq approches de détection des zones d'occlusion basées sur :

- l'analyse de l'histogramme des disparités dans une fenêtre de disparité,
- les variations du coefficient de corrélation,
- le double calcul de la disparité à partir de chacune des images,
- l'analyse de la contrainte d'ordre des disparités (pas forcément respectée lors d'une occlusion),
- les discontinuités de la disparité.

Ce dernier algorithme, ainsi que l'algorithme de double calcul de disparité apparaissent comme les plus prometteurs. Cependant, leur efficacité semble dépendre du paysage en présence (la position de l'occlusion, la présence d'objets fins, etc...). De plus, il n'y a pas eu d'étude concernant la performance de ces algorithmes en fonction de la méthode d'optimisation appliquée.

Les méthodes réduisant la sensibilité aux zones d'occlusion correspondent souvent à

l'adaptation de la mesure de similarité en fonction de ce problème. Chambon [51] utilise deux mesures combinées : le *zero mean normalized cross correlation coefficient* et le *smooth median absolute deviation*, afin d'améliorer la robustesse envers ces zones. Les filtrages de *rank* et *census* proposés par Zabih [37] permettent aussi une certaine robustesse aux *outliers* et donc souvent aux erreurs parfois ponctuelles générées par les zones d'occlusion (lorsque l'angle d'incidence reste faible).
Enfin, d'autres méthodes passent par l'utilisation de fenêtres adaptatives en fonction du degré de confiance de la corrélation, le but est d'optimiser les résultats de la corrélation près des zones d'occlusion [29].

Les méthodes modélisant les zones d'occlusion, c'est–à–dire permettant leur localisation en fonction du voisinage, intègrent les zones d'occlusion directement dans la recherche de la disparité en créant des fonctions de coût spécifiques à ces zones. Belhumeur [52] a implémenté un modèle complexe de fonction de coût incluant des termes pour les surfaces lisses, les limites d'objets et les pentes. Bobick et Intille [34] associent un coût avec les zones d'occlusion qu'ils réduisent ou augmentent en fonction de l'orientation des contours de bâtiments et donc de celle de la zone d'occlusion.

L'étude comparative de Egnal [50] ne considère que des méthodes de détection des zones d'occlusion. À notre connaissance, aucune étude ne permet de comparer les différentes techniques citées pour la prise en compte de ces zones (détection - sensibilité - modélisation). Cependant, les méthodes les plus efficaces, rapides et faciles à implémenter sont vraisemblablement les méthodes de détection des occlusions, même si elles nécessitent généralement un double calcul de la disparité ou un post–traitement des cartes de disparité générées.
Les méthodes réduisant la sensibilité aux occlusions montrent certaines performances mais ne permettent pas la localisation des occlusions et peuvent donc engendrer des erreurs dans le MNS, erreurs qui se propagent généralement ensuite dans la détection des changements de la scène.
Au contraire, les méthodes de modélisation s'avèrent très performantes pour la localisation mais elles sont particulièrement difficiles à implémenter.
Dans la chaîne de traitements développée, nous avons choisi une méthode de détection des zones d'occlusion appliquée après le calcul des cartes d'élévation par mise en correspondance. Cet

algorithme, basé sur des contraintes géométriques au niveau des discontinuités de l'image sera décrit dans le chapitre 3, section 3.2.

**Les géométries de mise en correspondance**

La géométrie de mise en correspondance correspond à la technique choisie pour la recherche des pixels homologues entre les images stéréoscopiques. Selon le nombre d'images disponibles et leur précision d'orientation, plusieurs géométries sont possibles.
Classiquement, cette recherche est basée sur la géométrie épipolaire (ou pseudo–épipolaire dans le cas des images satellites) comme représenté sur la figure 1.1(a). La recherche d'un pixel de l'image maître est effectuée le long de la ligne épipolaire correspondante sur l'image esclave. Une recherche bi–dimensionnelle (transversale à l'épipolaire) permet d'être plus robuste aux erreurs résiduelles dans les modèles géométriques des images.
Cette technique, encore très répandue, nécessite cependant un ré–échantillonnage épipolaire des images.

La seconde géométrie, très classique aujourd'hui, est la géométrie image (ou faisceaux), illustrée en figure 1.1(b). Le principe est la recherche du pixel homologue d'une image maître sur l'image esclave en localisant celui–ci grâce aux modèles géométriques des images (affinés préalablement) et pour chaque altitude définie dans un repère terrain [36], [32]. Une fois le pixel localisé sur l'image esclave, il est aussi possible d'effectuer la recherche selon une seconde dimension, dans l'espace image afin d'être, là encore, plus robuste aux erreurs géométriques des images.

Enfin, la géométrie dite terrain, représentée en figure 1.1(c), permet de retrouver, pour chaque altitude $z$ d'un point défini dans un repère terrain, les pixels de chaque image qui correspondent à ce point terrain grâce aux modèles géométriques des images. Un score de corrélation est alors calculé entre les fenêtres de pixels définies autour des pixels sélectionnés. L'altitude finale choisie est alors celle qui maximise le score de corrélation entre les fenêtres de pixels sélectionnées.
Cette technique permet le calcul direct de la carte de profondeur sur une grille terrain régulière, quel que soit le nombre d'images utilisé. De plus, toutes les images sont considérées ensemble, sans notion de maître et esclave, permettant un processus symétrique. Si cette technique se

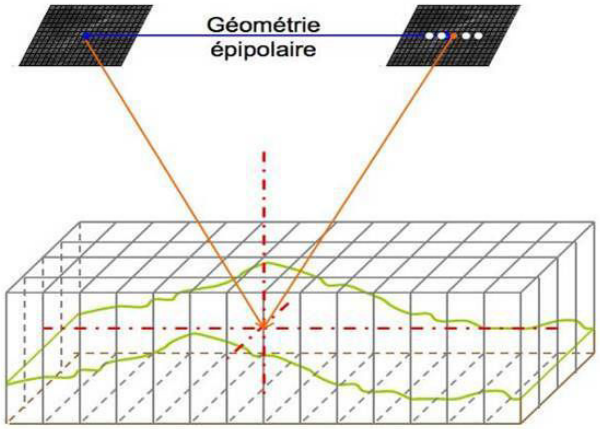
montre peu robuste aux erreurs géométriques des images ou aux occlusions lorsque seulement deux images sont utilisées, elle s'avère très avantageuse lorsque le nombre d'images augmente car elle est alors plus précise (notamment pour la prise en compte des zones d'occlusion) et plus rapide qu'une mise en correspondance d'images deux à deux.

Ces différentes géométries de mise en correspondance sont décrites en détail, avec leurs avantages et inconvénients, dans l'article de Pénard et al [48]. Les auteurs montrent notamment que la géométrie image présente généralement de meilleurs résultats que la géométrie terrain, c'est pourquoi cette dernière a été sélectionnée pour la mise en correspondance dans notre chaîne de traitements.

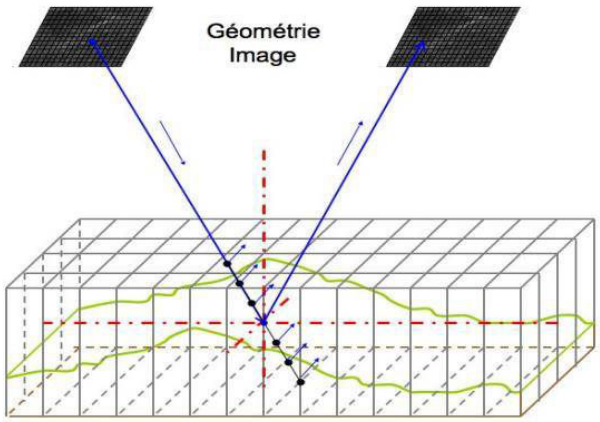
### 1.2.4 Les outils de restitution 3D existants

Aujourd'hui, de nombreux outils de reconstruction 3D existent sous forme de packages commerciaux ou open–source. On notera, parmi les packages commerciaux les plus populaires, Socet–set de BAE Systems, Erdas Imagine de Intergraph, Sat–PP de 4D Ixplorer, OrthoEngine de PCI Geomatics, Smart3DCapture de Acute3D ou encore Photo Modeler. Les performances de ces outils sont généralement bonnes mais le paramétrage peut être difficile surtout lorsque les codes sources ne sont pas accessibles, ce qui les rend peu exploitables dans un cadre de recherche.

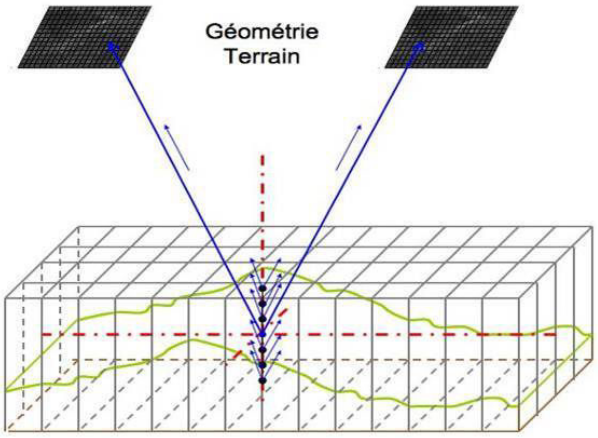
D'autre outils existent tels que PMVS, de Furukawa et Ponce [53], OpenCV library (contenant les librairies pour la mise en correspondance selon l'algorithme de Hirschmüller [31]) ou encore MicMac développé par l'IGN [32]. Ces outils sont libres de téléchargement mais parmi eux, seul MicMac est un outil totalement open–source.
Le téléchargement des codes sources de MicMac permet en effet le contrôle de tous les paramètres et la maîtrise interne complète du code. Ses différentes options, qui autorisent la mise en correspondance selon toutes les géométries citées précédemment, l'utilisation de plusieurs méthodes d'optimisation et d'interpolation et l'option multi–échelle en font un outil particulièrement adaptable à tous les contextes (géologique ou urbain, en imagerie spatiale ou terrestre).
C'est pourquoi l'outil MicMac est donc celui que nous avons sélectionné pour la mise en correspondance des images stéréoscopiques.




(a) Schéma de principe de la géométrie épipolaire




(b) Schéma de principe de la géométrie image




(c) Schéma de principe de la géométrie terrain

FIGURE 1.1 – *Schéma de principe des différentes géométries de mise en correspondance.*

## 1.3 Description générale de la chaîne de traitements développée

L'objectif fixé lors du travail de thèse est de développer une méthode totalement automatique pour la détection des changements d'élévation à partir d'images satellites stéréoscopiques.
Nous nous plaçons dans un contexte urbain, le but étant de retrouver tous les changements du bâti (construction, destruction ou modifications d'infrastructure). Les contraintes du système sont alors les suivantes :

- le système doit être entièrement automatique de l'acquisition des images à la génération des cartes de changements,
- la méthode doit être applicable quels que soient le ou les capteurs très haute résolution utilisés, qu'ils soient satellites ou aériens,
- les paramètres doivent être suffisamment génériques afin que la méthode soit applicable dans des contextes variés (milieu urbain, montagneux, désertique).

Afin de répondre au mieux à ces contraintes, une chaîne complète et innovante de traitements à été développée. Cette chaîne, illustrée en figure 1.2 comprend trois étapes majeures : la première étape (entourée en trait plein sur la figure) correspond à la spatio–triangulation ou recalage simultané de toutes les images disponibles, sans considération de leur date d'acquisition. Cette étape, décrite dans le chapitre 2, section 2.1, sera réalisée à l'aide d'outils existants permettant la recherche de points de liaison entre les images et l'affinage des modèles géométriques des images.

La seconde étape (entourée en pointillés fins) permet la génération de MNS précis pour chaque date. Cette étape, décrite dans le chapitre 3 consiste d'abord en une mise en correspondance des images à l'intérieur de chaque couple, chaque image étant successivement maître et esclave. Cette mise en correspondance, réalisée à l'aide de l'outil MicMac et par une méthode d'optimisation (voir section 3.1), est suivie par le basculement des MNS générés sur une grille régulière commune (voir section 3.2) puis par la fusion de tous les MNS obtenus avec un algorithme de labellisation par optimisation (voir section 3.3).
Ces deux dernières étapes sont issues de développements réalisés au cours de la thèse et

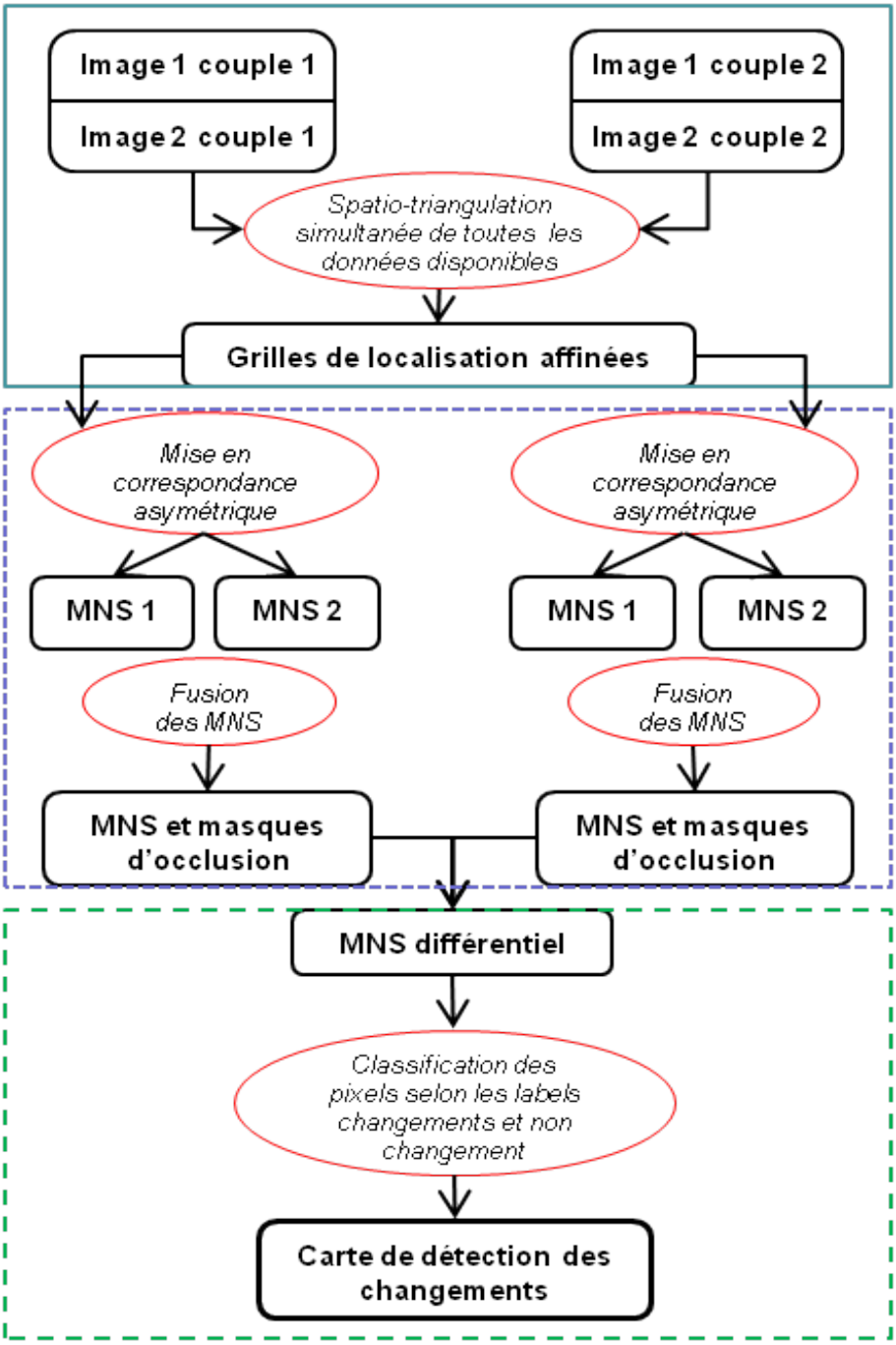



FIGURE 1.2 – *Illustration des différentes étapes de la chaîne globale de traitements.*

permettent de prendre en compte toutes les contraintes et sources d'erreurs inhérentes aux images satellites afin de générer des MNS précis en vue de la détection des changements d'élévation.

Ainsi, la dernière étape (entourée en pointillés larges) a pour objectif l'analyse de la différence des MNS afin de mettre en évidence les changements d'élévation les plus pertinents de la scène selon une méthode que nous avons développée. Cette méthode, décrite dans le chapitre 4, est basée sur la classification des pixels du MNS différentiel à travers un algorithme de labellisation par optimisation.

Nous noterons finalement que trois étapes de calcul réalisées dans cette chaîne de traitements sont basées sur une labellisation des pixels de l'image résolue par une méthode d'optimisation. Un algorithme "noyau" a ainsi été développé au cours de thèse et a été adapté pour la résolution de chacun de ces problèmes. Cet algorithme est décrit dans le chapitre 2, section 2.2.3.

## 1.4 Capteurs utilisés et données d'entrée

Au cours de ce travail, de nombreux jeux de données ont été utilisés afin d'expérimenter les méthodes sur différentes scènes et d'en analyser les résultats. Seules des images panchromatiques provenant de capteurs très haute résolution spatiale ont été employés : Ikonos, dont les images obtenues étaient ré–échantillonnées à 1 m de résolution spatiale, WorldView-1 et 2, produisant des images à environ 60 cm de résolution spatiale et Pléiades, produisant des images à environ 70 cm de résolution spatiale. Le tableau 1.1 résume toutes les données utilisées.
Bien que les images multispectrales puissent apporter des informations permettant de filtrer les cartes de détection de changements obtenues à la manière de Tian [8], nous avons choisi de développer notre chaîne de traitements à partir des images panchromatiques uniquement.
Tous ces capteurs sont de type push–broom : barrettes linéaires de détecteurs permettant l'acquisition de chaque ligne de l'image successivement grâce au défilement du paysage dans le plan focal.

Quatre zones géographiques ont été étudiées pour les différentes étapes de calcul de la méthode, en fonction des données disponibles.
Pour l'illustration de la chaîne complète, i.e. de la génération des MNS à la détection de changements, la ville de Phoenix, en Arizona (États-Unis) a été analysée entre les dates de 2008 et 2011. Cette zone permet d'évaluer les performances de la méthode lors de l'analyse de l'évolution ty-

| *Capteur* | *Date d'acquisition* | *Résolution spatiale (m)* | *Angle d'azimut (°)* | *Angle d'élévation (°)* | *b/h* |
|---|---|---|---|---|---|
| Données acquise sur la ville de Phoenix (Etats-Unis) | | | | | |
| WorldView-1 | 2008-10-06 | 0,59 | 51,5 | 64,1 | 0,67 |
| WorldView-1 | 2008-10-06 | 0,56 | 141,6 | 67,6 | |
| WorldView-1 | 2011-04-28 | 0,62 | 156,2 | 59,6 | 0,70 |
| WorldView-1 | 2011-04-28 | 0,56 | 65,0 | 67,6 | |
| Données acquise sur la ville de Toulouse (France) | | | | | |
| Pléiades | 2012-05-07 | 0,70 | 115,5 | 64,7 | 0,54 |
| Pléiades | 2012-05-07 | 0,70 | 88,9 | 64,6 | 0,55 |
| Pléiades | 2012-05-07 | 0,70 | 102,2 | 65,2 | 0,54 |
| Données acquise sur la ville de Chistchurch (Nouvelle Zélande) | | | | | |
| WorldView-1 | 2009-09-16 | 0,66 | 85,7 | 57,2 | 0,66 |
| WorldView-1 | 2009-09-16 | 0,58 | 66,3 | 23,7 | |
| WorldView-2 | 2011-08-30 | 0,60 | 199,5 | 57,3 | 0,66 |
| WorldView-2 | 2011-08-30 | 0,47 | 289,0 | 87,2 | |
| Données acquise sur la ville de Sendai (Japon) | | | | | |
| Ikonos | 2010-12-11 | 1 (reech.) | 64,9 | 64,65 | 0,79 |
| Ikonos | 2010-12-11 | 1 (reech.) | 149,15 | 60,66 | |
| Ikonos | 2011-08-13 | 1 (reech.) | 314,4 | 61,14 | 0,72 |
| Ikonos | 2011-08-13 | 1 (reech.) | 250,96 | 56,66 | |

TABLE 1.1 – *Métadonnées des images utilisées lors des différentes études.*

pique d'un milieu urbain. Dans ce même contexte, la ville de Christchurch, en Nouvelle Zélande a aussi été étudiée. L'intérêt de cette zone est la présence de deux acquisitions LiDAR effectuées au même endroit et à des dates proches des acquisitions satellitaires. Cette zone permet ainsi une quantification précise des performances des algorithmes mis en place.

Les données Pléiades, acquises sur la ville de Toulouse selon le mode tri–stéréoscopique, permettent d'illustrer l'intérêt de l'algorithme de génération des MNS pour le traitement d'acquisitions multi–stéréoscopiques.

La zone de Sendai (Japon), permet enfin de tester de la chaîne de traitements sur un cas concret de catastrophe majeure. En effet, les données disponibles ont été acquises avant et après le séisme suivi par le tsunami qui a ravagé la côte de Sendai le 11 mars 2011.

## 1.5 Conclusion

Ce chapitre présente tout d'abord de nombreuses méthodes de détection de changements à partir d'images satellites ou aériennes. Cependant, si les méthodes classiques basées sur la comparaison radiométrique des images montrent de bonnes performances, elles ne s'avèrent pas toujours efficaces du fait des nombreuses fausses alarmes qu'elles génèrent et sont souvent difficilement automatisables car trop d'*a priori* sur les changements sont nécessaires.
Les méthodes basées sur l'exploitation de l'information d'élévation de la scène paraissent alors mieux adaptées, notamment au contexte urbain. Bien que ces méthodes nécessitent aussi des informations *a priori* ou contextuelles, l'analyse de la différence entre deux cartes d'élévation montre de très bon résultats.

Dans cet objectif, de nombreuses méthodes de restitution 3D ont été étudiées afin de sélectionner une stratégie suffisamment robuste pour la mise en correspondance d'images satellites et la génération de cartes d'élévation. Nous avons vu que de nombreuses techniques permettent de prendre en compte les erreurs de recalage des images ou bien les zones d'occlusion, ces deux points étant particulièrement importants dans le cas des images satellites.

La chaîne de traitements réalisée lors de cette thèse est ainsi basée sur la génération de MNS à travers plusieurs étapes de calcul permettant d'améliorer leur précision grâce à une technique de mise en correspondance sélectionnée pour ses performances et son adaptabilité, suivi par des post–traitements permettant une meilleur prise en compte des erreurs d'occlusion.
La méthode de détection des changements développée est, quant à elle, une méthode innovante car totalement automatique permettant la détection des changements pertinents du bâti grâce à la seule utilisation de l'information d'élévation produite aux dates d'intérêt.

# Chapitre 2

# Outils et méthodes nécessaires à la chaîne de traitements

Ce chapitre présente les outils de calcul, utilisés en l'état ou bien implémentés, nécessaires à la chaîne de traitements développée.
La première étape de la chaîne de traitements consiste en un recalage simultané de toutes les données disponibles aux dates $t_1$ et $t_2$. Cette étape est basée sur des outils existants pour la recherche de points de liaison entre les images et l'affinage des modèles géométriques des images par ajustement de faisceaux. Cette étape est essentielle pour la génération des MNS et le calcul des MNS différentiels, calculs qui nécessitent une grande précision de recalage entre les images. Après un bref rappel de l'objectif et du principe de la spatio–triangulation des images, nous verrons quelques résultats obtenus après ce calcul et nous discuterons des erreurs résiduelles générées.

Nous verrons ensuite dans ce chapitre l'outil central nécessaire aux étapes de calcul réalisées dans notre chaîne de traitements. Cet outil est un algorithme d'optimisation semi-globale et permettant la résolution de problèmes de labellisation. Nous montrerons tout d'abord comment cet algorithme a été sélectionné parmi les nombreux algorithmes d'optimisation existants puis nous détaillerons son implémentation telle que nous l'avons réalisée.

## 2.1 Co–localisation des images : utilisation d'un outil de spatio–triangulation

### 2.1.1 Contexte et objectif

Les capteurs très haute résolution tels que les capteurs Pléiades, WorldView, Ikonos, GeoEye ou OrbView sont particulièrement agiles, facilitant les acquisitions stéréoscopiques. Les images peuvent alors être prises par vue latérale (*across track*) ou plus communément en mode avant–arrière le long de la trace du satellite (*along track*), permettant des acquisitions quasi–simultanées. Ces différentes conditions d'acquisition sont le point clé de la stéréoscopie.

Cependant, les modèles géométriques fournis avec les images et contenant les informations sur la position du satellite, son orientation et sa direction de visée lors de l'acquisition, ne sont pas suffisamment précis. En effet, pour des satellites très haute résolution tels que Pléiades, il faudrait atteindre une précision de restitution des angles de visée au micromètre pour que l'orientation du satellite soit connue à l'échelle du pixel. Ceci n'est aujourd'hui pas réalisable par les systèmes de contrôle [54].
La figure 2.1 présente les angles correspondant aux variations d'attitudes du satellite appelés roulis $\alpha R$, tangage $\alpha T$ et lacet $\alpha L$.

Ainsi, lorsque les modèles géométriques natifs associés aux images sont exploités, on observe que les pixels homologues (ou points de liaison) appartenant à différentes images (du même couple stéréoscopique ou non) et représentant un même point physique ne correspondent pas aux mêmes coordonnées terrain.
On observe des biais variables selon les satellites et la précision de leurs modèles géométriques : DigitalGlobe annonce une précision de localisation absolue au sol de 15 m pour QuickBird, 9 m pour Ikonos et 5 m pour WorldView tandis que Pléiades est annoncé à 12 m de précision [54].

Or, la mise en correspondance des images nécessite une précision subpixellique (typiquement inférieure à 0,5 m) de localisation relative des images. En effet, la technique employée, détaillée au chapitre 3, est basée sur le parcours des lignes de visée de chacun des pixels. De même, lors de l'étape de détection de changements, décrite dans le chapitre 4, le calcul de la différence des

MNS pixel à pixel est particulièrement sensible au recalage des MNS, lui–même directement lié au recalage des images d'origine.

Afin de s'assurer de la précision de co–localisation des images puis des MNS, nous avons choisi de procéder, avant tout autre traitement, à une spatio–triangulation [55] (plus connue sous le nom d'aéro–triangulation pour les images aériennes).

### 2.1.2 Principe de la spatio–triangulation

Le principe de la spatio–triangulation est de corriger les erreurs des modèles géométriques des images afin que les lignes de visée provenant des pixels homologues de chacune des images se croisent au mieux sur les mêmes coordonnées au sol [56]. Cet affinage des modèles peut être réalisé de façon absolue au sol si des points d'appuis sont disponibles (points repérés dans l'image et de coordonnées terrain connues) ou relative, c'est–à–dire que les images sont alors seulement co–localisées entre elles.
Le principe de la spatio–triangulation est illustré dans les figures 2.2(a) et 2.2(b) qui présentent l'orientation des images avant et après cette étape.

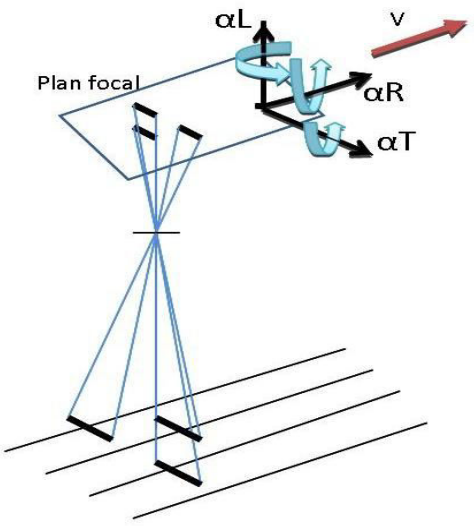
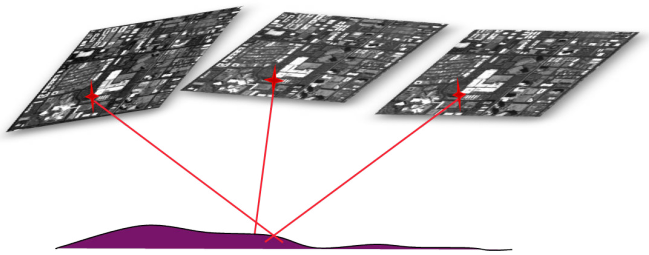
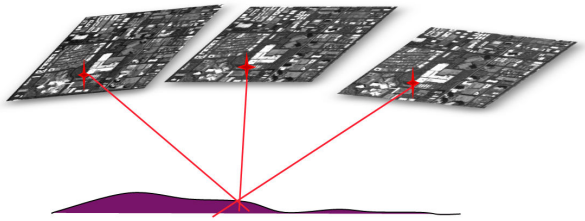



FIGURE 2.1 – *Illustration du système d'acquisition en barrette push–broom et des angles d'attitude du satellite.*



(a) Orientation et localisation des images avant spatio–triangulation.



(b) Orientation et localisation des images après spatio–triangulation.

FIGURE 2.2 – *Illustration du résultat de la spatio–triangulation pour le recalage des images.*

Selon les données fournies avec les images satellites, les modèles géométriques peuvent correspondre au modèle physique des images ou aux fractions rationnelles (RPC).
Le modèle physique est un modèle explicite, calculé en prenant en compte les paramètres de prise de vue du système à chaque instant et défini par les éphémérides du satellite (orbitographie, attitude et orientation du satellite à chaque instant).
Les fractions rationnelles permettent de modéliser la ligne de visée de chacun des pixels afin de retrouver les matrices de passage des coordonnées images aux coordonnées terrain pour n'importe quelle altitude [56].
En géométrie image, on définit par modèle direct, le modèle physique ou analytique, permettant le calcul des coordonnées terrain d'un point à partir de ses coordonnées image (et d'une altitude). Le modèle inverse permet de calculer les coordonnées image d'un point à partir de ses coordonnées terrain.

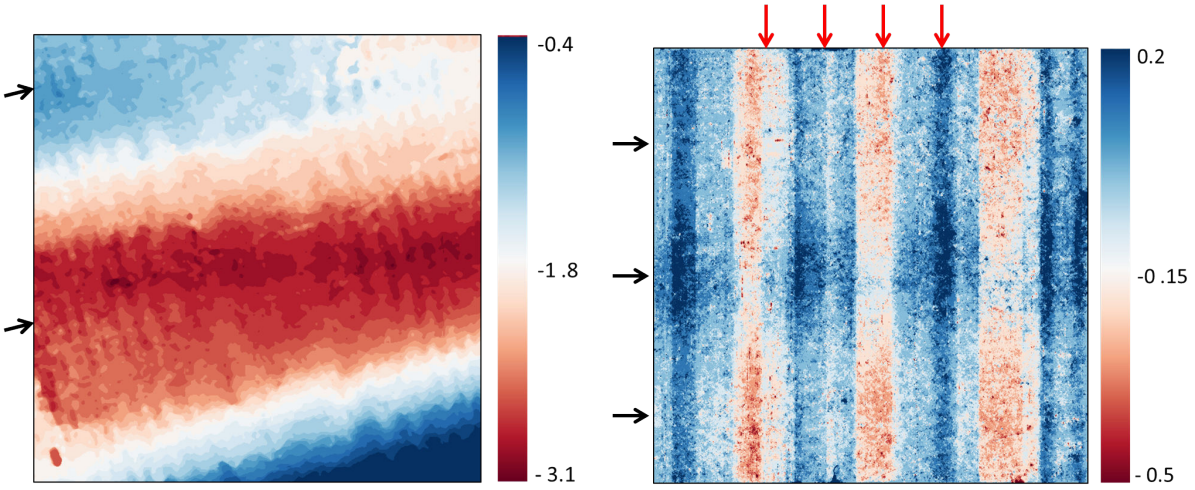
Ainsi, la spatio–triangulation consiste à affiner les modèles géométriques des images, et plus particulièrement les angles d'attitude : roulis, tangage et lacet, en leur introduisant un modèle

d'erreur. L'objectif est de corriger ces modèles afin de garantir la croisée des faisceaux provenant d'un jeu de points de liaison ou d'appuis. Cet objectif peut être atteint si les modèles d'erreurs introduits, sur l'instrument, l'orientation ou l'attitude du satellite, sont en adéquation avec la réalité [55]. Lorsque seuls des points de liaison sont disponibles, des contraintes de localisation absolue sont posées grâce à des *a priori* afin d'éviter que le modèle ne diverge.

Les points d'appuis étant des données rares voire inexistantes, toutes les images de la chaîne de traitements proposée sont recalées entre elles de façon relative.
Une première étape, avant l'affinage des modèles, consiste à rechercher des points de liaison entre les images. Cette étape est réalisée grâce à un outil de corrélation d'images développé au CEA [1]. Cet outil permet la corrélation non dense des images deux à deux afin de déterminer les points homologues les plus fiables (c'est–à–dire ceux montrant le meilleur score de corrélation). Seuls quelques centaines de points, bien répartis sur chaque image, sont nécessaires pour affiner les modèles.
L'étape d'ajustement de faisceaux est ensuite effectuée à l'aide du logiciel Euclidium, développé par le CNES et l'IGN.
Ces outils de corrélation et d'ajustement de faisceaux, déjà existants et validés, fonctionnent de façon totalement automatique pour tous type de capteurs.

Cependant, il apparaît que même après l'affinage des modèles géométriques des images, des erreurs résiduelles persistent. Ces erreurs sont généralement dues à des variations temporelles d'orientation du capteur lors de la prise de vue, impossibles à modéliser et à corriger par la spatio–triangulation du fait du faible nombre de degrés de liberté du modèle d'erreur (biais et dérive temporelle d'attitude).
Un autre défaut, inhérent aux traitements sol, peut aussi représenter une source d'erreur non corrigeable par la spatio–triangulation : c'est le défaut provenant de la simulation en "capteur parfait" des images. En effet, l'acquisition des images satellites est réalisée par segments, illustrés en figure 2.1. Lorsque l'image est produite, les segments de chaque ligne sont fusionnés afin de simuler un capteur linéaire monolitique. Or, des défauts de calibration sur les distances inter–barrettes peuvent subsister et engendrer des erreurs dans les modèles géométriques des images (RPC ou modèles physiques), erreurs qui ne peuvent être modélisées et corrigées par la

1. Cet outil a été développé en interne et aucune publication n'y fait référence.




(a) Défauts résiduels d'une image OrbView-3 après spatio–triangulation.

(b) Défauts résiduels d'une image WorldView-1 après spatio–triangulation.

FIGURE 2.3 – *Illustration du résultat de la spatio–triangulation pour le recalage des images (échelle en pixels).*

spatio–triangulation car elle n'autorise pas de degrés de liberté sur le plan focal.

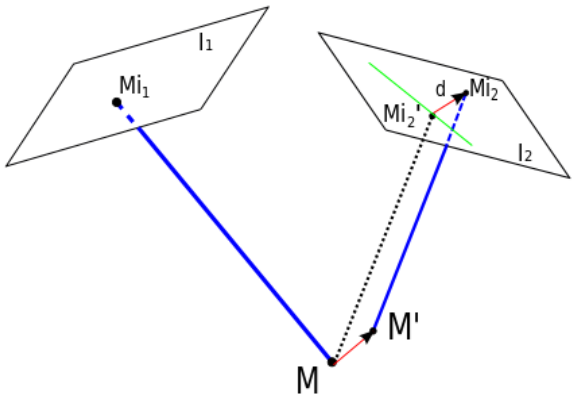



FIGURE 2.4 – *Illustration du calcul des résidus de la spatio–triangulation.*

Dans le but de quantifier ces erreurs, les distances résiduelles entre les faisceaux provenant de points homologues trouvés entre les deux images d'un couple OrbView-3 et d'un couple WorldView-1 ont été calculées (figures 2.3(a) et 2.3(b)).

La figure 2.4 permet d'illustrer le calcul de cette distance pour deux points homologues $Mi_1$ et $Mi_2$, appartenant à deux images différentes $I_1$ et $I_2$. Le recalage des images n'étant pas parfait, les faisceaux provenant de ces points ne se croisent pas. La plus petite distance entre ces faisceaux est représentée par la distance entre les points $M$ et $M'$ dans l'espace terrain. Projetée en géométrie image, cette distance correspond à la distance $d$ entre les points $Mi_2$ (point homologue à $Mi_1$) et $Mi'_2$ (point retrouvé par corrélation épipolaire entre les deux images). Cette distance, calculée après le recalage des images, représente l'erreur résiduelle de la spatio–triangulation.

Cette distance a été calculée après l'affinage des modèles géométriques d'un couple d'images OrbView–3 et d'un couple d'images WorldView–1.
Les figures 2.3(a) et 2.3(b) illustrent les résultats obtenus pour ces deux capteurs. On observe sur la figure 2.3(b) des artefacts linéaires très marqués formant des discontinuités dans l'image (repérés par les flèches rouges en haut de la figure 2.3(b)), d'amplitude d'environ 0,7 pixel pour les images WorldView-1. Ces défauts correspondent aux défauts capteurs non corrigés des images (ils ne sont pas visibles sur la figure 2.3(a)).
Les défauts d'attitude sont, quant à eux, visibles sous formes d'ondulations quasi horizontales et très marquées sur la figure 2.3(a) et horizontales sur la figure 2.3(b). Ces défauts sont repérés par les flèches noires à gauche de chacune des figures.

Ces résidus, d'ordre pixellique pour les images OrbView-3 mettent en évidence des erreurs d'orientation trop importantes sur ce capteur pour notre étude, tandis que les erreurs d'ordre subpixelliques du capteur WorldView-1 sont acceptables et autorisent la mise en correspondance des images pour la génération fiable de MNS.

## 2.2 Outil pour la résolution de problèmes de labellisation

### 2.2.1 Trois problèmes de labellisation

À l'intérieur de cette chaîne de traitements, trois étapes de calcul sont basées sur la résolution d'un problème de labellisation (voir section 1.3). C'est–à–dire qu'à chacune de ces

étapes de calcul, la valeur de chaque pixel de l'image de sortie résulte d'un choix réalisé entre plusieurs valeurs ou attributs possibles selon des critères calculés appelés attaches aux données. Dans une première étape de calcul, lors de la mise en correspondance des images, le pixel homologue à un pixel de l'image maître est choisi parmi plusieurs candidats en fonction d'un critère de similarité. La sélection du pixel homologue permet alors de déduire l'élévation du point terrain représenté par ces pixels homologues (voir chapitre 3, section 3.1).
Dans un second calcul, qui correspond à la fusion des MNS, les pixels du MNS final résultent du choix entre les différentes valeurs d'élévation proposées par chacun des pixels des MNS à fusionner. Ce choix est réalisé en fonction d'un critère de confiance calculé pour chacun des pixels des différents MNS (voir chapitre 3, section 3.3).
Enfin, lors de la détection des changements, c'est en fonction de la valeur du pixel dans la différence de MNS qu'un label de changement (positif, négatif ou non changement) est attribué au pixel de la carte finale des changements (voir chapitre 4).

Ces problématiques, bien que très différentes en termes de labels et d'objectifs, ont pour caractéristique commune que toutes les solutions ne sont pas équiprobables. En effet, à critères d'attache aux données identiques, on privilégie les solutions régulières. Cette régularité se traduit par le fait que, dans une image, la valeur d'un pixel n'est pas indépendante des valeurs des pixels qui l'entourent, de même, le label affecté à un pixel dépend des labels affectés aux pixels voisins. L'ensemble des labels peut ainsi être considéré comme un champ de Markov aléatoire [28, 42, 57, 58] et la probabilité d'affectation d'un label à un pixel dépend alors de la probabilité conjointe de l'appartenance du label à ce pixel et aux pixels de son voisinage. Le problème de labellisation n'est alors plus considéré comme un problème local mais peut être résolu de façon globale sur l'image.
Toutes les solutions de labellisation possibles sur l'image ne sont pas équiprobables. Cependant, il existe une solution optimale qui correspond *a priori* à une solution régulière sur l'image qui limite les variations de hautes fréquences, correspondant généralement à du bruit.
Cette formulation se traduit alors par l'introduction d'une contrainte de régularisation spatiale dans le calcul de labellisation des pixels.

Chacune des labellisations rencontrées dans la chaîne de traitements est alors résolue à

partir d'une même méthode basée sur la minimisation d'une fonctionnelle d'énergie globale $E$ sur l'ensemble des pixels $x$ d'une image $X$ et permettant d'assigner un label $n^j$ appartenant au jeu de labels $N = \{n^1, n^2, ..., n^z\}$, ($z$ étant le nombre total de labels) à chacun des pixels $x \in X$ selon la fonction suivante :

$$E(n) = \sum_{x \in X} C_{data}(n(x)) + \sum_{(x') \in V(x)} C_{reg}(n(x), n(x')). \tag{2.1}$$

$C_{data}(n(x))$ représente le coût pour l'affectation du label $n$ au pixel $x$ en fonction d'un critère de probabilité d'appartenance à ce label.

$C_{reg}(n(x), n(x'))$ correspond au coût pour l'affectation d'un des labels $n$ à deux pixels voisins $x$ et $x'$. Ce terme traduit la contrainte de régularisation.

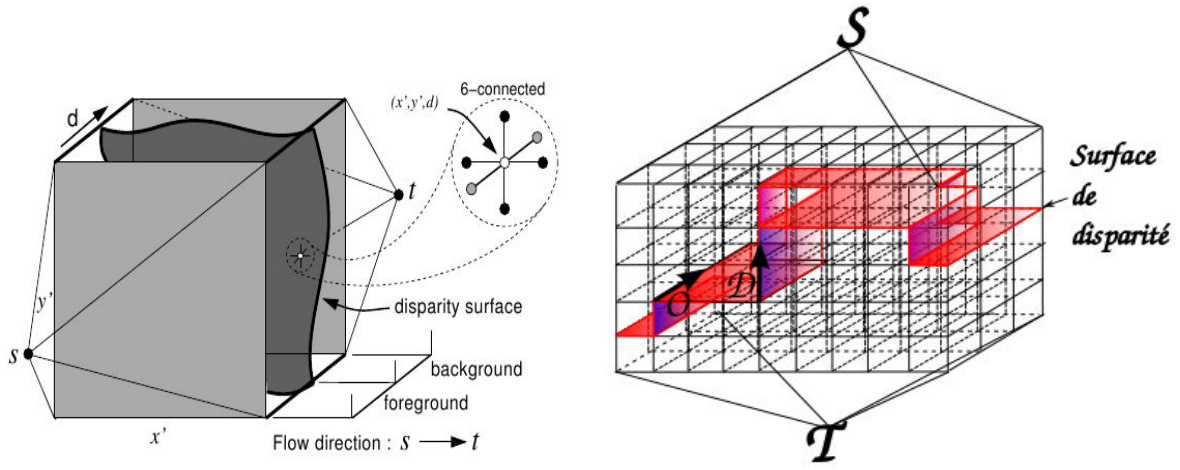
### 2.2.2 Les méthodes d'optimisation existantes

De nombreuses approches existent pour résoudre la minimisation d'une fonctionnelle d'énergie sur une image. Ce sujet a notamment été largement étudié pour résoudre le problème de la mise en correspondance d'images [28], comme nous l'avons vu dans le chapitre 1, section 1.2, mais cette approche est aussi très utilisée en restauration d'image ou en segmentation [12, 58].

Nous rappelons que les algorithmes de minimisation d'énergie peuvent être divisés en deux catégories : les algorithmes permettant de retrouver une solution exacte ou approchée en considérant toute l'image en une seule fois, et les algorithmes qui partitionnent l'image en une série de problèmes plus simples à résoudre (généralement en considérant l'image par séquences de pixels en une dimension) et qui permettent de retrouver des solutions minimales localement qui sont ensuite fusionnées afin d'obtenir le résultat global [43]. Ce sont les méthodes semi–globales [42].

#### Les méthodes globales

Les méthodes les plus populaires d'optimisation globale sont celles basées sur les *graph-cuts* [59], [58] et [41], ou encore le *belief propagation* [43, 60]. Ces méthodes sont parmi les plus efficaces comme l'a démontré Scharstein [28], à partir du benchmark Middelbury [61], puis




(a) Illustration de la construction des graph–cuts. Figure provenant de l'article de Roy et Cox [59]

(b) Illustration du graphe 3D et de la surface de disparité.

FIGURE 2.5 – *Illustration de la méthode des graph–cuts appliquée au calcul de la disparité.*

Schindler [42] sur des problèmes de classification d'images. D'autres méthodes, comme les recuits–simulés (*simulated annealing*), développés par Geman [57], sont souvent jugées moins efficaces et trop lourdes en temps de calcul, notamment parce que les labels de chacun des pixels sont recalculés à chaque itération afin de retrouver un minimum global [58].

**Formulation du problème de mise en correspondance à travers la théorie du *Maximum-flow*.** L'idée des *graph–cuts* est de ramener le problème de minimisation d'énergie à un problème de coupe minimale dans un graphe, ce qui, d'après le théorème de Ford-Fulkerson [62], revient à trouver le flot maximal (ou *MaxFlow*) à travers un graphe, problème pour lequel il existe des algorithmes de calcul exacts et efficaces.

Dans le cas d'un problème de mise en correspondance, Roy et Cox [59] ont développé un algorithme permettant de retrouver globalement et en une seule fois la surface de disparité (représentée en rouge sur la figure 2.5) pour un couple d'images stéréoscopiques.

Avec cet algorithme d'optimisation, et afin de satisfaire une contrainte de régularisation spatiale dans toutes les directions, la mise en correspondance ne peut pas se faire selon une seule dimension à travers la mise en correspondance classique des lignes épipolaires.

L'idée est alors de construire un graphe 3D $(x, y, z)$, $x$ et $y$ représentent la dimension image (ligne,colonne) et $z$ représente les hypothèses d'élévation données par la recherche sur les parallaxes. À ce graphe sont ajoutés une source $S$ et un puits $T$ permettant de former le graphe $G = (V, E)$ avec $V$ l'ensemble des sommets du graphe, c'est à dire la grille 3D formée entre $S$ et $T$, et $E$ l'ensemble des arêtes qui connectent tous les sommets de la grille (représentée sur la figure 2.5(b)).

A chaque arête est affecté un coût (ou une capacité de flot) :

- les arrêtes verticales (nommées $D$ sur la figure 2.5(b)) sont les arêtes dites de disparité et correspondent directement au coût de mise en correspondance : un coefficient de corrélation élevé entre deux pixels va permettre la saturation de l'arête correspondante par le flot maximal et inversement,
- les arêtes horizontales (nommées $O$ sur la figure 2.5(b)) sont les arêtes d'occlusion ($O$) qui contrôlent la régularisation et permettent de limiter les variations de la surface de disparité selon le poids de la régularisation,
- les arêtes connectées à la source ou au puits sont, quant à elles, de capacité infinie.

Toute solution $Z = f(x)$ représente une coupe minimale qui divise le graphe selon deux ensembles :

- $Z > f(x)$, c'est l'ensemble appelé *background* dans la figure 2.5(a),
- $Z < f(x)$, c'est l'ensemble appelé *foreground* dans la figure 2.5(a).

On montre que le coût de la coupe minimale définie précédemment est calculé exactement par la formule précédente (équation 2.1).

Ce problème peut donc être formulé comme un problème de flot maximum pour lequel Cox et Roy proposent une solution optimisée [59].

Kolmogorov [63] a cependant montré que l'application des *graph–cuts* est soumise à une condition dite de "sous–modularité" de la fonction d'énergie de régularisation. En effet, les poids des arêtes du graphe doivent être forcément positifs de la source vers le puits afin que l'algorithme converge, cette condition se traduit par l'équation suivante :

$$C_{reg}(0,0) + C_{reg}(1,1) \leq C_{reg}(0,1) + C_{reg}(1,0). \tag{2.2}$$

Avec $C_{reg}$ le coût d'affectation de deux labels binaires (0 ou 1) à deux pixels voisins. Si l'énergie

de régularisation ne satisfait pas ces contraintes, la solution optimale ne peut être garantie. Cette contrainte peut limiter la gamme de fonctions de coût possible pour la résolution d'un problème d'optimisation par *graph–cuts*.

**Autres approches d'optimisation globale** Un autre algorithme, largement utilisé pour l'optimisation globale est le *belief propagation* [39], [33] ou la version *loopy belief propagation* proposée par Felzenszwalb [43]. Cette méthode fonctionne par passage de "messages" entre le graphe défini par les quatre connexes d'un pixel. Chaque message est un vecteur de dimension égale au nombre de labels en jeu et fonction du coût minimum obtenu pour chaque label, considérant leurs probabilités sur les quatre connexes du pixel. Cependant, Kolmogorov montre que la convergence de cet algorithme n'est pas toujours assurée [64, 65] et la solution trouvée apparaît souvent de plus haute énergie que les *graph–cuts* et donc non optimale. De plus, le temps de calcul nécessaire reste particulièrement long [64], bien que la variante de Felzenszwalb permette de réduire ce temps [43].
Une méthode proche du *belief propagation* a été récemment introduite, le *tree-reweighted message passing* [65]. Cette méthode apparaît moins gourmande en énergie puisque les "messages" ne sont plus passés entre un graphe formé des quatre connexes mais à travers un graphe formé par un arbre. Kolmogorov [65] et Szeliski [64] ont montré que cette méthode était plus efficace que les *graph–cuts* sur les jeux de test de Middelbury mais reste néanmoins plus lente que les *graph–cuts*.

De manière générale, les *graph–cuts* et autres méthodes d'optimisation globales nécessitent de très grandes capacités mémoire (notamment lorsque plusieurs labels sont en jeu) et sont alors difficilement applicables sur des images complètes de grande emprise telles que les images satellites très haute résolution [42].

### Les méthodes semi–globales

D'autres procédés permettent la résolution de problèmes d'optimisation, non pas en recherchant une solution exacte ou approchée au minimum global sur toute l'image simultanément mais en fonctionnant itérativement sur la recherche de minimas locaux sur l'image à partir de

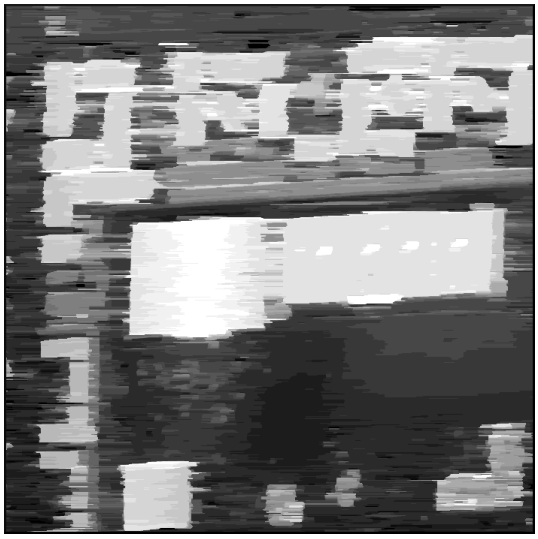


FIGURE 2.6 – *Illustration d'un effet de streaking le long des lignes d'une image traitée par programmation dynamique.*

fonctions qui permettent une approximation de la fonction globale. L'avantage de ces méthodes est de réduire largement la complexité des problèmes en les divisant en sous–problèmes plus faciles et rapides à résoudre [43].
La méthode la plus populaire est la programmation dynamique. Cette méthode est aujourd'hui appliquée dans de nombreux domaines de traitement d'image [31, 34, 44, 66].

Appliquée à une image, l'idée de base de la programmation dynamique consiste à optimiser une fonction d'énergie sur chacune des lignes indépendamment [34]. L'inconvénient majeur de cette méthode est le manque de cohérence entre chacune des lignes, créant des effets de *streaking* ou artefacts linéaires, le long des lignes dans le résultat lorsqu'un label a tendance à se propager le long de la ligne du fait de la régularisation spatiale.
Ce phénomène est illustré sur le MNS présenté en figure 2.6 où la propagation de valeurs d'élévation le long des lignes, sans cohérence avec les colonnes, est fortement marquée.

Afin de limiter ce problème, plusieurs approches ont été développées : Bobick et Intille [34], dans un contexte de mise en correspondance d'images, détectent préalablement les contours de l'image et adaptent les coûts le long de ces contours afin d'aligner les disparités dans l'image finale. La difficulté réside alors dans la détection précise des contours de l'image à prendre en

compte.

Ohta [67] procède d'abord à une recherche des disparités selon les lignes puis utilise les résultats obtenus afin de contraindre une recherche selon les colonnes de l'image. Les résultats obtenus sont alors additionnés pour sélectionner la disparité finale de chaque pixel.

Veksler [44] applique la programmation dynamique sur des structures en arbre, choisies en fonction des contours les plus importants de l'image. Cet algorithme permet alors de minimiser l'énergie selon deux dimensions afin de se rapprocher au mieux d'une optimisation globale. L'inconvénient de cette méthode réside ici dans le choix des structures d'arbre utilisées et de leur initialisation.

Hirschmüller [31] applique la programmation dynamique selon plusieurs directions sur l'image, indépendamment les unes des autres, puis additionne les coûts obtenus pour chaque label afin de calculer le coût final de chaque disparité pour chacun des pixels de l'image. Cet algorithme permet le calcul rapide de l'optimisation et prend en compte tout le voisinage des pixels puisque toutes les directions de l'image sont calculées. Cette optimisation quasi globale est appelée le *semi–global labelling* [31, 42].

Schindler, [42] compare plusieurs méthodes de classification d'images basées sur (i) un algorithme de graph–cuts tel que celui proposé par Boykov [58], (ii) un algorithme de *semi–global labelling* basé sur la méthode de Hirschmüller [31] et (iii) des méthodes de classification locale, sans contrainte spatiale.

Il apparaît d'abord que les méthodes d'optimisation globales sont plus efficaces que les méthodes de classification locales : les résultats obtenus montrent une précision générale de classification de 70,5% à 80% pour une méthode locale, 72% à 80% pour une méthode semi-globale et de 72,5% à 85% pour une méthode de *graph–cuts*. Cependant, si les méthodes graph–cuts présentent la meilleure efficacité, Schindler observe que les méthodes de *semi–global labelling* montrent des performances très proches tout en maintenant un temps de calcul inférieur, ce qui en fait une bonne alternative aux *graph–cuts* pour des labellisations sur de nombreuses étiquettes et une large emprise.

Le *semi–global labelling* apparaît donc comme une méthode d'optimisation suffisamment performante pour des questions de labellisation et très adaptée dans un contexte de traitement

rapide et automatique puisque ses performances en termes de temps de calcul et d'allocation mémoire sont supérieures à celles des méthodes globales. Ce type de méthode a donc été sélectionné pour la résolution des problèmes de labellisation rencontrés dans notre chaîne de traitements.

### 2.2.3 Description de l'algorithme sélectionné

L'algorithme de programmation dynamique utilisé dans ce travail est basé sur la programmation dynamique sur une séquence, algorithme décrit par Felzenszwalb [66]. Une séquence correspond alors à un sous–ensemble de pixels de l'image.

Afin de réduire les effets de *streaking* de la programmation dynamique, les minimisations sont effectuées indépendamment sur plusieurs directions de l'image : lignes, colonnes, diagonales et directions intermédiaires, telles que représentées sur la figure 2.7 et selon la technique proposée par Hirschmüller et Pierrot-Deseilligny [31, 32]. Tous les coûts, obtenus par label pour chaque pixel et selon chaque direction sont ensuite fusionnés afin d'obtenir la solution semi–globale sur toute l'image. De cette façon, et en parallélisant les calculs dans chaque direction, l'optimisation est effectuée efficacement selon un temps polynomial.

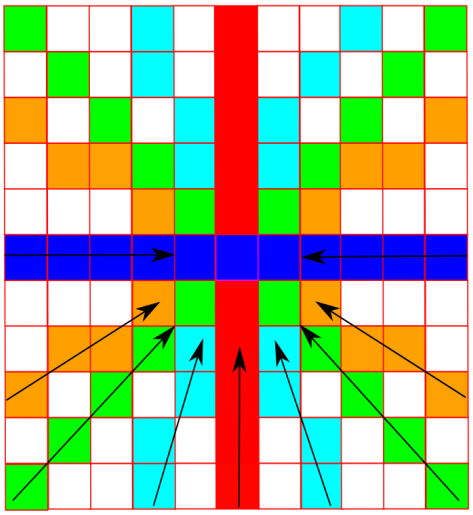


FIGURE 2.7 – *Illustration du parcours des différentes directions pour la programmation dynamique afin de calculer le coût "semi–global".*

**Algorithme de minimisation selon une direction**

Soit une séquence linéaire de pixels $X = \{x_0, x_1, ..., x_i, ..., x_f\}$ le long d'une direction de l'image. Le but est d'assigner à chacun des pixels $x_i$ un label $n(x_i)$ appartenant au jeu de labels $N = \{n^1, n^2, ..., n^z\}$, $z$ étant le nombre total de labels.

La solution $S$ de labellisation sur la séquence de $x_0$ à $x_f$ correspond alors à la solution qui minimise les coûts de labellisation en chaque pixel de la séquence, selon l'équation :

$$S(n(x_0), ..., n(x_f)) = \sum_{x_0}^{x_f} C_{\text{data}}(x, n(x_i)) + \lambda \sum_{x_0}^{x_{f-1}} C_{\text{reg}}((x_i, n(x_i)), (x_{i+1}, n(x_{i+1})). \quad (2.3)$$

$C_{\text{data}}(x, n(x_i))$ représente la fonction d'attache au données, c'est la probabilité que le label $n$ corresponde au pixel $x_i$ de la séquence.

$C_{\text{reg}}((x_i, n(x_i)), (x_{i+1}, n(x_{i+1}))$ représente le coût de régularisation entre deux pixels consécutifs $x_i$ et $x_{i+1}$ sur la séquence et portant un label $n$.
Cette fonction de régularisation est souvent calculée selon le modèle de Potts [68] dans lequel le coût affecté au label $n$ est égal à 0 si les deux labels consécutifs sont identiques et 1 sinon (équation 2.4).

$$C_{\text{reg}}((x_i, n(x_i)), (x_{i+1}, n(x_{i+1}))) = \begin{cases} 0 & \text{if } n(x_i) = n(x_{i+1}), \\ 1 & \text{if } n(x_i) \neq n(x_{i+1}). \end{cases} \quad (2.4)$$

Ce modèle est utilisé lors de l'étape de détection des changements (chapitre 4) mais nous verrons que les étapes de mise en correspondance (chapitre 3 section 3.1) et de fusion des MNS (section 3.3) nécessitent des fonctions de régularisation plus complexes.

La variable scalaire $\lambda$ représente, quant à elle, le poids accordé à la contrainte de régularisation : plus cette valeur est importante, plus la contrainte spatiale sera renforcée, c'est–à–dire que le coût d'affectation d'un label, différent du label précédent, sera augmenté. En pratique, nous verrons que cette variable est définie en fonction du résultat recherché qui doit être plus ou moins régulier selon le contexte.

Le principe de l'algorithme est de calculer le meilleur chemin passant par chaque couple pixel–label $(x_i, n^j)$, selon un parcours de la séquence de pixels de $x_0$ à $x_f$. L'objectif est de

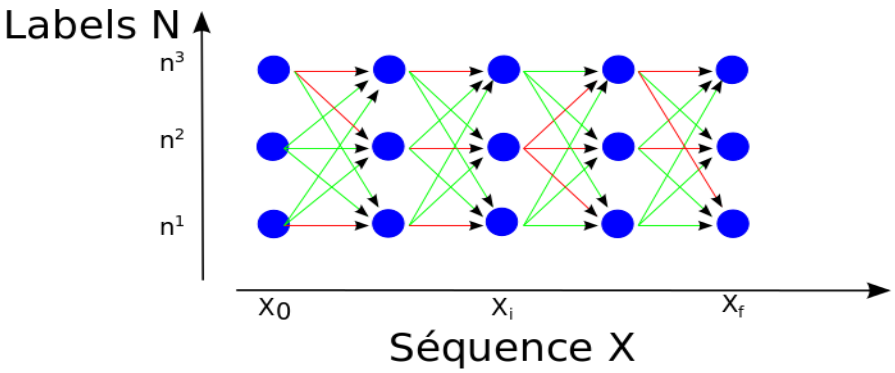



FIGURE 2.8 – *Illustration du calcul des coûts obtenus pour chaque label et chaque pixel de la séquence selon l'équation 2.5. Les flèches vertes représentent chacun des coûts calculés, les flèches rouges représentent le coût le plus faible obtenu entre un label d'un pixel et tous les labels du pixel précédent.*

calculer, pour un couple pixel–label $(x_i, n^j)$, le coût pour le couple suivant $(x_{i+1}, n^j)$ (figure 2.8).

Soit $A(x_0, n^j)$ le coût initial du label $n^j$ au pixel $x_0$ de la position 0 est défini par :

$$A(x_0, n^j) = C_{\text{data}}(x_0, n). \tag{2.5}$$

À la position $x_i$ de la séquence, le coût $A(x_i, n^j)$ pour lui assigner un label $n^j$ se calcule selon l'équation suivante :

$$A(x_i, n^j) = C_{\text{data}}(x_i, n^j) + \min_{n^k}\Big[A(x_{i-1}, n^k) + \lambda C_{\text{reg}}(n^j, n^k)\Big], \tag{2.6}$$

$n^j$ et $n^k$ deux labels appartenant au jeu de labels $N$. Ainsi, pour chacun des labels $n^j$ ($j \in [1 : z]$) proposés à chacun des pixels $x_i$, le coût obtenu correspond à tous les coûts cumulés du chemin de $x_0$ à $x_i$ qui minimisent le coût en $n^j$.

Dans la figure 2.8, les flèches vertes représentent tous les chemins possibles pour parvenir à chacun des labels. Le chemin amenant le coût minimal en chaque label de chaque pixel est représenté par les flèches rouges.

Une fois tous ces coûts calculés, le meilleur chemin de $x_0$ à $x_f$ correspond au chemin ayant

amené le label de coût minimal $n_{x_f}^{j*}$ au dernier pixel $x_f$, tel que :

$$n^*(x_f) = \underset{n}{\operatorname{argmin}}\, A(x_f, n(x_f))). \tag{2.7}$$

La matrice finale contenant la somme des coûts de chacun des labels et pour chaque pixel est alors parcourue afin de retrouver le label amenant le coût minimal sur chaque pixel.

**Généralisation de l'algorithme à plusieurs directions**

Le coût semi–global est défini comme le coût de chaque label obtenu à travers le sous–graphe déterminé par toutes les directions calculées dans l'image.
Selon cette définition, la minimisation est appliquée sur des séquences linéaires de pixels dans toutes les directions imposées et de façon indépendante entre les séquences comme illustré sur la figure 2.7. Les matrices de coûts, obtenues pour chaque pixel et selon chaque direction sont ensuite fusionnées afin d'obtenir un coût final pour chaque label et pour chaque pixel, tel que le décrivent Hirschmüller et Pierrot-Deseilligny [31, 32].

Or, les coûts obtenus dans les matrices ne peuvent être directement comparables entre eux puisque les coûts sont cumulés lors des parcours de chaque séquence de pixels qui ne possèdent pas forcément le même nombre de pixels. Afin de rendre possible cette comparaison de coût entre les labels d'un même pixel sur les différentes directions, il est nécessaire de les recalculer selon une même échelle. Cette opération se traduit par le calcul de la distance à la solution optimale de labellisation, $C(n_x)^d$ sur toutes les séquences obtenues, selon la formule :

$$C(x_i, n(x_i))^d = C(x_i, n(x_i)) - C(x_i, n^*(x_i)). \tag{2.8}$$

Ces distances à la solution optimale par label peuvent alors être additionnées pour chaque label et selon chacune des directions. Le label final $n^{d*}$ choisi pour chaque pixel $x_i$ et dans toutes les directions est finalement celui ayant le coût minimal, tel que :

$$n^{d*}(x_i) = \underset{n}{\operatorname{argmin}} \sum_{dirs} C(x_i, n(x_i))^d. \tag{2.9}$$

Le nombre de directions peut avoir un impact important sur le résultat final. Cependant, à partir d'un certain nombre de directions traitées, les différences entre les résultats diminuent fortement, elles sont presque nulles au delà de 12 directions. Or, le nombre de directions traitées

augmente sensiblement le temps de calcul de la méthode, il est donc nécessaire de trouver le nombre de directions adéquat afin d'obtenir le meilleur équilibre entre temps de calcul et précision du résultat.
Dans la plupart des problèmes traités dans ces travaux, un nombre de 8 ou 12 directions sera utilisé. Une comparaison des résultats obtenus avec 1, 4, 8, 12 et 16 directions est présentée dans le chapitre 5.

## 2.3 Conclusion

La spatio–triangulation permet l'affinage des modèles géométriques des images afin de corriger les erreurs d'orientation et d'attitude dues à une restitution imparfaite des informations sur la position du satellite, son orientation et sa direction de visée lors de l'acquisition. L'objectif de cette opération, réalisée simultanément sur toutes les images disponibles, est de permettre un recalage relatif précis entre toutes ces données.
Cette étape est réalisée à l'aide d'outils existants permettant la recherche fiable de points homologues puis l'ajustement des faisceaux.
Cependant, nous avons montré que des défauts, de capteur ou d'attitude, persistent dans les modèles affinés. Ces erreurs peuvent néanmoins être acceptables lorsqu'elles restent sub–pixelliques comme c'est le cas pour les images WorldView-1 que nous utilisons.

Le second outil que nous avons détaillé, au cours de ce chapitre, est un algorithme de programmation dynamique que nous avons implémenté et adapté pour le calcul de trois des traitements de la chaîne développée au cours de cette thèse : la mise en correspondance, la fusion des MNS et la détection des changements d'élévation.
Cet algorithme de programmation dynamique a été sélectionné pour sa facilité d'implémentation par rapport à d'autres méthodes d'optimisation globales classiques telles que les *graph–cuts* mais aussi et surtout pour la rapidité d'exécution des calculs et son efficacité reconnue.

# Chapitre 3

# Génération de Modèles Numériques de Surface sur une grille régulière à partir de couples stéréoscopiques

Comme nous l'avons vu dans le chapitre 1, de nombreuses techniques existent pour la génération de MNS en fonction des contraintes associées aux données d'entrée et des sources d'erreurs à prendre en compte. La stratégie que nous avons choisie est basée sur trois étapes de calcul. La première étape consiste en une mise en correspondance des images à l'aide d'un outil existant, MicMac, et dont les options que nous utilisons seront décrites dans une première partie de ce chapitre.

Puis nous détaillerons la méthode que nous avons développée afin de basculer les MNS sur une grille régulière terrain tout en localisant et filtrant certaines erreurs de mise en correspondance.

Dans une troisième partie, nous présenterons certaines erreurs de mise en correspondance non corrigées, notamment dues aux zones d'occlusion de l'image esclave et nous décrirons la méthode que nous avons développée pour résoudre ce problème et améliorer la précision des MNS obtenus. Cette méthode est basée sur la fusion des deux MNS générés à partir du même couple d'images stéréoscopiques par inversion des rôles maître et esclave des images. Nous détaillerons notamment l'algorithme permettant cette fusion à travers la labellisation des pixels du MNS final, contrôlée par une contrainte de régularisation spatiale.

Puis, dans une dernière partie, nous analyserons les résultats obtenus à travers la comparaison des MNS générés avec une vérité LiDAR, acquise à une date proche de nos données satellites, sur la ville de Christchurch en Nouvelle-Zélande.

## 3.1 Description de la stratégie de mise en correspondance adoptée

### 3.1.1 Technique de mise en correspondance

La méthode de mise en correspondance d'images que nous avons sélectionnée est basée sur la géométrie image (représentée figure 1.1(b), p34). Comme cité dans le chapitre 1, section 1.2.3, cette géométrie ne nécessite pas le ré–échantillonnage complet de l'image en géométrie épipolaire et se montre plus robuste que la géométrie terrain aux erreurs géométriques des images ainsi qu'aux zones d'occlusion lorsque seulement deux images sont en jeu [48]. Cette robustesse est un avantage certain pour notre chaîne de traitements dans laquelle toute erreur de MNS (planimétrique ou altimétrique) peut avoir des conséquences sur la carte finale des changements de la scène.

Le principe de la géométrie image réside dans le parcours des lignes de visée des pixels dont la modélisation a été affinée lors de l'étape de spatio-triangulation (chapitre 2,section 2.1).
Pour chacun des pixels de l'image maître (généralement l'image la plus au nadir) et de coordonnées $x_M(l_M, c_M)$, la ligne de visée est parcourue de façon discrète avec un pas altimétrique déterminé par l'utilisateur. Chaque pas définit un point $P$ dans l'espace, de coordonnées $(P_x, P_y, P_z)$. La ligne de visée de l'image esclave passant par ce point de l'espace est alors déterminée à partir du modèle géométrique inverse de l'image esclave et le pixel correspondant dans cette image, de coordonnée $(l_E, c_E)$ est localisé.
La corrélation entre les pixels est alors effectuée à travers des fenêtres de corrélation centrées sur chacun des pixels de l'image maître et de l'image esclave et de taille spécifiée par l'utilisateur (ici une fenêtre de rayon $R = 2$ pixels est utilisée). L'imagette de l'image esclave est alors ré–échantillonnée dans la géométrie de l'imagette maître. Lorsque le pixel homologue est sélectionné, on obtient alors, pour le pixel considéré dans l'image maître, une valeur directe de

l'altitude du point.

Toutefois, la génération de MNS dans la géométrie de l'image maître présente un inconvénient. En effet, puisque la recherche de l'élévation est réalisée à partir de l'image maître, la carte des élévations est générée dans la géométrie de l'image maître et il est alors nécessaire de les ré–échantillonner dans une nouvelle grille géographique régulière afin que les MNS produits à partir de différentes images maître soient comparables entre eux pixel à pixel.

### 3.1.2 Fonction d'optimisation et régularisation

Parmi les méthodes d'optimisation implémentées dans l'outil MicMac, nous avons choisi une méthode basée sur la programmation dynamique, notamment pour des raisons de temps de calcul (cf. chapitre 2 section 2.2.2). L'équation générale de cette optimisation est présentée en section 2.2.3

Lors de la mise en correspondance, les labels correspondent à toutes les élévations $z$ testées et le coût d'attache aux données $C_{data}(x_{M_i}, z(x_{M_i}))$, c'est–à–dire le coût pour assigner une élévation $z$ à un pixel $x_{M_i}$ de l'image maître, est une fonction du score de corrélation obtenu entre le pixel de l'image maître $x_{M_i}$ et le pixel homologue trouvé dans l'image esclave $x_E$. Il est ici calculé avec le NCC (équation 3.1) :

$$C_{data}(x_{M_i}, z(x_{M_i})) = NCC(x_{M_i}, x_E). \tag{3.1}$$

Soit deux vecteurs $u_M$ et $u_E$ constitués de l'ensemble des $q$ valeurs des vignettes de corrélation centrées en $x_{M_i}$ et en $x_E$, respectivement. $p_w^{ds}$ est une fonction de pondération, $w \in [1 : q]$. Pour chaque vecteur $u_M$ de valeurs $U_w$, on pose :

$$R(u_M) = \frac{\sum_{w=1}^{q} p_w^{ds} U_w}{\sum_{w=1}^{n} p_w^{ds}} \tag{3.2}$$

La définition est identique pour chaque vecteur $V_E$ de valeurs $V_w$.

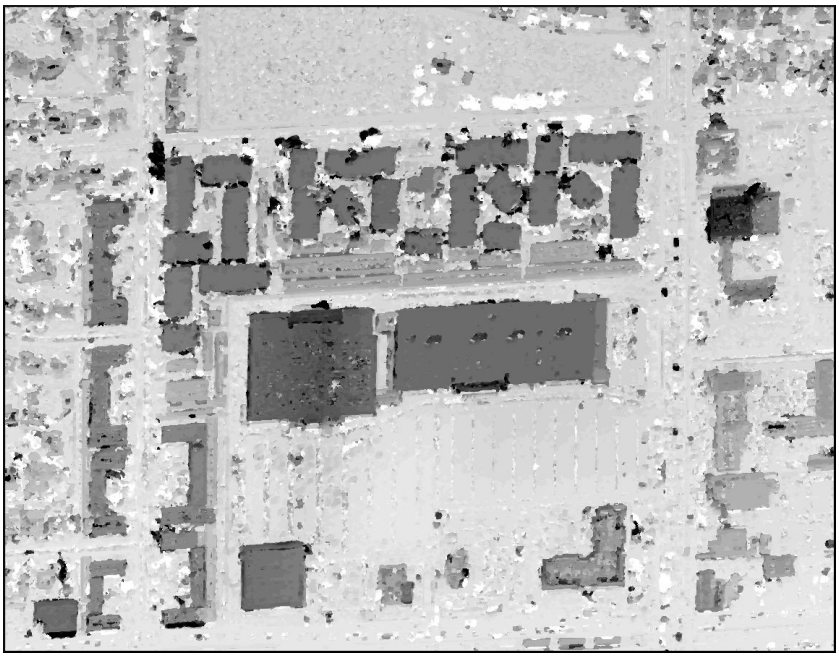
Le coefficient de corrélation s'écrit :

$$NCC(u_M, v_E) = \frac{R(u_M v_M) - R(u_M)R(v_E)}{\sqrt{(R(u_M^2) - R(u_M)^2) * (R(v_E^2) - R(v_E)^2)}} \tag{3.3}$$

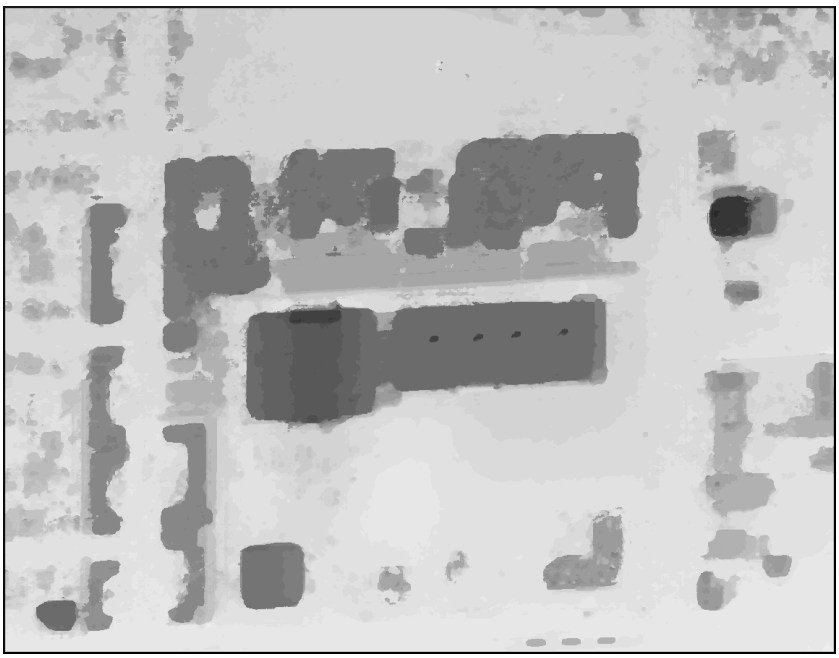
Lors de la génération de MNS, la régularisation a pour but de limiter les variations d'altitude. Ainsi, cette fonction est calculée à partir de la différence d'altitude obtenue entre deux pixels

consécutifs de l'image maître $x_{M_i}$ et $x_{M_{i-1}}$ selon l'équation suivante :

$$C_{reg}((x_{M_i}, z(x_{M_i})), (x_{M_{i-1}}, z(x_{M_{i-1}}))) = f(\mid z(x_{M_i}) - z(x_{M_{i-1}}) \mid). \tag{3.4}$$



(a) Carte des élévations d'un couple WorldView-1, régularisation $\lambda = 0.02$.



(b) Carte des élévations d'un couple WorldView-1, régularisation $\lambda = 1.0$.

FIGURE 3.1 – *Illustration de l'impact de la régularisation pour la génération de MNS.*

Le paramètre de régularisation est un paramètre particulièrement important dans la géné-

ration des MNS puisqu'il contrôle les variations d'altitudes des MNS générés. En effet, plus sa valeur est importante, plus le poids de la régularisation spatiale sera important, faisant disparaître les structures les plus fines (et le bruit) du MNS et flouter les bords de bâtiments. Ce phénomène est bien représenté en figure 3.1(b). Inversement, une faible régularisation va permettre d'observer les variations plus fines d'altitude (et ainsi de séparer les bâtiments très proches les uns des autres), mais en augmentant la présence de bruit (figure 3.1(a)).
Ainsi, en milieu urbain, milieu dans lequel les discontinuités et les variations fines d'élévation doivent être respectées afin d'obtenir un MNS précis au niveau des bâtiments, la régularisation est définie à une valeur faible, de même pour un milieu fortement escarpé. Inversement, en milieu rural, désertique ou vallonné, la régularisation peut être largement augmentée.
L'outil MicMac permet d'adapter aisément ce paramètre qui fait ainsi partie des paramètres de notre chaîne de traitements.

### 3.1.3 Prise en compte des zones d'occlusion

La méthode choisie pour la prise en compte des zones d'occlusion est basée sur une localisation post–traitement, après la génération de la carte d'élévation de la scène en géométrie image. En effet, lors du ré–échantillonnage de la carte des élévations sur une grille régulière, étape dite de "basculement", nous avons développé une méthode permettant de localiser ces zones en fonction des angles d'acquisition de l'image maître et de la présence de discontinuités.
Le basculement a ainsi un rôle majeur dans la génération du MNS final.

## 3.2 Développement d'une méthode de basculement des MNS sur une grille géographique régulière

Comme précisé précédemment, selon cette technique de mise en correspondance, la carte des élévations est calculée dans la géométrie de l'image maître. Afin de comparer des cartes d'élévations obtenues sur une même zone selon différentes dates, différents capteurs ou même selon différentes techniques de génération de MNS, il est nécessaire de ré–échantillonner ces cartes d'élévations selon une grille terrain régulière et commune à toutes les cartes produites. C'est l'étape de basculement.
Généralement, ce ré–échantillonnage est effectué par interpolation des points de la grille

régulière. Ainsi, l'élévation de chacun des points de cette grille régulière est interpolée à partir des coordonnées terrain des trois points les plus proches de lui provenant de la carte des élévations calculée. Pour ce faire, les pixels de la grille image sont parcourus par triplets. Cette technique d'interpolation est la plus commune, notamment parce que la maille d'origine est une maille carrée.

Le basculement que nous proposons est aussi basé sur une interpolation des points de la grille régulière à partir du nuage de points formé par la carte des élévations en géométrie image. Mais, avant chaque interpolation d'un point de la grille à partir des points les plus proches, certaines conditions sont vérifiées afin de déterminer si le point est valide et peut être calculé ou non. Ces conditions reposent sur des contraintes géométriques permettant la détection des façades et des zones d'occlusion et une contrainte qualitative sur le score de corrélation obtenu lors de la mise en correspondance. Elles ont pour but de limiter les erreurs sur le MNS final produit.

### 3.2.1 Interpolation des points de la grille régulière

L'étape de basculement repose sur une interpolation de l'élévation de chacun des points de la grille régulière à partir des trois points les plus proches formant un triangle autour du pixel considéré, le système est illustré en figure 3.2. Ce triplet de point provient du nuage de points obtenu lors de la mise en correspondance et dont les coordonnées planimétriques ont été calculées grâce aux modèles géométriques des images.

Les pixels de l'espace image, considérés par triplets, de coordonnées ligne $l$ et colonne $c$ tel que $P_1^I = (l, c)$, $P_2^I = (l, c+1)$ et $P_3^I = (l+1, c)$ correspondent alors à des coordonnées $x$ $y$ et $z$ dans l'espace terrain $T$ tel que $P_1^E = (x_1, y_1, z_1)$, $P_2^E = (x_2, y_2, z_2)$ et $P_3^E = (x_3, y_3, z_3)$ (sommets des triangles noirs sur la figure 3.2).

Si un point de la grille terrain régulière (en rouge sur la figure 3.2) se situe à l'intérieur du triangle, les points de coordonnées $P = (x_P, y_P, z_P)$ sont alors interpolés par une interpolation linéaire à partir des coordonnées des trois sommets du triangle $P_1^E$, $P_2^E$ et $P_3^E$.

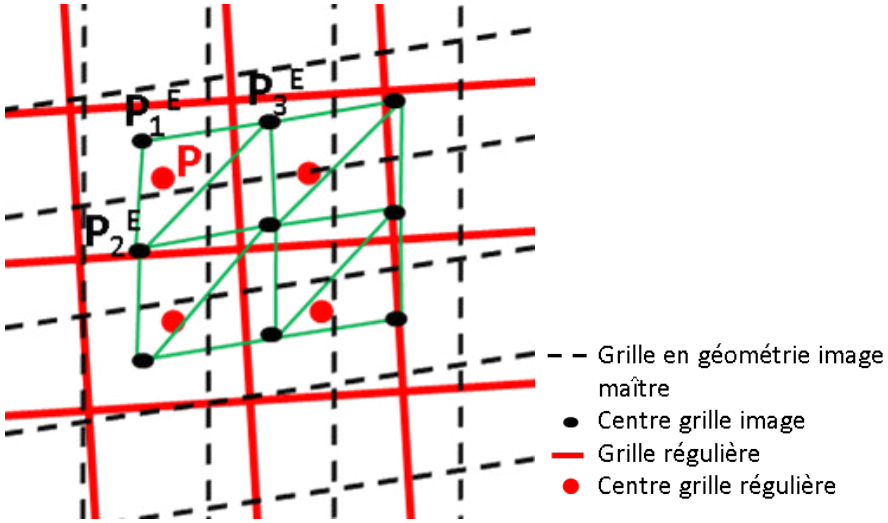



FIGURE 3.2 – *Principe de parcours de la grille image maître pour l'interpolation des points de la grille régulière.*

### 3.2.2 Conditions au calcul de l'élévation d'un point

Les images ayant été acquises avec un angle d'incidence pouvant être fort, deux difficultés doivent être considérées :

- les façades de bâtiments, visibles sur l'image maître, génèrent plusieurs points d'élévations différentes mais avec les mêmes coordonnées planimétriques,
- les zones d'occlusion des images doivent être localisées afin de ne pas interpoler les points correspondants.

Pour le premier cas, lors du calcul de l'élévation d'un point situé à l'intérieur d'un triangle d'interpolation tel que le point $P$ de la figure 3.2, si il apparaît qu'une élévation a déjà été déterminée à ces mêmes coordonnées planimétriques, cela peut signifier que ce point appartient à une façade de bâtiment. En effet, le long d'une façade, les points possèdent les mêmes coordonnées planimétriques mais des coordonnées altimétriques différentes.
Dans ce cas, seul le point correspondant à l'élévation maximum trouvé doit être pris en compte, ceci afin de retrouver l'élévation la plus haute du bâtiment (son toit). Cette technique de recherche de la plus haute élévation d'un objet est connue sous le nom de *Z-Buffer* dans la littérature [69].

Pour le second cas, les zones d'occlusion doivent être localisées afin de ne pas interpoler les points de ces zones ce qui entraînerait des erreurs.
Dans ce but, nous avons mis en place une technique de détection des zones d'occlusion basée sur leur localisation à partir de l'analyse des angles d'acquisition (élévation et azimut) de l'image maître. La figure 3.3 illustre le principe de localisation utilisé.

Sur la figure 3.3, les points $P_1^E$, $P_2^E$ et $P_3^E$ représentent les points dans l'espace terrain correspondant à un triplet de pixels de l'image maître. Le point $P$ correspond au point de la grille régulière à interpoler.
Dans une première étape, les deux points du triplet de points correspondant à l'élévation maximum et minimum dans le triangle sont déterminés, si la différence d'élévation $D_h$ obtenue sur le triangle est inférieure à 1 m (résolution altimétrique des MNS), le point $P$ de la grille terrain situé dans le triangle peut être calculé directement car le terrain est considéré comme quasiment plat à cet endroit.

Si une différence d'élévation plus importante est relevée, il est alors nécessaire de définir si une zone d'occlusion est présente à cet endroit et quelle est son emprise afin de décider si le triangle est viable pour le calcul d'un point.
En premier lieu, on définit le point le plus élevé ($P_1^E$ dans la figure 3.3) et le plus bas (point $P_3^E$) du triplet de points.
La droite $D_M$ reliant la projection de $P_1^E$, $P_1^{E'}$ sur le plan horizontal et le point $P_3^E$ est calculée. Cette droite représente la distance au sol entre le point le plus haut et le point le plus bas.
Puis, les angles d'azimut $a$ et d'incidence $i$ sont définis pour le point $P_1^E$. On détermine alors la droite $D_i$, projection sur le plan horizontal du vecteur défini dans l'espace par les angles d'azimut et d'incidence au point $P_1^E$. La droite $D_i$ représente alors l'orientation et la longueur au sol de la zone d'occlusion.
Le vecteur $D_M$ précédemment obtenu est alors projeté sur le vecteur $D_i$ en un vecteur $D_x$. Cette projection permet de situer le triplet de points considérés en fonction de la zone d'occlusion, i.e si les vecteurs sont opposés le point peut être calculé car il n'y a pas de zone d'occlusion (sur l'image maître). Si les vecteurs sont dans le même sens et que $D_x$ est plus grand que $\frac{Res_{plani}}{2}$, $Res_{plani}$ étant la résolution planimétrique du MNS, alors le point ne peut pas être calculé car il

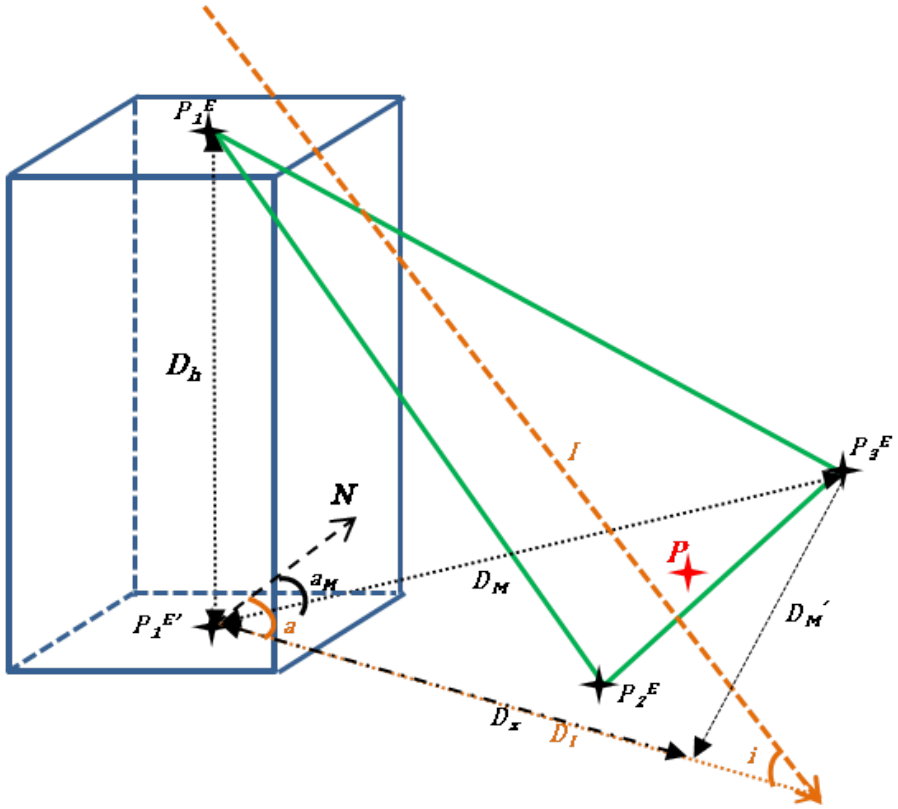



FIGURE 3.3 – *Illustration du principe de l'interpolation d'une grille régulière terrain à partir des points obtenus dans la géométrie de l'image maître.*

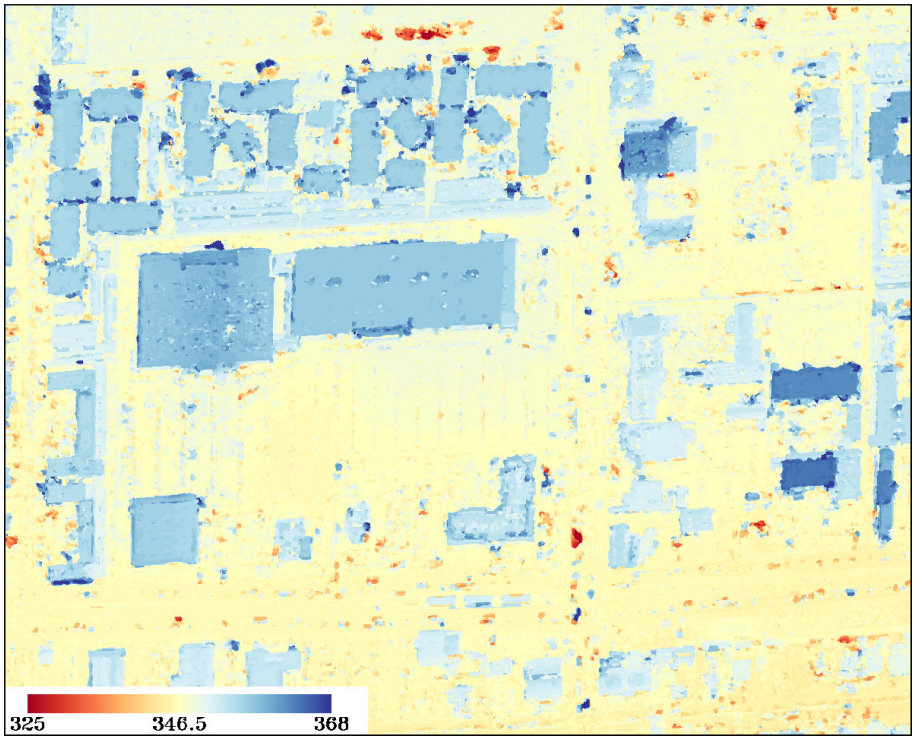
se situe dans la zone d'occlusion de l'image maître.

Ce principe permet la localisation rapide des zones d'occlusion qui pourront alors être prises en compte dans la suite des calculs.

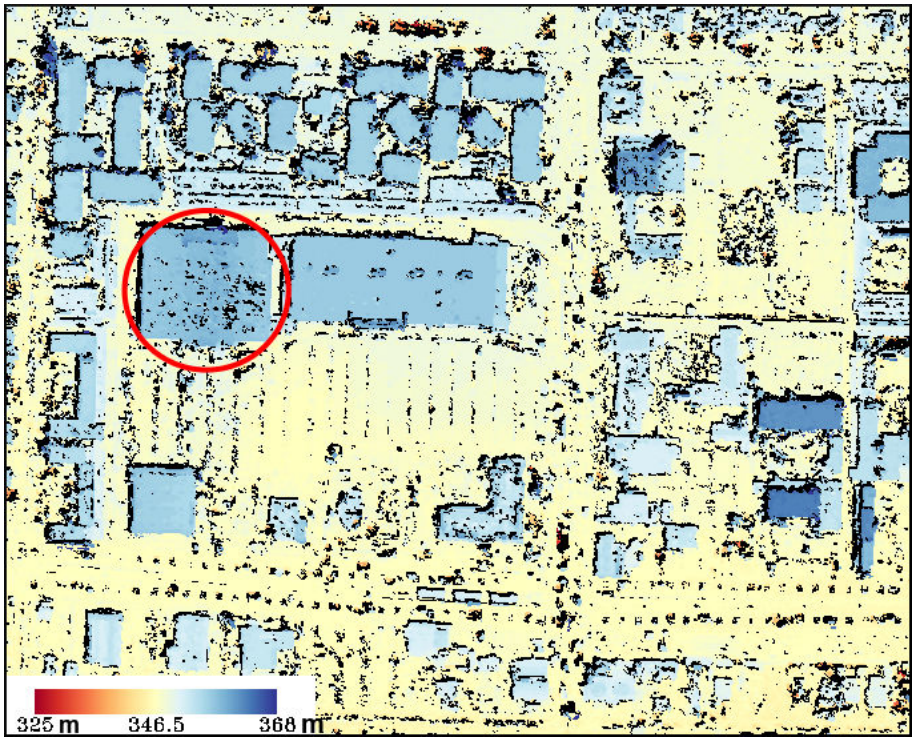
Enfin, une dernière contrainte sur le basculement consiste à ne pas prendre en compte, dans le calcul, les pixels de la carte des élévations dont le score de corrélation final, obtenu lors de la mise en correspondance, est inférieur à un certain seuil. L'objectif est ici de filtrer les points pour lesquels la corrélation a échoué, toujours afin de limiter les erreurs sur le MNS final.

Ce seuil sur le score de corrélation est défini ici à $2\sigma$, $\sigma$ étant l'écart–type obtenu à l'histogramme des scores de corrélation. Cette valeur est généralement proche de $0,45$.

La figure 3.4(b) illustre le résultat d'un basculement d'une carte des élévations présentée en figure 3.4(a) obtenue à partir d'un couple d'images WorldView-1 à 60 cm de résolution spatiale. Sur la carte basculée, les pixels noirs représentent les zones masquées lors du basculement, que




(a) Carte des élévations en géométrie image




(b) Carte des élévations en géométrie terrain

FIGURE 3.4 – *Illustration du résultat du basculement d'un MNS WorldView-1 à partir de la carte des élévations en géométrie image.*

ce soit à cause des zones d'occlusion de l'image maître ou d'un score de corrélation trop faible. Il faut noter que lors de toutes les étapes suivantes de notre chaîne de traitements, ces points ne sont jamais interpolés à partir de leurs voisins. En effet, leur interpolation serait susceptible de produire des erreurs dans le MNS, erreurs pouvant se répercuter dans la carte finale de détection de changements.

Cette technique de localisation des zones d'occlusion est basée uniquement sur une analyse géométrique, postérieure au calcul des élévations par mise en correspondance. Son principal inconvénient est de produire de nombreuses fausses zones d'occlusion. Ces erreurs proviennent le plus souvent d'erreurs de corrélation dans la carte des élévations originale. En effet, une erreur de corrélation génère, par définition, une élévation erronée et donc souvent une variation d'élévation entre deux pixels voisins, qui ne représente pas la réalité. Dans la méthode de détection des zones d'occlusion, cette variation est assimilée à la présence d'un bâtiment et la zone à proximité de la variation d'élévation est alors considérée comme une zone d'occlusion. Ainsi, de nombreux pixels du MNS basculés peuvent être masqués sans pour autant faire partie d'une zone d'occlusion.

Sur des zones homogènes de l'image, toit de bâtiment large et lisse par exemple tel que celui du bâtiment carré encerclé dans la figure 3.4(a), on observe de nombreuses petites zones noires masquées. Ces zones masquées proviennent très probablement de pixels dont le score de corrélation se situait en dessous du seuil fixé.

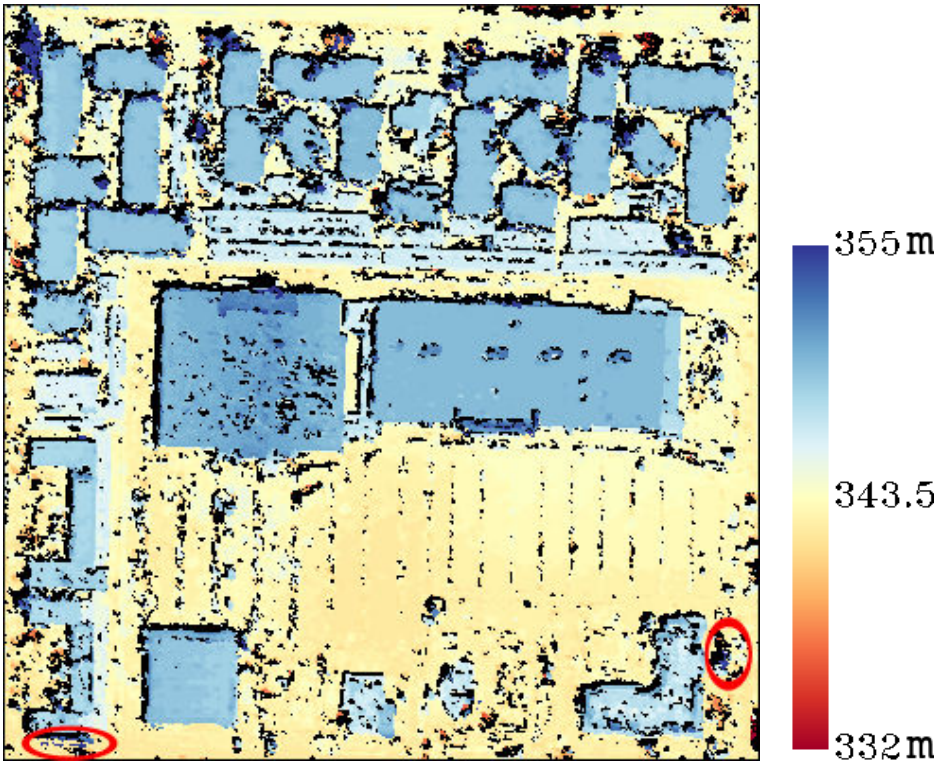
Dans les sections et chapitres suivants, ces zones d'occlusion et de mauvaise corrélation détectées et masquées lors du basculement seront regroupées et citées sous le terme de "masque du MNS".
Ce masque de MNS sera pris en compte lors de toutes les étapes suivantes de la chaîne de traitements.

## 3.3 Amélioration des MNS : développement d'une méthode de fusion des MNS

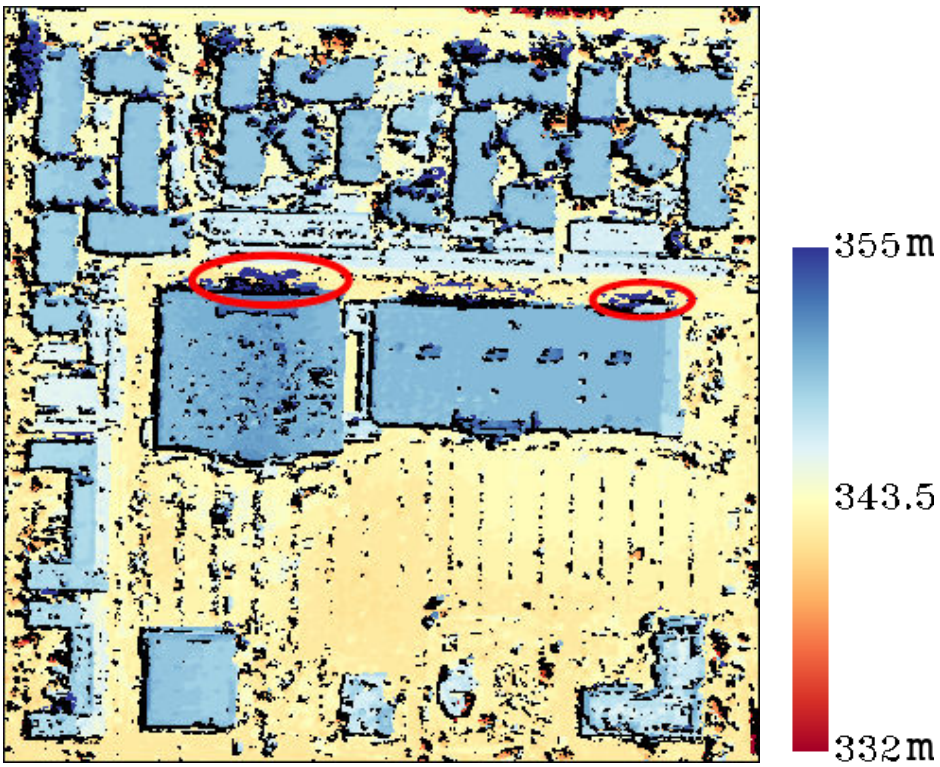
### 3.3.1 Erreurs des MNS asymétriques

Les figures 3.5(a) et 3.5(c) présentent les résultats de deux MNS calculés à partir du même couple stéréoscopique mais avec inversion des rôles maître et esclave de chaque image.
Sur ces images, les pixels noirs représentent les zones masquées car appartenant à des zones d'occlusion ou des zones de faible score de corrélation, calculées lors du basculement (section 3.2).
De façon générale, les deux MNS produits à partir du même couple stéréoscopique sont très similaires. On note, pour une zone de 2000 × 2000 pixels sur un couple d'image WorldView-1 acquis en 2008 sur la ville de Phoenix, une différence moyenne de 0,004 m avec un écart–type de 1,76 m. Cet écart-type résulte typiquement du bruit provenant de mauvaises corrélations ou de zones d'occlusion ayant généré des erreurs lors de la mise en correspondance des images, erreurs ensuite répercutées sur les MNS basculés.

On observe que, selon le MNS (et donc selon l'angle d'acquisition de l'image maître), les zones d'occlusion sont gérées et masquées plutôt au sud (figure 3.5(a)) ou au nord (figure 3.5(c)) des bâtiments.
De plus, comme détaillé précédemment, certaines erreurs de corrélation ont généré des variations d'élévation dans la carte des élévations qui ont elles–mêmes généré des occlusions virtuelles lors du basculement. Ces erreurs étant variables d'un MNS à l'autre, les zones d'occlusion générées sont aussi variables entre les MNS.
Enfin, des erreurs importantes de corrélation sont visibles dans les MNS. Ces erreurs sont variables en amplitude et en localisation selon le MNS, elles sont notamment visibles à proximité des bâtiments et encerclées en rouge sur la figure 3.5(a).
Afin d'obtenir un MNS cohérent et le plus précis possible, la solution proposée est de calculer les deux cartes d'élévation provenant du même couple avec inversion des rôles maître et esclave des images puis de fusionner ces MNS, préalablement basculés dans la même géométrie. L'objectif de cette fusion est de garder le maximum d'informations correctes provenant de l'un ou l'autre des MNS.




(a) $MNS_1$, couple stéréoscopique WorldView-1 du 24/06/2008.




(c) $MNS_2$, couple stéréoscopique WorldView-1 du 24/06/2008.

FIGURE 3.5 – *MNS asymétriques provenant du même couple d'images stéréoscopiques.*

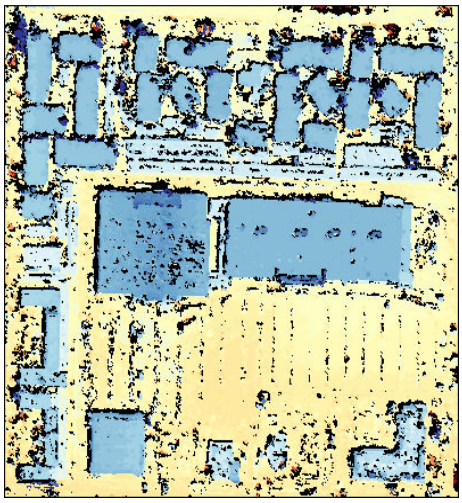
### 3.3.2 Méthode de fusion des MNS asymétriques

Plusieurs techniques peuvent être employées pour la fusion des deux MNS.
Une technique consiste à calculer, pour chaque pixel, la moyenne entre les deux valeurs obtenues. Si l'une des deux valeurs appartient à un des masques de MNS calculés lors du basculement, alors le pixel du MNS final est lui–même enregistré dans le masque final du MNS. Ce masque final représente alors l'union des deux masques générés pour chaque MNS.
Un résultat de cette fusion est présenté figure 3.6(c). Sur cette figure, les erreurs de corrélation encerclées précédemment n'ont pas étés corrigées, on observe même que le MNS fusionné possède toutes les erreurs provenant de chacun des MNS. De même, le masque final, union des deux masques des MNS est plus conséquent que les masques de chacun des MNS. Cette fusion par la moyenne entraîne donc la perte qualitative et quantitative d'information par rapport à chacun des MNS source.

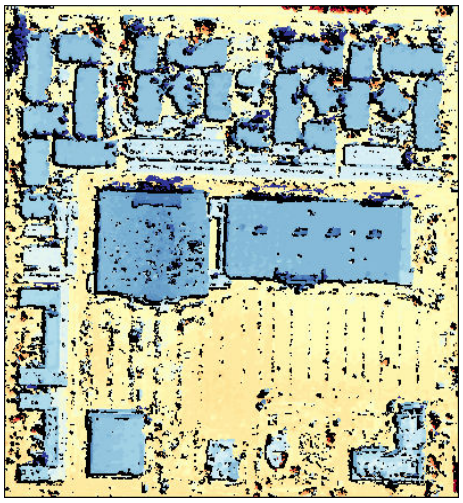
Afin de mieux prendre en compte les erreurs d'élévation de l'un ou l'autre des MNS tout en utilisant toute l'information spatiale correcte de chacun des MNS (c'est–à–dire en limitant les erreurs et les zones masquées), nous proposons de résoudre la fusion à l'aide d'une contrainte de régularisation spatiale.
La méthode proposée est basée sur un problème de labellisation permettant d'obtenir le MNS final : à chaque pixel $x_i$ du MNS final est attribué un label $m$ provenant du jeu de label $M \in \{\mathrm{MNS}_1, \mathrm{MNS}_2, \mathrm{Occlusion}\}$. La valeur du pixel attribuée au MNS final est alors celle qui correspond au label sélectionné. Les labels $\mathrm{MNS}_1$ et $\mathrm{MNS}_2$ correspondent, pour chaque pixel, aux valeurs d'élévation calculées lors de la génération de l'un ou l'autre MNS, tandis que le label Occlusion correspond à un label occlusion, attribué au pixel lorsque aucune valeur de MNS ne paraît cohérente.

Le choix d'un label, pour chacun des pixels du MNS final est effectué à partir des différentes données obtenues à la génération des MNS. Ces données sont :

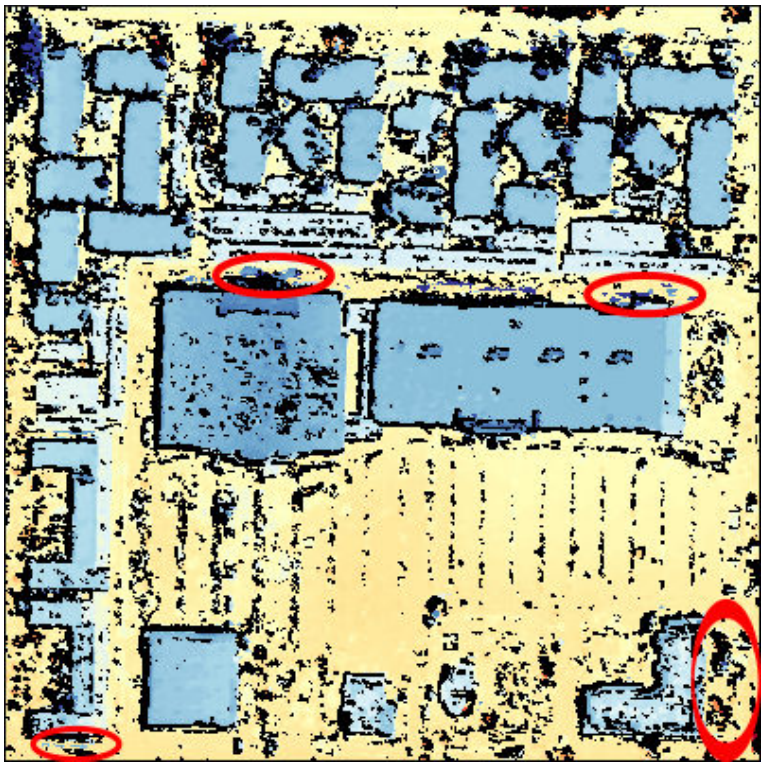
– le *Normalized Cross Correlation Coefficient*, NCC(présenté en section 3.1.2 et d'équation 3.3) obtenu lors de la mise en correspondance des images pour chacun des pixels (et ré–échantillonné dans la même géométrie que le MNS lors de l'étape de basculement, section



(a) $MNS_1$.



(b) $MNS_2$.



(c) Résultat de la fusion des MNS par une moyenne et l'intersection des zones d'occlusion.

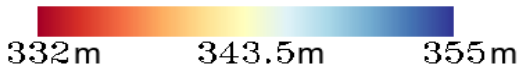



FIGURE 3.6 – *MNS provenant du même couple stéréoscopique et résultat de leur fusion par la moyenne.*

3.2),

- la différence d'élévation obtenue entre les deux MNS pour le même pixel,
- la régularisation spatiale qui dépend de la variation d'altitude entre deux pixels consécutifs dans le MNS final.

Le NCC est une donnée importante qui permet ici de donner un indice de confiance à l'altitude délivrée par chaque pixel de chacun des MNS : plus le NCC est important, plus la probabilité que la corrélation soit correcte en ce pixel est élevée. La comparaison des NCC provenant de chaque MNS permet alors d'indiquer quelle valeur de MNS paraît la plus pertinente.

De même, la différence obtenue entre les valeurs de MNS sur un même pixel permet de localiser des zones de mauvaise corrélation ou d'occlusion. En effet, l'élévation obtenue en un pixel doit être sensiblement la même pour les deux MNS, une différence d'élévation correspond alors généralement à une erreur sur l'un ou les deux MNS. C'est le principe utilisé par les techniques de *back matching* pour la détection des zones d'occlusion [49], [33].

Enfin, la présence d'une contrainte de régularisation spatiale permet de lisser le MNS final.

Comme détaillé au chapitre 2, section 2.2.1, ce problème de labellisation peut être résolu par une minimisation d'énergie dont l'algorithme est décrit en section 2.2.3 et l'équation générale est de type :

$$E(n) = \sum_{x \in X} C_{data}(n(x)) + \sum_{(x') \in V(x)} C_{reg}(n(x), n(x')). \tag{3.5}$$

Nous rappelons que chaque MNS possède déjà son propre masque, calculé lors de l'étape de basculement décrit en section 3.2. Les pixels appartenant à ce masque doivent alors être pris en compte lors du calcul du MNS final par fusion.

Ainsi, la valeur de tous les pixels de chaque MNS et correspondant à un masque est définie à $-9999$ lors de cette étape de fusion. Les valeurs correspondantes dans la carte des NCC sont, elles, définies à 0. Nous détaillerons tout au long de cette section, comment ces pixels sont pris en compte dans la fusion.

Dans cette étape de fusion, la fonction d'attache aux données, $C_{\text{Data}}(x_i, m(x_i))$, pour les labels $\text{MNS}_1$ et $\text{MNS}_2$ dépend du coefficient de corrélation (NCC) obtenu à la génération des

MNS : plus ce coefficient est important pour le label considéré $m$ au pixel $x_i$, plus la probabilité du label est élevée en ce pixel.

Concernant le label Occlusion, la fonction d'attache aux données dépend du NCC maximum obtenu entre les deux NCC correspondant à chacune des valeurs de MNS en présence et de la différence calculée entre les deux valeurs de MNS au pixel $x_i$.

Ainsi, lorsque l'un ou les deux MNS obtiennent un score de corrélation important, leur coût d'attache aux données est faible tandis que celui du label Occlusion est fort car il est calculé comme étant l'inverse du maximum de corrélation obtenu. La probabilité du label Occlusion est donc faible au pixel $x_i$. Le choix entre les labels $\mathrm{MNS}_1$ et $\mathrm{MNS}_2$ dépend alors de chacun de leur coefficient de corrélation et de la régularisation. Ce cas est représenté par un chemin de flèches vertes en larges pointillés sur la figure 3.7.

L'équation 3.6 résume le calcul des attaches aux données pour chacun des labels :

$$C_{\mathrm{Data}}(x_i, m(x_i)) = \begin{cases} 1 - \mathrm{NCC}(x_i, m(x_i)) & \text{si } m \in \{\mathrm{MNS}_1, \mathrm{MNS}_2\}, \\ max[\mathrm{NCC}(x_i, \mathrm{MNS}_1), \mathrm{NCC}(x_i, \mathrm{MNS}_2)] + F_1(x_i) & \text{si } m \in \{\mathrm{Occlusion}\}. \end{cases} \tag{3.6}$$

Dans la fonction d'attache aux données du label Occlusion, la fonction $F_1(x_i)$ correspond à la valeur absolue de la différence d'élévation relevée au pixel $x_i$ et normalisée entre 0 et 1 tel que :

$$F_1(x_i) = \frac{1}{1 + e^{-L(|Z(x_i, \mathrm{MNS}_1) - Z(x_i, \mathrm{MNS}_2)| - t_1)}}. \tag{3.7}$$

Le paramètre $L$ est défini à 1.

Cette fonction, non linéaire, permet d'affecter un coût suffisamment faible pour les différences supérieures à un seuil $t_1$ défini à 5 m, différences considérées ici comme importantes et donc pouvant provenir d'une occlusion. Des différences plus faibles peuvent en revanche provenir d'une faible erreur sur l'un des MNS qui ne doit pas pour autant privilégier le label Occlusion mais plutôt le label MNS le plus pertinent, en particulier si le NCC correspondant est fort. C'est pourquoi le NCC maximum est intégré dans la fonction d'attache aux données du label Occlusion : plus le maximum NCC est fort, plus la probabilité d'une occlusion est faible.

Dans le cas où l'une des valeurs de MNS au pixel $x_i$ correspond au masque natif du MNS,

par exemple pour le $\mathrm{MNS}_2$, on obtient : $\mathrm{NCC}(\mathrm{x_i}, \mathrm{MNS}_2) = 0$ et $\mathrm{MNS}_2(\mathrm{x_i}) = -9999$, donc :

$$\begin{aligned} &C_{\mathrm{Data}}(x_i, \mathrm{MNS}_1) = 1 - \mathrm{NCC}(\mathrm{x}, \mathrm{MNS}_1), \\ &C_{\mathrm{Data}}(x_i, \mathrm{MNS}_2) = 1, \\ &C_{\mathrm{Data}}(x_i, \mathrm{Occlusion}) = \mathrm{NCC}(\mathrm{x}, \mathrm{MNS}_1) + 0. \end{aligned} \tag{3.8}$$

Le pixel $x_i$ peut alors être labellisé $\mathrm{MNS}_1$ ou bien Occlusion selon la valeur de NCC obtenue en $x_i$ pour le $\mathrm{MNS}_1$. C'est le chemin représenté en flèches violettes et en petits pointillés dans la figure 3.7.

Dans le cas où les valeurs des deux MNS en $x_i$ appartiennent aux masques d'occlusion on obtient :

$$\begin{aligned} &C_{\mathrm{Data}}(x_i, \mathrm{MNS}_1) = 1, \\ &C_{\mathrm{Data}}(x_i, \mathrm{MNS}_2) = 1, \\ &C_{\mathrm{Data}}(x_i, \mathrm{Occlusion}) = 1. \end{aligned} \tag{3.9}$$

Dans ce cas particulier, la valeur du MNS final en $x_i$ est forcément égale à 0 (puisque tous les labels en présence correspondent à une valeur à 0) et le label sélectionné dépend alors seulement du terme de régularisation. C'est le chemin représenté en flèches oranges et lignes pleines dans la figure 3.7.

La fonction de régularisation est définie telle que :

$$C_{\mathrm{Reg}}((x_i, m(x_i)), (x_{i+1}, m(x_{i+1}))) = C^R, \tag{3.10}$$

avec $C^R$ le terme de régularisation qui dépend de la différence absolue entre les valeurs d'élévations $Z(x_i, m(x_i))$ et $Z(x_{i+1}, m(x_{i+1}))$ obtenues par les labels $m$ pour deux pixels consécutifs $x_i$ et $x_{i+1}$. Ce terme est calculé ainsi :

$$C^R = \begin{cases} F_2(\mid Z(x_i, m(x_i)) - Z(x_{i+1}, m(x_{i+1})) \mid) & \text{si } m(x_i) \text{ et } m(x_{i+1}) \in \{\mathrm{MNS}_1, \mathrm{MNS}_2\}, \\ \beta & \text{si } m(x_i) \in \{\mathrm{MNS}_1, \mathrm{MNS}_2\} \text{ et } m(x_{i+1}) \in \{\mathrm{Occlusion}\}, \\ \gamma & \text{si } m(x_i) \in \{\mathrm{Occlusion}\}. \end{cases} \tag{3.11}$$

avec $F_2$ définie par :

$$F_2(\mid Z(x_i, m(x_i)) - Z(x_{i+1}, m(x_{i+1})) \mid) = \frac{1}{1 + e^{-L(\mid Z(x_i, m(x_i)) - Z(x_{i+1}, m(xi+1)) \mid - t_2)}}. \tag{3.12}$$

Le paramètre $L$ est défini à 1. La valeur $t_2$ de la fonction $F_2$ est ici posée à 2.5 m (valeur seuil utilisée dans la méthode de détection de changements). L'objectif de cette fonction est de normaliser les différences d'élévations entre 0 et 1 afin que les valeurs de coût obtenues pour chaque label soient comparables entre elles.
$F_2$ est une fonction de régularisation entre deux labels de pixels consécutifs et permet ainsi de lisser le MNS final car la valeur du label choisi est alors celle qui se rapproche le plus de la valeur voisine.
L'utilisation d'une fonction sigmoïde permet de donner des coûts similaires aux labels amenant des variations d'élévation proches et ainsi éviter une fonction seuil qui attribuerait des coûts égaux à 0 ou 1 et orienterait alors trop fortement le choix final du label.

Lors du calcul de la fonction de régularisation :

- lorsque l'une des deux valeurs consécutives appartient à l'un des masques natifs des MNS, le coût calculé par la fonction $F_2$ est alors proche de 1, coût élevé permettant de limiter la probabilité de ce label,
- lorsque les deux valeurs de pixels correspondent aux masques natifs des MNS, $F_2$ est alors égal à 0. Les coûts de labellisation sont alors similaires en ce pixel pour les deux MNS.

Les variables $\beta$ et $\gamma$ correspondent à des valeurs fixées à une valeur moyenne de coût ($\beta = 0.5$ et $\gamma = 0.5$). Ces valeurs, très lâches, permettent d'attribuer des coûts neutres aux labels considérés, le coût final du label dépend alors principalement de l'attache aux données.

La figure 3.7 présente un schéma récapitulatif qui illustre l'orientation du choix de chacun des labels en fonction des données d'entrée (NCC et valeur de MNS) et des fonctions d'attache aux données utilisées. Sur la figure 3.7, les carrés bleus clair indiquent les labels gagnants pour lesquels le coût attribué est le plus faible en fonction du calcul de l'attache aux données mais aussi en fonction de la régularisation. Les labels entourés en rouge représentent les labels impossibles pour chacun des cas décrits.

En pratique, il est à noter que les fonctions d'attache aux données, de même que la régularisation, ne privilégient que rarement le label Occlusion. Cette caractéristique n'est cependant pas un inconvénient puisque l'objectif est de diminuer les occlusions, et notamment

les fausses occlusions provenant d'erreurs de corrélation, et les pixels de faible NCC afin de privilégier les valeurs correctes de chacun des MNS.

De même, on note que des coûts très similaires peuvent être affectés aux deux labels MNS. Or, comme précisé précédemment, en de nombreux pixels des MNS, les élévations calculées sont très proches car il n'y avait pas d'ambiguïtés dans la mise en correspondance. Dans ce cas, l'affectation de l'une ou l'autre valeur des MNS n'a pas d'impact réel sur le MNS final.

La figure 3.8(c) présente le résultat obtenu par la fusion par labellisation des deux MNS produits à partir du même couple stéréoscopique. Bien que toujours globalement similaire aux MNS, ce résultat ne présente plus les erreurs importantes de corrélation encerclées précédemment.

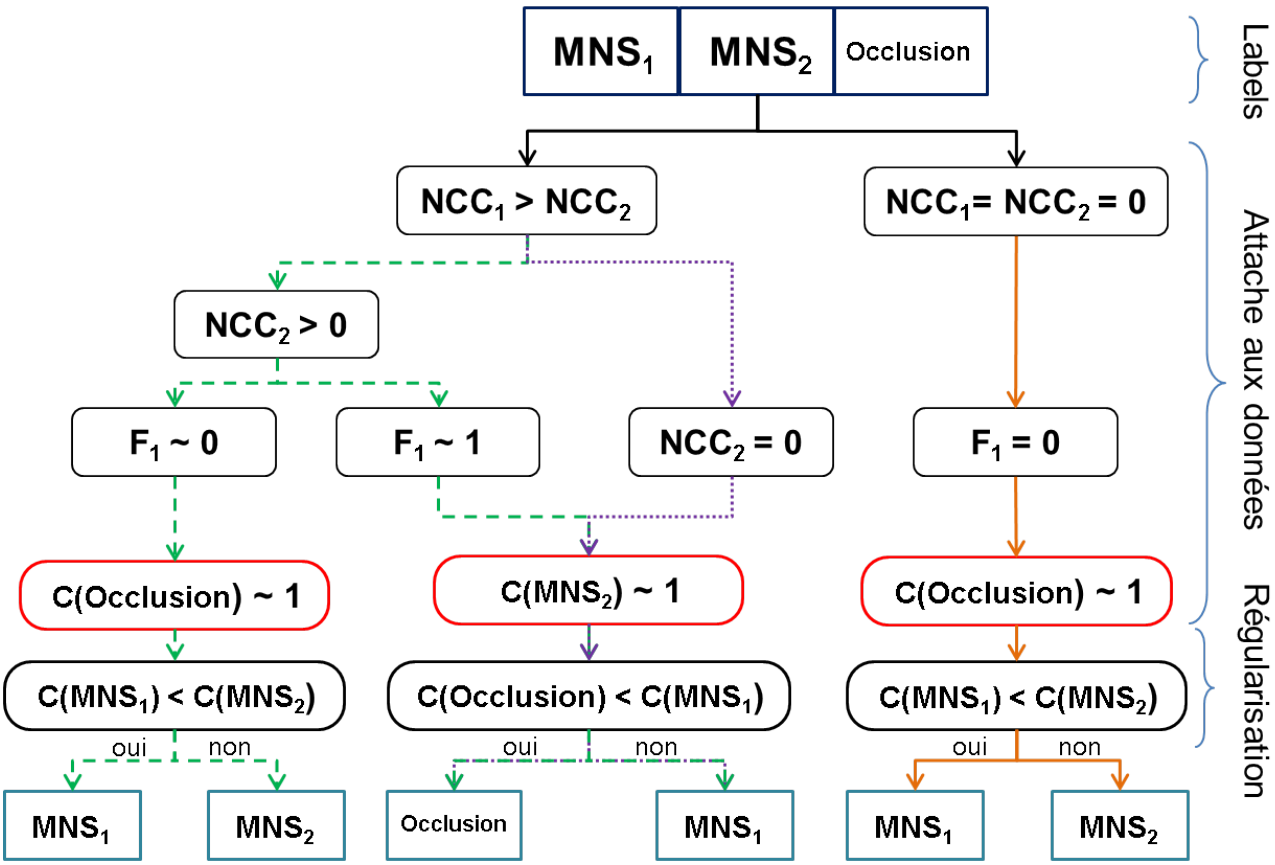
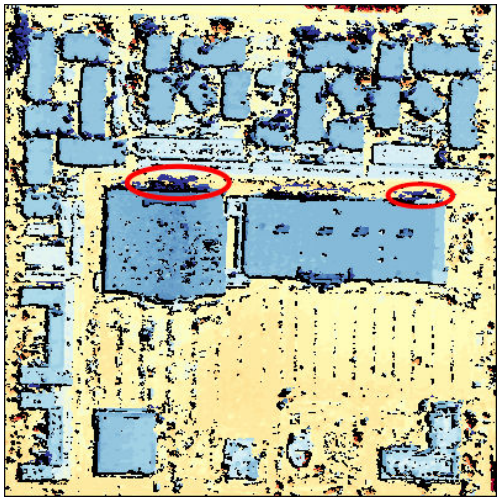
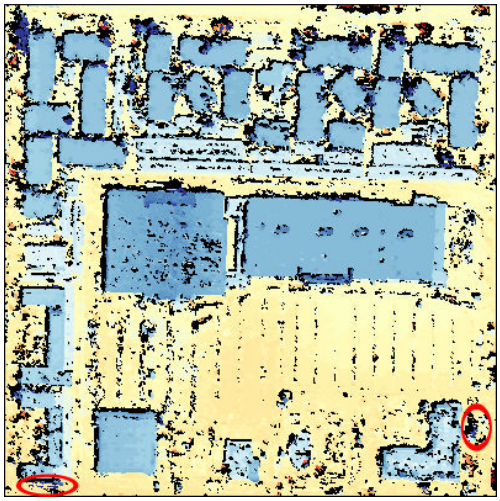


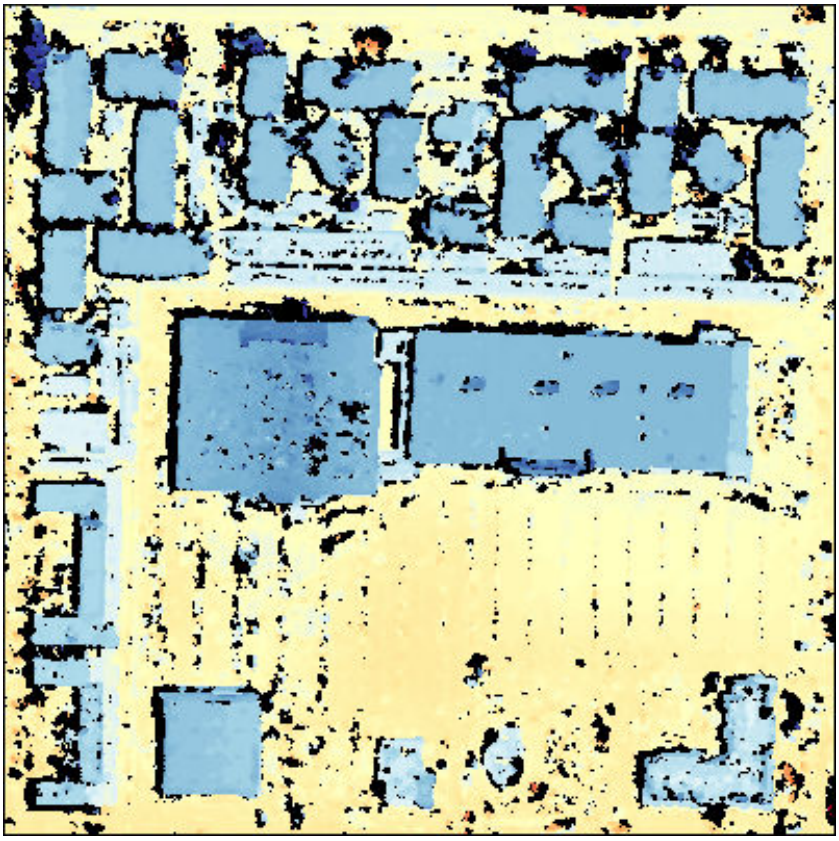

FIGURE 3.7 – *Schéma récapitulatif de la procédure de fusion des MNS en fonction des valeurs initiales et de l'attache aux données.*



(a) $MNS_1$



(b) $MNS_2$



(c) Résultat de la fusion par optimisation avec régularisation spatiale.

FIGURE 3.8 – *3.8(a) et 3.8(b) : MNS calculés à partir du même couple stéréoscopique et résultats de leur fusion par optimisation globale 3.8(c).*

Elles ont pu être supprimées grâce à cette méthode de fusion. De plus, les zones masquées sont réduites par rapport à la fusion par la moyenne présentée en figure 3.6(c), permettant d'obtenir un résultat contenant une plus grande quantité d'informations car c'est la somme des informations pertinentes des $MNS_1$ et $MNS_2$.

Enfin, il est à noter que, bien que ces zones masquées puissent être interpolées à partir des valeurs qui les entourent, nous avons choisi de ne pas le faire afin de ne pas générer d'erreurs d'interpolation qui pourraient engendrer des erreurs dans la carte finale des changements d'élévation. Ces zones sont alors considérées comme des zones de "non–information" et prises en compte lors de la détection des changements d'élévation.

## 3.4 Qualité et précision des MNS obtenus

Afin d'observer l'apport de la méthode de génération des MNS par fusion par rapport à la génération de MNS asymétriques par mise en correspondance simple, le MNS généré sur la ville de Christchurch (Nouvelle-Zélande) est comparé avec une vérité terrain obtenue avec un LiDAR. Ce LiDAR [1] a été acquis en 2011, année d'acquisition des images WorldView-2 sur la même zone (voir tableau 1.1, chapitre 1, section 1.4).
La comparaison a été effectuée sur une zone commune entre le LiDAR et les MNS d'environ $2300 \times 2600$ pixels.

### 3.4.1 Co–localisation LiDAR et MNS

Le LiDAR dont nous disposons a été ré–échantillonné à 1 m de résolution spatiale.
Comme discuté dans le chapitre 2, section 2.1, les images ayant permis la génération des MNS ont été affinées simultanément permettant leur co–localisation précise entre elles et ainsi, la co–localisation précise entre les MNS. Cependant, cette co–localisation est effectuée en relatif et la localisation absolue des modèles n'est pas garantie.
Ainsi, les MNS et le LiDAR ne peuvent être nativement co–localisés (on observe des biais

1. Pour des raisons de confidentialité, le LiDAR et les informations associées ne peuvent être présentés dans ce manuscrit. Nous nous excusons pour cette restriction auprès des lecteurs.

jusqu'à 12 m en longitude et 25 m en latitude entre le MNS et le LiDAR).
Afin de permettre la comparaison précise entre le LiDAR et les MNS, il est alors nécessaire de recaler ces données entre elles. Cette opération est effectuée de façon classique, en prenant des points de liaison entre les données (ces points sont sélectionnés au milieu des toits afin d'obtenir une meilleure précision). L'erreur résiduelle sur ces points est d'environ 0,7 pixels RMS (erreur moyenne quadratique). Puis, un modèle de déformation affine du MNS est calculé en fonction de ces points de liaison afin d'obtenir un MNS recalé (au pixel près) sur le LiDAR.

### 3.4.2 Critères de qualité

La comparaison entre le LiDAR et les MNS a été effectuée à plusieurs niveaux : une différence globale a été réalisée entre les données afin d'observer le bruit de ces dernières. Puis, une comparaison a été effectuée entre les zones masquées des MNS afin d'observer les erreurs résiduelles sur ces zones et leurs différences entre le MNS asymétrique et le MNS fusionné. Enfin, la comparaison de profils sélectionnés sur les différentes données va permettre d'observer comment les structures telles que les bords de bâtiments, les toits mais aussi les zones homogènes sont gérées dans les MNS.

#### Différence globale

Dans un premier temps, l'image des différences entre le LiDAR et le MNS asymétrique (MNS obtenu à partir de la géométrie image puis basculé dans un espace terrain)a été calculée. La moyenne des différences obtenue est d'environ $-0,4$ m avec un écart–type d'environ $3,1$ m. Cet écart–type est particulièrement important et illustre le bruit présent dans le MNS.

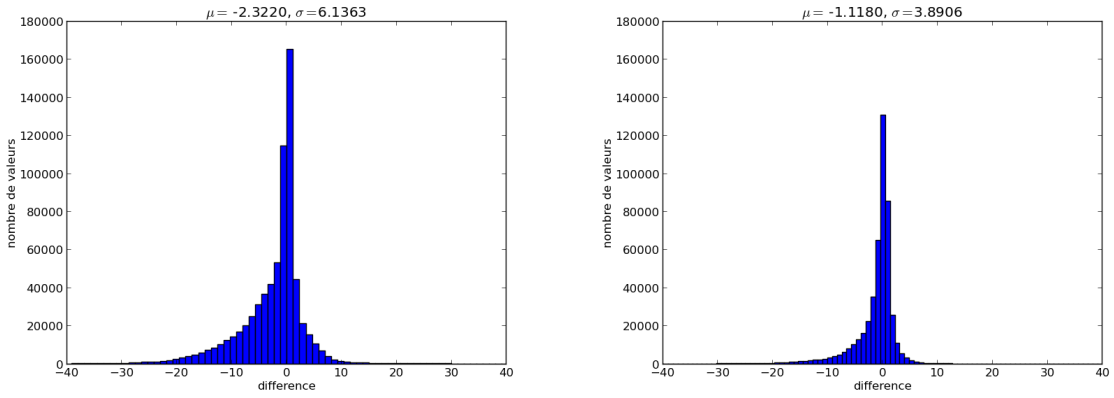
La différence calculée entre le MNS fusionné (par la technique décrite en section 3.3) montre, elle, une moyenne d'environ $-0,25$ m avec un écart–type d'environ $2,4$ m. Ces valeurs, largement plus faibles que les résultats obtenus avec le MNS asymétrique attestent que la proportion d'erreur a diminué sur le MNS fusionné.

**Différences entre les masques des MNS asymétrique et fusionné**

Les zones masquées sur le MNS asymétrique et pas sur le MNS fusionné, ou inversement, représentent les zones où la probabilité d'erreur est la plus forte.
Dans la section 3.3, nous expliquons que la fusion de deux MNS générés à partir du même couple stéréoscopique, mais avec inversion des rôles maître et esclave, permet de prendre en compte les zones d'occlusion situées sur les deux images du couple stéréoscopique.
Cependant, nous avons pu noter, dans la section 3.2, que les erreurs de corrélation générées lors de la mise en correspondance peuvent engendrer des fausses zones d'occlusion. Ainsi, le masque du MNS fusionné peut contenir des pixels faussement enregistrés comme occlusion.

La comparaison de la taille de ces masques donne une première idée sur la qualité des MNS et la précision des masques : en effet, on observe que 19,3% des pixels appartenant au MNS asymétrique sont enregistrés dans le masque, tandis que cette quantité baisse à 15,5% pour le MNS fusionné. Cette différence indique qu'une plus petite quantité de pixels a été enregistrée dans le masque du MNS fusionné.
Ce phénomène est paradoxal puisque le MNS fusionné permet d'enregistrer les occlusions appartenant aux deux images du couple stéréoscopique au contraire du MNS asymétrique. On peut donc en déduire que le masque du MNS fusionné est plus précis sur les zones d'occlusion et limite le nombre de pixels masqués.
De plus, les masques des MNS prennent aussi en compte les scores de corrélation les plus faibles obtenus lors de la mise en correspondance. Or, nous avons vu que la fusion des MNS permet de réduire le nombre de pixels masqués à cause de ce seuil. En effet, lorsqu'un pixel est masqué dans l'un des MNS mais possède un score de corrélation suffisamment élevé dans l'autre MNS tout en répondant à la contrainte de régularisation, cette valeur de pixel est alors privilégiée dans le MNS final plutôt que la valeur masquée. Ces contraintes de calcul sur le MNS final permettent ainsi de réduire les zones masquées.

Afin de comparer la qualité des masques des MNS asymétrique et fusionné, tous les pixels enregistrés dans un masque pour l'un des MNS et ayant une valeur définie dans l'autre MNS, sont comparés avec la vérité LiDAR.




(a) Histogramme des différences entre le LiDAR et les pixels masqués dans le MNS fusionné mais définis dans le MNS asymétrique

(b) Histogramme des différences entre le LiDAR et les pixels masqués dans le MNS asymétrique mais définis dans le MNS fusionné

FIGURE 3.9 – *3.9(a) et 3.9(b) : Histogrammes des cartes de différences calculées entre le liDAR et le MNS asymétrique et MNS fusionné, respectivement.*

L'histogramme présenté en figure 3.9(a) intègre les pixels masqués dans le MNS fusionné et ayant une valeur d'élévation définie dans le MNS asymétrique.
Ces pixels représentent 11,9% de l'image. On observe une moyenne des différences entre ces valeurs dans le MNS asymétrique et le LiDAR de $-2,32$ m avec un écart–type de $6,14$ m. L'histogramme confirme cette tendance de fortes différences entre ces valeurs et le LiDAR. Nous pouvons ainsi en déduire que ces valeurs définies dans le MNS asymétrique mais pas dans le MNS fusionné représentent principalement des erreurs d'élévation du MNS asymétrique.

L'histogramme présenté en figure 3.9(b) correspond aux pixels masqués dans le MNS asymétrique mais ayant une valeur définie dans le MNS fusionné.
On recense 8% des pixels de l'image appartenant à cette catégorie. La différence calculée entre ces pixels appartenant au MNS fusionné et la vérité LiDAR est en moyenne de $-1,1$ m avec un écart–type de $3,9$ m. Ces chiffres montrent un résultat bien meilleur sur ces pixels bien que des erreurs subsistent.
Ce résultat indique qu'un grand nombre de ces pixels, recensés dans le masque du MNS asymétrique, possédaient en réalité une altitude définie et correcte, car proche du LiDAR, dans l'autre MNS asymétrique. Ces valeurs sont donc maintenant des valeurs correctes dans le MNS

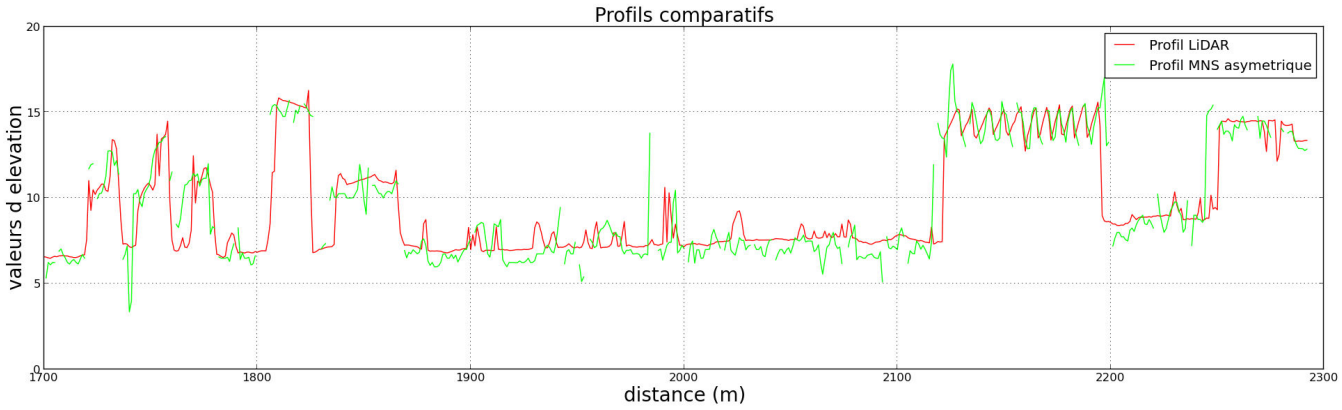
fusionné. Cependant, certaines erreurs résiduelles pourraient être corrigées, par exemple, grâce à l'utilisation de plus de deux images stéréoscopiques afin d'améliorer la précision de ce masque.

**Comparaison de profils d'élévation**

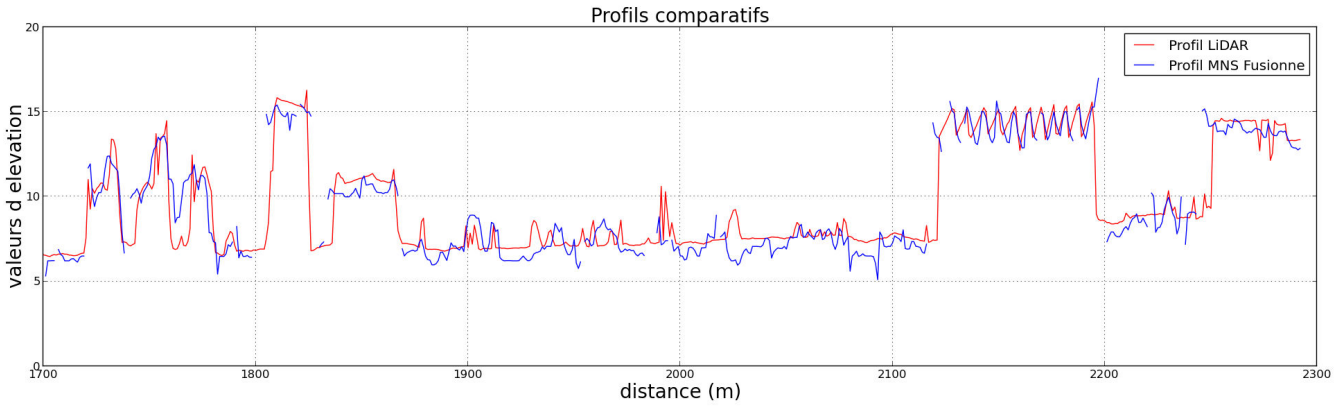
Les figures 3.10(a) et 3.10(b) permettent de comparer des profils d'élévation obtenus entre le LiDAR, le MNS asymétrique et le MNS fusionné.
Ce profil, qui s'étend sur une distance de 1000 m (la résolution spatiale des pixels étant de 1 m), montre de nombreux bâtiments, de hauteurs variables entre 4 et 9 m environ.
Globalement, les profils des MNS asymétrique et fusionné suivent la tendance du profil LiDAR, on observe cependant un biais entre le LiDAR et les MNS, ces derniers présentent en effet une altitude plus faible de plusieurs dizaine de centimètres d'après la zone basse entre 1900 m et 2100 m. Ce biais est aussi visible sur les bâtiments situés entre 1800 m et 1900 m. Ce biais reste néanmoins dans la résolution altimétrique donnée pour les MNS qui est de 1 m.

On observe, malgré le bruit présent sur les MNS, que le toit gondolé du bâtiment situé entre 2100 m et 2200 m est particulièrement bien rendu dans les deux MNS. De même, les bâtiments, pourtant proches les uns des autres entre 1700 m et 1800 m apparaissent très bien démarqués.

On note que le MNS asymétrique semble générer un bruit plus important, avec de plus fortes variations d'élévations, que celui généré par le MNS fusionné. Les bords de bâtiment (correspondant à des zones d'occlusion) sont aussi mieux gérés par le MNS fusionné. En effet, on observe, notamment pour les bâtiments entre 2100 m et 2300 m, que le MNS asymétrique a tendance à élargir les bâtiments et à générer des élévations très fortes aux abords des bâtiments. Ce phénomène est corrigé dans le MNS fusionné dans lequel tous les bords de bâtiments sont supprimés car considérés comme des zones d'occlusion.
Ce comportement montre l'intérêt de gérer les zones d'occlusion dans les MNS puisque des bords de bâtiments élargis sont de fortes sources de fausses alarmes dans la carte finale des changements.




(a) Comparaison des profils LiDAR et MNS asymétrique.




(b) Comparaison des profils LiDAR et MNS fusionné.

FIGURE 3.10 – *Profils d'élévation comparés entre le MNS asymétrique (en vert), le MNS fusionné (en bleu) et le Lidar (en rouge).*

## 3.5 Application au cas de la multi–stéréoscopie

L'algorithme que nous avons développé précédemment permet le calcul d'un MNS précis à partir d'un couple stéréoscopique. Or, les nouveaux satellites tels que Pléiades permettent aujourd'hui l'acquisition d'images en $n$–uplet stéréoscopiques, avec $n \in [2 : 25]$ pour Pléiades. Chaque image est alors acquise avec un angle d'incidence et d'azimut différent. Dans ce cas, il est nécessaire de prendre en compte tout ou partie des images produites pour le calcul d'un MNS.

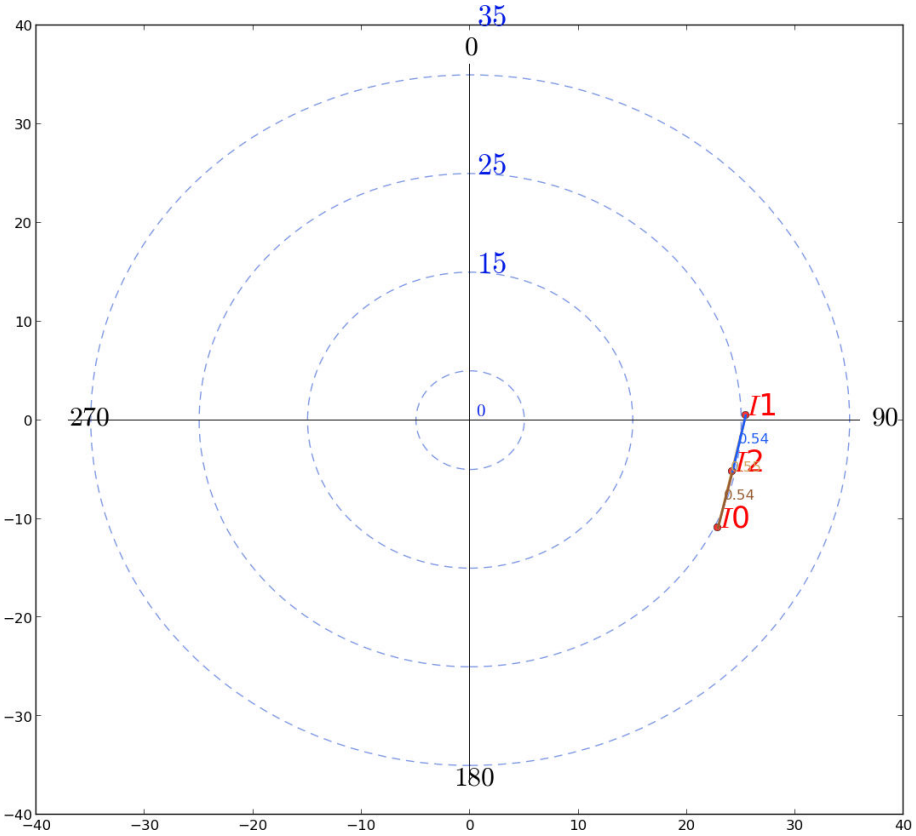
L'algorithme de fusion des MNS est alors une solution intéressante pour gérer les $n$-uplets stéréoscopiques.

Dans ce contexte, la solution proposée est la création de tous les MNS possibles provenant de la mise en correspondance de toutes les images 2 à 2 avec inversion des rôles maître et esclave et la fusion de tous les MNS obtenus afin de calculer un MNS final. Pour un $n$–uplet stéréoscopique, le nombre de MNS à générer est alors de $N_{MNS} = n * (n - 1)$. Le nombre de labels présents dans la fusion est ensuite égal au nombre total des MNS générés, additionné du label Occlusion.

Dans le calcul de la fusion, effectué selon l'algorithme décrit précédemment, la seule différence notable est le calcul de l'attache aux données pour le label Occlusion (équation 3.6). Dans le cas de la stéréoscopie simple, il est nécessaire de calculer la différence entre les deux valeurs de MNS obtenues au même pixel $x_i$. Dans le cas de la multi–stéréoscopie, le nombre de MNS en jeu étant supérieur à 2, toutes les différences entre les MNS générés sont calculées et l'attache aux données est alors fonction de la différence minimale obtenue (en dehors des différences égales à 0 qui proviennent de la différence entre deux masques de MNS).
Ce choix d'utiliser la différence minimum est orienté par le fait qu'en un pixel, les valeurs de MNS doivent être similaires. Ainsi, plus des valeurs seront proches, plus la probabilité que les valeurs de MNS soient pertinentes est élevée.

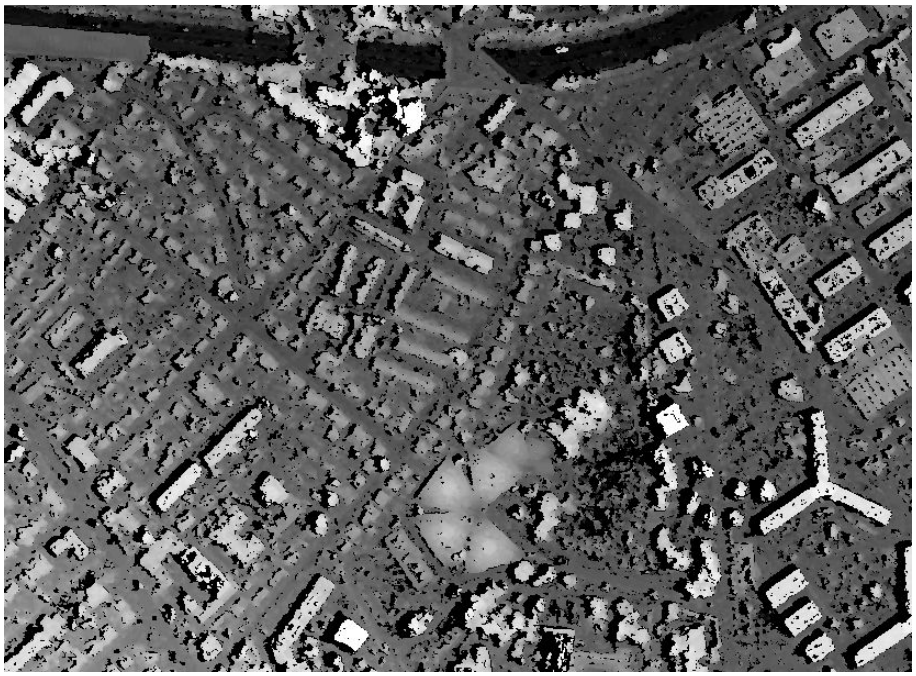
La figure 3.12 présente le résultat du calcul d'un MNS à partir de deux images d'un triplet stéréoscopique 3.12(a) et des trois images du triplet 3.12(b). Ce triplet d'images Pléiades a été acquis sur la ville de Toulouse en mai 2012 (figure 3.11(a)).
La figure 3.11(b) permet de montrer les conditions d'acquisition de ce triplet. Dans cette figure, la position d'un point sur le périmètre du cercle représente l'angle d'azimut de l'acquisition de l'image considérée tandis que la position sur le rayon du cercle représente l'angle d'incidence de l'acquisition.
Selon cette figure, on observe que le triplet stéréoscopique a été acquis avec un angle d'incidence assez élevé (environ 25°) et selon un seul passage du satellite. Ce système d'acquisition n'apparaît pas idéal pour limiter les zones d'occlusion puisque les angles azimutaux des trois images sont particulièrement proches entre eux.
En effet, le MNS calculé à partir de deux images contient de nombreuses zones masquées à cause des nombreuses zones d'occlusion, particulièrement importantes du fait des angles d'acquisition des images, mais aussi des erreurs de corrélation (sur les toits de bâtiments, par exemple). Tandis

(a) Partie d'une image panchromatique du triplet stéréoscopique Pléiades acquis sur la ville de Toulouse le 07/05/2012.

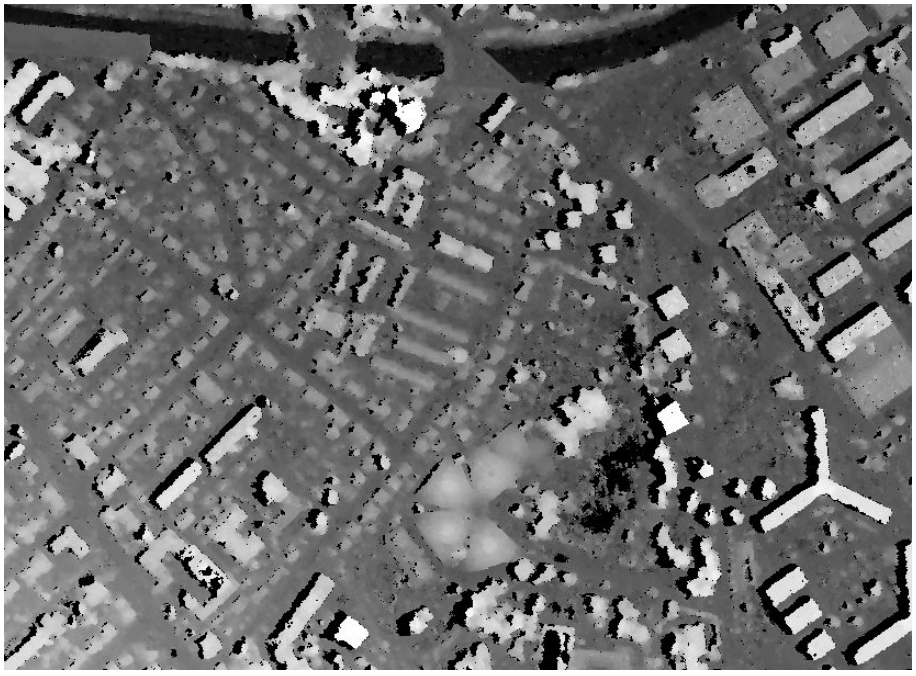



(b) Illustration de la géométrie d'acquisition du triplet stéréoscopique.

FIGURE 3.11 – *Illustration du triplet stéréoscopique Pléiades acquis sur la ville de Toulouse.*



(a) Toulouse, MNS fusionné à partir d'un couple stéréoscopique.



(b) Toulouse, MNS fusionné à partir d'un triplet stéréoscopique.

FIGURE 3.12 – *MNS calculés à partir de 2 images puis 3 images du même uplet stéréoscopique.*

que le MNS tri-stéréoscopiques apparaît beaucoup plus lisse et seules les vraies zones d'occlusion (zones vues par aucune des images), situées, comme attendues, au nord–ouest des bâtiments, sont masquées.

De façon quantitative, on observe que sur le MNS formé à partir de deux images stéréoscopiques, 19,2% des valeurs du MNS appartiennent au masque tandis que sur le MNS obtenu avec le triplet stéréoscopique complet, seule 8,3% des valeurs appartiennent au masque.

Ce résultat montre l'apport de la multi–stéréoscopie pour la génération des MNS, même lorsque les conditions d'acquisition stéréoscopiques ne sont pas idéales, c'est–à–dire que les images ne sont pas acquises selon des angles d'azimut très différents. Cette comparaison permet de démontrer non seulement la capacité de la méthode pour la multi–stéréoscopie mais aussi le fort intérêt de la multi-stéréoscopie pour la précision des MNS.

## 3.6 Conclusion

La méthode de génération développée dans notre chaîne de traitements est basée sur trois étapes de calcul. La première est la mise en correspondance des images, selon une géométrie image et donc basée sur le calcul de l'élévation de tous les points d'une image maître, calcul effectué par l'outil MicMac. La seconde étape est le basculement des élévations, de la géométrie image à une géométrie terrain définie. Cette opération, développée au cours de cette thèse, a permis de mettre en place une meilleure gestion des points terrain situés dans des zones d'occlusion de l'image maître, et de filtrer les points dont le score de corrélation est particulièrement bas.

Cette technique de mise en correspondance étant asymétrique, les erreurs dues notamment aux zones d'occlusion de l'image esclave ne sont pas prises en compte. La troisième étape de calcul a donc pour objectif de générer un MNS final dans lequel les zones d'occlusion appartenant à l'une ou l'autre des images du couple sont prises en compte tout en limitant les erreurs ponctuelles de corrélation présentes sur l'un des MNS seulement. Pour cela, la mise en correspondance est réalisée deux fois, avec inversion des images maître et esclave et les MNS obtenus sont fusionnés selon une technique formulée comme un problème de labellisation dans lequel les labels de chaque pixel correspondent aux valeurs de chacun des MNS en présence ou à un label occlusion. Ce problème est alors résolu à partir de l'algorithme de programmation dynamique implémenté, dont les fonctions de coûts ont été adaptées afin de privilégier les meilleurs coefficients de corrélation

obtenus et les valeurs permettant de lisser les variations du MNS final.

Tous les paramètres et seuils de la méthode complète de génération des MNS sont résumés dans le tableau 3.1.

La comparaison des MNS résultats avec un LiDAR acquis sur la même zone a permis de présenter l'apport de cette fusion de MNS par rapport à un MNS asymétrique, notamment envers les zones d'occlusion ou les zones de mauvaise corrélation.

Enfin, nous avons montré que cette méthode et l'algorithme développé étaient parfaitement adaptables au cas de la multi–stéréoscopie quels que soient les angles d'acquisition des images.

| **Etape** | **Paramètre** | **impact sur le résultat** | **Valeur recommandée** |
|---|---|---|---|
| MicMac | Régularisation des MNS | Impact important sur les variations du MNS | Dépend de la zone traitée : en milieu urbain $\lambda = 0.02$ |
| | Nombre de direction de calcul | Impact modéré dès 12 directions | 12 directions |
| Basculement | Résolution de la grille terrain du MNS | Impact sur la résolution du résultat | 2 fois la résolution native des images |
| | Seuil sur le score de corrélation | Impact sur la qualité du MNS résultat | $2\sigma$ |
| Fusion | Seuil $t_1$ | Impact sur la quantité de pixels définis en label occlusion | Dépend des scores de corrélation obtenus $t_1$ 5 m |
| | Seuil $t_2$ | Impact faible sur la régularisation du MNS | Défini à 2,5 m |
| | $\beta$ et $\gamma$ | Impact faible | Définies arbitrairement à 0,5 |
| | Nombre de directions de calcul | Impact très modéré sur le résultat | Généralement posé à 12 |
| | Régularisation | Impact modéré sur la régularisation du MNS | Défini expérimentalement à 5,0 |

TABLE 3.1 – *Tableau de synthèse des paramètres de la méthode complète de génération des MNS.*

# Chapitre 4

# Développement d'une méthode de détection des changements d'élévation

La dernière étape de la chaîne de traitements, développée au cours de cette thèse, consiste à détecter les changements d'élévation réels et pertinents de la scène. Dans un contexte urbain, ces changements correspondent à toutes les constructions, destructions ou modifications de bâtiments ou d'infrastructures apparues entre les deux dates d'intérêt.
La méthode employée pour la détection de changement est basée sur l'analyse de la différence des MNS obtenus pour les dates $t_1$ et $t_2$. Cependant, à la différence de Tian [26] qui filtre le MNS différentiel à partir d'opérations morphologiques et de connaissances *a priori* sur les tailles et formes des objets recherchés, notre analyse est basée sur une classification par une méthode d'optimisation semi–globale des pixels de l'image.

Dans cette section, nous présenterons tout d'abord l'intérêt du filtrage du MNS différentiel pour la mise en évidence des changements d'élévation puis nous détaillerons la méthode que nous avons développée afin de mettre en évidence les changements les plus pertinents.

## 4.1 Calcul et analyse du MNS différentiel

Comme décrit dans le schéma global de la chaîne de traitements (chapitre 1, section 1.3, figure 1.2), une fois les MNS générés pour les dates $t_1$ et $t_2$, leur différence $d_{MNS}$ est calculée telle que :

$$d_{MNS} = MNS_{t_2} - MNS_{t_1}. \tag{4.1}$$

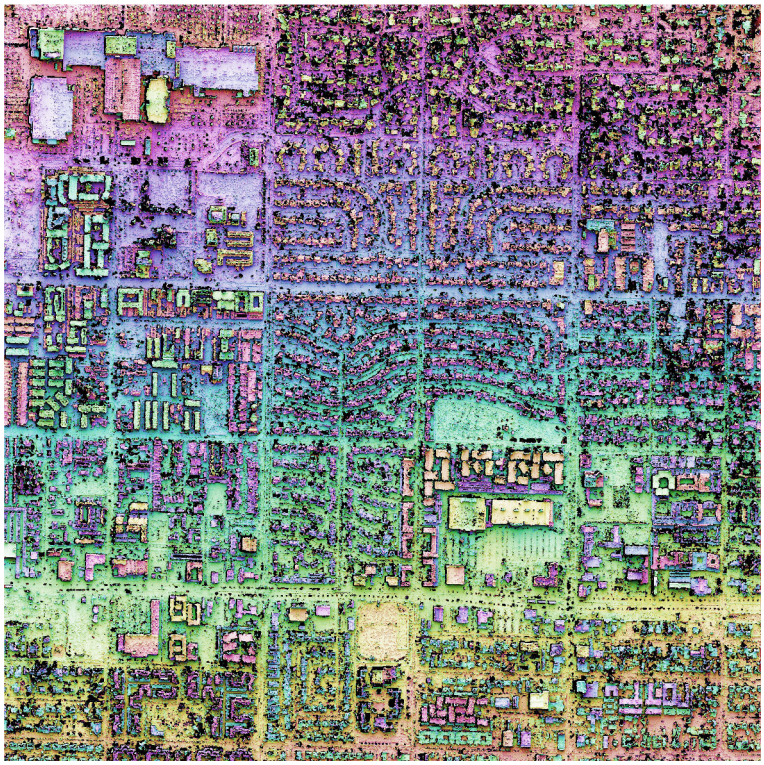
Ainsi, les constructions apparues entre $t_1$ et $t_2$ apparaissent avec des pixels de valeurs positives sur le MNS différentiel tandis que les destructions correspondent à des valeurs négatives.
L'union des zones masquées de chacun des MNS, sont, quant à elles définies à zéro dans le MNS différentiel.

Cependant, certaines erreurs résiduelles de corrélation restent attachées à chacun des MNS, comme nous l'avons illustré dans le chapitre 3, section 3.4 lors de la comparaison avec la vérité LiDAR et ce, malgré l'amélioration de la précision des MNS.
Ces erreurs de corrélation proviennent le plus souvent de la mise en correspondance sur des zones difficiles telles que :

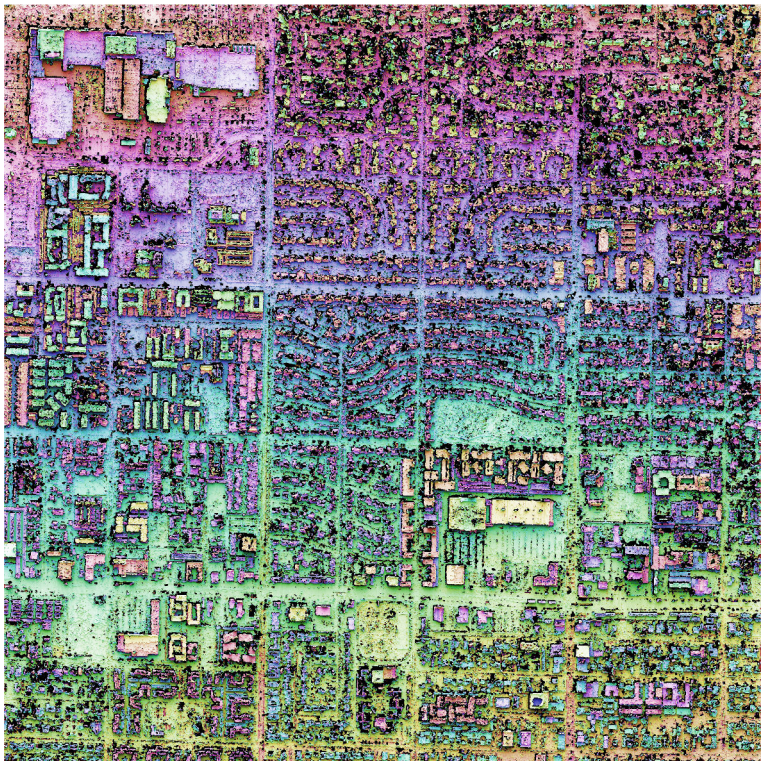
- les structures 3D complexes comme celles de la végétation,
- les zones homogènes ou très périodiques,
- les bords de bâtiments,
- les cibles mobiles,
- les zones d'occlusion.

Or, toute erreur dans un des MNS peut générer une différence d'élévation potentiellement importante dans le MNS différentiel et ainsi provoquer une fausse détection dans la carte finale des changements. La seule segmentation du MNS différentiel ne permet donc pas de mettre en évidence les changements pertinents d'élévation dans le bâti car le nombre d'alarmes de détection provenant des erreurs du MNS est alors trop grand.

Les figures 4.1(a) et 4.1(b) représentent les MNS obtenus sur la ville de Phoenix aux dates $t_1 = 2008$ et $t_2 = 2011$. La figure 4.2(a) montre le MNS différentiel calculé à partir de ces deux MNS. Enfin, la figure 4.2(b) présente un exemple de seuillage du MNS différentiel. Sur cet exemple toutes les différences d'élévation inférieures à -2,5 m sont représentées en rouge et celle supérieures à 2,5 m, en bleu.

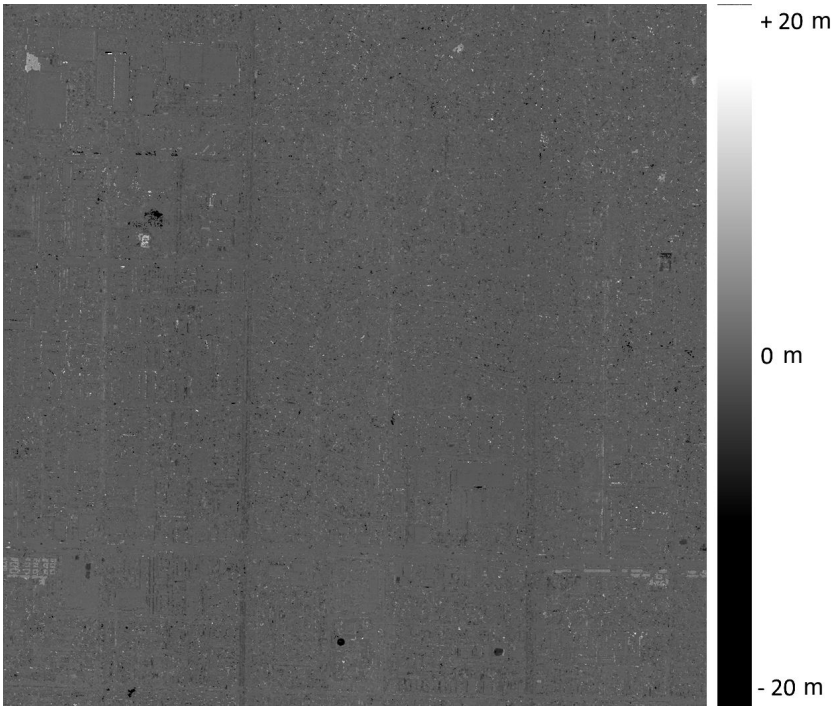


(a) MNS ombré calculé à partir des images WorldView-1 de 2008.

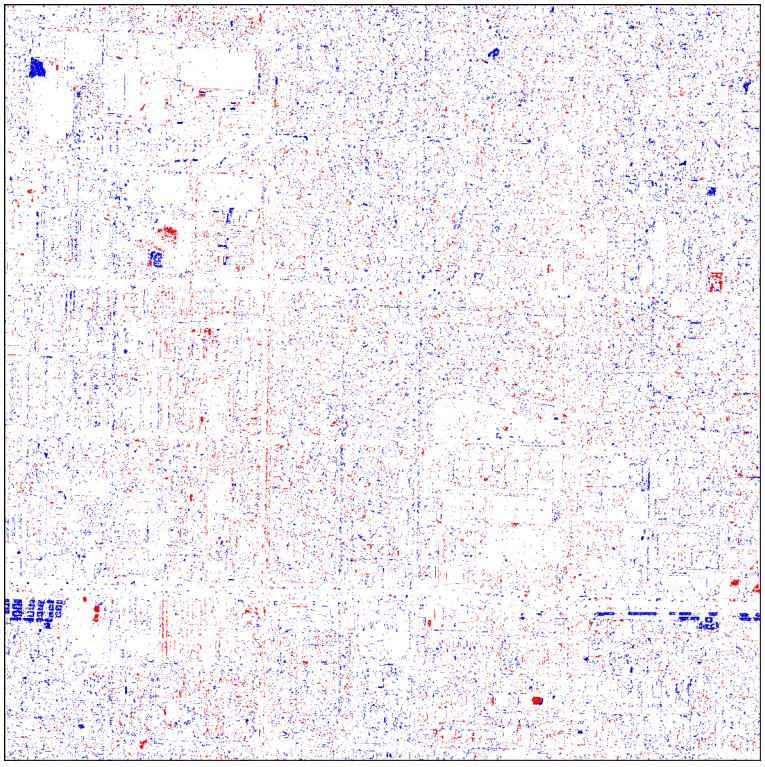


(b) MNS ombré calculé à partir des images WorldView-1 de 2011.

FIGURE 4.1 – *4.1(a) et 4.1(b) MNS de 2008 et 2011 calculés sur la zone 1 définie sur la ville de Phoenix.*




(a) MNS différentiel calculé à partir des MNS 2008 et 2011.



(b) MNS différentiel segmenté à 2,5 m (changements positifs en bleus) et à -2,5 m (changements négatifs en rouge).

FIGURE 4.2 – *4.2(a) MNS différentiel. 4.2(b) MNS différentiel seuillé.*

Notons que ce seuil à 2,5 m représente le seuil que nous utiliserons pour toutes les détections de changements effectuées à partir des MNS calculés à 1 m de résolution altimétrique et planimétrique. Ce seuil correspond à la hauteur standard d'un étage de bâtiment et paraît tout à fait adapté pour la résolution des MNS (1 m de résolution altimétrique et planimétrique).

Sur cette figure, de très nombreuses alarmes de changements sont visibles. Ces alarmes de changements, souvent représentées par quelques pixels, sont réparties sur toute l'image.

Les changements réels et pertinents du bâti sont pourtant aisément repérables (pour les plus importants) sur la carte seuillée. En effet, ces changements sont généralement représentés par des amas de pixels, homogènes spatialement et bien délimités contrairement au bruit.
C'est pourquoi, nous avons choisi de mettre en évidence ces changements pertinents à partir d'une technique de classification résolue à travers une optimisation globale du MNS différentiel, associée à une contrainte de régularisation spatiale. Cette technique permet en effet de prendre en compte la cohérence locale du MNS différentiel afin d'éliminer le bruit tout en conservant les changements cohérents du bâti.

## 4.2 Méthode de détection des changements

La méthode choisie pour la détection des changements d'élévation consiste en une classification des pixels du MNS différentiel. À chaque pixel est attribué un label $r$ provenant du jeu de labels $R$ et qui correspond à un changement positif (construction), à un changement négatif (destruction) ou à un non–changement, en fonction de sa valeur mais aussi de celle des ses voisins afin de satisfaire la contrainte de régularisation spatiale ($R \in \{\text{Changementpositif}, \text{Changementngatif}, \text{Nonchangement}\}$).
Ce problème s'apparente alors aux problèmes d'optimisations globales, introduits dans le chapitre 2.2, et résolus à partir de l'équation générale 2.2.3.

La fonction d'attache aux données utilisée dans cette optimisation, $C_{\text{Data}}(x_i, r(x_i))$, correspond au coût pour attribuer un label $r$ au pixel $x_i$ du MNS différentiel, $i$ étant la valeur de

différence d'élévation au pixel $x_i$. Ce coût, pour chaque label, dépend de la valeur de différence d'élévation obtenue dans le MNS différentiel et normalisée par une fonction sigmoïde (représentée en figure 4.3).

Il est défini par :

$$C_{\text{Data}}(x_i, r(x_i)) = \begin{cases} \frac{1}{1+\exp^{-L(|i|-T)}} & \text{si } r(x_i) = \{\text{non changement}\}, \\ 1 - \frac{1}{1+\exp^{-L(i-T)}} & \text{si } r(x_i) = \{\text{changement positif}\}, \\ 1 - \frac{1}{1+\exp^{-L(-i-T)}} & \text{si } r(x_i) = \{\text{changement négatif}\}. \end{cases} \tag{4.2}$$

Dans l'équation 4.2, le paramètre $T$ est fixé à 2,5 m, seuil de la détection de changements défini précédemment.

La normalisation des valeurs du MNS différentiel dans la fonction d'attache aux données permet de rendre comparable entre eux les coûts obtenus par un pixel et pour chaque label. Si de nombreuses techniques permettent la normalisation des valeurs, nous avons choisi une fonction sigmoïde qui permet un seuillage "flexible" des coûts de chacun des labels selon la valeur du pixel.

En effet, une fonction concave (ou "marche") calculerait un coût à 0 ou à 1 sans aucun intermédiaire pour chaque label, une fonction linéaire, au contraire, ne serait pas suffisamment déterminante pour des valeurs de pixels trop différentes du seuil fixé. Tandis que la fonction sigmoïde permet un seuillage plus doux.

Le paramètre $L$ est lié à la courbure de la fonction sigmoïde.

La figure 4.3 présente différentes sigmoïdes calculées avec des valeurs de $L$ allant de $L = 1,0$ à $L = 4,0$. D'après cette figure, on observe que la valeur de $L$ sélectionnée permet d'obtenir des coûts différents de 0 et 1 pour des valeurs de différence d'élévation situées entre 0,5 m et 4,5 m tandis qu'une valeur de $L$ plus forte telle que $L = 4,0$ restreint ce seuillage à des valeurs de différences entre 1,0 m et 4,0 m. Au contraire, une valeur de $L$ plus faible n'est pas suffisamment restrictive car on observe alors des valeurs de coût encore importantes (et donc une probabilité plus faible de changements positifs) pour une différence d'élévation de 4,0 m qui est pourtant une différence d'élévation non négligeable.

Le paramètre $L$ a ainsi été choisi à $L = 3,0$, cette valeur paraît plus adaptée aux différences d'élévation recherchées.

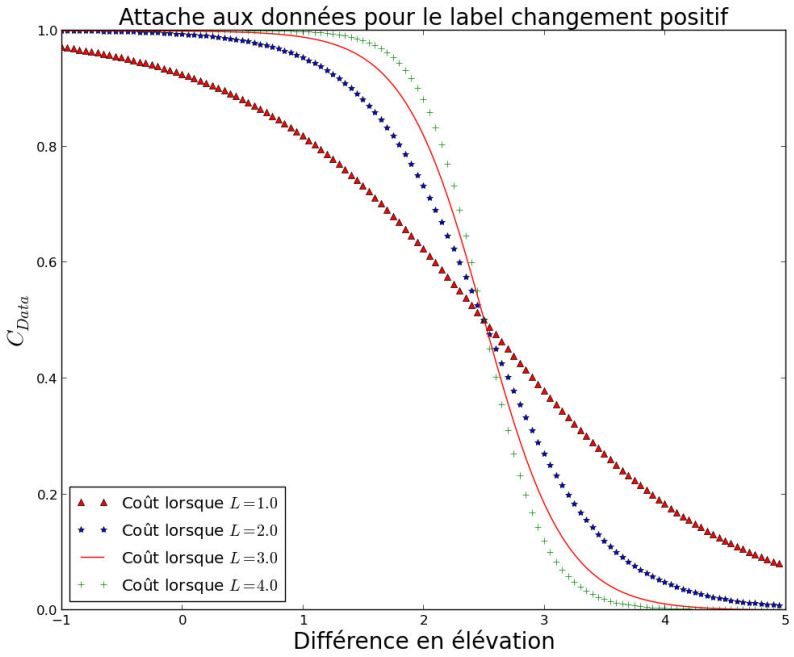



FIGURE 4.3 – *Fonction sigmoïde permettant le calcul des coûts d'attache aux données pour le label changement positif. Les quatre courbes illustrent les résultats pour différentes valeurs de L utilisées.*

Il faut noter cependant que l'impact de $L$ sur le résultat final est plutôt faible.

Nous rappelons que tous les pixels appartenant à l'un ou l'autre des masques des MNS générés à $t_1$ et $t_2$ sont définis à zéro dans le MNS différentiel. Leur probabilité de changement est donc fixée à zéro tandis que leur probabilité de non–changement est très élevée. Le label de ces pixels n'est cependant pas fixé de façon définitive, ainsi, selon la régularisation employée et le voisinage de ces pixels, leur labellisation en changement est autorisée si les conditions sont favorables.

Le terme de régularisation, $C_{\text{Reg}}((x_i, r(x_i)), (x_{i+1}, r(x_{i+1})))$ correspond à un coût de transition seulement et calculé selon le modèle de Potts [68]. Ce terme dépend ainsi de la variation de labels entre deux pixels consécutifs $x_i$ et $x_{i+1}$ de la séquence. Il est défini par :

$$C^T((x_i, r(x_i)), (x_{i+1}, r(x_{i+1}))) = \begin{cases} 0 & \text{si } r(x_i) = r(x_{i+1}) \\ 1 & \text{si } r(x_i) \neq r(x_{i+1}) \end{cases} \tag{4.3}$$

Enfin, le paramètre $\lambda$ représente, dans ce contexte, une variable particulièrement importante puisqu'elle définit le poids accordé à la fonction de régularisation par rapport à celui de la

fonction d'attache aux données. Plus cette valeur est élevée, plus la contrainte de régularisation sera importante et aura tendance à supprimer les variations fines de labels et donc les alarmes de changements les plus petites spatialement et inversement. Le paramètre $\lambda$ permet alors le contrôle de l'équilibre entre la précision de la détection et la sensibilité aux changements réels de la scène.

## 4.3 Conclusion

La méthode d'analyse du MNS différentiel que nous avons développée est basée sur la classification des pixels selon des labels changements positifs, changements négatifs ou non–changement en fonction de la valeur du pixel et de son voisinage, pris en compte à travers une contrainte de régularisation spatiale.
Les paramètres de la méthode sont décrits dans le tableau 4.1.
Cette méthode de filtrage global permet ainsi le contrôle de la précision et de la sensibilité du résultat grâce au paramètre de régularisation $\lambda$, contrôle d'autant plus important qu'il va permettre à l'utilisateur de moduler le résultat en fonction des changements recherchés.

Les résultats obtenus par cette méthode de détection des changements d'élévation et selon différents paramètres seront analysés dans la section 5

| **Paramètre** | **impact** | **Valeur** |
|---|---|---|
| Seuil $T$ | Impact important sur la hauteur des changements détectés | Dépend de la résolution alti. du MNS pour 1 m : seuil = 2,5 m |
| Régularisation | Impact important sur la précision et sensibilité du résultat | Dépend des résultats recherchés généralement entre 2 et 7 |
| Directions | Impact modéré sur le résultat | Généralement posée à 12 |
| Courbure L | Impact négligeable | Dépend du seuil en élévation des changements recherchés |

TABLE 4.1 – *Tableau de synthèse des paramètres nécessaires à la méthode de détection des changements d'élévation.*

# Chapitre 5

# Expérimentation et discussion des résultats de la méthode de détection des changements d'élévation

Afin d'analyser les performances de la méthode de détection des changements, trois sites sont analysés à travers la précision de la méthode (quantité de fausses alarmes détectées par rapport au nombre total de détections) et sa sensibilité, aussi appelée rappel, (quantité de bonnes détections retrouvées par rapport au nombre total de changements de la scène).
Les deux premiers sites sont représentés par plusieurs zones choisies sur les villes de Phoenix (Arizona, USA) et de Christchurch (Nouvelle–Zélande), villes dynamiques à développement continu.
Le troisième site se trouve dans la région de Tohoku (Japon), et plus particulièrement sur la ville de Sendai. Cette zone, qui a été dévastée par une catastrophe majeure, permet en effet d'analyser les résultats de la méthode dans un contexte de crise.

## 5.1 Présentation des zones de test

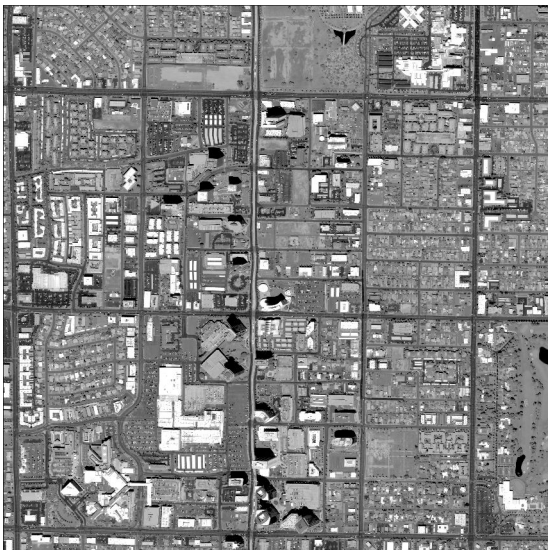
### 5.1.1 Phoenix

Sur la ville de Phoenix, deux couples stéréoscopiques ont été acquis en 2008 et 2011 par le capteur WorldView-1 (cf. tableau 1.1, chapitre 1, section 1.4).
Trois zones d'études ont été définies à partir de ces images. Ces zones, d'une taille de $2000 \times 2000$ pixels ont été sélectionnées sur l'image en fonction du nombre de changements qui ont été retrouvés mais aussi en fonction de la variété de bâtiments et infrastructures qu'elles contiennent. Les figures 5.1(a), 5.1(b) et 5.1(c) représentent ces différentes zones.

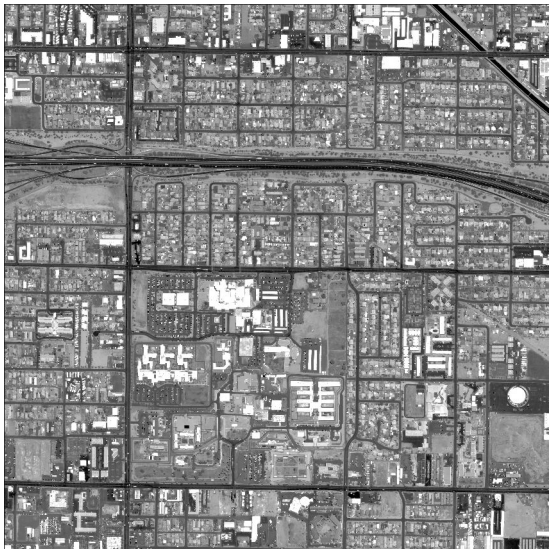
La zone 1 (figure 5.1(a)) comprend de petits pavillons résidentiels, de nombreux parcs ainsi que des quartiers d'affaires. La zone 2 (figure 5.1(b)) contient plutôt des quartiers d'affaires avec quelques très hauts bâtiments. Enfin, la zone 3 (figure 5.1(c)) présente plutôt des zones industrielles.
Sur ces zones, on note aussi la présence d'autoroutes avec de nombreux véhicules mobiles, de la végétation, ou encore des pavillons très proches les uns des autres. Tous ces éléments représentent des sources d'erreurs potentiellement importantes pour les MNS, qui se répercutent dans la détection de changements. Ces différentes difficultés permettent ainsi de tester au mieux les performances de notre chaîne de traitements.

Aucune vérité terrain de type cadastrale n'étant disponible sur la ville de Phoenix, une carte de référence a été générée manuellement pour chacune des trois zones d'étude. Plus précisément, cette carte a été créée par comparaison visuelle entre les images acquises aux dates $t_1$ et $t_2$. Tous les changements du bâti visibles ont été repérés et localisés dans cette carte de référence, qu'ils soient de quelques mètres carrés (taille d'un conteneur) ou de plusieurs centaines de mètres carrés (bâtiment industriel).
Cependant, afin de mieux analyser les résultats, les changements ont été divisés en deux catégories : les changements supérieurs ou égaux à $15 \times 15$ pixels, soit 100 $m^2$ pour des images à 60 cm de résolution, et tous les changements de l'image. Ce seuil représente la taille généralement détectable par les méthodes de détection de changements sur des images très haute résolution décrites par ailleurs [26]. Or, l'analyse des performances de la méthode lorsque toutes les tailles

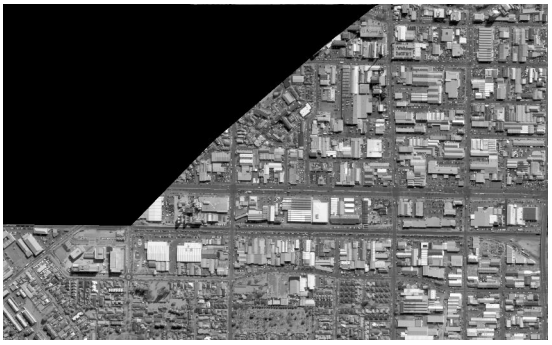
(a) Illustration de la zone #1 de Phoenix.



(b) Illustration de la zone #2 de Phoenix.

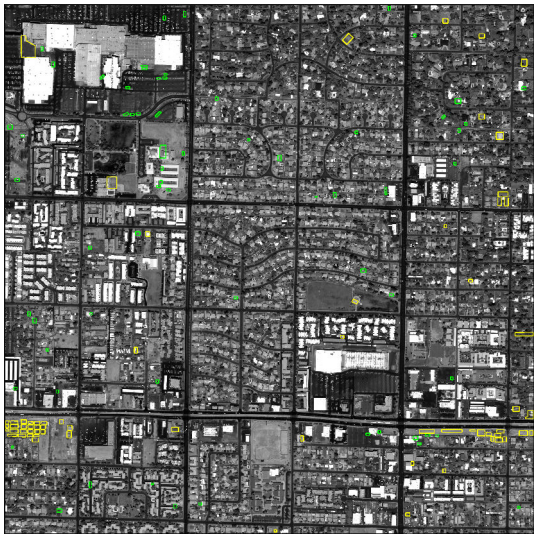


(c) Illustration de la zone #3 de Phoenix.

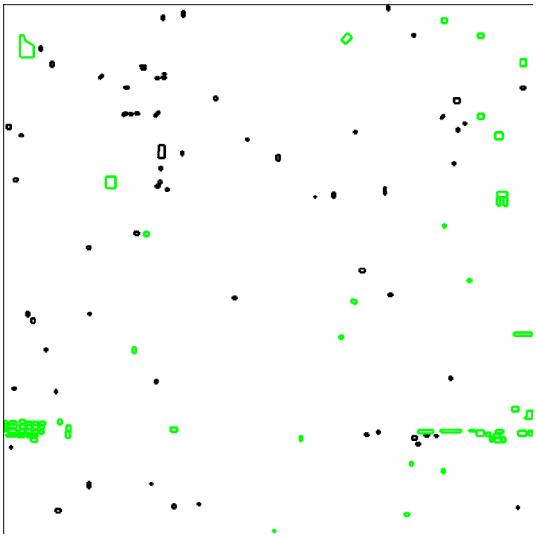


(d) Illustration de la zone de Christchurch.

FIGURE 5.1 – *Orthoimages provenant des images panchromatiques WorldView–1 représentant les trois zones testées sur la ville de Phoenix et la zone de Christchurch.*



(a) Orthoimage WorldView-1, 2008, Phoenix zone 1.



(b) Donnée de référence. En vert les changements de taille > 100 m$^2$, en noir les changements inférieurs à 100 m$^2$

FIGURE 5.2 – *5.2(a)Orthoimage calculée et 5.2(b) donnée de référence des changements de la zone 1 de Phoenix.*

de changements sont recherchées permet de mieux appréhender les limites de la méthode.
Au cours de ces analyses de sensibilité, nous parlerons plutôt de ce seuil à 100 m$^2$ mais nous précisons que ce seuil dépend avant tout de la résolution des images native. Pour des images haute résolution à 2 m, ce seuil serait calculé à 15×15 pixels soit 300m$^2$.

Finalement, sur la zone 1 de Phoenix, 126 changements de toutes tailles ont été recensés dont 55 changements supérieurs à 100 m$^2$ (figure 5.2(b)). Sur la zone 2 de Phoenix, 71 changements ont été recensés dont 22 larges changements. Enfin, sur la zone 3, 138 changements ont été retrouvés dont 34 supérieurs à 100 m$^2$.

### 5.1.2 Christchurch

Sur la ville de Christchurch, en Nouvelle-Zélande, nous disposons de deux acquisitions stéréoscopiques de 2009 et 2011 (cf. tableau 1.1, chapitre 1, section 1.4). A ces données s'ajoutent deux acquisitions LiDAR de 2010 et 2011.
La zone commune aux données optiques et LiDAR a été sélectionnée pour tester la méthode de détection de changement. Cette zone, d'une superficie de 1200 km$^2$ environ, est représentée en figure 5.1(d).

Le LiDAR a été préalablement ré–échantillonné à une résolution planimétrique identique à celle des MNS (1 m). Cependant, la précision d'acquisition du LiDAR est largement meilleure que celle des MNS générés. Les petites structures, les bâtiments et tous les objets de la scène sont donc plus nets [1].

Une détection des changements d'élévation entre les deux acquisitions LiDAR a donc été réalisée avec une régularisation faible ($\lambda = 2,0$) et un seuil classique à 2,5 m. L'objectif est de faire ressortir tous les changements de la scène.

Ainsi, les changements détectés entre les deux acquisitions LiDAR sont considérés comme suffisamment précis et exhaustifs pour représenter une vérité terrain qui permettra l'analyse des résultats obtenus sur cette zone par la détection de changements appliquée entre les MNS générés.

Cependant, les LiDAR ayant été acquis à des saisons différentes, une faible régularisation a aussi fait ressortir tous les changements liés à la végétation. Ces changements ont dû être manuellement supprimés de la vérité terrain générée.

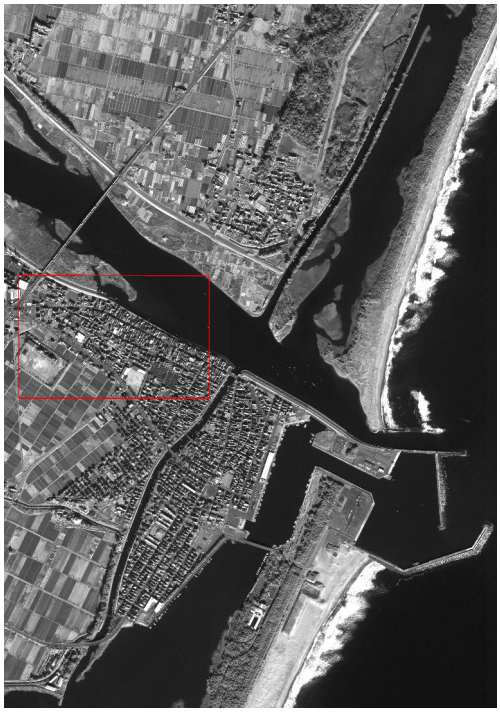
De plus, une année sépare la première acquisition satellite stéréoscopique de la première acquisition LiDAR (cf tableau 1.1, section 1.4). Afin de prendre en compte cet écart et les changements ayant eu lieu entre ces acquisitions, ces derniers ont été détectés en appliquant la méthode de détection entre l'acquisition satellite de 2009 et LiDAR de 2010, à faible régularisation spatiale. Puis, ces changements ont été manuellement filtrés des fausses alarmes telles que celles provenant de la végétation. La carte des changements ainsi obtenue a ensuite été additionnée à la vérité terrain.

Tout comme pour les trois zones de Phoenix, les changements repérés ont été divisés selon les deux catégories : ceux de taille supérieure à 100 $m^2$ et tous les changements. Finalement, 43 changements de toutes tailles ont étés repérés sur la zone, dont 35 de taille supérieure à 100 $m^2$.

### 5.1.3 Région de Tohoku : analyse d'une zone catastrophée

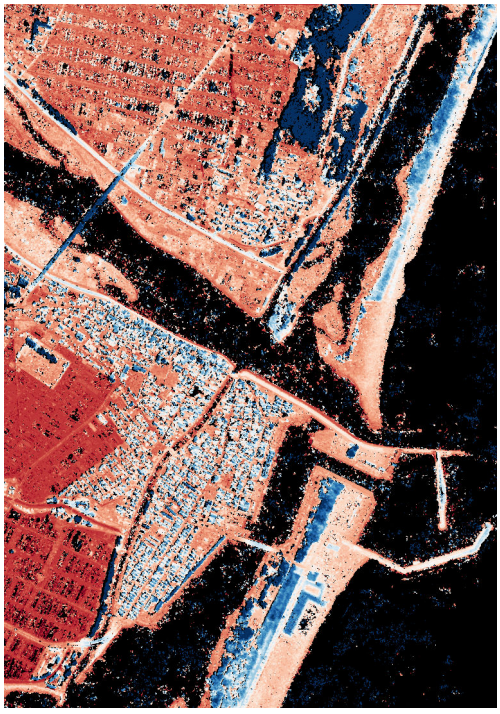
La région de Tohoku est située au Nord Est du Japon, sur la préfecture de Miyagi.

1. Nous rappelons que pour des raisons de confidentialité des donnnées, les données LiDAR ne peuvent être présentées ici.

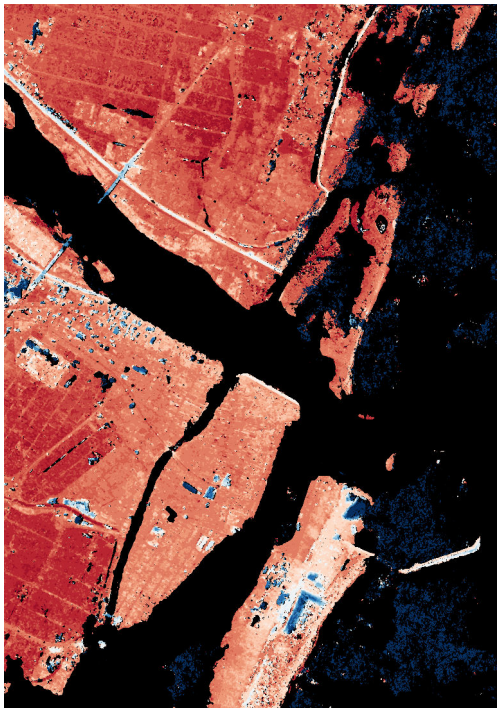


(a) Orthoimage Ikonos de 2010, Sendai.

(b) Orthoimage Ikonos de 2011, Sendai.



(c) MNS obtenu à partir du couple Ikonos 2010.



(d) MNS obtenu à partir du couple Ikonos 2011.

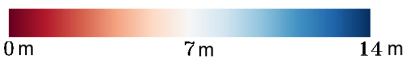



FIGURE 5.3 – *Ortho images et MNS calculés avant et après la catastrophe de Sendai du 11/03/2011.*

Cette région a été frappée le 11 mars 2011 par un séisme de magnitude 9 qui a été suivi par l'un des plus puissants tsunamis jamais connus au Japon puisque la vague a atteint 7,6 m de haut près de la préfecture de Miyagi et plus de 40 m plus au Nord [70].
Un rapport de police fait état de plus de 15 000 personnes décédées et 2800 disparus. Le tsunami a aussi généré de très importants dégâts avec près de 240 000 bâtiments détruits et 224 000 partiellement endommagés.

Les images présentées en figures 5.3(a) et 5.3(b) montrent une partie de la région touchée par cette catastrophe, avant et après le 11 mars 2011. Ces images proviennent de couples stéréoscopiques acquis par le capteur Ikonos le 11 décembre 2010 et le 13 août 2011 (cf. tableau 1.1, chapitre 1, section 1.4).
Contrairement aux zones de Phoenix et Christchurch précédemment présentées, la zone de Sendai montre un paysage totalement différent entre les acquisitions avant et après le passage du tsunami. La zone habitable a été presque totalement détruite de même que tous les terrains agricoles aux alentours. Ce type de paysage rendrait totalement impossible une détection de changements radiométriques qui génèrerait alors un très grand nombre de fausses alarmes.
Les MNS présentés en figures 5.3(c) et 5.3(d) font état de l'ampleur de la catastrophe et des destructions générées.
Afin de tester les performances du système pour la détection des dégâts, une large zone, d'environ 360 $km^2$ a été sélectionnée parmi les zones les plus touchées (représentée par le rectangle rouge dans les figures 5.3(a) et 5.3(b). Cette zone ainsi que la carte de référence des changements sont illustrées en figure 5.4.
On peut observer sur cette zone la destruction totale de la partie la plus proche de la côte, puis, plus au nord–ouest, des bâtiments qui semblent avoir résisté. Le nombre de bâtiments changés est alors largement supérieur à celui des bâtiments inchangés.
L'objectif est donc de tester notre chaîne de traitements sur une zone sinistrée afin d'observer les résultats obtenus en détection de changements du bâti.

Comme pour les zones urbaines précédentes, une carte de référence des changements a été réalisée sur la zone. Cependant, les dégâts importants et donc les changements massifs rendent difficile la quantification des résultats, notamment par le coefficient Kappa (que nous

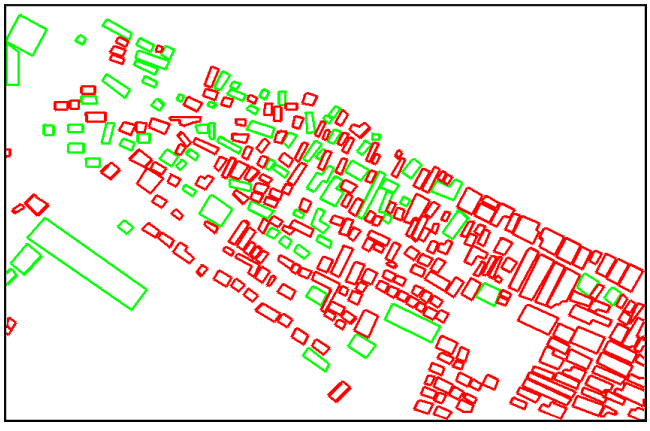
expliquerons section 5.2.2).

Pour résoudre ce problème et quantifier précisément les performances de la méthode, la carte de référence contient tous les changements (bâtiments détruits) mais aussi les non–changements (bâtiments intacts) de la scène. Au total, 220 bâtiments détruits ont été répertoriés et 76 bâtiments intacts.

Nous précisons cependant que cette carte de référence, réalisée manuellement à partir des images panchromatiques avant et après la catastrophe, regroupe parfois plusieurs bâtiments adossés les uns aux autres en un seul élément (détruit ou non), ce qui entraîne une diminution de sa précision.

(a) Orthoimage Ikonos de 2010, Sendai.

(b) Orthoimage Ikonos de 2011, Sendai.



(c) Carte de référence des bâtiments détruits (en rouge) et intacts(en vert).

FIGURE 5.4 – *Zone partiellement détruite par le Tsunami et quantifiée à travers la carte de référence des bâtiments détruits et intacts.*

## 5.2 Métriques d'analyse des résultats de la détection de changements

### 5.2.1 Variables calculées

Afin d'analyser les performances de la méthode, les alarmes, vraies ou fausses, détectées par la méthode sont comptées à l'échelle de l'objet et non du pixel. En effet, la qualité de nos données de référence, et notamment le détourage des changements dans les cartes de référence générées manuellement, ne permet pas une correspondance au pixel près entre les détections sur la carte des changements détectés et les changements détourés sur la carte de référence. De plus, la méthode de détection des changements mise en place est basée sur une régularisation spatiale ce qui tend à éroder les bords des alarmes de changements dans la carte résultat.
Le décompte des pixels bien ou mal classés ne représenterait alors pas la réalité de la détection.

Le décompte des vraies et fausses détections est donc effectué au niveau objet, c'est–à–dire que tout pixel isolé ou amas de pixels connexes (avec 8 connexes considérés autour de chaque pixel) rencontré dans la carte des changements résultante est compté comme une alarme de changement.
Ainsi, une vraie détection est confirmée dés lors qu'un, ou plusieurs pixels, labellisés en changement dans la carte des changements, correspondent à un changement réel dans la carte de référence. Inversement, si un pixel isolé, ou un amas de pixels, labellisé changement dans la carte des changements, ne correspond à aucun changement de la carte de référence, cette détection est considérée comme une fausse alarme.
Nous noterons enfin que dans la carte de référence des changements, les changements réels sont marqués sans considération de leur sens (changement positif ou négatif). Le décompte des changements réels bien détectés par la méthode se fait donc sans cette considération (on parlera des classes changement et non–changement). En effet, le décompte séparé des changements positifs et négatifs n'apporte pas d'indice sur les performances de la méthode.
La détection distincte de ces changements lors du calcul de la carte des changements est nécessaire au fonctionnement de la méthode de détection mais pas à l'analyse des résultats.

La quantification des résultats est effectuée à travers le calcul de plusieurs variables :

- TP (*True Positives* ou vrai positifs) : nombre de changements réels bien détectés,
- TN (*True Negatives* ou vrai négatifs) : nombre de non–changements bien détectés,
- FP (*False Positives* ou faux positifs) : nombre de non–changements détectés comme changement (aussi appelé fausse alarme),
- FN (*False Negatives* ou faux négatifs) : nombre de changements réels omis (ou omissions).

Le calcul de la quantité TN est difficile et n'a jamais été réalisé, à notre connaissance, à l'échelle objet dans la littérature. Pourtant cette variable est essentielle pour le calcul d'une de nos métriques d'analyse, le coefficient Kappa.
Lors de ce travail, nous avons choisi de calculer cette valeur selon une définition empirique, basée sur le nombre de pixels total de l'image divisé par la taille moyenne d'un objet (équation 5.1). De cette façon, l'image est considérée comme si elle était totalement constituée d'objets à détecter.

$$TN = \frac{(l \times c)}{moyenne_{objet}} - TP - FN - FP. \tag{5.1}$$

Dans l'équation 5.1, $l$ et $c$ représentent le nombre de lignes et de colonnes de l'image respectivement, $moyenne_{objet}$ représente la taille moyenne d'un objet d'une scène urbaine. Il est déterminé ici à environ $15 \times 15$ pixels, valeur considérée comme une taille moyenne de détections si toutes les détections (vraies et fausses) sont comptées à faible régularisation.

Avec cette définition particulière de TN, toute l'emprise de l'image est alors prise en compte, quelle que soit la densité des habitations. En effet, si le TN devait être représenté par le nombre de bâtiment uniquement, alors, dans une scène contenant moins de 10 bâtiments mais de nombreuses sources de fausses alarmes, le nombre de fausses alarmes serait particulièrement élevé et le TN très faible, diminuant la précision de la méthode sans pour autant représenter la réalité de la détection. La considération de la scène comme recouverte d'objets permet ainsi d'améliorer la précision du calcul des performances de la méthode.
Il est cependant à noter que cette variable TN sera utilisée dans une métrique particulière permettant la comparaison, entre elles, des classifications réalisées dans ce travail et la recherche des paramètres de la méthode les plus adaptés. La définition de TN est alors la même pour tous les résultats de classification qui seront comparés lors de ce travail.

### 5.2.2 Métriques d'analyse

Les métriques d'analyse des résultats utilisées sont des métriques standards qui permettent l'évaluation des performances de la détection de changements par rapport aux données de référence.

Les métriques les plus classiques correspondent aux taux de bonnes détections (ou *True Positive Rate*, TPR) et de fausses détections (ou *False Positive Rate*, FPR), calculées selon les équations 5.2.

$$TPR = \frac{TP}{TP + FN} \times 100, \qquad FPR = \frac{FP}{FP + TP} \times 100. \tag{5.2}$$

Une autre métrique importante est représentée par le coefficient Kappa (équation 5.3) qui correspond au calcul de l'accord entre la vérité terrain et le résultat de classification obtenu.
La première description et utilisation de ce coefficient provient de Cohen [71] et a permis de calculer le degré de cohérence entre deux jugements psychiatriques indépendants.
Plus récemment et dans notre domaine d'étude, ce coefficient a été employé notamment par Wilkinson [72] pour l'analyse et l'inter–comparaison des résultats de différentes méthodes de classification d'images recensés sur 15 ans. L'objectif de cette analyse était l'évaluation des performances de nombreuses méthodes de classification supervisées ou non, utilisant des données externes ou non et quel que soit le nombre de classes ou la résolution initiale de l'image. Dans ce contexte, l'utilisation du coefficient Kappa, dont les données d'entrées correspondent au nombre de classes et aux variables TP, TN, FP et FN décrites précédemment, apparaît alors particulièrement adaptée.

Ce coefficient est calculé selon l'équation suivante :

$$K = \frac{OA - P_e}{1 - P_e}, \tag{5.3}$$

$$OA = \frac{TP}{TP + FP + FN}, \tag{5.4}$$

$$P_e = \frac{M}{(TP + FP + TN + FN)^2}, \tag{5.5}$$

$$M = (TP + FP) \times (TP + FN) + (FN + TN) \times (FP + TN). \tag{5.6}$$

$OA$ correspond à la précision générale de la classification (*Overall Accuracy*). $P_e$ est décrit comme le taux de concordance aléatoire. Cette valeur est égale à 1 si la correspondance entre la vérité terrain et le résultat de classification obtenu est uniquement le fait du hasard et à 0 sinon.

L'avantage de cette métrique est d'être particulièrement adaptée lorsque les classes en présence sont déséquilibrées, c'est–à–dire, dans notre contexte, lorsque la classe des non–changements est très grande par rapport à la classe changement.

De plus, les taux de vraies et de fausses détections ne permettent pas de définir la régularisation qui entraînera le meilleur équilibre entre fausses alarmes et bonnes détections. En effet, ces métriques, indépendantes l'une de l'autre, indiquent le meilleur paramétrage de l'algorithme en fonction du taux de bonnes détections recherchées ou bien du taux de fausses alarmes maximum accepté.
Le coefficient Kappa prend en compte tous les termes de la matrice de confusion de la détection de changement (TP, TN, FP, FN). Cette métrique permet ainsi la comparaison précise de plusieurs classifications tout en mettant en évidence le paramétrage de l'algorithme le plus adapté, qui correspond alors à la meilleure correspondance entre la carte de référence des changements et la carte calculée et donc au meilleur équilibre entre bonnes détections et fausses alarmes. Le calcul du coefficient Kappa est donc particulièrement important pour l'analyse complète des résultats obtenus par la méthode de détection tandis que l'analyse des différents taux de bonne ou fausse détection est importante pour expliquer le comportement du Kappa.
Ces deux métriques sont très complémentaires entre elles.

## 5.3 Analyse de sensibilité des résultats obtenus

Dans cette section, plusieurs analyses de sensibilité sont effectuées. Elles concernent les différents paramètres de la méthode et leur impact sur les résultats mais aussi l'analyse de l'impact de la qualité des données d'entrée (MNS) ou des changements recherchés ou encore l'analyse des performances en fonction des caractéristiques de la zone étudiée.

Les cartes de détections de changements calculées en vue de ces analyses de sensibilité ont,

pour la plupart, été générées à partir de la zone 1 de Phoenix. Cette zone présente en effet de très nombreux changements et des sources réduites de fausses alarmes ce qui en fait une zone test intéressante pour la mesure de l'influence des paramètres ou des données d'entrée.

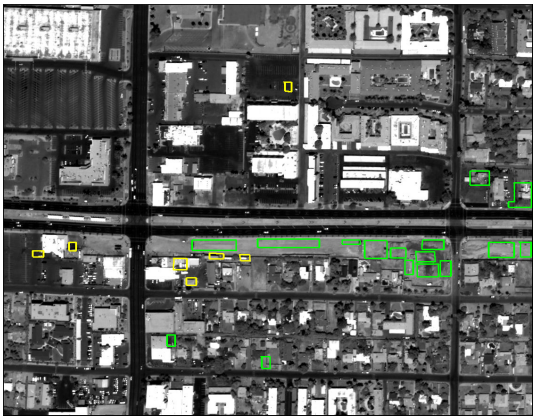
### 5.3.1 Influence des paramètres de la méthode sur les résultats

#### Nombre de directions de la programmation dynamique

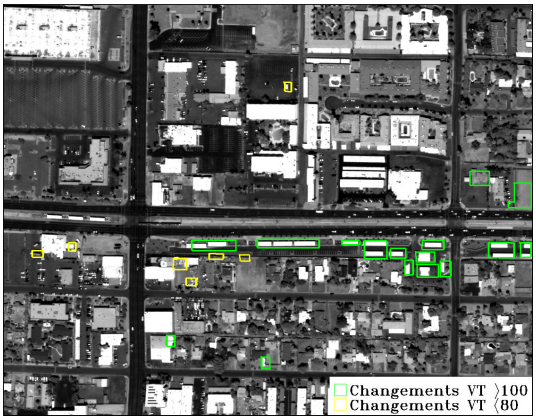
Le premier paramètre de la détection de changements que nous étudions est le nombre de directions utilisées dans le calcul. En effet, comme décrit dans le chapitre 2, section 2.2.3, l'algorithme de programmation dynamique, conçu au cours de cette thèse, nécessite le calcul des coûts de labellisation selon plusieurs directions dans l'image, coûts calculés indépendamment pour chaque direction puis fusionnés afin de retrouver la solution finale de labellisation. Nous avons montré en effet que l'utilisation de nombreuses directions permet de limiter les effets de *streaking* sur le résultat final.

Les figures 5.5(c), 5.5(d), 5.5(e) et 5.5(f) illustrent les résultats obtenus avec 1, 8, 12 et 16 directions respectivement et une régularisation moyenne $\lambda = 3,0$.
On observe que lorsqu'une seule direction est calculée, les effets de *streaking* sont particulièrement importants (visible le long des lignes de l'image 5.5(c)). D'après la figure 5.5(d), le calcul de 8 directions réduit l'effet de *streaking*. Puis, de 12 à 16 directions, le résultat est globalement identique et l'effet de *streaking* est quasi nul.

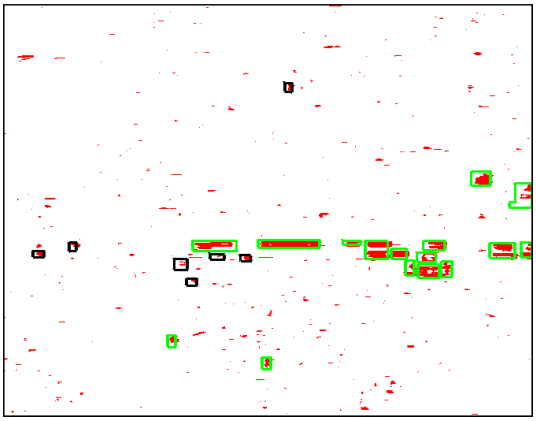
La figure 5.6 présente les taux de bonnes et fausses détections obtenues avec une régularisation fixée à $\lambda = 3,0$ et un nombre variable de directions de calcul, de 1 à 16 directions.
La régularisation a été fixée à une valeur relativement faible afin de bien observer l'impact de ce nombre de directions sur un grand nombre de détections.
Comme attendu d'après les figures précédentes, on observe une très forte diminution des fausses alarmes (mais aussi des bonnes détections) entre 1 et 4 directions de calcul. À partir de 4 directions, les taux diminuent de façon moins flagrante puis ils semblent se stabiliser entre 12 et 16 directions.



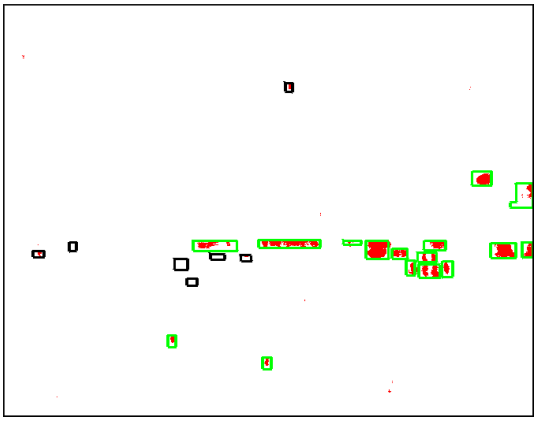
(a) Orthoimage WorldView-1, 2008.



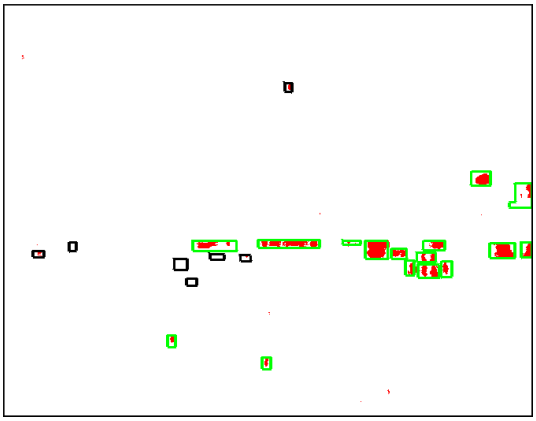

(b) Orthoimage WorldView-1, 2011.



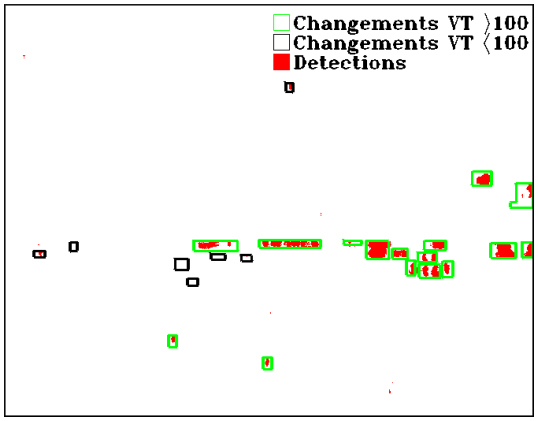
(c) Carte des changements calculée avec 1 direction de calcul.



(d) Carte des changements calculée avec 8 directions de calcul.



(e) Carte des changements calculée avec 12 directions de calcul.




(f) Carte des changements calculée avec 16 directions de calcul.

FIGURE 5.5 – *Cartes de détection de changements obtenues à partir d'un nombre variable de directions de calcul et une régularisation* $\lambda = 3,0$. *Résultats calculés sur la zone 1 de Phoenix.*

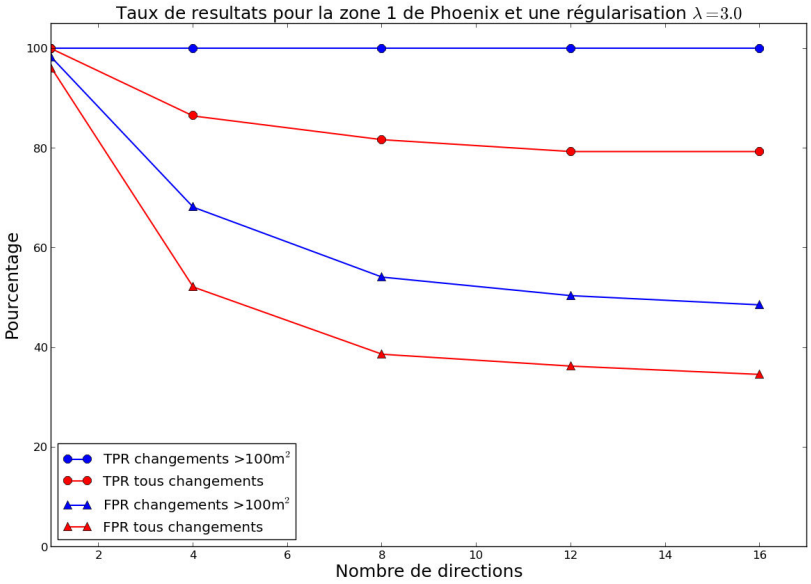



FIGURE 5.6 – *Illustration de l'impact du nombre de directions de calcul sur les taux de bonnes et fausses détections. Résultats calculés sur la zone 1 de Phoenix avec une régularisation* $\lambda = 3,0$.

La détection de changements est donc généralement calculée avec 12 directions. Le calcul à 16 directions n'apportant pas de différences significative au résultat mais augmente le temps de traitement.

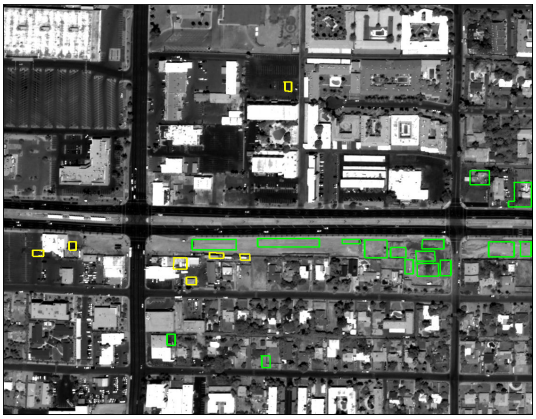
### Paramètre de régularisation

Le paramètre $\lambda$ est le paramètre ayant le plus d'influence sur le résultat de détection de changements.

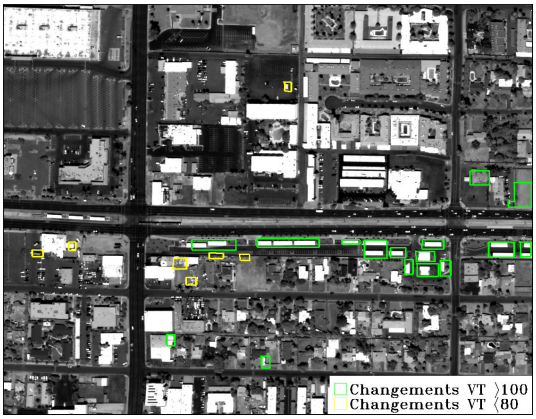
Les figures 5.7(c), 5.7(d), 5.7(e) et 5.7(f) illustrent qualitativement l'effet de ce paramètre. Chacune de ces figures représente une carte de détection de changements obtenue à partir d'un $\lambda$ égal à 2,0, 3,0, 4,5 et 6,0, respectivement. D'après ces figures, les alarmes de changements disparaissent en fonction de leur taille, que ce soit les fausses alarmes (situées principalement dans la carte obtenue avec une très faible régularisation) ou les vrais changements, et ce, de la plus faible à la plus forte régularisation. On note, par exemple, que de nombreux changements réels marqués dans la carte de référence ne sont pas détectés à $\lambda = 6,0$.

Afin d'observer les effets de la régularisation sur les résultats de détection de changement, les taux de bonnes détections (TPR) et de fausses alarmes (FPR) ont été tracés en fonction du

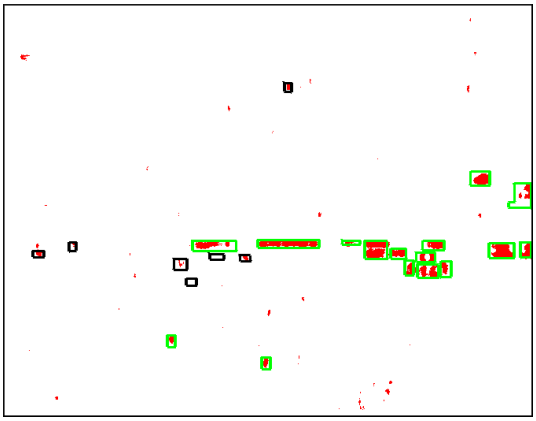
paramètre de régularisation sur les figures 5.8(a) à 5.8(d).



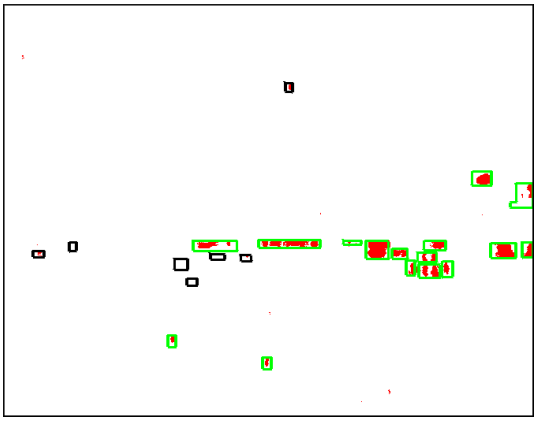
(a) Orthoimage WorldView-1 de 2008, Phoenix zone #1.



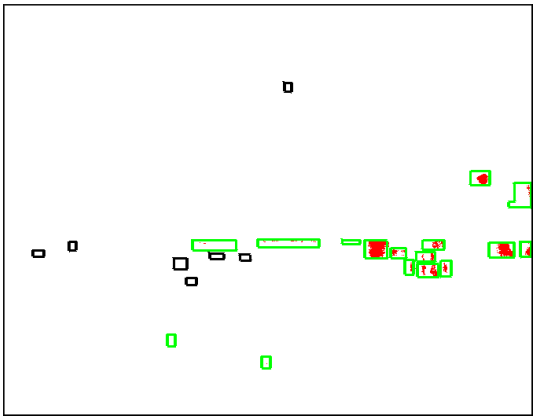

(b) Orthoimage WorldView-1 de 2011, Phoenix zone #1.



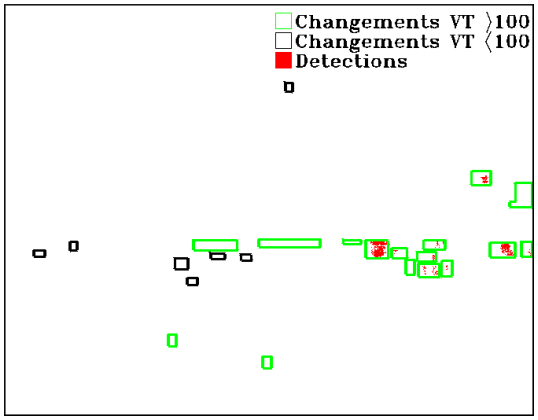
(c) Carte des détections obtenue avec $\lambda = 2$.



(d) Carte des détections obtenue avec $\lambda = 3$.



(e) Carte des détections obtenue avec $\lambda = 5$.




(f) Carte des détections obtenue avec $\lambda = 6$.

FIGURE 5.7 – *Cartes des détections calculée avec différentes valeurs de $\lambda$ et 12 directions. Résultats calculés sur la zone 1 de Phoenix avec une régularisation $\lambda = 3,0$*

Ces figures présentent le comportement des métriques avec l'augmentation de la régularisation.

Dans un premier temps, les TPR (figures 5.8(a) et 5.8(c)) marquent un palier proche de 100% de bonnes détections tandis que les taux de FPR, d'abord tout aussi haut, diminuent de façon presque linéaire. Puis, pour des $\lambda$ entre 2,5 et 3,5, le taux de bonnes détections commence à décroître mais plus faiblement que le taux de fausses alarmes qui baisse drastiquement. Puis la courbe des fausses alarmes atteint un pallier vers $\lambda = 4,0$ de même que celle des TPR qui ne décroit que faiblement.

Ce comportement résulte en fait de la suppression des petites fausses alarmes, souvent très nombreuses à faible régularisation, ce qui entraîne une très forte baisse des taux FPR. Les petits changements pertinents, eux aussi supprimés, sont cependant moins nombreux, expliquant la décroissance moins rapide de la courbe des TPR.

Lorsque toutes ces petites détections sont supprimées, les détections restantes sont alors des fausses alarmes ou des vraies détections de taille importante et donc très difficiles à supprimer à moins d'augmenter très fortement la régularisation, ceci explique le palier à la fin des courbes.

**Autres paramètres de la méthode**

Nous avons vu, dans la section 4, que d'autres paramètres interviennent dans la méthode de détection de changements développée. Ces paramètres sont notamment le seuil $T$, fixé à 2,5 m et la courbure $L$, fixée à 3,0.

Des analyses de sensibilité pourraient être réalisées pour observer les performances de la méthode en fonction de ces deux paramètres. Cependant, le seuil a été fixé ici d'après une valeur standard et en considérant la résolution altimétrique obtenue pour les MNS. C'est pourquoi aucune étude n'a été menée quant à la sensibilité de ce paramètre.

La courbure $L$ permet, comme nous l'avons montré, de contrôler les coûts pour chaque label en fonction de la proximité entre la valeur du pixel et le seuil. Ce paramètre n'a cependant que peu d'impact d'après les quelques tests que nous avons réalisés cependant aucune analyse complète n'a encore été effectuée.

### 5.3.2 Impact de la qualité des données d'entrée

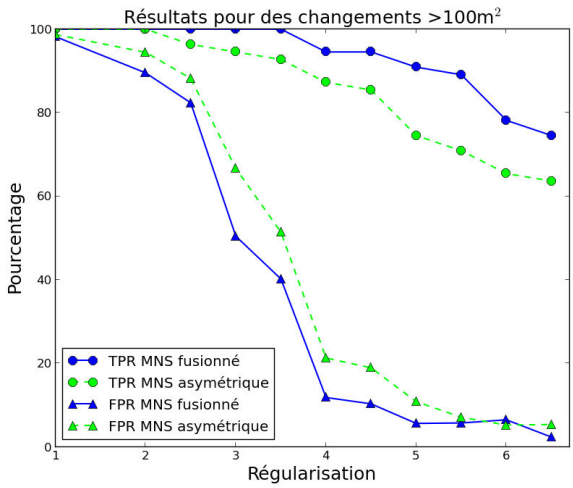
**Influence de la qualité des MNS**

Afin d'observer la sensibilité de la méthode de détection des changements envers les données d'entrée qui lui sont injectées, c'est–à–dire les MNS utilisés, l'algorithme a été testé sur des MNS de qualité différente : les MNS obtenus après une mise en correspondance asymétrique simple et les MNS obtenus après fusion de MNS asymétriques (technique décrite dans le chapitre 3, section 3.3). La qualité de ces MNS a été discutée dans section 3.4.

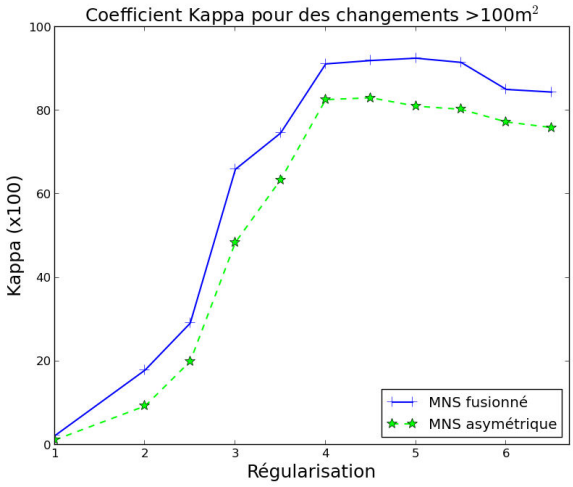
Les figures 5.8 présentent les résultats obtenus pour la zone 1 de Phoenix, dont les changements ont été détectés à partir des MNS asymétriques (courbes vertes et turquoises pointillées) et des MNS fusionnés (courbes bleues et rouges pleines).
Les résultats ont été calculés sur la zone 1 de Phoenix puis tracés en fonction du paramètre $\lambda$ utilisé.

Les figures 5.8(a) et 5.8(c) présentent les taux de bonnes et de fausses détections lorsque seuls les changements supérieurs à 100 $m^2$ sont considérés et lorsque tous les changements sont considérés.
D'après ces figures, les taux de bonnes détections obtenues avec les MNS fusionnés apparaissent sensiblement plus élevés que ceux obtenus avec les MNS asymétriques : une différence de 5% à 20% est notée. De même, les taux de fausses alarmes détectées à partir des MNS fusionnés sont plus faibles que ceux obtenus avec les MNS asymétriques. L'écart constaté est de 0 à plus de 10%.

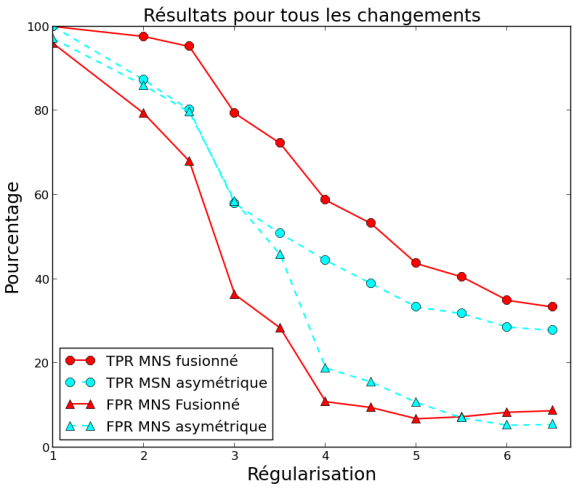
Les figures 5.8(b) et 5.8(d) présentent les coefficients Kappa obtenus pour tous les changements ou les changements supérieurs à 100 $m^2$ uniquement, pour le MNS fusionné et pour le MNS asymétrique présentés. D'après ces figures, le Kappa apparaît fortement corrélé aux taux de fausses et bonnes détections : plus les taux de vraies détections et de fausses détections sont proches l'un de l'autre (c'est–à–dire autant de bonnes détections que de fausses alarmes) plus le Kappa est faible, indiquant une mauvaise classification et inversement un taux de bonnes détections très supérieur à un taux de fausses alarmes entraine un fort coefficient Kappa et donc une bonne classification.



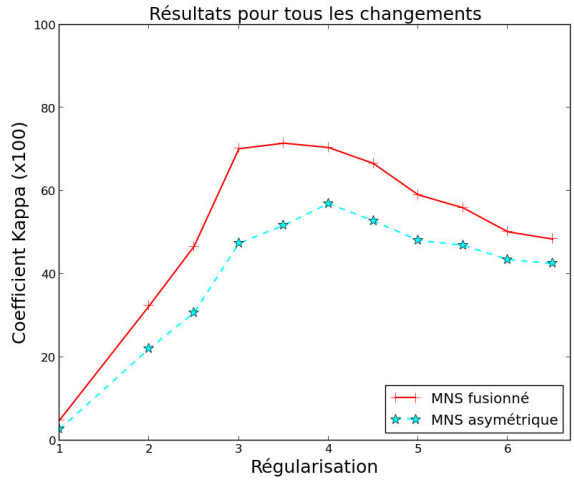

(a) Taux de vrais et faux positifs calculés pour les changements supérieurs à 100 $m^2$.




(b) Coefficient Kappa calculé pour les changements supérieurs à 100 $m^2$.




(c) Taux de vrais et faux positifs calculés pour tous les changements de la scène.




(d) Coefficient Kappa calculé pour tous les changements de la scène.

FIGURE 5.8 – *5.8(a) et 5.8(c) variation des taux de fausses et vraies détection en fonction de la régularisation λ. 5.8(b) et 5.8(d) : Variations du coefficient Kappa en fonction de la régularisation λ. Résultats calculés pour la zone 1 de Phoenix.*

Comme attendu ici, les valeurs de Kappa obtenues à partir des MNS fusionnés sont largement plus hautes que celles provenant des résultats de MNS asymétriques, quelle que soit la taille des changements recherchés.

D'après les différents taux TPR et FPR calculés, cela s'explique par le nombre important

des fausses alarmes par rapport au nombre total de détections lorsque les MNS utilisés sont asymétriques. En effet, les MNS fusionnés présentent une meilleure précision sur les fines structures de la scène et moins d'erreurs de corrélation, entraînant un taux de fausses alarmes plus faible. Cette différence a néanmoins tendance à disparaître avec des régularisations plus importantes puisque ce paramètre agit comme un filtre qui supprime les plus fines erreurs des MNS. Le résultat final à haute régularisation est alors très similaire que ce soit pour les MNS fusionnés ou pour les MNS asymétriques.

**Influence de la taille de changements recherchée**

Les figures 5.8 présentent les résultats obtenus pour la zone 1 de Phoenix lorsque l'on considère les changements supérieurs à 100 $m^2$ (figures 5.8(a) et 5.8(b)). Dans ce cas, le taux de bonnes détections observé est de 100% à plus de 75% pour des régularisations de 1 à 6,5.
À la valeur de Kappa maximum (soit à une régularisation à $\lambda = 5,0$ d'après la figure 5.8(b)), le taux de bonnes détections est de 90%, avec moins de 10% de fausses alarmes.

Lorsque les changements de toutes tailles sont considérés, la valeur maximale du kappa atteint seulement 0,76 pour une régularisation à $\lambda = 3,0$. D'après la figure 5.8(c), cela correspond à environ 80% des changements bien détectés et 35% de fausses alarmes.
En effet, les changements inférieurs à 100 $m^2$ représentent les deux tiers de tous les changements mais sont très difficiles à détecter et ont tendance à être supprimés rapidement lorsque la régularisation augmente. Une plus faible régularisation est donc nécessaire pour les repérer ce qui implique alors un taux de fausses alarmes résiduelles plus important et donc un coefficient Kappa plus faible.

Nous rappelons que la taille des changements recherchés dépend en premier lieu de la résolution des images (puis des MNS) en entrée de la chaîne de traitements. En effet, la taille de 100 $m^2$ (soit $15 \times 15$ pixels) est adaptée pour des images acquises à 60 cm de résolution permettant la génération de MNS à 1 m de résolution spatiale. Cette taille doit être redéfinie en fonction de la résolution des images d'entrée du système.

### 5.3.3 Intérêt d'une méthode globale : comparaison avec un filtrage local

Cette section présente la comparaison de la méthode de classification des changements employée avec une méthode plus classique, représentée par un filtrage morphologique. Ce type de filtrage local est en effet souvent utilisé dans la littérature et plus particulièrement par Tian [26] et Chaabouni [22] pour l'analyse de MNS différentiels en vue de la détection des changements d'élévation.

Le filtrage local testé est appliqué sur le résultat d'une segmentation (à +2,5 m et -2,5 m, valeur de seuil identique à celle utilisée dans le filtrage global) du MNS différentiel. Ce filtrage est basé sur l'application d'une opération morphologique sur l'image segmentée, sous forme d'une ouverture.

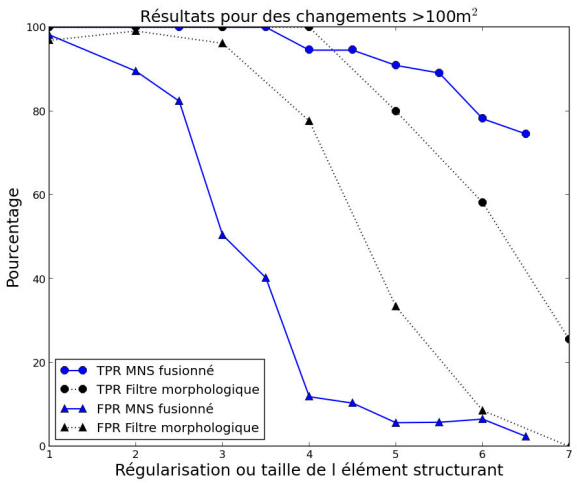
Pour rappel, une ouverture (binaire), en morphologie mathématique, correspond à l'érosion des amas de pixels connexes portant une valeur égale à 1 (donc un label changement dans notre cas) à l'aide d'un élément structurant, suivi par la dilatation de ces mêmes amas avec le même élément [73].

Cette opération permet de supprimer les détections (vraies ou fausses) de taille inférieure à l'objet structurant. Ainsi, la variation de la taille de l'objet structurant permet de contrôler le nombre de fausses détections mais aussi celui des vraies détections en fonction de leur taille. Cette technique se rapproche ainsi du filtrage global dont le paramètre de régularisation permet aussi le contrôle des taux de fausses alarmes et de vrais positifs.

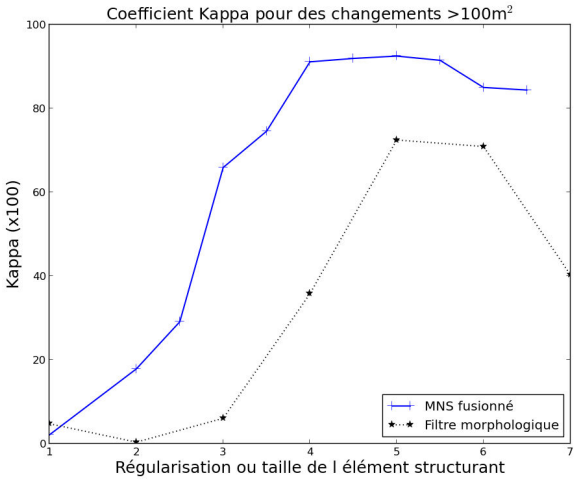
Les résultats de ce filtrage morphologique sont présentés dans les figures 5.9 et comparés aux résultats obtenus avec le filtrage global.

Sur les figures, les courbes noires représentent les résultats obtenus avec le filtrage morphologique. Les courbes bleues et rouges présentent les résultats obtenus avec le filtrage global pour les changements supérieurs à 100 $m^2$ et tous les changements, respectivement.

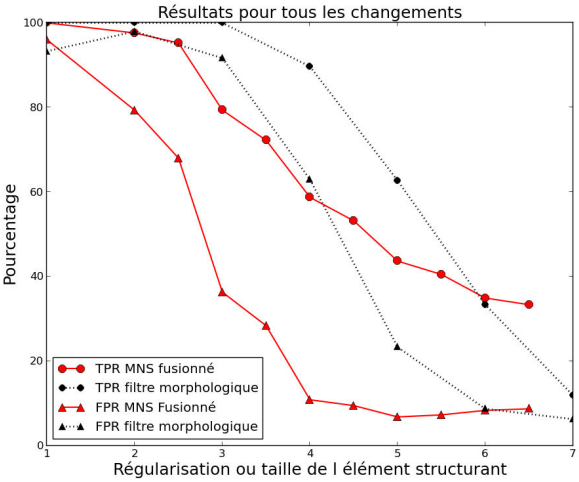
Lorsque seuls les changements importants (supérieurs à 100 $m^2$) sont considérés, les taux de bonnes détections des deux types de filtrage atteignent tous les deux les 100% (à faible régularisation) mais avec des taux de fausses alarmes de 78% environ pour le filtrage morphologique et de 50% pour le filtrage optimisé (figure 5.9(a)).



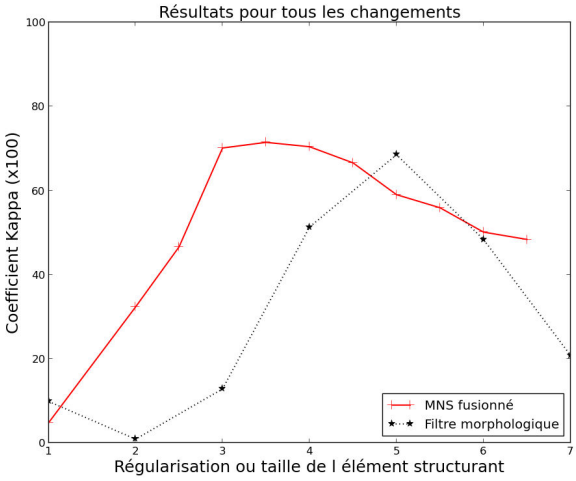

(a) Taux de vrais et faux positifs calculés pour les changements supérieurs à 100 $m^2$.




(b) Coefficient Kappa calculé pour les changements supérieurs 100 $m^2$.




(c) Taux de vrais et faux positifs calculés pour tous les changements de la scène.




(d) Coefficient Kappa calculé pour tous les changements de la scène.

FIGURE 5.9 – *5.9(a) et 5.9(c) : Variations des taux de fausses et vraies détections en fonction de la régularisation $\lambda$ ou de la taille de l'élément structurant. 5.9(b) et 5.9(d) : Variations du coefficient Kappa en fonction de la régularisation $\lambda$ ou de la taille de l'élément structurant. Résultats calculés sur la zone 1 de Phoenix*

Si un faible taux de fausses alarmes est recherché (moins de 10%), le taux de bonnes détections descend alors à 58% pour le filtrage morphologique tandis qu'il se maintient au–dessus des 90% pour le filtrage optimisé.

Les courbes Kappa (figure 5.9(b)) confirment une différence importante entre les résultats des

filtrages puisque la valeur maximale du Kappa atteint par le filtrage optimisé est de 0,92 tandis qu'elle atteint seulement 0,72 pour le filtrage morphologique.

Lorsque tous les changements sont considérés, le taux de bonnes détections du filtrage morphologique apparaît globalement plus élevé que celui du filtrage optimisé mais le taux de fausses alarmes est plutôt élevé par rapport à celui du filtrage optimisé (figure 5.9(c)).
Les courbes Kappa montrent en effet que certains résultats des filtrages optimal et morphologique sont comparables lorsqu'une régularisation de $\lambda = 4,0$ est utilisée pour le filtrage optimisé et un élément structurant de taille $5 \times 5$ est utilisé pour le filtrage morphologique. Avec ces paramètres, le filtrage optimisé présente un TPR d'environ 58% et un FPR d'environ 10% tandis que le filtrage morphologique présente un TPR d'environ 61% pour un taux FPR d'environ 20%, entraînant une valeur de Kappa de 0,68.
Cependant, pour une régularisation à $\lambda = 3,5$, régularisation donnant les meilleurs résultats pour le filtrage optimisé d'après les courbes Kappa en figure 5.9(d), le TPR obtenu est d'environ 72% pour un FPR de moins de 30% et une valeur de kappa à 0,71, légèrement meilleure que celle obtenu par filtrage morphologique.

Les résultats obtenus pour le filtrage morphologique apparaissent ainsi d'une qualité presque semblable à celle du filtrage optimisé lorsque tous les changements sont recherchés mais très en dessous des résultats du filtrage optimisé lorsque la détection se concentre sur les changements les plus larges uniquement. Les taux de bonnes détections et de fausses alarmes étant à chaque fois inférieurs en qualité à ceux du filtrage optimisé.
Si nous ne réfutons pas que les résultats obtenus, particulièrement pour les changements larges, peuvent être dus à l'utilisation non optimale des opérations morphologiques, nous montrons cependant ici que le filtrage optimisé et son paramétrage peuvent être utilisés de façon très simple et efficacement pour la recherche des changements quelle que soit leur taille. Tandis que les opérations morphologiques peuvent être complexes à paramétrer afin d'être adaptées aux changements recherchés et montrer des performances équivalentes.

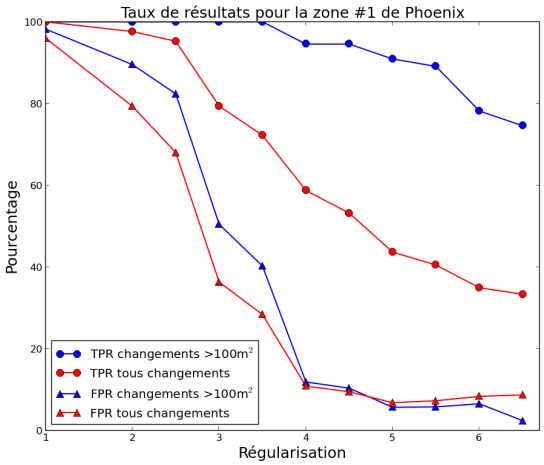
### 5.3.4 Influence du paysage de la zone d'étude

Dans cette section, les performances obtenues sur les différentes zones tests présentées au début de ce chapitre seront comparées entre elles afin d'analyser les résultats en fonction des zones, des difficultés qu'elles présentent et du nombre de fausses alarmes qu'elles génèrent.

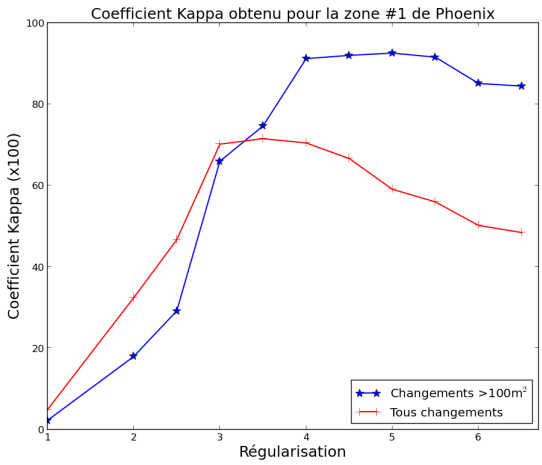
#### Phoenix

Les résultats obtenus sur les trois zones de Phoenix sont présentés en figure 5.10 avec, comme précédemment le calcul des taux de bonnes (TPR) et fausses (FPR) détections et le calcul du coefficient Kappa pour tous les changements et pour les changements supérieurs à 100 $m^2$.

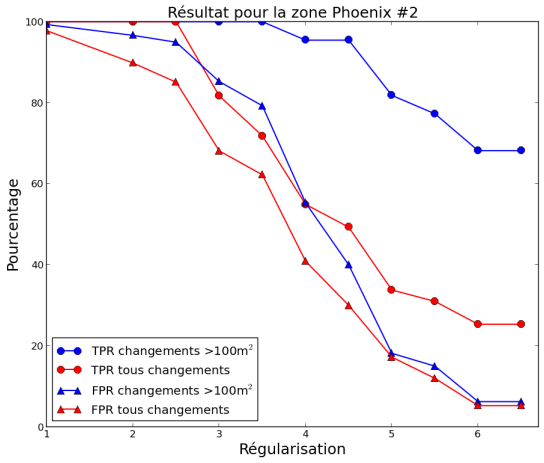
Lorsque seuls les grands changements sont pris en compte, globalement et selon les zones, les taux de bonnes détections se situent entre 100% et 75%, avec un taux de fausses alarmes variant alors de 100% à moins de 5%.
Malgré les variations de paysage de ces zones, la régularisation optimale pour ces changements se situe toujours approximativement à $\lambda = 5,0$, d'après les courbes Kappa présentées dans les figures 5.10(b), 5.10(d) et 5.10(f).
Ainsi, à cette régularisation, la valeur du Kappa est alors égale à 0,93 pour la zone 1, avec 90% de bonnes détections et 5,6% de fausses alarmes. Pour la zone 2, on obtient une valeur de Kappa égale à 0,82 pour $\lambda = 5,0$ pour des taux de 81% de bonnes détections et de 18% de fausses alarmes. Enfin, concernant la zone 3, la valeur du Kappa est à 0,87 pour $\lambda = 5,0$ et les taux de bonnes et fausses détections atteignent 79,5% et 3,57%, respectivement.
Lorsque tous les changements sont considérés, en revanche, la régularisation optimale est d'environ $\lambda = 3,5$. Les coefficients Kappa, calculés pour chacune des zones sont alors de 0,72, 0,50 et 0,59, respectivement pour les zones 1, 2 et 3. Les taux de bonnes détections atteignent alors 72% pour les zones 1 et 2 et 53% pour la zone 3 tandis que les taux de fausses alarmes sont de 28,3% pour la zone 1, 62% pour la zone 2 et 34% pour la zone 3.
D'après tous ces résultats, la zone 1 semble présenter les meilleurs taux de bonnes et fausses détections. Cette zone contient en fait de très nombreux changements (126) dont près de la moitié sont des changements de grande taille.



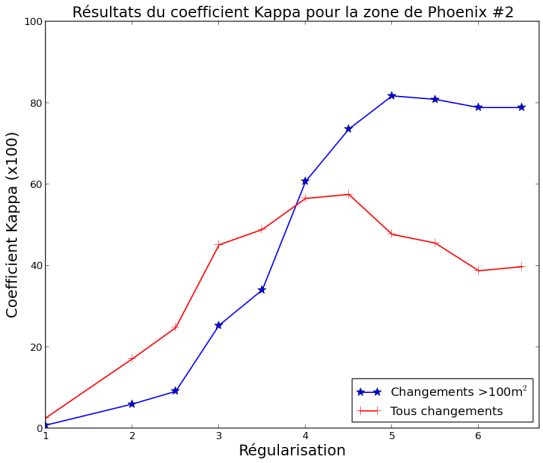

(a) Taux de vrais et faux positifs pour la zone #1.



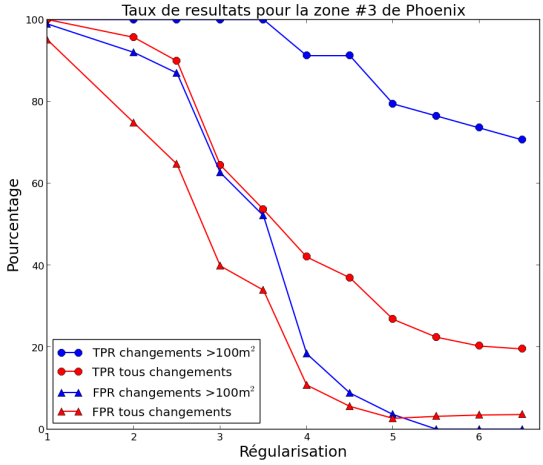

(b) Coefficient Kappa calculé pour la zone #1.



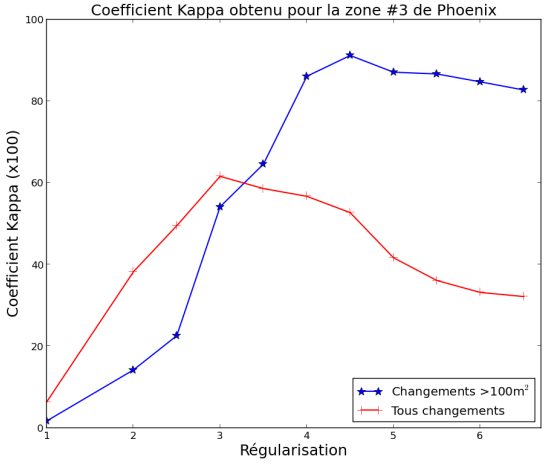

(c) Taux de vrais et faux positifs pour la zone #2.




(d) Coefficient Kappa calculé pour la zone #2.




(e) Taux de vrais et faux positifs pour la zone #3.




(f) Coefficient Kappa calculé pour la zone #3.

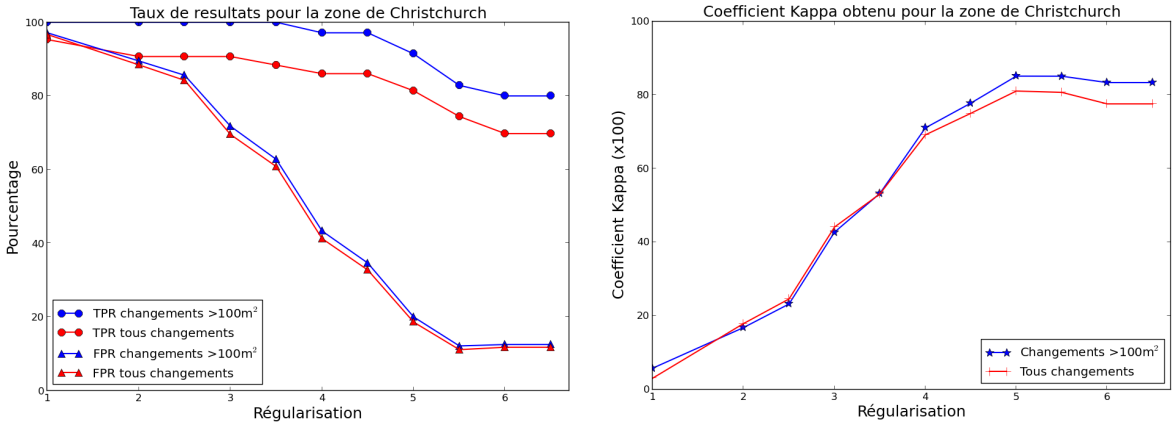
FIGURE 5.10 – *5.10(a), 5.10(c) et 5.10(e) Variations des taux de bonnes et fausses détections en fonction du λ. 5.10(b), 5.10(d) et 5.10(f) Variations du coefficient Kappa. Résultats calculés pour chacune des zones de Phoenix*

De plus, les sources de fausses alarmes sont assez réduites par rapport aux autres zones.
La zone 2 de Phoenix présente des résultats de moins bonne qualité. Ceci s'explique par le fait que cette zone contient de nombreux bâtiments particulièrement hauts. Ces bâtiments génèrent de grandes zones d'occlusion et donc de nombreuses erreurs qui ne sont pas totalement filtrées lors de la génération des MNS. De plus, cette zone contient plusieurs zones de végétation qui entraînent de nombreuses fausses alarmes. Enfin, seulement 71 changements sont recensés dans cette zone dont moins d'un tiers sont de grands changements bien détectables.
Finalement, la zone 3 présente des résultats situés qualitativement entre la zone 1 et la zone 2. Cette zone contient de nombreux changements (138) mais seul un quart correspond à des changements larges, ce qui explique les taux plus faibles de bonnes détections. Cette zone contient aussi une voie express avec de nombreux véhicules générant un certain nombre de fausses alarmes. Cependant, à forte régularisation, les nombreuses fausses alarmes générées, (qui sont de petite taille) sont supprimées permettant d'atteindre un très faible taux de fausses alarmes à $\lambda = 5$.

**Christchurch**

Les résultats obtenus pour la zone de Christchurch sont présentés en figure 5.11. Comme précédemment, les changements de taille supérieure à 100 $m^2$ sont représentés par la courbe bleue et les changements toutes tailles confondues par la courbe rouge. Au vu du très faible nombre de changements inférieurs à 100 $m^2$ (8 seulement), les courbes sont très proches, que ce soit les courbes de bonnes détections ou de fausses alarmes.
Dans les deux cas, le coefficient Kappa atteint plus de 0,8 (0,8 pour tous les changements et 0,85 pour les changements supérieurs à 100 $m^2$). Lorsque tous les changements sont considérés, ce coefficient apparaît très haut par rapport à celui obtenu sur les zones de Phoenix (il est de 0,72 dans le meilleur résultat de Phoenix). Ceci s'explique par le faible nombre de changements de petite taille. Lorsque seuls les grands changements sont considérés, le coefficient Kappa est de 0,85, coefficient proche des meilleurs résultats de Phoenix.

Cependant, d'après la figure 5.11(a), si le taux de bonne détection est haut (il va de 100% à 80% pour les changements larges), on observe un taux de fausses alarmes assez haut.




(a) Taux de vrais et faux positifs calculés pour la zone de Christchurch.

(b) Coefficient Kappa calculés pour la zone de Christchurch.

FIGURE 5.11 – *5.11(a) : Variations des taux de bonnes et fausses détections en fonction du $\lambda$. 5.11(b) : Variation du coefficient Kappa. Résultats calculés pour la zone de Christchurch.*

En effet, la zone sélectionnée dans Christchurch contient beaucoup de végétation sous la forme d'arbres imposants et très proches des habitations. Cet élément est souvent source de fausses alarmes, notamment lorsque des arbres sont coupés ou des haies sont plantées entre deux acquisitions.

Cette zone, assez différente en terme de paysage et contenant des défauts plus prononcés que les zones de Phoenix, montre tout de même une régularisation optimale aux environs d'un $\lambda = 5,0$ pour des changements supérieurs à 100 m$^2$. Cette valeur de régularisation, identique à celle trouvée pour les 3 zones de Phoenix apparaît ainsi bien adaptée au contexte urbain.

## 5.4 Analyse des résultats sur la zone de Sendai

Les résultats obtenus sur la zone de Sendai sont présentés en figure 5.12. Sur cette figure, la courbe des bonnes détections est tracée en bleue, la courbe des fausses alarmes en rouge et la courbe du coefficient Kappa en vert.
Contrairement aux zones présentées précédemment, la courbe des bonnes détections décroît doucement, de façon presque linéaire, tandis que la courbe des fausses alarmes, très basse même à faible régularisation, ne décroît que très peu puis semble se stabiliser à moins de 10% de fausses alarmes.

La courbe Kappa, quant à elle, présente un maximum à environ 6,5 pour une régularisation à $\lambda = 3,0$.

Le fort taux de bonnes détections, même à haute régularisation, s'explique par le nombre très important de changements contigus de la scène. En effet, près de 300 bâtiments ont été recensés sur une zone de 270 km$^2$ soit une densité d'habitation de plus d'un bâtiment pour 1 000 m$^2$. Densité indiquant des bâtiments très proches entre eux.
De plus, ces bâtiments sont très peu élevés par rapport à ceux des villes précédemment étudiées puisque leur hauteur n'excède pas 5 m.
Or, la méthode développée, basée sur la notion de cohérence locale traduite par l'utilisation d'une régularisation spatiale, a tendance à lisser les variations de labels hautes fréquences. Dans le cas du suivi urbain, cette caractéristique permet de supprimer les fines variations de labels qui représentent alors généralement des fausses alarmes puisque la grande majorité de l'image est labellisée non–changement.
Dans le cas de Sendai, au contraire, la majorité de l'image est labellisée en changement. Ainsi, à l'intérieur de la zone détruite, ce sont les pixels labellisés non–changements qui apparaissent comme des variations fines de labels et qui auront donc tendance à être supprimés.
Ce phénomène tend à limiter la présence de fausses alarmes de changements sous forme d'amas de pixels isolés et à maintenir un taux de bonnes détections très haut.

Ainsi, la méthode développée ne permet pas de discriminer chacun des bâtiments détruits dans la zone totalement sinistrée mais elle va plutôt délimiter cette zone sinistrée. Toutefois, la méthode peut s'avérer très intéressante pour localiser les bâtiments détruits isolés et plus difficiles à retrouver visuellement car noyés au milieu des très nombreux changements radiométriques.
La courbe Kappa présente une forte augmentation partant d'une régularisation $\lambda = 1,0$ à $\lambda = 3,5$, augmentation due à la baisse du nombre de fausses alarmes et au maintien des bonnes détections à un taux élevé. Puis, entre $\lambda = 3,5$ et $\lambda = 5,0$, le kappa montre une diminution qui s'explique par un palier et même une légère augmentation du nombre de fausses alarmes (phénomène qui apparait lorsque la régularisation va diviser une alarme importante de changement en deux alarmes distinctes) et la baisse du taux de bonnes détections.

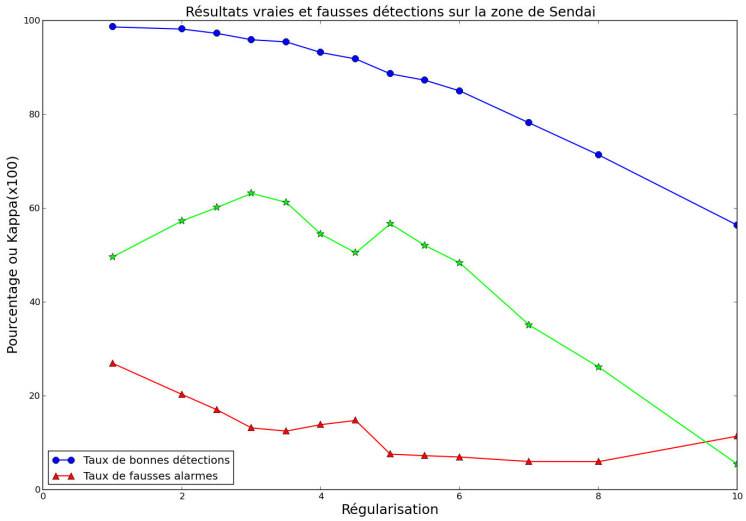



FIGURE 5.12 – *Courbes des taux de bonnes détections, fausses alarmes et coefficient Kappa.*

Enfin, lorsque la régularisation atteint $\lambda = 5,0$, le Kappa baisse de façon quasi linéaire en suivant la tendance du taux de bonnes détections.

Finalement, les courbes présentées montrent de nouveau une régularisation optimale proche de $\lambda = 3,0$ lorsque des changements de toutes les tailles (mais majoritairement inférieurs à 100 $m^2$) sont recherchés, même dans un contexte aussi particulier que celui d'une destruction massive du bâti et surtout avec des images provenant d'un capteur très haute résolution différent de celui utilisé pour les zones expérimentales précédentes.

La figure 5.13 présente la carte des changements obtenue sur une emprise importante de la région de Sendai. Cette carte a été obtenue avec une régularisation $\lambda = 3,0$, régularisation optimale définie précédemment. L'ampleur de la destruction est particulièrement visible sur cette carte qui laisse pourtant apparaître des zones en changements positifs. Ces zones ne sont pas des erreurs mais correspondent en fait à des amas, très probablement de débris, apparus entre les acquisitions avant et après la catastrophe.
Nous rappelons que l'image post–évènement a été acquise 5 mois après la catastrophe. La zone détruite a donc été en grande partie déblayée durant ces 5 mois. Cet intervalle d'acquisition peut alors représenter un inconvénient pour la mesure de la performance de notre méthode, le résultat de la carte de changement serait peut être différent si la zone n'avait pas été déblayée.

(a) Ortho image Ikonos de 2010.

(b) Ortho image Ikonos de 2011.

(c) Carte des changements d'élévation obtenue avec $\lambda = 3$.

FIGURE 5.13 – *Images ortho rectifiées de la région de Tohoku et carte des changements résultante.*

Cette carte de changements met aussi clairement en évidence les destructions apparues dans la végétation, notamment sur la partie sud de l'image où une forêt entière a été dévastée par le tsunami. Bien que les changements sur la végétation soient généralement comptés en fausses alarmes, ceux–ci apparaissent pertinents puisque les dégâts environnementaux peuvent aussi avoir un fort impact économique pour une région et nécessiter une évaluation.

## 5.5 Conclusion

Ce chapitre a permis de tester les performances de la méthode développée pour la détection des changements d'élévation ainsi que sa sensibilité envers les paramètres utilisés ou les données d'entrée.

Ces évaluations ont été réalisées grâce à la comparaison des cartes de changements obtenues avec des cartes de références générées manuellement (ou à l'aide d'un LiDAR) et à travers l'utilisation de plusieurs métriques classiques de la littérature : les taux de bonnes détections et de fausses alarmes ainsi que le coefficient Kappa.

Comme attendu, la régularisation est le paramètre ayant la plus forte influence sur le résultat. Les différents cas test présentés (Phoenix, Christchurch et Sendai) ont permis de montrer que lorsque des changements de toutes tailles sont recherchés, la régularisation optimale se situe à environ $\lambda = 3,0$ et permet d'obtenir 100% des changements détectés pour 70% à 80% de fausses alarmes.

Lorsque seuls les changements supérieurs à 100 $m^2$ sont ciblés, une régularisation $\lambda = 5,0$ est recommandé et permet d'obtenir plus de 80% de bonnes détections et moins de 20% de fausses alarmes.

Nous avons aussi pu observer que la qualité du MNS avait un impact important sur la qualité du résultat, notamment à faible régularisation lorsque les changements de petite taille sont considérés et que le taux de fausses alarmes générées par les erreurs des MNS est élevé.

Enfin, dans le cas d'une scène présentant des changements très important dus à une catastrophe majeure, nous avons pu montrer les performances de la méthode en termes de bonnes détections et de fausses alarmes tout en confirmant une régularisation optimale à $\lambda = 3,0$ lorsque les changements de toutes les tailles sont recherchés. Si ces analyses permettent déjà une bonne évaluation des performances de la méthode envers différents paramètres, elles pourraient

cependant être complétées par des analyses portant sur l'impact de la résolution native et la qualité (contraste et piqué) des images d'entrée sur la taille des changements détectables et leur taux de détection.

# Chapitre 6

# Conclusion générale

## 6.1 Conclusion

Nous présentons, dans ce travail de thèse, une chaîne de traitements automatique, non supervisée et efficace pour la recherche des changements d'élévation apparus entre deux dates d'intérêt. Cette chaîne est basée sur la génération de MNS à chaque date à partir de couples d'images satellites stéréoscopiques très haute résolution puis sur l'analyse de leur différences.
L'un des avantages de cette chaîne est d'être totalement automatique, c'est–à–dire qu'aucune information externe n'est nécessaire. Cet avantage permet de garantir l'efficacité de la méthode sur n'importe quelle zone du globe, pourvu que des couples stéréoscopiques soient présents aux dates d'intérêt.

Nous avons montré au cours de cette thèse que l'un des éléments les plus sensibles de la méthode est la qualité du MNS, c'est pourquoi une grande partie de ce travail est dédiée à l'amélioration de la précision des MNS produits par imagerie spatiale.
Une technique innovante, basée sur la fusion de tous les MNS asymétriques réalisables à partir d'un même couple stéréoscopique a été développée dans cet objectif. Cette méthode s'est avérée particulièrement efficace pour diminuer les erreurs de corrélations dans le MNS final, qu'elles proviennent de la mise en correspondance sur des objets difficiles de la scène ou des zones d'occlusion de l'une ou l'autre image du couple. Enfin, nous avons montré que cette méthode était aussi avantageuse pour traiter la multi–stéréoscopie.

La méthode de détection des changements d'élévation développée dans la chaîne de traitements est basée sur une classification optimisée de tous les pixels du MNS différentiel, classification qui dépend de la valeur de différence d'élévation des pixels mais aussi d'une contrainte spatiale permettant de limiter les variations de labels affectés à des pixels connexes. Cette méthode de classification permet ainsi de mettre en avant les changements cohérents de la scène tout en supprimant le bruit de corrélation tandis que l'équilibre fausses alarmes et bonnes détections est alors contrôlé par le paramètre de régularisation $\lambda$.
Cette méthode a été testée sur plusieurs zones différentes et dans un contexte de recherche des changements urbains typiques d'une ville. Ces zones ont été sélectionnées dans les villes de Phoenix (Arizona, États-Unis) et de Christchurch (Nouvelle-Zélande).

Les résultats montrent que, lorsque un taux de détection de plus de 80% est attendu et pour des changements de taille supérieure à 15×15 pixels, les taux de fausses alarmes varient alors de 20% à 35% seulement et avec une régularisation optimale définie à $\lambda = 5,0$.
Lorsque des changements de toutes tailles sont recherchés, la régularisation optimale est alors abaissée à environ $\lambda = 3,0$ et permet la détection de 60% à 80% des changements avec de 40% à 60% de fausses alarmes.
Ces résultats sont cependant particulièrement sensibles aux sources d'erreurs de corrélation présentes dans la scène.
Dans un contexte plus particulier de destruction urbaine importante, après une catastrophe majeure, nous montrons que les mêmes régularisations peuvent s'appliquer et entraînent un taux de plus de 95% de bonnes détections et moins de 10% de fausses alarmes, valeurs obtenues pour la région de Tohoku, touchée par un tsunami le 11 mars 2011.

La chaîne de traitement développée nécessite cependant l'utilisation de nombreux paramètres, que ce soit pour la génération des MNS ou la détection des changements. On montre pourtant que, quel que soit le contexte ou le capteur très haute résolution (ville de Phoenix en image WorldView-1 ou région détruite de Sendai en images Ikonos) les même paramètres peuvent être utilisés. Nous montrons ainsi que seule la résolution du capteur ou le contexte de la zone d'étude (urbaine ou rurale vallonnée) doivent être pris en compte pour la génération des MNS tandis que la résolution des images, la taille des changements ou la précision recherchée

sont les seules véritables contraintes du système de détection de changements.

Enfin, le système fonctionne sur un seul et même “noyau” algorithmique, noyau que nous avons réalisé de façon à le rendre adaptable à de nombreux problèmes de labellisation tandis que les problèmes rencontrés dans la chaîne de traitements sont eux–mêmes tous formulés en termes de labellisation des pixels d'une image selon des contraintes de probabilité et une contrainte de régularisation spatiale.

## 6.2 Perspectives

Nous pouvons mettre en évidence trois perspectives d'évolution à fort intérêt pour la chaîne de traitements développée.

La première consiste en la création d'un masque de végétation. En effet, nous avons montré, lors des étapes de génération de MNS, que la mise en correspondance avait une forte tendance à produire des erreurs sur des textures 3D aussi complexes que la végétation. Dans l'étape de basculement des MNS, la végétation est partiellement supprimée grâce aux contraintes sur le score de corrélation et dans l'étape suivante de fusion des MNS, d'autres erreurs dues à la végétation peuvent être éliminées grâce à la régularisation sur le MNS.

Mais ces erreurs restent néanmoins une des sources principales de fausses alarmes dans la détection des changements, notamment lorsque les couples ou n-uplets stéréoscopiques n'ont pas été acquis à la même saison et que la végétation est alors plus ou moins développée entre les images.

Ce phénomène a notamment été illustré sur les cartes de changements obtenues sur la zone de Christchurch. Cependant, nous avons aussi pu montrer les dégâts importants sur la végétation après la catastrophe de Sendai grâce aux changements détectés sur la végétation. Les changements apparus sur la végétation peuvent ainsi, selon le contexte ou l'application demandée, être recherchés ou être considérés en fausses alarmes.

C'est pourquoi une perspective intéressante à ce travail est la création de masques de végétation afin de diminuer les fausses alarmes générées ou bien afin de mettre en évidence les alarmes de changements que la végétation produit.

Afin d'éviter les cas où une zone de végétation serait transformée en zone bâtie ou inversement, il serait alors nécessaire de générer des masques de végétation correspondant à chacune des dates d'étude.
Deux solutions pourraient ensuite être envisagées pour la prise en compte de ces masques dans la détection des changements. Une première approche serait le filtrage post–traitement des alarmes de changements lorsque qu'une alarme correspond, à chacune des dates, à une zone de végétation. Une seconde approche, certainement plus précise, consisterait à considérer la végétation comme un label à part entière lors de la classification des pixels.

Une autre perspective d'amélioration de la précision des résultats peut être représentée par le couplage de la méthode de détection des changements d'élévation, telle que nous l'avons conçue, avec une méthode de détection des changements radiométriques.
Nous avons précisé, au début de ce travail, que les méthodes de détection des changements radiométriques étaient très sensibles aux conditions d'acquisitions des images et détectaient ainsi des changements d'illumination non pertinents pour notre contexte. Cependant, nous avons aussi montré que notre méthode peut aussi générer des fausses alarmes mais sur des zones trop homogènes ou trop périodiques qui sont une gêne pour la mise en correspondance. Ainsi, le filtrage, par exemple, des alarmes de changements 3D par une méthode de détection des changements radiométriques pourrait s'avérer très avantageux pour la précision des résultats finaux.

Enfin, une dernière perspective à ce travail est la génération de MNS à partir d'images satellites monososcopiques, n'ayant pas été acquises simultanément. En effet, la chaîne de traitements développée est basée sur l'utilisation de couple ou uplets stéréoscopiques nécessaires pour les deux dates $t_1$ et $t_2$. Cette condition, requise pour tout traitement, n'est pourtant pas toujours satisfaite sur toutes les régions du globe car si les satellites aujourd'hui font preuve d'une grande réactivité et peuvent acquérir des images stéréoscopiques dans un court délai et n'importe où, les archives stéréoscopiques ne sont, elles, pas toujours disponibles.
Une piste importante consisterait donc à développer une méthode capable de générer des MNS à partir de nombreuses images monoscopiques acquises dans les mois qui entourent la date d'intérêt $t_1$ et avec n'importe quel capteur très haute résolution. La mise en correspondance de

toutes les images nécessiterait alors la prise en compte de leurs différences qu'elles soient dues à l'acquisition des images comme les différences d'illumination ou aux changements survenus entre les acquisitions.

Cette piste présente ainsi un fort intérêt dans de nombreux domaines nécessitant la création de MNS.

# Table des figures


1.1 *Schéma de principe des différentes géométries de mise en correspondance.* . . . . 34

1.2 *Illustration des différentes étapes de la chaîne globale de traitements.* . . . . . 36

2.1 *Illustration du système d'acquisition en barrette push–broom et des angles d'attitude du satellite.* . . . . . 43

2.2 *Illustration du résultat de la spatio–triangulation pour le recalage des images.* . . 44

2.3 *Illustration du résultat de la spatio–triangulation pour le recalage des images (échelle en pixels).* . . . . . 46

2.4 *Illustration du calcul des résidus de la spatio–triangulation.* . . . . . 46

2.5 *Illustration de la méthode des graph–cuts appliquée à une image.* . . . . . 50

2.6 *Illustration d'un effet de streaking le long des lignes d'une image traitée par programmation dynamique.* . . . . . 53

2.7 *Illustration du parcours des différentes directions pour la programmation dynamique afin de calculer le coût "semi–global".* . . . . . 55

2.8 *Illustration du calcul des coûts obtenus pour chaque label et chaque pixel de la séquence selon l'équation 2.5.* . . . . . 57

3.1 *Illustration de l'impact de la régularisation pour la génération de MNS.* . . . . . 64

3.2 *Principe de parcours de la grille image maître pour l'interpolation des points de la grille régulière.* . . . . . 67

3.3 *Illustration du principe de l'interpolation d'une grille régulière terrain à partir des points obtenus dans la géométrie de l'image maître.* . . . . . 69

3.4 *Illustration du résultat du basculement d'un MNS WorldView-1 à partir de la carte des élévations en géométrie image.* . . . . . 70

3.5 *MNS asymétriques provenant du même couple d'images stéréoscopiques.* . . . . . 73
3.6 *MNS provenant du même couple stéréoscopique et résultat de leur fusion par la moyenne.* . . . . . 75
3.7 *Schéma récapitulatif de la procédure de fusion des MNS en fonction des valeurs initiales et de l'attache aux données.* . . . . . 80
3.8 *3.8(a) et 3.8(b) : MNS calculés à partir du même couple stéréoscopique et résultats de leur fusion par optimisation globale 3.8(c).* . . . . . 81
3.9 *3.9(a) et 3.9(b) : Histogrammes des cartes de différences calculées entre le liDAR et le MNS asymétrique et MNS fusionné, respectivement.* . . . . . 85
3.10 *Profils d'élévation comparés entre le MNS asymétrique (en vert), le MNS fusionné (en bleu) et le Lidar (en rouge).* . . . . . 87
3.11 *Illustration du triplet stéréoscopique Pléiades acquis sur la ville de Toulouse.* . . 89
3.12 *MNS calculés à partir de 2 images puis 3 images du même uplet stéréoscopique.* . 90

4.1 *4.1(a) et 4.1(b) MNS de 2008 et 2011 calculés sur la zone 1 définie sur la ville de Phoenix.* . . . . . 95
4.2 *4.2(a) MNS différentiel. 4.2(b) MNS différentiel seuillé.* . . . . . 96
4.3 *Fonction sigmoïdale permettant le calcul des coûts d'attache aux données pour le label changement positif.* . . . . . 99

5.1 *Orthoimages provenant des images panchromatiques WorldView–1 représentant les trois zones testées sur la ville de Phoenix et la zone de Christchurch.* . . . . . 103
5.2 *5.2(a)Orthoimage calculée et 5.2(b) donnée de référence des changements de la zone 1 de Phoenix.* . . . . . 104
5.3 *Ortho images et MNS calculés avant et après la catastrophe de Sendai du 11/03/2011.* . . . . . 106
5.4 *Zone partiellement détruite par le Tsunami et quantifiée à travers la carte de référence des bâtiments détruits et intacts.* . . . . . 108
5.5 *Cartes de détection de changement obtenues à partir d'un nombre variable de directions de calcul et une régularisation* $\lambda = 3,0$. . . . . 114
5.6 *Illustration de l'impact du nombre de directions de calcul sur les taux de bonnes et fausses détections.* . . . . . 115

5.7 *Cartes des détections calculée avec différentes valeurs de λ et 12 directions.* . . . 116

5.8 *5.8(a) et 5.8(c) : Variations des taux de fausses et vraies détections en fonction de la régularisation λ. 5.8(b) et 5.8(d) : Variation du coefficient Kappa en fonction de la régularisation λ.* . . . . . 119

5.9 *5.9(a) et 5.9(c) : Variations des taux de fausses et vraies détections en fonction de la régularisation λ ou de la taille de l'élément structurant. 5.9(b) et 5.9(d) : Variations du coefficient Kappa en fonction de la régularisation λ ou de la taille de l'élément structurant.* . . . . . 122

5.10 *5.10(a), 5.10(c) et 5.10(e) Variations des taux de bonnes et fausses détections en fonction du λ ; 5.10(b), 5.10(d) et 5.10(f) Variations du coefficient Kappa pour chacune des zones test de Phoenix.* . . . . . 125

5.11 *5.11(a) : Variations des taux de bonnes et fausses détections en fonction du λ. 5.11(b) : Variation du coefficient Kappa. Résultats calculés pour la zone de Christchurch.* . . . . . 127

5.12 *Courbes des taux de bonnes détections, fausses alarmes et coefficient Kappa.* . . . 129

5.13 *Images ortho rectifiées de la région de Tohoku et carte des changements résultante.* 130

# Liste des tableaux

1.1 *Métadonnées des images utilisées lors des différentes études.* . . . . . 38

3.1 *Tableau de synthèse des paramètres de la méthode complète de génération des MNS.* 92

4.1 *Tableau de synthèse des paramètres nécessaires à la méthode de détection des changements d'élévation.* . . . . . 100

# Bibliographie


[1] E. Berthier, Y. Arnaud, R. Kumar, S. Ahmad, P. Wagnon, and P. Chevallier. Remote sensing estimates of glacier mass balances in the Himachal Pradesh (Western Himalaya, India). In *Remote Sensing of Environment*, vol. 108 no. 3 pp. 327–338, 2007.

[2] P. Baldi, M. Fabris, M. Marsella, and R. Monticelli. Monitoring the morphological evolution of the sciara del fuoco during the 2002 – 2003 Stromboli eruption using multi–temporal photogrammetry. In *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 59 no. 4 pp. 199–211, 2005.

[3] M. Koga and A. Iwasaki. Improving the measurement accuracy of three-dimensional topography changes using optical satellite stereo image data. In *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49 no. 8 pp. 2918–2923, Aug. 2011.

[4] K. Tsutsui, S. Rokugawa, H. Nakagawa, S. Miyazaki, C. Cheng, T. Shiraishi, and Shiun-Der Yang. Detection and volume estimation of large-scale landslides based on elevation–change analysis using DEMs extracted from high–resolution satellite stereo imagery. In *IEEE Transactions on Geoscience and Remote Sensing* , vol. 45 no. 6 pp. 1681–1696, Jun. 2007.

[5] N. Champion, D. Boldo, M. Pierrot-Deseilligny, and G. Stamon. 2D Change detection from satellite imagery : performance analysis and impact of the spatial resolution of input images. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1421–1424, Jul. 2011.

[6] P. Reinartz, J. Tian, and A.A. Nielsen. Building damage assessment after the earthquake in Haiti using two post–event satellite stereo imagery and dsms. In *Joint Urban Remote Sensing Event (JURSE)*, pp. 057–060, 2013.

[7] R.J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms : a systematic survey. In *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 294–307, Mar. 2005.

[8] J. Tian, H. Chaabouni-Chouayakh, P. Reinartz, T. Krauss, and P. d Angelo. Automatic 3D change detection based on optical satellite stereo imagery. In *ISPRS TC VII Symposium, Vienna*, 2010.

[9] P. J. Deer. Digital change detection techniques : Civilian and military applications. In *International Symposium on Spectral Sensing Research*, 1995.

[10] A. Singh. Review article digital change detection techniques using remotely–sensed data. In *International Journal of Remote Sensing*, vol. 10, no. 6, pp. 989–1003, 1989.

[11] M. Volpi, D. Tuia, M. Kanevski, F. Bovolo, and L. Bruzzone. Supervised change detection in VHR images : a comparative analysis. In *IEEE International Workshop on Machine Learning for Signal Processing*, pp. 1–6, Sept. 2009.

[12] L. Bruzzone and F. Bovolo. A conceptual framework for change detection in very high resolution remote sensing images. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 2555–2558, Jul. 2010.

[13] P.L. Rosin and E. Ioannidis. Evaluation of global image thresholding for change detection. In *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2345–2356, 2003.

[14] F. Bovolo, L. Bruzzone, and M. Marconcini. A novel approach to unsupervised change detection based on a semi–supervised SVM and similarity measure. In *IEEE Transaction on Geoscience and Remote Sensing Letters*, vol. 46, no. 7, pp. 2070–2082, 2008.

[15] D.H.A. Al-Khudhairy, I. Caravaggi, and S. Clada. Structural damage assessments from Ikonos data using change detection, object–oriented segmentation and classification techniques. In *Photogrammetric Engineering and Remote Sensing*, vol. 71, pp. 825–837, 2005.

[16] L. Bruzzone and D.F. Prieto. Automatic analysis of the difference image for unsupervised change detection. In *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[17] T.T. Vu, M. Matsuoka, and F. Yamazaki. Lidar–based change detection of buildings in dense urban areas. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 5, pp. 3413–3416, Sep. 2004.

[18] G. Sithole and G. Vosselman. Automatic structure detection in a point–cloud of an urban landscape. In *2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, pp 67–71, May 2003.

[19] F. Rottensteiner. Automated updating of building data bases from digital surface models and multi–spectral images : Potential and limitations. In *ISPRS Congress, Beijing, China*, vol. 37, pp. 265–270, Dec. 2008.

[20] G. Vosselman, B. G. H. Gorte, and G. Sithole. Change detection for updating medium scale maps using laser altimetry. In *Proceedings of the 20th ISPRS Congress*, pp. 12–23, 2004.

[21] L. Matikainen, J. Hyyppa, E. Ahokas, L. Markelin, and H. Kaartinen. Automatic detection of buildings and changes in buildings for updating of maps. In *Remote Sensing*, vol. 2, no. 5, pp. 1217–1248, 2010.

[22] H. Chaabouni-Chouayakh, P. d'Angelo, T. Krauss, and P. Reinartz. Automatic urban area monitoring using digital surface models and shape features. In *Joint Urban Remote Sensing Event (JURSE)*, pp. 85–88, Apr. 2011.

[23] F. Jung. Detecting building changes from multitemporal aerial stereopairs. In *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, no. 4, pp. 187–201, 2004.

[24] J. Tian, H. Chaabouni-Chouayakh, and P. Reinartz. 3D Building change detection from high resolution spaceborne stereo imagery. In *International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM)*, pp. 1–7, 2011.

[25] J. A. Marchant and C. M. Onyango. Shadow-invariant classification for scenes illuminated by daylight. In *Journal of the Optical Society of America A*, vol. 17, no. 11, pp. 1952–1961, Nov. 2000.

[26] J. Tian, S. Cui, and P. Reinartz. Building change detection based on satellite stereo imagery and digital surface models. In *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–12, 2013.

[27] A.A. Nielsen. The regularized iteratively reweighted mad method for change detection in multi- and hyperspectral data. In *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 463–478, 2007.

[28] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two–frame stereo correspondence algorithms. In *International Journal of Computer Vision*, vol. 47, pp. 7–42, 2002.

[29] M.Z. Brown, D. Burschka, and G.D. Hager. Advances in computational stereo. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993–1008, 2003.

[30] N. Lazaros, G.C. Sirakoulis, and A. Gasteratos. Review of stereo vision algorithms : from software to hardware. In *International Journal of Optomechatronics*, vol. 2, no. 4, pp. 435–462, 2008.

[31] H. Hirschmuller. Accurate and efficient stereo processing by semi–global matching and mutual information. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 807–814, Jun. 2005.

[32] M. Pierrot-Deseilligny and N. Paparoditis. A multiresolution and optimization–based image matching approach : an application to surface reconstruction from SPOT5-HRS stereo imagery. In *Internationales Archives of Photogrammetry and Remote Sensing*, 2006.

[33] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister. Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 492–504, 2009.

[34] A.F. Bobick and S.S. Intille. Large occlusion stereo. In *International Journal of Computer Vision*, pp. 181–200, 1999.

[35] C. Baillard and A. Zisserman. Automatic reconstruction of piecewise planar models from multiple views. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 559–65, Jun. 1999.

[36] L. Zhang and A. Gruen. Multi–image matching for dsm generation from ikonos imagery. In *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 60, no. 3, pp. 195–211, 2006.

[37] R. Zabih and J. Woodfill. Non–parametric local transforms for computing visual correspondence. In *Computer Vision ECCV'94*, pp. 151–158, 1994.

[38] A. Klaus, M. Sormann, and K. Karner. Segment-based stereo matching using belief propagation and a self–adapting dissimilarity measure. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CPVR)*, vol. 2, Jun. 2006.

[39] J. Sun, Y. Li, S.B. Kang, and H.Y. Shum. Symmetric stereo matching for occlusion handling. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 399–406, 2005.

[40] M. Sizintsev and R.P. Wildes. Coarse-to-fine stereo vision with accurate 3d boundaries. In *Image and Vision Computing*, vol. 28, no. 3, pp. 352–366, 2010.

[41] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[42] K Schindler. An overview and comparison of smooth labeling methods for land–cover classification. In *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4534–4545, Nov. 2012.

[43] P.F. Felzenszwalb and D.R. Huttenlocher. Efficient belief propagation for early vision. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 261–268, Jul. 2004.

[44] O. Veksler. Stereo correspondence by dynamic programming on a tree. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CPVR)*, vol. 2, pp. 384–390, Jun. 2005.

[45] H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, Jun. 2007.

[46] S. Birchfield and C. Tomasi. Depth discontinuities by pixel–to–pixel stereo. In *International Journal of Computer Vision*, vol. 35, pp. 269–293, 1999.

[47] H. Hirschmuller, P. R Innocent, and J. Garibaldi. Real–time correlation based stereo vision with reduced border errors. In *International Journal of Computer Vision*, vol. 47, no. 3, pp. 229–246, 2002.

[48] M. Pierrot-Deseilligny L. Penard, N. Paparoditis. 3d building facade reconstruction under mesh form from multiple wide angle views. In *Proceedings of the 20th ISPRS Congress*, 2005.

[49] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. In *Machine Vision and Applications*, vol. 6, pp. 35–49, 1993.

[50] G. Egnal and R.P. Wildes. Detecting binocular half-occlusions : empirical comparisons of five approaches. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1127–1133, 2002.

[51] S. Chambon and A. Crouzil. Towards correlation-based matching algorithms that are robust near occlusions. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, vol. 3, pp. 20–23, 2004.

[52] P. N. Belhumeur. A bayesian approach to binocular steropsis. In *International Journal of Computer Vision*, vol. 19, no. 3, pp. 237–260, 1996.

[53] Y. Furukawa and J. Ponce. Accurate dense and robust multi–view stereopsis. In *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1362–1376, 2010.

[54] F De Lussy, D. Greslou, C Dechoz, V. Amberg, J. M. Delvit, L. Lebegue, G. Blanchet, and S. Fourest. Pleiades HR in flight geometrical calibration : location and mapping of the focal plane. In *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 39, pp. 519–523, 2012.

[55] J. Grodecki and G. Dial. Block adjustment of high-resolution satellite images by rational polynomials. In *Photogrammetric Engineering and Remote Sensing*, vol. 69, pp. 59–68, 2003.

[56] C.S. Fraser, G. Dial, and J. Grodecki. Sensor orientation via RPCs. In *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 60, pp. 182–194, 2006.

[57] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[58] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph–cuts. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[59] S. Roy and I.J. Cox. A maximum–flow formulation of the n-camera stereo correspondence problem. In *Sixth International Conference on Computer Vision*, pp. 492–499, Jan. 1998.

[60] Y. Weiss and W.T. Freeman. On the optimaly of solutions of the max-product belief propagation algorithm in arbitrary graphs. In *IEEE Transactions Information Theory*, vol. 47, pp. 723–735, 2001.

[61] Daniel Scharstein and Richard Szeliski. vision.middlebury, 2012.

[62] L. Randolph Ford and D.R. Fulkerson. *A simple algorithm for finding maximal network flows and an application to the Hitchcock problem*. Rand Corporation, 1955.

[63] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph–cuts. In *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 26, pp. 147–159, 2004.

[64] R. Szeliski, R. Zabih, D. Scharstein, O Veksler, V Kolmogorov, A. Agarwala, and C. Rother. A comparative study of energy minimization methods for markov random fields. In *Computer Vision-ECCV*, pp. 16–29, 2006.

[65] V. Kolmogorov. Convergent tree–reweighted message passing for energy minimization. In *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1568–1583, 2006.

[66] P.F. Felzenszwalb and R. Zabih. Dynamic programming and graph algorithms in computer vision. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 721–740, Apr. 2011.

[67] Y. Otha and T. Kanade. Stereo by intra- and inter–scanline search using dynamic programming. In *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 2, pp. 139–154, 1985.

[68] R. B. Potts. Some generalized order–disorder transformations. *Cambridge Philos. Soc. Math. Proc.*, vol. 48, pp. 106–109, 1952.

[69] A. Fahmi, J. Jansa, C. Ries, et al. The generation of true orthophotos using a 3D building model in conjunction with conventional DTM. In *International Archives of Photogrammetry and Remote Sensing*, vol. 32, 1998.

[70] G. Singh, Y. Yamaguchi, W.-M. Boerner, and Sang-Eun Park. Monitoring of the March 11, 2011, off-tohoku 9.0 earthquake with super–tsunami disaster by implementing fully polarimetric high–resolution polsar techniques. In *Proceedings of the IEEE*, vol. 101, no. 3, pp. 831–846, 2013.

[71] J. Cohen. A coefficient of agreement for nominal scales. In *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[72] G.G. Wilkinson. Results and implications of a study of fifteen years of satellite image classification experiments. In *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 433–440, Mar. 2005.

[73] R.M. Haralick, S.R. Sternberg, and X. Zhuang. Image analysis using mathematical morphology. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 532–550, 1987.