HAL
open science

Modèles a contrario pour la détection générique de changements en imagerie satellitaire très haute résolution

Vincent Vidal

▶ **To cite this version:**

Vincent Vidal. Modèles a contrario pour la détection générique de changements en imagerie satellitaire très haute résolution. Statistiques [math.ST]. Université Paris Cité, 2020. Français. NNT : 2020UNIP7208 . tel-03354090

HAL Id: tel-03354090

https://theses.hal.science/tel-03354090

Submitted on 24 Sep 2021

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.

Université de Paris

École doctorale 386 : Sciences Mathématiques de Paris Centre

Laboratoire MAP5 (CNRS UMR 8145)

Modèles a contrario pour la détection générique de changements en imagerie satellitaire très haute résolution

par Vincent Vidal

Dirigée par

Lionel Moisan

Thèse de doctorat en Mathématiques Appliquées

Soutenance prévue le 20 novembre 2020 devant un jury composé de

Andrés Almansa	*Directeur de recherche CNRS*, Université de Paris	Président du Jury
Tugdual Ceillier	*Docteur*, Earthcube	Examinateur
Agnès Desolneux	*Directrice de recherche CNRS*, ENS Paris-Saclay	Examinateur
Yann Gousseau	*Professeur*, Telecom Paris	Rapporteur
Jordi Inglada	*Ingénieur de recherche CNES*, UMR CESBIO	Rapporteur
Sylvie Le Hégarat-Mascle	*Professeur*, Université Paris-Saclay	Examinateur
Lionel Moisan	*Professeur*, Université de Paris	Directeur de thèse

Modèles a contrario pour la détection générique de changements en imagerie satellitaire très haute résolution

Résumé : Dans cette thèse, nous proposons plusieurs méthodes de détection de changements génériques, adaptées à l'aide à la photo-interprétation. Nous introduisons différents modèles de changements, dit « primaires », présentant chacun une approche différente de la notion de changement et reposant sur le cadre statistique dit « a contrario », qui permet un contrôle du nombre moyen de fausses alarmes. Pour le premier modèle, qui soustrait directement des descripteurs calculés ponctuellement sur chacune des images, nous proposons une estimation du nombre de fausses alarmes en présence de bruit gaussien, qui s'avère exacte dans le cas de descripteurs linéaires. Le deuxième modèle primaire repose sur une comparaison d'histogrammes locaux, qui par sa robustesse permet de s'affranchir de la nécessité d'effectuer une mise en correspondance préalable précise entre les deux images à comparer. Le troisième modèle proposé est une extension de l'algorithme de détection d'anomalies de Reed-Xiaoli à un couple d'images et des distributions locales modélisées par des mélanges de gaussiennes. Les deux derniers modèles primaires intègrent une invariance à certains changements globaux, et permettent ainsi la comparaison d'images radiométriquement différentes, problème peu étudié dans la littérature. Ainsi, le modèle par « destin commun » examine la présence d'anomalies locales dans la statistique jointe des deux images à comparer, et permet donc de s'affranchir d'éventuelles transformations systématiques entre les deux images. Le modèle par « reconstruction indirecte », quant à lui, utilise une distance entre patchs pour construire une version de la première image dans le référentiel radiométrique de la seconde, ce qui permet ensuite de les comparer directement. Dans la dernière partie, nous nous intéressons à la construction d'un détecteur de changements spécifiés à partir d'un petit nombre d'exemples. À l'aide d'un algorithme de type *gradient boosting* appliqué à des arbres de régression, nous réalisons l'agrégation de détecteurs obtenus en combinant les modèles primaires précédents avec différents descripteurs. L'intérêt de cette méthode est illustré sur un cas réel de données satellitaires.

Mots clefs : Traitement d'images, Télédétection, détection de changements, cadre a-contrario

---

A contrario models for generic change detection in very high resolution satellite imagery

Abstract: In this thesis, we propose several methods for the detection of generic changes, suitable for photo-interpretation support. We introduce various models of change detection, called "primary", each presenting a different approach to the notion of change and based on the statistical framework called "a contrario", which allows a control of the average number of false alarms. For the first model, which directly subtracts descriptors computed punctually on each image, we propose an estimation of the number of false alarms in the presence of Gaussian noise, which proves to be exact in the case of linear descriptors. The second primary model is based on a comparison of local histograms, which by its robustness avoids the need for a precise prior registration between the two images. The third proposed model is an extension of the Reed-Xiaoli anomaly detection algorithm to a pair of images and local distributions modeled by Gaussian mixtures. The last two primary models integrate an invariance to some global changes, and thus allow the comparison of radiometrically different images, a problem little studied in the literature. In this way, the "common fate" model examines the presence of local anomalies in the joint statistics of the two compared images, and thus makes it possible to free oneself from systematic transformations between the two images. The "indirect reconstruction" model uses a distance between patches to construct a version of the first image in the radiometric frame of reference of the second, which then allows a direct comparison. In the last part, we focus on the construction of a detector of specified changes from a small number of examples. Using a gradient boosting algorithm applied to regression trees, we perform the aggregation of detectors obtained by combining the previous primary models with different descriptors. The interest of this method is illustrated on a real case of satellite data.

Keywords: Image analysis, remote sensing, change detection, a-contrario framework

Table des matières

Chapitre 1 **Introduction** **9**

1.1 L'imagerie satellitaire, contexte et enjeux 9
1.1.1 Contexte historique 9
1.1.2 Détection de changements et photo-interprétation . . 10
1.1.3 Caractéristique des satellites 13

1.2 Détection de changements par mesure directe 19
1.2.1 Définition du problème 19
1.2.2 Modèle de changements par comparaison ponctuelle . 21
1.2.3 Modèle de changements par comparaison d'histogramme 25

1.3 Modèles pour la détection indirecte de changements 28
1.3.1 Modélisation locale du contexte 28
1.3.2 Modèle de changements par destin commun 29
1.3.3 Modèle de changements par reconstruction indirecte . 32

1.4 Détection supervisée de changements 35

1.5 Plan du manuscrit 36

1.6 Publications 38

Chapitre 2 **Réflexions autour de la notion de changement** **39**

2.1 Cause des changements 40
2.1.1 Introduction 40
2.1.2 Source de changements dans la chaîne de traitement . 42
2.1.3 Une séparation causes et effets difficile 45

2.2 Segmentation sémantique pour la détection de changements 47
2.2.1 Changements pertinents 47
2.2.2 Utilisation de l'interprétation sémantique 49
2.2.3 Limitation de l'approche sémantique 49

2.3 L'anormalité comme critère de pertinence 51
2.3.1 Pertinence contextuelle 51
2.3.2 Rareté des changements pertinents 53

2.4 Point de vue adopté 55
2.4.1 Définition de la pertinence par l'utilisateur 55
2.4.2 Hypothèses du problème 56
2.4.3 Images synthétiques utilisées en test 57

Chapitre 3 **Détection de changements par mesure directe** **61**

3.1 Formalisme a contrario 62
3.1.1 Définitions et propriétés 62
3.1.2 Cas pratique 68

3.2 Prétraitement et conventions prises 70
3.2.1 Incohérences spatiales et recalage d'images 70
3.2.2 Variation de contraste et recalage en intensité 74
3.2.3 Conventions générales adoptées 75

3.3 Descripteurs d'images 76
3.3.1 Généralité sur les descripteurs 76
3.3.2 Descripteurs d'intensité 78
3.3.3 Descripteurs géométriques 79
3.3.4 Descripteurs de texture 86

3.4 Modèle de changements par comparaison ponctuelle 88
3.4.1 Introduction 88
3.4.2 Modélisation d'images 89
3.4.3 Résultats et limitations 100
3.4.4 Extensions du modèle 104

3.5 Modèle de changements par comparaison d'histogramme . . 110
3.5.1 Introduction 110
3.5.2 Formalisation a contrario 111
3.5.3 Résultats et limitation du modèle 115

Chapitre 4 **Modèles pour la détection indirecte de changements** **121**

4.1 Modélisation locale du contexte 122
4.1.1 Introduction 122
4.1.2 Explicitation du modèle 123
4.1.3 Résultats et limitation du modèle 130

4.2 Modèle de changements par destin commun 135
4.2.1 Introduction 135
4.2.2 Formalisation du problème 137
4.2.3 Résultats et limitations du modèle 143

4.3 Modèle de changements par reconstruction indirecte 148
4.3.1 Introduction 148
4.3.2 Formalisation du modèle 149
4.3.3 Résultats et limitations du modèle 158

Chapitre 5 **Fusion de détecteurs primaires** **167**

5.1 Retour sur les détecteurs primaires 168
5.1.1 Choix des détecteurs primaires 168
5.1.2 Détections primaires 174

5.2 Agrégation de détecteurs primaires 180
5.2.1 Problème de classification 180

5.2.2 Méthodes de classification 182
5.2.3 Choix pratique 188
5.3 Évaluation des performances d'un détecteur 189
5.3.1 Algorithme de post-processing 189
5.3.2 Mesures de performance 190
5.4 Détecteur agrégé sur un cas réel 197
5.4.1 Description des données 197
5.4.2 Résultats obtenus 198
5.4.3 Étude quantitative 206

CHAPITRE 6 **Conclusion** **211**
6.1 Modèles primaires 211
6.2 Utilisation de l'apprentissage continue 212
6.3 Détection de changements dans des séries temporelles 212

CHAPITRE A **Annexes** **215**
A.1 Propriété des nombres de fausses alarmes 215
A.2 Recalage d'intensités affine 218
A.3 Conditionnement gaussien 218

Introduction 1

1.1 L'imagerie satellitaire, contexte et enjeux

1.1.1 Contexte historique

La télédétection Depuis la fin du XXe siècle, les satellites d'observation de la Terre se sont imposés comme un outil et une source d'informations incontournable et leurs données sont quotidiennement utilisées, que ce soit dans un cadre scientifique, en météorologie, climatologie, océanologie, etc, ou dans un cadre plus pratique, par exemple pour la cartographie, l'aménagement du territoire, le suivi des ouragans, la gestion des ressources naturelles, ou bien entendu pour la sécurité du territoire et l'armée.

Avant d'introduire le problème étudié dans cette thèse, décrivons brièvement les moments historiques clés [Moo79, LMS97, KKH95] qui ont marqué l'avènement de la télédétection, correspondant à l'ensemble des techniques d'observation de la Terre.

Premiers satellites d'observation terrestre La première photographie depuis le ciel est prise en 1858 à bord d'un ballon par Félix Tournachon, dit Nadar. La photographie aérienne se développe alors avec pour principaux objectifs des applications cartographiques et militaires, utilisant pour cela divers instruments comme des ballons, cerfs-volants, avions ou encore des pigeons photographes. Avec son utilisation massive de l'imagerie aérienne, la première guerre mondiale marque les débuts de la « photo-interprétation », correspondant à l'analyse thématique d'une image, dépassant la simple reconnaissance des éléments s'y présentant.

L'intérêt des images aériennes dans un cadre scientifique apparaît dans les années 1930. Le 11 novembre 1935, le ballon stratosphérique « Explorer II » parvient à prendre des images à environ 22 kilomètres d'altitude et la première photo prise depuis l'espace est obtenue au cours du vol suborbital du missile « V-2 No. 13 », le 24 octobre 1946, à une altitude d'environ 105 kilomètres. Si le premier satellite artificiel mis en orbite circulaire autour de la Terre, « Sputnik 1 », est lancé le 4 Octobre 1957, il faut attendre le 1er avril 1960 pour voir la première photographie prise en orbite, obtenue par « TIROS-1 » (Television and Infrared Observation Satellite) (voir figure 1.1). Il s'agit du premier satellite munie d'un capteur optique, entièrement dédié à l'observation de la Terre, et qui, pendant ses 79 jours de fonctionnement, a permis l'obtention de plusieurs dizaines de milliers d'images, marquant alors un tournant dans le domaine de la télédétection.

Entre 1960 et 1970, de nombreux satellites sont lancés par de multiples agences spatiales nationales. Ces satellites, majoritairement météorologiques de très basse résolution (un pixel représentant une surface de l'ordre du kilomètre carré) permettent l'acquisition d'une grande quantité de données, nourrissant la communauté scientifique et permettant, en 1962, l'établissement du premier symposium ERIM (Environmental Research Institute of Michigan) sur le thème de la télédétection appliquée à l'environ-

nement.

Programme Landsat Au début des années 1970, la NASA et l'USGS (United States Geological Survey) initient leur programme « Earth Resources Technology Satellites Program » (renommé plus tard « Landsat Program »), ayant pour vocation la conception et mise en orbite de satellites d'observation de la terre (pris en charge par la NASA) ainsi que la conservation et diffusion libre des données obtenues (pris en charge par l'« United States Geological Survey » (USGS)), entre autres dans la communauté scientifique. Le premier satellite, « Landsat 1 », est lancé en 1972 et marque un tournant dans le monde scientifique, en fournissant gratuitement une couverture multispectrale (rouge, vert, bleu, et infrarouge) régulière (tous les 18 jours) de l'intégralité des terres émergées, et à une « haute résolution » (un pixel représentant une surface d'environ 80×80 mètres carrés). Au total, sept satellites sont mis en orbite entre 1975 et 2013 (les satellites « Landsat 7 » et « Landsat 8 » étant toujours en fonctionnement à ce jour [1]), et fournissent des données servant de support à un grand nombre d'articles scientifiques. Ce programme est toujours maintenu à ce jour, un neuvième satellite étant prévu pour 2021.

Contexte actuel À partir de 1970, de plus en plus de satellites sont mis en orbite, avec une accélération dans les années 1990 avec l'arrivée de plusieurs acteurs commerciaux dans le domaine du spatial (comme Airbus, Digital Globe et GeoEye [2]). La résolution des images (c'est-à-dire la taille de zone correspondant à un pixel) s'améliore au fur et à mesure des années [BS15], passant de 80 m par pixel en 1972 (avec Landat-1), à 10 m en 1987 (avec SPOT 1), puis 6 m en 1995 (avec IRS-1C), 1 m en 1999 (avec IKONOS), 50 cm en 2007 (avec WorldView 1) et récemment 31 cm en 2016 (avec WorldView-4).

Dans les années 2000, il s'opère un changement de point de vue. Il ne s'agit plus d'améliorer la résolution des images, mais d'augmenter le nombre d'images capturées. L'apparition des « nanosatellites », satellites d'une dizaine de centimètres de large fournissant une résolution d'images de l'ordre du mètre, réduit fortement le coût de mise en orbite d'un satellite et démocratise l'univers de l'imagerie spatiale, permettant l'arrivée de nombreuses entreprises et start-up, comme Planet Lab avec des constellations de satellites comme Dove (3 mètres de résolution) et SkySat (80 et 50 centimètres de résolution), et aujourd'hui, plusieurs centaines de satellites orbitent autour de la Terre, capturant quotidiennement l'intégralité du globe.

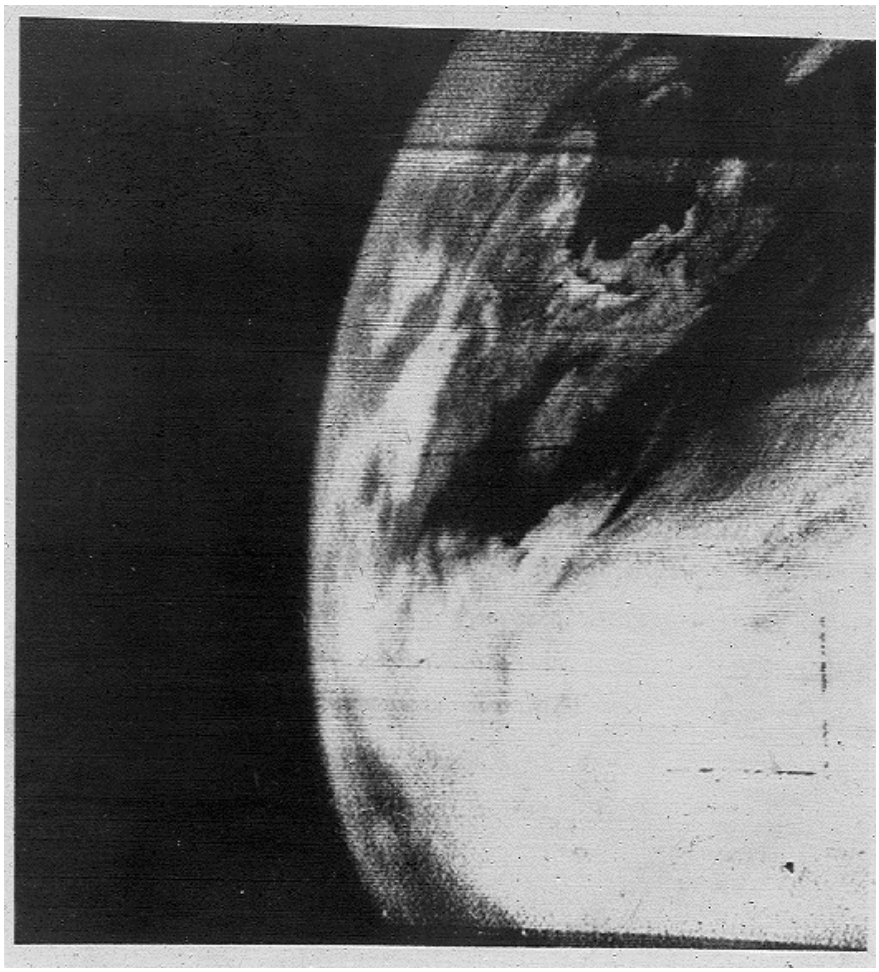
1.1.2 Détection de changements et photo-interprétation

Introduction Au cours des dernières années, avec l'augmentation du nombre de satellites, l'amélioration de la résolution spatiale, qui peut aujourd'hui facilement atteindre le mètre, et la forte diminution du temps d'attente entre deux images d'une même zone géographique, la télédétection apparaît comme une solution à de plus en plus de problèmes de surveillance, de détection et de management. La quantité de

1. La mise en orbite de « Landsat 6 » a échoué à cause d'un problème technique.
2. GeoEye est racheté par Digital Globe en 2013

(a) Image de la Terre, prise à 22 km d'altitude, le 11 novembre 1935 par le ballon stratosphérique Explorer II.

(b) Première image prise depuis l'espace, à 105 km d'altitude, lors du vol suborbital du missile « V-2 No. 13 » le 24 octobre 1946.



(c) La première image satellite, prise le 1er avril 1960 par le satellite TIROS-1, à une altitude d'environ 650 km.

Figure 1.1 – Trois images de la Terre ayant marqué l'univers de la télédétection.

données utilisées augmentant fortement, il devient d'autant plus nécessaire de pouvoir traiter ces données de manière automatique et les algorithmes de traitements d'images adoptent alors une position centrale dans le domaine de la télédétection.

Dans cette thèse, nous nous intéressons au problème de la détection de changements en télédétection, en particulier dans le cas d'images satellites très haute résolution. Ce problème consiste à détecter les changements survenant sur Terre à partir de deux images d'une même zone géographique prises à des instants différents. Il s'agit d'un problème très général, les changements considérés pouvant être multiples et variés. Il peut s'agir, par exemple, d'apparitions de véhicules, du déplacement d'une grue dans un chantier, du changement de couleur d'un champ, de la destruction d'un bâtiment, ou encore de l'élargissement d'un fleuve ou d'une la variation d'une texture au sol. Par ailleurs, les images utilisées peuvent être acquises indépendamment de différentes manières, par satellite, par avion, ou même par drone, et présenter un nombre arbitraire de canaux, de quelques bandes pour une image standard (souvent rouge, vert et bleu), à plus d'une centaine pour une image « hyperspectrale ».

Intérêt de la détection de changements Les problèmes appréhendés par la détection de changements sont multiples. On trouve bien évidemment des applications de surveillance, par exemple d'exploitations agricoles ou d'oléoducs, cherchant à détecter les changements pour pouvoir répondre à d'éventuels problèmes le plus rapidement possible. C'est par exemple le cas de la détection de fuites de pétrole ou de feux de forêt. D'autres applications reposent plus spécifiquement sur la localisation et l'étendue des changements. Il peut s'agir d'évaluer les dommages causés par une catastrophe naturelle (inondation, glissement de terrain, tremblement de terre, etc), de surveiller l'évolution de ressources naturelles (par exemple en s'intéressant à la déforestation).

La détection de changements est aussi utilisée pour envisager des modifications de cartographie, notamment relative à l'aménagement du territoire. Il est alors intéressant de remarquer que ces problèmes n'exigent pas de se ramener à la détection de changements. En effet, il serait possible d'évaluer la couverture du sol (qui correspond à la segmentation du territoire en zone de : forêts, champs, ville, désert, étendue d'eau, etc) à partir d'une seule image. Il est toutefois beaucoup plus rapide de mettre à jour une cartographie déjà existante, en utilisant un algorithme de détection de changements pour connaître l'emplacement des modifications à apporter.

La photo-interprétation La détection de changements apparaît comme particulièrement utile dans le cadre de la photo-interprétation. Celle-ci correspond à l'analyse, centrée autour d'une thématique particulière, d'une image aérienne ou satellitaire. Il ne s'agit pas d'une simple description des objets présents dans la scène observée, mais d'une réelle compréhension de la scène obtenue en recoupant les informations obtenues à partir des images étudiées avec d'autres informations préalables, comme la connaissance topographique et géographique de la zone étudiée, ou encore son contexte social, économique et politique.

Les études de photo-interprétation peuvent être diverses et variées. Il peut s'agir, par exemple, d'évaluer les risques de fractures géologiques dûs à l'installation d'un barrage [GLR+78], de faire un suivi de l'occupation du sol et de cartographie d'espaces

agricoles, naturels ou urbanisés, d'effectuer une étude de la croissance économique d'une ville [HU75], de surveiller les zones d'orpaillage illégal, d'effectuer un suivi du réchauffement climatique, ou encore, dans un cadre militaire, d'apporter des informations stratégiques nécessaires à la gestion des conflits.

Dans le cadre actuel où le nombre d'images satellites ne cesse d'augmenter, il devient particulièrement important d'apporter un moyen d'accélérer le traitement des images. S'il n'est pas envisageable de répondre automatiquement à chaque tâche que doit résoudre un photo-interprète, ces dernières demandant l'utilisation d'informations extérieures aux images, l'extraction automatique de certaines informations permet d'accélérer grandement son travail.

La détection de changements apparaît alors comme particulièrement adapté au cas de surveillance d'une zone sur une période donnée. Si chaque nouvelle image nécessite a priori une nouvelle étude, la connaissance et visualisation des principaux changements peuvent alors permettre de restreindre cette étude aux parties des images qui ont changé et qui présentent donc de nouvelles informations.

Cadre facilité pour la détection de changements Remarquons que le problème de détection de changements est particulièrement adapté à l'imagerie satellitaire. En effet, les images prises sur Terre sont prises au sein d'une scène 3D et subissent alors de nombreux effets d'occlusion, correspondant à un objet disparaissant derrière un autre, et de parallaxe, correspondant à un décalage relatif de deux objets. qui complexifient énormément le problème de détection de changements. Dans le cadre de l'imagerie satellite, la distance de l'appareil optique à la scène est telle que cette dernière apparaît plus comme un bas-relief, limitant grandement les effets évoqués précédemment. Par ailleurs, en considérant deux images prises par un même satellite, la distance à la scène peut être vue comme constante et un même objet apparaît ainsi avec une même échelle dans les deux images.

1.1.3 Caractéristique des satellites

Avant de décrire plus particulièrement les méthodes de détection de changements, présentons brièvement les principales caractéristiques des images satellites. Nous présentons ici tout d'abord les caractéristiques des satellites et de leurs orbites, puis introduisons celles des systèmes optiques et des images satellites.

1.1.3.1 Trajectoire et angle de vue

Orbite, fauchée et temps de revisite L'orbite d'un satellite correspond à la trajectoire que celui-ci effectue autour de la Terre. Les satellites ne disposant pas de réel système de propulsion, leur orbite est donc elliptique et fortement contrainte par la loi de la gravitation et les lois de Kepler. En supposant une trajectoire circulaire (ce qui est souvent le cas), la trajectoire du satellite est alors entièrement déterminée par sa hauteur h (par rapport à la surface de la terre) et par l'angle d'inclinaison orbitale α existant entre le plan de l'équateur et celui de l'orbite. Les autres grandeurs comme sa vitesse v de déplacement et sa période T (correspondant au temps nécessaire pour

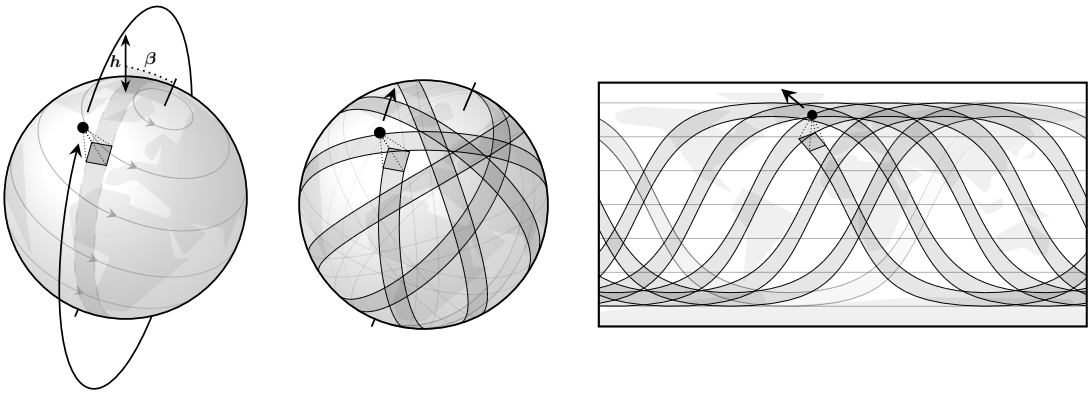


Figure 1.2 – Représentation schématique d'une orbite. On représente ici successivement l'orbite du satellite, puis le domaine capturé au fur et à mesure du temps, le satellite tournant autour de la Terre, cette dernière tournant autour de son axe, et ce sur une représentation 3D de la Terre ainsi que sur une représentation planisphérique. Sur le schéma de gauche, nous avons fait figurer la hauteur h du satellite ainsi qu'un angle $\beta = \alpha - 90°$, où α est l'angle d'incidence orbitale.

effectuer une rotation autour de la Terre) sont alors données par

$$\frac{V^2}{2} = \frac{GM_T}{R_T + h} \qquad \text{et} \qquad T = 2\pi\sqrt{\frac{(R_T + h)^3}{GM_T}}, \tag{1.1}$$

où G est la constante de gravitation et M_T et R_T sont la masse et le rayon de la terre.

La Terre étant en rotation autour de son axe, la trajectoire du satellite projeté sur la surface terrestre n'est plus circulaire et peut recouvrir l'intégralité du globe. Nous avons représenté un exemple de trajectoire dans la figure 1.2.

La période de rotation T n'a alors plus réellement d'intérêt et on définit alors le « cycle orbital » comme la période de cette trajectoire projetée, c'est-à-dire le temps que met le satellite à repasser au-dessus d'un même point. Cette période pourrait correspondre au temps d'attente entre deux images d'une même localisation, si l'on négligeait la fauchée du satellite, qui correspond à la largeur de la zone observée effectivement sur le sol terrestre Cette fauchée étant souvent relativement étendue (d'autant plus que la hauteur est grande), il n'est pas nécessaire que le satellite repasse exactement par une même position pour pouvoir reprendre une image d'une même zone. On peut alors définir une notion de temps de revisite comme la durée entre deux vues successives d'un même point à la surface de la Terre. Ce temps de revisite est un paramètre important, puisqu'il détermine la période d'obtention d'images d'un même site.

Orbites géostationnaires et héliosynchrone Il existe plusieurs cas particuliers d'orbites particulièrement utilisées. La première est l'orbite géostationnaire. Elle concerne les satellites dont le plan de l'orbite correspond au plan de l'équateur et dont la période coïncide avec la période de rotation de la Terre (contraignant sa hauteur à valoir environ 35 800 kilomètres). Les satellites géostationnaires ont alors une position fixe dans le référentiel terrestre, constamment au-dessus d'une même position.

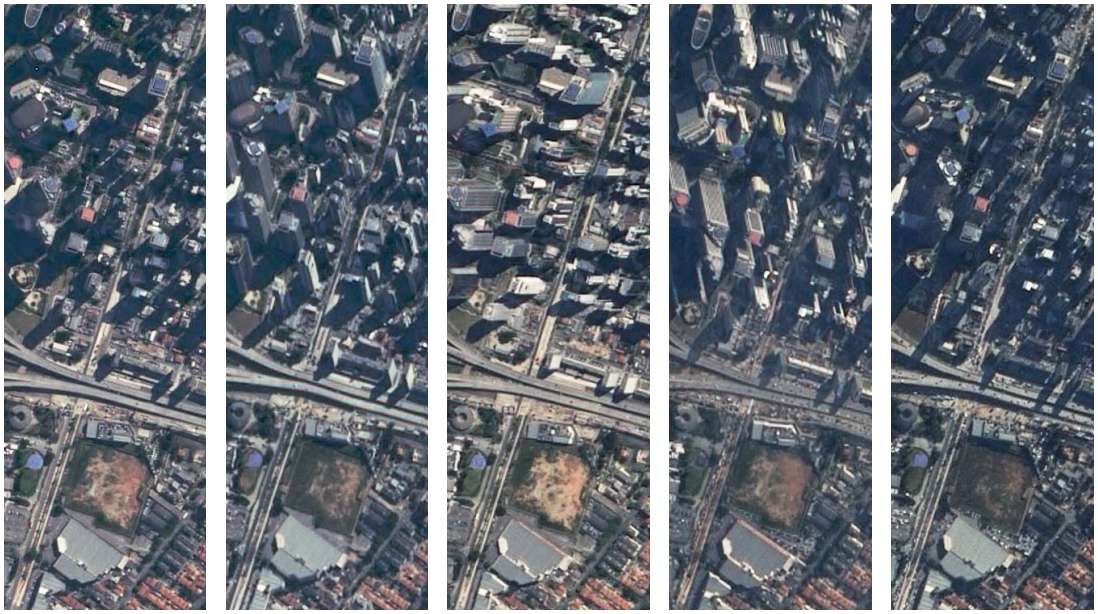


Figure 1.3 – Cinq images satellites présentant un même emplacement sous différents angles incidents. De gauche à droite, l'image est prise au nadir (à la verticale) puis avec un dépointage vers le nord, l'est, le sud et enfin le nord-est. Dans la dernière image, l'angle de vue s'approche beaucoup de l'angle solaire et les façades des immeubles, intégralement dans l'ombre, n'apparaissent donc que très peu.

La seconde est l'orbite héliosynchrone et concerne les satellites restant constamment à la même heure solaire. Un satellite héliosynchrone possède l'avantage de toujours capturer une zone avec les mêmes conditions d'illumination solaire, ce qui peut être très intéressant lorsqu'il s'agit de comparer plusieurs images successives. Cette orbite est donc particulièrement utilisée. Notons que, pour cela, elles demandent de vérifier des conditions précises sur la hauteur h et l'angle d'incidence α [Boa04], et se situent donc à des altitudes comprises entre 400 et 1000 kilomètres (pour une période comprise entre 1h30 et 2h et un temps de revisite d'environ deux semaines).

Dépointage Les satellites possèdent des systèmes de stabilisation leur permettant d'effectuer des rotations sur eux-mêmes pour ajuster leur angle de vue et le maintenir constamment dirigé vers la Terre. Certains satellites utilisent ce dispositif pour pouvoir observer la Terre autrement qu'à la verticale (au nadir). On parle alors de dépointage.

Notons toutefois qu'en se dépointant, les satellites capturent des images qui subissent une certaine déformation perspective ainsi que des effets d'occlusion, dus au phénomène de parallaxe. Nous avons affiché, dans la figure 1.3, un exemple d'une même zone prise avec plusieurs angles incidents différents.

Figure 1.4 – Exemple d'images satellites très haute résolution. Il s'agit d'une image prise par le satellite Pléiades-HR 1A, de taille 23082×22080 pixels avec une résolution de 50 cm par pixel. Pour mieux apprécier la taille de cette image, nous avons affiché tout d'abord son intégralité puis deux agrandissements successifs, en représentant l'emprunte de ces derniers par des carrés noirs.

1.1.3.2 Système optique et résolutions

Dans le cadre de cette thèse, nous nous intéressons en particulier à l'imagerie satellitaire optique. Notons cependant qu'il en existe d'autres types, comme l'imagerie SAR (Radar à Synthèse d'Ouverture), qui repose sur l'émission d'un signal radar et sur la mesure des signaux réfléchis, ou l'image Lidar (LIght Detection And Ranging), reposant sur l'émission et mesure d'un signal optique (presque toujours un laser) pour en déduire la distance de chaque point au système d'acquisition.

Il est important de noter que les images satellites se distinguent des images prises sur Terre sur plusieurs points. Tout d'abord, elles sont souvent très grandes, pouvant aller jusqu'à plusieurs dizaines de milliers de pixels de large. À cause de cela, nous ne pouvons jamais directement utiliser une image dans son intégralité. Cette dernière est donc découpée en sous-images, de l'ordre du millier de pixels de large, puis traitée, et les résultats sont éventuellement fusionnés pour reconstruire une image de résultat correspondant à l'image d'origine. Par ailleurs, contrairement à une prise de vue sur Terre, où un même objet peut apparaître à plusieurs échelles, en fonction de sa distance au système optique, dans le cas de l'imagerie satellitaire, cette distance est relativement fixe, dépendant uniquement de la hauteur du satellite. De plus, beaucoup de satellites ne se restreignent pas aux simples canaux « Rouge », « Vert » et « Bleu ». Détaillons alors brièvement quelques caractéristiques principales d'une image satellite, à savoir

— son nombre de canaux et les longueurs d'onde auxquelles ils correspondent ;
— sa résolution radiométrique ;
— et sa résolution spatiale.

Longueurs d'onde mesurées Lors de la capture d'une image, les capteurs optiques du satellite mesurent le rayonnement de la scène dans une ou plusieurs bandes de fréquence (ou, de manière équivalent, de longueur d'onde). Ce rayonnement peut correspondre à un flux réfléchi provenant d'une source lumineuse (presque exclusive-

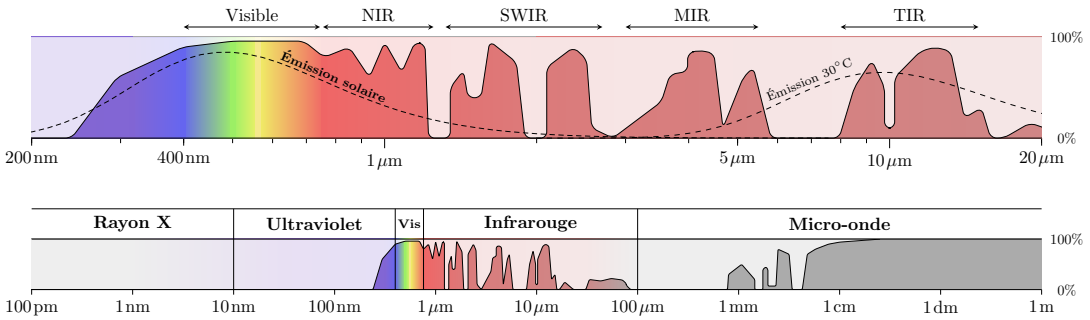


Figure 1.5 – Représentation schématique de la transmittance de l'atmosphère en fonction de la longueur d'onde. La transmittance correspond à la fraction du signal qui traverse effectivement l'atmosphère, sans être réfléchie ou absorbée. Les longueurs d'onde utilisables en télédétection (et en télécommunication) correspondent donc aux domaines fréquentiels de transmittance élevée. La courbe supérieure correspond à un agrandissement de la courbe inférieure, au niveau du domaine visible et infrarouge. Sur la courbe supérieure, nous avons fait figurer les domaines correspondant au visible, NIR, SWIR, MIR et TIR, ainsi que le spectre d'émission solaire et thermique (à 30° C).

ment le soleil dans notre cas) ou à un rayonnement propre (en infrarouge lointain ou thermique). Si les longueurs d'onde mesurées peuvent varier en fonction du système optique, on retrouve très souvent les mêmes types de bandes de fréquences, comme

- les trois bandes optiques, Bleu, Vert et Rouge, correspondant à des longueurs d'onde respective de l'ordre de 475 nm, 550 nm et 660 nm ;
- l'infrarouge proche (NIR), entre 750 nm et 1300 nm ;
- les bandes « SWIR » (Short-Wavelength InfraRed), situées entre 1400 nm et 3000 nm ;
- l'infrarouge moyen (MIR ou MWIR), situé entre 3000 nm et 5000 nm ;
- ou encore l'infrarouge thermique (TIR), situé entre 8 μm et 15 μm.

Notons que, comme montré dans la figure 1.5, ces domaines correspondent aux principaux domaines fréquentiels qui ne sont pas trop absorbés ou réfléchis par l'atmosphère et qui correspondent à un rayonnement solaire réfléchi ou à un rayonnement thermique.

La majorité des satellites possèdent aujourd'hui plusieurs capteurs et les images obtenues peuvent donc avoir plusieurs canaux, chaque canal correspondant à une bande de fréquence. On parle alors d'images multispectrales pour des images possédant jusqu'à une vingtaine de canaux et d'images hyperspectrales lorsque ce nombre est de l'ordre de la centaine ou plus.

Il faut aussi noter l'existence de la bande « panchromatique », mesurée sur un domaine fréquentiel large (en visible et proche infrarouge) et dont le flux de photons plus important permet d'obtenir une résolution spatiale plus élevée.

Résolution radiométrique La résolution radiométrique est définie comme la capacité d'un système d'imagerie à enregistrer de nombreux niveaux de luminosité et est donc reliée au nombre de bits utilisés pour le stockage ces valeurs de luminosité. Si le codage des images se fait sur 8, 16 ou 32 bits, dans la pratique, certains capteurs pos-

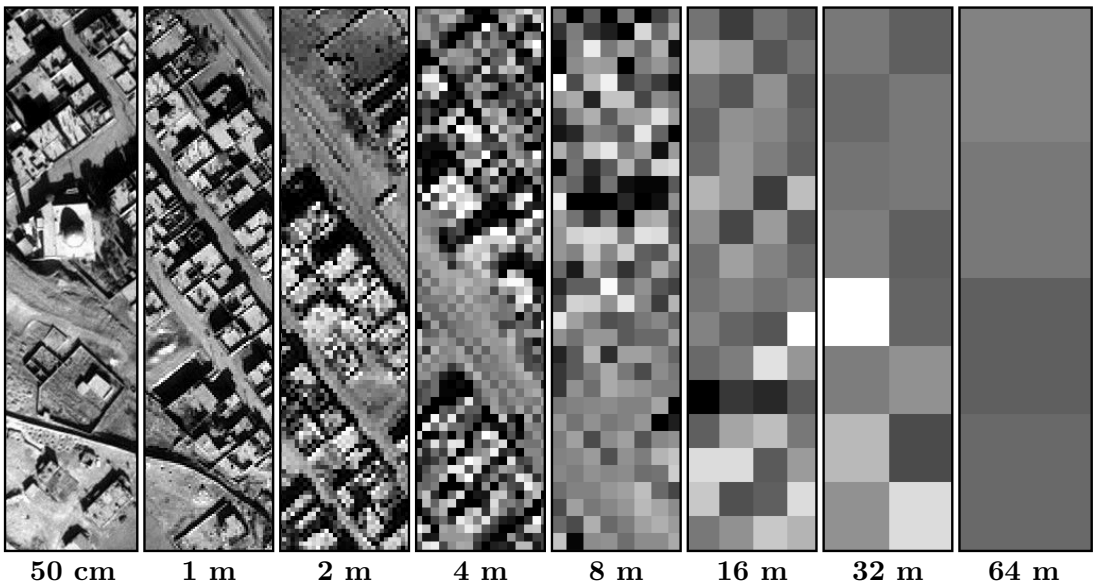


Figure 1.6 – Représentation d'une image satellite à différentes résolutions spatiales, allant de 50 centimètres par pixels à 64 mètres par pixels.

sèdent des résolutions radiométriques natives intermédiaires, qui sont alors stockées sur un nombre de bits supérieur. C'est par exemple le cas de certains capteurs du satellite Landsat 8, fournissant nativement des signaux sur 12 bits, mais encodés sur 16.

Résolution spatiale Pour finir, la résolution spatiale de l'image correspond à la taille que représente un pixel. Pour des images satellites, elle peut aller de quelques dizaines de centimètres jusqu'à plusieurs kilomètres dans le cas des satellites météorologique. On parle alors ici de

— très haute résolution pour des résolutions inférieures au mètre ;
— haute résolution pour des résolutions de l'ordre du mètre ;
— moyenne résolution pour des résolutions de l'ordre de la dizaine de mètres ;
— et de basse résolution pour le reste.

Nous montrons dans la figure 1.6, un exemple d'image satellite à différentes résolutions spatiales.

Il est important de noter que, la Terre étant sphérique, la surface au sol que représente un pixel n'est pas la même pour tous les pixels, celle-ci étant minimale au nadir (à la verticale du satellite) et augmentant à mesure que l'on s'en éloigne. Dans la pratique, les images étant réinterpolées pour être toutes placées dans un même système de coordonnées et suréchantillonées pour limiter certains artefacts lors du traitement, les images que nous considérerons ici auront des pixels de taille constante.

1.2 Détection de changements par mesure directe

1.2.1 Définition du problème

On s'intéresse au problème de la détection générique de changements entre deux images satellites, en particulier dans le cas de la très haute résolution, avec pour but l'aide à la photo-interprétation.

Le problème peut être décrit de la manière suivante : on se donne deux images satellites représentant une même zone géographique à deux temps différents, ces images pouvant éventuellement être prises par deux satellites différents, à deux résolutions spatiales différentes, ou correspondant à des longueurs d'onde différentes, et il s'agit alors de détecter l'ensemble des changements intervenant entre ces deux images. Comme nous l'avons mentionné précédemment, le terme de changements peut désigner de nombreuses choses, comme l'apparition, la disparition ou la translation d'objet, une modification de couleur, taille, forme ou orientation d'un autre. Un changement peut a priori être de n'importe quelle taille et intervenir sur une zone réduite de l'image aussi bien que sur son intégralité. On envisage ici la détection de changements dans le cadre de l'aide à la photo-interprétation et on souhaite détecter les changements pour orienter et prioriser l'attention d'un photo-interprète devant étudier la paire d'images. De ce fait, on s'intéresse à un problème de détection générique de changements, sans se donner d'a priori sur le type de changements recherchés.

Notons que la notion de « changement » n'est pas réellement définie, ce qui constitue la difficulté principale du problème de détection de changements. Plusieurs méthodes approchent ce problème en se restreignant à des changements très spécifiques, comme l'apparition ou disparition d'habitations [BDZ12], l'apparition de fuite de pétrole [CLPT11] en encore la déforestation [Mal80]. Si cette approche peut parfaitement convenir à des utilisations spécifiques, elles ne sont pas envisageables dans le cadre de l'aide générique à la photo-interprétation où le type de changements à détecter peut dépendre fortement du cas d'usage.

Autres problèmes de détection de changements Avant toute chose, distinguons le problème étudié ici d'autres problème de détection de changements.

On trouve tout d'abord la détection de changements dans des séries temporelles. Il s'agit alors de détecter les changements dans un ensemble de plusieurs images (au moins trois), ce qui démultiplie les définitions possibles d'un changement, ce dernier pouvant n'exister que dans une seule image ou s'inscrire plus généralement dans un contexte temporel, si l'on prend en compte des effets de périodicité dans les images (par exemple dus aux saisons). La détection de changements est aussi étudiée pour des vidéos et il s'agit alors presque exclusivement du traitement de vidéos de surveillance [HN00, GJP+12].

On peut aussi distinguer la détection de changements dans des paires d'images multispectrales [HS18], possédant un nombre réduit de bandes de fréquences (en général trois ou quatre), de celle appliquée à des paires d'images hyperspectrales, présentant parfois plus d'une centaine de canaux. D'une part, ces dernières étant souvent de résolution moyenne (plus d'une dizaine de mètres par pixel), elles ne subissent que très peu les effets de parallaxe. D'autre part, la très grande dimension associée aux intensités

des pixels doit aussi nécessairement être prise en compte d'un point de vue complexité algorithmique.

Distinction pixellique et objets Une grande partie des auteurs en détection de changements dans des paires d'images distinguent deux approches principales : une approche pixellique, que l'on retrouve sous les noms de « Direct Comparison », de « Direct Approaches », ou encore de « Pixel-Based », et une approche dite « Object-based ». Cette distinction se retrouve naturellement dans les principales revues de l'état de l'art, comme celles de Coppin et coauteurs [CJN+04], de Chen et coauteurs [CHCW12], de İlsever and Ünsalan [İÜ12], Hussain et coauteurs [HCC+13], ou encore celle de Ban and Yousif [BY16].

D'un côté, les méthodes pixelliques envisagent les changements en comparant les deux images, pixels à pixels, considérant ces derniers comme indépendant. De l'autre, les approches orientées objets cherche à tirer partie de la cohérence entre les pixels voisins, par exemple, en segmentant l'image en différents objets puis en comparant ces objets entre les deux images.

Cette séparation binaire nous semble particulièrement limitée et ne prend pas en compte certaines approches, comme l'approche de Hong and Vatsavai [HV16] comparant les distributions locales des deux images, ou encore celle de Gueguen et coauteurs [GSP11] utilisant la distribution empirique jointe évaluée sur l'intégralité du domaine des images.

Regroupement des approches Nous proposons de regrouper différemment les différentes approches.

— Nous présentons tout d'abord, dans la section 1.2.2, les approches que nous qualifierons de *pixelliques*, et qui correspondent aux modèles dit « Pixel-based ». Ici, les différents pixels sont considérés comme indépendants et il s'agit de comparer leurs valeurs dans les deux images. Ces méthodes sont particulièrement pertinentes dans le cadre de la détection de changements en résolution basse ou moyenne. Notons que nous verrons certains modèles « Object-based » comme une extension des modèles pixelliques, cherchant à tirer parti de la corrélation spatiale des pixels pour limiter l'influence des décalages spatiaux faibles ou de petites anomalies.

— Nous présentons ensuite les approches par *représentations locales*. Il s'agit ici de prendre en compte une certaine variabilité de la scène entre les deux images, que ce soit en comparant les distributions locales des images (voir section 1.2.3) ou en effectuant une modélisation de l'image sans changement (voir section 1.3.1). Ces méthodes subissent très peu les problèmes de décalages spatiaux et sont donc particulièrement adaptées à l'imagerie très haute résolution.

— Enfin, nous présentons les approches prenant en compte d'éventuelles différences radiométriques entre les deux images, que ce soit utilisant des mesures de similarités issues de la théorie de l'information (voir section 1.3.2) ou en cherchant à transformer les deux images pour les ramener dans un même « espace radiométrique » (voir section 1.3.3).

1.2.2 Modèle de changements par comparaison ponctuelle

L'approche pixellique Comme nous l'avons dit précédemment, la grande majorité des algorithmes de détection de changements utilise une « approche pixellique », en considérant les changements comme des variations significatives des valeurs des pixels entre les deux images. Cette approche peut se décomposer en deux étapes. Tout d'abord, il s'agit de calculer, à partir de chaque image, une certaine représentation de ces dernières, mettant en valeur certaines caractéristiques et permettant de se ramener à des images plus facilement comparables. Ces représentations peuvent correspondre aux canaux bruts des images, à un ensemble de descripteurs ou même à des transformations de l'espace. Il s'agit ensuite d'évaluer, pour chaque pixel du domaine, si la différence de ces deux représentations correspond à un changement ou pas.

Formellement, étant donné deux images I_1 et I_2, les modèles pixelliques peuvent s'écrire sous la forme

$$S\Big(f(I_1)_p - f(I_2)_p\Big)\,, \tag{1.2}$$

où f est une fonction correspondant à la représentation d'image choisie, X_p est la valeur de l'image X au pixel p et où S est une fonction de décision, à valeur dans $[0, 1]$, une valeur 1 indiquant que le pixel p est considéré comme un changement. Notons que, dans la plupart des cas, la fonction de décision S se ramène à un seuillage de la norme de la différence, en testant ainsi si

$$\Big\|f(I_1)_p - f(I_2)_p\Big\|_2 \geq \lambda, \tag{1.3}$$

pour $\lambda > 0$ une valeur de seuil donnée.

Représentation des images Les représentations d'images peuvent être séparées en plusieurs catégories.

Tout d'abord, il est important de citer les « indices spécifiques ». Dans la littérature, les approches pixelliques sont beaucoup utilisées pour la détection de changements en imagerie multispectrale de résolution moyenne ou basse, relatif à des études sur la modification de la couverture du sol (extension des forêts, lacs, exploitations agricoles, villes, etc). Ces modèles de changements se concentrent presque exclusivement sur le choix d'un bon descripteur, fusionnant les différents canaux (principalement Rouge, Vert, Bleu, NIR et SWIR) pour mettre en valeur différents éléments, comme des zones de végétation, de mer ou encore des zones urbaines. On peut citer ici l'indice le plus utilisé, le NDVI (Normalized Difference Vegetation Index) [Uns07] qui s'écrit

$$\text{NDVI} \;=\; \frac{\text{NIR} - R}{\text{NIR} + R}. \tag{1.4}$$

Nous reviendrons brièvement sur ces descripteurs dans la section 3.3.2.

En parallèle de ces indices très spécifiques à la télédétection et à des cas précis de détection de changements, on trouve des descripteurs d'images plus génériques utilisés dans de multiples problèmes de traitement d'images. On trouve par exemple des méthodes utilisant les coefficients de texture d'Haralick [TEK11]. Récemment, quelques méthodes utilisent des réseaux de neurones convolutionnels en guise de descripteurs

d'image, comme celle de El Amin et coauteurs [EALW17] ou celle de Saha et coauteurs [SBB19]. Ces réseaux de neurones sont entraînés sur des images naturelles, non satellite, et souvent pour résoudre un problème de détection d'objets, et semblent, malgré cela, s'adapter relativement bien à l'imagerie satellite, comme le montre Penatti et coauteurs [PNDS15], et au problème de détection de changements.

Pour finir, certaines transformations d'images sont aussi utilisées. Certaines méthodes considèrent la différence normalisée, cité par Coppin et Bauer [CB94] ou le rapport des intensités, utilisé par Rignot et Van Zyl [RVZ93], s'écrivant respectivement, pour un pixel p,

$$\frac{I_1(p) - I_2(p)}{I_1(p) + I_2(p)} \qquad \text{et} \qquad \frac{I_1(p)}{I_2(p)}, \tag{1.5}$$

que l'on peut voir, en quelque sorte, comme des différences de représentation d'images, avec

$$\frac{I_1(p)}{I_1(p) + I_2(p)} - \frac{I_2(p)}{I_1(p) + I_2(p)} \qquad \text{et} \qquad \exp\Big(\log I_1(p) - \log I_2(p)\Big), \tag{1.6}$$

en acceptant ici que la représentation f dépendent des deux images. Deng et coauteurs [DWDQ08] utilisent une décomposition en composantes principales pour pouvoir réduire le nombre de canaux des images tout en restant le plus possible adapté aux données, tandis que Nielsen et coauteur [NCS98] considèrent transformation MAD (Multivariate alteration detection), cherchant à maximiser la corrélation de I_1 et I_2 en appliquant une transformation linéaire différente à chaque image.

Critère de décision Intéressons-nous à présent à la seconde phase de l'approche pixellique, c'est-à-dire au critère de décision utilisé pour déterminer si une différence de représentation, au niveau d'un pixel, correspond à un changement ou pas. Ici aussi, on peut différencier plusieurs approches.

La première approche correspond à un simple seuillage de la différence par une valeur fixée en amont de manière empirique [TDBPP01, GCC+11, FCL+10, CLPT11].

D'autres approches consistent à se donner un ensemble de données où les changements ont été marqués, et à utiliser ces données pour apprendre le critère de décision. Le problème est alors traité comme un simple problème de classification sur la différence des deux images. C'est le cas de Bovolo et coauteurs [BBM08], qui utilisent une « Support Vector Machine » (SVM), cherchant à séparer les changements et les non-changements par un hyperplan, ou de Camps-Valls et coauteurs [CVGCMM+08] qui utilisent un équivalent sphérique comme un « Support Vector Domain Description » (SVDD), cherchant à isoler les non-changements au sein d'une boule minimale.

Pour éviter de devoir utiliser des données labellisées, Bovolo et Bruzzone [BB11] tentent de modéliser les distributions associées aux changements, $p(\cdot|C)$, et aux non-changements, $p(\cdot|N)$, puis de définir les changements de manière bayesienne lorsque

$$\frac{p(x_p|C)}{p(x_p|N)} \;\geq\; 1, \tag{1.7}$$

la valeur x_p correspondant à la différence des deux images au pixel p. Bruzzone propose ainsi de modéliser toutes les données par un mélange de deux gaussiennes (à l'aide

d'un algorithme EM), une première gaussienne correspondant aux changements, une seconde correspondant aux non-changements.

Mentionnons brièvement de la méthode dite de « Change Vector Analysis » (CVA), beaucoup utilisée et citée en détection de changements. Cette méthode, introduite en 1980 par Malila [Mal80], rajoute, après un premier critère de changements sur la norme de la différence, une phase de segmentation des différents types de changements en utilisant la « direction du vecteur de changements », donnée par exemple par

$$\alpha_p \;=\; \mathrm{Arccos}\left(\frac{x_p^\top x^0}{\|x_p\|_2\,\|x^0\|_2}\right), \tag{1.8}$$

où x_p est la différence au pixel p, et x^0 est un vecteur de référence [BMB12].

Cas de la haute résolution Notons que l'approche pixellique est très sensible au décalage spatial entre les images, particulièrement présent en haute résolution. Pour prendre en compte d'éventuels décalages, Thonfeld et coauteurs [TFBM16] introduisent le « Robust Change Vector Analysis » (RCVA), qui consiste à remplacer la différence simple des deux images par une version robuste aux légères translations, en considérant, pour X_1 et X_2 les deux images étudiées et p un pixel,

$$X_D^a(p) \;=\; \min_{\|p-q\|_\infty \le w} I_2(p) - I_1(q) \qquad \text{et} \qquad X_D^b(p) \;=\; \min_{\|p-q\|_\infty \le w} I_1(p) - I_2(q), \tag{1.9}$$

pour $w \in \mathbb{N}^*$ et en définissant le vecteur de différence par

$$X_D(p) \;=\; X_D^a(p)\mathbb{1}_{\{X_D^a(p)>0\}} - X_D^b(p)\mathbb{1}_{\{X_D^a(p)\le 0\}}. \tag{1.10}$$

Plusieurs autres méthodes, proposent de prendre en compte les corrélations spatiales existant entre les différents pixels pour limiter l'impact des petites erreurs dues entre autres aux décalages. Li et coauteur [LLZ+16] introduisent, par exemple, l'« Object-oriented Change Vector Analysis » (OCVA) qui utilise une segmentation des images comme descripteurs du « Change Vector Analysis », et permet ainsi de prendre en compte le regroupement des pixels en objets. Plusieurs auteurs considèrent des champs de Markov pour prendre en compte les corrélations spatiales des pixels, comme Singh et coauteurs [SKZ14] ou encore Benedek et coauteurs [BSK+15].

D'un autre point de vue, Dalla Mura [DMBBB08] propose d'en tenir compte au moyen de filtres morphologiques.

Comparaison ponctuelle a contrario Le plus souvent, les méthodes présentées précédemment nécessitent soit de disposer d'un ensemble de données labellisées permettant de distinguer les changements des non-changements, soit de disposer d'une estimation à peu près fiable de la distribution associée aux changements. Ces deux suppositions peuvent poser problème. D'une part, pour pouvoir espérer apprendre à distinguer les changements des non-changements, il est nécessaire de disposer d'exemples suffisamment représentatifs de l'ensemble des changements possibles. Et si le nombre d'images satellites a fortement augmenté dans les dernières années, ce nombre reste

limité et il est difficilement envisageable d'espérer constituer une base d'apprentissage suffisamment représentative. De la même manière, il est difficile d'envisager une modélisation crédible de l'ensemble des changements, à moins de se retreindre à des changements très particuliers.

Une manière d'éviter ces deux suppositions est de se placer dans le cadre statistique « a contrario », introduit par Desolneux et coauteur au début des années 2000 [DMM00] et beaucoup utilisé dans de nombreux problèmes de vision par ordinateur, comme l'estimation d'alignements de points, la détection de tâches dans des textures de bruit coloré (Grosjean et Moisan [GM08]), la détection d'anomalies (Ehret et coauteurs [EDMD18]), mais aussi la détection de changements en télédétection [LHMS04, RMLHM10, DDG+14, LGT19].

Le cadre « a contrario » repose sur le principe de Helmholtz, qui stipule que « on ne perçoit rien dans une image de bruit » [Att54], ou plus précisément que l'oeil perçoit des structures lorsque ces dernières s'éloignent trop grandement de l'aléatoire. Il s'agit alors de fournir une modélisation du « bruit » (ici, la situation ordinaire, sans changement) et d'estimer, dans les images données, l'écart obtenu vis-à-vis de ce modèle dit « naïf ». En considérant alors la fonction de décision, appelée par la suite « Nombre de Fausses Alarmes » (NFA),

$$\mathrm{NFA}(x_p) \;=\; N\,\mathbb{P}(\,\mu(X_p) \geq \mu(x_p)\,)\,, \tag{1.11}$$

pour x_p la différence des deux images observées au pixel p, X_p une variable aléatoire modélisant cette différence sans changement, μ une fonction réelle arbitraire évaluant cette différence (on pourra par exemple prendre $\mu(x) = \|x\|_2$) et N le nombre total de pixels, le cadre « a contrario » nous assure alors qu'en détectant tous les pixels p vérifiant $\mathrm{NFA}(x_p) \leq \varepsilon$, pour une valeur $\epsilon \geq 0$ donnée, on est assuré de ne pas, en moyenne, détecter plus de ε fausses alarmes.

Dans la section 3.4, nous reprenons l'approche pixellique dans le cadre statistique « a contrario ». Nous faisons la supposition que, dans le cas où il n'existe aucun changement entre les deux images, les variations mesurées correspondent à un bruit gaussien, indépendant pour chaque pixel. Nous obtenons une expression exacte du nombre moyen de fausses alarmes,

$$\mathrm{NFA}(x_p) \;=\; \frac{N}{\Gamma(\frac{K}{2})}\Gamma\left(\frac{K}{2}, \frac{\|x_p\|_2^2}{2\sigma^2}\right), \tag{1.12}$$

où K est le nombre de canaux des images, N le nombre de pixels, σ^2 est la variance du bruit au niveau de chaque pixel et $\Gamma(\cdot)$ et $\Gamma(\cdot,\cdot)$ sont respectivement les fonctions d'Euler gamma et gamma incomplète. Nous proposons un moyen d'étendre ce modèle à l'utilisation de descripteurs stationnaires, f, c'est-à-dire lorsque

$$x_p \;=\; f(I_2)_p - f(I_1)_p, \tag{1.13}$$

pour une fonction f associée aux descripteurs. Nous montrons qu'il est possible de substituer simplement $\|x_p\|_2^2$ par $\|\Sigma_f^{-1/2} x_p\|_2^2$ dans l'expression du nombre de fausses alarmes précédent, où $\Sigma_f^{-1/2}$ correspond à la racine de l'inverse de la matrice de covariance des descripteurs f, tout en conservant un nombre de fausses alarmes, exact dans le cas des descripteurs linéaires.

Pour prendre en compte d'éventuelles incohérences spatiales entre les deux images, notamment dans le cadre de l'imagerie très haute résolution, nous proposons ensuite d'adapter le modèle en minimisant ponctuellement la différence des deux images sur un voisinage du pixel étudié, en considérant

$$x_p = \min_{\|\tau\| \leq \tau_{\max}} f(X_2)_p - f(X_1)_{p+\tau}, \tag{1.14}$$

où $\tau_{\max} \in \mathbb{N}$ est une constante majorant l'amplitude maximale de la translation pris en compte. Ce modèle permet alors de détecter toutes les variations significatives des descripteurs en chaque pixel, en étant robuste aux translations faibles et tout en restant inscrit dans le cadre statistique a contrario.

Il possède toutefois quelques limitations. D'une part, il est préférable de ne pas considérer une robustesse aux translations trop importante, le temps de calcul devenant rapidement important et les résultats devenant difficilement interprétables. Aussi, ce modèle fonctionne relativement mal en présence de forte parallaxe, notamment dans les scènes urbaines. D'autre part, en considérant un modèle donnant un rôle symétrique aux deux images et en se restreignant à la différence des deux images, ce modèle ne parvient pas à distinguer une apparition d'une disparition, ou une apparition d'un élargissement d'objet.

1.2.3 Modèle de changements par comparaison d'histogramme

Faiblesse de l'approche pixellique Comme nous l'avons déjà souligné précédemment, les approches pixelliques apparaissent comme performantes dans le cas de la résolution moyenne et basse, où les effets d'un changement d'angle de vue n'ont que très peu d'influence sur la parallaxe et où il est donc envisageable de recaler presque parfaitement les images étudiées. Toutefois, ce recalage spatial des deux images est souvent impossible à réaliser parfaitement en résolution plus élevée. Ceci est particulièrement vrai en très haute résolution, où une légère variation de l'angle de vue peut provoquer un décalage conséquent des bâtiments, arbres et autres objets verticaux, pouvant parfois atteindre la dizaine de pixels.

Pour résoudre ce problème et proposer des détecteurs robustes aux décalages spatiaux, plusieurs méthodes sont proposées dans la littérature.

Approche pixelliques robustes Une première approche, déjà mentionnée précédemment, consiste à utiliser des méthodes pixelliques prenant en compte les corrélations spatiales entre les différents pixels, dans le but de fournir une détection plus robuste aux problèmes de l'imagerie haute résolution. Toutefois, si ces méthodes parviennent à limiter l'impact des petites erreurs dues aux décalages de quelques pixels, elles ne sont cependant pas réellement efficaces dans des cas de forts décalages,q pouvant parfois atteindre la dizaine de pixels.

Comparaison des distributions locales Une seconde approche consiste à comparer les distributions locales des images, plutôt que les simples valeurs associées aux différents pixels. Cette approche peut être décrite en deux temps. Il s'agit tout d'abord

de représenter ou de modéliser la distribution locale des deux images au voisinage d'un point, puis de considérer une mesure de similarité entre ces deux représentations, que l'on peut alors seuiller pour en déduire l'ensemble des changements. Il est intéressant de remarquer que cette description est en réalité identique à l'approche pixellique, si l'on considère comme descripteurs d'image la représentation de la distribution locale.

Pour estimer la similarité entre les distributions locales, la grande majorité des approches repose sur la divergence de Kullback-Leibler [MDP+06, IM07, VG12, HV16], que l'on peut écrire

$$\mathrm{KL}(f\|g) \;=\; \int_{\mathbb{R}^n} f(x)\log\frac{f(x)}{g(x)}dx, \tag{1.15}$$

où f et g sont les fonctions de densités associées aux distributions que l'on souhaite comparer. Cette mesure n'étant pas symétrique, sa version symétrisée,

$$D(f,g) \;=\; \mathrm{KL}(f\|g) + \mathrm{KL}(g\|f)\,, \tag{1.16}$$

est presque toujours utilisée.

Pour pouvoir effectivement calculer cette divergence, plusieurs méthodes proposent de modéliser les distributions locales par des lois usuelles, possédant une expression explicite de la divergence de Kullback-Leibler. Hong and Vatsavai [HV16] considèrent des modélisations gaussiennes, la mesure précédente s'écrivant alors

$$D(g_1,g_2) \;=\; \frac{\sigma_1^4+\sigma_2^4+(\mu_1-\mu_2)^2(\sigma_1^2+\sigma_2^2)}{2\sigma_1^2\sigma_2^2} - 1 \tag{1.17}$$

lorsque g_1 et g_2 correspondent à des distributions gaussiennes de paramètres respectifs (μ_1,σ_1^2) et (μ_2,σ_2^2). Inglada et Mercier [IM07] proposent une modélisation locale par des fonctions de Pearson (une famille de fonctions comprenant entre autres les distributions gaussiennes, gamma et beta). Ils proposent aussi d'estimer la divergence de Kullback-Leibler à partir des premiers cumulants des distributions, ces derniers étant des coefficients similaires aux premiers moments. Notons que cette extension aux fonctions de Pearson est particulièrement pertinente dans le cas des images radar, où le bruit présent dans les images ne correspond pas du tout à un bruit gaussien.

Notons par ailleurs que d'autres mesures de similarité entre distributions existent et sont régulièrement utilisées en traitement d'images. On peut citer la distance de Bhattacharyya [SUD10] (qui n'est pas une distance), la distance d'Hellinger [EA96], la α/β divergence [CCA11], ou encore la distance de Wasserstein [WJS16]. Ces mesures ne semblent toutefois pas avoir été utilisées dans le cadre de la détection de changements.

Critère de décision De la même manière que pour les approches pixelliques, une fois toutes les mesures de similarité calculées, il s'agit alors de trouver un critère de décision, permettant de calculer un masque binaire représentant les changements à partir d'une image d'erreur.

Pour cela, comme pour ce qui précède, plusieurs auteurs proposent de fixer une valeur de seuil à partir de données de référence, comme Mercier et coauteurs [MDP+06]. Hong et Vatsavai [HV16] proposent de distinguer les changements des non-changements en appliquant un algorithme de clustering, apprenant un modèle de mélange de gaussiennes.

Utilisation des distributions empiriques Une dernière approche, très peu utilisée, consiste à considérer directement les distributions empiriques, sans modélisation, et à les comparer au moyen de la distance de Kolmogorov-Smirnov [Egh79, TZH11].

Étant donné un ensemble de pixels D sur lequel on envisage les distributions locales des deux images I_1 et I_2, il s'agit alors de définir les deux fonctions de répartition empirique

$$F_i(t) \;=\; \frac{1}{|D|}\sum_{p\in D} \mathbb{1}_{\{I_i(p)\leq t\}}, \tag{1.18}$$

pour $i \in \{1,2\}$, $t \in \mathbb{R}$ et en notant $|D|$ le cardinal de D, puis de considérer la distance entre ces distributions

$$d(F_1,F_2) \;=\; \max_{t\in\mathbb{R}} |F_1(t) - F_2(t)|. \tag{1.19}$$

Eghbali [Egh79] propose alors d'utiliser le test de Kolmogorov-Smirnov comme critère de décision et d'en déduire une valeur de seuil, ce dernier donnant une estimation de la probabilité que $d(F_1,F_2)$ soit inférieure à une constante, lorsque F_1 et F_2 suivent une même distribution et que le nombre de pixels considérés est suffisamment grand.

Dans la section 3.5, nous reprenons cette approche utilisant la « distance de Kolmogorov-Smirnov » dans le cadre statistique a contrario. L'utilisation des distributions empiriques est particulièrement intéressante puisqu'elle permet d'éviter une étape de modélisation, qui peut manquer de pertinence lorsque le voisinage sur lequel on estime la distribution locale est trop petit.

Dans un premier temps, nous reprenons le modèle d'Eghbali [Egh79] et utilisons le test de Kolmogorov-Smirnov pour en déduire une expression approchant un nombre de fausses alarmes

$$\mathrm{NFA}(p) \;=\; 4\,N\sum_{k=1}^{\infty}(-1)^{k-1}e^{-\frac{1}{2}n\|F_1-F_2\|_\infty^2 k^2}, \tag{1.20}$$

où N est le nombre total de pixels dans les images et n est le nombre de pixels utilisés pour estimer la fonction de répartition empirique.

Par la suite, nous présentons une seconde expression

$$\mathrm{NFA}(p) \;=\; 2N\sum_{i=1}^{[n/k_p]}\frac{(-1)^{i+1}n!^2}{(n+ik_p)!(n-ik_p)!}, \quad k_p = \Big[\, n\,\|F_1-F_2\|_\infty \Big], \tag{1.21}$$

où $[\cdot]$ représente la partie entière inférieure, correspondant cette fois ci à un nombre de fausses alarmes, sans hypothèse sur le nombre de pixels considérés pour l'estimation de la distribution empirique, alors que le test de Kolmogorov-Smirnov ne donne qu'une estimation asymptotiques.

Le modèle obtenu permet alors bien de détecter des changements tout en étant robuste aux décalages locaux, ces décalages ne devant toutefois pas dépasser la largeur du voisinage utilisé pour estimer la distribution empirique. En revanche, l'utilisation de fonction de répartition contraint ce modèle à n'être utilisé que sur des images à un seul canal (ou sur des descripteurs d'images de dimension 1). Par ailleurs, en comparant des représentations d'images à l'échelle d'un voisinage des pixels, le résultat obtenu présente une résolution amoindrie, le masque de détection apparaissant comme flou.

1.3 Modèles pour la détection indirecte de changements

1.3.1 Modélisation locale du contexte

Estimation du signal de fond et image résiduelle Plusieurs travaux étudient à un problème se rapprochant du problème de détection, la détection de cible dans du bruit. Le problème est le suivant : étant donné plusieurs images de référence, ne contenant que du bruit, ainsi qu'une dernière image contenant éventuellement certaines structures, comment alors de détecter ces structures. Notons que ce problème peut être considéré comme un cas particulier de la détection de changements, où ces derniers correspondent nécessairement à des apparitions d'objets n'étant pas présents dans les images de référence.

Reed et Yu [RY90] s'intéressent à ce problème dans le cas où la forme de la cible est connue. Ils proposent de modéliser le fond, présent dans les images de référence, par un bruit gaussien, et obtiennent alors, à partir du test de vraisemblance généralisé [VT04, p.86-96], le critère de décision

$$r(X) \;:=\; \frac{(Xs)^{\mathsf{T}}(XX^{\mathsf{T}})^{-1}Xs}{s^{\mathsf{T}}s} \;\overset{?}{\gtrless}\; \lambda, \tag{1.22}$$

où X correspond à l'image, s à la forme supposée de la cible et λ à une valeur de seuil donnée. Cette expression est reprise par Stein et coauteurs [SBH+02] dans le cas où les pixels sont considérés séparément, ce dernier obtenant le critère

$$\mathrm{RX}(x) \;:=\; (x-\mu)^{\mathsf{T}}M^{-1}(x-\mu) \;\overset{?}{\gtrless}\; \lambda, \tag{1.23}$$

où μ et M sont respectivement la moyenne et la matrice de covariance du bruit, estimées empiriquement sur des données de référence.

Cet algorithme est beaucoup repris pour la détection d'anomalies, sous le nom d'algorithme de Reed-Xiaoli (ou algorithme RX), en considérant comme données de références, pour évaluer un pixel p, l'ensemble des pixels situés dans une couronne autour de p, en particulier dans des images multispectrales ou hyperspectrales [KN05, ZKAE16]. Cet algorithme est aussi proposé par Zhou et coauteurs [ZKAE16] comme solution pour la détection de changements, en détectant les changements comme des anomalies dans la différence des deux images, anomalies détectées au moyen de l'algorithme RX. Si les auteurs proposent d'effectuer une transformation en amont de la différence, pour limiter l'influence d'une variation de contraste entre les deux images, cette méthode reste particulièrement sensible aux décalages spatiaux et donc relativement limitée en imagerie très haute résolution.

Modélisation du contexte Dans la section 4.1, nous considérons une méthode de détection de changements reposant sur cette approche de modélisation du « fond » de l'image, reprenant l'algorithme de Reed-Xiaoli. À l'instar du modèle précédent, nous supposons que les deux images possèdent les mêmes distributions locales. Seulement, nous évaluons dans quelle mesure les pixels de la seconde image respectent bien les distributions locales de la première image, plutôt que de comparer directement ces dernières.

Étant donné un pixel p, nous modélisons la distribution locale de la première image sur un voisinage de p, avant de confronter la valeur de p dans la seconde image avec la distribution estimée. Ce faisant, nous proposons une extension de l'algorithme de Reed-Xiaoli à des paires d'images et pour des distributions correspondant à des mélanges de gaussiennes, dont la fonction de densité s'écrit sous la forme

$$f_{R,\alpha,\rho,\Sigma}(x) \;=\; \sum_{r=1}^{R} \frac{\alpha_r}{\sqrt{(2\pi)^K \,|\Sigma_r|}} \exp\!\left(-\frac{1}{2}\,(x-\rho_r)^{\mathsf{T}}\,\Sigma_r^{-1}\,(x-\rho_r)\right), \tag{1.24}$$

pour tout $x \in \mathbb{R}^K$, K représentant le nombre de canaux des images, et les constantes R et α_r, μ_r, Σ_r (pour $r \in [\![1, R]\!]$) correspondant aux paramètres du modèle de mélange de gaussiennes. En s'appuyant sur l'expression de la vraisemblance, nous construisons un expression s'inscrivant dans le cadre statistique a contrario,

$$\mathrm{NFA}(p) \;=\; \frac{N}{\Gamma\!\left(\frac{K}{2}\right)} \Gamma\!\left(\frac{K}{2}, \frac{\min_r (d_{\rho_r^p,\Sigma_r^p}(I_2(p))^2 + m_r^p) - \max_r m_r^p}{2}\right), \tag{1.25}$$

où $m_r^p = \log|\Sigma_r^p| - 2\log\alpha_r^p$ et où $d_{\rho,\Sigma}(x)^2 = (x-\rho)^{\mathsf{T}}\Sigma^{-1}(x-\rho)$ est la distance de Mahalanobis. Le modèle obtenu possède alors une certaine robustesse aux décalages locaux et se distingue, par ailleurs, de la majorité des modèles classique de détection de changements par son asymétrie. En donnant deux rôles différents aux deux images, une servant de référence, l'autre étant évaluée, le modèle parvient alors à ne détecter que les apparitions, les disparitions d'objets n'étant pas perçues. Notons que les disparitions sont facilement détectables en échangeant simplement les rôles des deux images. Il est de plus naturellement robuste aux décalages modérés entre les deux images, tout en conservant une certaine finesse de détection, que l'on perdait en comparant directement les distributions locales.

1.3.2 Modèle de changements par destin commun

Images radiométriquement différentes Dans certains cas, on peut vouloir chercher des changements entre deux images qui ne sont pas comparables. Cela peut être dû à plusieurs choses, par exemple à l'utilisation de deux longueurs d'onde différentes ou de deux systèmes d'acquisitions différents (si l'on compare par exemple une image visible avec une image SAR), et l'on parle alors de paires multimodale d'images . Cela peut aussi arriver en considérant deux images prises par deux satellites différents, qui peuvent alors présenter des dynamiques différentes. C'est aussi particulièrement le cas en imagerie thermique, où l'on peut souvent voir des inversions de contraste entre certains objets (principalement au niveau des étendues d'eau). Par la suite, nous qualifierons de telles paires d'images non directement comparables de « radiométriquement différentes ».

Notons que le cas des paires d'images multimodale est particulièrement étudié en imagerie médicale, cette dernière comprenant un grand nombre de systèmes d'acquisitions différents (IRM, échographie, radiographie, tomographie, etc). Leurs études est beaucoup moins présente en imagerie satellite.

Similarité d'informations Pour comparer des images radiométriquement différentes, une approche consiste à utiliser des mesures de similarité locale issue de la théorie de l'information, ces dernières étant relativement invariantes aux variations radiométriques.

En considérant les deux images, I_1 et I_2, comme des vecteurs aléatoires discrets, ces mesures reposent sur les distributions marginales associées, p_1 et p_2, ainsi que sur la distribution jointe $p_{1,2}$, ces distributions étant estimées localement sur des voisinages de pixels.

Il existe un grand nombre de mesures de similarité. La plus commune est l'information mutuelle,

$$\mathrm{I}(I_1, I_2) \;=\; \sum_{x,y} p_{1,2}(x,y) \log \frac{p_{1,2}(x,y)}{p_1(x)p_2(y)}, \tag{1.26}$$

la somme étant effectuée sur l'ensemble des valeurs possibles des deux images, utilisée par exemple par Aktar et coauteurs [AMHS16]. L'information mutuelle permet de quantifier la dépendance des deux images et ainsi, la présence d'un changement devrait faire baisser l'information mutuelle. Plusieurs autres mesures existent et sont comparées par Alberga [Alb09], comme la distance à l'indépendance

$$\mathrm{DTI}(I_1, I_2) \;=\; \sum_{x,y} \frac{(p_{1,2}(x,y) - p_1(x)p_2(y))^2}{p_1(x)p_2(y)}, \tag{1.27}$$

le « Cluster Reward Algorithm »

$$\mathrm{CRA}(I_1, I_2) \;=\; \frac{\sum_{x,y} p_{1,2}(x,y)^2 - \sum_{x,y} p_1(x)^2 p_2(y)^2}{\sqrt{\sum_{x,y} p_1(x)^2 p_2(y)^2} - \sum_{x,y} p_1(x)^2 p_2(y)^2}. \tag{1.28}$$

Notons que, de la même manière que pour les approches pixelliques ou par comparaison de représentation locales, il est alors nécessaire de se donner un critère de décision sur les valeurs de similarités obtenues.

Considérations globales En parallèle des méthodes utilisant des mesures de similarités locales, d'autres méthodes considèrent les distributions jointes et marginales sur l'ensemble des deux images. Theiler et coauteurs [TP06] proposent de modéliser la loi jointe $p_{1,2}$ des deux images (I_1 et I_2) par une distribution gaussienne puis de seuiller la grandeur

$$\frac{p_{1,2}(I_1(p), I_2(p))}{p_1(I_1(p))\, p_2(I_2(p))}, \tag{1.29}$$

quantifiant l'indépendance des lois marginales, pour en déduire les changements.

Avec une approche similaire, Gueguen et coauteurs [GSP11] proposent d'utiliser une version régularisée de la loi jointe empirique,

$$p_{1,2}(x,y) \;=\; \frac{1}{Z}\left(1 + \sum_p \mathbb{1}_{\{(I_1(p), I_2(p)) = (x,y)\}}\right), \tag{1.30}$$

où n et m sont les nombres de canaux respectifs des images I_1 et I_2 et Z étant un coefficient de normalisation, et utilisent, comme indicateur de changements au voisinage d'un pixel p, une mesure reposant sur une version locale l'information mutuelle,

$$l_\alpha(p) \;=\; -\frac{1}{|N_p|}\sum_{q\in N_p}\log\frac{p_{1,2}(I_1(p),I_2(p))^{1+\alpha}}{p_1(I_1(p))\,p_2(I_2(p))}, \tag{1.31}$$

où N_p est un voisinage centré sur le pixel p et $\alpha \in [0,1]$.

Fléau de la dimension Lorsque l'on considère des distributions empiriques, il est important de faire attention à la dimension de l'espace, le nombre de points nécessaire pour pouvoir fidèlement estimer la distribution augmentant exponentiellement avec la dimension. Autrement dit une augmentation de la dimension de l'espace a tendance à isoler les différents points s'y trouvant. Ce problème, connu sous le nom de « fléau de la dimension », est particulièrement critique lorsque l'on considère des mesures comme l'information mutuelle. En effet, pour un ensemble de n échantillons $(x_i, y_i) \in \mathbb{R}^{n+m}$, $i \in [\![1, N]\!]$, augmenter les dimensions n et m revient à isoler chaque x_i et y_i. La distribution jointe et ses marginales s'apparentent alors à de simples fonctions indicatrices

$$p_{1,2}(x,y) \;=\; \frac{1}{N}\mathbb{1}_{\{(x,y)\in\{(x_i,y_i)\}_i\}} \;=\; \begin{cases} \frac{1}{N} & \text{si } (x,y)\in\{(x_i,y_i)\}_i \\ 0 & \text{sinon} \end{cases}, \tag{1.32}$$

$$p_1(x) \;=\; \frac{1}{N}\mathbb{1}_{\{x\in\{x_i\}_i\}} \qquad \text{et} \qquad p_2(x) \;=\; \frac{1}{N}\mathbb{1}_{\{y\in\{y_i\}_i\}}, \tag{1.33}$$

et l'information mutuelle est alors triviale

$$\mathrm{I}(I_1,I_2) \;=\; \sum_{x,y} p_{1,2}(x,y)\log\frac{p_{1,2}(x,y)}{p_1(x)p_2(y)} \;=\; \sum_{i=1}^{N}\frac{1}{N}\log\frac{N^2}{N} \;=\; \log N. \tag{1.34}$$

Pour que l'information mutuelle ait un sens, il est souvent nécessaire de limiter le nombre de valeurs prises par chaque pixel, en quantifiant au préalable les images.

Modélisation par destin commun Dans la section 4.2, nous présentons un modèle a contrario de changements pouvant être appliqué à des images radiométriquement différentes. Les intensités locales des groupements de pixels sont ici confrontées à la distribution empirique globale, estimée sur les images entières. Seulement, contrairement aux méthodes précédentes, nous évaluons ici le nombre de pixels dont les intensités varient d'un sous-ensemble F, dans I_1, à un second sous-ensemble G, dans I_2, F et G correspondant à des sous-ensembles des valeurs possibles des pixels, ces derniers étant alors vus comme partageant un même « destin ». Cette forme de quantification, binaire dans chaque image, permet de ne pas subir le fléau de la dimension lors de l'estimation de la distribution globale.

Nous proposons une première expression de NFA, en faisant l'hypothèse que les pixels sont indépendants et en évaluant un voisinage de pixels E,

$$\mathrm{NFA}(E,F,G) \;=\; N^2\,\mathcal{B}\left(\left|E\cap D_1^F\right|, \left|E\cap D_1^F\cap D_2^G\right|, \frac{\left|D_2^G\cap D_1^F\right|}{\left|D_1^F\right|}\right), \tag{1.35}$$

où D_1^F et D_2^G correspondent aux pixels p tels que $I_1(p) \in F$ et $I_2(p) \in G$ et où $\mathcal{B}$ est la queue de la distribution binomiale, définie pour $p \in [0,1]$, $M \in \mathbb{N}^*$ et $k \in [\![0, N]\!]$, par

$$\mathcal{B}(M,k,p) \;=\; \sum_{i=k}^{M} \binom{M}{i} p^i (1-p)^{M-i}. \tag{1.36}$$

Nous montrons dans un second temps qu'il est possible de prendre en compte les corrélations spatiales des pixels en modélisant localement le nombre de pixels partageant le même « destin radiométrique » par une distribution en mélange de gaussiennes, de densité

$$f_{R,\alpha,\rho,\Sigma}(x) = \sum_{r=1}^{R} \alpha_r g_r(x) \qquad \text{avec} \qquad g_r(x) = \frac{1}{\sqrt{(2\pi)^K |\Sigma_r|}} e^{-\frac{1}{2}(x-\rho_r)^\intercal \Sigma_r^{-1}(x-\rho_r)}, \tag{1.37}$$

où R et α_r, μ_r, Σ_r (pour $r \in [\![1, R]\!]$) correspondent aux paramètres du modèle de mélange de gaussiennes. Nous obtenons alors un second NFA,

$$\mathrm{NFA}(E,F,G) \;=\; \frac{N^2}{2} \sum_{r=1}^{R} \left(\frac{\alpha_r g_r(|E \cap D_1^F|)}{\sum_{r=1}^{R} \alpha_r g_r(|E \cap D_1^F|)} \operatorname{erfc}\left(\frac{|E \cap D_1^F \cap D_2^G| - \mu_{y|x}^r}{\sqrt{2\sigma_{y|x}^r}} \right) \right), \tag{1.38}$$

où $\sigma_{y|x}^r$ et $\mu_{y|x}^r$ correspondent aux paramètres de la gaussienne g_r conditionnée suivant la première composante, c'est-à-dire

$$\sigma_{y|x}^r \;=\; \sigma_y^r - \frac{{\sigma_{xy}^r}^2}{\sigma_x^r} \qquad \text{et} \qquad \mu_{y|x}^r \;=\; \mu_y^r - \frac{\sigma_{xy}^r}{\sigma_x^r}(|E \cap D_1^F| - \mu_x^r), \tag{1.39}$$

en écrivant

$$\mu_r = (\mu_x^r, \mu_y^r) \qquad \text{et} \qquad \Sigma_r = \begin{pmatrix} \sigma_x^r & \sigma_{xy}^r \\ \sigma_{yx}^r & \sigma_y^r \end{pmatrix}. \tag{1.40}$$

Le modèle de détection de changements obtenu présente alors une certaine invariance aux transformations de changement de contraste des images, permettant de comparer des images radiométriquement différentes. Cependant, s'il parvient à éviter le fléau de la dimension, que subissent les modèles reposant sur les mesures d'information mutuelle, l'augmentation du nombre de canaux des images provoque une augmentation significative du temps de calcul qui le rend difficilement utilisable pour des images présentant un trop grand nombre de canaux.

1.3.3 Modèle de changements par reconstruction indirecte

Transformation d'images Une autre approche pour pouvoir comparer des images radiométriquement différentes consiste à effectuer une transformation des images pour les ramener dans un même « espace radiométrique », rendant possible une comparaison directe. Remarquons que l'on peut voir les différences radiométriques comme un décalage d'images au niveau des intensités, les transformations considérées ici pouvant alors être vues comme des recalages radiométriques.

Dans certains cas simples, l'utilisation de descripteurs d'images invariant par certaines fonctions de changements de contraste (comme le gradient) peuvent parfois suffire. Plusieurs méthodes proposent de recaler les intensités des deux images. Schaum et Stocker [SS04] proposent de recaler les intensités de manière affine en optimisant certaines mesures comme l'erreur quadratique moyenne ou la corrélation entre les images (en retrouvant alors un équivalent de la transformation MAD [NCS98]). Notons que ces méthodes sont particulièrement utilisées en imagerie hyperspectrale, la transformation utilisée réduisant souvent le nombre de bandes [NCS98].

Dans le contexte de l'étude d'une série temporelle bien recalée (un même pixel dans chaque image correspondant à la même position géographique) Carlotto [Car00] propose de reconstruire la dernière image comme une combinaison particulière (non linéaire) des images précédentes, en minimisant l'erreur quadratique moyenne. Remarquons que, en regroupant les images servant de référence en une seule image multispectrale (ou hyperspectrale pour des longues séries temporelles), cette méthode peut être considérée comme traitant des paires d'images.

Toutes ces méthodes présentent le désavantage de devoir disposer d'un grand nombre d'images ou de bandes de référence, la nouvelle image étant supposée pouvoir être exprimée à partir d'elles. Aussi, ces méthodes restent limitées à des « variations radiométriques » relativement simples, et ne sont a priori pas adaptées à la comparaison de deux images monospectrales radiométriquement différentes.

Utilisation de la redondance des structures Pour palier ce problème, nous proposons, dans la section 4.3, d'utiliser la redondance des structures, présente naturellement dans les images.

Cette redondance est très utilisée dans la littérature, par exemple en débruitage d'image par Buades et coauteurs [BCM05], mais aussi en détection d'anomalies par Zontak et Cohen [ZC10] ou par Davy et coauteurs [DEMD18] ou encore en régularisation de problèmes inverses par Peyré et coauteurs [PBC08]. Elle est popularisée par l'algorithme de débruitage « Non-local means », introduit par Buades et coauteurs [BCM05]. Il s'agit alors de débruiter une image I, en définissant pour tout pixel p,

$$\widetilde{I}_p \;=\; \frac{\sum_q w_{p,q} I_q}{\sum_q w_{p,q}} \qquad \text{avec} \qquad w_{p,q} \;=\; \exp\left(-\frac{\left\|I_{[p]} - I_{[q]}\right\|_{2,a}^2}{h^2}\right), \tag{1.41}$$

où $h^2 > 0$, $[p]$ et $[q]$ sont des patchs carrés centrés respectivement sur les pixels p et q, $I_{[p]}$ est la restriction de I au patch $[p]$ et où $\|X\|_{2,a}$ est la norme 2 de l'image X pondérée par un noyau gaussien de variance a^2,

$$\|X\|_{2,a}^2 \;=\; \frac{1}{2\pi a^2} \sum_p e^{-\frac{\|p\|_2^2}{2a^2}} \|X_p\|_2^2. \tag{1.42}$$

Zontak et Cohen [ZC10] proposent d'utiliser cette redondance pour la détection d'anomalie lorsque l'on possède une image de référence. Il s'agit alors, dans un premier temps, de reconstruire chaque patch $[p]$ de l'image étudiée I à partir des n patchs $[p_i]$

($i \in [\![1, n]\!]$) les plus proches dans l'image de référence J, en définissant

$$\overline{I_{[p]}} = \frac{\sum_{i=1}^{n} w_{p,p_i} J_{[p_i]}}{\sum_{i=1}^{n} w_{p,p_i}} \qquad \text{où} \qquad w_{p,p_i} = \exp\left(-\frac{\left\|I_{[p]} - J_{[p_i]}\right\|_2^2}{h^2}\right), \tag{1.43}$$

puis de moyenner simplement les patchs reconstruits pour obtenir une nouvelle image $\overline{I}$ sans anomalie, les anomalies étant alors détectés dans l'image résiduelle $|I - \overline{I}|$. Cette approche est reprise par Davy et coauteurs [DEMD18] dans pour la détection d'anomalie sans image de référence, en cherchant les patchs voisins les plus proches au sein de l'image étudiée. Notons que cela suppose que les anomalies sont suffisamment rares pour que les patchs voisins puissent être considérés sans anomalie.

Reconstruction indirecte Dans la section 4.3, nous proposons une adaptation du processus de reconstruction précédent, permettant, à partir d'une paire d'images, d'exprimer la structure de voisinage de la première, J, dans l'espace radiométrique de la seconde, I, en reconstruisant tout d'abord chaque patch $[p]$ avec

$$\overline{I_{[p]}} = \frac{\sum_{i=1}^{n} w_{p,p_i} I_{[p_i]}}{\sum_{i=1}^{n} w_{p,p_i}} \qquad \text{où} \qquad w_{p,p_i} = \exp\left(-\frac{\left\|J_{[p]} - J_{[p_i]}\right\|_2^2}{h^2}\right), \tag{1.44}$$

les patchs voisins $[p_i]$ les plus proches étant estimés sur J, h étant une constante positive fixée, puis en moyennant ces patchs pour reconstruire une image $\overline{I}$. En supposant que la structure de voisinage de J reste présente dans I (deux patchs plus proches voisins dans J restant proches dans I), nous montrons que l'erreur de reconstruction $\|\overline{I}_p - I_p\|_2$ peut être modélisée par une distribution gaussienne centrée de variance

$$\sigma_p^2 = \sigma^2 + \frac{\sigma^2}{|P_0|^2} \sum_{s \in \mathbb{Z}^2} \left(\sum_{\tau \in P_0} \sum_{[q] \in \mathcal{N}_{p-\tau}(u)} w_{p-\tau,q} \mathbb{1}_{\{q+\tau=s\}} \right)^2, \tag{1.45}$$

où P_0 est le patch centré en 0 et σ^2 correspond à la variance du bruit dans l'image I. Nous en déduisons un nombre de fausses alarmes,

$$\mathrm{NFA}(p, u, v) = \frac{N}{\Gamma(\frac{n}{2})} \Gamma\left(\frac{n}{2}, \frac{\left\|\overline{I}_p - I_p\right\|_2^2}{2\sigma_p^2}\right), \tag{1.46}$$

où n est le nombre de canaux de I et N son nombre de pixels.

Nous montrons par la suite qu'il est possible de prendre en compte d'éventuels décalages spatiaux entre les images, en permettant de légères translations des patchs $[p_i]$ au moment de la reconstruction, ainsi que d'éventuels patchs sous-représentés à l'aide d'un biais $\lambda > 0$ vers le patch central. Le processus de reconstruction s'obtient alors avec

$$\overline{I_{[p]}} = \frac{\lambda I_{[p]} + \sum_{i=1}^{n} w_{p,p_i+\tau_i^*} I_{[q]}}{\lambda + \sum_{i=1}^{n} w_{p,p_i+\tau_i^*}} \qquad \text{avec} \qquad \tau_q^* = \underset{\|\tau\| \leq 2\tau_{\max}}{\operatorname{argmin}} \left\|I_{[p]} - I_{[q+\tau]}\right\|_2, \tag{1.47}$$

où la constante $\tau_{\max}$ correspond au décalage maximal pris en compte. Le modèle de détection de changements obtenu permet, à l'instar du modèle précédent, de comparer des images radiométriquement différentes. Cependant, en se ramenant à un nombre réduit de voisins, il est ici possible de considérer des images possédant un grand nombre de canaux sans trop augmenter le temps de calcul nécessaire.

1.4 Détection supervisée de changements

Détecteurs supervisés La détection de changements est un problème mal défini, la notion de changements étant difficilement définissable précisément de manière générale. De ce fait, plusieurs méthodes reposent sur l'utilisation d'exemples de changements et de non-changements, et approchent la détection de changements comme un problème d'apprentissage statistique.

Cet apprentissage peut être effectué à plusieurs étapes du détecteur de changements. Il peut avoir lieu dans une phase d'extraction des descripteurs d'images. Il s'agit de trouver la meilleure représentation des images, pour faciliter leur comparaison. Zhao et coauteurs [ZWGL17] proposent d'utiliser un même réseau de neurones convolutif, appliqué à chaque image, pour extraire des descripteurs adaptés au problème de détection de changements, et une simple comparaison des descripteurs obtenus permet d'en déduire les changements.

Les données peuvent être utilisées pour apprendre directement un modèle de changement. Volpi et coauteurs [VTB+13] proposent de concaténer les canaux des deux images et d'utiliser l'image fusionnée pour apprendre une SVM. Plusieurs structures de réseaux de neurones sont envisagées pour apprendre directement ce modèle. Dai and Khorram [DK99] considère un réseau de neurones prenant directement en entrée les deux images, Mou et coauteurs [MBZ18] proposer d'appliquer successivement un réseaux convolutifs et un réseau récurent. On peut encore citer Zhan et coauteurs [ZFY+17] qui proposent d'utiliser des réseaux siamois. Ces approches de modélisation supervisée ne sont pas exclusives à la détection d'objets. Dans un contexte de détection d'objet, Inglada [Ing07] propose d'utiliser une SVM sur un grand nombre de descripteurs sélectionnés manuellement.

Un grand nombre de méthodes proposent d'utiliser des données labellisées en aval de la comparaison des deux images. Il s'agit alors majoritairement d'utiliser des méthodes de classification sur la différence des deux images, et on retrouve des méthodes déjà mentionnées dans la section sur les approches pixelliques. Bovolo et coauteurs [BBM08] proposent d'utiliser une SVM, Camps-Valls et coauteurs [CVGCMM+08] considèrent un « Support Vector Domain Description », Im and Jensen [IJ05] utilisent des arbres de décision. Notons que certaines méthodes proposent un cadre semi-supervisé, où seule une partie des données est labellisée [BP00].

Agrégation supervisée de détecteurs L'approche supervisée est parfois utilisée pour agréger plusieurs détecteurs de changements. Il s'agit alors de construire un détecteur de changements en combinant les apports de différentes représentations et modèles.

Plusieurs méthodes sont proposées dans la littérature. La première consiste simplement à effectuer un simple vote majoritaire, comme le propose Bo et Wang [BW05], en appliquant à chaque pixel de l'image l'ensemble des détecteurs et en conservant le résultat le plus représenté. D'autres méthodes plus élaborées existent. Le Hégarat et Seltz [LHMS04] utilisent la théorie des croyances de Dempster-Shafer. Cette théorie repose sur l'existence d'une fonction de masse, attribuant à chaque « état de changement » possible une valeur entre 0 et 1 et représentant une quantité de preuves penchant vers cet état. L'avantage de cette représentation est qu'elle permet facilement d'agréger des fonctions de masses en utilisant la règle de combinaison de Dempster, fournissant ainsi une manière naturelle d'agréger des détecteurs. Certaines méthodes, comme celle de Du et coauteur [DLG+12], reposent sur la théorie des mesures floues en utilisant l'intégration de Sugeno pour agréger les différents détecteurs.

Dans le chapitre 5.2, nous proposons d'agréger différents détecteurs de changements au sein d'un cadre d'apprentissage statistique classique, en utilisant un algorithme de *gradient boosting* reposant sur des arbres de décision. Nous obtenons ainsi un détecteur de changements fusionnant différentes modélisations et capable de s'adapter à un grand nombre de cas d'usages, sans nécessiter un trop grand nombre de données labellisées.

1.5 Plan du manuscrit

Chapitre 2 Dans ce chapitre, nous menons une réflexion sur la notion de changements dans une paire d'images satellites, en particulier dans le cadre de l'aide à la photo-interprétation. Le problème de la détection de changements étant mal défini et très peu discuté dans la littérature, nous profitons de ce chapitre pour examiner différentes approches de la notion de changements, dans le but de clarifier autant que possible le problème de la détection générique de changement et de construire les bases des différents modèles apportés dans les chapitres suivants. Nous menons cette réflexion en trois parties, tout d'abord, nous envisageant les changements à travers leur cause, ce qui nous amène à isoler les changements externes à la scène. Nous discutons ensuite de l'utilisation d'une interprétation sémantique pour évaluer la pertinence des différents changements, et enfin, nous introduisons une notion de contexte local ou global des changements, faisant le lien avec la détection d'anomalie.

Chapitre 3 Dans ce chapitre, nous introduisons dans un premier temps le modèle « a contrario », constituant le cadre statistique utilisé pour définir les différents modèles « primaires ». Ce cadre repose sur le calcul d'un nombre de fausses alarmes permettant le contrôle du nombre moyen de faux positifs. Dans un second temps, nous présentons des recalages spatiaux et en intensité, qu'il est souvent préférable d'effectuer en amont de la détection de changements, puis nous revenons rapidement sur les principaux descripteurs utilisés en traitement d'image.

Nous présentons alors successivement les deux premiers modèles de détection de changements. Le premier, qualifié de « Modèle par comparaison ponctuelle », repose sur la norme de la différence entre des descripteurs extraits des deux images. Nous proposons ainsi une version généraliste du modèle de détection de changements le plus répandu, tout en apportant une certaine robustesse aux faibles décalages entre

les deux images. Le second modèle, dit « Modèle par comparaison d'histogramme » se concentre sur la comparaison des distributions locales au travers de leur fonction de répartition. En utilisant une alternative au test de Kolmogorov-Smirnov, nous aboutissons à un nombre de fausses alarmes exact permettant de limiter grandement les effets d'éventuels décalages spatiaux entre les images. De plus, l'utilisation du cadre a contrario permet ici de simplifier le choix du seuil de détectabilité, problème récurrent des approches de détection de changements par comparaison des distributions locales.

Chapitre 4 Dans ce chapitre, pour présentons trois modèles a contrario de détection de changements supplémentaires, s'écartant de la comparaison habituelle des deux images. Pour le premier modèle, nous proposons une extension d'un algorithme de détection d'anomalie de Reed-Xiaoli à un couple d'images et des distributions locales modélisées par des mélanges de gaussiennes. Nous obtenons ainsi un modèle de détection de changements robustes aux décalages spatiaux, tout en conservant une résolution spatiale comparable à celle de l'image, résolution que perdent les modèles de détections de changements par comparaison des distributions locales.

Nous construisons les deux autres modèles dans le cadre de la détection de changements dans des paires d'images radiométriquement différents et donc non comparables directement. Le modèle par « destin commun » examine la présence d'anomalies locales dans la statistique jointe des deux images à comparer et permet donc de s'affranchir d'éventuelles transformations systématiques entre les deux images. Le modèle par « reconstruction indirecte », quant à lui, utilise une distance entre patchs pour construire une version de la première image dans le référentiel radiométrique de la seconde, ce qui permet ensuite de les comparer directement. De plus, ce modèle dispose d'une complexité temporelle relativement peu impactée par une augmentation de la dimension des images et permet donc l'évaluation de changements dans des séries temporelles de plus de deux images.

Chapitre 5 Après un rapide retour sur les différents descripteurs et détecteurs primaires qu'ils permettent d'obtenir, nous présentons dans ce chapitre le problème de l'agrégation de détecteurs comme un problème de classification supervisé avant de faire un rapide tour des principales méthodes de classification existantes. Nous discutons brièvement des post-traitement et de la pertinence des mesures de performances dans le cadre de la détection de changements.

Nous appliquons alors le procédé d'agrégation des modèles primaires dans le cas particulier d'un nombre très réduit de données labellisées provenant de deux paires d'images satellites et obtenons des résultats prometteurs sur les paires d'images testées, montrant ainsi la complémentarité des modèles de détection de changements développés.

1.6 Publications

Le processus de fusion exposé dans le chapitre 5 et utilisant les quatre premiers modèles de détection de changements a été présenté dans l'article de conférence

Aggregated primary detectors for generic change detection in satellite images (Vincent Vidal, Matthieu Limbert, Tudgual Ceillier et Lionel Moisan), *proceedings of IEEE International Symposium on Geoscience and Remote Sensing (IGARSS), pp. 59-62*, 2019

Le modèle de détection de changements par reconstruction indirect, introduit dans la section 4.3 a été présenté dans l'article

Détection indirecte de changements en imagerie satellitaire haute résolution (Vincent Vidal et Lionel Moisan), *Actes du Gretsi*, 2019

Réflexions autour de la notion de changement 2

Table des matières

2.1 Cause des changements 40
2.2 Segmentation sémantique pour la détection de changements 47
2.3 L'anormalité comme critère de pertinence 51
2.4 Point de vue adopté 55

Contenu Dans ce chapitre, nous menons une réflexion sur la notion de changement dans une paire d'images satellites, en particulier dans le cadre de l'aide à la photo-interprétation. Nous menons cette réflexion en trois parties, en envisageant tout d'abord les changements à travers leur cause, puis en discutant de l'utilisation d'une interprétation sémantique pour évaluer la pertinence des différents changements, et enfin en s'intéressant au contexte local et global de ces derniers.

2.1 Cause des changements

2.1.1 Introduction

Explicitation du problème On s'intéresse ici au problème de la détection de changements génériques, tout particulièrement dans le cas de l'aide à la photo-interprétation. Le problème est le suivant : étant données plusieurs images satellites d'une même zone géographique, prises à des instants différents, on s'intéresse au moyen d'accélérer le travail d'un photo-interprète devant étudier ces images. Pour cela, nous souhaitons détecter les différents changements intervenant au sein des images, dans le but de focaliser son attention sur les zones importantes. Les tâches d'un photo-interprète étant diverses et variées, nous chercherons à détecter les changements dans une optique la plus générale possible, et ce, pour pouvoir s'adapter facilement à un maximum de situations.

Pour simplifier, on se place dans le cadre d'une paire d'images et on cherche, a priori, à détecter tous les changements présents dans une des deux images par rapport à la seconde. La principale difficulté du problème de détection de changements réside dans la définition qu'il faut donner à la notion de changement, et, dans ce chapitre, nous menons une réflexion montrant la complexité de l'élaboration d'une telle définition.

Exemple préalable Pour mieux appréhender la complexité derrière la définition d'un changement, considérons un exemple réel, présenté dans la figure 2.1.

Le premier changement que l'on y voit est l'apparition, la disparition et le déplacement des camions au niveau des deux parkings et sur la route. On peut noter que certains véhicules changent de teinte, indiquant vraisemblablement un remplacement par un autre véhicule. Il est intéressant de remarquer l'apparition d'un camion de la même teinte que la route, particulièrement difficile à voir, si ce n'est grâce à son ombre. Un autre changement facilement visible correspond aux taches, non identifiées, disparaissant au niveau du petit parking et de la route avoisinante. On peut aussi apercevoir un assombrissement global de la route, qui pourrait être causé par une rénovation.

Ces changements, facilement visibles entre les deux images, ne sont cependant pas les seuls changements présents ici. Il existe tout d'abord un décalage spatial local entre les deux images. En effet, si certaines zones semblent être bien recalées, un même pixel correspondant à la même position dans les deux images, d'autres, comme au niveau de la route dans le bas de l'image, subissent des décalages de l'ordre d'une petite dizaine de pixels. En plus de cette translation, il apparaît dans les zones de végétation une variation non négligeable du contraste, difficilement explicable au premier abord. Pour finir, on peut noter l'existence d'un léger changement de contraste sur l'ensemble de l'image, ainsi que la présence de petites variations d'intensité au niveau de chaque pixel (dues au bruit de mesure).

Changements externes Dans cette liste, non exhaustive, des différences entre les deux images données dans la Figure 2.1, apparaît une distinction assez nette entre deux grands types de changements. D'un côté, on trouve les changements correspondant à des modifications physiques de la « scène », comme les apparitions et disparitions, mais aussi l'assombrissement de la route, qui pourrait correspondre à une rénovation

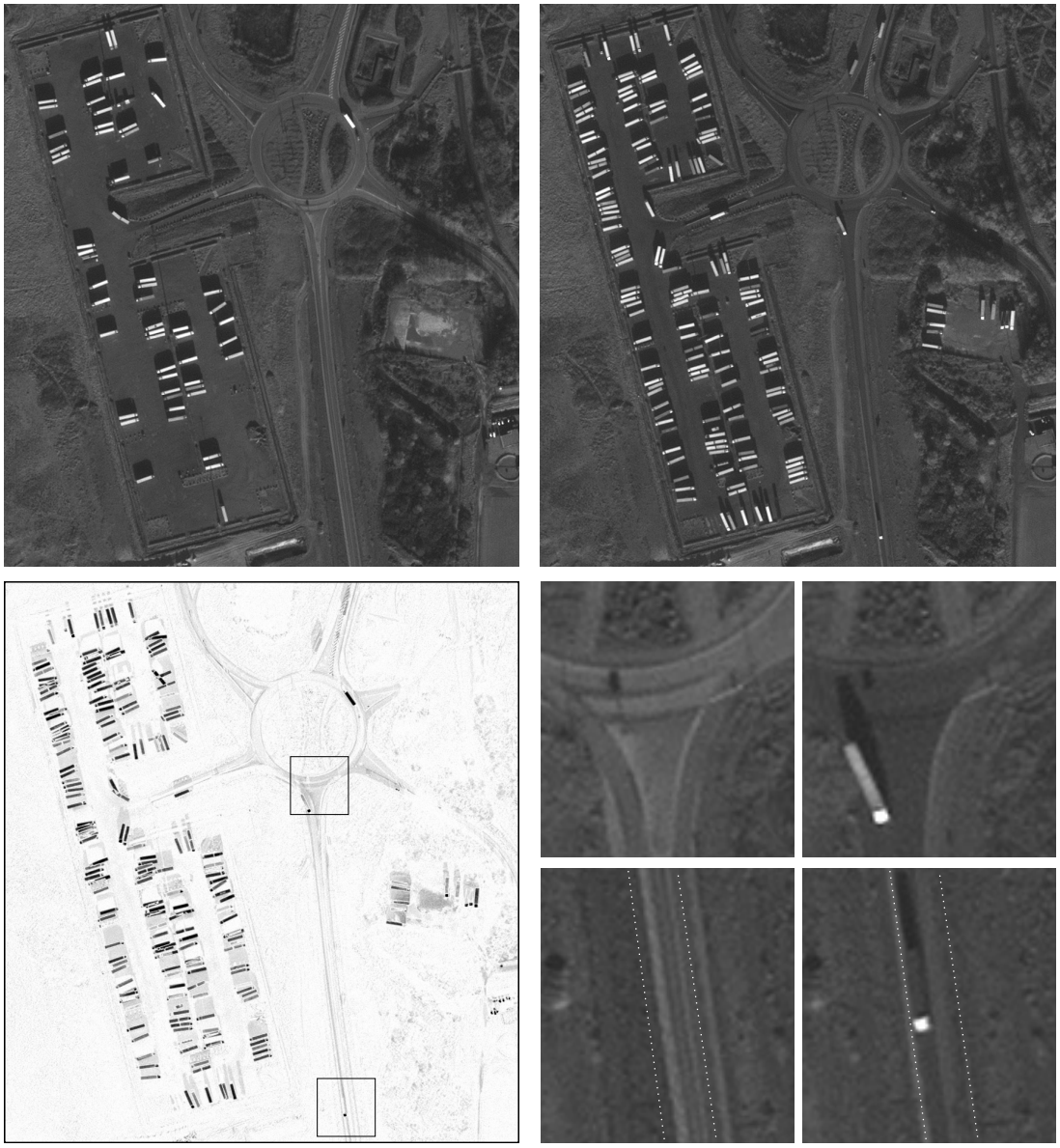


Figure 2.1 – Paire d'images acquises respectivement par les satellites « Pléiades-HR 1A » et « Pléiades-HR 1B ». On présente ici successivement les deux images, en nuance de gris, la valeur absolue de leur différence (d'autant plus foncée que la différence est importante), ainsi que des extraits des deux images, dont les positions sont encadrées en noir dans l'image de différence. Dans le deuxième extrait, nous avons affiché en pointillés blancs l'emplacement de la route, dans la seconde image, pour pouvoir mieux apprécier le décalage local présent entre les deux images. Il est présenté ici une scène rurale où apparaît et disparaît un ensemble de véhicules. On peut aussi observer un assombrissement de la route (représentée dans le premier extrait) et une tache blanche apparaissant dans le petit parking sur la droite de l'image. En plus de ces changements, on peut remarquer un certain nombre de changements « externes » à la scène et donc que l'on pourrait considérer comme non pertinents, comme un décalage local (visible dans le second extrait), un léger changement d'intensité global (visible dans l'image de différence, légèrement grisée) ou encore le bruit au niveau de chaque pixel.

de cette dernière. D'un autre côté, on trouverait les changements dus à des éléments présents en dehors de la scène, comme le bruit dû au capteur, la variation globale des intensités, les décalages spatiaux locaux, dus à un changement dans le positionnement du satellite, ou encore les déplacements d'ombres dus au déplacement du soleil (non présents dans la paire d'images présentée en exemple). Ces changements, que nous qualifierons « d'externes », sont la plupart du temps peu pertinents dans le cas de l'aide à la photo-interprétation puisqu'ils n'apportent a priori aucune information réelle quant à la scène en elle-même. Notons que ça n'est pas toujours le cas, une variation d'ombres d'un bâtiment, par exemple, pouvant indiquer un changement de hauteur de ce dernier.

2.1.2 Source de changements dans la chaîne de traitement

Pour avoir une meilleure vision des changements externes, revenons rapidement sur la chaîne de traitement que subit une image satellite, entre sa capture par un système optique et les utilisations des divers détecteurs de changements que l'on souhaite lui appliquer. Nous ferons notamment bien attention à envisager les principaux effets et artefacts que chaque traitement peut engendrer sur les images, et nous nous reposerons sur les différentes étapes telles qu'elles sont listées dans l'article de synthèse de Hussain et co-auteurs [HCC+13].

Bruit d'acquisition On s'intéresse ici au système optique capturant l'image. Notons que le cas des images radar est très différent et ne sera pas considéré ici. Une modélisation précise du système d'acquisition optique a été faite dans la thèse d'Aguerrebere [Agu14]. Pour résumer, chaque pixel est composé d'un capteur comptant les photons incidents sous forme d'électrons. La charge des électrons est alors mesurée, quantifiée et stockée. Pour un pixel, la valeur stockée, noté ici Z, peut être écrite sous la forme

$$Z = f\Bigg(\underbrace{g_{\text{out}}\Big(g_{\text{cv}}(\overbrace{C+D}^{\text{Nombre d'électrons}}) + N_{\text{reset}}\Big)}_{\text{Potentiel mesuré}} + \overbrace{N_{\text{out}}}^{\text{Bruit de transmission}} + \underbrace{Q}_{\text{Erreur de quantification}}\Bigg), \tag{2.1}$$

où :

- C représente le nombre d'électrons correspondant au flux de photons incidents, et D un nombre d'électrons s'excitant par erreur au niveau des impuretés du matériau (tous deux suivant des lois de Poisson) ;
- g_{cv} est un coefficient capacitif correspondant à la conversion de la charge en potentiel, N_{reset} est le bruit de lecture (assimilable à un modèle gaussien) et g_{out} est un coefficient de gain ;
- N_{out} est un bruit ajouté lors de la transmission du signal (assimilable à un modèle gaussien) et Q correspond à l'erreur faite lors de la quantification du signal (qui suit donc une loi uniforme) ;
- et f est la fonction de réponse de la caméra, le plus souvent linéaire jusqu'à un certain seuil, puis constante.

Dans le cas de scène suffisamment éclairée, les variables de Poisson de C et D sont très souvent modélisées par des lois gaussiennes et, dans la pratique, le bruit de quan-

tification Q est négligeable si l'on stocke la valeur du pixel sur un nombre suffisant de bits.

De cette manière, en dehors de la zone de saturation du capteur, on pourra considérer que Z suit une loi gaussienne. Ainsi, on trouve deux premiers types de changements externes, le bruit présent de manière indépendante dans les intensités de chaque pixel et la saturation de certains pixels, entre autres provoqués par des reflets solaires (dont on montre un exemple dans la Figure 2.4).

Corrections radiométriques Les corrections radiométriques concernent les corrections des effets et artefacts dus à l'appareil optique. Les différents paramètres capacitifs et de gain exposés dans l'équation (2.1) sont propres à chaque pixel et demande dont une calibration de l'appareil optique pour pouvoir homogénéiser la sensibilité des différents pixels. Avec le vieillissement et de la dégradation des composantes, il est toutefois nécessaire de prendre en compte des variations de sensibilité des capteurs au cours du temps, ainsi que d'éventuelles défaillances de certains pixels. Il faut aussi tenir compte des caractéristiques propres du capteur utilisé. Par exemple, dans le cas d'un capteur « Push broom », où chaque ligne de l'image est acquise successivement par une unique ligne de pixels, une inhomogénéité des pixels provoquent l'apparition de structures de lignes dans l'image (effet de « striping »), qu'il est alors nécessaire de corriger [Tei86]. Il s'agit enfin de corriger les effets de vignettage [KL12], qui correspondent à un assombrissement radial des bords de l'image.

Corrections atmosphériques Les corrections atmosphériques correspondent à la prise en compte des effets de l'atmosphère sur les intensités mesurées. En effet, en traversant l'atmosphère, les radiations sont en partie réfléchies et diffusées par les différentes couches de l'atmosphère ainsi que par la présence d'aérosol, de brume, etc, et voient donc leur intensité fortement impactée. Un modèle atmosphérique est alors utilisé pour corriger ces effets [VV99] et simuler une image prise sans atmosphère.

Corrections topographiques Les corrections topographiques correspondent à une prise en compte des effets de réflexion de l'onde solaire sur la surface terrestre, et notamment de la manière dont cette réflexion dépend de l'angle solaire incident, de l'orientation locale de la surface terrestre et de l'angle de vue. Cette correction se fait notamment en connaissant l'angle solaire, et à l'aide d'un « Modèle Numérique de Terrain », donnant plus ou moins grossièrement l'élévation du terrain en chaque point, ce qui permet ainsi d'estimer la normale à la surface. Notons que ces corrections sont négligeables lorsque les angles incidents sont faibles, la surface terrestre ayant alors un comportement lambertien, mais deviennent nécessaires lorsque ces angles augmentent [RKK09].

Ce modèle d'élévation n'étant pas parfait, il peut subsister des phénomènes de variation de contraste au sein des images. Notons que, si nous avons préféré les séparer, le plus souvent, ces trois corrections, radiométriques, atmosphériques et topographiques sont confondus sous l'appellation « Corrections radiométriques ».

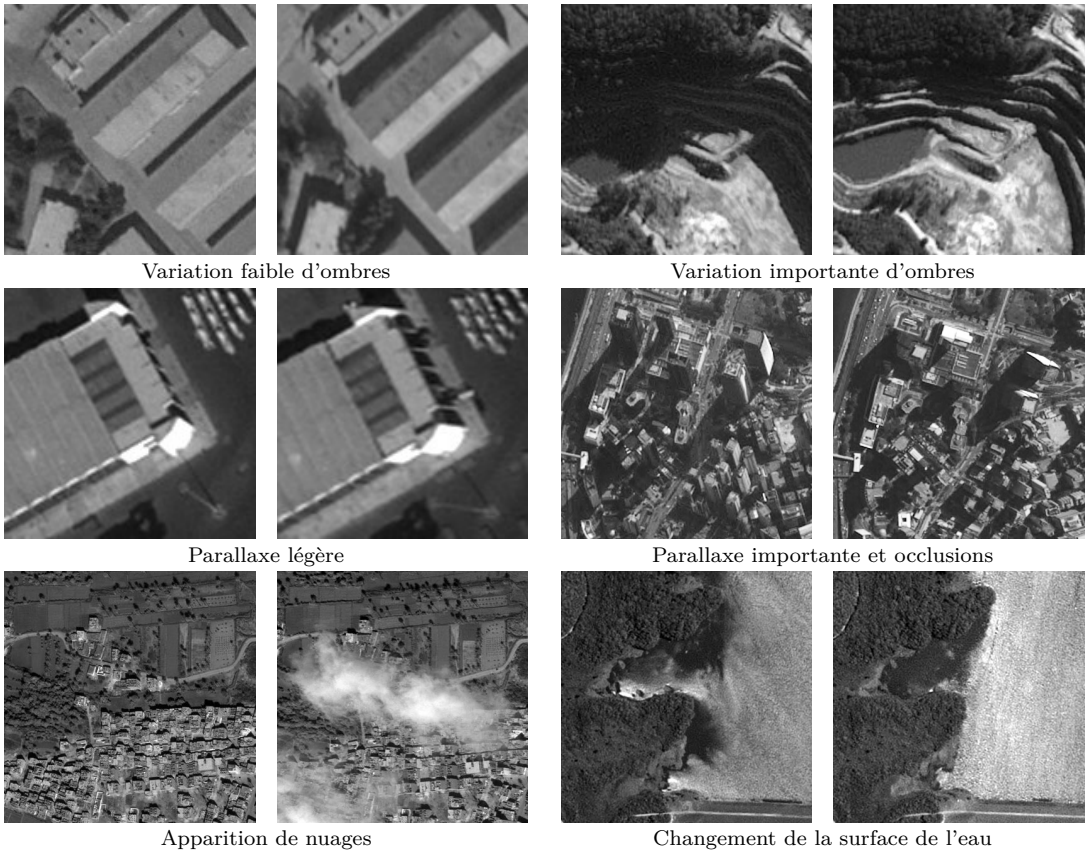


Figure 2.2 – Paires d'images prises par divers satellites, illustrant certains changements externes. On présente successivement deux variations (faible puis importante) d'ombre, des effets de parallaxe légère (au niveau de la façade ainsi que d'un poteau dans le coin inférieur droit), puis importante avec des occlusions, l'apparition d'un nuage et la modification de la texture de surface d'une étendue d'eau à cause du vent. Notons que, si les quatre premiers changements sont indéniablement externes à la scène, le caractère externe des deux derniers est plus discutable.

Orthorectification L'orthorectification est un procédé consistant à supprimer les effets de perspective dans les images satellites, pour se ramener à une image plane d'échelle constante. Cette transformation est estimée en s'appuyant sur les informations de prises de vue (notamment l'angle de vue et l'emprunte au sol) ainsi que sur le modèle d'élévation du sol, prenant aussi en compte le caractère sphérique de la terre. Les images sont de plus projetées dans un même système de coordonnées, utilisant la latitude et la longitude comme axes principaux.

Le modèle d'élévation étant imparfait, deux images orthorectifiées de la même zone peuvent présenter des déformations, une même position spatiale pouvant subir, entre les deux images, un décalage allant régulièrement jusqu'à une dizaine de pixels.

Par ailleurs, l'orthorectification consiste en un processus de réinterpolation de l'image et de ce fait, il pourra rajouter aussi des effets de repliement spectral (aliasing), de ringing ou de flou, en fonction de la méthode de réinterpolation choisie. Toutefois,

mis à part le flou, qui pourra s'apparenter à une variation de résolution au niveau des pixels, nous considérerons souvent ces derniers comme relativement négligeables.

Bilan des changements dus au traitement Ainsi, au vu de ces différentes corrections, nous nous attendrons éventuellement aux changements externes suivants,

— un bruit au niveau de chaque pixel
— des variations de contraste locales
— des décalages spatiaux locaux.

En ce qui concerne les différents artefacts (entre autres saturation, « striping », repliement spectral, flou), ils sont le plus souvent négligeables et les cas où ils apparaissent significativement sont rares.

Autres sources de changements externes Tous les changements externes ne sont cependant pas provoqués par des corrections imparfaites. Comme nous pouvons le voir dans les exemples de paires d'images données Figure 2.2.

Les différentes positions du soleil peuvent tout d'abord provoquer des changements de positionnement des ombres. S'ils ne sont pas problématiques pour des variations légères, cela devient un réel problème pour des variations plus importantes. En plus d'assombrir localement l'image, les ombres provoquent très souvent une perte locale du contraste, qui peut, pour une image trop fortement quantifiée (par exemple en étant codée sur 8 bits), complexifier grandement la détection de changements.

De même, la variation de position du satellite vis-à-vis de la scène étudiée provoque d'autres effets de parallaxe, plus complexes qu'un simple décalage local. Dans les cas d'une parallaxe légère, il s'agit principalement d'un allongement des façades des bâtiments, tandis qu'une parallaxe importante peut provoquer des occlusions, voire l'apparition de nouvelles façades (comme montré dans le quatrième exemple de la Figure 2.2).

Pour finir, certains objets, comme des nuages, des vagues, ou des textures d'ondulation à la surface de l'eau, peuvent n'apparaître que très peu pertinents dans l'étude de la scène et pourraient ainsi être considérés comme des changements externes.

2.1.3 Une séparation causes et effets difficile

Utilité des changements externes Comme nous l'avons fait remarquer précédemment, il peut sembler naturel d'écarter les changements externes, notamment dans le cadre de l'aide à la photo-interprétation.

Toutefois, d'une part, certains changements externes peuvent être intéressants. Ainsi, on peut imaginer que certains cas d'usage, en météorologie par exemple, demandent de s'intéresser aux nuages. D'autres part, certains changements externes peuvent mettre en valeurs des changements au sein de la scène, qui auraient été difficilement visibles sans. Nous avons affiché, dans la Figure 2.3, deux exemples illustrant l'importance que peut avoir, par exemple, la présence d'une ombre.

Une remontée compromise vers les causes Cependant, il faut remarquer que la notion de changements externes se base sur la cause des changements. Or, on ne peut

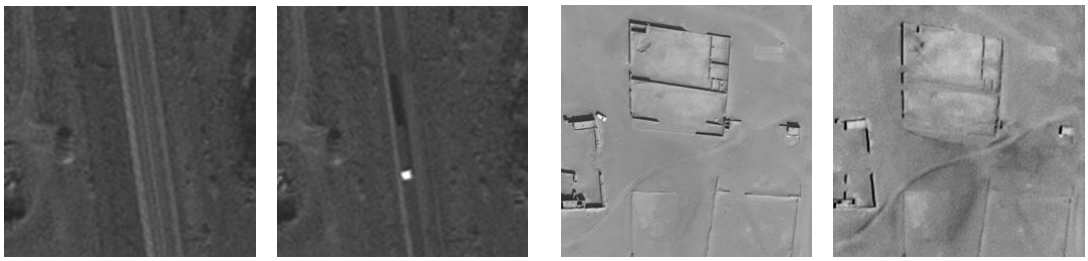


Figure 2.3 – Exemples illustrant à quel point les ombres peuvent être décisives pour la détection de changements. Dans la paire de gauche, la présence de l'ombre permet de repérer l'apparition d'un camion de même intensité que la route. Sans elle, il aurait été impossible de voir autre chose qu'un carré blanc. Dans la seconde paire, les ombres de la seconde image sont dirigées horizontalement et de ce fait, les murs horizontaux semblent avoir disparu. Il est alors possible de voir à quel point l'ombre permet de différencier un mur d'un léger relief.

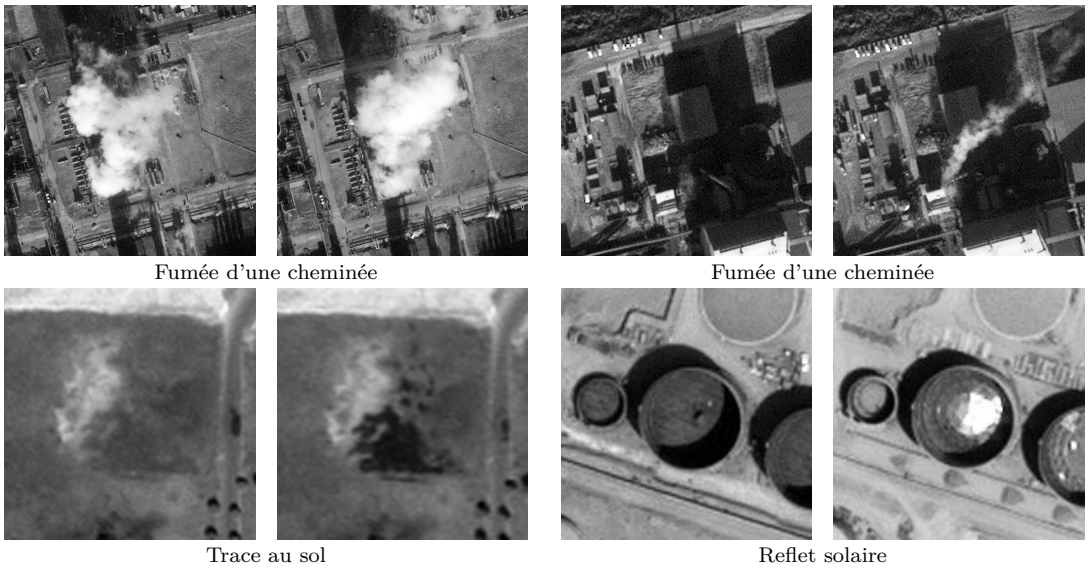


Figure 2.4 – Exemples de changements illustrant la difficulté d'interprétation de certains changements. Dans la première paire, la fumée d'une cheminée pourrait être confondue avec un nuage. Notons que c'est en remarquant que cette fumée est présente sur plusieurs images que l'on peut en fait conclure à la présence d'une cheminée. La deuxième paire présente aussi une fumée qui pourrait correspondre à un nuage. Dans la troisième paire, une tache au sol pourrait être confondue avec l'ombre d'un nuage. Dans la quatrième paire, le reflet solaire qui sature les pixels et l'ombre observée ne sont explicables qu'en comprenant la géométrie particulière du silo, le toit étant sphérique et pouvant descendre à l'intérieur de la structure.

appréhender les changements que par leurs effets, les causes n'étant que rarement mesurables. Il peut donc apparaître illusoire de vouloir définir un changement uniquement par cette dernière.

De plus, si remonter à la cause à partir des effets peut sembler faisable dans des cas simples, ça n'est pas du tout évident dans le cas général. Pour se convaincre de cela, il suffit de reprendre l'exemple de l'apparition du camion, présentée dans la Figure 2.3. En l'absence d'ombre, ce qui aurait pu arriver avec un soleil au zénith, il n'aurait pas été possible d'apercevoir autre chose que l'apparition d'un carré blanc. Par ailleurs, dans le cas des images satellites de résolution moyenne, un pixel correspondant à une zone d'une centaine de mètres, il est difficile de fournir une interprétation des changements observés sans informations extérieures, ces derniers correspondant souvent à des taches.

Même en très haute résolution, il est possible de trouver des changements possédant a priori plusieurs interprétations possibles, une tache pouvant ressembler à une ombre, un déplacement s'apparentant à un effet de la parallaxe, ou encore un reflet solaire s'identifiant à l'apparition d'un objet. Nous avons exhibé dans la Figure 2.4, des exemples de changement dont l'interprétation nous ont d'emblée semblé incertaine.

Un critère non suffisant Ainsi, si les changements externes apparaissent la plupart du temps comme peu pertinents, nous n'avons accès qu'aux effets et non aux causes des changements et il est donc impossible de discriminer parfaitement un changement présent dans la scène d'un changement externe, ce dernier étant défini d'un point de vu causal. Il n'est donc pas réellement envisageable de fonder la notion de changement entièrement sur le caractère externe de celui-ci.

2.2 Segmentation sémantique pour la détection de changements

2.2.1 Changements pertinents

Omniprésence des changements Pour continuer, intéressons-nous à un second exemple de paire d'images satellites, présenté dans la Figure 2.5. En dehors des changements dus à la parallaxe (visibles au niveau des différents bâtiments et dont on présente un extrait dans la Figure 2.2 sous le nom « Parallaxe légère »), on y trouve beaucoup d'apparitions et de disparitions d'objets de tailles diverses : des bateaux, des camions, des voitures, des conteneurs, ainsi que des taches. Remarquons alors que ces changements interviennent un peu partout dans l'image, en excluant la mer, particulièrement vide. En fait, comme c'est très souvent le cas, et comme on pourra encore le vérifier dans l'exemple suivant, donné dans la Figure 2.7, les changements sont presque omniprésents dans les images.

Nécessité de la pertinence Cette omniprésence des changements pose une interrogation : un détecteur de changements qui détecterait rigoureusement tous les changements serait-il réellement utile ? En effet, on s'intéresse ici à la détection de changements génériques, en particulier dans le cadre de l'aide à la photo-interprétation. Il s'agit de permettre aux photo-interprètes de mieux focaliser leur attention sur les

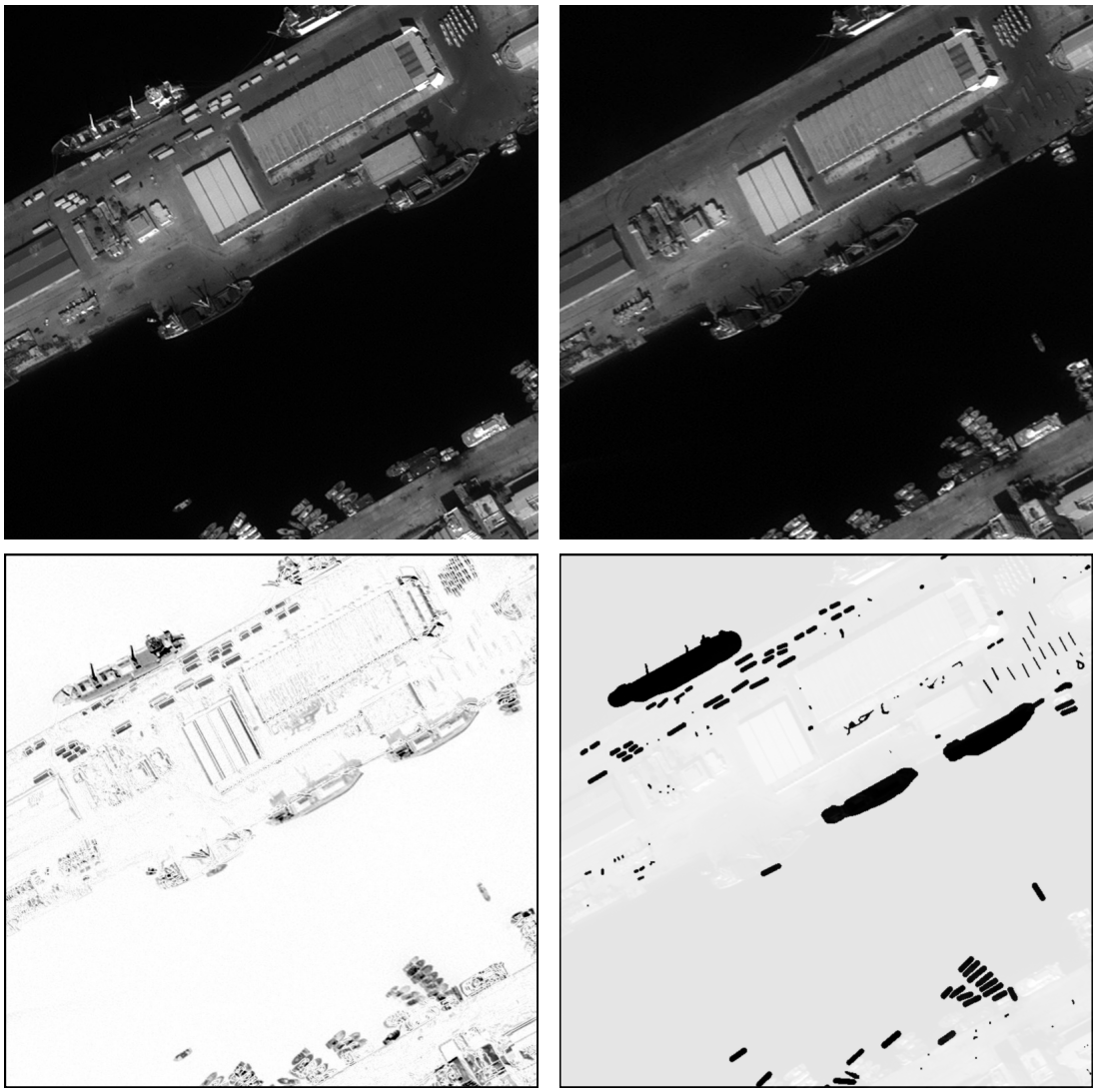


Figure 2.5 – Paire d'images acquises respectivement par les satellites « Pléiades-HR 1B » et « Pléiades-HR 1A ». On présente ici successivement les deux images, en nuance de gris, la valeur absolue de leur différence (d'autant plus foncée que la différence est importante), ainsi qu'une représentation schématique des différentes apparitions et disparitions d'objets présentes dans la scène. Dans cette dernière image, les changements sont représentés en noir, sur une version blanchie de la seconde image de la paire. Dans cette paire, on peut repérer un peu partout, en dehors de la mer, un grand nombre d'apparitions et de disparitions d'objets de diverses tailles : des gros bateaux, des plus petits, des camions, des voitures, des conteneurs, ainsi que diverses taches. Cette omniprésence des changements nous force alors à en écarter certains pour ne détecter que les changements les plus « pertinents ». Notons qu'un extrait de ces images est présenté dans la Figure 2.2 sous le nom « Parallaxe légère ».

différents changements, pour accélérer l'étude des images. Ainsi, en détectant tout le domaine, on mettrait l'accent sur l'intégralité de l'image, et on ne mettrait donc rien en valeur. Pour éviter cela, il serait ainsi nécessaire de mettre de côté certains changements, que nous jugerions non pertinents. Avec cette idée en tête, le problème n'est donc pas de définir ce qu'est un changement mais de réussir à définir une notion de pertinence sur l'ensemble des changements.

2.2.2 Utilisation de l'interprétation sémantique

Pertinence sémantique au cas par cas On s'intéresse alors ici aux manières de définir une notion de pertinence sur l'ensemble des changements. De manière naturelle, on pourrait essayer de se reposer sur une interprétation sémantique des différents objets présents dans la scène. En reprenant l'exemple donné dans la Figure 2.5, on pourrait ainsi considérer les petites taches comme non pertinentes, les apparitions de gros bateaux comme particulièrement intéressantes, et continuer ainsi de suite à définir la pertinence d'une apparition pour chaque type d'objet. Une telle méthode exigerait alors d'effectuer une sémantisation/classification préalable des différents objets présents.

Les méthodes « post-classification » Cette vision des choses est au coeur de certains algorithmes de détection de changements de la littérature, couramment qualifiés de « post-classification » ou encore de « class-object » [HCC+13, CHCW12]. Ces méthodes reposent sur une sémantisation de chaque image pour simplifier la comparaison des images à proprement parler. Les changements sont, dans ce cas, vus comme des transitions de pixels, d'une classe sémantique vers une autre. En procédant ainsi, le problème de détection de changements se ramène à un problème de segmentation d'image, beaucoup plus étudié.

Ces méthodes sont particulièrement utilisées pour la détection de changements dans le plan d'occupation du sol (ou « landcover »), et en résolution moyenne ou basse [YSLB05, GDMDD09, RAK+13, DBD06]. Chaque image est alors segmentée selon diverses classes (comme « forêt dense », « neige », « urbain », « cultures », « mer », etc) et, comme dit précédemment, chaque changement de classe sera considéré comme un changement par l'algorithme.

Notons que ces méthodes sont très peu appliquées à l'imagerie haute et très haute résolution, possédant une résolution pixellique inférieure au mètre, malgré les résultats de segmentation sémantique obtenus sur des scènes naturelles (hors satellite) [LSD15]. Nous donnerons dans la section suivante quelques pistes d'explication de ce phénomène.

2.2.3 Limitation de l'approche sémantique

Si une segmentation sémantique peut apparaître comme une représentation pertinente des images, dans le cadre de la détection de changements, elle possède cependant un certain nombre de limitations.

Figure 2.6 – Paire d'images acquises par un satellite de Planet Lab, illustrant la difficulté de distinguer les différents objets présents dans la scène. Il est en effet très difficile de distinguer un bâtiment d'un chantier ou d'un terrain en friche, ou encore un champ d'un jardin. L'utilisation d'une méthode « post-classification » semble ici compromise.

Une robustesse sémantique problématique Tout d'abord, la représentation sémantique présente un certain nombre de robustesses à l'égard de certaines variations dans les images, comme les variations de contraste, de teinte, et dans le cas des images satellites, variation d'orientation. Ces robustesses semblent particulièrement appréciables au vu des changements externes exposés dans la section 2.1.2, et permettraient de simplifier grandement le problème de détection de changements. Seulement, il n'est pas certain que toutes ces robustesses soient réellement souhaitables, car elles permettent de masquer, par exemple, des changements de teinte ou d'orientation, que l'on pourrait vouloir détecter.

Sémantisation totale impossible Si l'on souhaite à la fois profiter des robustesses de la sémantisation, sans masquer les variations au sein d'une même classe, il serait nécessaire d'y rajouter certaines informations, par exemple la couleur, la forme, la taille, l'orientation, l'état de délabrement, etc. Ces informations dépendent cependant assez fortement des objets auxquels elles se rapportent, les différentes formes d'une voiture se démarquant complètement de celles d'un bateau, d'un bâtiment ou encore d'un arbre. Il serait alors nécessaire de démultiplier le nombre de classes prises en compte dans la sémantisation.

Seulement, il faudrait aussi disposer de suffisamment de données labélisées, un aussi grand nombre de classes allant difficilement de pair avec des méthodes non supervisées. Si de telles données existent en grand nombre pour des images naturelles, prises au sol, le nombre d'images satellite labellisées reste aujourd'hui extrêmement limité, à l'exception peut-être de certaines classes d'objets très spécifiques comme les véhicules ou les bateaux. En pratique, il serait donc difficile d'obtenir des performances de segmentations sémantiques satisfaisantes, ce qui limiterait énormément l'intérêt de ces modèles « post-classification ».

Notons par ailleurs que cette sémantisation ne permettrait a priori pas de traiter des données dont l'interprétation sémantique est incertaine. C'est par exemple le cas en résolution moyenne, comme on peut le voir dans la Figure 2.6, où la majorité des objets apparaît comme des taches et une approche sémantique n'aurait alors qu'un intérêt très limité.

La sémantisation comme descripteur Il apparaît ainsi que se reposer entièrement sur une sémantisation de la scène pour résoudre le problème de détection de changements génériques n'est pas vraiment envisageable, et ce tant que l'on ne disposera pas au moins d'un nombre suffisamment important de données labélisées. Malgré cela, elle reste tout de même très utile dans certains cas particuliers, comme la détection de changements d'occupation des sols, traitée par les méthodes « post-classification ».

Il est intéressant de remarquer qu'une segmentation (sémantique ou pas) ne nous épargne pas de devoir comparer les couches obtenues à partir des deux images. La segmentation sémantique apparaît alors plus comme une forme « haut niveau » de descripteur d'image qu'une méthode de détection de changements. Par la suite, nous tenterons donc de nous détacher de cette vision sémantique, pour rester le plus général possible.

2.3 L'anormalité comme critère de pertinence

2.3.1 Pertinence contextuelle

Exemple préalable Continuons alors de réfléchir à la notion de pertinence d'un changement et à la manière dont on pourrait la caractériser, en s'intéressant à une dernière paire d'images, présentée dans la figure 2.7. Dans cette scène urbaine, avant toute chose, on peut noter la présence de nombreux changements correspondant à des effets de parallaxe et à des variations importantes d'ombres, particulièrement visibles dans l'image de différence.

On peut y trouver ensuite un grand nombre d'apparitions et de disparitions de véhicules, sur la route, dans les parkings, dans les quartiers résidentiels, mais aussi sur l'étendue de terre au milieu de l'image. On y trouve aussi quelques autres changements « uniques », détaillés dans la légende de la figure 2.7. Pour finir, il peut être pertinent de signaler que la première image est légèrement floue par rapport à la deuxième, et qu'il existe un changement de contraste local entre les deux images, notamment visible au niveau des toits dans le coin supérieur gauche.

Influence du contexte spatial Dans cet exemple, les différentes apparitions et disparitions de véhicules peuvent nous sembler peu pertinentes, très certainement à cause de leur important nombre au sein de l'image. À cause de sa surreprésentation, ce type de changement pourrait être considéré comme banal et donc moins intéressant pour un photo-interprète. Cette vision est renforcée par les autres changements presque « uniques », comme les deux reconstructions de bâtiments ou encore l'apparition du toit, et qui semblent alors particulièrement pertinents. Notons que cet argument pourrait aussi donner une raison supplémentaire de ne pas détecter les variations d'ombres, ces dernières étant présentes en trop grand nombre dans tout le domaine.

D'un autre côté, il peut être intéressant de remarquer que toutes les apparitions et disparitions de véhicules ne semblent pas avoir la même pertinence. Par exemple, il pourrait être important de détecter l'apparition du camion sur l'étendue de terre au milieu de l'image, ce dernier étant le seul à ne pas apparaître sur une route ou dans un parking. On comprend ici très bien qu'un véhicule apparaissant au milieu de nulle part

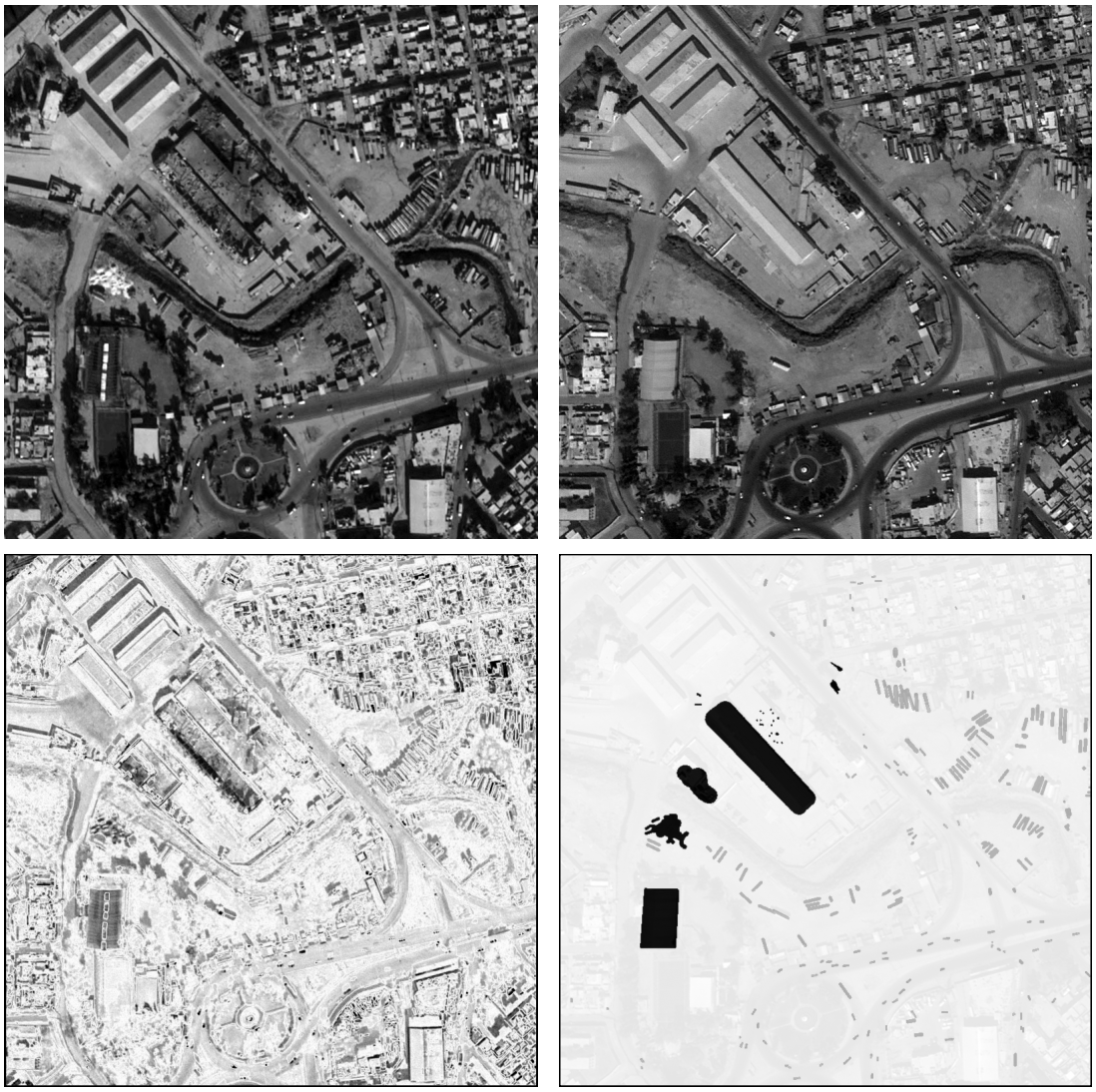


Figure 2.7 – Paire d'images acquises respectivement par les satellites « Pléiades » et « WorldView-2 ». On présente ici successivement les deux images, la valeur absolue de leur différence (d'autant plus foncée que la différence est importante), ainsi qu'une représentation schématique des différents changements présents dans la scène, en excluant ceux dus à la parallaxe et aux variations d'ombres, particulièrement importantes entre ces images et bien visibles dans l'image de différence. Dans la dernière image, nous avons fait figurer en gris toutes les apparitions et disparitions de véhicules, qui, à cause de leur surreprésentation, semblent perdre en pertinence face aux autres changements, affichés en noir et correspondant (de haut en bas) à l'apparition de taches sur le sol, à la reconstruction de deux bâtiments détruits, à la disparition d'un ensemble d'objets blancs difficilement identifiables et à l'apparition du toit d'un troisième bâtiment.

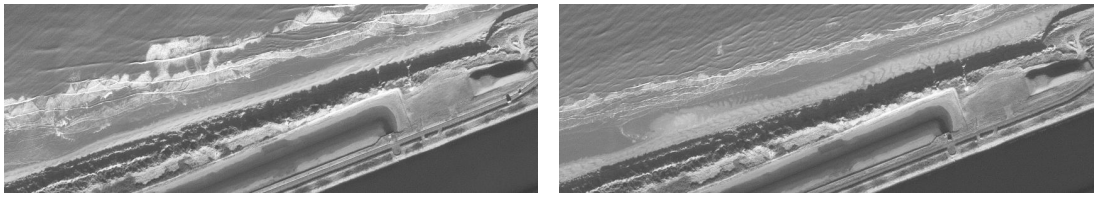


Figure 2.8 – Paire d'images illustrant la non pertinence de changements au sein d'une macro-texture. Ici les changements de texture à la surface de l'eau et dans la disposition des vagues au niveau du rivage peuvent nous apparaître comme des simples variations d'une même texture et ainsi nous sembler peu intéressants.

peut être plus pertinent, dans le cadre de la détection de changements, qu'un véhicule apparaissant au milieu d'autres véhicules, sur un parking. Ainsi, la pertinence d'un changement ne dépendrait pas seulement de l'objet auquel il se rapporte, mais aussi du contexte, aussi bien global (en s'intéressant notamment aux autres changements de même type dans l'image) que local.

Cette idée de contexte local expliquerait bien la non-pertinence ressentie dans le cadre de changements au sein d'une texture, et ce pour toutes échelles de textures :

— les micro-textures, comme les changements dans la végétation (dus à la parallaxe) de la Figure 2.1 ou encore un changement de texture à la surface de l'eau, comme dans la dernière paire de la Figure 2.2 ;
— et les macro-texture, comme le changement de fumée dans la Figure 2.4, les changements de véhicules dans les parkings des figures 2.1 et 2.7, ou encore le changement de vagues sur le rivage, présenté dans la Figure 2.8.

2.3.2 Rareté des changements pertinents

Changement global et changement local L'idée de contexte global, défini précédemment, nous amène alors à définir deux types de changements, les changements globaux et les changements locaux. Un changement global correspondrait à un changement influençant chaque pixel du domaine. Il pourra par exemple s'agir d'une variation globale d'intensité, de contraste, voire l'apparition d'un nuage recouvrant la majorité du domaine. À l'opposé, un changement local correspondrait à un changement n'intervenant que sur un voisinage spatial réduit de l'image, comme l'apparition d'un objet, n'apparaissant pas ailleurs dans l'image.

Notons que ces deux définitions n'englobent pas l'ensemble des changements. Les multiples apparitions de véhicules dans la figure 2.7 donnent un exemple de changement à la fois non global, ces derniers n'impactant qu'une petite quantité de pixels, et non locale, les apparitions intervenant en de multiples endroits.

Cette idée semble bien suivre l'optique de la détection de changements dans le cas de l'aide à la photo-interprétation : puisqu'il s'agit d'aider un utilisateur à voir les changements les plus pertinents, il semble logique de ne pas détecter un changement global, ce dernier étant suffisamment visible pour ne pas avoir besoin d'être souligné. De plus, comme nous l'avons dit précédemment, il peut être important d'éviter un trop grand nombre de détections pour ne pas diluer l'attention de l'utilisateur. Un

changement pourra donc sembler d'autant plus pertinent qu'il est local, c'est à dire qu'il correspond à un « type de changements » peu représenté dans les images.

Groupement de changements Il est important de noter que ces deux concepts dépendent énormément de la manière dont on regroupe les différents changements. En reconsidérant l'exemple donné dans la Figure 2.7, on peut considérer l'apparition du camion blanc sur l'étendue de terre au milieu de l'image comme un changement local, du fait qu'il est le seul à apparaître ailleurs que sur une route ou un parking. On pourrait cependant décider de regrouper toutes les apparitions et disparitions de véhicules et ne pas aboutir alors à un changement local.

Le problème de groupement d'objets, qui apparaît ici, est un problème complexe et beaucoup étudié, notamment au sein de la théorie de la Gestalt. Comme le montre Kanizsa dans son livre « La grammaire du voir » [Kan80], s'il semble exister de nombreuses règles régissant la manière dont la perception humaine regroupe les différents objets, aucune d'elles n'apporte de solution simple, la taille, la couleur, la forme, ou même la proximité ne prévalant sur les autres. Ainsi, si le critère de localité semble a priori prometteur pour définir un changement pertinent, il reste très dépendant des situations observées et donc très dépendant du cas d'usage.

Remarquons par ailleurs que les notions de localité et de globalité sont aussi très dépendantes de l'échelle, un changement global à très haute résolution pouvant perdre sa globalité à une résolution plus basse et même devenir local à faible résolution.

Lien avec la détection d'anomalies Il peut être intéressant de remarquer que l'utilisation d'une notion de localité pour définir un changement pertinent peut faire écho au problème de la détection d'anomalies. Dans les deux cas, il s'agit de détecter un phénomène rare, s'éloignant d'une situation jugée « normale », très subjectif et dépendant du cas d'usage. Notons que la détection de changements se confronte cependant au problème des incohérences spatiales, qui n'existe pas en détection d'anomalies. De plus, une anomalie au sein d'une paire d'images n'est pas nécessairement un changement (il suffit de s'imaginer le cas de deux images identiques présentant une anomalie). S'il n'est donc pas envisageable d'appliquer directement les méthodes de détection d'anomalies au problème de détection de changements, sans prendre en compte les spécificités de ce dernier, la vision d'un changement comme une anomalie de correspondance entre les deux images apparaît comme particulièrement intéressante.

Limitation de cette approche Il est important de noter que l'approche des changements pertinents en tant d'anomalies oriente fortement ces derniers vers les changements locaux. Et si cette approche peut sembler particulièrement pertinente, notamment dans le cadre de la détection de changements pour l'aide à la photo-interprétation, il est important de ne pas s'y restreindre. En effet, rien ne dit qu'un changement non-local (et non global), comme l'apparition des véhicules dans l'exemple de la Figure 2.7, n'intéresserait pas un photo-interprète pour un cas spécifique.

2.4 Point de vue adopté

2.4.1 Définition de la pertinence par l'utilisateur

Dans les sections précédentes, nous avons mené une réflexion autour de la notion de changement, en nous intéressant notamment à ce qui pourrait ou non justifier la pertinence d'un changement. Concluons à présent en précisant le point de vue adopté pour aborder le problème de la détection de changements génériques, sur lequel sont basés les différents modèles.

Pertinence au cas par cas Tout d'abord, et comme nous l'avons déjà mentionné, de par l'omniprésence des changements entre les images, le problème consiste ici à ne détecter que les changements pertinents. Par ailleurs, si certains changements, issus d'éléments externes à la scène, pourraient être considérés comme nécessairement non pertinents, il est dans la pratique très difficile, voire impossible, de pouvoir réellement attribuer les effets observés à des causes précises.

De plus, si l'idée de détecter des changements locaux nous semble particulièrement appropriée, la grande variabilité dans les cas d'utilisation d'un détecteur de changements nous pousse à ne pas nous fonder uniquement sur ce concept. Nous ne nous donnons donc aucun apriori sur ce qu'est ou n'est pas un changement, tant sur le côté externe, sémantique ou local. Nous supposerons ainsi que la pertinence des différents changements dépend directement du cas d'usage, ce dernier étant fixé par l'utilisateur. De ce fait, il s'agit ici de construire un détecteur de changements qui puisse s'adapter à de multiples définitions d'un changement pertinent.

Notons toutefois que le bruit présent dans l'image, au niveau de chaque pixel, ne sera, lui, jamais considéré comme un changement pertinent.

Algorithme automatique ou algorithme interactif Pour prendre en compte la grande variabilité envisagée dans la définition d'un changement pertinent, nous nous sommes tournés vers les méthodes d'apprentissage statistique. Il est alors nécessaire de s'interroger sur la manière qu'a l'utilisateur de transmettre sa définition de la pertinence d'un changement.

Le temps accordé à un photo-interprète pour effectuer une tâche donnée est souvent assez limité (parfois de l'ordre de quelques heures) et de ce fait, nous supposerons comme très réduit le nombre d'apports de l'utilisateur, au sein du processus d'apprentissage. Par ailleurs, on peut imaginer plusieurs types d'interactions entre l'utilisateur et l'algorithme.

— Tout d'abord, l'utilisateur pourrait sélectionner, à sa convenance, un détecteur de changements parmi plusieurs mis à sa disposition, ou, de manière similaire, choisir un jeu de paramètres donnant le détecteur qui, selon lui, correspondrait le mieux à sa tâche. Si cette méthode présente l'avantage de ne pas nécessiter d'apprentissage et peut donc être appliquée très rapidement, elle possède cependant plusieurs grosses limitations. Tout d'abord, cela demande à l'utilisateur d'avoir une connaissance précise des différents détecteurs ou paramètres, complexifiant ainsi beaucoup l'utilisation du programme. Par ailleurs, au vu de la difficulté apparente de la tâche, on peut imaginer qu'il serait nécessaire de

considérer de nombreux détecteurs ou paramètres et on peut alors douter qu'un détecteur choisi à la main puisse être optimal.

— D'un autre point de vue, on peut imaginer que l'utilisateur choisisse dans l'image quelques exemples de changements pertinents et non pertinents pour sa tâche et les transmette alors à l'algorithme, pour construire par apprentissage statistique supervisé un détecteur de changements, le plus adapté possible.

— Enfin, on pourrait envisager que l'algorithme propose lui-même un certain nombre de paires d'images que l'utilisateur devra classifier en changement pertinent ou non, le processus s'apparentant à de l'apprentissage actif.

Si ce dernier point pourrait être réellement pertinent dans le cadre de l'aide à la photo-interprétation, nous souhaitons nous concentrer sur la détection de changements et non sur l'apprentissage actif, et de ce fait, nous nous placerons dans le cadre, plus simple, de l'apprentissage supervisé classique du second cas. On supposera ainsi disposer d'un certain nombre de données labélisées par l'utilisateur, sans possibilité d'obtenir plus d'informations de sa part.

2.4.2 Hypothèses du problème

Type du résultat Avant toute chose, intéressons-nous au type de sortie que l'on attend des différents détecteurs. Comme dans tout problème de détection d'objets, on peut avoir plusieurs formats de résultat. Tout d'abord, l'algorithme peut rendre, en sortie, une carte de segmentation, explicitant pour chaque pixel s'il fait partie ou non d'un changement pertinent. Toutefois, dans la pratique, une telle précision quant à la forme des changements n'est souvent pas nécessaire, l'utilisateur pouvant très bien apprécier ces dernières, à l'oeil, directement sur les images et de simples boîtes englobant les changements suffiraient. Dans certains cas, toutefois, une segmentation précise des changements est nécessaire, par exemple pour l'évaluation des dommages lors d'une catastrophe naturelle, où il peut être important de connaître l'étendue spatiale des dommages.

De ce fait, nous nous sommes placés dans le cadre de la segmentation des changements, où à chaque pixel doit être attribuée une information de pertinence.

Par ailleurs, dans le cas d'alertes automatiques, s'il est important de disposer d'une information binaire sur la présence ou non de changements pertinents, dans le cadre de l'aide à la photo-interprétation, il peut être intéressant de pouvoir disposer d'une valeur continue de pertinence pour les changements détectés. Cela permet à l'utilisateur d'étudier directement les changements les plus pertinents, ou au contraire, de pouvoir vérifier si les changements détectés les moins pertinents ne sont pas des faux positifs. De ce fait, on suppose ici que les détecteurs de changements fournissent des valeurs continues pour la pertinence de chaque pixel, en représentant ainsi le résultat de la détection sous la forme d'une image à valeurs réelles.

Changements pris en compte séparément Un certain nombre de changements mentionnés précédemment font souvent l'objet d'études spécifiques.

Il y a tout d'abord les changements causés par une forte parallaxe, apparaissant principalement dans les milieux urbains. On peut voir quelques exemples de ces chan-

gements dans la Figure 2.2. Ces changements sont vraiment difficiles à prendre en compte et nécessiteraient certainement pour cela une modélisation en 3D de la scène. Par ailleurs, ils vont de pair avec des phénomènes d'occlusion, empêchant toute conclusion quant à la présence de changements pour les zones non observées dans l'une des images. Pour vraiment prendre en compte les changements dus à une parallaxe forte, il serait alors aussi nécessaire de détecter les occlusions séparément du reste, et de les marquer comme des « zones de non-information ».

Un autre changement fréquent est l'apparition de nuages. Si les nuages sont souvent faciles à détecter, on pourrait les considérer aussi comme une source d'occlusion. Ainsi, de la même manière que précédemment, on pourrait vouloir les détecter séparément.

Par ailleurs, on peut aussi vouloir prendre en compte séparément les grandes variations d'ombres, comme on peut en voir dans la Figure 2.2. Le problème de la détection et de la correction des ombres est un sujet beaucoup traité dans la littérature [Dar05, SYW+13, KMB15], présentant de bons résultats dans la majorité des cas. D'une part, et comme nous l'avons dit précédemment, les ombres correspondent à un changement d'illumination de la scène et de ce fait, ne devraient a priori pas être considérées comme pertinentes. De plus, la dégradation du contraste qu'elles provoquent sur l'image pourrait masquer certains changements et de ce fait être considérée comme une occlusion. Il serait alors pertinent de vouloir au moins les détecter séparément, voire de les « corriger ». D'autre part, comme nous l'avons déjà montré et illustré précédemment, dans la Figure 2.3, dans certains cas, les ombres permettent de trahir la présence d'un objet, difficilement visible autrement, et il pourrait ainsi être pertinent de détecter l'apparition d'une ombre. Il peut ainsi être raisonnable de vouloir traiter le cas des grandes variations d'ombres séparément.

Ainsi, dans toute la suite, nous ne prendrons donc pas en compte ces changements (associés aux nuages, à la parallaxe forte et aux ombres) et nous supposerons qu'ils ont été traités en pré ou post-traitement.

Non-utilisation des descripteurs sémantiques Comme nous l'avons mentionné précédemment, l'utilisation de descripteurs sémantiques reste trop spécifique à certains cas d'usage. Par la suite, nous ne considérerons donc pas de descripteurs sémantiques.

2.4.3 Images synthétiques utilisées en test

Pour évaluer les différents modèles de changements, il est nécessaire de pouvoir les évaluer sur différentes paires d'images. Seulement, et comme nous l'avons mentionné précédemment, les images réelles sont le plus souvent très difficiles à décrire et présentent de trop nombreux types de changements différents en même temps. En utilisant des images réelles, il sera donc très difficile de pouvoir évaluer séparément la manière dont un modèle appréhende les différents types de changements. Nous avons donc construit plusieurs paires d'images synthétiques, chacune représentant un type de changement différent, afin de faciliter l'évaluation des différents modèles. Décrivons brièvement ces différentes paires, représentés dans la figure 2.9.

1. La première paire représente une scène portuaire, où apparaissent plusieurs camions et disparaît un véhicule. En dehors du bruit, ces apparitions et dispa-

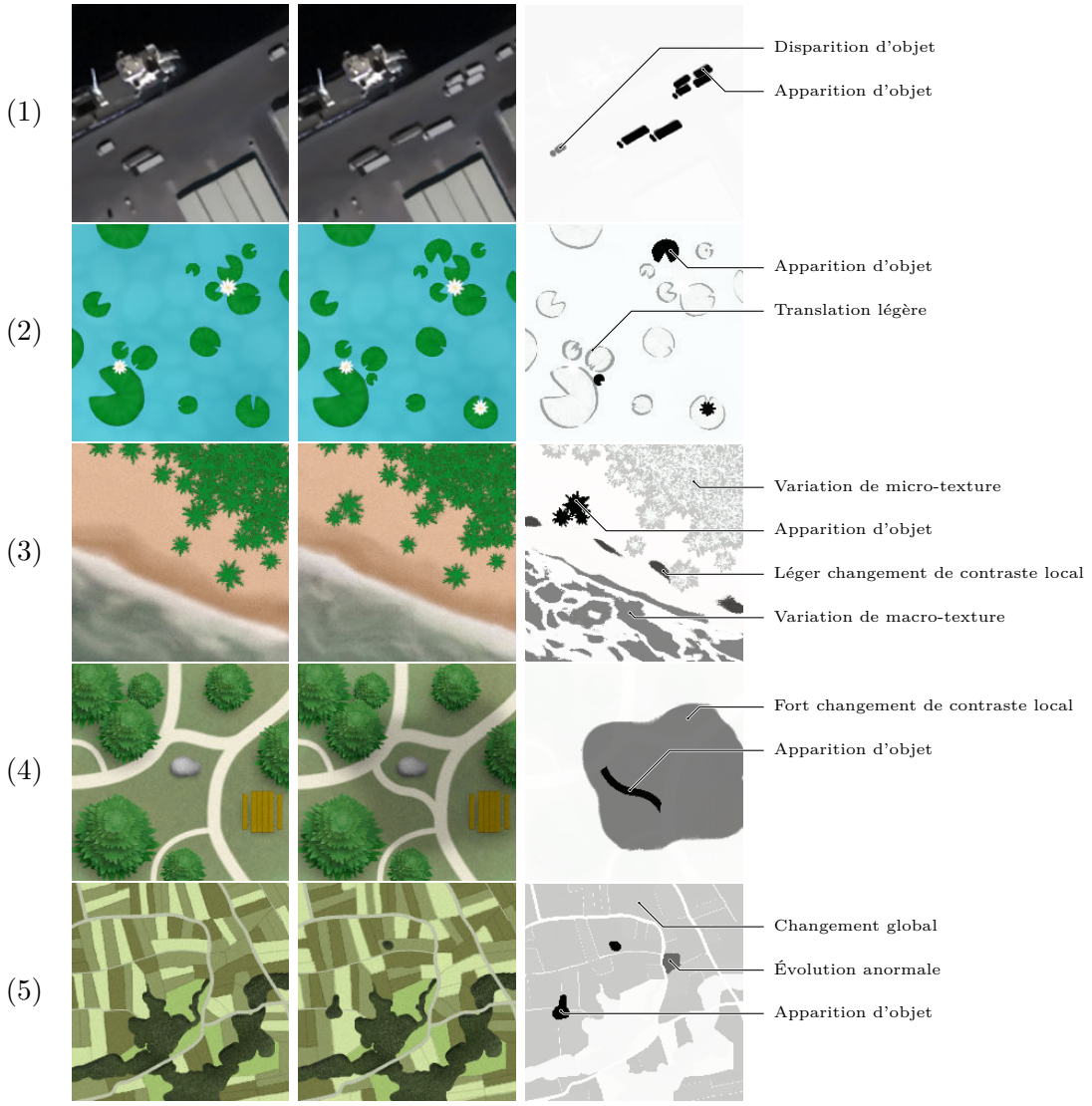


Figure 2.9 – Paires d'images synthétiques utilisées pour évaluer les différents modèles de changements. Chaque ligne correspond à une paire et présente successivement la première et la deuxième image de la paire, suivis d'une image en noir et blanc indiquant grossièrement les emplacements des changements, chaque nuance de gris correspondant à un certain type de changement.

ritions constituent les seuls changements présents. Notons qu'un des camions qui apparaît a sensiblement la même intensité que celle du sol.

2. La deuxième paire introduit des translations faibles et représente une étendue d'eau sur laquelle sont représentés de manière stylisée des nénuphars et des fleurs. Ici, trois objets (une fleur et deux nénuphars) apparaissent, et chaque nénuphar se translate légèrement.
3. La troisième paire introduit des variations de textures et représente une plage avec des palmiers au bord de la mer. On y compte quatre types de changements, à savoir tout d'abord l'apparition d'un petit groupe d'arbres. La « forêt » présente une variation de micro-texture, tandis que la mer présente une variation de macro-texture. Pour finir, on trouve un léger changement de contraste au niveau de la zone de transition entre mer et sable.
4. La quatrième paire introduit un changement local de contraste et représente des sentiers entourés de quelques arbres, d'une table et d'un rocher. On trouve deux changements, l'apparition d'un sentier au centre de l'image, ainsi que celle de l'ombre d'un nuage provoquant un changement de contraste local.
5. La cinquième paire introduit un changement global et représente un ensemble de champs évoluant de manière simultanée. Ici, les intensités des champs respectent une rotation des intensités, les intensités claires devenant moyennes, celles-ci devenant foncées et ces dernières devenant claires. On trouve alors trois types de changement. On a tout d'abord l'apparition de deux taches, puis la rotation des intensités des champs. Enfin, une des parcelles ne respecte pas la même évolution que les autres, en passant d'une intensité moyenne à une intensité claire.

Détection de changements par mesure directe 3

Table des matières

3.1 Formalisme a contrario 62
3.2 Prétraitement et conventions prises 70
3.3 Descripteurs d'images 76
3.4 Modèle de changements par comparaison ponctuelle 88
3.5 Modèle de changements par comparaison d'histogramme 110

CONTENU Ce chapitre introduit deux premiers modèles primaires de changements, en explicitant au préalable le cadre statistique a contrario ainsi que les diverses notations, prétraitement et descripteurs d'images utilisés. Pour le premier modèle primaire, nous reprenons le modèle le plus répandu, évaluant simplement la différence des deux images dans un modèle a contrario, en proposant un moyen de limiter les effets d'un décalage entre les images. nous proposons un deuxième modèle basé sur une comparaison locale des histogrammes à l'aide de la distance de Kolmogorov-Smirnov.

3.1 Formalisme a contrario

3.1.1 Définitions et propriétés

3.1.1.1 Définition formelle

Pour construire les différents détecteurs primaires, nous allons nous placer dans le cadre statistique « a contrario », introduit par Desolneux et coauteurs [DMM00, DMM07]. Ce cadre est basé sur le principe de Helmholtz, qui stipule que « on ne perçoit rien dans une image de bruit » [Att54], ou plus précisément que l'œil perçoit des structures lorsque ces dernières s'éloignent trop grandement de l'aléatoire. La méthodologie a contrario consiste alors à donner à chaque structure possible une forme de significativité vis-à-vis d'une modélisation de l'aléatoire. Cette modélisation, souvent appelée modèle H_0, sera nommée ici « modèle naïf », et représentera l'idée de ce qui est « normal », c'est-à-dire ici sans aucun changement. L'intérêt du formalisme a contrario est que l'on ne repose sur aucune modélisation de changement. Son utilisation permet donc de limiter grandement la quantité d'informations a priori nécessaire pour construire un détecteur.

Nombre de Fausses Alarmes D'un point de vue formel, on observe ici des données au sein desquelles nous allons chercher à détecter des structures. Ces données sont ici modélisées par une variable aléatoire X, et on note $\mathcal{X}$ l'espace dans lequel elle évolue. Dans notre cas, X modélisera la paire d'images étudiée dans laquelle nous cherchons à détecter des changements et ainsi on peut se donner $\mathcal{X} = \mathbb{R}^N$, pour N le nombre de pixels dans la paire d'images. Le modèle naïf, correspondant à la situation « normale », sera alors vu comme une distribution H_0 sur cet espace $\mathcal{X}$, une distribution que sera censée suivre X s'il n'existe aucun changement. Dans la pratique, il n'est pas nécessaire de posséder un modèle très précis et, le plus souvent, une modélisation simple suffit, ce qui explique l'appellation « modèle naïf ». On définit alors la propriété de fausse alarme de la manière suivante.

Définition 3.1.1 (Nombre de fausses alarmes [DMM07])**.** *Pour $\mathcal{X}$ un ensemble, H_0 une distribution sur cet ensemble et $\mathcal{I}$ un ensemble fini, on appelle Nombre de Fausses Alarmes associé au modèle naïf $(\mathcal{X}, \mathcal{I})$, souvent réduit simplement à NFA, toute fonction F définie sur $\mathcal{I} \times \mathcal{X}$ à valeur dans $\mathbb{R}^+$ et vérifiant, pour tout $\varepsilon > 0$, l'inégalité*

$$\mathbb{E}_{X \sim H_0}\Big[\,\mathrm{Card}\,\{i \in \mathcal{I}, F(i, X) \leq \varepsilon\}\,\Big] \;\leq\; \varepsilon. \tag{3.1}$$

Cette inégalité sera nommée par la suite « propriété du nombre de fausses alarmes ».

L'ensemble $\mathcal{I}$ représentant ici l'ensemble de toutes les structures potentielles, que l'on souhaite éventuellement détecter. Très souvent, dans notre cas, ces structures correspondront à des positions dans l'image et de ce fait, on associera la plupart du temps l'ensemble $\mathcal{I}$ à l'ensemble des positions pixelliques.

Notons que cette propriété est parfois écrite en considérant une variable X_i par structure $i \in \mathcal{I}$, en demandant alors la propriété $\mathbb{E}\big[\,\mathrm{Card}\,\{i \in \mathcal{I}, F(i, X_i) \leq \varepsilon\}\,\big] \leq \varepsilon$. Ces deux versions reviennent bien évidemment au même.

Significativité d'une structure La valeur d'un NFA représente une sorte de significativité des différentes structures vis-à-vis du modèle naïf. Pour F, un nombre de fausses alarmes, et des données observées $x \in \mathcal{X}$, la structure i est dite ε-significative lorsque $F(i, x) \leq \varepsilon$, cette structure étant d'autant plus significative que la valeur du NFA associé est proche de zéro.

Avant de continuer, revenons rapidement sur la propriété dite du nombre de fausses alarmes, donnée dans l'équation (3.1). Il s'agit de compter le nombre moyen de structures dont la significativité dépasse une certaine valeur ε. Pour respecter la propriété, ce nombre moyen ne doit alors pas dépasser ε, autrement dit, en moyenne, il doit exister au plus ε structures qui soient ε-significatives, ou, d'un autre point de vue, une structure ε-significative ne doit apparaître en moyenne pas plus de ε fois dans une image de bruit (bruit correspondant au modèle naïf). Un NFA permet ainsi de donner à chaque structure une majoration du nombre moyen de structures ayant au moins la même « qualité de structure ».

Détecteur a contrario La construction d'un détecteur à partir d'un nombre de fausses alarmes se fait alors directement en seuillant ce dernier et en détectant ainsi toutes les structures dont la significativité est au moins égale au niveau ε du seuil choisi. En procédant ainsi, on s'assure théoriquement de majorer par ε le nombre moyen de fausses alarmes, c'est-à-dire le nombre moyen de changements détectés à tort. Si la valeur $\varepsilon = 1$ est souvent prise comme valeur de seuil (voir [DMM07]), dans la pratique, nous avons considéré une valeur plus faible, par exemple $\varepsilon = 10^{-10}$, permettant de rajouter une certaine robustesse aux modèles. Notons que cette valeur peut sembler très proche de 0 au premier abord mais s'avère en réalité relativement peu contraignante dans le cadre d'un détecteur a contrario.

Probabilité de fausse alarme et p-valeur Avant de s'intéresser à la manière dont on construit un nombre de fausses alarmes, il peut être intéressant d'introduire la notion de p-valeur. La p-valeur est définie comme étant la probabilité d'obtenir un résultat au moins aussi important que la mesure observée, sous l'hypothèse d'un modèle H_0, c'est-à-dire

$$\text{p-valeur}(y) \;=\; \mathbb{P}\big(\, Y \geq y \,\big)\,, \tag{3.2}$$

pour Y une variable aléatoire réelle et $y \in \mathbb{R}$. Il est intéressant de remarquer que cette définition aboutit à une propriété de sous-uniformité de la fonction de p-valeur, similaire à celle exigée d'un NFA dans l'équation (3.1),

$$\forall \varepsilon \geq 0, \quad \mathbb{P}_{X \sim H_0}\big(\,\text{p-valeur}(X) \leq \varepsilon\,\big) \;\leq\; \varepsilon. \tag{3.3}$$

La p-valeur représente ainsi une notion de probabilité de fausse alarme. En effet, pour une réalisation x d'une variable aléatoire suivant le modèle H_0 et en détectant cette dernière si p-valeur$(x) \leq \varepsilon$, pour $\varepsilon > 0$, on s'assure alors que la probabilité de détection soit au plus égale à ε. Le nombre de fausses alarmes apparaît ainsi comme une extension de la notion de p-valeur, exigeant que l'on prenne en compte l'existence de tests multiples pour majorer, non pas simplement la probabilité de fausse alarme, mais le nombre de fausses alarmes.

3.1.1.2 Construction d'un Nombre de Fausses Alarmes

Théorème fondamental Dans la pratique, la construction d'un nombre de fausses alarmes repose sur le théorème suivant.

Théorème 3.1.2 (Construction d'un NFA [DMM07])**.** *Soit $X \in \mathcal{X}$ une variable aléatoire suivant une distribution H_0 et soit un ensemble de structures $\mathcal{I}$. Alors, pour toute fonction $\mu : \mathcal{I} \times \mathcal{X} \to \mathbb{R}$ et tout ensemble de réels positifs $(\eta_i)_{i\in\mathcal{I}}$, la fonction définie pour tout $x \in \mathcal{X}$ et $i \in \mathcal{I}$ par*

$$\mathrm{NFA}_{\mu,\eta}(i,x) \;=\; \eta_i \,.\, \mathbb{P}_{X\sim H_0}\big(\,\mu(i,X) \geq \mu(i,x)\,\big)\,, \tag{3.4}$$

est un nombre de fausses alarmes dès lors que

$$\sum_{i\in\mathcal{I}} \eta_i^{-1} \;\leq\; 1. \tag{3.5}$$

Par la suite, nous qualifierons la fonction μ de « mesure de structure » et les η_i, pour $i \in \mathcal{I}$, de coefficients de pondération.

Nous donnons la démonstration de ce théorème en annexe A.1. Notons que ce théorème se retrouve souvent sous une forme légèrement différente, mais équivalente. On la retrouve par exemple, dans le livre de Desnoleux et coauteur [DMM07] sous la forme « $\eta_i\,\mathbb{P}\big(\,X_i \geq x_i\,\big)$ », en considérant une variable X_i par structure $i \in \mathcal{I}$. Nous préférerons cependant l'expression donnée dans l'équation (3.4) qui présente l'avantage de bien distinguer ce qui est lié à la modélisation des données, avec le modèle naïf, et ce qui est lié à ce que l'on mesure sur les données, au travers de la mesure de structure.

Choix de la pondération Revenons rapidement sur l'intérêt pratique de la pondération. La pondération permet de donner des importances différentes à certaines structures. On peut ainsi imaginer prendre en compte, dans la pondération, des aprioris quant aux structures à détecter.

Toutefois, dans notre cas, les structures correspondront dans la plupart des cas à des positions dans l'image. Aussi, ne voulant rajouter aucun apriori sur la position des changements, nous choisirons dans ces cas une pondération uniforme, avec ainsi $\eta_i = |\mathcal{I}|$ pour tout $i \in \mathcal{I}$.

Valeur extrêmes et représentation machine Il est important de noter que, dans la pratique, il n'est pas rare que les valeurs prises par les p-valeurs et NFA soient extrêmement faibles, souvent inférieures à 10^{-100}, parfois même à 10^{-1000}. Hors, les précisions standard des différents langages de programmation ne sont souvent pas suffisantes pour pouvoir stocker ces valeurs sans arrondir à zéro le résultat. La majorité d'entre eux utilisent en effet la norme « IEEE 754 » sur l'arithmétique à virgule flottante [Flo19], dont les plus petites valeurs strictement positives représentables sont $2^{-1022} \simeq 10^{-308}$ pour une représentation des nombres flottants sur 64 bits.

De ce fait, dans la pratique, il faut plutôt chercher à calculer le logarithme du NFA et l'on pourra plutôt considérer, comme mesure de significativité, la grandeur « $-\log \mathrm{NFA}(i,X)$ », que l'on cherchera à estimer sans problème d'arrondi à zéro.

3.1.1.3 Optimalité et approximations d'un NFA

Nombre exact de fausses alarmes Il est important de remarquer que la propriété de nombre de fausses alarmes ne permet que de majorer le nombre de fausses alarmes, sans rien demander sur le nombre réel de fausses alarmes sous la condition H_0. Entre autre, remarquons que la fonction définie pour tout $x \in \mathcal{X}$ et $i \in \mathcal{I}$ par $\mathrm{NFA}_0(i,x) = |\mathcal{I}|$ est bien un nombre de fausses alarmes, puisque pour tout $\varepsilon > 0$,

$$\mathbb{E}_{X\sim H_0}\Big[\,\mathrm{Card}\,\{i \in \mathcal{I}, \mathrm{NFA}_0(i,X) \leq \varepsilon\}\,\Big] \;=\; |\mathcal{I}|\;\mathbb{1}_{\{\varepsilon \geq |\mathcal{I}|\}} \;\leq\; \varepsilon. \tag{3.6}$$

Ainsi, dans la pratique, il n'est pas suffisant de vérifier la propriété de nombre de fausses alarmes, donnée dans l'équation (3.1). Il s'agit aussi d'être le plus proche possible d'un « vrai » nombre de fausses alarmes, vérifiant

$$\forall \varepsilon > 0, \qquad \mathbb{E}_{X\sim H_0}\Big[\,\mathrm{Card}\,\{i \in \mathcal{I}, \mathrm{NFA}(i,X) \leq \varepsilon\}\,\Big] \;\simeq\; \varepsilon.$$

Optimalité d'un NFA L'hypothèse précédente n'ayant pas de sens lorsque la variable X est discrète, nous définissons plutôt un critère de optimalité d'un NFA de la manière suivante.

Définition 3.1.3 (Optimalité d'un NFA)**.** *Soit $X \in \mathcal{X}$ une variable aléatoire suivant une distribution H_0 et soit un ensemble de structures $\mathcal{I}$. Un nombre de fausses alarmes F relatif au modèle naïf $(X,\mathcal{I})$ est dit optimal lorsqu'aucun nombre de fausses alarmes relatif à $(X,\mathcal{I})$ ne peut être inférieur à F sans lui être égal sur un ensemble de probabilité 1, c'est-à-dire lorsque, pour toute fonction G de $\mathcal{X} \times \mathcal{I}$ dans $\mathbb{R}$,*

$$\begin{cases} G \text{ est un NFA} \\ G \leq F \end{cases} \quad\Longrightarrow\quad \forall i \in \mathcal{I},\;\; \mathbb{P}\Big(\,G(i,X) = F(i,X)\,\Big) = 1. \tag{3.7}$$

Dans la pratique, les nombres de fausses alarmes obtenus par le théorème 3.1.2 permettent souvent de trouver une bonne approximation du nombre réel de fausses alarmes. Dans notre cas, on peut montrer le résultat suivant.

Propriété 3.1.4 (Critère d'optimalité d'un NFA)**.** *En gardant les notations du théorème 3.1.2, si chaque coefficients de pondération η_i est égal à $\mathrm{Card}(\mathcal{I})$ et pour tout $i \in \mathcal{I}$, la variable aléatoire $\mu(i,X)$ possède une fonction de densité, alors la fonction $\mathrm{NFA}_{\mu,\eta}$ est un NFA optimal.*

De plus, sous ces conditions, $\mathrm{NFA}_{\mu,\eta}$ correspond bien au nombre moyen de fausses alarmes, dans le sens où

$$\forall \varepsilon \in [0, \mathrm{Card}(\mathcal{I})]\,, \qquad \mathbb{E}_{X\sim H_0}\Big[\,\mathrm{Card}\,\{i \in \mathcal{I}, \mathrm{NFA}_{\mu,\eta}(i,X) \leq \varepsilon\}\,\Big] \;=\; \varepsilon. \tag{3.8}$$

Démonstration. Cette propriété est démontrée en annexe A.1 □

Majoration d'un NFA La propriété précédente nous incite fortement à « toujours » considérer des nombres de fausses alarmes en utilisant le théorème 3.1.2. Toutefois, il est parfois impossible de trouver une valeur explicite et calculable de l'expression donnée dans l'équation (3.4). Il est alors possible d'approcher la valeur du NFA par majoration, tout en conservant la propriété du nombre de fausses alarmes, donnée équation (3.1). On peut en effet montrer la propriété suivante.

Propriété 3.1.5. *Soit $X \in \mathcal{X}$ une variable aléatoire suivant une distribution H_0 et $\mathcal{I}$ un ensemble quelconque. Si F et G sont deux fonctions de $\mathcal{I} \times \mathcal{X}$ dans $\mathbb{R}$, alors*

$$\begin{cases} F \text{ est un NFA} \\ F \leq G \end{cases} \implies G \text{ est un NFA}, \tag{3.9}$$

les NFA étant associés au modèle naïf $(X, \mathcal{I})$.

Démonstration. En effet, puisque pour tout $i \in \mathcal{I}$ et tout $x \in \mathcal{X}$, $G(i,x) \geq F(i,x)$, alors pour tout $\varepsilon > 0$,

$$\{i, G(i, X_i) \leq \varepsilon\} \subseteq \{i, F(i, X_i) \leq \varepsilon\},$$

ce qui nous donne directement une majoration de leur cardinal et nous donne donc bien

$$\mathbb{E}\Big[\operatorname{Card}\{i, G(i, X_i) \leq \varepsilon\}\Big] \leq \mathbb{E}\Big[\operatorname{Card}\{i, F(i, X_i) \leq \varepsilon\}\Big] \leq \varepsilon.$$

□

Ainsi, pour construire un nombre de fausses alarmes, on pourra partir de l'expression donnée équation (3.4), puis agir par majoration jusqu'à obtenir une expression calculable ou estimable, en faisant attention à ne pas effectuer de majorations trop grossières.

3.1.1.4 Fusion de NFA

Minimum et maximum de NFA Énonçons deux propriétés relatives à la fusion de nombre de fausses alarmes.

Propriété 3.1.6 (Maximum de NFA)**.** *Soit $\mathcal{X}$ un ensemble muni d'une distribution H_0 et $\mathcal{I}$ un ensemble quelconque. Si F_1, F_2, ..., F_n sont des fonctions de $\mathcal{I} \times \mathcal{X}$ dans $\mathbb{R}$, pour $n \in \mathbb{N}^*$, alors*

$$F_1, \ldots, F_n \text{ sont des NFA} \implies \max(F_1, \ldots, F_n) \text{ est un NFA}, \tag{3.10}$$

tous les NFA étant associés au modèle naïf $(X, \mathcal{I})$.

Démonstration. Il suffit de remarquer que $F_1 \leq \max(F_1, \ldots, F_n)$ et d'utiliser la propriété 3.1.5. □

Propriété 3.1.7 (Minimum pondéré de NFA)**.** *Soit $\mathcal{X}$ un ensemble muni d'une distribution H_0 et $\mathcal{I}$ un ensemble quelconque. Si F_1, F_2, ..., F_n sont des fonctions de $\mathcal{I} \times \mathcal{X}$ dans $\mathbb{R}$, pour $n \in \mathbb{N}^*$, et a_1, ..., a_n sont des réels positifs vérifiant $\sum_i a_i^{-1} \leq 1$, alors*

$$F_1, \ldots, F_n \text{ sont des NFA} \implies \min(a_1 F_1, \ldots, a_n F_n) \text{ est un NFA}, \tag{3.11}$$

tous les NFA étant associés au modèle naïf $(X, \mathcal{I})$.

Démonstration. Considérons des réels positifs $a_1, \ldots, a_n$, dont la somme des inverses est inférieure ou égal à 1 et posons $F = \min(a_1 F_1, \ldots, a_n F_n)$. Pour tout $x \in \mathcal{X}$ et $\varepsilon > 0$, on a alors,

$$\Big\{ i \in \mathcal{I}, F(i,x) \leq \varepsilon \Big\} \;=\; \bigcup_{k=1}^{n} \Big\{ i \in \mathcal{I}, a_k F_k(i,x) \leq \varepsilon \Big\}, \tag{3.12}$$

et donc

$$\mathrm{Card}\Big(\Big\{ i \in \mathcal{I}, F(i,x) \leq \varepsilon \Big\}\Big) \;\leq\; \sum_{k=1}^{n} \mathrm{Card}\Big(\Big\{ i \in \mathcal{I}, a_k F_k(i,x) \leq \varepsilon \Big\}\Big). \tag{3.13}$$

Pour $X \in \mathcal{X}$ une variable aléatoire de distribution $\mathcal{H}_0$ et en appliquant l'espérance, les F_i étant des nombres de fausses alarmes, on obtient bien, comme annoncé,

$$\mathbb{E}\Big[\, \mathrm{Card}\Big(\Big\{ i \in \mathcal{I}, F(i,X) \leq \varepsilon \Big\}\Big) \,\Big] \;\leq\; \sum_{k=1}^{n} \frac{\varepsilon}{a_k} \;\leq\; \varepsilon. \tag{3.14}$$

□

Intersection et union de masques de détection Ces deux propriétés sont particulièrement utiles lorsque l'on cherche à considérer l'union ou l'intersection de masques de détections obtenus en seuillant des nombres de fausses alarmes.

Pour $n \in \mathbb{N}$, considérons n nombres de fausses alarmes, $F_1, \ldots, F_n$, associés à un même modèle naïf $(\mathcal{X}, \mathcal{I})$, et que l'on évalue sur un objet $x \in \mathcal{X}$, chacun donnant un ensemble de structures détectées

$$A_j(\varepsilon) \;=\; \{ i \in \mathcal{I},\; F_j(i,x) \leq \varepsilon \},$$

pour $j \in [\![1,n]\!]$ et $\varepsilon \geq 0$. En considérant le maximum des NFA,

$$G(i,x) \;=\; \max_{j \in [\![1,n]\!]} F_j(i,x) \tag{3.15}$$

pour tout $i \in \mathcal{I}$ et $x \in \mathcal{X}$, l'ensemble des structures détectées par G apparaît comme l'intersection des $A_j(\varepsilon)$, le détecteur associé à G détectant une structure lorsque chacun des F_j la détecte. De manière similaire, on peut considérer le minimum pondéré des NFA,

$$H(i,x) \;=\; \min_{j \in [\![1,n]\!]} n \times F_j(i,x) \tag{3.16}$$

pour tout $i \in \mathcal{I}$ et $x \in \mathcal{X}$, qui permet, en quelque sorte, de retrouver l'ensemble des structures détectées par au moins un des F_j.

Remarquons que le nombre de fausses alarmes H correspond ici à une fusion de détecteurs plus « contraignants », les structures détectées correspondant à l'union des $A_j(\varepsilon/n)$. Notons toutefois que les NFA ont souvent des valeurs très faibles, évaluées en échelle logarithmique, et que le facteur n a donc souvent une influence limitée.

3.1.2 Cas pratique

3.1.2.1 Résumé de la construction a contrario

Reprenons ici le processus complet de construction d'un nombre de fausses alarmes et du détecteur a contrario qui en découle, et ce dans une optique de détection de changements.

Modèle naïf La construction d'un modèle a contrario commence par la définition du modèle naïf. Comme dit précédemment, il s'agit de définir l'objet étudié ainsi que la manière dont il dépend du bruit. Dans le cadre de la détection de changements au sein d'une paire d'image, les données correspondront soit directement aux deux images que l'on modélisera de manière jointe, soit uniquement à la seconde, en prenant la première image comme référence. Dans les deux cas, il faudra au moins prendre en compte le bruit d'acquisition, mentionné dans la section 2.1.2.

Ensemble de structures On définit ensuite l'ensemble $\mathcal{I}$ des structures envisagées comme potentielles détections. Dans la pratique, il s'agit de considérer l'ensemble des paramètres décrivant un changement potentiel, comme sa position ou sa taille. Dans la plupart des cas, nous fixerons la forme des structures en faisant uniquement varier leur position, en prenant ainsi $\mathcal{I}$ comme étant égal à l'ensemble des pixels.

Pondération On choisit des coefficients de pondération η_i, pour $i \in \mathcal{I}$, strictement positifs. Si chaque structure est équivalente, on considérera une pondération uniforme. Notons que la contrainte sur la valeur de la somme des inverses des coefficients de pondérations, donnée équation (3.5), se réduit en fait à un simple facteur de normalisation, l'ensemble $\mathcal{I}$ étant nécessairement fini dans la pratique.

Mesure de structure On définit alors la mesure de structure μ, de $\mathcal{I} \times \mathcal{X}$ dans $\mathbb{R}$, caractérisant les structures que l'on cherche à détecter. Dans le cadre de la détection de changement, il s'agira de mesurer ce qui pourrait trahir la présence de changement, comme une variation locale d'intensité ou encore un écart des intensités local vis-à-vis de statistiques globales.

Estimation du nombre de fausses alarmes On cherche alors à trouver une expression calculable ou estimable du nombre de fausses alarmes donné équation (3.4), en partant de la p-valeur associée à $\mu(i, X)$ lorsque $X \sim H_0$, et agissant par majoration. Autrement dit, on cherche à trouver une fonction F^* calculable (ou estimable) de sorte que, pour tout $\delta \in \mathbb{R}$,

$$\mathbb{P}_{X \sim H_0}\big(\, \mu(i, X) \geq \delta \,\big) \;\leq\; F^*(i, \delta). \tag{3.17}$$

On en déduit alors le nombre de fausses alarmes, majorant l'expression donnée équation (3.4),

$$\mathrm{NFA}\,(i, x) \;=\; \eta_i \,.\, F^*(i, \mu(i, x)). \tag{3.18}$$

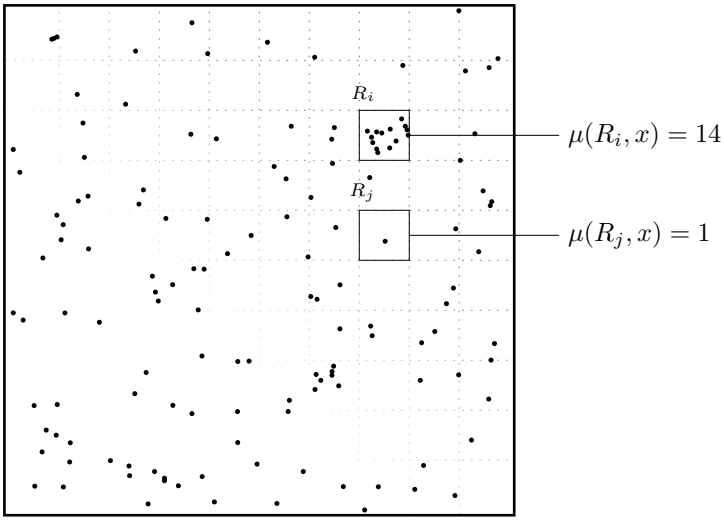


Figure 3.1 – Schéma représentant le problème de détection de groupement de points. On représente ici le carré $[0,1]^2$ et chaque point représente un des x_i, pour $i \in [\![1, N]\!]$. Nous faisons figurer partiellement et en pointillé la limite des régions constituant l'ensemble des structures $\mathcal{I}$ et affiché la mesure de structure pour deux d'entre eux.

Seuillage au niveau ε Pour obtenir un masque de détection, il suffit alors de seuiller les valeurs de NFA par une valeur seuil $\varepsilon > 0$ donnée. Si la valeur $\varepsilon = 1$ semble convenir dans les exemples jouets où le modèle naïf correspond parfaitement aux données, dans la pratique, il peut être préférable de compenser les erreurs de modélisation par une valeur seuil inférieure à 1, par exemple 10^{-10}, que nous considérons dans les différents exemples.

3.1.2.2 Exemple : Détection de groupements de points

Pour mieux comprendre la méthode a contrario, intéressons-nous à un exemple concret : le problème de la détection de groupements de points. On se donne un ensemble de N points, $x = \{x_1, x_2, \ldots, x_n\}$, dans le carré $[0,1]^2$ et on souhaite détecter des groupements de points. Construisons alors un détecteur a contrario en suivant les étapes données dans la section 3.1.2.1.

Modèle naïf L'espace des objets à considérer est l'ensemble des N-uplets de points, $\mathcal{X} = \left([0,1]^2\right)^N$, et on considère comme modèle naïf que les points sont tirés indépendamment de manière uniforme sur le carré : $H_0 = \mathcal{U}_{nif}(\mathcal{X})$. En ce qui concerne l'ensemble $\mathcal{I}$ des structures à envisager, nous allons simplement considérer un ensemble de sous-rectangles de $[0,1]^2$, correspondant aux positions éventuelles des clusters. Partitionnons ici $[0,1]^2$ en petits carrés de taille $\frac{1}{10} \times \frac{1}{10}$, et considérons ces carrés comme l'ensemble des structures potentielles :

$$\mathcal{I} \;=\; \left\{ \left[\tfrac{k}{10}, \tfrac{k+1}{10}\right] \times \left[\tfrac{l}{10}, \tfrac{l+1}{10}\right],\ (k,l) \in [\![0,9]\!]^2 \right\},$$

comme nous le représentons dans la Figure 3.1. Ces carrés étant de taille fixe, nous prendrons une pondération uniforme, ce qui nous donne $\eta_R = 100$ pour tout $R \in \mathcal{I}$.

Mesure de structure et expression du NFA En ce qui concerne la mesure de structure, pour chaque rectangle $R \in \mathcal{I}$, nous compterons le nombre de points x_i tombant dans R, c'est-à-dire

$$\mu(R, x) \;=\; \sum_{i=1}^{N} \mathbb{1}_{\{x_i \in R\}}. \tag{3.19}$$

Il reste alors à calculer $\mathbb{P}(\,\mu(R, X) \geq \delta\,)$ pour $\delta \geq 0$ et lorsque $X \sim H_0$. Remarquons alors que, les points étant répartis de manière uniforme et indépendante, $\mu(R, X)$ suit une loi binomiale de paramètres N et $\frac{1}{100}$. De ce fait, on trouve directement

$$\mathbb{P}(\,\mu(R, X) \geq \delta\,) \;=\; \mathcal{B}\Big(N, \delta, \tfrac{1}{100}\Big)\,, \tag{3.20}$$

où $\mathcal{B}(N, k, p)$ est la queue de la distribution binomiale de paramètres $N \in \mathbb{N}$ et $p \in [0, 1]$, en sommant à partir de la valeur $k \in [\![0, N]\!]$, c'est-à-dire

$$\mathcal{B}(N, k, p) \;=\; \sum_{i=k}^{N} \binom{N}{i} p^i (1-p)^{N-i}. \tag{3.21}$$

On en déduit alors le nombre de fausses alarmes, pour $R \in \mathcal{I}$,

$$\mathrm{NFA}(R, x) \;=\; 100.\mathcal{B}\left(N, \sum_{i=1}^{N} \mathbb{1}_{\{x_i \in R\}}, \tfrac{1}{100}\right). \tag{3.22}$$

En seuillant alors cette valeur de NFA, pour une valeur de seuil $\varepsilon = 1$, on trouve alors les clusters détectés. Nous affichons dans la Figure 3.2 le résultat obtenu pour la répartition de point donnée dans la Figure 3.1.

3.2 Prétraitement et conventions prises

Avant de rentrer d'expliciter les différents modèles primaires de changements, revenons formellement sur les différents traitements qu'il est nécessaire ou utile d'effectuer au préalable.

3.2.1 Incohérences spatiales et recalage d'images

Incohérences spatiales entre les images Comme nous l'avons signalé dans le chapitre précédent, en particulier dans la section 2.1.2, dans la plupart des cas, l'imagerie satellite souffre d'un problème d'incohérences spatiales. En effet, si la phase d'orthorectification (explicitée page 44) permet théoriquement de ramener dans un même référentiel pixellique deux images pouvant être prises à des positions et avec inclinaisons différentes, les erreurs et/ou approximations (principalement sur le modèle d'élévation du sol) qui sont faites peuvent faire subsister des incohérences spatiales.

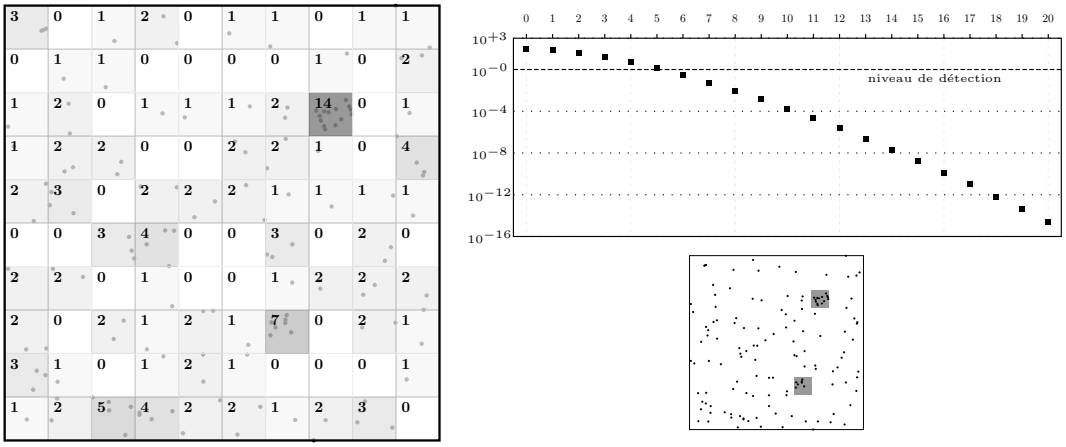


Figure 3.2 – Résultat de l'algorithme de détection de groupements de points. À gauche, nous affichons, à l'intérieur de chaque carré R représentant une détection potentielle, la valeur de la mesure de structure, $\mu(R, x)$. À droite figure le graphe représentant le nombre de fausses alarmes en fonction de la valeur de $\mu(R, x)$ ainsi que le résultat obtenu en seuillant le NFA à une valeur de 1, les carrés aboutissant à des détections étant affichés grisés.

Notons que ces incohérences spatiales concernent d'autant plus les images que leur résolution spatiale est grande. Elles seront donc très présentes dans des images très hautes résolutions et très peu visibles dans des images moyenne ou basse résolution.

Formalisons alors le problème et considérons une zone géographique que l'on souhaite étudier, vue comme un sous-domaine de $\mathbb{R}^2$ et que nous noterons ici $\mathcal{R}_0$. Chaque image u peut être vue comme une fonction d'un domaine rectangulaire $\mathcal{D}_u \in \mathbb{Z}^2$, à valeur dans $\mathbb{R}^K$, K étant le nombre de canaux de l'image. Dans le cas d'une orthorectification parfaite, pour chaque pixel $p \in \mathcal{D}_u$, l'intensité u_p correspondrait exactement à l'emplacement $p \in \mathcal{R}_0$. Toutefois, cette correction n'étant pas parfaite, il serait nécessaire de prendre en compte une déformation d'image $T : \mathcal{D}_u \to \mathcal{R}_0$ de sorte que u_p, l'intensité de u au pixel p, corresponde en réalité à l'emplacement $T(p) \in \mathcal{R}_0$. l'orthorectification est alors d'autant meilleure que T est proche de l'identité.

Décalage local Dans le cadre où l'on dispose de plusieurs images, u et v, de domaine $\mathcal{D}_u$ et $\mathcal{D}_v$, et si l'on suppose que les transformations respectives T_u et T_v sont suffisamment régulières pour être interpolées de manière inversible (ce qui suppose entre autres qu'il n'existe aucune occlusion dans les images), les décalages de chaque image par rapport à la zone géographique reviennent à un décalage $T : \mathcal{D}_u \to \mathbb{R}^2$ entre les deux images, de sorte qu'un pixel $p \in \mathcal{D}_u$ dans la première image corresponde à une position $T(p)$ dans la seconde. Remarquons bien ici que rien ne force T à n'avoir que des valeurs entières.

Dans un cadre idéal, les deux orthorectifications seraient parfaites et la transformation T serait l'identité, simplifiant alors largement la comparaison des deux images. Dans toute la suite du document, cette condition sera introduite sous l'appellation

d'« images parfaitement recalées ». Notons que cela n'est malheureusement pas le cas dans la pratique et il est alors possible de définir une notion de « décalage local » en considérant $T(p) - p$, pour tout pixel $p \in \mathcal{D}_u$.

Problème de recalage d'images Ainsi, dans le cas général, on dispose initialement d'une paire d'image, décalées spatialement et définies sur deux domaines différents ($\mathcal{D}_u$ et $\mathcal{D}_v$). Pour simplifier alors le problème de détection de changement, nous allons chercher à modifier une des deux images de sorte qu'elles soient alors recalées « au mieux », sur un même domaine.

Il s'agit là d'un problème classique de recalage d'image, très étudié entre autres pour des paires d'images stéréoscopiques (notamment pour la reconstruction de nuage de point 3D) ainsi qu'en biologie. Remarquons cependant que l'on a ici affaire à un problème simplifié, et cela pour plusieurs raisons. Tout d'abord, contrairement à l'imagerie médicale, qui demande très souvent de modéliser finement les déformations, nous n'avons ici affaire qu'à des déformations perspectives (si on néglige les effets de déformations optiques, cités dans la section 2.1.2), donnant ainsi une direction commune aux décalages $T(p) - p$, pour tout $p \in \mathcal{D}_u$. Par ailleurs, contrairement à l'étude des paires d'images stéréoscopiques « standard », l'imagerie satellite présente l'avantage de disposer d'une scène très « plane », ce qui limite très fortement les effets d'occlusion et nous cantonne la plupart du temps à des déformations continues du plan, le tout en conservant une résolution à peu près constante dans toute l'image. De plus, il est ici possible d'utiliser les méta-données de l'image, contenant entre autres la position du satellite, l'angle de vue ainsi que l'heure et la date de la prise de vue.

Remarquons pour finir que le recalage que nous souhaitons faire ici intervient après les différentes corrections et l'orthorectification et ainsi, on peut supposer que les images ont déjà la même orientation et que le décalage local présent entre les deux images est limité (dans la pratique, inférieur à une vingtaine de pixel).

Nécessité d'un recalage multi-échelle Avant de continuer, il est important de noter que les images satellites « bruts » sont particulièrement grandes, classiquement jusqu'à 20000×20000 pixels. Ces « macro-images », comme nous les appellerons par la suite, sont donc difficilement traitables dans leur globalité. Pour des raisons de mémoire vive limitée, nous découperons les paires de macro-images en une multitude de paires de sous-images, que nous qualifierons par la suite simplement d'« images ».

Par ailleurs, l'orthorectification ayant pour but de recaler à l'échelle des macro-images, on peut alors comprendre pourquoi il est nécessaire d'effectuer un second recalage à l'échelle d'une sous-image de l'ordre de 1000×1000 pixels. Nous profiterons de ce découpage en sous-images, alors que nous qualifierons par la suite simplement d'« images », pour effectuer un second recalage plus précis.

Recalage uniquement en translation En ce qui concerne le recalage, il n'est pas nécessairement préférable d'utiliser un recalage le plus évolué possible du fait des masquages de changements que cela pourrait provoquer. En effet, un recalage précis pourrait ne pas faire de distinction entre un effet de parallaxe et une simple translation d'objets, et corrigerait donc cette dernière, masquant ainsi un changement potentiel.

Notons par ailleurs qu'un tel recalage demanderait pour être effectivement appliqué à l'image, que cette dernière soit réinterpolée, ce qui rajoute au problème le choix de la méthode d'interpolation, une « mauvaise » interpolation pouvant entraîner l'apparition d'artefacts qui complexifierait la détection de changements [TBU00, Tsa03, PMV00].

Pour ces deux raisons, nous ne considérerons qu'un recalage par translation pour les paires d'images étudiées. Étant donné un sous-domaine $\mathcal{D}$ de $\mathcal{D}_u$, il s'agit de chercher une translation τ^* minimisant une mesure de distance entre u et une version translatée de v. Ici, nous considérons ici la corrélation des deux images, en définissant alors

$$\tau^* = \underset{\tau}{\arg\max} \frac{\sum_{p\in D} (u_p - \overline{u})(v_{p+\tau} - \overline{v}_\tau)}{\sqrt{\sum_{p\in D} (u_p - \overline{u})^2 \sum_{p\in D} (v_{p+\tau} - \overline{v}_\tau)^2}}, \tag{3.23}$$

où

$$\overline{u} = \frac{1}{|D|}\sum_{p\in D} u_p \qquad \text{et} \qquad \overline{v}_\tau = \frac{1}{|D|}\sum_{p\in D} v_{p+\tau}. \tag{3.24}$$

Le maximum est ici pris pour $\|\tau\|_\infty \leq \tau_{\max}$, $\tau_{\max}$ étant une constante majorant l'amplitude maximale du décalage local, et que nous avons fixée arbitrairement à 20. En translatant v de τ^* et en restreignant ces images à $\mathcal{D}$, on obtient alors deux images de même domaine et dont les incohérences locales sont réduites.

Remarquons que, pour pouvoir trouver τ^*, il est nécessaire que le domaine $\mathcal{D}$ soit suffisamment loin des « bords » de $\mathcal{D}_v$, puisque $v_{q+\tau}$ doit être défini pour tout $p \in \mathcal{D}$ et pour tout $\tau \in [-\tau_{\max}, \tau_{\max}]^2$. Dans toute la suite, nous noterons u et v, les deux images recalées et partageant un même domaine $\mathcal{D}$.

Convention aux bords du domaine Il apparaît ici qu'il est souvent nécessaire de pouvoir disposer d'informations dans un certain contexte autour des pixels étudiés, que ce soit pour recaler les images (comme vu précédemment), pour calculer des descripteurs d'images (voir section 3.3.3) ou encore pour pouvoir modéliser les distributions locales des images (comme dans les sections 3.5 ou 4.1).

Si l'on peut parfois adapter les modèles ou algorithmes en fonction de la taille du contexte disponible, cela complexifie inutilement tant les algorithmes que les notations. Dans la pratique, nous ne chercherons pas à détecter les changements trop proches des bords du domaine, en définissant une marge (dont la largeur dépend des descripteurs et des modèles utilisés) et dans laquelle on ne demande aucun résultat de l'algorithme. Cette décision est justifiée par plusieurs points :

— tout d'abord, de part la taille des macro-images (plusieurs milliards de pixels), il est presque toujours possible, si l'on souhaite étudier une zone en particulier, d'utiliser un domaine $\mathcal{D}$ centré sur cette dernière de sorte qu'aucun bord ne pose problème ;

— d'autre part, dans le cadre de l'aide à la photo-interprétation, il n'est pas forcément nécessaire de chercher à obtenir un résultat sur l'intégralité de la zone, et il n'est pas trop dérangeant de laisser de côté les bords des macro-images, ces derniers ne représentant qu'un faible pourcentage du domaine.

Ainsi, nous ne considérerons ici aucun problème dû aux bords et nous supposerons pouvoir toujours disposer d'un contexte arbitrairement grand autour des pixels étudiés. Par ailleurs, pour éviter de devoir formellement se restreindre à un domaine érodé et pour simplifier les notations, nous considérerons que les images sont étendues à $\mathbb{Z}^2$, nulle en dehors de $\mathcal{D}$, de sorte que l'on puisse considérer leurs valeurs en tout point sans problème de définition.

3.2.2 Variation de contraste et recalage en intensité

Incohérence en intensité Comme annoncé dans le chapitre précédent et de manière similaire aux incohérences spatiales, lorsque l'on considère deux images satellites, prises dans des conditions différentes et parfois même par des satellites différents, il existe parfois des incohérences en termes d'intensité entre les différentes images étudiées.

Ces incohérences peuvent être dues à une mauvaise correction (radiométrique, atmosphérique et topographique, comme présenté dans la section 2.1.2) mais aussi à d'autres facteurs, comme la présence de nuage ou d'aérosols, ou encore des variations de température de l'air dans le cas de l'imagerie thermique, ou encore simplement à des variations au niveau des capteurs utilsés, si les images sont prises par des satellites différents.

Notons toutefois que, contrairement aux incohérences spatiales qui touchent la grande majorité des images, les incohérences en intensité sont plus rares et le plus souvent relativement faibles.

Changement de contraste Pour formaliser le problème, considérons deux images u et v, définies sur un même domaine $\mathcal{D} \subset \mathbb{Z}^2$, à valeur dans $\mathbb{R}^n$, $n \in \mathbb{N}^*$ étant un nombre commun de canaux, que l'on supposera correspondre à des longueurs d'ondes similaires dans les deux images.

On parlera alors de changement de contraste entre les deux images lorsqu'il existe une fonction $f : \mathbb{R}^n \to \mathbb{R}^n$ faisant correspondre les intensités des deux images, c'est-à-dire que (sans changement), pour tout $p \in \mathcal{D}$,

$$v_p \;=\; f(u_p), \tag{3.25}$$

cette égalité étant définie au bruit sur les images prêt.

Cette fonction de changement de contraste peut dans la pratique dépendre du pixel p (par exemple lorsque des aérosols ou une ombre ne sont présents que dans une partie de l'image), auquel cas nous parlerons de « changement de contraste local ». Dans le cas contraire, où f sera indépendante de p, nous parlerons de « changement de contraste global » dans le cas contraire.

Recalage en intensité envisagé De la même manière que pour les incohérences spatiales, il est nécessaire de se demander si l'on souhaite effectuer un recalage en intensité et avec quel niveau de « précision ». Un recalage trop précis pourrait en effet masquer certains changements, entre autres les changements de couleurs.

Dans la pratique, nous ne considérerons que les recalages globaux, considérant la fonction de changement de contraste f indépendant du pixel considéré et nous nous

limiterons aux fonctions affines de la forme $f(x) = ax + b$ pour $x \in \mathbb{R}^n$, e $a, b \in \mathbb{R}$. En cherchant alors la fonction minimisant la différence quadratique, c'est-à-dire pour

$$(a^*, b^*) \;=\; \underset{(a,b)\in\mathbb{R}^2}{\operatorname{argmax}} \sum_{p\in\mathcal{D}} \|u_p - av_p - b\|_2^2\,, \tag{3.26}$$

on trouve facilement (voir la démonstration dans l'annexe A.2) l'expression explicite suivante,

$$a^* \;=\; \frac{\displaystyle\sum_{p\in\mathcal{D}} v_p^{\mathsf{T}} u_p - |\mathcal{D}|\, \overline{v}^{\mathsf{T}}\overline{u}}{\|v\|_F^2 - |\mathcal{D}|\, \|\overline{v}\|_2^2}, \qquad \text{et} \qquad b^* \;=\; \overline{u} - a^*\overline{v}, \tag{3.27}$$

où $\|\cdot\|$ est la norme de Frobenius et où $\overline{u}$ et $\overline{v}$ sont les vecteurs moyens associés à u et v, c'est-à-dire

$$\overline{u} \;=\; \frac{1}{|\mathcal{D}|}\sum_{p\in\mathcal{D}} u_p \qquad \text{et} \qquad \overline{v} \;=\; \frac{1}{|\mathcal{D}|}\sum_{p\in\mathcal{D}} v_p. \tag{3.28}$$

Notons qu'il est aussi possible d'effectuer un tel recalage sur chaque composante prise séparément.

3.2.3 Conventions générales adoptées

Images et domaine étudiées Pour résumer, dans toute la suite, on considérera que les différentes corrections abordées dans la section 2.1.2, ainsi que les deux recalages, spatial et en intensité, ont déjà été appliqués. De ce fait, on suppose ici que l'on dispose de deux images u et v, représentant la même zone géographique à deux temps différents, définies sur un même domaine $\mathcal{D} \subset \mathbb{Z}^2$ (mais étendues à $\mathbb{Z}^2$ pour éviter les problèmes de définitions). Chaque image sera vue comme une fonction du domaine $\mathcal{D}$ dans $\mathbb{R}^K$, K représentant son nombre de canaux. Dans certain cas, il sera aussi plus simple de voir les images comme des vecteurs de $\mathbb{R}^{|\mathcal{D}|\times K}$ et nous nous permettrons de confondre les deux visions, en notant par exemple indifféremment u_p ou $u(p)$ la valeur d'une image u en un pixel $p \in \mathcal{D}$.

Par ailleurs, ces deux images pourront éventuellement toujours présenter des incohérences spatiales, le décalage local ou global existant entre les images pouvant alors être considéré comme relativement limité.

Note sur le recalage des images multimodales Il est important de noter que les deux derniers modèles permettront aux deux images de posséder un nombre différent de canaux. Dans ces conditions, les deux recalages dont il a été question ne sont plus directement applicables et il serait théoriquement nécessaire de présenter un algorithme de recalage permettant de prendre en entrée de telles paires d'images. Toutefois, dans la pratique, nous n'étudierons ici que des paires d'images possédant un même nombre de canaux, le problème de recalage multimodale constituant par ailleurs une réelle difficulté, nous ne l'aborderons pas ici.

Variables aléatoires Pour finir, fixons ici les différentes notations qui seront reprises dans toute la suite. Puisque l'on manipule des vecteurs aléatoires, nous prendrons la convention de nommer en majuscule les variables aléatoires modélisant les différentes grandeurs étudiées et en minuscule (avec la même lettre) les observations ou instances de ces variables. Ainsi, on notera u et v les deux images effectivement mesurées et U et V les variables aléatoires modélisant respectivement u et v.

3.3 Descripteurs d'images

3.3.1 Généralité sur les descripteurs

Introduction Avant de nous intéresser aux modèles primaires, intéressons-nous brièvement aux descripteurs d'images. Les descripteurs d'images peuvent être vus comme des manières de représenter l'information présente dans les différents canaux d'une image sous un autre point de vue. L'utilisation de descripteurs d'images possède plusieurs avantages.

— Tout d'abord, un descripteur permet de focaliser l'attention de la personne ou algorithme traitant les images sur certains éléments en particulier. Ainsi, par exemple, l'utilisation du « Normalized Difference Vegetation Index » fera largement ressortir les zones de végétation.

— Dans le même temps, en se focalisant sur un type d'information, un descripteur aura tendance à en masquer d'autres. Cela peut avoir un grand intérêt si ces informations amoindries sont jugées comme non pertinentes. Ainsi, par exemple, le descripteur de la norme du gradient, qui permet de visualiser facilement les contours, masquera la valeur de l'intensité locale et de ce fait, il pourra être utilisé pour comparer plus facilement des images radiométriquement différentes.

— Pour finir, il est possible en utilisant des descripteurs pour réduire le nombre de canaux, de condenser la quantité d'information à traiter. Ainsi, par exemple, la décomposition en composantes principales permet de réduire la dimension des données tout en perdant le moins possible d'information.

Formalisation d'un descripteur Formalisons la notion de descripteur. Étant donné une image I définie sur $\mathcal{D}$ et à valeur dans $\mathbb{R}^K$, pour $K > 0$, que l'on verra comme un vecteur de $\mathbb{R}^{|\mathcal{D}|\times K}$, nous définissons un descripteur d'image comme une fonction

$$f : \mathbb{R}^{|\mathcal{D}|\times K} \to \mathbb{R}^{|\mathcal{D}|\times K'}, \tag{3.29}$$

où K' est la dimension du descripteur f. Rappelons bien que, comme annoncé dans la section 3.2.1, nous ne considérerons pas ici les problèmes de définition qui pourraient survenir au bord du domaine $\mathcal{D}$. Dans la pratique, on ne pourrait obtenir qu'un descripteur sur une version érodée du domaine $\mathcal{D}$.

Par ailleurs, dans le cadre de l'étude des images satellite, nous pouvons cependant rajouter un certain nombre de contraintes dans cette définition.

Stationnarité et localité des descripteurs Tout d'abord, on demande que les descripteurs ne dépendent pas de la position dans l'image. En effet, si dans certaines

images, par exemple de caméra de surveillance, il existe des a priori sur la position des différents objets, comme le ciel en haut, le sol en bas, ou éventuellement que l'information pertinente est sensiblement au centre de l'image, ça n'est pas le cas pour des images satellites, qui ne présentant en général aucun a priori quant à la localisation de l'information. On demandera ainsi au descripteur d'être « stationnaire », dans le sens où une version translatée de l'image devra donner une version translatée des descripteurs. Nous supposerons de plus que les descripteurs sont « locaux », dans le sens où la valeur d'un descripteur associé à un pixel p ne devra pas dépendre de pixels trop éloignés de ce dernier.

Nous pouvons alors réussir ces contraintes sous la forme de l'existence d'un voisinage de 0, P_0, et d'une fonction f_0 définie sur $\mathbb{R}^{|P_0|\times K}$ de sorte que la valeur du descripteur $f(I)$ au niveau d'un pixel p ne dépende que des valeurs de I sur une version translatée de P_0 et centrée en p, noté ici $p + P_0$, c'est-à-dire

$$\forall p \in \mathcal{D}, \qquad f(I)_p \;=\; f_0(I_{|p+P_0}), \tag{3.30}$$

en notant I_{p+P_0} la restriction de I au domaine $p + P_0$, vu comme un vecteur de $\mathbb{R}^{|P_0|}$.

Stabilité des descripteurs Une autre contrainte que l'on se donne est la « stabilité » des descripteurs, autrement dit, on s'attend à ce que deux images presque identiques en termes de valeurs des pixels aient des descripteurs presque identiques. Formellement, il s'agit de demander à f_0 d'être au moins continue, si ce n'est α-lipschitzienne, c'est-à-dire vérifiant

$$\left\| f_0(I_{|p+P_0} + \varepsilon_{|p+P_0}) - f_0(I_{|p+P_0}) \right\|_2 \leq \alpha \left\| \varepsilon_{|p+P_0} \right\|_2, \tag{3.31}$$

pour tout image de bruit ε, $\alpha > 0$ étant fixé. Cette condition s'explique par le fait que l'on ne souhaite pas que le bruit présent dans les images puisse avoir une importance sur la valeur des descripteurs, ces derniers étant censés représenter l'information présente dans les images et non dans le bruit.

Notons que si cette condition ne semble pas être très contraignante, elle empêche pourtant de considérer certains descripteurs, comme l'orientation du gradient (que ce soit à valeur dans $[0, 2\pi]$ ou dans la sphère unité), à cause de la discontinuité que l'on aurait pour une image parfaitement homogène. Pour considérer cette orientation, il sera ainsi nécessaire de la lisser au voisinage de 0, comme nous l'expliquons dans la section 3.3.3.1.

Les trois types de descripteurs Nous allons ici décrire quelques-uns des principaux descripteurs utilisés en imagerie satellitaire et ainsi que ceux que nous utiliserons par la suite. Nous allons séparer ces descripteurs en plusieurs catégories, à savoir

- les descripteurs « d'intensité » regroupant l'information multispectrale et donc basés en grande majorité sur les différents canaux de l'image,
- les descripteurs « géométriques », ayant trait à la disposition spatiale de l'information, c'est-à-dire aux contours, formes ou autre, et ce de manière indépendante de la longueur d'onde utilisée.
- et les descripteurs de textures.

3.3.2 Descripteurs d'intensité

Intensité brut Le premier descripteur d'image que nous allons considérer est simplement la valeur brute des différents canaux de l'image. Il ne s'agit pas à proprement parler d'un descripteur, toutefois, dans le cas d'images multispectrales ou hyperspectrales, possédant un nombre conséquent de canaux, il peut être nécessaire de n'en sélectionner qu'un nombre réduit. S'il est possible d'effectuer une sélection à la main, il existe aussi de nombreuses méthodes pour effectuer cette sélection automatiquement [SB01, JKC13]. Toutefois, et puisque l'on ne s'intéressera ici qu'à des images possédant un nombre faible de canaux, nous ne nous arrêterons pas plus longtemps sur ce problème.

Nuance de gris Dans le cas d'une image multispectrale, il est très courant de vouloir fusionner les différents canaux en un seul. Cette réduction de dimension est particulièrement pertinente dans le cadre du traitement des images satellites, où les trois bandes de fréquence (« Rouge », « Vert » et « Bleu ») semblent n'apporter souvent que peu d'informations supplémentaires par rapport à leur fusion. La grande majorité des images que l'on étudie ici étant des images dans le spectre visible, on se restreindra à la nuance de gris en tant que fusion des trois canaux « Rouge », « Vert » et « Bleu », notés respectivement par la suite R, V et B. Il existe dans la littérature un grand nombre de moyens de calculer la nuance de gris, que nous noterons G. On peut citer par exemple

$$G \;=\; \frac{1}{3}\Big(R + V + B\Big), \tag{3.32}$$

ou encore

$$G \;=\; \frac{1}{2}\Big(\max\left(R, V, B\right) + \min\left(R, V, B\right)\Big). \tag{3.33}$$

Nous choisirons ici la combinaison

$$G \;=\; 0{,}2126\, R + 0{,}7152\, V + 0{,}0722\, B, \tag{3.34}$$

selon les recommandations « UIT-R BT.709-6 » proposées par l'Union internationale des télécommunications [BT15].

Indicateurs spécifiques Il existe un grand nombre de descripteurs cherchant à mettre en valeur différents types d'environnement. Ces descripteurs reposent sur différentes longueurs d'onde, entre autres le visible (R, V et B) ainsi que le proche infrarouge (NIR ou SWIR).

Le plus utilisé, le « Normalized Difference Vegetation Index » (NDVI) [RHSD74], s'écrit

$$\mathrm{NDVI} \;=\; \frac{\mathrm{NIR} - R}{\mathrm{NIR} + R}, \tag{3.35}$$

et fait ressortir la végétation. Parmi la multitude d'indicateurs de zones urbaines, on peut citer le « Normalized Difference Built-up Index » (NDBI) et le « Built-up Index » (BU) [Xu08],

$$\mathrm{NDBI} \;=\; \frac{\mathrm{SWIR} - \mathrm{NIR}}{\mathrm{SWIR} + \mathrm{NIR}} \qquad \text{et} \qquad \mathrm{BU} \;=\; \mathrm{NDBI} - \mathrm{NDVI}. \tag{3.36}$$

On peut aussi citer des indicateurs d'étendue d'eau, comme le « Normalized Difference Water Index » (NDWI) ou le « Modified Normalized Difference Water Index » (MNDWI) [Gao96],

$$\text{NDWI} = \frac{\text{Green} - \text{NIR}}{\text{Green} + \text{NIR}} \qquad \text{et} \qquad \text{MNDWI} = \frac{\text{Green} - \text{SWIR}}{\text{Green} + \text{SWIR}}. \tag{3.37}$$

Ces descripteurs sont utilisés dans des cas de détection, segmentation ou classification d'objets ou de changements, très orientés d'un point de vue sémantique. De ce fait, pour rester le plus général possible, nous ne considérerons pas ces descripteurs.

3.3.3 Descripteurs géométriques

On s'intéresse ici à des descripteurs capturant une information ayant trait à la disposition spatiale de l'information, c'est-à-dire aux contours, formes ou autre, et ce de manière indépendante de la longueur d'onde utilisée. De ce fait, tous les descripteurs que nous allons présenter ici sont extraits à partir d'un unique canal. Dans le cas d'images multispectrales, on peut alors fusionner les différents canaux et appliquer les descripteurs sur la nuance de gris obtenue. On pourrait aussi les extraire indépendamment sur plusieurs canaux pour obtenir plusieurs descripteurs différents.

Dans le cadre de notre problème, les images multispectrales seront presque exclusivement des images visibles. Nous utiliserons alors la nuance de gris comme base des descripteurs géométriques, puisque, comme nous l'avons déjà fait remarquer, les informations géométriques sont, dans la pratique, sensiblement identiques dans les trois canaux « Rouge », « Vert » et « Bleu », dans le cas d'images satellites.

Pour pouvoir montrer l'intérêt des descripteurs géométriques, nous les appliquerons sur une image synthétique, présentée dans la figure 3.3. Pour éviter des problèmes de bords, nous appliquerons les descripteurs à une image agrandie et nous n'afficherons les descripteurs que sur une version réduite de l'image, obtenue en rognant les marges de 200 pixels.

3.3.3.1 Gradients

Approximation discrète Le premier descripteur géométrique que l'on considère est le gradient. Il existe de nombreuses méthodes permettant d'estimer le gradient d'une image. La plus utilisée est l'estimation du gradient par un produit de convolution par les filtres de Sobel [SF68],

$$\nabla_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad \text{et} \qquad \nabla_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \tag{3.38}$$

respectivement pour le gradient horizontal et le gradient vertical. Si d'autres filtres sont éventuellement utilisés, comme

$$\nabla'_x = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$$

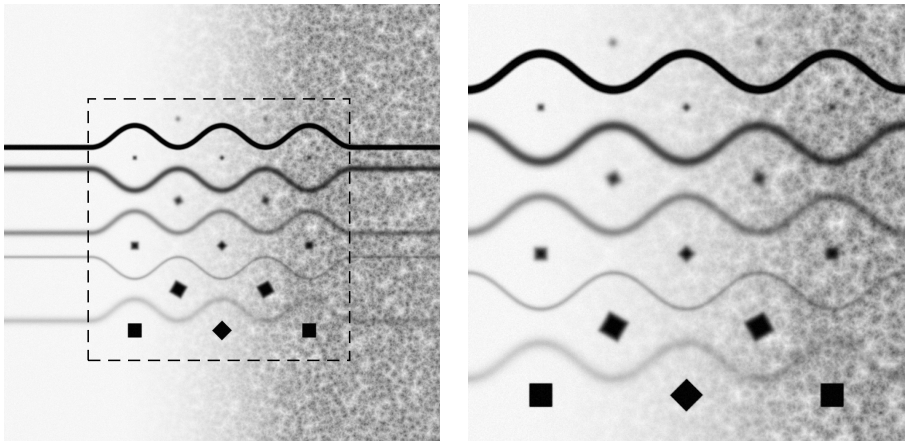


Image d'origine Image rognée

Figure 3.3 – Image synthétique utilisée pour représenter les différents descripteurs géométriques. Pour éviter des problèmes de bords, nous appliquerons les descripteurs à l'image entière, à gauche, et nous n'afficherons les descripteurs que sur l'image rognée, affichée à droite. Nous avons symbolisé le domaine de l'image de droite par une ligne en pointillé dans l'image de gauche.

ou simplement le filtre correspondant à l'opérateur de dérivation discrète,

$$\nabla''_x \;=\; \begin{bmatrix} 0 & -1 & 1 \end{bmatrix},$$

nous choisirons dans la pratique les filtres de Sobel, qui apportent plus de robustesse à des faibles perturbations. En effet, on veut voir ∇_x comme un lissage de ∇''_x, avec

$$\nabla'_x \;=\; \nabla''_x * \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \qquad \text{et} \qquad \nabla_x \;=\; \nabla'_x * \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix},$$

pour $*$ l'opérateur correspondant au produit de convolution.

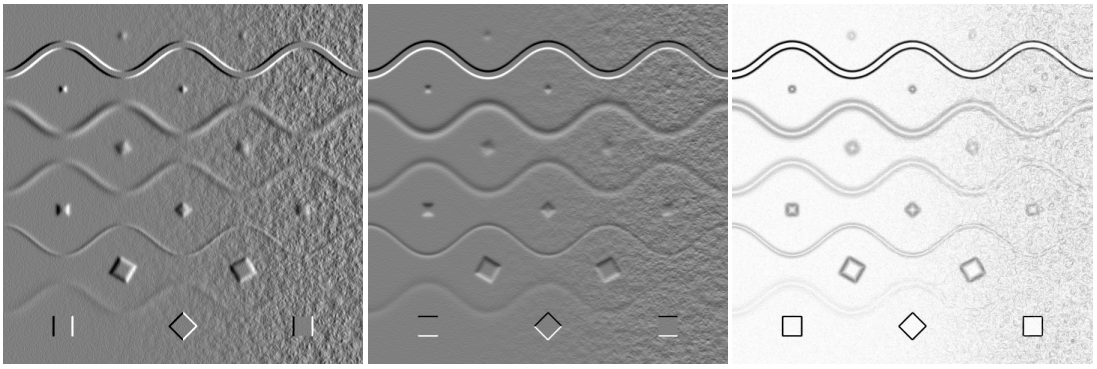
Normalisation pour une estimation réelle du gradient Intéressons-nous à une éventuelle normalisation de ces filtres. D'un côté, on peut vouloir estimer réellement le gradient de l'image. Pour un signal f défini sur $\mathbb{R}^2$, à valeurs réelles et de classe $\mathcal{C}^2$, en effectuant un développement limité à l'ordre 2, on trouve facilement que, pour $x, y \in \mathbb{R}$ et $h \in \mathbb{R}$,

$$f(x+h,y) - f(x-h,y) \;=\; 2h\,\mathrm{grad}_x f(x,y) + o(h^2),$$

et ainsi, en sommant cette expression respectivement 1, 2 et 1 fois, pour $y-h$, y et $y+h$, on trouve facilement que

$$(\nabla_x * f)(x,y) \;=\; 8h\,\mathrm{grad}_x f(x,y) + o(h^2),$$

en notant abusivement $\nabla_x * f$, la somme discrète de versions translatées f suivant le filtre ∇_x. En normalisant les deux filtres ∇_x et ∇_y par $\frac{1}{8}$, on estime alors bien le gradient de l'image à l'ordre 2.



Gradient X ($\nabla_x I$) Gradient Y ($\nabla_Y I$) Norme du gradient ($\|\nabla I\|_2$)

Figure 3.4 – Descripteurs d'image associés au gradient, appliqués sur l'image présentée dans la Figure 3.3. On présente ici successivement la composante horizontale du gradient, sa composante verticale et sa norme. Dans chacune des images, un pixel est d'autant plus sombre que le descripteur associé a une valeur importante. On notera que la valeur grise dans le coin haut gauche des deux premières images correspond à une valeur nulle, les couleurs claires représentant alors des valeurs négatives.

Normalisation pour une conservation du niveau de bruit D'un autre côté il peut être utile de disposer d'une estimation de la variance du bruit dans les différents descripteurs. De ce fait, on peut vouloir normaliser les deux filtres de sorte de conserver une même variance de bruit dans les descripteurs. On peut alors facilement montrer qu'il suffit de normaliser chaque filtre par sa norme 2, ici $\frac{1}{2\sqrt{3}}$. L'interprétation du gradient n'ayant pas une réelle importance ici, nous utiliserons cette normalisation, que nous continuerons d'appeler le « gradient », malgré le mauvais coefficient multiplicatif.

Norme du gradient Étant donnée une image I, on notera abusivement son gradient (à une constante multiplicative près), $\nabla I = (\nabla_x I, \nabla_y I)$, ici,

$$\nabla_x I = \frac{1}{2\sqrt{3}} \nabla_x * I \qquad \text{et} \qquad \nabla_y I = \frac{1}{2\sqrt{3}} \nabla_y * I, \tag{3.39}$$

et on définit alors la norme du gradient au pixel $p \in \mathcal{D}$, par

$$N(I)_p \;=\; \|\nabla I_p\|_2 \;=\; \sqrt{\nabla_x I_p^2 + \nabla_y I_p^2}. \tag{3.40}$$

Nous avons affiché dans la Figure 3.4 ces trois descripteurs, appliqués à l'image présentée dans la Figure 3.3.

Direction du gradient Comme annoncé précédemment, nous ne considérerons pas directement la direction du gradient, celle-ci présentant une discontinuité lorsque ∇I est nul. Pour pouvoir l'utiliser, il est alors nécessaire de la lisser en considérant, par exemple,

$$D(I)_p \;=\; \frac{\nabla I_p}{\|\nabla I_p\|_2} \min\left(\frac{\|\nabla I_p\|_2}{\alpha}, 0\right), \tag{3.41}$$

pour $\alpha > 0$ et qui correspond bien à la direction du gradient lorsque $\|\nabla I_p\|_2 \geq \alpha$, tout en étant continu au voisinage de $\nabla I_p = 0$.

3.3.3.2 Tenseur de structure

Définition du tenseur Le tenseur de structure est une matrice représentant le gradient à une échelle donnée [KWA11]. Il est basé sur le gradient orienté définit précédemment, $\nabla I = (\nabla_x I, \nabla_y I)$, que l'on représente sous la forme de la matrice des seconds moments,

$$\forall p \in \mathcal{D}, \qquad M_p \;=\; \nabla I_p \cdot \nabla I_p^{\mathsf{T}} \;=\; \begin{pmatrix} {\nabla_x I_p}^2 & \nabla_x I_p \cdot \nabla_y I_p \\ \nabla_x I_p \cdot \nabla_y I_p & {\nabla_y I_p}^2 \end{pmatrix}. \tag{3.42}$$

Écrit sous cette forme, la direction et la norme du gradient sont représentées au travers des vecteurs et valeurs propres. Il suffit en effet de remarquer que, par construction, cette matrice est de rang 1 et que $M_p \nabla I_p = \|\nabla I_p\|_2^2 \, \nabla I_p$. Si l'on perd l'orientation du gradient, la direction ne pouvant être retrouvée qu'au signe près, cette représentation possède l'énorme avantage de pouvoir être moyennée sans destruction d'information, contrairement au gradient.

Le tenseur de structure se définit alors comme une moyenne locale pondérée de ces matrices de second moment du gradient. Pour une fenêtre w, définie sur $\mathbb{Z}^2$, dans la pratique une gaussienne, le tenseur de structure $S(p)$ s'écrira ainsi

$$S_w(p) \;=\; \sum_r w(p-r) M_r \;\stackrel{\text{def}}{=}\; \begin{pmatrix} s_1(p) & s_3(p) \\ s_3(p) & s_2(p) \end{pmatrix}, \tag{3.43}$$

pour $s_1(p)$, $s_2(p)$ et $s_3(p)$ des réels représentant ainsi le tenseur « brut ».

Valeurs propres Pour une direction u de la sphère unité de $\mathbb{R}^2$, la valeur de $u^{\mathsf{T}} S_w(p) u$ va alors représenter une sorte d'intensité du gradient dans la direction u au voisinage du pixel p. Ainsi, pour $p \in \mathcal{D}$, en considérant les deux valeurs propres du tenseur, $\lambda_1(p) \geq \lambda_2(p)$, on pourra facilement évaluer l'isotropie de la distribution locale du gradient :

- si $\lambda_2(p) = 0$, le gradient est exclusivement orienté selon une seule direction,
- si $0 < \lambda_2(p) < \lambda_1(p)$, le gradient possède une direction privilégiée, et le champ est d'autant plus isotrope que les deux valeurs propres sont proches,
- et si $\lambda_2(p) = \lambda_1(p)$, le gradient est alors réparti de manière homogène, sans direction privilégiée.

On pourrait aussi considérer le premier vecteur propre, $V_1(p)$, représentant la direction principale du champ de gradient. Toutefois, tout comme la direction du gradient, cette direction principale présente une instabilité en 0 et ainsi, comme nous l'avons signalé précédemment, nous ne le considérerons donc pas comme un descripteur valable.

Notons que le tenseur étant de dimension 2×2, on peut trouver facilement des expressions explicites de $\lambda_1(p)$, $\lambda_2(p)$ et $V_1(p)$, à savoir (sans écrire toutes les dépendances en p pour une question de lisibilité)

$$V_1(p) \;=\; \frac{1}{Z} \begin{pmatrix} 2 s_3 \\ s_2 - s_1 + \sqrt{(s_2 - s_1)^2 + 4 {s_3}^2} \end{pmatrix}, \tag{3.44}$$

pour Z un coefficient de sorte que $\|V_1(p)\|_2 = 1$, et

$$\begin{aligned} \lambda_1(p) &= \tfrac{1}{2}\left(s_2 + s_1 + \sqrt{(s_2 - s_1)^2 + 4{s_3}^2}\right) \\ \lambda_2(p) &= \tfrac{1}{2}\left(s_2 + s_1 - \sqrt{(s_2 - s_1)^2 + 4{s_3}^2}\right). \end{aligned} \tag{3.45}$$

Nous avons représenté dans la Figure 3.5 la représentation de ces valeurs propres dans le cas de l'image présentée dans la Figure 3.3.

Cohérence et Cornerness Il est possible de définir d'autres descripteurs basés sur le tenseur de structure. Le premier que nous citerons, appelé « cohérence » [Jah93] quantifie l'anisotropie du gradient, caractérisant entre autres les bords, et s'écrit

$$c(p) = \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2. \tag{3.46}$$

Il faut noter que ce descripteur présente une discontinuité en 0, discontinuité que l'on peut régulariser en considérant par exemple

$$c_\alpha(p) = \min\left(0, \frac{\lambda_1}{\alpha}\right) \times \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2, \tag{3.47}$$

pour $\alpha > 0$, ce qui annulera progressivement la cohérence dans un voisinage proche de 0.

Les deux autres descripteurs dont nous allons parler représentent tous deux une notion de coin ou de tâche dans l'image et seront donc tous deux qualifiés de « cornerness ». Ce sont les grandeurs utilisées dans le détecteur de coins de Harris [HS+88], qui s'écrivent respectivement

$$H_1(p) = \frac{\det S_w}{\operatorname{Tr} S_w} = \frac{\lambda_1\lambda_2}{\lambda_1 + \lambda_2} = \frac{s_1 s_2 - s_3^2}{s_1 + s_2}, \tag{3.48}$$

$$H_2(p) = \det S_w - k \operatorname{Tr} {S_w}^2 = s_1 s_2 - s_3^2 - k\,(s_1 + s_2)^2. \tag{3.49}$$

3.3.3.3 Steerable filters et descripteur de structures linéaires

Définition Les descripteurs précédents permettent de mettre en valeur les contours en général, les bords, les coins ainsi que les tâches. On souhaite alors avoir un descripteur mettant en valeur les structures linéaires, que l'on retrouve régulièrement en imagerie satellitaire, par exemple au niveau des routes. Pour cela, nous nous sommes basés sur les « Steerable Gaussian filters » [FA91].

On considère comme filtre de base la dérivée seconde d'une gaussienne anisotrope,

$$G_\alpha(x, y) = \frac{\partial^2}{\partial x^2}\left(\frac{\alpha}{2\pi} e^{-\frac{1}{2}\left(x^2 + \alpha^2 y^2\right)}\right) = \alpha \frac{x^2 - 1}{2\pi} e^{-\frac{1}{2}\left(x^2 + \alpha^2 y^2\right)}, \tag{3.50}$$

pour $\alpha > 0$. Pour un angle $\theta \in [0, 2\pi]$ et un facteur d'échelle $\sigma > 0$, on définit alors la version tournée et mise à l'échelle du filtre par

$$G_{\alpha,\theta,\sigma}(x, y) = \sigma^2 G_\alpha\left(\frac{1}{\sigma} R_\theta(x, y)\right), \tag{3.51}$$

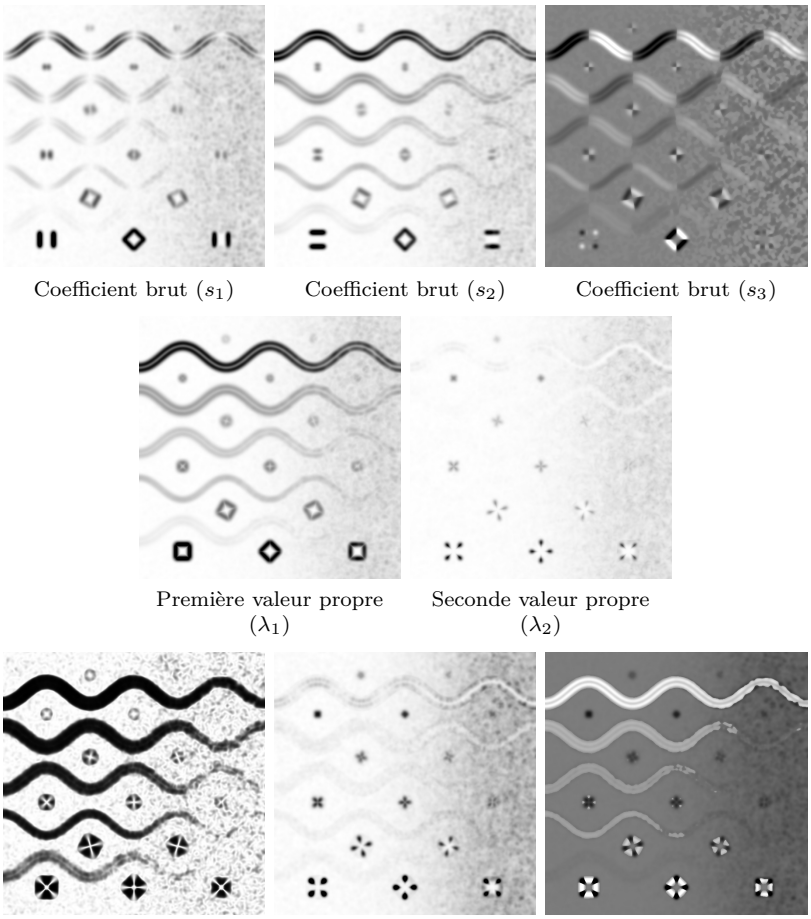


Figure 3.5 – Représentation des différents descripteurs associés au tenseur de structure, appliqués sur l'image de la Figure 3.3. On représente ici successivement les trois composantes du tenseur brut, ses deux valeurs propres et les trois descripteurs basés sur le tenseur, à savoir la cohérence, et les deux mesures de « cornerness ». Le tenseur est estimé ici en utilisant une fenêtre gaussienne de variance 10 pixels. On retrouve bien ici, dans les coefficients bruts, les composantes horizontales (s_1), verticales (s_2) du gradient. En ce qui concerne les valeurs propres et comme attendu, la première valeur propre semble marquer les zones non homogènes (bords et coins) tandis que la seconde permet de discriminer les coins des bords. Par ailleurs, on peut ici voir l'intérêt des descripteurs de « cohérence » et « cornerness » du tenseur. Le premier parvient à mettre en valeur les bords, et ce de manière bien plus visible que la norme du gradient (présenté dans la figure 3.4) ou la première valeur propre, qui s'estompe plus rapidement en présence de flou ou d'un manque de contraste. De même, les mesures de « cornerness » soulignent les structures de coin dans l'image, et ce de manière plus importante que la seconde valeur propre.

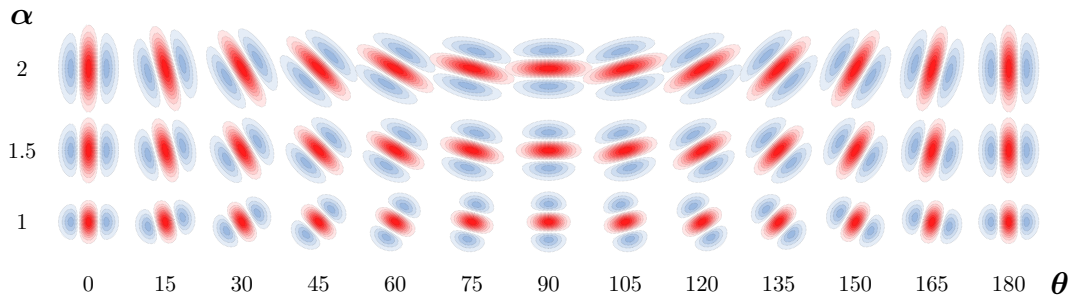


Figure 3.6 – Représentation des filtres $G_{\alpha,\theta,\sigma}$ pour différentes valeurs de α et θ. La couleur rouge (au centre de chaque filtre) correspond aux valeurs négatives tandis que la couleur bleue (sur les côtés de chaque filtre) correspond à des valeurs positives.

où R_θ est la rotation d'angle θ,

$$R_\theta(x,y) \;=\; \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{3.52}$$

Nous avons représenté ces filtres dans la Figure 3.6, pour différents angles et coefficients α. Pour simplifier les notations, fixons alors α et permettons-nous de ne plus noter systématiquement les dépendances en α. On définit alors la réponse $F_{\theta,\sigma}(p)$ d'une image I au pixel $p \in \mathcal{D}$ comme le produit de convolution entre cette image et une version échantillonnée de $G_{\alpha,\theta,\sigma}$,

$$F_{\theta,\sigma}(p) \;=\; \sum_{q\in\mathcal{D}} I(q)\, G_{\alpha,\theta,\sigma}(p-q). \tag{3.53}$$

L'intérêt des « Steerable gaussien filters » est de ne pas avoir à calculer cette réponse pour tous les angles possibles, mais de pouvoir exprimer $F_{\theta,\sigma}(p)$, pour un angle θ quelconque, comme une combinaison linéaire de réponses pour seulement quelques angles fixés. Le temps de calcul des descripteurs n'étant pas un point critique, nous ne discuterons pas des méthodes de calculs efficaces qui pourraient exister [FA91].

Descripteur de structure linéaire On pourrait considérer, comme descripteur, les réponses de l'image aux différents filtres, pour différentes échelles et orientations, seulement il peut être préférable de ne pas augmenter la dimension du descripteur, ce qui augmenterait aussi le temps de calcul des différents modèles de changement.

Puisque l'on cherche ici à mettre en valeur les structures linéaires, sans particulièrement s'intéresser à leur orientation ou échelle, nous pouvons condenser l'information, en considérant la représentation matricielle utilisée pour le tenseur de structure. En notant alors $e_\theta = (\cos\theta, \sin\theta)$ le vecteur unitaire de direction θ, et pour une échelle $\sigma > 0$ donnée, on définit alors $S_\sigma(p)$ la matrice regroupant les réponses du filtre suivant les différentes directions,

$$S_\sigma(p) \;=\; \int_{[0,\pi]} |F_{\theta,\sigma}(p)|\, .e_\theta e_\theta^\top \mathrm{d}\theta. \tag{3.54}$$

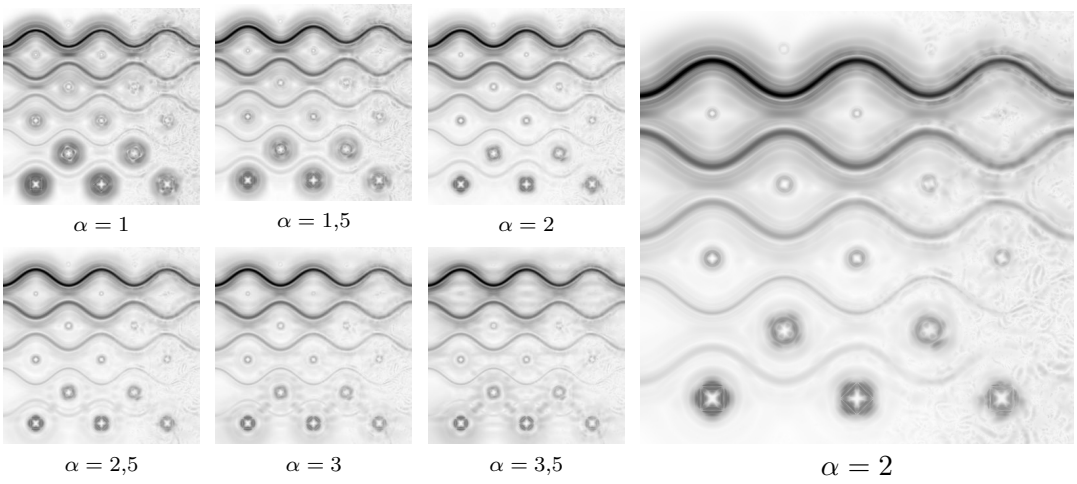


Figure 3.7 – Représentation du descripteur de structure linéaire, appliqué sur l'image de la Figure 3.3, pour différents paramètres α et en considérant le maximum sur des échelles σ allant de 1 à 28. Il apparaît que, lorsque α est trop petit ou trop grand, une auréole apparaît autour des différents objets. Nous avons affiché une seconde fois et agrandie, l'image obtenue pour $\alpha = 2$, cette valeur nous semblant être la plus pertinente sur cette image.

On pourrait reconsidérer les mêmes descripteurs que ceux définis pour le tenseur de structure, en considérant $\lambda_1(\sigma)$ et $\lambda_2(\sigma)$ les valeurs propres de cette matrice. Toutefois, remarquons que l'on cherche ici à détecter les structures linéaires et de ce fait, on s'intéresse en particulier aux réponses présentant une direction privilégiée (avec donc un grand écart entre les valeurs propres) et on considèrera, à l'échelle σ,

$$D_\sigma(p) \;=\; |\lambda_1(\sigma) - \lambda_2(\sigma)| \;=\; \sqrt{\left(s_2(\sigma) - s_1(\sigma)\right)^2 + 4s_3(\sigma)^2}, \tag{3.55}$$

en reprenant l'écriture $S_\sigma(p) = \begin{pmatrix} s_1(p) & s_3(p) \\ s_3(p) & s_2(p) \end{pmatrix}$. Pour éviter alors de se donner une échelle fixe, nous prendrons comme descripteur principal,

$$D(p) \;=\; \max_\sigma D_\sigma(p). \tag{3.56}$$

Dans la pratique, nous prendrons ce maximum sur un ensemble discret de valeurs, en incrémentant multiplicativement d'un facteur $\sqrt{2}$ entre chaque valeur. Par ailleurs, nous avons affiché dans la figure 3.7 les résultats obtenus pour différentes valeurs de α. Nous avons alors choisi de fixer α égal à 2, cette valeur semblant être la plus pertinente sur l'image utilisée en test.

3.3.4 Descripteurs de texture

Descripteurs d'Haralick Les descripteurs d'Haralick sont des descripteurs classiques de texture, proposés en 1973 par Haralick [HSD73]. Ils reposent sur les matrices de co-ocurrence. Étant donnée une image I et un vecteur $\Delta \in \mathbb{Z}^2$, ces matrices, dont la

largeur correspond au nombre de valeurs prises dans I, comptent pour chaque couple de valeurs possibles (a, b), le nombre de couples de pixels $(p, p+\Delta)$ présentant respectivement ces valeurs, c'est-à-dire

$$P_\Delta(a,b) \;=\; \frac{1}{Z}\,\mathrm{Card}\Big(\Big\{p \in \mathcal{D},\ (I_p, I_{p+\Delta}) = (a,b)\Big\}\Big)\,, \tag{3.57}$$

où Z est une constante de normalisation de sorte que la somme des coefficients de P_Δ soit égale à 1. Haralick propose alors une vingtaine de descripteurs de texture associés à chaque matrice de co-occurence, comme (sans les citer tous) le second moment angulaire, le contraste, l'entropie, l'homogénéité ou la dissimilarité, définies respectivement par

$$\begin{aligned}
\text{Second Moment Angulaire} \;&=\; \sum_{a,b} P_\Delta(a,b)^2,\\
\text{Contraste} \;&=\; \sum_{a,b} (a-b)^2 P_\Delta(a,b),\\
\text{Entropie} \;&=\; -\sum_{a,b} P_\Delta(a,b)\log P_\Delta(a,b),\\
\text{Homognéité} \;&=\; \sum_{a,b} \frac{P_\Delta(a,b)}{1+(a-b)^2},\\
\text{Dissimilarité} \;&=\; \sum_{a,b} |a-b| P_\Delta(a,b).
\end{aligned}$$

Dans la littérature, on retrouve souvent ces différents descripteurs pour les matrices de co-occurence associées aux directions $\Delta_1 = (d, 0)$, $\Delta_2 = (d, d)$, $\Delta_3 = (-d, d)$ et $\Delta_4 = (-d, -d)$. L'utilisation de plusieurs directions Δ, souvent nécessaire pour pouvoir caractériser les textures, entraîne une hausse significative de la dimension du descripteur de texture. Pour cette raison, nous avons préféré ne pas l'utiliser dans la pratique.

Descripteur d'ondelettes Pour constituer des descripteurs de textures, nous utilisons la décomposition en ondelettes des images. Introduit en 1984 par Alex Grossman et Jean Morlet [GM84], puis exprimé sous une forme multi-échelle par Yves Meyer et Stéphane Mallat en 1988 [Mal89], la décomposition en ondelettes consiste à capturer les informations d'oscillation présentes dans l'image, de manière locale et multi-échelle. Cette décomposition est largement utilisée dans de nombreux problèmes de traitement d'image, par exemple le débruitage, la compression d'image [CDVLL98], les problèmes de déconvolution [NCB04], ou encore l'analyse et la classification de texture [CK93].

La décomposition en ondelettes est constituée de trois parties, que nous pourrons qualifier d'horizontale, verticale et diagonale, construites à l'aide de trois filtres φ_{H}, φ_{V} et φ_{D}, correspondant à des fonctions de $\mathbb{R}^2$ dans $\mathbb{R}$. Ces trois filtres sont appliqués à l'image pour différentes positions et échelles. Étant donné un filtre $l \in \{H, V, D\}$, une position $p \in \mathcal{D}$ et une échelle $k \in \mathbb{N}$, on définit la réponse de l'image I à cette position et échelle par

$$f_l^k(p) \;=\; 2^{-k} \sum_{q \in D} \varphi_l\Big(\frac{p-q}{2^k}\Big)\, I(q) \;=\; \Big\langle I, \varphi_{k,p}^l \Big\rangle\,, \tag{3.58}$$

en l'écrivant comme un produit scalaire par une « ondelette », $\varphi^l_{k,p}$. Tout l'intérêt de la décomposition en ondelettes est alors que l'on peut choisir les filtres φ_{H}, φ_{V} et φ_{D} de sorte que les ondelettes $\varphi^l_{k,p}$ soient orthonormales, pour tout $l \in \{H, V, D\}$, $k \in \mathbb{N}$ et $p = (i\,2^{-k}, j\,2^{-k})$, avec $i, j \in \mathbb{Z}$.

Étant donné une taille de voisinage $d > 0$ et une échelle maximale, $k_m \geq 0$, nous allons alors considérer, comme descripteurs, « l'énergie » des ondelettes au voisinage d'un pixel $p = (p_x, p_y) \in \mathcal{D}$ et ce pour chaque filtre l et chaque échelle $k \in [\![0, k_m]\!]$, c'est-à-dire

$$F_{l,k}(p)^2 \;=\; \frac{1}{(1+2d/2^k)^2} \sum_{i=-d/2^k}^{d/2^k} \sum_{j=-d/2^k}^{d/2^k} f^k_l(p_x + i2^k, p_y + j2^k)^2. \tag{3.59}$$

Dans la pratique, nous prendrons un voisinage de rayon $d = 16$, les trois premières échelles ($k_m = 2$), qui donnera ainsi un descripteur de dimension neuf. En ce qui concerne le choix des ondelettes, nous prendrons de manière assez arbitraire les ondelettes de Daubechies [Dau92].

3.4 Modèle de changements par comparaison ponctuelle

3.4.1 Introduction

Rappel de la méthode Comme nous l'avons dit précédemment, nous ne cherchons pas ici à donner une définition générale de la notion de changement. Nous cherchons simplement à construire le premier des modèles « primaires », qui plus tard, fusionnés, serviront à discriminer les différents types de changements. De ce fait, il est important de garder en tête que la définition d'un changement que l'on prend ici est différente de celles utilisées dans les sections suivantes, pour les autres modèles primaires.

Philosophie du modèle Pour ce premier modèle, nous nous sommes concentrés sur une idée simple, à savoir qu'un changement correspondrait à une variation importante des valeurs des pixels entre les deux images. En effet, la quasi-totalité des changements que l'on peut imaginer, une couleur qui change, une tâche qui apparaît, un contour qui disparaît, une variation de texture, etc, peut être vu au travers de la variation significative des valeurs des pixels. La seule source de variabilité que l'on considère comme n'étant pas un changement est ici le bruit présent dans les images.

La comparaison directe dans la littérature Comme nous l'avons mentionné dans l'introduction, cette vision du changement comme variation importante des valeurs des pixels constitue l'approche la plus répandue dans la littérature. Si les différentes méthodes qui s'y rapportent peuvent sembler s'appuyer sur des concepts différents, elles suivent, dans la plupart des cas, deux phases distinctes :

— une phrase de prétraitement (recalage d'histogramme, normalisation moyenne variance, etc) et d'extraction de descripteurs d'image adaptés ;
— une phase de comparaison, consistant à décider du caractère de changement à partir de la différence des deux images, et à laquelle nous nous intéressons particulièrement ici.

Nous nous intéressons en particulier à la phase de comparaison dans un cadre non supervisé. Si l'on écarte les méthodes qui se contentent de fixer une valeur de seuil sur la norme de la différence des deux images, les approches aboutissant à un critère de décision à partir de la différence des deux images reposent sur la modélisation de cette différence par une variable aléatoire X. Deux hypothèses sont alors considérées, l'hypothèse H_0 où X ne correspond pas à un changement et l'hypothèse H_1 où X correspond à un changement. Il est alors possible d'utiliser le ratio de vraisemblance pour en déduire les changements,

$$\frac{p(x|H_0)}{p(x|H_1)}, \tag{3.60}$$

la valeur x correspondant à la différence mesurée des deux images et p étant la mesure de probabilité associée à X. Le problème est qu'il est très difficile d'estimer la valeur de $p(x|H_1)$. La méthode la plus répandue pour cela est de modéliser l'ensemble des différences sur toutes la paire d'image par un mélange deux gaussiennes, une première gaussienne correspondant aux changements, une seconde correspondant aux non-changements, comme le suggère Bruzzone et Prieto [BP00].

Cependant, cette modélisation reste approximative, d'autant plus que les changements envisagés ne concernent qu'un nombre réduit de pixels. Nous proposons de nous placer dans le cadre a contrario, qui présente l'avantage de n'utiliser que l'hypothèse H_0, facilement modélisable, pour en déduire un critère de décision.

3.4.2 Modélisation d'images

3.4.2.1 Modèle monospectral

Dans un premier, plaçons-nous dans le cadre d'images monospectrales. On suppose ainsi que chacune des deux images u et v ne disposent que d'une seule bande de fréquence. On prend donc

$$u, v : \mathcal{D} \to \mathbb{R},$$

ainsi que U et V, les représentations aléatoires de u et v. Par la suite, nous nous permettrons de confondre les fonctions de $\mathcal{D}$ dans $\mathbb{R}$ avec les vecteurs de $\mathbb{R}^{|\mathcal{D}|}$.

Modèle de bruit On considère ici, sans présence de changement, que les deux images correspondent à deux versions bruitées indépendamment d'une même image de base. Comme nous l'avons dit dans la section 2.1.2, le bruit d'acquisition présent au niveau de l'intensité de chaque pixel peut, sans trop d'erreur, modélisé par une loi gaussienne. On peut alors supposer qu'il existe une image $I_0 \in \mathbb{R}^{|\mathcal{D}|}$ telle que

$$U, V \;\sim\; \mathcal{N}\Big(I_0, \sigma_0^2 I_{|\mathcal{D}|}\Big)\,,$$

en notant σ_0^2 la variance commune du bruit et $I_{|\mathcal{D}|}$ la matrice identité de dimension $|\mathcal{D}|$. De part l'indépendance supposée de U et V, et en considérant $\varepsilon = V - U$, il apparaît que $\varepsilon \sim \mathcal{N}(0, 2\sigma_0^2 I_N)$ et on peut alors résumer le modèle de bruit sous la forme

$$\forall p \in \mathcal{D}, \qquad V_p \;=\; U_p + \varepsilon_p \qquad \text{où} \qquad \varepsilon_p \;\sim\; \mathcal{N}\Big(0, \sigma^2\Big)\,, \tag{3.61}$$

pour $\sigma^2 = 2\sigma_0^2$. Notons que l'on aurait parfaitement pu considérer deux variances différentes pour les deux images, σ_u^2 et σ_v^2, ce qui aurait simplement abouti à $\sigma^2 = \sigma_u^2 + \sigma_v^2$.

Cadre a contrario Comme nous l'avons précisé dans la section 3.1.1, nous nous plaçons dans le formalisme a contrario. Il s'agit donc de s'assurer que les fonctions de « régression de changement », ici F, que l'on seuillera pour en déduire les changements, respectent la propriété de majoration du nombre de fausses alarmes donnée dans l'équation (3.1), que l'on peut écrire ici

$$\forall \varepsilon > 0, \qquad \mathbb{E}_{(U,V)\sim H_0}\Big[\,\mathrm{Card}\,\{p \in \mathcal{D}, F(p,U,V) \leq \varepsilon\}\,\Big] \;\leq\; \varepsilon, \tag{3.62}$$

le modèle naïf H_0 étant ici donné par le modèle de bruit explicité dans l'équation (3.61) et l'ensemble des structures possibles (noté précédemment $\mathcal{I}$) correspondant à l'ensemble des pixels. On montre alors la proposition suivante.

Proposition 3.4.1. *En considérant erfc, la fonction d'erreur complémentaire, définies pour $x \geq 0$ par*

$$\mathrm{erfc}(x) \;=\; \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-t^2} dt, \tag{3.63}$$

la fonction définie pour tout $p \in \mathcal{D}$ par

$$\mathrm{NFA}(p,u,v) \;=\; |\mathcal{D}|\,\mathrm{erfc}\Big(\frac{|v_p - u_p|}{\sigma\sqrt{2}}\Big) \tag{3.64}$$

est un nombre de fausses alarmes optimal (selon la définition 3.1.3) pour le modèle naïf explicité dans l'équation (3.61).

Démonstration. Considérons la mesure de structure suivante,

$$\mu(p,u,v) \;=\; |v_p - u_p|, \tag{3.65}$$

ainsi qu'un ensemble de coefficients de pondération $\eta_p = |\mathcal{D}|$. L'inverse des η_p se sommant bien à 1, pour $p \in \mathcal{D}$, le théorème 3.1.2, relatif à la construction des nombres de fausses alarmes, nous assure alors que la fonction

$$F(p,u,v) \;=\; |\mathcal{D}|\,\mathbb{P}\big(\,\mu(p,U,V) \geq \mu(p,u,v)\,\big)\,,$$

pour $p \in \mathcal{D}$, est un nombre de fausses alarmes pour le modèle naïf explicité dans l'équation (3.61). Il reste alors à calculer cette probabilité. D'après le modèle naïf,

$$\mathbb{P}\big(\,\mu(p,U,V) \geq \mu(p,u,v)\,\big) \;=\; \mathbb{P}\big(\,|\varepsilon_p| \geq \mu(p,u,v)\,\big)\,,$$

où ε_p suit une distribution gaussienne centrée de variance σ^2. En notant $\delta = \mu(p,u,v)$ la mesure effectivement calculée, on trouve facilement que

$$\begin{aligned}
\mathbb{P}\big(\,|\varepsilon_p| \geq \delta\,\big) \;&=\; \frac{1}{\sqrt{2\pi\sigma^2}}\, 2\int_\delta^{+\infty} e^{-\frac{1}{2\sigma^2}x^2} dx \\
&=\; \frac{2}{\sqrt{2\pi\sigma^2}} \int_{\frac{\delta}{\sqrt{2}\sigma}}^{+\infty} e^{-t^2} \sigma\sqrt{2}\, dt \\
&=\; \mathrm{erfc}\Big(\tfrac{\delta}{\sigma\sqrt{2}}\Big) \;=\; 1 - \mathrm{erf}\Big(\tfrac{\delta}{\sigma\sqrt{2}}\Big)\,.
\end{aligned}$$

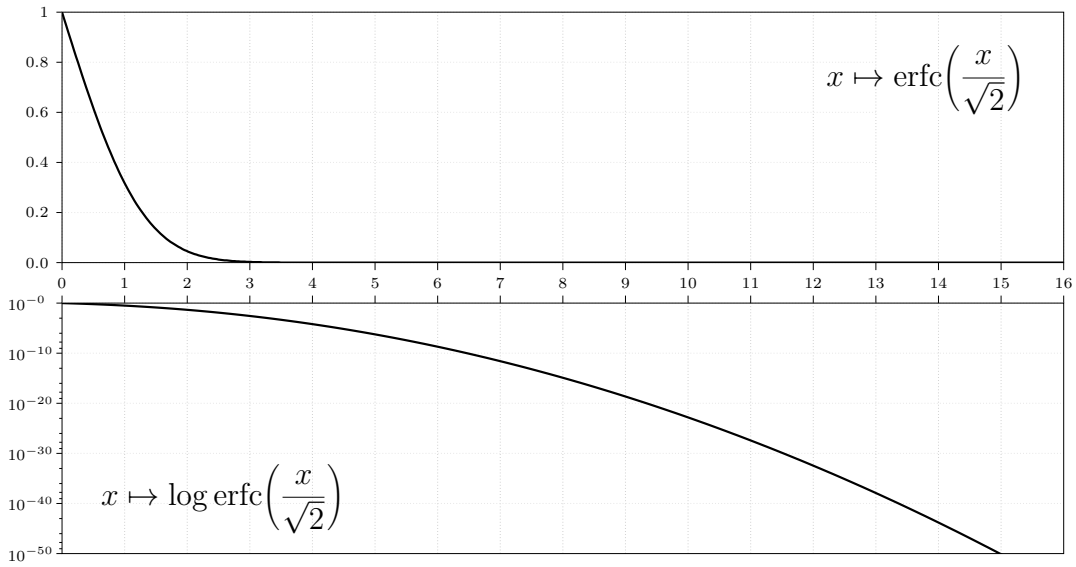


Figure 3.8 – Courbes représentatives de la fonction erreur avec et sans logarithme. Ces courbes représentent donc la répartition de la norme d'une variable gaussienne de variance 1 et correspodent donc au nombre de fausses alarmes donnée dans l'équation (3.64) à un coefficient multiplicatif près. On voit bien ici que, lorsque x augmente, la valeur du NFA devient rapidement très proche de 0 et peut ainsi demander une attention particulière pour ne pas faire d'erreur d'arrondi.

Et ainsi, pour tout $p \in \mathcal{D}$,

$$\begin{aligned} F(p,u,v) &= \eta_i \, \mathbb{P}\big(\mu(p,U,V) \geq \mu(p,u,v) \big) \\ &= |\mathcal{D}| \operatorname{erfc}\Big(\tfrac{\mu(p,u,v)}{\sigma\sqrt{2}}\Big) \\ &= |\mathcal{D}| \operatorname{erfc}\Big(\tfrac{|v_p - u_p|}{\sigma\sqrt{2}}\Big), \end{aligned} \tag{3.66}$$

qui est bien un nombre de fausses alarmes, comme annoncé. Par ailleurs, l'optimalité de ce dernier est assurée par la propriété 3.1.4, la variable aléatoire $\mu(p,U,V) = |\varepsilon_p|$ étant à densité. □

Erreur numérique et précision machine Comme on peut le voir dans la figure 3.8, les valeurs prises par ce nombre de fausses alarmes tendent très rapidement vers 0 lorsque la variation d'intensité entre les deux images augmente. Dans la pratique, il est donc préférable de calculer plutôt son logarithme,

$$-\log \mathrm{NFA}(p,u,v) \;=\; -\log|\mathcal{D}| - \log \operatorname{erfc}\left(\frac{|v_p - u_p|}{\sigma\sqrt{2}}\right),$$

pour éviter tout erreur d'arrondis à zéro. Il faut alors faire attention lors du calcul de $\log \operatorname{erfc}$. Pour une valeur x trop grande, il n'est pas possible l'appliquer le logarithme à

$\mathrm{erfc}(x)$, ce dernier étant arrondis à 0. Nous estimons ici cette expression en effectuant un développement asymptotique de la fonction erfc au voisinage de l'infini,

$$\mathrm{erfc}(x) \;=\; e^{-x^2}\frac{1}{\sqrt{\pi}}\sum_{n=0}^{\infty}\frac{(-1)^n(2n)!}{4^n\,n!}\frac{1}{x^{2n+1}},$$

en prenant

$$\log\mathrm{erfc}(x) \;\simeq\; -x^2-\frac{1}{2}\log(\pi)-\log\left(\sum_{n=0}^{A}\frac{(-1)^n(2n)!}{n!\,4^n}\frac{1}{x^{2n+1}}\right), \tag{3.67}$$

pour $A \in \mathbb{N}$ une constante suffisamment grande (fixée à 8 dans la pratique), qui ne pose plus de problème de précision.

3.4.2.2 Modèle multispectral avec bruit blanc

Considérons maintenant le cas d'images multispectrales, c'est-à-dire possédant K bandes de fréquences, pour $K \in \mathbb{N}^*$, avec ainsi

$$u, v : \mathcal{D} \to \mathbb{R}^K,$$

que nous pourrons voir, sous forme de vecteurs, comme des éléments de $\mathbb{R}^{|\mathcal{D}|\times K}$. Notons ici qu'il est nécessaire que les deux images possèdent le même nombre de canaux pour que nous puissions effectivement les comparer.

Modèle naïf Le modèle développé précédemment n'est que très peu impacté par le caractère multispectral des images. Le seul changement réel réside dans la dimension du bruit existant pour chaque pixel, qui passe de 1 à K. On fait ici la supposition que les différentes composantes du bruit, correspondant aux différents canaux des images, sont indépendantes et de même variance. Pour un pixel $p \in \mathcal{D}$, on écrit alors le modèle naïf sous la forme

$$V_p \;=\; U_p+\varepsilon_p \qquad \text{où} \qquad \varepsilon_p \sim \mathcal{N}\Big(0,\sigma^2 I_K\Big). \tag{3.68}$$

On peut alors étendre la proposition 3.4.1 aux cas d'images multispectrales en utilisant le théorème suivant.

Théorème 3.4.2 (Répartition d'une variable gaussienne centrée isotrope [Rub60])**.** *Soit $X \in \mathbb{R}^n$ une variable gaussienne, centrée et de matrice de covariance $\sigma^2 I_n$, alors pour tout $\delta \geq 0$,*

$$\mathbb{P}(\,\|X\|_2 \leq \delta\,) \;=\; \frac{\gamma\Big(\frac{K}{2},\frac{\delta^2}{2\sigma^2}\Big)}{\Gamma\Big(\frac{K}{2}\Big)} \qquad \textit{et} \qquad \mathbb{P}(\,\|X\|_2 \geq \delta\,) \;=\; \frac{\Gamma\Big(\frac{K}{2},\frac{\delta^2}{2\sigma^2}\Big)}{\Gamma\Big(\frac{K}{2}\Big)}, \tag{3.69}$$

où les deux fonctions Γ et la fonction γ correspondent respectivement à la fonction gamma, définie pour tout $a > 0$ par

$$\Gamma(a) \;=\; \int_0^{\infty} t^{a-1}e^{-t}dt, \tag{3.70}$$

et aux fonctions gamma incomplètes respectivement supérieure et inférieure, définies pour tout $a > 0$ et $x \geq 0$ par

$$\gamma(a,x) \;=\; \int_0^x t^{a-1}e^{-t}dt \qquad \text{et} \qquad \Gamma(a,x) \;=\; \int_x^\infty t^{a-1}e^{-t}dt. \tag{3.71}$$

Démonstration. En notant $\mathcal{B}(0,\delta)$ la boule centrée en 0 et de rayon δ, et S_{n-1} la sphère unité en dimension n, on obtient, en effectuant un changement de variable sphérique,

$$\begin{aligned}
\mathbb{P}(\,\|X\|_2 \leq \delta\,) &= \frac{1}{(2\pi\sigma^2)^{n/2}}\int_{\mathcal{B}(0,\delta)} e^{-\frac{\|x\|_2^2}{2\sigma^2}}dx. \\
&= \frac{1}{(2\pi)^{n/2}}\int_{\mathcal{B}(0,\delta/\sigma)} e^{-\frac{\|x'\|_2^2}{2}}dx'. \\
&= \frac{1}{(2\pi)^{n/2}}\int_{r=0}^{\delta/\sigma}\int_{\theta\in S_{n-1}} e^{-\frac{r^2}{2}}r^{n-1}d\theta dr \\
&= \frac{|S_{n-1}|}{(2\pi)^{n/2}}\int_{t=0}^{\delta^2/2\sigma^2} e^{-t}(2t)^{\frac{n}{2}-1}dt \qquad \text{avec } 2t{=}r^2 \\
&= \tfrac{1}{\Gamma(\frac{n}{2})}\,\gamma\Big(\tfrac{n}{2},\tfrac{\delta^2}{2}\Big),
\end{aligned}$$

où $|S_{n-1}| = 2\sqrt{\pi}^n\Gamma(\frac{n}{2})^{-1}$ est le volume de S_{n-1}. La seconde égalité s'obtient alors directement par propriété des fonctions gamma incomplètes, et on trouve bien

$$\mathbb{P}(\,\|X\|_2 \geq \delta\,) \;=\; 1 - \frac{\gamma\Big(\frac{n}{2},\frac{\delta^2}{2\sigma^2}\Big)}{\Gamma(\frac{n}{2})} \;=\; \frac{\Gamma\Big(\frac{n}{2},\frac{\delta^2}{2\sigma^2}\Big)}{\Gamma(\frac{n}{2})}. \tag{3.72}$$

□

Proposition 3.4.3. *La fonction définie pour tout $p \in \mathcal{D}$ par*

$$\mathrm{NFA}(p,u,v) \;=\; \frac{|\mathcal{D}|}{\Gamma(\frac{K}{2})}\Gamma\left(\frac{K}{2},\frac{\|v_p-u_p\|_2^2}{2\sigma^2}\right), \tag{3.73}$$

est un nombre de fausses alarmes optimal (selon la définition 3.1.3) pour le modèle naïf défini dans l'équation (3.68).

Démonstration. Considérons la mesure de structures suivante,

$$\mu(p,u,v) = \|v_p - u_p\|_2\,.$$

Le théorème 3.1.2 nous assure alors que la fonction

$$F(p,u,v) \;=\; |\mathcal{D}|\,\mathbb{P}(\,\mu(p,U,V) \geq \mu(p,u,v)\,)\,,$$

pour $p \in \mathcal{D}$, est un nombre de fausses alarmes pour le modèle naïf donnée dans l'équation (3.68). Comme précédemment, il s'agit alors de calculer, pour $\delta = \mu(p,u,v)$ donné, la probabilité

$$\mathbb{P}(\,\mu(p,U,V) \geq \delta\,) \;=\; \mathbb{P}(\,\|\varepsilon_p\|_2 \geq \delta\,)\,,$$

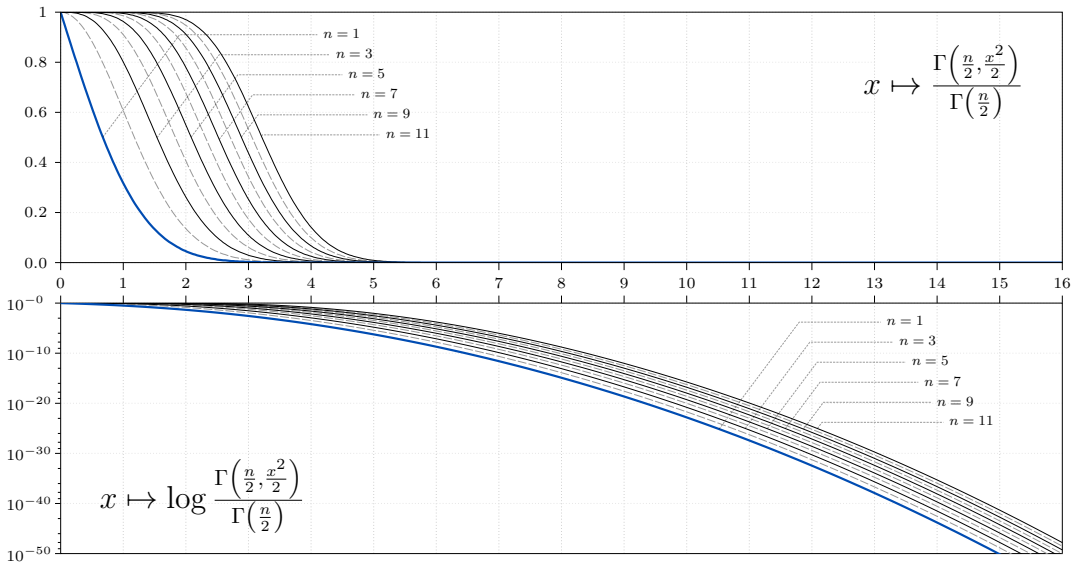


Figure 3.9 – Courbes représentatives de la fonction gamma incomplète supérieure pour différent paramètre de dimension, avec et sans logarithme. Ces courbes représentent donc la répartition de la norme d'une variable gaussienne de variance 1 et de dimension n. La courbe bleue, épaisse, correspond au paramètre $n = 1$ et représente donc aussi la fonction $x \mapsto \mathrm{erfc}(x/\sqrt{2})$. Les autres courbes correspondent successivement aux valeurs de n compris entre 2 et 11, ces courbes étant rangées de manière croissante. Nous avons affiché ces courbes en alternant courbes pleines et en pointillés pour une meilleure lecture et n'avons indiqué la valeur de n que pour les nombres impairs pour ne pas surcharger la figure. On voit bien ici à quel point le nombre de fausses alarmes se converge vite vers 0 lorsque x, représentant ici la norme de la différence des deux images, augmente.

où ε_p suit une distribution gaussienne centrée isotrope en dimension K. Le théorème 3.4.2 nous donne alors directement

$$F(p, u, v) \;=\; \frac{|\mathcal{D}|}{\Gamma(\frac{K}{2})}\Gamma\left(\frac{K}{2}, \frac{\|v_p - u_p\|_2^2}{2\sigma^2}\right), \tag{3.74}$$

ce qui correspond bien à l'expression annoncée.

Par ailleurs, puisque la variable aléatoire $\mu(p, U, V)$ est à densité, la propriété 3.1.4 nous assure l'optimalité de ce NFA. □

Erreur numérique et précision machine À nouveau ici, les valeurs du NFA peuvent être particulièrement faibles, souvent inférieur à 10^{-100}, comme le montre la figure 3.9. Dans la pratique, il est donc préférable de calculer le logarithme du NFA,

$$-\log \mathrm{NFA}(p, u, v) \;=\; \log \Gamma\left(\frac{K}{2}\right) - \log |\mathcal{D}| - \log \Gamma\left(\frac{K}{2}, \frac{\|v_p - u_p\|_2^2}{2\sigma^2}\right).$$

Pour pouvoir évaluer ces fonctions sans problème d'arrondis à zéro, nous avons utilisé la méthode de Abergel et co-auteur [AM16], dont l'implémentation est suffisamment rapide pour pouvoir être appliquée à chaque pixel de l'image, avec un faible coût total.

3.4.2.3 Modèle multispectral avec bruit corrélé

Propriété étendue au cas d'un bruit corrélé On se place dans le même cadre que précédemment, avec ainsi

$$u, v : \mathcal{D} \to \mathbb{R}^K,$$

pour $K \in \mathbb{N}^*$. Seulement, on tient ici compte des corrélations des différentes composantes du bruit en écrivant le modèle naïf sous la forme

$$V_p \;=\; U_p + \varepsilon_p \qquad \text{où} \qquad \varepsilon_p \sim \mathcal{N}(0, \Sigma_0)\,, \tag{3.75}$$

où Σ_0 est une matrice symétrique définie positive de dimension K. On peut alors étendre la proposition 3.4.3 de la manière suivante.

Proposition 3.4.4. *La fonction définie pour tout $p \in \mathcal{D}$ par*

$$\mathrm{NFA}\,(p, u, v) \;=\; \frac{|\mathcal{D}|}{\Gamma\left(\frac{K}{2}\right)} \Gamma\left(\frac{K}{2}, \frac{1}{2}\,(u_p - v_p)^\mathsf{T}\,\Sigma_0^{-1}(u_p - v_p)\right), \tag{3.76}$$

est un nombre de fausses alarmes optimal (selon la définition 3.1.3) pour le modèle naïf défini dans l'équation (3.75), Σ_0 étant la matrice de covariance du bruit présente dans cette même équation.

Démonstration. On considère ici la mesure de structure

$$\mu(p, u, v) \;=\; (u_p - v_p)^\mathsf{T}\,\Sigma_0^{-1}(u_p - v_p) \tag{3.77}$$

et on applique le théorème 3.1.2, qui nous assure que

$$F(p, u, v) \;=\; |\mathcal{D}|\,\mathbb{P}(\,\mu(p, U, V) \geq \mu(p, u, v)\,)\,,$$

pour $p \in \mathcal{D}$, est un nombre de fausses alarmes. Puisque Σ_0 est une symétrique définie positive, il existe une unique matrice symétrique définie positive $\Sigma_0^{-1/2}$ vérifiant $({\Sigma_0}^{-1/2})^2 = {\Sigma_0}^{-1}$, et on peut alors écrire

$$\mu(p, U, V) \;=\; \left\|\Sigma_0^{-1/2}\,(V_p - U_p)\right\|_2^2 \;=\; \left\|\Sigma_0^{-1/2}\varepsilon_p\right\|_2^2. \tag{3.78}$$

Or, puisque ε_p suit une distribution gaussienne centrée, de matrice de covariance Σ_0, alors $\Sigma_0^{-1/2}\varepsilon_p$ est une variable gaussienne de matrice de covariance

$$\Sigma_0^{-1/2}\,\Sigma_0\,\left(\Sigma_0^{-1/2}\right)^\mathsf{T} \;=\; I_K,$$

où I_K est la matrice identité de dimension K. En notant $\delta = \mu(p, u, v)$, le théorème 3.4.2 nous donne alors

$$F(p, u, v) \;=\; |\mathcal{D}|\,\mathbb{P}\left(\left\|\Sigma_0^{-1/2}\varepsilon_p\right\|_2^2 \geq \delta\right) \;=\; \frac{|\mathcal{D}|}{\Gamma\left(\frac{K}{2}\right)}\Gamma\left(\frac{K}{2}, \frac{\sqrt{\delta}}{2}\right)$$

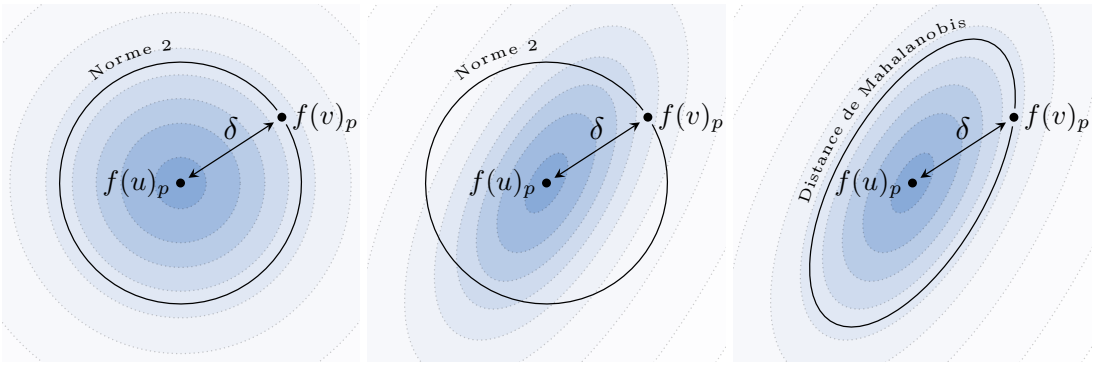


Figure 3.10 – Schéma représentant l'intérêt de la distance de Mahalanobis devant la norme 2 dans le cas d'un bruit corrélé. Sur chacune des figures sont affichés des points correspondant aux descripteurs $f(u)_p$ et $f(v)_p$ ainsi qu'une représentation de la distance δ qui les sépare. Les cercles ou ellipses représentent des lignes de niveau de la distribution de $f(V)_p$, c'est-à-dire des courbes passant par tous les points partageant une même valeur de densité de probabilité, les couleurs remplissant ces courbes étant d'autant plus foncées que la valeur de densité est élevée. Nous avons représenté en trait plein et passant par v_p la sphère centrée en u_p et de rayon δ correspondant à la métrique choisie. De gauche à droite, on trouve respectivement : la norme 2 pour une distribution gaussienne indépendante ; la norme 2 pour une distribution gaussienne corrélée ; la distance de Mahalanobis pour une distribution gaussienne corrélée. Notons que le premier cas, le trait correspond aussi à la sphère pour la distance de Mahalanobis.

et on retrouve bien le résultat annoncé. Notons que, comme précédemment, l'optimalité de ce NFA est assurée par la propriété 3.1.4, la variable aléatoire $\mu(p, U, V)$ étant à densité.

□

Distance de Mahalanobis La mesure utilisé ici correspond à la distance de Mahalanobis, que l'on peut noter

$$d_\Sigma(x, y) \;=\; \sqrt{(x-y)^\mathsf{T}\,\Sigma^{-1}(x-y)}, \tag{3.79}$$

pour $x, y \in \mathbb{R}^K$ et Σ, une matrice symétrique définie positive. La distance L^2 n'est pas capable ici de prendre en compte les différentes corrélation des différentes composantes, contrairement à la distance de Mahalanobis. En reconsidérant $\Sigma^{-1/2}$, la matrice racine carré de Σ^{-1} définie dans la démonstration précédente et vérifiant $(\Sigma^{-1/2})^2 = \Sigma^{-1}$, et en écrivant

$$d_\Sigma(x, y) \;=\; \left\|\Sigma^{-1/2}(x-y)\right\|_2, \tag{3.80}$$

il apparaît que la distance de Mahalanobis correspond à une déformation de la norme 2, suivant la transformation correspondant à $\Sigma^{1/2}$, comme nous l'avons représenté dans la figure 3.10. En prenant $\Sigma = \Sigma_0$, cette déformation permet de décorréler les différentes composantes de l'espace.

Utilisation de la norme infinie Si la distance de Mahalanobis exige de considérer la norme 2 dans l'équation (3.77), remarquons que rien n'empêche en réalité de considérer d'autres normes. Ainsi, on obtient ainsi une expression différente en considérant la norme infinie.

Proposition 3.4.5. *La fonction définie pour tout $p \in \mathcal{D}$ par*

$$\mathrm{NFA}\,(p,u,v) \;=\; |\mathcal{D}| \left(1 - \mathrm{erf}\left(\frac{1}{\sqrt{2}}\left\|\Sigma_0^{-1/2}\,(v_p - u_p)\right\|_\infty\right)^K\right), \tag{3.81}$$

où

$$\mathrm{erf}(x) \;=\; \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt, \tag{3.82}$$

est un nombre de fausses alarmes pour le modèle naïf défini dans l'équation (3.75).

Démonstration. Soit

$$\mu(p,u,v) \;=\; \left\|\Sigma_0^{-1/2}\,(v_p - u_p)\right\|_\infty, \tag{3.83}$$

on a alors, pour $\delta \geq 0$ et en posant $\widetilde{\varepsilon}_p = \Sigma_0^{-1/2}\,(v_p - u_p)$ (qui suit une distribution gaussienne centrée réduite),

$$\begin{aligned}
\mathbb{P}(\,\|\varepsilon_p\|_\infty \geq \delta\,) &= 1 - \mathbb{P}\left(\bigcap_{i\in[\![1,K]\!]} |\varepsilon_{p,i}| \leq \delta\right) \\
&= 1 - \prod_{i\in[\![1,K]\!]} \mathbb{P}(\,|\varepsilon_{p,i}| \leq \delta\,) \\
&= 1 - \mathrm{erf}\left(\frac{\delta}{\sqrt{2}}\right)^K,
\end{aligned} \tag{3.84}$$

et le théorème 3.1.2 nous assure donc bien le résultat annoncé. □

3.4.2.4 Modèle multispectrale avec descripteurs linéaires

L'utilisation directe des intensités brutes des différents canaux des images est dans la pratique d'un intérêt limitée. Pour dépasser cela, on se donne toujours deux images u et v, mais cette fois-ci, on considère de plus un ensemble de K' descripteurs, $K' \in \mathbb{N}^*$, stationnaires et locaux, comme précisé dans la section 3.3.1, que l'on suppose de plus linéaires.

La modélisation du bruit sur les images, constituant le modèle naïf de notre cadre a contrario, ne change pas et on suppose, comme précédemment, que

$$V_p \;=\; U_p + \varepsilon_p \qquad \text{où} \qquad \varepsilon_p \sim \mathcal{N}(0,\Sigma_0)\,, \tag{3.85}$$

où Σ_0 est une matrice symétrique définie positive de dimension K.

Descripteurs en dimension 1 Dans un premier temps, pour simplifier les notations, on suppose que les images sont monospectrales et donc à valeur dans $\mathbb{R}$, avec ainsi $K = 1$.

Comme précisé dans la section 3.3.1, on suppose que ces descripteurs sont stationnaires et locaux, c'est-à-dire qu'il existe une distance $d > 0$ de sorte que les valeurs des descripteurs au niveau d'un pixel $p \in \mathcal{D}$ ne dépendent que de celles des pixels voisins situés à une distance d'au plus d, et ce de manière indépendante de la position p. En notant alors P_0 le carré de $\mathbb{Z}^2$ centré en 0 et de côté $2d+1$, les descripteurs étant linéaires, on peut montrer que chaque descripteur, que nous noterons f^k pour $k \in [\![1, K']\!]$, peut alors être représenté comme un produit de convolution par un noyau $W^k \in \mathbb{R}^{|P_0|}$, c'est-à-dire que pour tout $p \in \mathcal{D}$,

$$f^k(u)_p \;=\; (W^k * u)_p \;=\; \sum_{q \in P_0} u_{p-q} W^k_q. \tag{3.86}$$

Notons bien que, comme annoncé dans la section 3.2.1, nous supposons qu'il n'y a aucun problème de bord et que les images sont aussi définies en dehors de $\mathcal{D}$. Par la suite, on notera $f(u)_p = \left(f^k(u)_p\right)_k \in \mathbb{R}^{K'}$ la concaténation de ces descripteurs.

Descripteurs en dimension K Dans le cas général, où chaque image possède K canaux, et où u_p, v_p et ε_p sont dans $\mathbb{R}^K$, chaque descripteur f^k, pour $k \in [\![1, K']\!]$, peut être représenté par un noyau W^k de taille P_0 et de dimension K de sorte que, pour tout $p \in \mathcal{D}$

$$f^k(u)_p \;=\; \sum_{q \in P_0} u_{p-q}{}^{\mathsf{T}} W^k_q \;=\; \sum_{q \in P_0} \sum_{c=1}^{K} u_{p-q,c} W^k_{q,c}, \tag{3.87}$$

ce qui correspond à une sorte de produit de convolution sur des images en dimension K.

On peut alors étendre la proposition 3.4.4 dans cas où l'on utilise des descripteurs linéaires.

Proposition 3.4.6. *Soit f^k, pour $k \in [\![1, K']\!]$, des descripteurs linéairement indépendants f^k, tels que définis précédemment dans l'équation (3.87). En considérant la concaténation de ces descripteurs, $f(u)_p = \left(f^k(u)_p\right)_{k \in [\![1,K']\!]}$, pour tout $p \in \mathcal{D}$, et en définissant*

$$\widetilde{\Sigma}_f \;=\; \left(\sum_{p \in P_0} W^{k\,\mathsf{T}}_p \Sigma_0 W^l_p \right)_{k,l \in [\![1,K']\!]}, \tag{3.88}$$

les noyaux W^k et P_0 représentant les descripteurs (voir équation (3.87)), la fonction définie pour tout $p \in \mathcal{D}$ par

$$\mathrm{NFA}\,(p, u, v) \;=\; \frac{|\mathcal{D}|}{\Gamma\left(\frac{K}{2}\right)} \Gamma\left(\frac{K}{2}, \frac{1}{2} \left\| \widetilde{\Sigma}_f^{-1/2} (f(v)_p - f(u)_p) \right\|_2^2 \right) \tag{3.89}$$

est un nombre de fausses alarmes optimal (selon la définition 3.1.3) pour le modèle naïf défini dans l'équation (3.75).

Démonstration. Avant toute chose, remarquons que la matrice $\widetilde{\Sigma}_f$ est bien symétrique définie positive, ce qui nous assure que $\Sigma_f^{-1/2}$ est bien définie. En effet, il suffit de remarquer que cette matrice s'écrit comme la matrice de Gram associée à la famille des descripteurs $(f^k)_{k\in[\![1,K']\!]}$, en considérant comme produit scalaire

$$\left\langle f^k, f^l \right\rangle \;=\; \sum_{p\in P_0} {W_p^k}^{\mathsf{T}} \Sigma_0 W_p^l, \tag{3.90}$$

qui est bien une forme bilinéaire symétrique définie positive, la matrice Σ_0 étant une matrice symétrique définie positive. La famille des $(f^k)_{k\in[\![1,K']\!]}$ étant linéairement indépendante par hypothèse, sa matrice de Gram, $\widetilde{\Sigma}_f$, est donc bien symétrique définie positive.

Considérons Δ, la différence des descripteurs de deux images aléatoire U et V tirées suivant le modèle naïf donné dans l'équation (3.75), c'est-à-dire

$$(\Delta_p)_k \;=\; f^k(V)_p - f^k(U)_p \;=\; f^k(\varepsilon)_p \;=\; \sum_{q\in P_0} {W_q^k}^{\mathsf{T}} \varepsilon_{p-q}, \tag{3.91}$$

pour tout pixel $p \in \mathcal{D}$ et tout $k \in [\![1, K']\!]$. Les descripteurs étant linéaires, Δ_p apparaît comme une somme de variables aléatoires gaussiennes centrées et suit donc une distribution gaussienne centrée. Notons $\widetilde{\Sigma}_f$ sa matrice de covariance. Pour tout $k, l \in [\![1, K']\!]$, on a

$$\begin{aligned}
(\widetilde{\Sigma}_f)_{k,l} &= \mathbb{E}\big[\,(\Delta_p)_k\,(\Delta_p)_l\,\big] \\
&= \mathbb{E}\left[\left(\sum_{q\in P_0} {W_q^k}^{\mathsf{T}} \varepsilon_{p-q}\right)\left(\sum_{r\in P_0} {\varepsilon_{p-r}}^{\mathsf{T}} W_r^l\right)\right] \\
&= \sum_{q,r\in P_0} \mathbb{E}\Big[\,{W_q^k}^{\mathsf{T}} \varepsilon_{p-q} {\varepsilon_{p-r}}^{\mathsf{T}} W_r^l\,\Big] \\
&= \sum_{q,r\in P_0} {W_q^k}^{\mathsf{T}} \mathbb{E}\Big[\,\varepsilon_{p-q} {\varepsilon_{p-r}}^{\mathsf{T}}\,\Big]\, W_r^l \\
&= \sum_{q,r\in P_0} {W_q^k}^{\mathsf{T}} \Sigma_0 W_r^l \mathbb{1}_{\{q=r\}} \;=\; \sum_{p\in P_0} {W_p^k}^{\mathsf{T}} \Sigma_0 W_p^l\ .
\end{aligned} \tag{3.92}$$

Pour trouver ce résultat, il suffit de remarquer que, la fonction f étant linéaire, la distribution de $(f(U), f(V))$ restera gaussienne. Aussi, il suffit alors d'estimer la matrice de covariance, notée $\widetilde{\Sigma}_f$, de cette nouvelle distribution, pour pouvoir alors directement appliquer la proposition 3.4.4. □

Cas d'un bruit blanc Dans la majorité des cas, on peut approcher le bruit des images satellites par un bruit blanc, nous donnant ainsi $\Sigma_0 = \sigma^2 I_K$, pour $\sigma^2 > 0$. Dans ce cas, il est possible de séparer les contributions du bruit et des descripteurs dans la matrice $\widetilde{\Sigma}_f$ et en posant

$$\Sigma_f \;=\; \left(\sum_{p\in P_0} {W_p^k}^{\mathsf{T}} W_p^l\right)_{k,l\in[\![1,K']\!]}. \tag{3.93}$$

on a alors $\widetilde{\Sigma}_f = \sigma^2 \Sigma_f$ et le nombre de fausses alarmes donnés dans l'équation 3.89 peut alors s'écrire

$$\mathrm{NFA}\,(p,u,v) \;=\; \frac{|\mathcal{D}|}{\Gamma\left(\frac{K}{2}\right)} \Gamma\left(\frac{K}{2}, \frac{1}{2\sigma^2} \left\| \Sigma_f^{-1/2} (f(v)_p - f(u)_p) \right\|_2^2 \right). \tag{3.94}$$

3.4.3 Résultats et limitations

3.4.3.1 Résultats sur les exemples synthétiques

Descripteurs utilisés Pour évaluer les résultats de ce modèle, nous nous sommes restreint à trois descripteurs différents, à savoir : la nuance de gris (voir section 3.3.2), qui donne un descripteur linéaire de dimension 1 ; l'estimation du gradient orienté, calculé à partir des filtres de Sobel (voir section 3.3.3.1), qui donne un descripteur linéaire de dimension 2 ; un descripteur de textures, de dimension 9, basé sur des coefficients d'ondelettes (voir section 3.3.4), qui est quadratique et dont l'utilisation sera justifiée plus tard, dans la section 3.4.4.2. Nous nous limitons à ces trois descripteurs pour plusieurs raisons. D'une part, il est difficile de s'intéresser à tous les détecteurs obtenus en combinant le modèle par comparaison ponctuelle avec chaque descripteur possible. D'autre part, en choisissant un descripteur basé sur l'intensité brute, un descripteur « géométrique » et un descripteur de texture, on garde tout de même une vision relativement globale du comportement du détecteur pour différents types de descripteurs.

Remarques générales Nous avons appliqué ce modèle sur les exemples synthétiques présentés dans la Figure 2.9 et affiché les résultats dans la Figure 3.11, en seuillant les valeurs du NFA à 10^{-10}.

Remarquons tout d'abord que, par construction, ce modèle détecte tous les pixels dont les descripteurs présentent des variations significatives entre les deux images. Ainsi il ne possédera aucune robustesse quant à des éventuels décalages, variation de texture, changement de contraste ou autre, et les seules robustesses qu'auront les détecteurs basés sur ce modèle, seront celles apportées par les descripteurs. Ainsi, comme attendu, le détecteur basé sur l'intensité en nuance de gris détecte à peu près tous les changements, que cela soit les apparitions d'objets dans chacune des paires d'images, les translations légères de la paire 2, le changement de contraste local de la paire 3, les variations de textures de la paire 4 ainsi que les changements globaux de la paire 5.

De manière tout aussi attendue, l'utilisation du gradient permet de ne pas détecter le changement de contraste local présent dans la paire 3. En effet, le gradient ne permettant que d'évaluer localement les variations spatiales d'intensités, il sera insensible à des changements n'incluant pas de variation de bords. Cela explique aussi pourquoi seuls les bords des champs dans la paire 5 sont détectés. Il est intéressant de noter que la variation de macro-texture de la paire 4 n'est pas non plus détectée, pour des raisons similaires.

En ce qui concerne le descripteur de texture, les résultats peuvent sembler plus étonnants. En effet, si le détecteur permet bien de détecter les différentes apparitions d'objets dans les paires 1, 2, 3 et 4, ça n'est déjà pas le cas dans la paire 5, où le

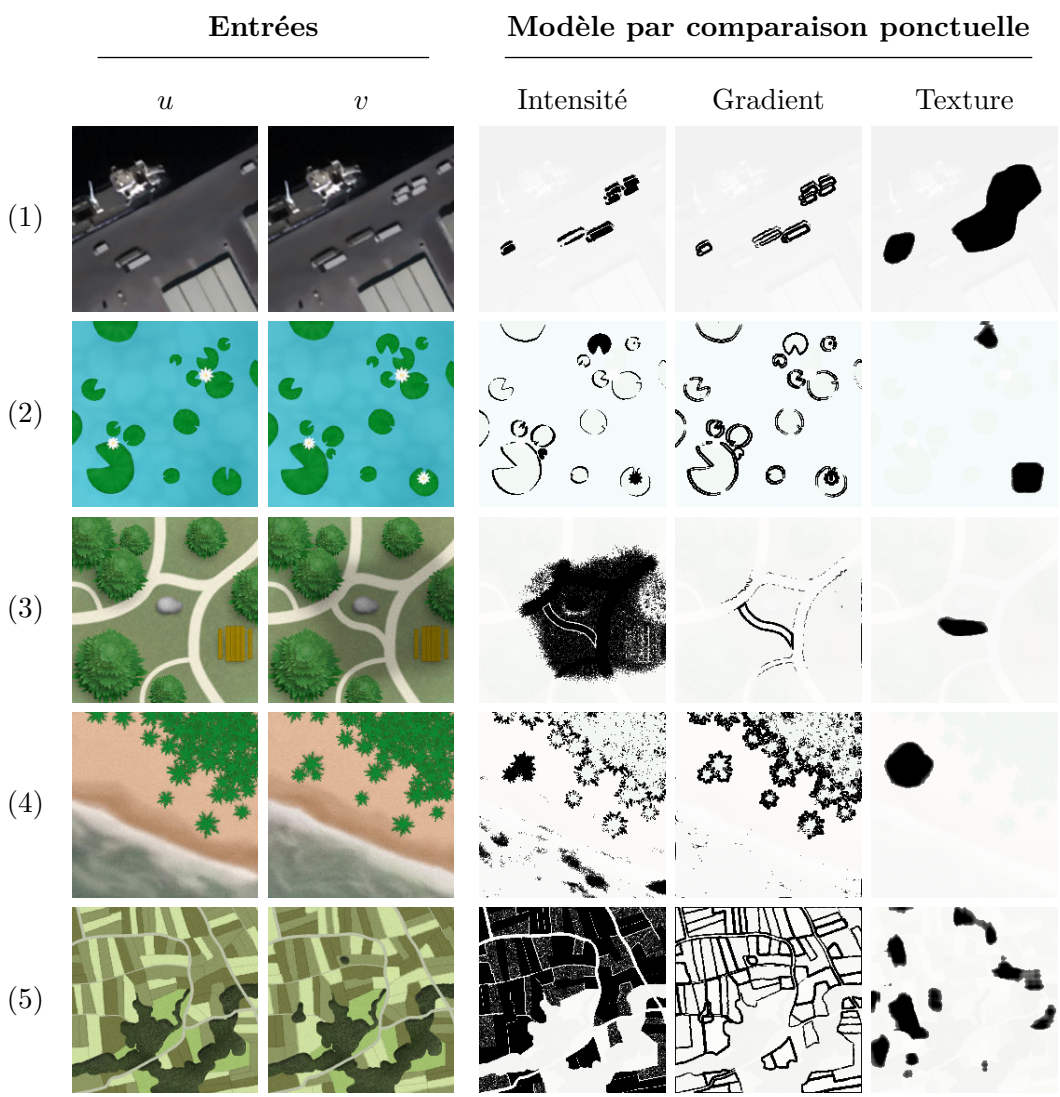


Figure 3.11 – Résultats du modèle par comparaison ponctuelle sur les exemples synthétiques présentés dans la Figure 2.9. Trois types de descripteurs ont été utilisés : l'intensité en nuance de gris, une estimation du gradient et un descripteur de texture, de dimension 9, basé sur la décomposition de l'image en ondelettes. Chaque ligne représente un exemple et est constituée successivement des deux images u et v suivies des résultats du modèle pour chacun des descripteurs considérés. Dans les images de résultat, nous avons représenté en blanc les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$, le tout affiché sur une version assombrie de la seconde image.

détecteur rate l'apparition d'un petit objet. Par ailleurs, le détecteur présente bien des robustesses aux variations de texture, seulement il semble aussi robuste aux décalages faibles de la paire 2 et aux variations de bords de la paire 5. Tout cela peut être expliqué par l'étendue spatiale du descripteur utilisé.

Étendue spatiale et taille de changement Considérons le cas simple d'un descripteur moyennant l'intensité de l'image sur N pixels, $N \in \mathbb{N}$ et considérons les N pixels qui moyennés donneront le descripteur d'un pixel p, $p \in \mathcal{D}$. Supposons à présent que, parmi ces N pixels, k d'entre eux, $k \in [\![0, N]\!]$, subissent une hausse d'intensité d'une valeur de $\delta > 0$. Ici, l'erreur mesurée sera de $k\delta/N$ et en supposant un bruit pixellique de variance $\sigma^2 > 0$, la variance de cette mesure est alors de $N\sigma^2/N^2$. On peut alors directement, à partir de l'équation (3.66), en déduire une l'expression du nombre de fausses alarmes obtenue pour cette hausse des intensités sur k pixel parmi les N pixels dont on moyenne les intensités, et que l'on peut écrire

$$|\mathcal{D}| \operatorname{erfc}\left(\frac{k\delta}{N} \times \frac{\sqrt{N}}{\sigma\sqrt{2}}\right) = |\mathcal{D}| \operatorname{erfc}\left(\frac{k\delta}{\sqrt{2N}\sigma}\right).$$

Ce changement étant d'autant plus détecté que cette valeur est proche de 0, c'est-à-dire que $k\delta/\sqrt{2N}\sigma$ est grand, il apparaît alors directement qu'en choisissant une valeur de N élevé, il est possible de faire baisser suffisamment la significativité des « petits » changements, ne faisant intervenir qu'un nombre réduit de pixels, pour que ces derniers ne soient pas détectés.

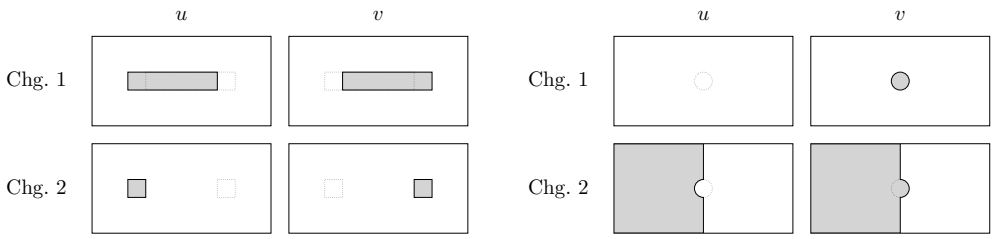
Dans le cas du descripteur de texture, celui-ci étant calculé sur une étendue 32×32 pixels, il n'est donc pas étonnant de ne pas observer de détections de taille réduite.

3.4.3.2 Limitations du modèle

Remarque générale sur la recherche de limitations Lorsque l'on cherche ici à exhiber une limitation du modèle, il faut garder en tête que l'on ne se donne aucune définition précise de ce que doit être ou pas un changement. De ce fait, exhiber un cas où le modèle, pour un descripteur donné, ne donne aucune détection ne suffit pas à montrer une de ses failles. Pour cela, une première méthode serait présenter des exemples étant soit toujours (ou jamais) détectés, peu importe les descripteurs utilisés. Toutefois, à cause du niveau de bruit variable et de la multitude de descripteurs possibles, il peut être difficile d'exhiber de tels changements.

Nous nous ramenons à une seconde méthode. Il s'agit alors de présenter des paires de changements ne pouvant être distingués par le modèle, ce dernier ne pouvant pas détecter l'un sans détecter l'autre.

Symétrie du modèle Une première limitation découle de la symétrie du modèle. Au vu de la mesure de structure donnée équation (3.77), le modèle par comparaison ponctuelle apparaît comme un modèle symétrique, sans le sens où inverser les deux images ne modifierait en rien le résultat obtenu. De ce fait, il apparaît directement que le modèle est incapable de distinguer une hausse d'une baisse d'intensité, et ainsi une apparition d'une disparition.



(a) Changements indistinguables basés sur une translation.

(b) Changements indistinguables basés sur une extension.

Figure 3.12 – Deux exemples schématiques de paires de changements indistinguable pour le modèle de changement par comparaison ponctuelle, dans le sens où l'on obtiendrait des résultats identiques en appliquant le modèle à l'une ou l'autre des paires d'images. Dans chacun des deux cas, chaque ligne représente un changement au travers des deux images u et v. Pour faciliter la visualisation des changements, nous avons fait figurer en pointillé, sur chaque image, l'emplacement des objets présents sur la seconde image. Dans les deux cas, il n'est pas possible pour le modèle de prendre en compte le contexte, si ce dernier est constant entre les deux images.

Indistinguabilité entre décalage et apparition D'un autre côté, à cause de la linéarité des descripteurs que l'on utilise, le modèle ne peut percevoir que la différence, $u - v$, des deux images. Il suffit, pour voir cela, de reprendre la mesure de structure donnée dans l'équation (3.77) et d'écrire

$$\mu(p, u, v) \;=\; \left\| \Sigma_0^{-1/2} \left(f(v)_p - f(u)_p \right) \right\|_2 \;=\; \left\| \Sigma_0^{-1/2} \left(f(v-u)_p \right) \right\|_2 .$$

Ainsi, ce modèle, dans sa version de base, ne pourra pas distinguer deux paires d'images présentant une même différence entre elles. En se basant sur cette remarque, nous avons alors construit un exemple, présenté dans la Figure 3.12, mettant en valeur une limitation de ce modèle. Regardons le premier exemple. Dans la première paire, on présente un objet se décalant légèrement, tandis que dans la seconde paire, la partie centrale n'existant pas, il s'agirait plus probablement d'une disparition d'un côté et une apparition de l'autre. Ici, la partie centrale de l'objet n'est en fait pas perçue par le modèle et de ce fait, il n'a aucun moyen de distinguer une translation faible d'une apparition et disparition (ou d'une translation beaucoup plus importante).

Le deuxième exemple présente un cas similaire. Dans chacune des deux paires, un objet circulaire apparaît au milieu de l'image. Seulement dans un cas, cet objet est isolé, tandis que dans le second cas, l'objet apparaît comme une extension d'une structure déjà existante. Le contexte existant dans les deux images n'étant pas perçu par le modèle, il est impossible pour ce modèle de distinguer une apparition d'une « extension ».

3.4.4 Extensions du modèle

3.4.4.1 Prise en compte des décalages spatiaux faibles

Introduction Le modèle développé ici suppose un recalage parfait des deux images, qui, dans la pratique, n'existe presque jamais en imagerie satellite, sauf cas très particuliers (par exemples deux images prises avec très exactement le même angle, depuis un satellite geostationnaire). Comme nous l'avons fait remarquer précédemment, dans la section 3.2.1, les différentes corrections effectuées en prétraitement (la correction topographique, l'orthorectification, voir le second recalage spatial) ne parviennent jamais à recaler parfaitement les images et la majeure partie du temps, il persiste, dans les paires d'images, un décalage allant souvent jusqu'à quelques pixels.

Aussi, il peut être important de rendre le modèle par comparaison ponctuelle un minimum robuste aux décalages spatiaux faibles.

Adaptation du NFA Pour prendre en compte ces éventuelles incohérences spatiales, nous avons considéré que le décalage s'exprime sous la forme d'un champ de déformation $T : \mathcal{D} \to \mathcal{D}$ de sorte que

$$\forall p \in \mathcal{D}, \quad V_p \;=\; U_{p+T(p)} + \varepsilon_p. \tag{3.95}$$

Notons que cette modélisation est assez grossière et ne prend pas du tout en compte l'existence d'une composante sous-pixellique au décalage entre les deux images.

Par ailleurs, rappelons que l'on ne s'intéresse pas aux problèmes qu'il pourrait y avoir pour la définition du champ sur les bords du domaine $\mathcal{D}$ (voir section 3.2.1). On considère alors que ce champ varie suffisamment lentement à l'échelle du pixel pour que l'on puisse le considérer comme constant localement. De plus, grâce au recalage en prétraitement, on supposera que $T(p)$ est majoré pour tout pixel p par une constante $\tau_m \in \mathbb{N}$. On souhaiterait alors considérer la mesure

$$\mu^*(p,u,v) \;=\; \left\|\Sigma_f^{-1/2}(f(v)_p - f(u)_{p+T(p)})\right\|_2 \;=\; \left\|\Sigma_f^{-1/2} f(\varepsilon)_p\right\|, \tag{3.96}$$

qui permettrait de retrouver exactement la même expression du NFA. Seulement, sans informations sur la valeur du champ $T(p)$, nous sommes obligés de tester toutes les valeurs possibles, en considérant

$$\mu(p,u,v) \;=\; \min_{\|\tau\| \leq \tau_m} \left\|\Sigma_f^{-1/2}(f(v)_p - f(u)_{p+\tau})\right\|_2, \tag{3.97}$$

$\|\cdot\|$ étant une norme quelconque, dans la pratique égale à la norme infinie (pour simplifier l'implémentation des modèles). Toutefois, il est intéressant de remarquer que, comme $\mu(p,u,v) \leq \mu^*(p,u,v)$, on a

$$\mathbb{P}(\,\mu(p,u,v) \geq \delta\,) \;\leq\; \mathbb{P}(\,\mu^*(p,u,v) \geq \delta\,), \tag{3.98}$$

et ainsi, d'après la propriété 3.1.2 sur la majoration des nombres de fausses alarmes, il est possible de reprendre l'expression de NFA donnée dans l'équation (3.88), pour obtenir un nouveau nombre de fausses alarmes,

$$\mathrm{NFA}\,(p,u,v) \;=\; \frac{|\mathcal{D}|}{\Gamma\left(\frac{K}{2}\right)} \Gamma\left(\frac{K}{2}, \frac{1}{2\sigma^2} \min_{\|\tau\|_\infty \leq \tau_m} \left\|\Sigma_f^{-1/2}(f(v)_p - f(u)_{p+\tau})\right\|_2\right). \tag{3.99}$$

Toutefois, contrairement aux nombres de fausses alarmes définit précédemment, celui-ci n'est plus optimal à cause de la majoration effectuée.

Retour sur la majoration Cette majoration est en réalité moins grossière qu'il n'y paraît, car en général, le minimum sur τ est atteint en $T(p)$. Par ailleurs, cette majoration est d'autant plus grossière que le contexte local autour du pixel p considéré est homogène. Pour voir cela, il suffit de reprendre l'expression de la mesure de structure (nous supposerons $\Sigma_f = I_{K'}$ pour ne pas alourdir inutilement les notations),

$$\begin{aligned}
\mu(p,u,v) &= \min_{\|\tau\|_\infty \leq \tau_m} \|f(v)_p - f(u)_{p+\tau}\|_2 \\
&= \min_{\|\tau\|_\infty \leq \tau_m} \left\| f(u)_{p+T(p)} + f(\varepsilon)_{p+T(p)} - f(u)_{p+\tau} \right\|_2 \\
&= \min_{\|\tau\|_\infty \leq \tau_m} \left\| \Delta_{T(p),\tau} + f(\varepsilon)_{p+\tau} \right\|_2 ,
\end{aligned} \tag{3.100}$$

avec $\Delta_{T(p),\tau} = f(u)_{p+T(p)} - f(u)_{p+\tau}$ et donc $\Delta_{T(p),T(p)} = 0$. Ainsi, plus la zone sera « contrastée », dans le sens où les $\Delta_{T(p),\tau}$ sont grand, plus les termes $f(\varepsilon)_{p+\tau}$ seront négligeables devant $\Delta_{T(p),\tau}$, et plus la probabilité d'avoir $\mu(p,u,v) = \|f(\varepsilon)_{p+T(p)}\|_2$ sera importante. Ainsi dans notre cas, rajouter une robustesse aux translations dans la mesure de structure aura peu d'impact dans les zones contrastées, tandis qu'elle aura tendance à faire baisser le nombre de détections dans les zones homogènes. Cette baisse de la sensibilité dans les zones homogènes n'est pas réellement un problème pour plusieurs raisons. Tout d'abord, la sensibilité que l'on a dans la pratique reste convenable. Par ailleurs, les changements difficiles à détecter ne se produisent souvent pas dans les zones homogènes, mais au contraire dans les zones présentant beaucoup d'information et donc relativement contrastées.

Asymétrie du détecteur Il est important de noter que cette robustesse aux faibles translations apporte une asymétrie au modèle. En effet, en intervertissant les deux images, l'image sur laquelle est estimée la meilleure translation τ change, et on obtient alors un second modèle, légèrement différent du premier.

Remarquons que cette robustesse permet théoriquement de lever les limitations données dans la figure 3.12. En effet, en optimisant la translation, on permet alors au modèle d'évaluer le contexte autour du pixel étudié. Ainsi, dans la seconde paire de changements, présentée dans cette figure, en considérant τ_m suffisamment grand pour englober l'ensemble du disque, le modèle détectera le changement 1 sans problème, tandis que dans le cadre du deuxième changement, il sera possible d'associer le disque blanc (dans u) et le disque noir (dans v) avec l'un ou l'autre des deux demi-plans.

Toutefois, dans la pratique, on observe une dégradation des résultats et une forte hausse du temps de calcul, lorsque τ_m augmente. Aussi, nous nous limiterons à des valeurs de translation relativement faibles.

De plus, s'il peut être intéressant de disposer de deux modèles, apportant des informations distinctes, il est toutefois possible de fusionner ces deux modèles en considérant le double du minimum de leur valeur de significativité, comme le permet la propriété 3.1.7. Dans la pratique, nous n'utiliserons que la version symétrique de ce détecteur, pour ne pas trop augmenter le nombre de modèles disponible.

Résultat sur les exemples Nous avons appliqué le modèle par comparaison ponctuelle avec robustesse aux décalages spatiaux faibles sur les exemples synthétiques présentés dans la Figure 2.9 et affiché les résultats dans la Figure 3.13 en utilisant comme descripteur l'intensité en nuance de gris. Comme attendu, la prise en compte des décalages spatiaux faibles permet bien de supprimer les détections dues aux translations dans la paire 2. Toutefois cette robustesse, quand appliquée avec un seul pixel, entraîne aussi une disparition des changements au sein des textures (voir paire 4) ou au niveau des zones de gradient (voir paires 3 et 5). Pour limiter cela, on peut utiliser un descripteur de plus grande taille, en moyennant, par exemple, les intensités sur un voisinage du pixel, comme nous l'avons fait dans la figure 3.13.

Si cette extension du modèle semble surmonter la limitation quant à l'indistinguabilité des décalages et des apparitions/disparitions, il faut bien garder en tête que cette robustesse a un coût et multiplie le temps d'exécution par $(2\tau_m + 1)^2$ ce qui n'est plus négligeable lorsque τ_m devient grand et que l'on doit détecter des changements dans des macro-images de très grande taille. De ce fait, dans la pratique, nous utiliserons cette robustesse avec un paramètre τ_m relativement faible (que l'on prendra égal à 2 pixels) uniquement pour limiter les effets des incohérences spatiales.

3.4.4.2 Cas des descripteurs non linéaires

Introduction Comme nous l'avons dit précédemment, la restriction de ce modèle aux descripteurs linéaires constitue une limitation majeure de ce modèle, la majorité des descripteurs présentés dans la section 3.3 n'étant pas linéaires.

Comme précédemment, plaçons-nous dans le cadre de l'étude d'une paire d'images multispectrales, avec ainsi

$$u, v : \mathcal{D} \to \mathbb{R}^K,$$

avec $K \in \mathbb{N}^*$, et considérons un bruit pixellique sur les images, que nous supposerons blanc, pour simplifier le problème, avec ainsi, pour tout $p \in \mathcal{D}$,

$$V_p \;=\; U_p + \varepsilon_p \qquad \text{où} \qquad \varepsilon_p \sim \mathcal{N}\Big(0, \sigma^2 I_K\Big)\,. \tag{3.101}$$

Considérons, de plus, un descripteur, f, stationnaire et local (comme annoncé dans la section 3.3.1), que l'on représentera sous la forme présentée dans l'équation (3.30),

$$\forall p \in \mathcal{D}, \qquad f(U)_p \;=\; f_0(U_{|p+P_0}), \tag{3.102}$$

où P_0 est un patch centré en 0 et f_0 est une fonction de $\mathbb{R}^{K|P_0|}$ dans $\mathbb{R}^{K'}$, représentant localement f, avec $K' \in \mathbb{N}^*$, la dimension du descripteur f, et en notant U_{p+P_0} la restriction de U au domaine $p + P_0$, vu comme un vecteur de $\mathbb{R}^{K|P_0|}$.

Descripteur lipschitzien S'il n'est plus possible de trouver une expression exacte d'un nombre de fausses alarmes dans le cas d'un descripteur quelconque, on peut toutefois

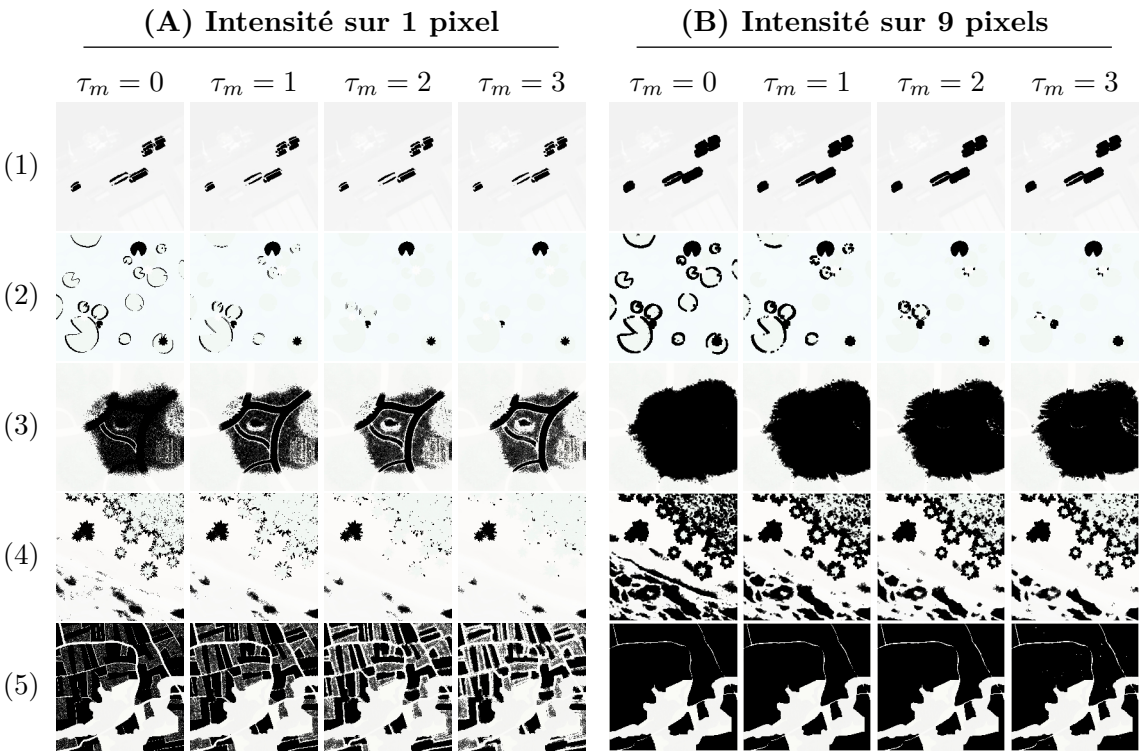


Figure 3.13 – Résultat du modèle par comparaison ponctuelle sur les exemples synthétiques présentés dans la Figure 2.9, en utilisant la robustesse légère aux décalages spatiaux et une version symétrisée du modèle. Dans l'expérience (A), nous avons ici utilisé l'intensité en nuance de gris standard. Dans l'expérience (B), nous l'avons moyennée sur un patch de taille 3×3. Dans chaque expérience, chaque sous-colonne représente une valeur de τ_m différente, et chaque ligne correspond à une paire d'images synthétiques. Dans chaque image de résultat, nous avons représenté en blanc les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$, le tout affiché sur une version assombrie de la seconde image. Comme attendu, dans les deux expériences, augmenter la valeur de τ_m permet bien ici de limiter les détections dues aux translations dans la paire 2. Il apparaît toutefois, dans l'expérience (A), qu'augmenter la valeur de τ_m entraîne aussi une « dégradation » des résultats, particulièrement visible dans les paires 3, 4 et 5, les changements n'étant plus détectés à proximité des bords ou au sein des textures. Si cette dégradation reste visible, dans l'expérience (B), notamment dans les paires 3 et 4, elle reste très limitée.

Proposition 3.4.7. *Pour tout descripteur f, local, stationnaire, si f est α-lipschitzien, pour $\alpha > 0$, alors la fonction définie pour tout $p \in \mathcal{D}$ par*

$$\mathrm{NFA}(p,u,v) \;=\; \frac{|\mathcal{D}|}{\Gamma\left(\frac{K|P_0|}{2}\right)}\Gamma\left(\frac{K|P_0|}{2}, \frac{\|f(u)_p - f(v)_p\|_2^2}{2\alpha^2\sigma^2}\right). \tag{3.103}$$

est un nombre de fausses alarmes pour le modèle naïf défini dans l'équation (3.101)*, pour P_0 le support de la représentation locale de f, tel que défini dans l'équation* (3.30)*.*

Démonstration. Comme précédemment, nous allons appliquer le théorème 3.1.2 en utilisant comme mesure de structure,

$$\mu(p,u,v) \;=\; \|f(u)_p - f(v)_p\|_2 \,. \tag{3.104}$$

Par définition, puisque f est α-lipschitzienne, on a

$$\|f(u)_p - f(v)_p\|_2 \;=\; \|f_0(u_{p+P_0}) - f_0(v_{p+P_0})\|_2 \;\leq\; \alpha\,\|(u-v)_{p+P_0}\|_2 \tag{3.105}$$

et ainsi

$$\mathbb{P}(\,\|f(U)_p - f(V)_p\|_2 \geq \delta\,) \;\leq\; \mathbb{P}(\,\alpha\,\|(V-U)_{p+P_0}\|_2 \geq \delta\,)\,. \tag{3.106}$$

Par hypothèse, La variable $(V-U)_{p+P_0}$ suit une distribution gaussienne centrée, de variance σ^2 et de dimension $K|P_0|$, on peut appliquer le théorème 3.4.2, qui donne

$$\mathbb{P}(\,\|f(U)_p - f(V)_p\|_2 \geq \delta\,) \;\leq\; \frac{\Gamma\left(\frac{K|P_0|}{2}, \frac{\delta^2}{2\alpha^2\sigma^2}\right)}{\Gamma\left(\frac{K|P_0|}{2}\right)}, \tag{3.107}$$

et on retrouve bien l'expression annoncée. □

Remarquons que le terme majorant, correspondant au nombre de fausses alarmes annoncé, est d'autant plus grand, et donc d'autant moins significatif, que la constante α et la taille du contexte $|P_0|$ sont grandes. De ce fait, pour obtenir une majoration la plus fine possible, il s'agira de réduire au maximum ces deux grandeurs. S'il n'est pas possible de réellement réduire la constante α sans changer de descripteurs, comme nous allons le voir, il est parfois possible de faire baisser fictivement la taille $|P_0|$.

Réduction de la dimension sous-jacente Prenons comme exemple le cas de la norme du gradient, estimé à l'aide des filtres de Sobel, comme explicité dans la section 3.3.3.1,

$$f(U)_p \;=\; f_0(U_{p+P_0}) \;=\; \|\nabla U_p\|_2 \;=\; \sqrt{(\nabla_x U)_p^2 + (\nabla_y U)_p^2},$$

avec $(\nabla_x U, \nabla_y U)$ le gradient de U, défini dans l'équation (3.39). On utilise ici théoriquement un support de taille 3×3 et, le centre n'étant pas considéré, on trouve donc $|P_0| = 8$ (avec ici $K = 1$). Toutefois, puisque $f(U)_p$ dépend directement de $\nabla_x U_p$ et $\nabla_y U_p$, apparaissant comme des variables aléatoires gaussiennes, on peut alors réduire fictivement $|P_0|$, la « dimension sous-jacente » de f, à 2.

Formellement, il s'agit d'écrire f comme la composition d'une application linéaire avec une fonction quelconque, c'est-à-dire d'écrire $f_0(x) = \tilde{f}_0(Rx)$, pour $x \in \mathbb{R}^{K|P_0|}$, $R \in \mathbb{R}^{r \times K|P_0|}$, $r \in \mathbb{N}^*$, et $\tilde{f}_0$ une fonction de $\mathbb{R}^r$ dans $\mathbb{R}^{K'}$. On a alors

$$\|f(U)_p - f(V)_p\|_2 \;=\; \|f_0(U_{p+P_0}) - f_0(V_{p+P_0})\|_2 \;\leq\; \tilde{\alpha}\, \|R\,(U-V)_{p+P_0}\|_2\,, \tag{3.108}$$

pour $\tilde{\alpha}$ la constante de lipschitz de $\tilde{f}_0$, et ainsi, pour tout $\delta > 0$,

$$\mathbb{P}(\,\|f(U)_p - f(V)_p\|_2 \geq \delta\,) \;\leq\; \mathbb{P}\Big(\,\|R\,(U-V)_{p+P_0}\|_2 \geq \tilde{\alpha}^{-1}\delta\,\Big)\,. \tag{3.109}$$

Puisque $(U-V)_{p+P_0}$ suit une distribution gaussienne de matrice de covariance $\sigma^2 I_{K|P_0|}$, $R\,(U-V)_{p+P_0}$ suit une distribution gaussienne de matrice de covariance $\sigma^2 R^\mathsf{T} R$.

Dans le cas où les colonnes de R sont orthonormées (ce qui est souvent le cas), la matrice de covariance s'écrit donc simplement $\sigma^2 I_{Kr}$ et on peut alors directement appliquer le théorème 3.4.2 pour en déduire la majoration

$$\mathbb{P}(\,\|f(U)_p - f(V)_p\|_2 \geq \delta\,) \;\leq\; \frac{\Gamma\Big(\frac{Kr}{2}, \frac{\delta^2}{2\tilde{\alpha}^2\sigma^2}\Big)}{\Gamma\Big(\frac{Kr}{2}\Big)}, \tag{3.110}$$

pour $\delta > 0$, et ainsi en déduire un nouveau nombre de fausses alarmes

$$\mathrm{NFA}(p,u,v) \;=\; \frac{|\mathcal{D}|}{\Gamma\Big(\frac{Kr}{2}\Big)}\Gamma\left(\frac{Kr}{2}, \frac{\|f(u)_p - f(v)_p\|_2^2}{2\tilde{\alpha}^2\sigma^2}\right). \tag{3.111}$$

Dans le cas où les colonnes de R ne sont pas orthonormées, il est possible de trouver une majoration de la probabilité $\mathbb{P}(\,\|R\,(U-V)_{p+P_0}\|_2 \geq \alpha^{-1}\delta\,)$ à l'aide du lemme suivant.

Lemme 3.4.8. *Soit $X \in \mathbb{R}^n$, une variable gaussienne, centrée et de matrice de covariance Σ, définie positive. Pour tout $\delta > 0$,*

$$\mathbb{P}(\,\|X\|_2 \geq \delta\,) \;\leq\; \frac{1}{\Gamma(\frac{n}{2})}\Gamma\left(\frac{n}{2}, \frac{\delta^2}{2{\lambda_\Sigma}^2}\right), \tag{3.112}$$

où λ_Σ est la plus grande valeur propre de Σ.

Démonstration. Pour montrer cela, considérons $\Sigma^{-1/2}$ la racine carrée de Σ^{-1}, c'est-à-dire l'unique la matrice symétrique définie positive vérifiant $(\Sigma^{-1/2})^2 = \Sigma^{-1}$. Ainsi, en posant $Y = \Sigma^{-1/2}X$, Y suit une distribution gaussienne de matrice de covariance $\Sigma^{-1/2}\Sigma\Sigma^{-1/2} = I_n$ et, en écrivant $X = \sigma^{1/2}Y$, on trouve

$$\mathbb{P}(\,\|X\|_2 \geq \delta\,) \;=\; \mathbb{P}\Big(\,Y^\mathsf{T}\Sigma Y \geq \delta\,\Big)\,.$$

En considérant λ_Σ la plus grande valeur propre de Σ, on a $Y^\mathsf{T}\Sigma Y \leq \lambda_\Sigma\,\|Y\|_2$ et ainsi, on obtient bien

$$\mathbb{P}(\,\|X\|_2 \geq \delta\,) \;\leq\; \mathbb{P}(\,\lambda_\Sigma\,\|Y\| \geq \delta\,) \;=\; \frac{1}{\Gamma(\frac{n}{2})}\Gamma\left(\frac{n}{2}, \frac{\delta^2}{2{\lambda_\Sigma}^2}\right),$$

en appliquant le théorème 3.4.2. □

On obtient ainsi le nombre de fausses alarmes

$$\mathrm{NFA}(p,u,v) \;=\; \frac{|\mathcal{D}|}{\Gamma\left(\frac{Kr}{2}\right)}\Gamma\left(\frac{Kr}{2}, \frac{\|f(u)_p - f(v)_p\|_2^2}{2\tilde{\alpha}^2\sigma^2{\lambda_{R^\intercal R}}^2}\right). \tag{3.113}$$

Dans les deux cas, on retrouve une expression similaire à celle donnée dans la proposition 3.4.7, mais utilisant d'autres les paramètres Kr et $\tilde{\alpha}\lambda_{R^\intercal R}$ au lieu de $K|P_0|$ et α.

3.5 Modèle de changements par comparaison d'histogramme

3.5.1 Introduction

Philosophie du modèle Le modèle précédent, par comparaison ponctuelle, possédait l'avantage de détecter à peu près tous les changements, mais montrait des limites en présence d'incohérences spatiales. Cela s'explique par la prise en compte trop stricte de la spatialisation des pixels, en exigeant de retrouver la même information à la même place. Pour construire le modèle suivant, nous allons remettre en cause la nécessité de cette spatialisation. Pour cela, plutôt que de comparer les valeurs des pixels, nous allons comparer les distributions locales des deux images. De cette manière, la position exacte des pixels n'est pas prise en compte. Nous avons considéré

La comparaison de distributions locales Plusieurs auteurs de détection de changements proposent de comparer les distributions locales des images. Il s'agit alors de considérer un patch commun aux deux images et d'appliquer une mesure de similarité entre les valeurs des deux images sur ce patch, puis de déduire les changements en seuillant cette mesure de similarité.

Ces méthodes se différencient sur la mesure utilisée. Hong and Vatsavai [HV16] se ramène à la mesure de Kullback-Leibler appliquée à des modélisations gaussiennes des distributions locales. Inglada et Mercier [IM07] proposent deux approches, tout d'abord d'utiliser cette même distance en modélisant les distributions par des fonction de Pearson, ou d'utiliser les premiers cumulants pour approcher la distance de Kullback-Leibler sans devoir effectuer de modélisation. Mercier et coauteurs [MDP+06] décident d'utiliser l'information mutuelle, qu'ils approchent numériquement à l'aide de copules, généralisant la notion de fonction de répartition pour une dimension supérieure à 1.

Le problème principal de ces approches réside dans le choix du seuil à appliquer après avoir appliquée la mesure de similarité. Par ailleurs, la modélisation utilisée dans certaines approches rajoute un apriori sur les distributions observées. Pour limiter au maximum ces aprioris, nous considérons ici directement les distributions empiriques. Par ailleurs, les images étant souvent comparées sur des patchs de taille réduite (comprenant une centaine de pixels), nous nous sommes intéressés à des distribution en dimension 1, qui se ramène ainsi à l'étude de leur fonction de répartition. Une telle approche a déjà été proposée dans la littérature, par Eghbali [Egh79] et reprise par Tang et coauteurs [TZH11], et consiste à comparer les deux fonctions de répartition au moyen

de la distance de Kolmogorov-Smirnov, puis d'utiliser le théorème de Kolmogorov-Smirnov pour en déduire une valeur du seuil permettant de limiter la probabilité de fausse alarme. Notons que le théorème utilisé ne donne qu'une propriété asymptotique, lorsque la taille du patch utilisé tend vers l'infini.

Nous reprenons ici tout d'abord l'expression obtenue avec le théorème de Kolmogorov-Smirnov, puis donnons une nouvelle expression traduisant exactement la probabilité de fausses alarmes et qui ne nécessite aucune supposition sur la taille des patchs utilisé pour estimer les distributions empiriques.

3.5.2 Formalisation a contrario

Modélisation des images et modèle naïf Comme pour le modèle précédent, on considère ici deux images u et v définies sur un même domaine $\mathcal{D} \subset \mathbb{Z}^2$. Cependant, contrairement au modèle précédent, on suppose ici que les deux images ne possèdent qu'un seul canal, avec ainsi

$$u, v \ : \mathcal{D} \to \mathbb{R}, \tag{3.114}$$

sans toutefois poser de condition sur le modèle de bruit, le canal utilisé pouvant ainsi éventuellement correspondre à un des descripteurs présentés dans la section 3.3.

Pour définir le modèle naïf, on suppose que, s'il n'y a aucun changement entre les deux images, ces dernières possèdent une même distribution locale, en supposant de plus les pixels indépendants. Formellement, il s'agit de considérer, pour un rayon donné, $d \in \mathbb{N}^*$, et pour tout pixel du domaine, $p \in \mathcal{D}$, que tous les pixels situés dans la boule centrée en p et de rayon d suivent indépendamment une même distribution, $\mathcal{F}_p$, c'est à dire

$$\forall p \in \mathcal{D},\ \forall q \in \mathcal{B}(p,d), \qquad U_q, V_q \sim \mathcal{F}_p, \tag{3.115}$$

où $\mathcal{B}(p,d)$ est la boule centrée en p et de rayon d, pour une norme quelconque (la norme infinie dans la pratique),

$$\mathcal{B}(p,d) \ = \ \{q \in \mathcal{D},\ \|p-q\| \leq d\}\,. \tag{3.116}$$

Notons que, comme précédemment, on ne s'intéresse pas ici aux problèmes que pourrait poser cette définition sur les bords du domaine $\mathcal{D}$ (voir section 3.2.1).

3.5.2.1 Test de Kolmogorov-Smirnov

Pour comparer les deux distributions, il est possible d'utiliser le test de Kolmogorov-Smirnov.

Théorème 3.5.1 (Test de Kolmogorov-Smirnov [Kol33])**.** *Soit $(X_i)_{i\in\mathbb{N}}$ un ensemble de variables aléatoires réelles, indépendantes et identiquement distribuées selon une distribution $\mathcal{F}$, de fonction de répartition F. En considérant alors F_n, la fonction de répartition empirique estimée sur les $n \in \mathbb{N}$ premières variables, c'est-à-dire, pour tout $t \in \mathbb{R}$,*

$$F_n(t) \ = \ \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq t\}}, \tag{3.117}$$

alors la suite des fonctions de répartition empirique F_n converge uniformément, presque sûrement, vers la fonction de répartition réelle F et de plus, pour tout réel $c > 0$,

$$\mathbb{P}\Big(\|F_n - F\|_\infty \geq \frac{c}{\sqrt{n}} \Big) \underset{n\to\infty}{\longrightarrow} 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2c^2k^2}. \tag{3.118}$$

Le test de Kolmogorov-Smirnov permet de quantifier à quel point un ensemble d'échantillons suit une distribution donnée, étant donnée un nombre d'échantillons suffisamment grand. Et il est alors possible de construire un nombre de fausses alarmes à partir de ce théorème.

Kolmogorov-Smirnof et nombre de fausses alarmes On considère les distributions empiriques au voisinage de $p \in \mathcal{D}$, associées à chaque image, que nous noterons $F_U(p)$ et $F_V(p)$ respectivement pour U et V, c'est-à-dire

$$\forall t \in \mathbb{R}, \qquad F_U(p) \;=\; \frac{1}{n}\sum_{q\in\mathcal{B}(p,d)} \mathbb{1}_{\{U_q \leq t\}}, \tag{3.119}$$

en notant $n = |\mathcal{B}(p,d)|$ le nombre de pixels dans le voisinage de p, qui correspond au nombre d'échantillon que l'on possède pour estimer la fonction de répartition. On définit alors la mesure de structure entre les deux images, au pixel p, par

$$\mu(p,u,v) \;=\; \|F_v(p) - F_u(p)\|_\infty\,. \tag{3.120}$$

On peut alors montrer la proposition suivante.

Proposition 3.5.2. *En considérant la fonction définie pour tout $p \in \mathcal{D}$ par*

$$\mathrm{NFA}(p,u,v) \;=\; 4\,|\mathcal{D}|\,\sum_{k=1}^{\infty}(-1)^{k-1}e^{-\frac{1}{2}n\|F_v(p)-F_u(p)\|_\infty^2 k^2}, \tag{3.121}$$

où $F_u(p)$ et $F_v(p)$ sont les fonctions de répartition empirique estimées sur une voisinage de p (voir équation 3.118) et n est la taille de ce voisinage, alors cette fonction approche un nombre de fausses alarmes pour le modèle naïf explicité dans l'équation (3.115) *lorsque $|\mathcal{D}|$ est suffisamment grand.*

En effet, d'après le théorème 3.1.2, la fonction

$$(p,u,v) \;\mapsto\; |\mathcal{D}|\,\mathbb{P}\big(\,\mu(p,U,V) \geq \mu(p,u,v)\,\big) \tag{3.122}$$

est un nombre de fausses alarmes, et il suffit alors de montrer que cette fonction minore la fonction donnée dans la propriété pour pouvoir conclure à l'aide de la propriété 3.1.5.

Lemme 3.5.3. *Soit $n \in \mathbb{N}$ et $X_1,\ldots, X_n$ et $Y_1, \ldots, Y_n$, des variables aléatoires, indépendantes et identiquement distribuées suivant une même distribution de fonction de répartition F. En notant F_n et G_n les fonctions de répartition empiriques respectives des X_i et Y_i, alors pour tout $\delta \geq 0$,*

$$\mathbb{P}\big(\,\|F_n - G_n\|_\infty \geq \delta\,\big) \;\leq\; 2\,\mathbb{P}\Big(\,\|F_n - F\|_\infty \geq \frac{\delta}{2}\Big)\,. \tag{3.123}$$

Démonstration. Pour montrer cela, il suffit d'appliquer l'inégalité triangulaire,

$$\|F_n - G_n\|_\infty \;\leq\; \|F_n - F\|_\infty + \|F - G_n\|_\infty\,,$$

ce qui nous donne bien, pour $\delta \geq 0$,

$$\begin{aligned}
\mathbb{P}(\,\|F_n - G_n\|_\infty \geq \delta\,) &\leq \mathbb{P}(\,\|F_n - F\|_\infty + \|F - G_n\|_\infty \geq \delta\,) \\
&\leq \mathbb{P}\Big(\,\big(\|F_n - F\|_\infty \geq \tfrac{1}{2}\delta\big) \cup \big(\|F - G_n\|_\infty \geq \tfrac{1}{2}\delta\big)\,\Big) \\
&\leq \mathbb{P}\Big(\,\|F_n - F\|_\infty \geq \tfrac{1}{2}\delta\,\Big) + \mathbb{P}\Big(\,\|F - G_n\|_\infty \geq \tfrac{1}{2}\delta\,\Big) \\
&\leq 2\,\mathbb{P}\Big(\,\|F_n - F\|_\infty \geq \tfrac{1}{2}\delta\,\Big)\,,
\end{aligned}$$

la deuxième égalité étant due au fait que, pour tout réels a et b, $a + b \geq \delta$ implique nécessairement $a \geq \frac{1}{2}\delta$ ou $b \geq \frac{1}{2}\delta$. □

En notant $\delta = \mu(p, u, v)$ et F la fonction de répartition de la distribution réelle, donnée équation (3.115), le lemme précédent nous donne

$$\begin{aligned}
\mathbb{P}(\,\mu(p, U, V) \geq \delta\,) &= \mathbb{P}(\,\|F_V(p) - F_U(p)\|_\infty \geq \delta\,) \\
&\leq 2\,\mathbb{P}\Big(\,\|F_V(p) - F(p)\|_\infty \geq \tfrac{\delta\sqrt{n}}{2\sqrt{n}}\,\Big)\,.
\end{aligned} \tag{3.124}$$

On peut alors appliquer le test de Kolmogorov-Smirnof (théorème 3.5.1) pour conclure que, si $\mathcal{D}$ est suffisamment grand,

$$2\,\mathbb{P}\left(\,\|F_V(p) - F(p)\|_\infty \geq \frac{\delta\sqrt{n}}{2\sqrt{n}}\right) \;\simeq\; 4\sum_{k=1}^{\infty}(-1)^{k-1}e^{-\frac{1}{2}n\delta^2k^2}\,, \tag{3.125}$$

et on a donc bien le résultat annoncé.

Prise en compte de l'imprécision numérique Comme nous l'avons déjà précisé, les valeurs des nombres de fausses alarmes peuvent dans la pratique avoir des valeurs très faibles, parfois inférieures à 10^{-100}, ce qui peut alors dépasser la précision machine. Pour cela, nous calculerons plutôt le logarithme de cette expression.

$$-\log \mathrm{NFA}(p, u, v) \;=\; -\log 2 - \log|\mathcal{D}| - \log\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2n\|F_v(p)-F_u(p)\|_\infty^2 k^2}. \tag{3.126}$$

Intéressons-nous tout d'abord à l'estimation de la somme de la série. En posant alors $\alpha = 2n\,\|F_v(p) - F_u(p)\|_\infty^2$ et en utilisant les propriétés sur les séries alternées, on peut alors majorer l'erreur commise en arrêtant la somme au rang $K > 0$,

$$\left|\sum_{k=1}^{K}(-1)^{k-1}e^{-\alpha k^2} - \sum_{k=1}^{\infty}(-1)^{k-1}e^{-\alpha k^2}\right| \;\leq\; e^{-\alpha(K+1)^2}.$$

Ainsi, si l'on souhaite avoir une précision de l'ordre de $e^{-\varepsilon}$, pour $\varepsilon > 0$, il suffit alors que $\alpha(K+1)^2$ soit supérieur à ε et on sommera alors jusqu'au rang K^*,

$$K^* \;=\; \left\lceil\sqrt{\frac{\varepsilon}{\alpha}} - 1\right\rceil \;=\; \left\lceil\sqrt{\frac{\varepsilon}{2n\,\|F_v(p) - F_u(p)\|_\infty^2}} - 1\right\rceil\,, \tag{3.127}$$

pour $\lceil x \rceil$ la partie entière supérieure de $x \in \mathbb{R}$. Par ailleurs, pour éviter des problèmes de précision machine, on prendra soin d'écrire

$$\log \sum_{k=1}^{K^*} (-1)^{k-1} e^{-\alpha k^2} \;=\; -\alpha + \log \left(1 + \sum_{k=2}^{K^*} (-1)^{k-1} e^{-\alpha(k^2-1)} \right),$$

qui permettra d'éviter que des arrondis à 0 donnent des valeurs infinies, pour des valeurs de α trop grandes.

3.5.2.2 Expression exacte

Nombre de fausses alarmes exact Il est cependant possible de trouver une expression exacte de la probabilité donnée dans l'équation 3.122 à l'aide de la proposition suivante.

Proposition 3.5.4 ([PG81]). *Soit $n \in \mathbb{N}$ et $X_1, \ldots, X_n$ et $Y_1, \ldots, Y_n$, des variables aléatoires, indépendantes et identiquement distribuées suivant une même distribution. En notant F_n et G_n les fonctions de répartition empiriques respectives des X_i et Y_i, alors pour tout $k \in \mathbb{N}^*$,*

$$\mathbb{P}\Big(\|F_n - G_n\|_\infty \geq \frac{k}{n} \Big) \;=\; 2 \sum_{i=1}^{[n/k]} \frac{(-1)^{i+1} n!^2}{(n+ik)!(n-ik)!}, \tag{3.128}$$

où $[n/k]$ est la partie entière inférieure de n/k.

Cette proposition permet d'obtenir directement la proposition suivante

Proposition 3.5.5. *Soit $n \in \mathbb{N}$ et N_p un voisinage de p contenant n pixel, pour tout $p \in \mathcal{D}$. En considérant comme modèle naïf que pour tout p, tous les U_q et V_q pour $q \in N_p$ suivent une même distribution, alors la fonction suivante est un nombre de fausses alarmes,*

$$\mathrm{NFA}(p,u,v) \;=\; 2|\mathcal{D}| \sum_{i=1}^{[n/k(u,v)]} \frac{(-1)^{i+1} n!^2}{(n+i\,k(u,v))!(n-i\,k(u,v))!}, \tag{3.129}$$

où $k(u,v) = [n\,\|F_v(p) - F_u(p)\|_\infty]$, $F_u(p)$ et $F_v(p)$ sont les fonctions de répartition empirique de u et v estimées sur N_p (voir équation 3.118) et $p \in \mathcal{D}$.

Démonstration. Pour montrer cela, il suffit d'appliquer le théorème 3.1.2, relatif à la construction des nombres de fausses alarmes, à la mesure $\mu(p,u,v) = \|F_v(p) - F_u(p)\|_\infty$, qui nous assure que la fonction

$$F(p,u,v) \;=\; |\mathcal{D}|\,\mathbb{P}(\,\mu(p,U,V) \geq \mu(p,u,v)\,) \;=\; |\mathcal{D}|\,\mathbb{P}\Big(\|F_V(p) - F_U(p)\|_\infty \geq \frac{k(u,v)}{n} \Big)$$

est bien un nombre de fausses alarmes. Il suffit alors d'appliquer la proposition 3.5.4 pour conclure. □

Par la suite, nous utilisons ce nombre de fausses alarmes, qui présente l'avantage de ne faire aucune hypothèse quant à la taille du voisinage sur lequel sont calculées les distributions empiriques.

Calcul numérique Notons que, tout comme l'expression précédente, il est nécessaire de faire attention aux problèmes de représentation des nombres réels sur un nombre fini de bits. En posant

$$u_i = \frac{n!^2}{(n+ik)!(n-ik)!} = \prod_{j=1}^{ik} \frac{n+1-j}{n+j} = \frac{\Gamma(n+1)^2}{\Gamma(n+ik+1)\Gamma(n-ik+1)}, \quad (3.130)$$

on fera alors attention à calculer le logarithme de ce nombre de la manière suivante,

$$\log \mathrm{NFA}(p,u,v) = \log|\mathcal{D}| + \log 2u_1 + \log\left(1 + \sum_{i=2}^{[n/k]} (-1)^{i+1} \prod_{j=k+1}^{ik} \frac{n+1-j}{n+j}\right), \quad (3.131)$$

avec

$$\log u_1 = 2\log\Gamma(n+1) - \log\Gamma(n-k+1) - \log\Gamma(n+k+1). \quad (3.132)$$

3.5.3 Résultats et limitation du modèle

3.5.3.1 Résultats sur les exemples synthétiques

Comment mentionné précédemment, nous utilisons ici le détecteur issus du nombre de fausses alarmes données dans la proposition 3.5.5.

De la même manière que pour le modèle précédent, nous avons testé ce modèle sur les exemples synthétiques présentés dans la figure 2.9, en ne considérant que quelques descripteurs d'images. Ce modèle n'acceptant que les descripteurs de dimension 1, nous avons choisi de nous restreindre à la nuance de gris et à la norme du gradient, que nous avons jugés comme étant les descripteurs les plus simples et permettant de ne pas confondre l'apport du modèle avec ceux des descripteurs. Nous avons affiché les résultats obtenus dans la Figure 3.14.

Remarques générales dans le cas de la nuance de gris Comparons alors les résultats obtenus à ceux du modèle par comparaison ponctuelle, et présentés dans la figure 3.11, dans un premier temps lorsque l'on utilise la nuance de gris comme descripteur. Il est tout d'abord intéressant de remarquer que l'on obtient des résultats similaires pour les paires 1, 3 et 5. On détecte en effet bien les différentes apparitions d'objets, le changement de contraste de la paire 3 ainsi que le changement global de la paire 5. De plus, comme attendu, les décalages légers de la paire 2 ne sont pas détectés ici. Remarquons cependant une petite détection dans le cas $n = 11 \times 11$ dans la partie inférieure gauche de cette même paire. Cette détection peut s'expliquer par l'écartement de deux objets à proximité. Dans la première image, ils sont suffisamment proches pour être tous deux dans un même contexte d'estimation de la distribution empirique. En s'éloignant, les deux objets se retrouvent trop éloignées, et, en ne pouvant voir qu'un seul des deux objets à la fois, la distribution empirique s'en trouve modifiée. C'est la raison pour laquelle cette détection n'apparaît pas pour une taille de contexte plus petite, les deux objets n'étant alors jamais présents dans un même contexte.

Intéressons-nous alors plus précisément à la paire 4. Cette paire présente (en dehors du bruit) quatre changements principaux, à savoir l'apparition d'un objet en haut à

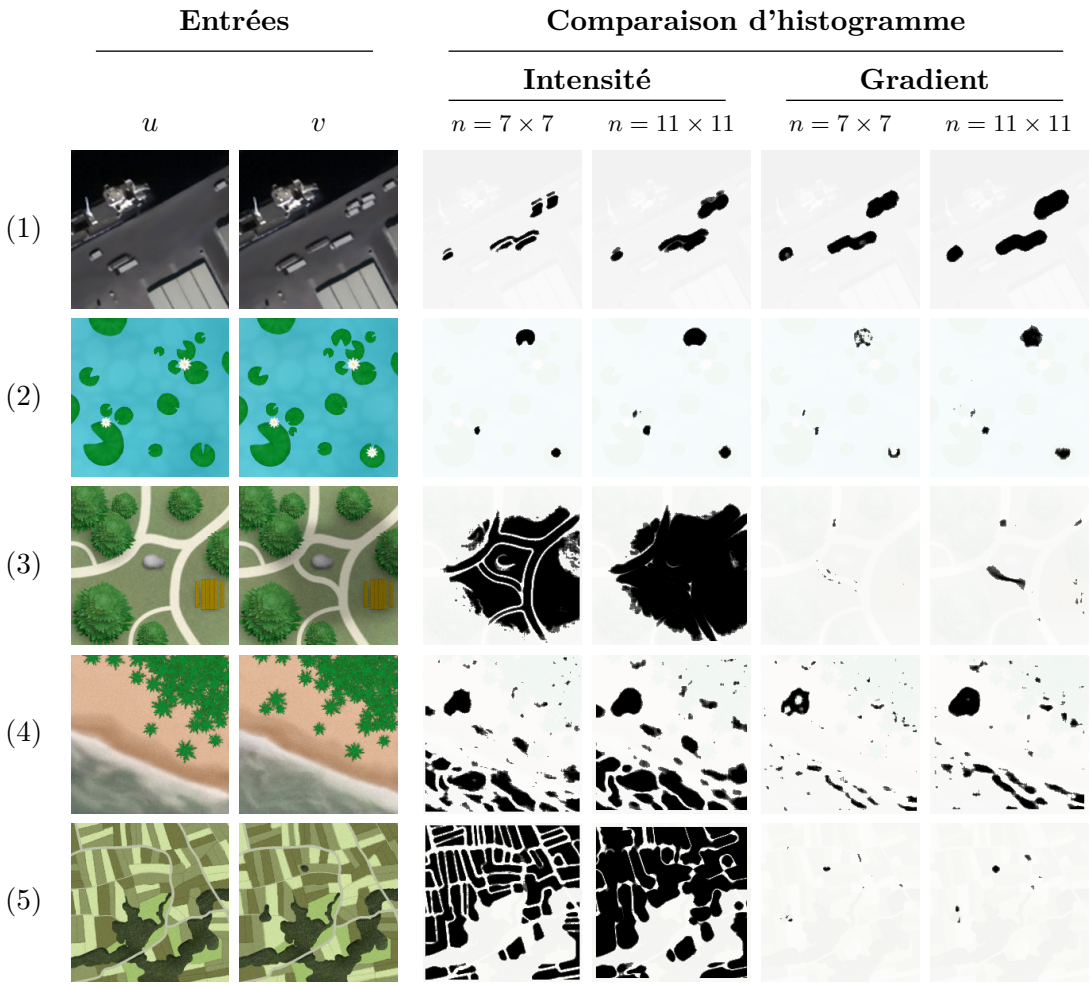


Figure 3.14 – Résultats du modèle par comparaison d'histogramme sur les exemples synthétiques présentés dans la Figure 2.9. Deux types de descripteurs ont été utilisés, l'intensité en nuance de gris et la norme du gradient, et pour chacun d'entre eux, deux tailles de contexte ont été utilisées pour estimer la distribution empirique, à savoir 7×7 et 11×11. Chaque ligne représente un exemple et est constituée successivement des deux images u et v suivies des résultats du modèle avec chacun des descripteurs. Dans les images de résultat, nous avons représenté en blanc les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$, le tout affiché sur une version assombrie de la seconde image.

gauche, une variation de micro-texture dans les arbres, une variation de macro-texture dans la mer ainsi qu'une variation légère de contraste le long du rivage. On détecte ici l'apparition d'objet, la variation de macro-texture ainsi que la légère variation de contraste, tandis que la variation de micro-texture n'est détectée qu'à certains endroits. Si la détection de la macro-texture s'explique facilement par la modification qu'elle apporte à la distribution locale, celles présentes au niveau de la micro-texture semblent plus arbitraires. Les détections présentes aux bords de la texture correspondent en fait à des zones où la limite de texture s'est déplacée, modifiant ainsi la distribution locale. Pour les détections, peu nombreuses, au sein de la texture, il s'agit très certainement de zones présentant des variations particulièrement importantes de la texture. Nous rediscuterons plus tard de la détection de la variation légère de contraste dans les limitations.

Utilisation du gradient En ce qui concerne l'utilisation du gradient, notons tout d'abord que ce dernier apporte toujours une robustesse aux changements de contraste et de ce fait, les changements qui s'y rapportent dans les paires 3 et 4 ne sont plus détectés. Après, il est intéressant de regarder de plus près la paire 5. L'utilisation du gradient semble en effet permettre de ne plus détecter le changement global. La non-detection est en fait dû à la taille des changements qui ont lieu ici. En effet, en regardant le gradient, on ne peut voir que les bords des champs, ne représentant qu'un faible pourcentage des pixels, et il semble alors que ce pourcentage ne soit pas ici suffisant pour entraîner une détection.

3.5.3.2 Limitations du modèle

Dimension limitée Une des premières limitations que possède ce modèle est qu'on ne peut utiliser que des descripteurs de dimension 1. En se basant sur les fonctions de répartition, qui ne peuvent être définies que pour des variables aléatoires de dimension 1, on contraint ce modèle à n'être utilisé que pour des images dans $\mathbb{R}$. On peut imaginer contourner cette limitation en effectuant une réduction de dimension avant d'appliquer le modèle, par exemple à l'aide d'une décomposition en composantes principales, cependant, cela reste une limitation forte de ce modèle.

Sensibilité aux changements de contraste Une seconde limitation importante de ce modèle est sa sensibilité aux changements de contraste, et ce dès lors qu'ils dépassent le niveau de bruit. Effet, en s'intéressant aux fonctions de répartition, la valeur des différents pixels dans les images n'a en fait aucune importance et l'ordre relative de ces valeurs, les une par rapport aux autres, est prise en compte.

Pour mieux voir cela, considérons $N > 0$ échantillons $X = \{X_i\}_{i\in[\![1,N]\!]} \in \mathbb{R}^N$ et notons F_X la fonction de répartition empirique de X. Pour toute fonction réelle strictement croissante (et donc inversible) h et tout $t \in \mathbb{R}$, on a

$$F_{h(X)}(t) \;=\; \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{h(X_i)\leq t\}} \;=\; \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i\leq h^{-1}(t)\}} \;=\; F_X(h^{-1}(t)). \tag{3.133}$$

Il apparaît alors que, étant donné d'autres échantillons Y,

$$\left\| F_{h(X)} - F_{h(Y)} \right\|_\infty \;=\; \left\| (F_X - F_Y) \circ f^{-1} \right\|_\infty \;=\; \| F_X - F_Y \|_\infty \,,$$

et ainsi, seul l'ordre relatif des X_i et Y_i a une importance dans cette distance. De ce fait, dans des zones homogènes, le modèle par comparaison d'histogramme ne pourra pas distinguer une variation de contraste faible d'une variation de contraste importante, et ce, dès lors que la variation d'intensité sera supérieure au bruit dans l'image. Ce qui explique que, dans la paire 4 de la figure 3.14, un léger changement de contraste a suffi à provoquer une détection par ce modèle.

Résolution spatial amoindrie D'autre part, il est important de noter que, en perdant l'information de localisation à l'intérieur du contexte, on perd aussi, d'une certaine manière, cette même spatialisation au niveau du résultat du détecteur. On peut observer cela dans la Figure 3.14, en comparant les résultats pour les deux tailles de contexte donnée : en passant d'un contexte de largeur 7 à 11, on peut observer une augmentation de la taille des détections.

Dans la pratique, cette baisse de la résolution spatiale du résultat n'est pas réellement importante. On rappelle en effet que l'on cherche ici à détecter des changements et non à les segmenter. Malgré tout, une mauvaise résolution spatiale du résultat pourrait éventuellement gêner, lors de la fusion des détecteurs, que nous présenterons dans la section 5.2.

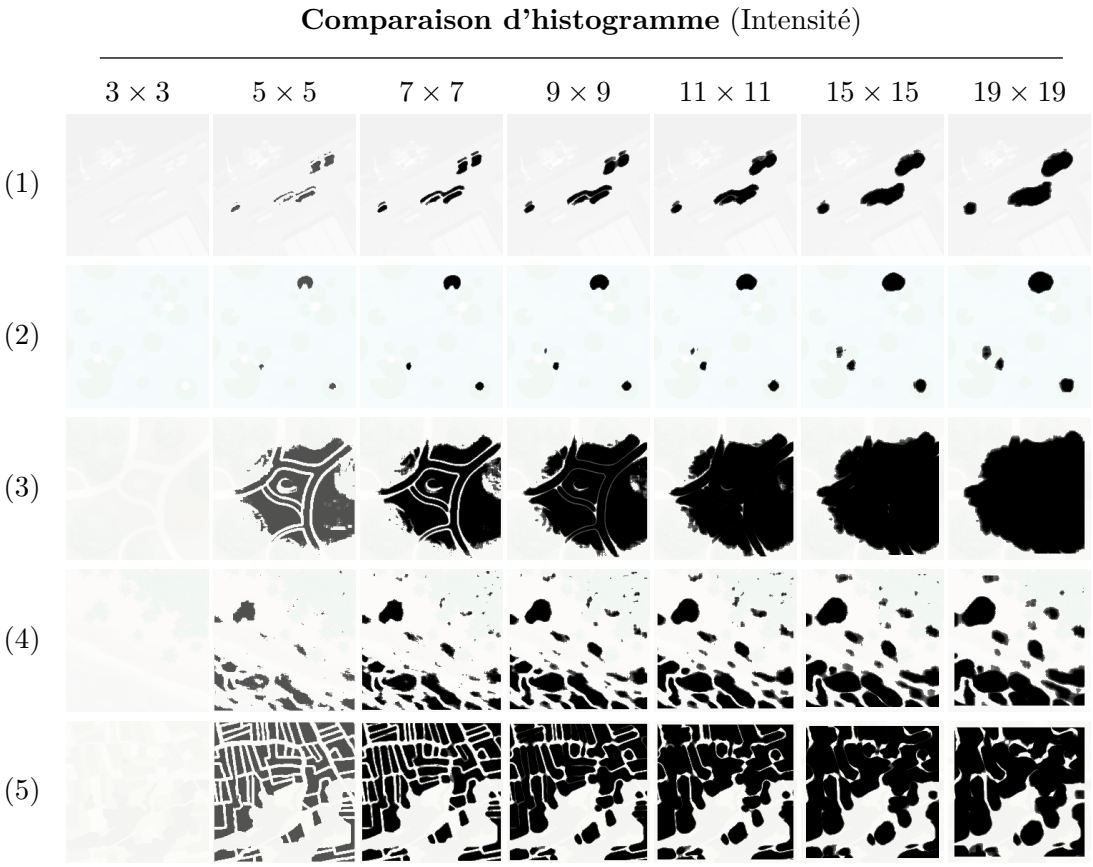


Figure 3.15 – Résultats du modèle par comparaison d'histogramme sur les exemples synthétiques présentés dans la Figure 2.9, en fonction de la taille du contexte utilisé. Chaque ligne représente un exemple et est composé des résultats obtenus pour les différentes tailles de contexte, allant de 3×3 à 19×19. Dans chaque image, nous avons représenté en blanc les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$, le tout affiché sur une version assombrie de la seconde image. On peut remarquer ici qu'en augmentant la taille du contexte utilisé, le masque de détection perd en finesse. Par ailleurs, il est intéressant de remarquer que, pour un contexte trop petit (ici de taille 3×3), le modèle ne renvoie aucune détection.

Modèles pour la détection indirecte de changements

4

Table des matières

4.1 Modélisation locale du contexte 122
4.2 Modèle de changements par destin commun 135
4.3 Modèle de changements par reconstruction indirecte 148

Contenu Dans ce chapitre, pour présentons trois modèles différents pour la détection de changements, s'écartant de la comparaison habituelle des deux images, en proposant pour chacun une majoration duq nombre de fausses alarmes associé. Pour le premier modèle, nous proposons une extension de l'algorithme de Reed-Xiaoli pour des modèles de mélange de gaussiennes, en évaluant la cohérence des distributions locales entre les deux images. Pour le second modèle, nous présentons une méthode robuste aux changements globaux, évaluant la cohérence de statistiques locales jointes vis-à-vis des statistiques globales. Enfin, pour le troisième modèle, nous nous inspirons de l'algorithme de détection d'anomalies de Davy et coauteurs [DEMD18], et utilisons un processus de reconstruction croisée pour pouvoir comparer les structures de voisinages des deux images, en intégrant une certaine robustesse aux incohérences spatiales.

4.1 Modélisation locale du contexte

4.1.1 Introduction

Comparaison ponctuelle et confrontation statistique Comme nous l'avons mentionné précédemment, le problème du modèle par comparaison directe est sa prise en compte trop stricte de la localisation des pixels, causant de nombreuses fausses détections en présence d'incohérences spatiales. Le modèle par histogramme, en considérant une mesure de distance entre les distributions locales plutôt qu'entre les intensités, nous a permis de comparer localement les images tout en restant relativement robuste aux décalages spatiaux, mais au prix d'une moindre résolution du masque de détection, provoquant entre autres la disparition de changements localisés dans des zones trop peu étendues.

Dans ces deux modèles, il s'agit en fait de comparer directement les images, en utilisant des représentations locales, à l'échelle du pixel, ou à celle d'un voisinage. Dans les modèles que nous présentons ici, nous cherchons plutôt à confronter les deux images. Il ne s'agira donc plus de savoir si, étant donné une certaine représentation, les images sont proches, mais de savoir si ces dernières sont compatibles entre elles, étant donné une certaine modélisation.

Philosophie du modèle Dans ce premier modèle de changements par confrontations statistiques, nous reprenons le postula précédent : sans changement, les deux images partageraient une même distribution locale. Seulement, plutôt que d'effectuer une simple comparaison, nous évaluons ici la vraisemblance des pixels d'une des images vis-à-vis de la distribution de la seconde. Pour envisager les changements au niveau d'un pixel, dans un premier temps, il s'agit de modéliser la distribution de la première image sur un voisinage du pixel étudié, ici par un modèle de mélange de gaussiennes, puis, dans un second temps, d'estimer dans quelle mesure l'intensité de ce pixel la seconde image respecte bien cette distribution.

Algorithme de Reed-Xiaoli Cette idée se rapproche beaucoup de l'algorithme de Reed-Xiaoli, utilisé en détection d'anomalies. Présenté par Reed et Yu dans les années 90 [RY90], il s'agit initialement de détecter, au sein d'une image, un objet de forme donnée mais de signature spectrale inconnue.

Dans un contexte général de détection de changements ou d'anomalies, il n'est pas réaliste de supposer la forme s connue, et l'algorithme de Reed-Xiaoli utilise en réalité la formulation donnée par Stein et coauteurs [SBH+02] qui prend le cas particulier où la forme recherchée est un unique pixel.

Cet algorithme de détection d'anomalie peut être décrit de la manière suivante. Pour évaluer le caractère anormal d'un pixel p, il s'agit de considérer l'ensemble des pixels situées dans une couronne autour de p, puis de modéliser leur intensité par une distribution gaussienne de moyenne μ et de matrice de covariance Σ. Une anomalie est alors détectée lorsque le critère de décision

$$\mathrm{RX}(x) \;:=\; (x-\mu)^{\mathsf{T}} M^{-1}(x-\mu), \tag{4.1}$$

où x est la valeur mesurée au pixel p, est inférieur à un certain seuil.

Nous proposons ici d'étendre cet algorithme au cas d'une paire d'image et à des distributions locales modélisées par des mélanges de gaussiennes, et en utilisant le cadre statistique a contrario.

4.1.2 Explicitation du modèle

4.1.2.1 Modélisation des images

Introduction Comme dans les modèles précédents, on considère deux images u et v, définies sur un même domaine $\mathcal{D} \subset \mathbb{Z}^2$, et possédant un nombre arbitraire de canaux, $K > 0$. Nous verrons ainsi chaque image comme une fonction de $\mathcal{D}$ dans $\mathbb{R}^K$, en notant u_p et v_p les vecteurs d'intensité dans $\mathbb{R}^K$ respectivement associés à u et v, pour chaque pixel $p \in \mathcal{D}$.

De plus, contrairement au premier modèle, par « comparaison ponctuelle », et à l'instar de celui par « comparaison d'histogramme », on ne demande ici aucune propriété quant aux différents canaux des images. S'il est donc possible de considérer les canaux bruts, on peut parfaitement imaginer utiliser un ensemble de descripteurs. Dans ce cas, K correspondra à la dimension totale des descripteurs, le nombre de canaux initiaux n'étant pas spécifié. La seule contrainte que l'on conserve ici est la correspondance entre les canaux des images, en supposant alors un même nombre K de canaux (ou de descripteurs) dans chaque image.

Choix du voisinage pour l'estimation de la distribution Étant donné un pixel $p \in \mathcal{D}$, il s'agit alors de modéliser la distribution locale de l'image u au voisinage de p. Pour cela, il est tout d'abord nécessaire de se donner un voisinage de p sur lequel évaluer cette distribution. Dans notre cas, nous avons simplement décidé de considérer un carré de largeur $2d+1$, pour $d > 0$, centré en p. Précisons que, comme nous l'avons signalé dans la section 3.2.1, nous ne considérerons pas ici les problèmes dus aux bords du domaine $\mathcal{D}$, en supposant formellement que les images sont aussi définies en dehors de $\mathcal{D}$.

Notons que l'on pourrait envisager d'adapter les voisinages à la géométrie de l'image, par exemple à l'aide de voisinages adaptatifs [DGP06] ou encore de superpixels [ASS+12]. En privilégiant, dans le voisinage, des pixels « similaires » à u_p, on biaiserait cependant l'estimation de la distribution vers des valeurs proches de celle u_p. Ce biais peut être intéressant dans certains cas, puisqu'il permettrait une meilleure estimation de la distribution au voisinage de u_p.

Toutefois, un tel biais peut s'avérer problématique en présence d'incohérences spatiales. Prenons le cas de deux images identiques présentant deux domaines A et B et envisageons un pixel situé, dans la première image, à la frontière des deux domaines, dans le domaine A. Avec cette configuration, un voisinage de taille fixe pris centré sur le pixel étudié tombera à cheval sur les deux domaines et entraînera donc une modélisation commune à A et B, tandis qu'un voisinage adaptatif privilégiera des pixels du domaine A, entraînant une modélisation plus fine du domaine A, sans modéliser réellement le domaine B. L'utilisation du voisinage adaptatif apparaît donc particulièrement pertinent si l'on est certain que le pixel étudié correspondra au même domaine A dans la seconde image. Seulement, si les deux images sont décalées, rien n'assure

que le pixel étudié ne soit pas situé du mauvais côté de la frontière dans la seconde image, au sein du domaine B.

Les décalages spatiaux étant presque toujours présent en imagerie satellite très haute résolution, nous avons préféré considérer une forme de voisinage indépendante des images.

Modèle de mélange de gaussiennes Pour modéliser la distribution locale de u au voisinage du pixel p, nous utilisons un mélange de gaussiennes, dont la fonction de densité s'écrit sous la forme

$$\forall x \in \mathbb{R}^K, \quad f(x|\alpha, \rho, \Sigma) \;=\; \sum_{r=1}^{R} \alpha_r \, g(x|\rho_r, \Sigma_r) \tag{4.2}$$

avec $R \in \mathbb{N}^*$, le nombre de gaussiennes, $\alpha_r \in \mathbb{R}^+$, $\rho_r \in \mathbb{R}^K$ et $\Sigma_r \in \mathbb{R}^{K\times K}$, respectivement la pondération, la moyenne et la matrice de covariance (que l'on suppose définie positive) de chaque gaussienne $r \in [\![1, R]\!]$ et en notant $g(x|\rho, \Sigma)$ la fonction de densité d'une gaussienne de moyenne ρ et de matrice de covariance Σ, c'est-à-dire

$$g(x|\rho, \Sigma) \;=\; (2\pi)^{-\frac{K}{2}} \, |\Sigma|^{-\frac{1}{2}} \exp\!\left(-\frac{1}{2}\,(x-\rho)^{\mathsf{T}}\,\Sigma^{-1}\,(x-\rho)\right), \tag{4.3}$$

en notant $|\Sigma|$ le déterminant de la matrice Σ. Nous avons représenté dans la figure 4.1 un exemple d'une telle distribution.

Notons qu'un tirage d'une variable aléatoire X suivant un tel mélange de gaussiennes peut être vu successivement comme le choix $S \in [\![1, R]\!]$ d'une des gaussiennes, tiré selon la pondération α, suivi d'un tirage suivant la gaussienne précédemment choisie. Ainsi définie, la densité de la variable X, conditionnée par l'événement $S = r$, pour $r \in [\![1, R]\!]$, correspond exactement à la densité de la s-ième gaussienne, $g(\cdot|\mu_r, \Sigma_r)$. Notons que l'on retombe alors bien sur la définition du mélange de gaussiennes,

$$f(x|\alpha, \rho, \Sigma) \;=\; \sum_{r=1}^{R} \mathbb{P}(\,S = r\,)\, f(x|\alpha, \rho, \Sigma, S = r) \;=\; \sum_{r=1}^{R} \alpha_r \, g(x|\rho_r, \Sigma_r), \tag{4.4}$$

en notant abusivement $f(x|\alpha, \rho, \Sigma, S = r)$ la densité de la variable X conditionnée par $S = r$.

Estimation par algorithme d'Espérance-Maximisation Pour un pixel p donné, on cherche à estimer la distribution de u sur un voisinage de ce pixel, que nous noterons P. Pour cela, il est naturel d'utiliser un algorithme d'Espérance-Maximisation (algorithme EM) [B+98]. Pour une valeur de R fixé et en notant $\theta = (\alpha_r, \mu_r, \Sigma_r)_{1\leq r\leq R}$ l'ensemble des paramètres du modèle, que l'on cherche à estimer, il s'agit de retrouver les paramètres θ^* expliquant au mieux les valeurs des u_q, pour $q \in P$, au sens de la maximisation de la vraisemblance L, c'est-à-dire

$$\theta^* \;=\; \arg\max_{\theta} L(u|\theta) \;=\; \arg\max_{\theta} \sum_{q\in P} \log f(u_q|\theta). \tag{4.5}$$

L'algorithme EM consiste alors à itérer deux phrases, dites d'espérance et de maximisation.

— La phase d'espérance (E) consiste, étant donné des paramètres θ^t, correspondant à une itération $t \in \mathbb{N}$, à estimer pour chaque $r \in [\![1, R]\!]$ et $q \in P$, la probabilité $w_{q,r}^{(t)}$ que u_q appartienne à la r-ième gaussienne,

$$w_{q,r}^{(t)} = \frac{\alpha_r^{(t)} g(u_q|\mu_r^{(t)}, \Sigma_r^{(t)})}{\sum_{r'=1}^{R} \alpha_{r'}^{(t)} g(u_q|\mu_r^{(t)}, \Sigma_{r'}^{(t)})}. \tag{4.6}$$

— La phrase de minimisation (M) consiste alors à trouver les nouveaux paramètres $\theta^{(t+1)}$ optimaux étant données les probabilités $w_{q,r}^{(t)}$,

$$\alpha_r^{(t+1)} = \frac{1}{|P|} \sum_{q \in P} w_{q,r}^{(t)} \tag{4.7}$$

$$\mu_r^{(t+1)} = \frac{1}{\sum_{q \in P} w_{q,r}} \sum_{q \in P} w_{q,r}^{(t)} u_q \tag{4.8}$$

$$\Sigma_r^{(t+1)} = \frac{1}{\sum_{q \in P} w_{q,r}} \sum_{q \in P} w_{q,r}^{(t)} (u_q - \mu_r^{(t+1)})(u_q - \mu_r^{(t+1)})^{\mathsf{T}}, \tag{4.9}$$

pour tout $r \in [\![1, R]\!]$.

Il est important de noter que cet algorithme nécessite d'avoir au préalable une valeur de R. S'il existe plusieurs méthodes de sélection de modèle permettant de trouver « la meilleure » valeur, notamment par pénalisation de la vraisemblance [Ler92, HPZ17], nous fixerons ici le nombre de gaussiennes sans chercher à l'estimer.

4.1.2.2 Expression du nombre de fausses alarmes

Modèle a contrario On suppose que l'on a modélisé la distribution de u, sur le voisinage P d'un pixel $p \in \mathcal{D}$, par un modèle de mélange de $R > 0$ gaussiennes, de paramètres $\theta = (\alpha_r, \mu_r, \Sigma_r)_{1 \leq r \leq R}$, en reprenant les notations précédentes.

Nous allons ici supposer que, en l'absence de changement, la variable aléatoire V_p, modélisant v_p, suit la distribution du mélange de gaussiennes précédemment estimée. On peut alors montrer la proposition suivante.

Proposition 4.1.1. *On se donne comme modèle naïf que, pour tout $p \in \mathcal{D}$, V_p a pour distribution un mélange de gaussienne de paramètres $\theta^p = (\alpha_r^p, \mu_r^p, \Sigma_r^p)_{1 \leq r \leq R}$, pour $R \in \mathbb{N}^*$. Alors la fonction suivante est un nombre de fausses alarmes,*

$$\mathrm{NFA}(p, u, v) = \frac{|\mathcal{D}|}{\Gamma\left(\frac{K}{2}\right)} \Gamma\left(\frac{K}{2}, \frac{\min_r (d_{\rho_r^p, \Sigma_r^p}(v_p)^2 + m_r^p) - \max_r m_r^p}{2}\right), \tag{4.10}$$

où $m_r^p = \log |\Sigma_r^p| - 2 \log \alpha_r^p$ et où $d_{\rho,\Sigma}(x)^2 = (x-\rho)^{\mathsf{T}} \Sigma^{-1} (x-\rho)$ est la distance de Mahalanobis.

Démonstration. Dans toute la suite, nous omettrons les dépendances en p des différents paramètres, pour ne pas alourdir inutilement les notations. Considérons la mesure de structure suivante

$$\mu(p,u,v) \;=\; \min_{1\le r\le R}\Big(m_r + d_{\rho_r,\Sigma_r}(v_p)^2\Big)\,, \tag{4.11}$$

où $m_r = \log|\Sigma_r| - 2\log\alpha_r$. Le théorème 3.1.2, relatif à la construction de nombres de fausses alarmes, nous assure alors que la fonction

$$F(p,u,v) \;=\; |\mathcal{D}|\;\mathbb{P}\big(\,\mu(p,u,V)\ge\mu(p,u,v)\,\big) \tag{4.12}$$

est bien un nombre de fausses alarmes. Il reste alors à montrer que l'on peut majorer cette fonction par l'expression annoncée pour pouvoir conclure à l'aide de la proposition 3.1.5 relative à la majoration des nombres de fausses alarmes.

Avant de continuer et comme nous l'avons mentionné dans la section 3.1.1.3, il est préférable, lorsque l'on cherche à majorer des nombres de fausses alarmes, d'éviter des majorations trop grossière. De ce fait, nous avons décidé de numéroter et de commenter en fin de section, chaque majoration effectuée ici.

Soit $\delta > 0$, on cherche à majorer la probabilité $\mathbb{P}\big(\,\mu(p,u,V)\ge\delta\,\big)$. Puisque V_p suit un mélange de gaussienne, considérons la variable aléatoire $S\in[\![1,R]\!]$ représentant l'indice de la gaussienne ayant engendré V_p et, comme dans l'équation (4.4), conditionnons par la valeur de S,

$$\begin{aligned}
\mathbb{P}\big(\,\mu(p,u,V)\ge\delta\,\big) &= \sum_{s=1}^{R}\mathbb{P}\big(\,S=s\,\big)\,\mathbb{P}\Big(\,\mu(p,u,V)\ge\delta\;\big|\;S=s\,\Big)\\
&= \sum_{s=1}^{R}\alpha_s\,\mathbb{P}\Big(\min_{1\le r\le R}\Big(m_r+d_{\rho_r,\Sigma_r}(V_p)^2\Big)\ge\delta\;\big|\;S=s\Big)\\
&= \sum_{s=1}^{R}\alpha_s\,\mathbb{P}\Bigg(\bigcap_{1\le r\le R}\Big(d_{\Sigma_r,\mu_r}(V_p)^2\ge\delta-m_r\Big)\;\Big|\;S=s\Bigg)\,.
\end{aligned} \tag{4.13}$$

Pour chaque terme $s\in[\![1,R]\!]$, on peut effectuer une première majoration

$$\mathbb{P}\Bigg(\bigcap_{1\le r\le R}\Big(d_{\Sigma_r,\mu_r}(V_p)^2\ge\delta-m_r\Big)\;\Big|\;S=s\Bigg)\;\le\;\mathbb{P}\Big(\,d_{\Sigma_s,\rho_s}(V_p)^2\ge\delta-m_s\;\big|\;S=s\,\Big)\,, \tag{4.14}$$

constant à négliger les contraintes des gaussiennes autres que celle pour $r=s$. En conditionnant par $S=s$, V_p suit une distribution gaussienne de même paramètres que ceux de la norme de Mahalanobis utilisée ici. Dès lors et de la même manière que pour la proposition 3.4.4, on peut voir $d_{\Sigma_s,\mu_s}(V_p)$ comme la norme euclidienne d'une variable gaussienne centrée réduite, de même dimension que V_p, et ainsi, pour tout $s\in[\![1,R]\!]$,

$$\mathbb{P}\Big(\,d_{\Sigma_s,\rho_s}(V_p)^2\ge\delta-m_s\;\big|\;S=s\,\Big)\;=\;\frac{1}{\Gamma\left(\frac{K}{2}\right)}\,\Gamma\Big(\frac{K}{2},\frac{\delta-m_s}{2}\Big)\,, \tag{4.15}$$

le paramètre K correspondant au nombre de composantes des images. On trouve donc

$$\mathbb{P}\big(\,\mu(V_p)\ge\delta\,\big)\;\le\;\frac{1}{\Gamma\left(\frac{K}{2}\right)}\sum_{r=1}^{R}\alpha_r.\Gamma\Big(\frac{K}{2},\frac{\delta-m_r}{2}\Big)\,. \tag{4.16}$$

La fonction gamma incomplète supérieure étant décroissante, on peut effectuer une seconde majoration de cette expression

$$\begin{aligned}\frac{1}{\Gamma\left(\frac{K}{2}\right)}\sum_{r=1}^{R}\alpha_r.\Gamma\left(\frac{K}{2},\frac{\delta-m_r}{2}\right) &\overset{(2)}{\leq} \frac{1}{\Gamma\left(\frac{K}{2}\right)}\max_{r\in[\![1,R]\!]}\Gamma\left(\frac{K}{2},\frac{\delta-m_r}{2}\right),\\ &\leq \frac{1}{\Gamma\left(\frac{K}{2}\right)}\Gamma\left(\frac{K}{2},\frac{\delta-\max_r m_r}{2}\right),\end{aligned} \tag{4.17}$$

qui correspond bien à l'expression annoncée. □

Calcul numérique Comme nous l'avons déjà signalé précédemment, il est préférable dans la pratique de calculer le logarithme des nombres de fausses alarmes pour éviter des erreurs d'arrondis à zéro. Pour cela, il est à nouveau nécessaire de pouvoir évaluer les logarithmes des fonctions gamma et gamma incomplète avec une grande précision. Nous utilisons pour cela la méthode de Abergel et co-auteur [AM16].

4.1.2.3 Remarques

Mesure de structure considérée Pour pouvoir estimer la cohérence de la valeur de v_p vis-à-vis de la distribution estimée sur u, la mesure de structure utilisée dans la propriété précédente repose sur la mesure de log-vraisemblance,

$$L(v_p) = \log\left(\sum_{r=1}^{R}\alpha_r g_r(v_p)\right), \tag{4.18}$$

en notant $g_r(v_p) = g(v_p|\rho_p,\Sigma_p)$ la fonction de densité gaussienne associée à la r-ième gaussienne (voir équation (4.3)). La vraisemblance n'est pas utilisée directement, puisqu'elle ne permet pas d'obtenir une majoration facilement utilisable de la probabilité

$$\mathbb{P}(\,\mu(p,u,V)\geq\delta\,),$$

pour $\delta>0$. La mesure de structure considérée s'écrit comme une approximation de la vraisemblance,

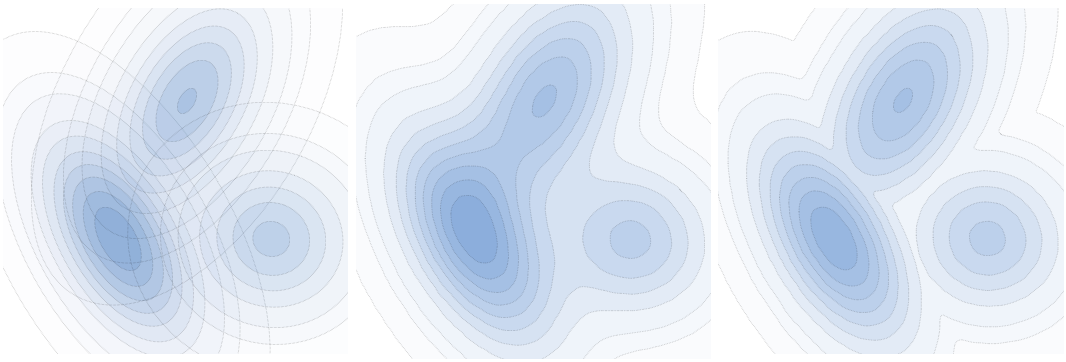
$$L(v_p) = \log\left(\sum_{r=1}^{R}\alpha_r g_r(v_p)\right) \simeq \max_{1\leq r\leq R}\log\left(\alpha_r g_r(v_p)\right) \overset{\text{def}}{=} \widetilde{L}(v_p), \tag{4.19}$$

avec

$$\begin{aligned}-2\widetilde{L}(x) &= \min_{1\leq r\leq R}\left(K\log 2\pi+\log|\Sigma_r|-2\log\alpha_r+d_{\rho_r,\Sigma_r}(x)^2\right)\\ &= K\log 2\pi+\mu(p,u,v).\end{aligned} \tag{4.20}$$

La densité de probabilité correspondant à la vraisemblance approchée $\widetilde{L}$ a été représentée dans la figure 4.1 au côté de celle du mélange de gaussiennes initial.

Notons bien que cette approximation n'intervient aucunement dans le cadre a contrario, la mesure de structure pouvant être choisie arbitrairement. Néanmoins, il



Les différentes gaussiennes apparaissant dans le mélange | Le mélange de gaussiennes | Le maximum des gaussiennes

Figure 4.1 – Représentation schématique d'un mélange de gaussiennes. Nous avons représenté successivement les différentes gaussiennes faisant partie du mélange, le mélange de gaussiennes, et la distribution correspondant au maximum des gaussiennes et correspondant à la distribution obtenue après approximation dans l'équation (4.19). Dans chaque cas, les densités de probabilité sont représentées au moyen de lignes de niveau liant tous les points de même valeur de densité, en coloriant chaque zone alors définie par une couleur d'autant plus foncée que la densité dans cette zone est élevée.

peut être pertinent de commenter cette approximation. Comme on peut le voir dans la figure 4.2, en considérant le maximum des gaussiennes à la place de leur somme, on sous-estime la densité des points situés à cheval sur plusieurs gaussiennes. Cette approximation apparaît donc d'autant plus valide que les gaussiennes sont loin les unes des autres.

Retour sur les approximations effectuées Comme nous l'avons fait remarquer dans la section 3.1.1.3, il est important, lorsque l'on cherche à estimer la probabilité de fausses alarmes, de ne pas effectuer de majorations trop grossières. Reprenons alors rapidement les deux majorations permettant d'obtenir l'expression du nombre de fausses alarmes données dans la proposition 4.1.1.

La première majoration, faite dans l'équation (4.14), correspond à une majoration séparée de la contribution de chaque gaussienne. Comme nous l'avons représenté dans la figure 4.2. Il s'agit ici de négliger la probabilité, en étant généré par une des gaussiennes, de tomber proche d'une autre. De la même manière que précédemment, cette approximation est d'autant plus valable que les différentes gaussiennes sont éloignées les unes des autres, leur densité respective ne se chevauchant pas trop.

La seconde majoration, effectuée dans l'équation (4.16), revient à considérer que le point que l'on observe a été généré par la gaussienne la plus probable. Notons que cette dernière majoration n'est pas forcément nécessaire et ne permet que d'accélérer le calcul en se ramenant de R à un seul calcul de la fonction gamma incomplète.

Régularisation des distributions Il peut apparaître un problème lors de l'estimation du modèle de mélange de gaussiennes, plus particulièrement au niveau de l'estimation des différentes matrices de covariance. Dans certains cas, notamment pour

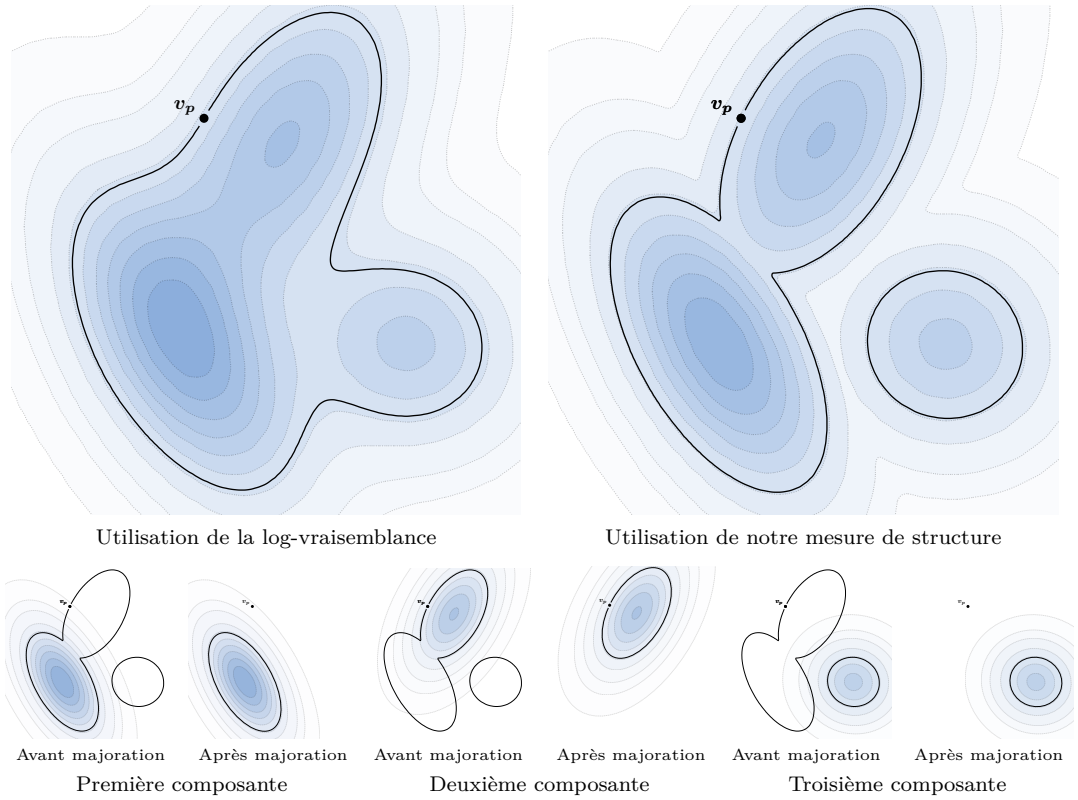


Figure 4.2 – Représentation des différentes probabilités intervenant dans le calcul du nombre de fausses alarmes du modèle de détection de changements par modélisation locale du contexte. Sur la première ligne, nous avons représenté successivement la distribution en mélange de gaussienne, correspondant au modèle naïf, puis son approximation, donnée dans l'équation (4.19). Sur la seconde ligne, nous avons représenté les trois composantes de cette probabilité, telles que détaillées dans l'équation (4.13). Dans chaque cas, nous avons représenté la probabilité avant et après la majoration effectuée dans l'équation (4.14). Dans chaque image, nous avons affiché en trait noir la limite du domaine correspondant aux différentes probabilités représentées (ce dernier étant situé à l'extérieur de cette limite) et en couleur la densité de probabilité intégrée, représentée par ses lignes de niveau.

des zones très homogènes, il peut arriver que l'utilisation d'un trop grand nombre de gaussiennes, avec peu de points, entraîne l'apparition de gaussiennes dégénérées, dont la matrice de covariance n'est pas inversible..

Pour éviter cela, on peut se donner une borne inférieure, $\lambda_{\min}$, sur les valeurs propres des matrices de covariances des différentes gaussiennes intervenant dans le mélange. Pour ce faire, nous pouvons reprendre l'algorithme EM d'estimation du modèle et rajouter, à chaque étape, une contrainte sur la valeur minimale des valeurs propres. Pour une matrice de covariance Σ, il suffira alors d'effectuer une décomposition en valeurs propres, $\Sigma = PDP^{\mathsf{T}}$, pour P orthogonale et D diagonale, et de considérer $\widetilde{\Sigma}^{\lambda} = P\widetilde{D}^{\lambda}P^{\mathsf{T}}$, où $\widetilde{D}^{\lambda}$ est la matrice diagonale obtenue en rehaussant à λ les valeurs trop faibles, c'est à dire $\widetilde{D}^{\lambda}_{ii} = \min(D_{ii}, \lambda_{\min})$ pour tout $i \in [\![1, K]\!]$.

Cette diagonalisation pouvant être un peu coûteuse, nous n'effectuerons ce rehaussement des valeurs propres qu'une seule fois, après convergence de l'algorithme EM.

Asymétrie du détecteur et symétrisation Il est important de remarquer que, contrairement aux modèles précédents, ce modèle par modélisation locale du contexte est fondamentalement asymétrique. En effet, on évalue ici les intensités d'une image (ici v) à partir de la distribution locale d'une autre, utilisée comme référence (ici u).

Il est toutefois possible de symétriser ce modèle en utilisant la propriété 3.1.7. On pourra ainsi considérer une version symétrique du modèle par modélisation locale du contexte en posant

$$\mathrm{NFA}_{\mathrm{sym}}(p, u, v) \;=\; 2\min\Big(\mathrm{NFA}(p, u, v), \mathrm{NFA}(p, v, u)\Big)\,. \tag{4.21}$$

4.1.3 Résultats et limitation du modèle

4.1.3.1 Résultats sur les exemples synthétiques

Résultats généraux De la même manière que pour les modèles précédents, nous avons évalué les résultats de ce modèle sur le jeu de données synthétiques, présenté dans la figure 2.9, et ce en se restreignant à deux descripteurs d'images élémentaires, à savoir la nuance de gris (voir section 3.3.2) et l'estimation du gradient orienté, calculé à partir des filtres de Sobel (voir section 3.3.3.1). Nous avons représenté ces résultats dans la figure 4.3.

Tout d'abord, il est intéressant de remarquer l'asymétrie du modèle, particulièrement visible dans la paire 1, où il détecte séparément les apparitions et les disparitions. Comme attendu, le modèle détecte ici tous les pixels dont l'intensité se distingue de celles présentées dans le voisinage de l'image prise en référence. Ainsi, on détecte bien la majorité des apparitions d'objets dans les différents exemples, et ce, sans détecter les translations légères de la paire 2 ou les variations de textures de la paire 4. Il est intéressant de remarquer que l'apparition du petit nénuphar, dans le coin inférieur gauche de la paire 2, n'est presque pas détecté, sa proximité avec d'autres nénuphars étant compris par le modèle comme une translation. De plus, certains éléments de la texture de mer de la paire 4 sont détectés par le modèle, cette texture présentant une échelle trop grande devant la taille du contexte choisie ici. Cet exemple permet de se rendre

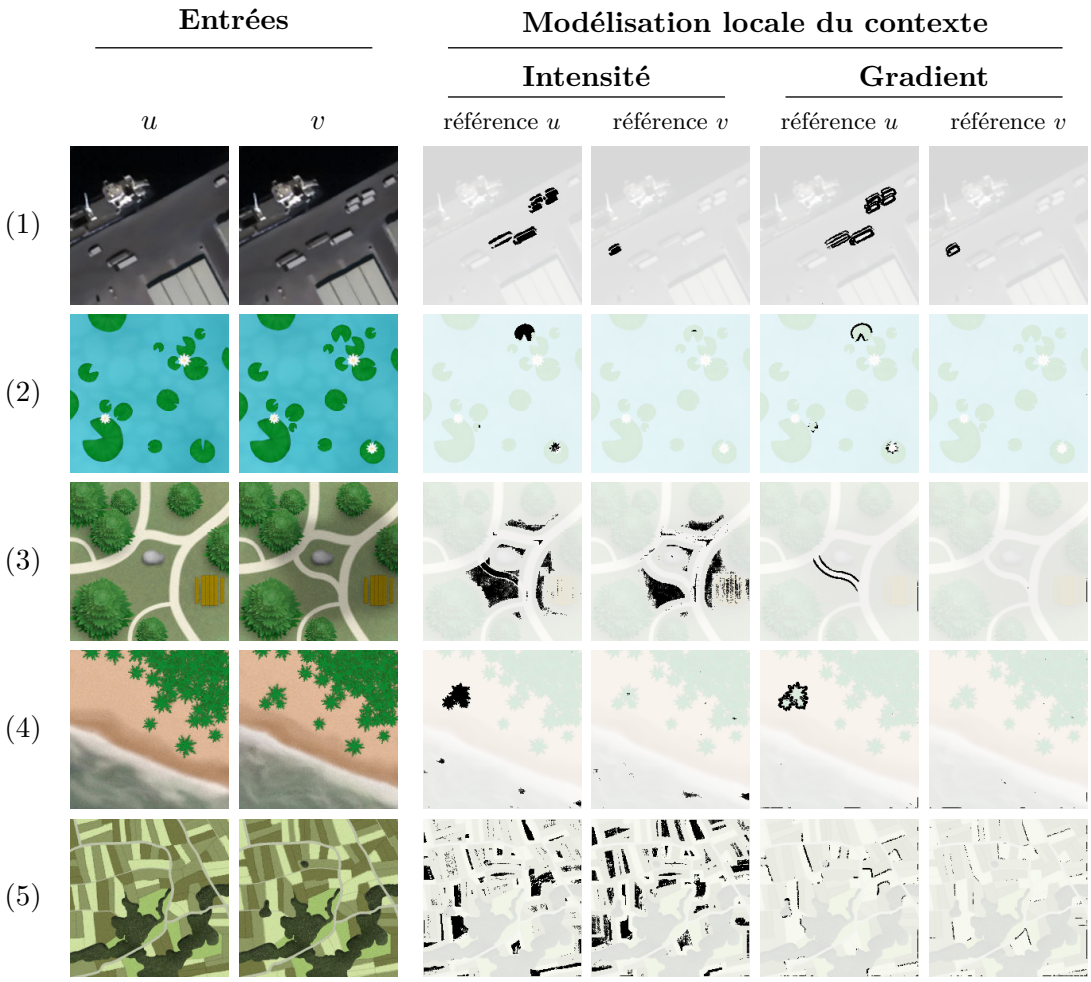


Figure 4.3 – Résultats du modèle de détection de changements par modélisation locale du contexte sur les exemples synthétiques présentés figure 2.9, en utilisant comme nombre de gaussiennes $R = 3$, ainsi qu'un contexte carré de taille 11×11. Nous affichons les résultats pour deux types de descripteurs, l'intensité en nuance de gris et le gradient orienté, en considérant l'une ou l'autre des images comme référence. Chaque ligne représente un exemple et est constituée successivement des deux images u et v suivies des résultats du modèle pour chacun des descripteurs considérés. Dans les images de résultat, nous avons représenté en noir les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$, le tout affiché sur une version éclaircie de la seconde image, v. Comme attendu ici, le modèle ne détecte ni les translations de la paire 2, ni les variations de micro-textures de la paire 4. On voit de plus, dans la paire 1, que le modèle parvient à distinguer les apparitions des disparitions d'objets. On peut remarquer que l'utilisation du gradient permet de rendre le modèle robuste aux changements de contraste local de la paire 3, au prix d'une détection réduite au bord des objets.

compte de l'adéquation nécessaire entre la taille du contexte et la taille caractéristique des objets et textures que l'on souhaite appréhender.

Par ailleurs, si, comme attendu, le modèle détecte le changement d'illumination de la paire 3, il est toutefois important de remarquer qu'il n'est pas sensible aux légers changements de contraste de la paire 4, contrairement au modèle par comparaison d'histogramme. Pour finir, dans la paire 5, l'utilisation de l'intensité provoque des détections un peu partout dans l'image, le modèle ne reposant que sur un contexte local variant beaucoup entre les deux images. Par ailleurs, tout comme pour le modèle par comparaison d'histogramme, et pour les mêmes raisons, l'utilisation du gradient a tendance à masquer la plupart des changements.

Distinction entre apparitions et disparitions Comme nous l'avons déjà signalé, le modèle est ici capable de distinguer une apparition d'une disparition. En effet, d'un côté, en modélisant tout d'abord le sol, il est relativement facile de détecter les apparitions d'objets dont l'intensité s'écarte significativement de celle du sol. De l'autre côté, dans le cas d'une disparition d'objet, le modèle va modéliser un contexte présentant l'objet en question ainsi que, si la taille du contexte est suffisamment grande, une partie du sol. Dès lors, l'évaluation du sol dans la seconde image ne provoquera a priori pas de détection.

Remarquons par ailleurs que ce modèle permet de distinguer l'apparition d'un objet « isolé », de celle à proximité d'objets similaires, ce que ne permettaient pas les modèles précédents. Cela apparaît clairement dans la paire 2, où l'apparition du petit nénuphar n'est que très peu détectée, et où celle du plus gros est tronquée en partie.

Influence de la taille du contexte Pour finir, intéressons-nous à la manière dont la taille du contexte influe sur la détection de ce modèle. Nous avons affiché, dans la figure 4.4, les résultats obtenus en faisant varier la taille du contexte utilisé pour estimer la distribution locale dans l'image de référence. Tout d'abord, remarquons à quel point la finesse de la détection est conservée en augmentant la taille du contexte, contrairement au modèle par comparaison d'histogramme qui avait tendance à diminuer, d'une certaine manière, la résolution du résultat (voir figure 3.15).

Par ailleurs, observons dans la paire 2 à quel point, en augmentant la taille du contexte, la petite apparition (dans le coin inférieur gauche) disparaît progressivement. Le même phénomène se produit dans le haut de l'image, où la détection y apparaissant est progressivement rognée. De manière similaire, dans la paire 4, il est intéressant d'observer qu'un contexte de taille réduite permet rapidement d'appréhender les variations de micro-textures, tandis qu'une macro-texture en nécessitera un plus large pour éviter des détections dues à ses variations. Ainsi, la taille du contexte correspond en quelque sorte à l'échelle à laquelle l'algorithme va évaluer les différents objets présents dans la scène.

4.1.3.2 Limitations du modèle

Changement au sein d'une texture Tout d'abord, comme nous l'avons fait remarquer précédemment, le modèle de changements par modélisation locale du contexte

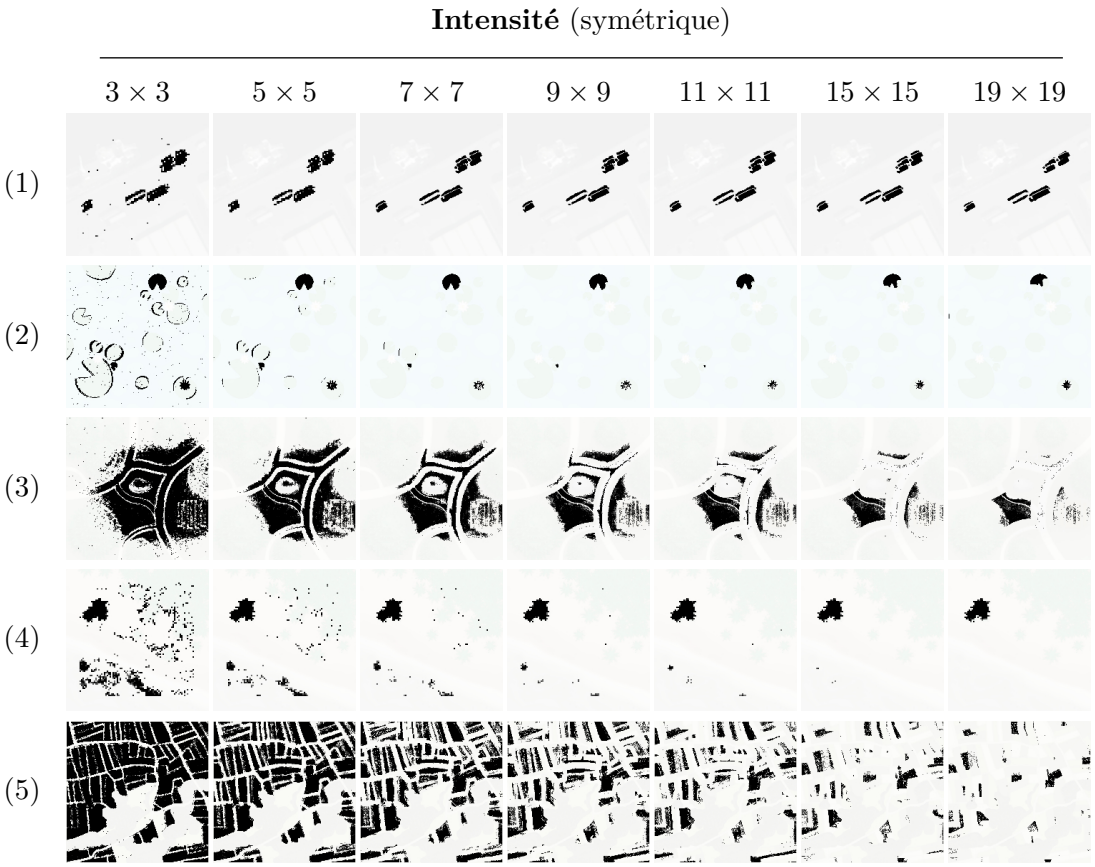


Figure 4.4 – Résultats du modèle de détection de changements par modélisation locale du contexte sur les exemples synthétiques présentés dans la figure 2.9, en fonction de la taille du contexte utilisé. Nous avons utilisé la version symétrisée du modèle, en considérant l'intensité en nuance de gris, un mélange de 3 gaussiennes, et des contextes carrés de taille allant de 3×3 à 19×19. Chaque ligne représente un exemple et est composée des résultats obtenus pour les différentes tailles de contexte. Dans chaque image, nous avons représenté en blanc les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$, le tout affiché sur une version assombrie de la seconde image. Il apparaît qu'en augmentant la taille du contexte, le modèle devient de plus en plus robuste aux décalages spatiaux, comme on peut le voir dans la paire 2, ainsi qu'aux variations de textures, comme on peut le voir dans la paire 4. Remarquons toutefois, qu'en élargissant trop ce contexte, on ne détecte plus certains changements, comme l'apparition du petit nénuphar dans la paire 2, qui n'est plus du tout détecté à partir d'une taille de 15×15 pixels.

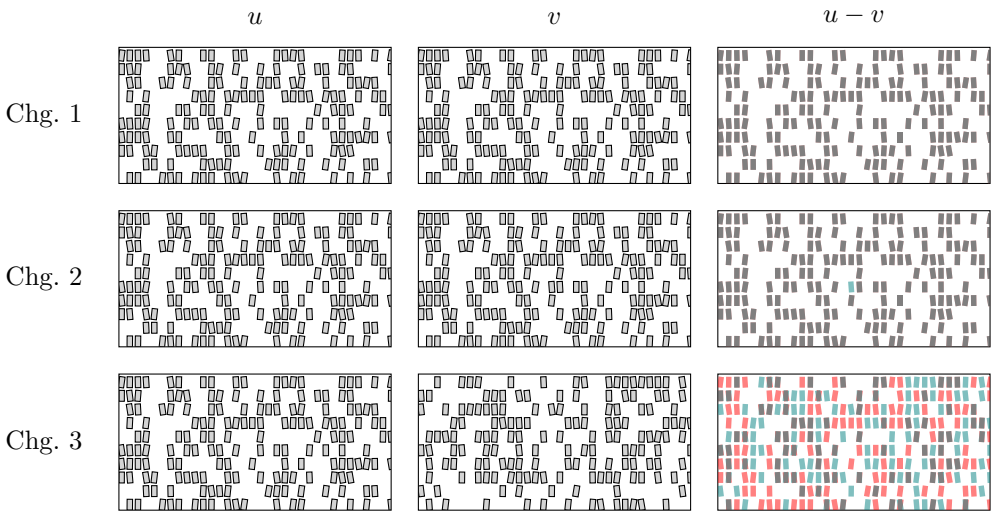


Figure 4.5 – Exemples de changements indistinguables par le modèle de détection de changements par modélisation locale du contexte, illustrant la limitation de ce modèle vis-à-vis des apparitions d'objets au sein d'une texture. Chaque ligne représente une paire d'image. Pour faciliter la visualisation des changements, nous avons représenté dans une troisième colonne, une représentation de la « somme colorée » des deux images. En rouge (●) sont affichés les objets présents uniquement dans la première images, en bleu (●), ceux présents uniquement dans la seconde, et en gris (●), ceux présents en même temps dans les deux. Dans la première paire, les deux images sont identiques. Dans la deuxième, on retrouve les mêmes images que dans la première paire, à l'exception d'un objet apparaissant au centre de la seconde image. Dans la troisième paire, chaque image présente un tirage aléatoire différente de la texture. Les objets apparaissant et disparaissant étant toujours situés à proximité d'objets similaires, le modèle de détection de changements par modélisation locale du contexte sera incapable de distinguer ces trois paires.

sera incapable de détecter une apparition d'un objet, si ce dernier intervient trop proche d'un objet similaire. À cause de cela, et comme présenté dans la figure 4.5, il sera donc impossible de distinguer le cas d'une même texture présente dans les deux images, une modification locale au sein de celle ci et une variation complète de la texture.

Il est cependant important de noter que cette limitation ne concerne que le modèle par modélisation locale du contexte et que l'utilisation du modèle par comparaison ponctuelle permet de résoudre ce problème.

Modélisation imparfaite Il est important de noter que l'on ne dispose dans la pratique que d'un nombre réduit de points pour pouvoir estimer la distribution locale des images. Si la modélisation paramétrique, ici par un mélange de gaussiennes, permet de compenser ce faible nombre, il est important de garder en tête quelques limitations dues à cette modélisation réduite.

Cela concerne tout d'abord les zones présentant un grand nombre d'objets d'inten-

sités différentes. Dans ces cas, la distribution locale que l'on cherchera à estimer sera alors trop complexe pour pouvoir être modélisée avec un nombre réduit de gaussiennes. Le même problème intervient aussi si l'on augmente la dimension des descripteurs utilisés. En effet, en augmentant la dimension de l'espace, on augmente en même temps la quantité de points nécessaire pour pouvoir estimer proprement la distribution locale.

Dans un cas comme dans l'autre, le risque est alors d'obtenir une modélisation grossière. Notons que, a priori, cette limitation aura plus tendance à masquer des changements qu'à en faire apparaître de nouveaux, ce qui peut sembler relativement acceptable dans certains cas d'usage.

Forte complexité temporelle Une autre limitation de ce modèle réside dans le temps de calcul. Si les modèles précédents pouvaient traiter une image très rapidement, la nécessité d'apprendre un modèle de mélange de gaussiennes différent pour chaque pixel demande beaucoup de temps de calcul.

Il est toutefois possible d'accélérer cette estimation au pris d'un biais relativement négligeable. Lors de l'initialisation de chaque algorithme EM, utilisé pour estimer la distribution locale, on pourrait utiliser les modèles déjà estimés sur les patchs voisins, qui devrait donner des modèles relativement similaires et donc demander moins d'itérations pour converger. Il faut toutefois garder en tête qu'en faisant ainsi, on introduit un biais dans l'estimation du modèle. En effet, l'algorithme EM ne converge pas vers une unique distribution, plusieurs itérations menant parfois à plusieurs solutions différentes. En fournissant une initialisation à l'algorithme EM, on risque ainsi de biaiser la distribution estimée vers la distribution fournie.

Notons que nous n'avons pas utilisé ce procédé pour obtenir les résultats que nous présentons ici.

Vision uniquement locale Pour finir, la limitation principale de ce modèle, que possèdent aussi les deux modèles précédent, est que l'on se restreint à une évaluation locale de l'image. De ce fait, il est impossible de comprendre les notions de changements globaux et locaux, présentés dans la paire 5 des exemples synthétiques, visible dans la figure 4.3.

4.2 Modèle de changements par destin commun

4.2.1 Introduction

Dans les modèles précédents, que ce soit le modèle par comparaison ponctuelle, comparaison d'histogramme ou par modélisation locale du contexte, on ne dispose que d'une compréhension locale des deux images, réduite à un voisinage autour des pixels étudiés. Ces approches locales ne permettent ne permettent pas d'appréhender globalement les images et ne peuvent donc pas distinguer les changements locaux (apparaissant en un faible nombre d'emplacements au sein du domaine), les changements globaux (qui concernent tous les pixels), et les autres changements (présents en de multiples endroits, mais ne concernant pas l'intégralité des pixels), ces types de changements étant discutés dans la section 2.3.2.

Images radiométriquement différentes On se place ici dans un cadre comprenant des changements globaux et on considère deux images, pouvant être « radiométriquement différentes », dans le sens où il n'est a priori pas possible de les comparer directement. Dans la pratique, ces différences radiométriques peuvent être dues à l'utilisation de longueurs d'onde différentes (si on cherche par exemple à comparer une image visible et une image infrarouge), à des différences au niveau des systèmes optiques, deux satellites différents entraînant deux dynamiques d'images différentes, ou encore à des fortes variations d'intensité d'une image sur l'autre (comme c'est particulièrement le cas en infrarouge thermique).

Considération globale Pour ce modèle, nous visualisons les changements comme des groupements de pixels dont la variation d'intensité entre les deux images se distingue des autres pixels. Plusieurs méthodes utilisent déjà cette idée d'évaluer la transition d'intensité des pixels vis-à-vis de la distribution globale estimée sur la totalité des images, en particulier celles de Theiler et coauteur [TP06], et Gueguen et coauteurs [GSP11]. Dans les deux cas, il s'agit tout d'abord de modéliser la loi jointe $p_{U,V}$ des deux images U et V, puis d'évaluer localement des groupes de pixels vis-à-vis de la distribution estimée.

Theiler et coauteur proposent de modéliser la distribution globale par une distribution gaussienne, puis d'évaluer chaque pixel $p \in \mathcal{D}$ en seuillant la grandeur

$$\frac{p_{U,V}(u_p, v_p)}{p_U(u_p)p_V(v_p)}, \tag{4.22}$$

pour en déduire les changements. L'avantage de la modélisation gaussienne est qu'elle permet d'obtenir des expressions explicites des différentes probabilités intervenant dans la grandeur à seuiller. Toutefois, la simplicité du modèle gaussien ne permet pas de modéliser finement cette distribution jointe.

À l'inverse, Gueguen et coauteurs [GSP11] proposent d'utiliser la loi jointe empirique,

$$p_{U,V}(x, y) \;=\; \frac{1}{Z}\left(1 + \sum_{p\in\mathcal{D}} \mathbb{1}_{\{(u_p,v_p)=(x,y)\}}\right), \tag{4.23}$$

pour $x \in \mathbb{R}^n$ et $y \in \mathbb{R}^m$, n et m étant les nombres de canaux respectifs des images u et v et où Z est un coefficient de normalisation. S'il est alors possible d'obtenir une distribution jointe aussi précise que possible, il est toutefois nécessaire de disposer d'un grand nombre d'échantillons pour obtenir une distribution empirique représentative de la loi estimée. Par ailleurs, Gueguen et coauteurs proposent d'utiliser un critère de décision reposant sur l'information mutuelle et la théorie de l'information. Comme nous l'avons déjà signalé dans l'introduction, il est alors critique de disposer de suffisamment de données pour bien estimer la distribution jointe.

Destin radiométrique Pour limiter les problèmes liés à l'information mutuel, nous proposons ici de quantifier fortement l'information modélisée. Nous nous intéressons ainsi à l'ensemble des pixels partageant plus ou moins une même intensité dans l'image u. Sans changement, on s'attend à ce que ces pixels évoluent de manière coordonnée,

en voyant les changement comme des groupements spatiaux de pixels dont l'intensité commune dans l'image v se distingue du « destin radiométrique » standard des autres pixels.

4.2.2 Formalisation du problème

4.2.2.1 Modélisation des images et mesure de structure

Formalisation Comme dans les modèles précédents, on considère deux images u et v, définies sur un même domaine $\mathcal{D} \subset \mathbb{Z}^2$. Toutefois, contrairement aux modèles précédents, on ne suppose plus nécessairement que les canaux des deux images se correspondent. On note alors K_U et K_V les nombres de canaux respectifs de u et v, avec ainsi

$$u : \mathcal{D} \to \mathbb{R}^{K_U} \qquad \text{et} \qquad v : \mathcal{D} \to \mathbb{R}^{K_V}, \tag{4.24}$$

que l'on verra souvent comme des vecteurs, respectivement dans $\mathbb{R}^{|\mathcal{D}| \times K_U}$ et $\mathbb{R}^{|\mathcal{D}| \times K_V}$. De même que précédemment, on notera U et V les variables aléatoires modélisant u et v. Les deux images jointes (U, V) sont vu ici comme un tirage d'un champ de Markov stationnaire. Un changement sera alors vu comme un ensemble de pixels $E \subset \mathcal{D}$ dont les valeurs dans les deux images ne correspondent pas aux statistiques globales de la paire d'image. De plus, pour rajouter une asymétrie dans le modèle, nous nous intéresserons particulièrement à la distribution de V conditionnée par l'observation faite de U.

Binarisation de l'information Fixons dès à présent un ensemble de pixels $E \subset \mathcal{D}$, correspondant potentiellement à un changement et notons $X_{|E}$ la restriction à E d'une image X définie sur $\mathcal{D}$. Comme dit précédemment, nous nous intéressons à la loi de $V_{|E}$ conditionnée par $U_{|E} = u_{|E}$. Il s'agit d'une distribution dans un espace de grande dimension, $\mathbb{R}^{|E| \times (K_U + K_V)}$. Pour éviter le fléau de la dimension, qui demanderait un très grand nombre de point pour pouvoir l'estimer, nous restreignons ici l'information à modéliser, en quantifiant les descripteurs des deux images. Étant donné deux sous-ensembles de l'espace des descripteurs, $F \subset \mathbb{R}^{K_U}$ et $G \subset \mathbb{R}^{K_V}$, on considère alors, pour tout $p \in \mathcal{D}$,

$$u_p^F = \mathbb{1}_{\{u_p \in F\}} \qquad \text{et} \qquad v_p^G = \mathbb{1}_{\{v_p \in G\}}, \tag{4.25}$$

ainsi que U^F et V^G, définies de manière similaire. On se ramène ainsi à l'étude de $(U^F, V^G)_{|E}$, à valeur dans $\{0, 1\}^{2|\mathcal{D}|}$.

Ensemble de structures potentielles Dans le cadre du formalisme a contrario, il est nécessaire de fixer l'ensemble des structures envisagés comme potentielles détections. Dans le cadre des modèles précédents, il s'agissait simplement des positions pixelliques, s'apparentant alors simplement au domaine $\mathcal{D}$. Toutefois ici, la situation est un peu plus compliquée. Une structure potentielle étant définie a priori par un triplet (E, F, G), il apparaît comme infaisable dans la pratique de considérer réellement tous les triplets possibles. Aussi, nous nous restreignons à un sous ensemble de ces triplets.

En ce qui concerne les ensembles E, nous considérons, pour chaque pixel p, le patch E_p centré sur p et dont nous aurons fixé la taille en amont.

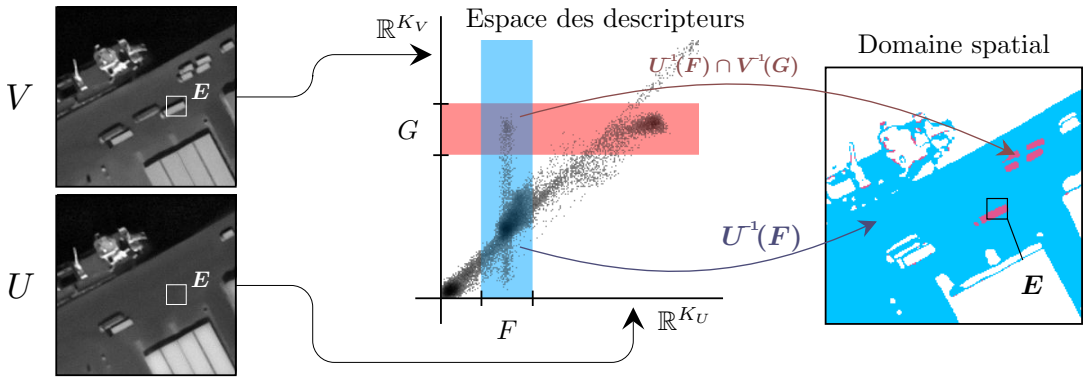


Figure 4.6 – Représentation schématique des différents objets définis dans le cadre du modèle de détection de changements par destin commun. On représente notamment trois ensembles E, F et G correspondant à une structure possible, les domaines pixelliques $u^{-1}(F)$ et $v^{-1}(G)$ correspondants à F et G.

Pour le choix des domaines de descripteurs F et G, on pourrait simplement considérer tous les ensembles possibles, à une ou plusieurs tailles fixées, éventuellement en se donnant un « pas d'échantillonage ». Toutefois, cela amènerait un très grand nombre de couples (F, G), dont la plupart ne contiendrait que très peu de pixels et n'aurait donc que très peu d'intérêt. Nous avons donc considéré les couples (F_q, G_q), indexés par $q \in \mathcal{D}$, et centrés respectivement sur les valeurs de u_q et v_q, dont le rayon, noté η_u et η_v aura été fixé en amont, c'est-à-dire

$$F_p \;=\; \mathcal{B}_{K_U}(u_p, \eta_u) \qquad \text{et} \qquad G_p \;=\; \mathcal{B}_{K_V}(v_p, \eta_v), \tag{4.26}$$

où $\mathcal{B}_n(x, \delta)$ est la boule de $\mathbb{R}^n$, centrée sur $x \in \mathbb{R}^n$ et de rayon $\delta > 0$. Nous discuterons plus tard du choix des différents rayons η_u et η_v.

De cette manière, l'ensemble des structures envisagées ici s'écrit

$$\mathcal{I} \;=\; \left\{(E_p, F_q, G_q),\ (p, q) \in \mathcal{D}^2\right\}, \tag{4.27}$$

pour un total de $|\mathcal{D}|^2$ structures.

Mesure de structure Pour réduire encore la dimension de l'espace et simplifier la modélisation de la loi jointe, nous avons décidé, il est intéressant de ne pas considérer l'information de localisation au sein de E et de n'utiliser ainsi que l'histogramme de $(U^F, V^G)_{|E}$. Notons que, ces deux images étant binaires, l'histogramme est donc réduit à 4 valeurs. En conditionnant par la valeur de $U^F_{|E}$, il s'agit donc de compter le nombre de pixels de E, dont les valeurs passent de F dans u à G dans v, c'est-à-dire

$$\mu(E, F, G; u, v) \;=\; \sum_{p \in E} u_p^F v_p^G \;=\; \mathrm{Card}(u^{-1}(F) \cap v^{-1}(G) \cap E)\,, \tag{4.28}$$

en notant $u^{-1}(F)$ et $v^{-1}(G)$ les domaines pixelliques associés aux ensembles F et G,

$$u^{-1}(F) = \{p \in \mathcal{D}', u_p \in F\} \qquad \text{et} \qquad v^{-1}(G) = \{p \in \mathcal{D}', v_p \in G\}\,. \tag{4.29}$$

Nous avons représenté dans la figure 4.7 les différentes grandeurs intervenant dans ce modèle de détection de changements.

4.2.2.2 Modèle de changements avec indépendance des pixels

Proposition 4.2.1. *Soit $\mathcal{I}$ un ensemble fini de triplets de $\mathcal{P}(\mathcal{D}) \times \mathcal{P}(\mathbb{R}^{K_U}) \times \mathcal{P}(\mathbb{R}^{K_V})$. En considérant, comme modèle naïf, que la distribution de V_p sachant U_p est égale à sa distribution empirique, c'est à dire entre autres que pour tout $(E, F, G) \in \mathcal{I}$ et tout $p \in \mathcal{D}$,*

$$\mathbb{P}\Big(V_p \in G \,\big|\, U_p \in F \Big) \;=\; \frac{\sum_{p\in\mathcal{D}} \mathbb{1}_{\{v_p \in G, u_p \in F\}}}{\sum_{p\in\mathcal{D}} \mathbb{1}_{\{u_p \in F\}}} \tag{4.30}$$

alors la fonction définie pour tout $(E, F, G) \in \mathcal{I}$ par

$$\mathrm{NFA}(E, F, G; u, v) \;=\; |\mathcal{D}|^2\, \mathcal{B}\left(|E \cap u^{-1}(F)|\,, |E \cap u^{-1}(F) \cap v^{-1}(G)|\,, \frac{|v^{-1}(G) \cap u^{-1}(F)|}{|u^{-1}(F)|} \right), \tag{4.31}$$

est un nombre de fausses alarmes, où $\mathcal{B}$ est la queue de la distribution binomiale, définie pour $p \in [0, 1]$, $N \in \mathbb{N}^*$ et $k \in [\![0, N]\!]$, par

$$\mathcal{B}(N, k, p) \;=\; \sum_{i=k}^{N} \binom{N}{i} p^i (1-p)^{N-i}. \tag{4.32}$$

Démonstration. Le théorème 3.1.2, relatif à la construction des nombres de fausses alarmes, appliqué à la mesure de structure donnée précédemment,

$$\mu(E, F, G; u, v) \;=\; \mathrm{Card}(E \cap u^{-1}(F) \cap v^{-1}(G))\,, \tag{4.33}$$

nous assure que la fonction

$$F(E, F, G, u, v) \;=\; |\mathcal{D}|^2\, \mathbb{P}\Big(\mu(E, F, G; U, V) \geq \mu(E, F, G; u, v) \,\big|\, U = u \Big)\,, \tag{4.34}$$

pour tout $(E, F, G) \in \mathcal{I}$, est un nombre de fausses alarmes. Il s'agit alors de calculer cette probabilité de fausses alarmes.

Soit $(E, F, G) \in \mathcal{I}$. Remarquons alors que $\mu(E, F, G; U, V)$ correspond au nombre de pixels de $E \cap u^{-1}(F)$ dont l'intensité dans la seconde image appartient à G. Par ailleurs, par hypothèse du modèle naïf, la probabilité de $V_p \in G$ sachant $U_p \in F$ est indépendante de p et donnée par la distribution empirique

$$\alpha \;\stackrel{\text{def}}{=}\; \mathbb{P}\Big(V_p \in G \,\big|\, U_p \in F \Big) \;=\; \frac{|v^{-1}(G) \cap u^{-1}(F)|}{|u^{-1}(F)|}. \tag{4.35}$$

Il apparaît donc que la variable $\mu(E, F, G; U, V)$ suit une loi binomiale avec $|E \cap u^{-1}(F)|$ tests et une probabilité α de succès.

Pour $\delta \geq 0$, la probabilité $\mathbb{P}(\mu(E, F, G; U, V) \geq \delta \mid U = u)$ apparaît donc comme la queue d'une distribution binomiale de paramètre $|E \cap u^{-1}(F)|$ et α et on retrouve bien l'expression annoncée. $\square$

Estimation et précision numérique Comme nous l'avons déjà mentionné plusieurs fois, les valeurs nombres de fausses alarmes peuvent être très faible et il est donc préférable de calculer leur logarithme pour éviter des problèmes d'arrondis à zéros. Il est pour cela nécessaire de pouvoir estimer le logarithme de la queue de la distribution binomiale,

$$\log \mathcal{B}(N,k,p) \;=\; \log \sum_{i=k}^{N} \binom{N}{i} p^i (1-p)^{N-i} \;=\; \log \sum_{i=k}^{N} u_i, \tag{4.36}$$

pour $N \in \mathbb{N}$, $k \in [\![0, N]\!]$ et $p \in [0,1]$, et en notant $u_i = \binom{N}{i} p^i (1-p)^{N-i}$. Un problème d'arrondi intervient lorsque N est grand et k trop proche de N, les termes u_i, pour $i \geq k$, étant alors trop proche de zéros. Pour éviter cela, il est possible d'écrire cette expression sous la forme

$$\log \mathcal{B}(N,k,p) \;=\; \log u_k + \log\left(1 + R_k\right) \qquad \text{où} \qquad R_k \;=\; \sum_{i=1}^{N-k} \frac{u_{k+i}}{u_k}. \tag{4.37}$$

Le premier terme peut être calculé explicitement,

$$\begin{aligned} \log u_k \;=\; & \log \Gamma(N+1) - \log \Gamma(k+1) - \log \Gamma(N-k+1) \\ & + k \log p + (N-k) \log(1-p). \end{aligned} \tag{4.38}$$

Le second terme ne pose plus de problème d'arrondi et peut ainsi être calculé directement en utilisant

$$R_k \;=\; \sum_{i=1}^{N-k} \frac{u_{k+i}}{u_k} \;=\; \sum_{i=1}^{N-k} \frac{(N-k)!\,k!}{(N-k-i)!\,(k+i)!} \left(\frac{p}{1-p}\right)^i. \tag{4.39}$$

4.2.2.3 Modélisation jointe par un mélange de gaussiennes

Simplification des notations Pour un triplet $(E,F,G) \in \mathcal{I}$, on considère les deux variables aléatoires

$$X = \mathrm{Card}(E \cap U^{-1}(F)) \qquad \text{et} \qquad Y = \mathrm{Card}(E \cap U^{-1}(F) \cap V^{-1}(G)), \tag{4.40}$$

ainsi que x^0 et y^0, les valeurs effectivement mesurées, correspondant à u et v. Ces variables permettent d'écrire de manière simplifiée la probabilité

$$\mathbb{P}\Big(\mu(E,F,G;U,V) \geq \delta \;\big|\; U = u\Big) \;=\; \mathbb{P}\Big(Y \geq \delta \;\big|\; X = x^0\Big), \tag{4.41}$$

pour $\delta \geq 0$. Nous avons représenté ces deux grandeurs dans la figure 4.7.

Notons bien que ces quatre variables dépendent du triplet (E,F,G) choisi, mais pour ne pas alourdir les notations, nous ne noterons pas explicitement ces dépendances.

Par ailleurs, le champs (U,V) étant supposé stationnaire, les distributions de X et Y sont donc invariantes par translation de E. De ce fait, on a accès à de multiples tirages de (X,Y), en translatant l'ensemble E dans le domaine $\mathcal{D}$. En notant, pour tout pixel $p \in \mathcal{D}$,

$$x_p = \mathrm{Card}(E_p \cap U^{-1}(F)) \qquad \text{et} \qquad y_p = \mathrm{Card}(E_p \cap U^{-1}(F) \cap V^{-1}(G)), \tag{4.42}$$

où E_p est la version translatée de E centrée en p, chaque (x_p, y_p) correspond alors à un tirage différente de la variable (X,Y).

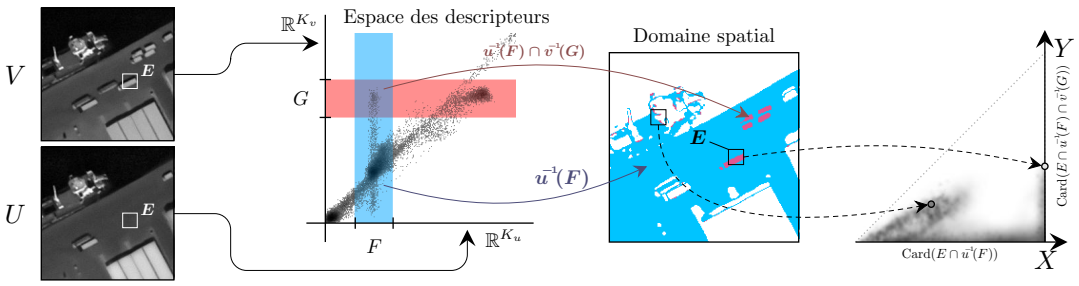


Figure 4.7 – Représentation schématique des différents objets définis dans le cadre du modèle de détection de changements par destin commun. On reprend ici le schéma de la figure 4.6 en rajoutant les deux grandeurs X et Y correspondant respectivement au cardinal de $E \cap u^{-1}(F)$ et de $E \cap u^{-1}(F) \cap v^{-1}(G)$.

Distribution en mélange de gaussiennes Plutôt que de rajouter une hypothèse d'indépendance sur le champ associé à (U, V), il est possible de modéliser directement la distribution de (X, Y) au moyen des (x_p, y_p), pour $p \in \mathcal{D}$. Pour cela, nous utilisons un modèle de mélange de gaussiennes, déjà évoqué dans la section 4.1.2.1, de fonction de densité de probabilité

$$p(x,y) \;=\; \sum_{r=1}^{R} \alpha_r g_r(x,y) \qquad \text{où} \qquad g_r(x,y) \;=\; \frac{e^{-\frac{1}{2}\left(\binom{x}{y}-\mu_r\right)^{\mathsf{T}}\Sigma_r^{-1}\left(\binom{x}{y}-\mu_r\right)}}{2\pi\sqrt{|\Sigma_r|}}, \tag{4.43}$$

où $x, y \in [\![0, |E|]\!]^2$, $|\Sigma|$ est le déterminant de la matrice Σ, et R est le nombre de gaussiennes utilisées, g_r représentant la densité de la r-ième gaussienne, de moyenne μ_r et de matrice de covariance Σ_r.

Lemme 4.2.2. *Soit deux variables aléatoires X et Y à valeur dans $\mathbb{R}$. Si la variable jointe (X, Y) suit une distribution gaussienne de moyenne et matrice de covariance*

$$\mu = (\mu_x, \mu_y) \qquad \textit{et} \qquad \Sigma = \begin{pmatrix} \sigma_x & \sigma_{xy} \\ \sigma_{xy} & \sigma_y \end{pmatrix}. \tag{4.44}$$

alors pour tout $x \in \mathbb{R}$, la variable Y conditionnée par $X = x$ suit une distribution gaussienne de paramètre $\mu_{|x}$ et $\sigma_{|x}$, avec

$$\mu_{|x} \;=\; \mu_y + \frac{\sigma_{xy}}{\sigma_x}(x - \mu_x) \qquad \textit{et} \qquad \sigma_{|x} \;=\; \sigma_y - \frac{\sigma_{xy}^2}{\sigma_x}. \tag{4.45}$$

Démonstration. Nous donnons la démonstration de ce lemme en annexe, section A.3. □

Proposition 4.2.3. *Soit $\mathcal{I}$ un ensemble fini de triplets de $\mathcal{P}(\mathcal{D}) \times \mathcal{P}(\mathbb{R}^{K_U}) \times \mathcal{P}(\mathbb{R}^{K_V})$. On considère, comme modèle naïf, que pour tout triplet $\gamma = (E, F, G) \in \mathcal{I}$, le couple (X, Y) correspondant a pour distribution un mélange de gaussiennes de densité*

$$p^\gamma(x,y) \;=\; \sum_{r=1}^{R} \alpha_r^\gamma g_r^\gamma(x,y), \tag{4.46}$$

où R est le nombre de gaussiennes considérés et g_r^γ est la densité gaussienne de moyenne μ^γ et de matrice de covariance Σ^γ. Considérons la fonction définie pour tout triplet $\gamma = (E, F, G) \in \mathcal{I}$ par

$$\mathrm{NFA}(\gamma; u, v) \;=\; \frac{|\mathcal{D}|^2}{2} \sum_{r=1}^{R} \left(\frac{\alpha_r^\gamma g_r^\gamma(x^0)}{\sum\limits_{r=1}^{R} \alpha_r^\gamma g_r^\gamma(x^0)} \operatorname{erfc}\left(\frac{y^0 - \mu_{|x^0}^\gamma}{\sqrt{2\sigma_{|x^0}^\gamma}} \right) \right), \tag{4.47}$$

où

$$x^0 = \mathrm{Card}(E \cap u^{-1}(F)) \qquad \textit{et} \qquad y^0 = \mathrm{Card}(E \cap u^{-1}(F) \cap v^{-1}(G)),$$

et en notant $\mu_{|x^0}^\gamma$ et $\sigma_{|x^0}^\gamma$ la moyenne et variance de la distribution gaussienne conditionnée définie dans le lemme 4.2.2. Cette fonction est un nombre de fausses alarmes.

Démonstration. Le théorème 3.1.2, relatif à la construction des nombres de fausses alarmes, appliqué à la mesure de structure donnée précédemment,

$$\mu(E, F, G; u, v) \;=\; \mathrm{Card}(E \cap u^{-1}(F) \cap v^{-1}(G)) \;=\; X, \tag{4.48}$$

nous assure que la fonction

$$\begin{aligned} F(E, F, G, u, v) &= |\mathcal{D}|^2\, \mathbb{P}\Big(\mu(E, F, G; U, V) \geq \mu(E, F, G; u, v) \;\big|\; U = u \Big) \\ &= |\mathcal{D}|^2\, \mathbb{P}\big(Y \geq \delta \mid X = x^0 \big), \end{aligned} \tag{4.49}$$

pour tout $(E, F, G) \in \mathcal{I}$, est un nombre de fausses alarmes. De plus, pour $\gamma = (E, F, G) \in \mathcal{I}$, on a

$$\begin{aligned} \mathbb{P}\Big(Y \geq y^0 \;\big|\; X = x^0 \Big) &= \int_{y \geq y^0} p^\gamma(|x^0) dy \\ &= \frac{1}{p^\gamma(x^0)} \int_{y \geq y^0} p^\gamma(y, x^0) dy \\ &= \frac{\sum\limits_{r=1}^{R} \left(\alpha_r^\gamma g_r^\gamma(x^0) \int_{y \geq y^0} g_r^\gamma(|x^0) dy \right)}{\sum\limits_{r=1}^{R} \alpha_r^\gamma g_r^\gamma(x^0)}, \end{aligned} \tag{4.50}$$

en notant $p(x)$ et $p(y|x)$ les densités marginale et conditionnelle, associées à $p(x, y)$ et $g_r(x)$ et $g_r(y|x)$ celles associées à $g_r(x, y)$. D'après le lemme 4.2.2, $g_r(y|x)$ correspond à la fonction de densité d'une gaussienne de paramètre $\mu_{|x^0}^\gamma$ et $\sigma_{|x^0}^\gamma$ et ainsi, de manière similaire à la preuve de la proposition 3.4.1, on obtient

$$\int_{y \geq y^0} g_r^\gamma(|x^0) dy \;=\; \frac{1}{2} \operatorname{erfc}\left(\frac{y^0 - \mu_{|x^0}^\gamma}{\sqrt{2\sigma_{|x^0}^\gamma}} \right). \tag{4.51}$$

On trouve bien le résultat annoncé,

$$F(E, F, G, u, v) \;=\; \frac{|\mathcal{D}|^2}{2} \sum_{r=1}^{R} \left(\frac{\alpha_r^\gamma g_r^\gamma(x^0)}{\sum\limits_{r=1}^{R} \alpha_r^\gamma g_r^\gamma(x^0)} \operatorname{erfc}\left(\frac{y^0 - \mu_{|x^0}^\gamma}{\sqrt{2\sigma_{|x^0}^\gamma}} \right) \right). \tag{4.52}$$

□

4.2.3 Résultats et limitations du modèle

4.2.3.1 Résultats sur les exemples synthétiques

Asymétrie du détecteur et symétrisation Le modèle de détection de changements par destin commun est volontairement asymétrique, puisqu'on évalue une image (par défaut v) par rapport à une image de base (par défaut u). Il est toutefois possible de symétriser ce modèle en considérant, comme avec le modèle précédent,

$$\mathrm{NFA}_{\mathrm{sym}}(E,F,G;u,v) \;=\; 2\min\Big(\mathrm{NFA}(E,F,G;u,v),\mathrm{NFA}(E,G,F;v,u)\Big)\,, \tag{4.53}$$

qui est bien un nombre de fausses alarmes d'après la propriété 3.1.7.

Paramètres et structures envisagées En ce qui concerne l'ensemble des structures considérées ici, nous nous sommes fixés trois rayons, un rayon $r > 0$ pour E, un rayon $\eta_u > 0$ pour F et un rayon $\eta_v > 0$ pour G, et avons considéré, alors comme structures potentielles, l'ensemble des triplets

$$(E_p, F_q, G_q) \;=\; \Big(\mathcal{B}_2(p,r), \mathcal{B}_{K_U}(u_q,\eta_u), \mathcal{B}_{K_V}(v_q,\eta_v)\Big) \tag{4.54}$$

indicés pour $(p,q) \in \mathcal{D}^2$, en notant $\mathcal{B}_n(x,\delta)$ la boule de $\mathbb{R}^n$, centrée sur $x \in \mathbb{R}^n$ et de rayon $\delta > 0$. Pour les résultats obtenus, nous avons pris $r = 3$, donnant des ensembles E de taille 7×7, et comme rayon η_u et η_v un pourcentage de l'amplitude maximal de l'image,

$$\eta_u \;=\; \max_{p,q\in\mathcal{D}} \gamma \left\|u_p - u_q\right\|_2\,, \tag{4.55}$$

avec, dans la pratique, $\gamma \in [10\%, 40\%]$. De la même manière que pour les modèles précédents, nous avons alors appliqué ce modèle au jeu de données synthétiques, présenté dans la figure 2.9, et avons affiché les résultats obtenus dans la figure 4.8. Notons que nous nous sommes ici restreint à la nuance de gris.

Remarques générales Tout d'abord, il est intéressant de remarquer que l'on détecte bien la majorité des apparitions et disparitions locales au sein des différentes paires d'images, que ce soit les véhicules dans la paire 1, les nénuphars dans la paire 2, le sentier dans la paire 3, le groupe de palmiers dans la paire 4 ou encore les taches dans la paire 5. Par ailleurs, comme attendu, on ne détecte plus le changement de contraste local dû à l'ombre dans la paire 3, ou la rotation des cultures de la paire 5, en détectant tout de même celui dont l'évolution est anormale.

Il est intéressant de remarquer que ce modèle parvient à être robuste aux faibles décalages spatiaux de la paire 2. Cela est dû à la présente de multiples décalages dans toutes l'images. Une transition « eau » vers « nénuphar » étant perçue comme rare, mais pas inexistante, il est alors possible d'expliquer une translation faible, n'impactant qu'un nombre réduit de pixels. Notons cependant bien qu'il ne s'agit pas a proprement parler d'une robustesse à la translation, puisque tout d'abord, dans ces mêmes conditions, le modèle ne détectera pas une apparition d'objets dont l'échelle correspond à celle des translations. D'autre part, si l'image ne présente un décalage local que pour un seul objet, ce dernier sera a priori détecté.

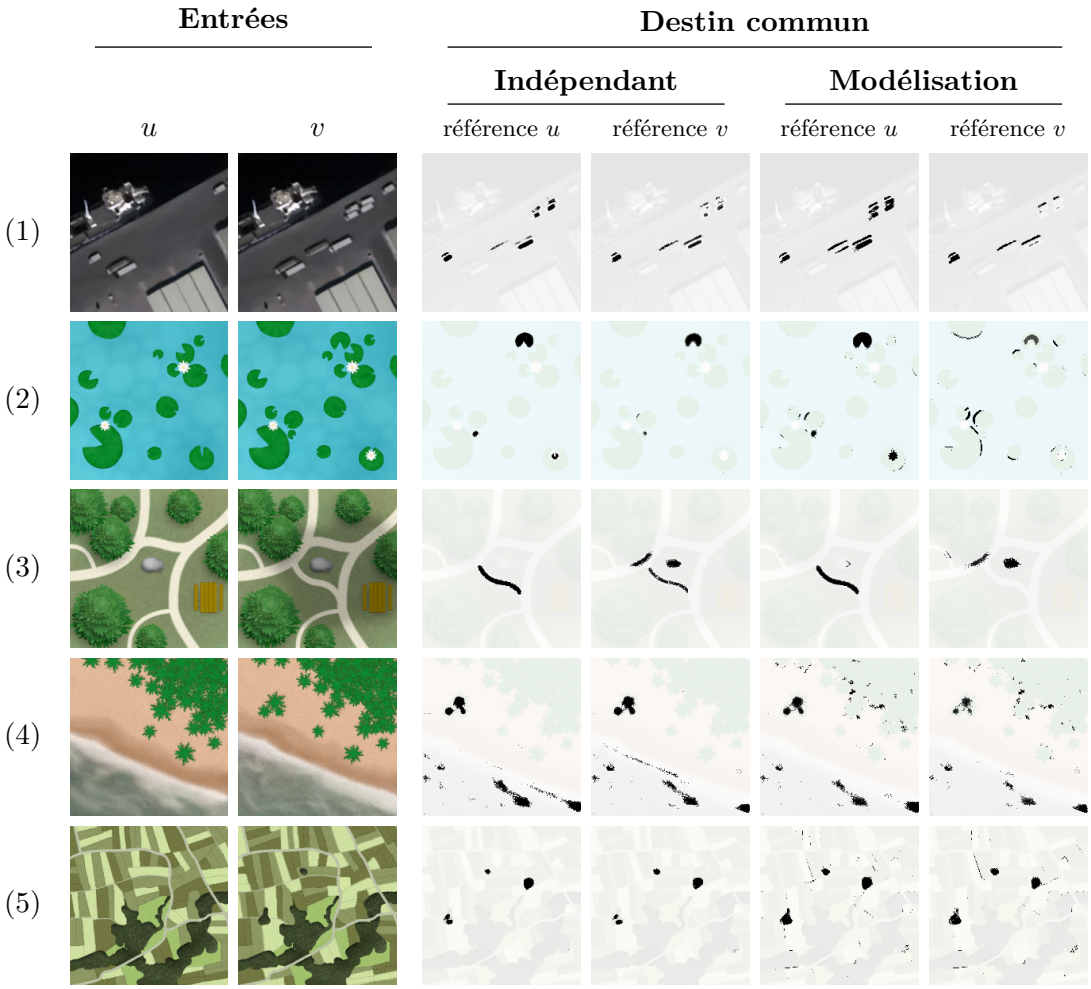


Figure 4.8 – Résultats du modèle de détection de changements par destin commun sur les exemples synthétiques présentés dans la figure 2.9, en prenant E de taille 7×7, F et G de largeur correspondant à 20% de l'amplitude maximale des images et en utilisant comme descripteur la nuance de gris. Les résultats sont affichés pour le modèle indépendant et pour la modélisation en mélange de gaussienne (en utilisant une unique gaussienne), et dans chaque cas, en conditionnant suivant l'une puis l'autre des images. Chaque ligne correspond à un exemple synthétique et dans chaque image de résultat, nous avons représenté en noir les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$, le tout affiché sur une version éclaircie de la seconde image. Le modèle détecte bien les différentes apparitions et disparitions présentes dans les divers exemples. Comme attendu, il parvient à ne pas détecter le changement de contraste local de la paire 3, ou encore la rotation des cultures de la paire 5, tout en détectant le champ évoluant anormalement par rapport aux autres (correspondant à la détection la plus à droite dans les résultats).

Résultats sur les micro et macro-textures Il peut être intéressant de s'attarder quelque peu sur la paire 4. Tout d'abord, on peut observer que les variations de la micro-texture au niveau des arbres sont relativement bien comprises par le modèle, au moins en son centre. Cela s'explique facilement par le caractère très répétitif de la texture, et ce à une petite échelle. Dans ce cadre très proche d'un modèle indépendant, il n'est donc pas surprenant de voir que la modélisation indépendante ne donne aucune détection. Il est intéressant de remarquer que, pour cette micro-texture et dans le cas de la modélisation gaussienne, la majorité des détections se trouvent à la périphérie de la texture, où cette dernière présente d'importantes variations.

Intéressons-nous à présent à la variation de macro-texture, présente au niveau de la mer. Ici, chacune des deux modélisations (indépendante et gaussienne) produit plusieurs détections, correspondant chacune à une structure se distinguant au sein de la texture. Dans le cadre du modèle gaussien, la principale raison à cela est que l'on ne dispose pas d'une texture assez grande pour que l'algorithme puisse réellement estimer la distribution locale de la texture, et on peut ainsi supposer que fournir une image plus grande aurait permis de mieux estimer la distribution et ainsi d'éviter ces détections. En revanche, dans le cadre du modèle indépendant, augmenter la taille de l'image ne suffirait certainement pas. En effet, dans les macro-textures, l'existence de structures à plus grande échelle, en particulier des zones relativement homogènes, remet complètement en question l'hypothèse d'indépendance, une telle homogénéité locale étant très peu probable sous cette hypothèse.

4.2.3.2 Retour sur le choix des paramètres

Taille du contexte spatial Intéressons nous à l'influence de la taille des ensembles E sur les résultats de l'algorithme. Nous avons affiché ces derniers dans la figure 4.9, en faisant varier cette taille entre 5×5 et 27×27.

Tout d'abord, en ce qui concerne le modèle indépendant, il est intéressant de remarquer que, pour une largeur de E trop importante, le modèle détecte une grande majorité de l'image, en particulier les zones homogènes. Le problème est en fait le même que celui exposé précédemment pour les macro-textures. Prenons pour l'exemple, un triplet (E, F, G) correspondant à un élément du fond dans les deux images et supposons qu'une grande proportion du fond α reste constante. Dans ce cas et si l'on suppose les pixels indépendants, même si α est très proche de 1, il est en réalité statistiquement improbable qu'aucun pixel de E ne change. On peut voir ici taille de E comme l'échelle à laquelle le modèle observe les structures locales, et il peut ainsi sembler normal que le modèle indépendant devienne d'autant moins pertinent que celle-ci augmente. Notons bien que la valeur limite de 11×11, qui semble apparaître ici, dépend en réalité de la valeur du seuil choisi sur le nombre de fausses alarmes.

Dans le cadre de la modélisation gaussienne, au contraire, il apparaît qu'augmenter trop cette taille a tendance à faire diminuer le nombre de détections, allant jusqu'à faire disparaître les différentes apparitions d'objets. Cela peut s'expliquer par la modélisation gaussienne que l'on effectue ici. En effet, en augmentant la taille de E on cherche alors à modéliser les variables X et Y (définies dans l'équation (4.40)), sur un domaine plus grand, et cela sans augmenter le nombre de points dont on dispose. On

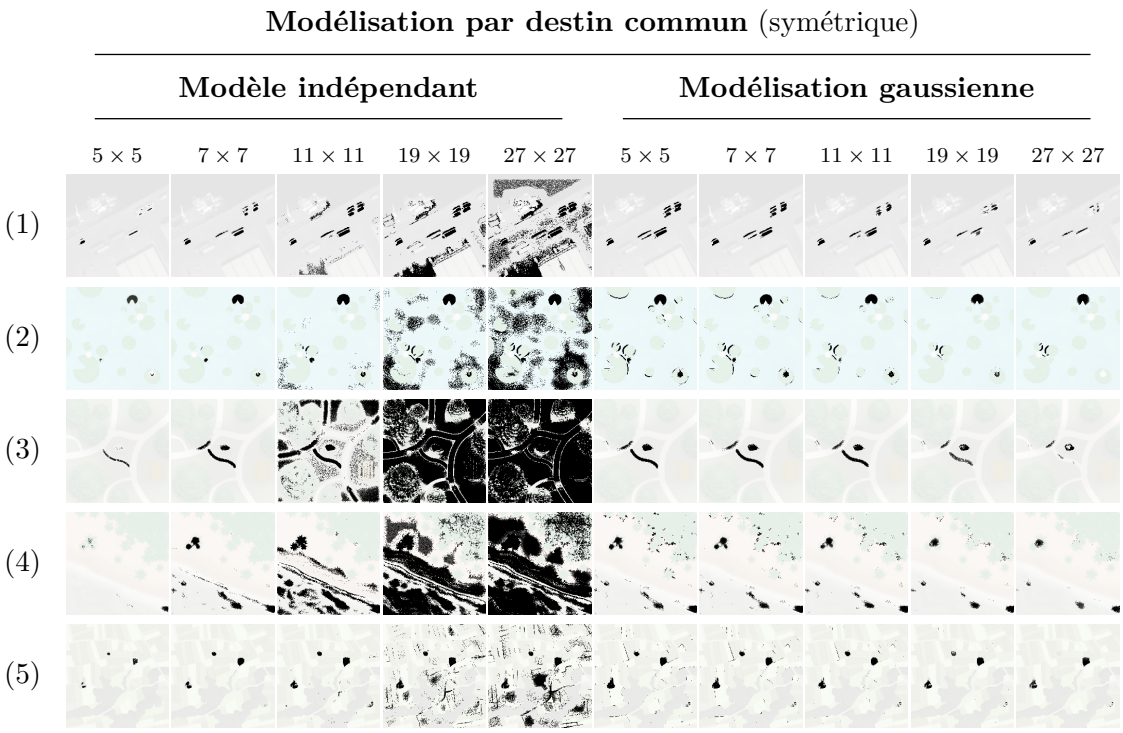


Figure 4.9 – Influence de la taille de l'ensemble E sur les résultats du modèle de détection de changements par destin commun sur les exemples synthétiques présentés dans la figure 2.9. Nous avons considéré les images en nuance de gris, en faisant varier la taille de E et en prenant F et G de largeur correspondant à 20% de l'amplitude maximale des images. Chaque ligne correspond à un exemple synthétique, présentant successivement les résultats du modèle indépendant, puis du modèle gaussien (utilisant une unique gaussienne). On applique ici une version symétrique de l'algorithme, obtenue en prenant le maximum des NFA, considérant l'une ou l'autre des images comme référence. En ce qui concerne le modèle indépendant, en augmentant la taille de E, on perçoit de plus en plus des structures dépendantes, ce qui entre en contradiction avec l'indépendance supposée des pixels. On peut ainsi observer les limites du modèle indépendant, pour E supérieur à 19×19, notamment au niveau des zones homogènes qui sont alors détectées, malgré un coefficient α proche de 1 (voir équation (4.35)). En ce qui concerne la modélisation gaussienne, en augmentant la taille de E, on voit au contraire le nombre de détections baisser. En effet, augmenter la largeur de E a pour effet d'augmenter la variabilité des points à modéliser. Le modèle s'en retrouve alors d'autant plus grossier et aura donc tendance à expliquer plus de changements. Dans les images de résultat, nous avons représenté en noir les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$, le tout affiché sur une version éclaircie de la seconde image.

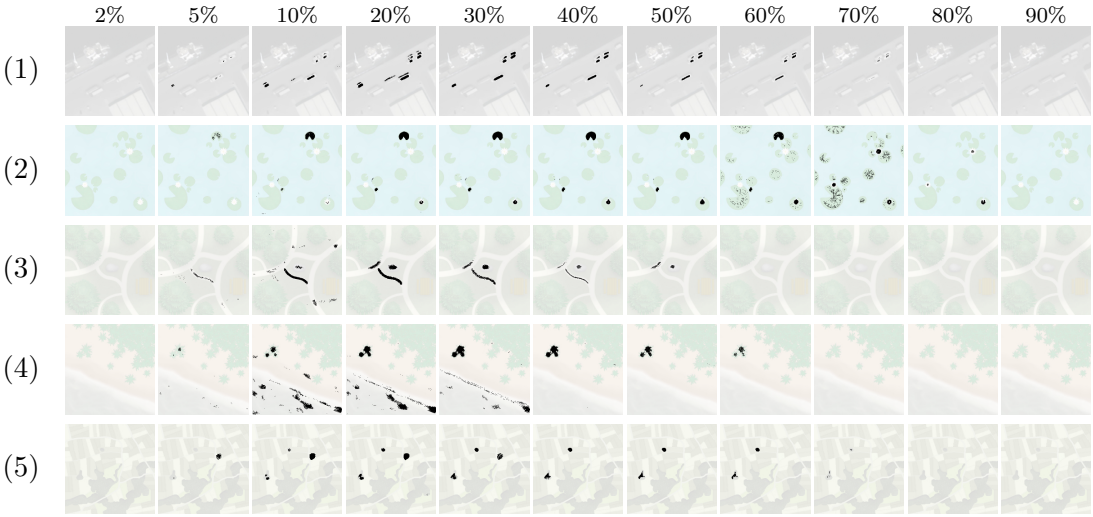


Figure 4.10 – Influence des largeurs des contextes F et G, notées η_u et η_v, sur les résultats du modèle de détection de changements par destin commun, sur les exemples synthétiques présentés dans la figure 2.9. Nous avons ici considéré les images en nuance de gris, en prenant les ensembles E de taille 7×7 et en considérant uniquement une version symétrique du modèle indépendant (le modèle gaussien présente de résultats très similaire). Ici, les valeurs de η_u et η_v sont choisies égales et nous donnons leur valeur en pourcentage de l'amplitude totale des deux images, allant de 2% à 90%. Dans les images de résultat, nous avons représenté en noir les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$, le tout affiché sur une version éclaircie de la seconde image. On voit ici que le modèle par destin commun ne détecte plus rien dès lors que la largeur du contexte est trop petite ou trop grande. La raison est simple : pour détecter une structure, il est nécessaire que le contexte soit suffisamment large pour inclure des structures similaires, sans être trop large pour ne pas inclure une trop grande partie de l'image. On peut par ailleurs apercevoir des détections étranges dans la paire 2, lorsque la largeur atteint les 70%.

s'attend alors à ce que le modèle appris soit plus grossier et donc aussi plus permissif, faisant ainsi baisser le nombre de détections.

Dans tous les cas, il est donc nécessaire limiter la taille des ensembles E utilisés, d'une part, pour éviter de trop prendre en compte des corrélations spatiales qui remettraient en cause l'indépendance supposée des pixels, d'autre part, pour éviter une modélisation trop imprécise.

Choix des tailles des contextes F et G Discutons à présent du choix des paramètres η_u et η_v, correspondant aux rayons des ensembles F et G. Nous avons affiché dans la figure 4.10 les résultats du modèle pour différentes valeurs de ces paramètres, supposés égaux. Il apparaît ici que pour pouvoir détecter une structure, il est nécessaire que le contexte utilisé pour les descripteurs soit suffisamment large pour inclure des structures similaires, sans être trop large pour ne pas inclure une trop grande partie de l'image. Il est donc important, pour pouvoir obtenir des détections, de prendre une valeur de ce paramètre qui soit adaptée aux images, tant d'un point de vue global que

d'un point de vue local.

Pour choisir un paramètre adapté, on peut imaginer plusieurs solutions. La première est de considérer un rayon adaptatif, de sorte, par exemple, que pour chaque $q \in \mathcal{D}$, les domaines associés à F_q et G_q (définis dans l'équation (4.54)), correspondent à un certain pourcentage de l'image.

Une seconde solution, plus simple, est de considérer plusieurs valeurs différentes de ces paramètres, et de fusionner les NFA correspondant en prenant le maximum, comme indiqué dans la section 3.1.1.3. Il faut noter que la dimension des descripteurs utilisés risque de beaucoup influer sur les rayons à prendre, les points étant de plus en plus isolé au fur et à mesure que la dimension augmente.

Complexité temporelle limitante Pour finir, il est important de prendre conscience d'une grande limitation du modèle de changements par destin commun, à savoir sa complexité temporelle. Il est en effet nécessaire, pour chaque pixel $q \in \mathcal{D}$, d'estimer une statistique globale au niveau de l'image, et d'envisager à chaque fois tous les sous-ensembles E possibles, et cela, pour chaque taille du contexte spatial E et chaque couple de rayons (η_u, η_v) choisi. De ce fait, nous nous limitons à seulement quelques valeurs des rayons η_u et η_v, que nous supposons tout deux égaux. Par ailleurs, cette complexité dépendant aussi beaucoup de la dimension des descripteurs utilisés, nous nous limiterons dans la pratique à des descripteurs de dimension au plus 3.

4.3 Modèle de changements par reconstruction indirecte

4.3.1 Introduction

Si le modèle de changements par destin commun apparaît comme bien robuste aux changements globaux, son temps d'exécution constitue une forte limitation. Par ailleurs, il est judicieux de disposer de plusieurs modèles robustes aux changements globaux, pour apporter plus d'expressivité aux détecteurs fusionnés, construits à partir des modèles primaires. Nous construisons donc un second modèle, toujours robuste aux changements globaux et pouvant entre autres s'appliquer sans problème à des descripteurs multidimensionnels en un temps raisonnable.

Pour cela, plutôt que d'évaluer la distribution des pixels dans l'ensemble de l'image, nous nous restreignons à l'étude des K plus proches voisins. En l'absence de changement, nous supposons que des structures similaires évoluent de la même manière entre les deux images. Cette approche repose sur la redondance des structures dans les images naturelles, un même motif se présentant souvent en de multiples endroits.

Utilisation de la redondance des structures L'utilisation de la redondance des structures dans les images naturelles est très utilisée en traitement d'image. L'approche de Zontak et Cohen [ZC10] est particulièrement intéressante. Il s'agit de résoudre un problème de détection d'anomalie dans une image en utilisant une image de référence, problème pouvant s'apparenter à de la détection de changements. Étant donnée une image sans anomalie J et une nouvelle image I, présentant toutes les deux un même motif, l'algorithme consiste tout d'abord à chercher, pour chaque patch de I, les K

plus plus proches patchs voisins dans J, pour $K \in \mathbb{N}^*$, puis de reconstruire le patch de I à partir de ses patchs voisins. Formellement, pour un patch $I_{[p]}$, de patchs voisins $J_{[p_i]}$, pour $i \in [\![1, K]\!]$, le patch reconstruit s'écrit sous la forme

$$\overline{I_{[p]}} = \frac{\sum_{i=1}^{n} w_{p,p_i} I_{[p_i]}}{\sum_{i=1}^{n} w_{p,p_i}} \qquad \text{où} \qquad w_{p,q} = \exp\left(-\frac{\left\|I_{[p]} - I_{[q]}\right\|_2^2}{h^2}\right), \tag{4.56}$$

pour h^2 un coefficient positif. En moyennant alors tous les patchs de I reconstruits de la sorte, on obtient une nouvelle image $\widetilde{I}$ sans anomalie et il est possible de détecter les anomalies en envisageant l'erreur de reconstruction $|I - \widetilde{I}|$.

Cet algorithme est repris par Davy et coauteur [DEMD18] dans le cas d'une unique image, en reconstruisant l'image à partir de ses propres patchs.

Nous adaptons cet algorithme à la détection de changements dans le cas d'une paire d'images radiométriquement différentes. Pour cela, nous utilisons la structure de similarité obtenue dans la première image pour reconstruire une image, présentant cette même structure mais utilisant la dynamique de la seconde image. Ce faisant, il nous est alors possible de comparer les deux structures de similarité au sein d'un même espace radiométrique.

4.3.2 Formalisation du modèle

4.3.2.1 Structure de similarité

Introduction Comme pour le modèle précédent, on se donne deux images u et v, définies sur un même domaine $\mathcal{D} \subset \mathbb{Z}^2$, que l'on suppose éventuellement différentes radiométriquement, dans le sens où il n'est pas nécessairement possible de comparer directement leurs intensités, et on note K_U et K_V leur nombre de canaux respectifs, avec ainsi

$$u : \mathcal{D} \to \mathbb{R}^{K_U} \qquad \text{et} \qquad v : \mathcal{D} \to \mathbb{R}^{K_V}, \tag{4.57}$$

que l'on pourra voir comme des vecteurs, respectivement dans $\mathbb{R}^{|\mathcal{D}| \times K_U}$ et $\mathbb{R}^{|\mathcal{D}| \times K_V}$.

Nous construisons ici une image $\widetilde{v}$, possédant les mêmes K_V canaux que v, mais capturant la structure de similarité interne de u. Cette image $\widetilde{v}$ permet ainsi d'exprimer la géométrie de u, mais dans l'espace radiométrique de v.

Patchs et restriction de patchs Dans l'ensemble de cette section, nous supposons fixée une forme $P_0 \subset \mathbb{Z}^2$ (dans la pratique, un carré centré en 0), et nous noterons $[p]$ le domaine pixellique, qualifié de patch, de forme P_0 centré au pixel $p \in \mathcal{D}$, c'est-à-dire

$$[p] = \{p + q,\ q \in P_0\}. \tag{4.58}$$

Notons bien que, comme précédemment et discuté dans la section 3.2.1, nous supposons qu'il n'existe pas de problèmes aux bords du domaine $\mathcal{D}$ et que chacune des images est aussi définie en dehors de $\mathcal{D}$ de sorte que l'on pourra, sans problème de définition, considérer les intensités des images sur les différents patchs $[p]$, pour tout $p \in \mathcal{D}$. Étant donné une image $X \in \mathbb{R}^{|\mathcal{D}| \times n}$, pour $n \in \mathbb{N}^*$, on notera alors $X_{[p]}$ le vecteur de $\mathbb{R}^{|P_0| \times n}$

correspondant à la restriction de X au patch $[p]$, et $(X_{[p]})_\tau$ sa valeur au pixel $\tau \in P_0$, c'est-à-dire, formellement,

$$X_{[p]} \;=\; (X_{p+\tau})_{\tau \in P_0}\,. \tag{4.59}$$

Structure de voisinage et mesure de similarité La structure de similarité interne d'une image u est envisagée au travers des voisinages des différents patchs. Étant donné un patch $[p]$, pour $p \in \mathcal{D}$, on note $\mathcal{N}_p^K(u)$ l'ensemble de ses K plus proches patchs voisins dans l'image u, en prenant comme distance de similarité la distance L^2, c'est-à-dire

$$\mathcal{N}_p^K(u) \;=\; \underset{\substack{S\subset\mathcal{D}\\ |S|=K}}{\operatorname{argmin}} \sum_{q\in S} \|u_{[p]} - u_{[q]}\|_2, \tag{4.60}$$

en notant $|S|$ le cardinal de l'ensemble S. À chaque couple $([p],[q])$ de patchs voisins, pour $p,q \in \mathcal{D}$, est alors associé un coefficient de similarité, noté $w_{p,q}(u)$, quantifiant la distance entre ces deux patchs, et qui sera utilisé pour la construction de l'image $\widetilde{v}$. Nous avons pris, de manière classique, une pondération gaussienne, en considérant

$$w_{p,q}(u) \;=\; \exp\!\left(-\frac{1}{h^2}\left\|u_{[p]} - u_{[q]}\right\|_2^2\right), \tag{4.61}$$

avec h^2 la variance du noyau utilisé, dont la valeur dépend en pratique de la variance du bruit dans l'image et de la taille du patch P_0.

On peut chercher à rajouter certaines contraintes de positionnement sur les patchs voisins. Il peut s'agir d'exclure les patchs voisins trop proches spatialement du patch considéré, en exigeant ainsi que

$$\forall [q] \in \mathcal{N}_p^K(u), \qquad \|p-q\|_2 \geq R_{\min}, \tag{4.62}$$

pour $R_{\min} > 0$ ou encore d'imposer que les voisins d'un même patch ne soient pas trop proches, c'est à dire

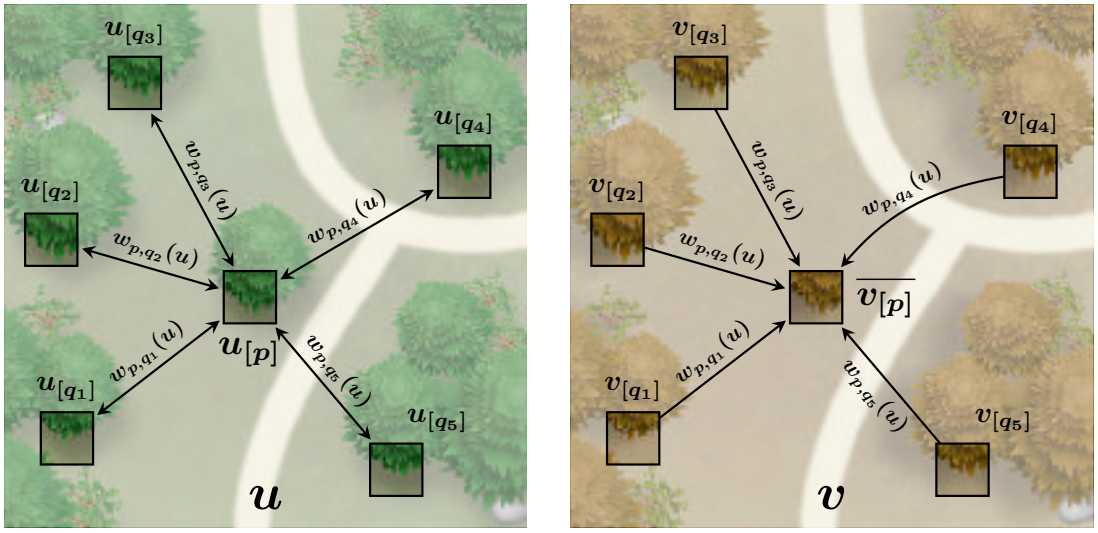
$$\forall [q],[q'] \in \mathcal{N}_p^K(u), \qquad q \neq q' \;\Rightarrow\; \|q-q'\|_2 \geq R'_{\min}, \tag{4.63}$$

pour $R'_{\min} > 0$. Notons que si la première contrainte peut facilement être prise en compte, la deuxième demanderait une étude supplémentaire.

4.3.2.2 Reconstruction d'image

Reconstruction d'un patch Étant donné la structure de similarité de l'image u, donnée par les $\mathcal{N}_p^K(u)$ pour chaque $p \in \mathcal{D}$ et par les coefficients de similarité $w_{p,q}$, pour $q \in \mathcal{N}_p^K(u)$, il s'agit alors de reconstruire une image $\widetilde{v}$ capturant cette structure, et ce, à partir des intensités de l'image v. Pour cela, il s'agit tout d'abord de reconstruire séparément chaque patch $[p]$ de v, séparément, en considérant

$$\overline{v_{[p]}} \;=\; \sum_{[q]\in\mathcal{N}_p^K(u)} \widetilde{w}_{p,q}(u)\, v_{[q]}, \tag{4.64}$$



Structure de similarité **Reconstruction du patch**

Figure 4.11 – Représentation du processus de reconstruction d'un patch $[p]$ en utilisant la structure de similarité de u et les patchs de v. Dans la première image, on représente la structure de similarité correspondant aux différents patchs $u_{[q_i]}$, voisins du patch central $u_{[p]}$, ainsi qu'aux coefficients de similarités $w_{p,q_i}(u)$ associé à chacun. Dans la seconde image, on représente le patch $\overline{v_{[p]}}$, reconstruit dans v à partir de la structure mesurée dans u.

où $\widetilde{w}_{p,q}(u)$ est un coefficient de pondération des patchs calculé à partir des $w_{p,q}(u)$, défini pour tout $p, q \in \mathcal{D}$ par

$$\widetilde{w}_{p,q}(u) \;=\; \frac{w_{p,q}(u)}{\sum\limits_{[q']\in\mathcal{N}_p^K(u)} w_{p,q'}(u)}. \tag{4.65}$$

Nous avons représenté schématiquement le processus de reconstruction d'un patch dans la figure 4.11.

Reconstruction par moyennage de patch Pour reconstruire l'image entière $\widetilde{v}$, il s'agit alors de fusionner les reconstructions des différents patchs reconstruits, $\overline{v_{[p]}}$. Pour ce faire, nous avons décidé d'agir en moyennant simplement les différents patchs. On reconstruit alors l'image entière $\widetilde{v}$ par moyennage des patchs, c'est-à-dire, pour $p \in \mathcal{D}$,

$$\widetilde{v}_p \;=\; \frac{1}{|P_0|}\sum_{\tau\in P_0}\left(\overline{v_{[p-\tau]}}\right)_\tau. \tag{4.66}$$

En utilisant la définition donnée dans l'équation (4.64), il est possible de trouver une expression de $\widetilde{v}$,

$$\widetilde{v}_p \;=\; \frac{1}{|P_0|}\sum_{\tau\in P_0}\sum_{[q]\in\mathcal{N}_{p-\tau}^K(u)} \widetilde{w}_{p-\tau,q}(u)\; v_{q+\tau}. \tag{4.67}$$

4.3.2.3 Erreur de reconstruction et modèle a contrario

Formalisme a contrario La transformation de l'image u en l'image $\widetilde{v}$ permet d'exprimer la structure de similarité de u dans l'espace radiométrique de v, et donc de comparer, d'une certaine manière, les structures de u et de v en comparant directement $\widetilde{v}$ et v.

Explicitons les caractéristiques de ce modèle a contrario, résumé dans la section 3.1.2.1. On envisage ici chaque pixel du domaine $\mathcal{D}$ comme une structure potentielle, en leur donnant une pondération uniforme, chacun d'eux ayant une importance a priori similaire.

Modélisation de la seconde image Pour modèle naïf, on considère que la structure de voisinage mesurée dans l'image u se retrouve dans l'image v. Plus formellement, on suppose que V s'écrit sous la forme

$$V \;=\; V^0 + \varepsilon, \tag{4.68}$$

où V^0 est une image aléatoire dépendant de u et capturant les variations radiométriques intervenant dans l'image V, et où ε prend en compte le bruit intrinsèque de l'image. Comme justifié dans la section 2.1.2, on modélise la distribution de ε par une loi gaussienne centrée et isotrope, de variance σ^2, c'est-à-dire que pour tout $p \in \mathcal{D}$,

$$\varepsilon_p \sim \mathcal{N}\Big(0, \sigma^2 I_{K_V}\Big), \tag{4.69}$$

où I_{K_V} est la matrice identité de dimension K_V. On considère alors que deux patchs voisins possèdent nécessairement la même distribution gaussienne, ce que nous écrirons, en utilisant V^0, de la manière suivante,

$$\forall p \in \mathcal{D}, \quad \forall q \in \mathcal{N}_p^K(u), \qquad V^0_{[p]} = V^0_{[q]}, \tag{4.70}$$

et que l'on peut aussi écrire sous la forme

$$\forall p \in \mathcal{D},\ \forall q \in \mathcal{N}_p^K(u),\ \forall \tau \in P_0, \qquad V_{q+\tau} - V_{p+\tau} = \varepsilon_{q+\tau} - \varepsilon_{p+\tau} \sim \mathcal{N}\Big(0, 2\sigma^2\Big). \tag{4.71}$$

La décomposition de V en une partie de bruit et une partie structurée permet de plus facilement prendre en compte les différentes corrélation existant entre les patchs. Considérer simplement que $V_{[p]} - V_{[q]}$ correspond à un bruit gaussien n'aurait pas fourni une compréhension claire de la corrélation entre deux patchs qui s'intersectent.

Modèle a contrario On peut alors montrer le résultat suivant.

Proposition 4.3.1. *En considérant $\widetilde{v}$ l'image reconstruite à partir de v et de la structure de voisinage de u, comme explicité dans l'équation (4.67), la fonction définie pour tout $p \in \mathcal{D}$ par*

$$\mathrm{NFA}(p, u, v) \;=\; \frac{|\mathcal{D}|}{\Gamma\Big(\frac{K_V}{2}\Big)} \Gamma\left(\frac{K_V}{2}, \frac{\|\widetilde{v}_p - v_p\|_2^2}{2\alpha_p \sigma^2}\right), \tag{4.72}$$

où

$$\alpha_p \;=\; 1 + \frac{1}{|P_0|^2} \sum_{s\in\mathcal{D}} \left(\sum_{\tau\in P_0} \sum_{[q]\in\mathcal{N}^K_{p-\tau}(u)} \widetilde{w}_{p-\tau,q}(u) \mathbb{1}_{\{q+\tau=s\}} \right)^2, \tag{4.73}$$

est un nombre de fausses alarmes pour le modèle naïf donné dans l'équation (4.71)*.*

Démonstration. Le théorème 3.1.2 relatif à la construction d'un nombre de fausses alarmes, appliqué à la mesure de structure

$$\mu(p,u,v) \;=\; \|\widetilde{v}_p - v_p\|_2\,, \tag{4.74}$$

nous assure que la fonction

$$F(p,u,V) \;=\; |\mathcal{D}|\,\mathbb{P}(\,\mu(p,U,V) \geq \mu(p,u,v)\,) \tag{4.75}$$

est bien un nombre de fausses alarmes. D'après la propriété 3.1.5 relatif à la majoration des NFA, il suffit alors de montrer que l'expression annoncée majore cette probabilité de fausse alarme. En reprenant l'équation (4.67), on peut écrire l'erreur de reconstruction sous la forme

$$\widetilde{V}_p - V_p \;=\; \frac{1}{|P_0|} \sum_{\tau\in P_0} \sum_{[q]\in\mathcal{N}^K_{p-\tau}(u)} \widetilde{w}_{p-\tau,q}(u)\;\left(V_p - V_{q+\tau}\right). \tag{4.76}$$

À ce stade, remarquons qu'il n'est pas possible que $q+\tau$ soit égal à p lorsque $[q]$ est un patch voisin de $[p-\tau]$. En effet, si $p = q+\tau$, alors $q = p-\tau$, ce qui n'est pas possible, les K patchs voisins de $[p-\tau]$ ne pouvant être égaux à $[p-\tau]$.

En utilisant le modèle naïf, donnée équation (4.71), on trouve

$$\begin{aligned} \widetilde{V}_p - V_p \;&=\; \frac{1}{|P_0|} \sum_{\tau\in P_0} \sum_{[q]\in\mathcal{N}^K_{p-\tau}(u)} \widetilde{w}_{p-\tau,q}(u)\;\left(\varepsilon_p - \varepsilon_{q+\tau}\right) \\ &=\; \varepsilon_p - \frac{1}{|P_0|} \sum_{\tau\in P_0} \sum_{[q]\in\mathcal{N}^K_{p-\tau}(u)} \widetilde{w}_{p-\tau,q}(u)\,\varepsilon_{q+\tau}. \end{aligned} \tag{4.77}$$

La variable ε étant supposée suivre une distribution gaussienne de moyenne nulle, il apparaît directement que l'erreur de reconstruction $\widetilde{V}_p - V_p$ suit aussi une distribution gaussienne centrée. Les différents ε_q, pour $q \in \mathcal{D}$ étant indépendants, on peut calculer la variance σ_p^2 de cette mesure en les regroupant par valeur de $q+\tau$. En écrivant

$$\widetilde{V}_p - V_p \;=\; \varepsilon_p - \frac{1}{|P_0|} \sum_{s\in\mathcal{D}} \left(\sum_{\tau\in P_0} \sum_{[q]\in\mathcal{N}^K_{p-\tau}(u)} \widetilde{w}_{p-\tau,q}(u) \mathbb{1}_{\{q+\tau=s\}} \right) \varepsilon_s, \tag{4.78}$$

et en notant à nouveau que le facteur est nul pour $s = p$, on trouve

$$\sigma_p^2 \;=\; \sigma^2 + \frac{1}{|P_0|^2} \sum_{s\in\mathcal{D}} \left(\sum_{\tau\in P_0} \sum_{[q]\in\mathcal{N}^K_{p-\tau}(u)} \widetilde{w}_{p-\tau,q}(u) \mathbb{1}_{\{q+\tau=s\}} \right)^2 \sigma^2. \tag{4.79}$$

On peut alors appliquer le théorème 3.4.2 qui nous donne, pour tout $\delta > 0$,

$$\mathbb{P}\Big(\left\| \widetilde{V}_p - V_p \right\|_2 \geq \delta \Big) \;=\; \frac{|\mathcal{D}|}{\Gamma\left(\frac{K_V}{2}\right)} \Gamma\left(\frac{K_V}{2}, \frac{\delta^2}{2\sigma_p^2} \right). \tag{4.80}$$

On retrouve bien le résultat annoncé, avec $\alpha_p = \sigma_p / \sigma$. □

Encadrement de la variance On peut montrer qu'il existe une expression simplifiée, en majorant la variance de l'erreur de reconstruction.

Corollaire 4.3.2. *En reprenant les notations et le modèle naïf de la proposition 4.3.1, la fonction définie pour tout $p \in \mathcal{D}$ par*

$$\mathrm{NFA}(p, u, v) \;=\; \frac{|\mathcal{D}|}{\Gamma\left(\frac{K_V}{2}\right)} \Gamma\left(\frac{K_V}{2}, \frac{\|\widetilde{v}_p - v_p\|_2^2}{4\sigma^2} \right), \tag{4.81}$$

est un nombre de fausses alarmes.

Démonstration. En reprenant les notations de la proposition précédente, on a

$$\alpha_p \;=\; 1 + \frac{1}{|P_0|^2} \sum_{s \in \mathcal{D}} \left(\sum_{\tau \in P_0} \sum_{[q] \in \mathcal{N}^K_{p-\tau}(u)} \widetilde{w}_{p-\tau,q}(u) \mathbb{1}_{\{q+\tau=s\}} \right)^2 .$$

Posons, pour tout $s \in \mathcal{D}$,

$$c_s \;=\; \frac{1}{|P_0|} \sum_{\tau \in P_0} \sum_{[q] \in \mathcal{N}^K_{p-\tau}(u)} \widetilde{w}_{p-\tau,q}(u) \mathbb{1}_{\{q+\tau=s\}}$$

et remarquons que tous les c_s sont positifs et se somme à 1. On a donc

$$0 \;\leq\; \sum_{s \in \mathcal{D}} c_s^2 \;\leq\; \left(\sum_{s \in \mathcal{D}} c_s \right)^2 \;=\; 1, \tag{4.82}$$

par convexité de la fonction carrée. On trouve donc directement l'encadrement suivant

$$\sigma^2 \;\leq\; \sigma_p^2 \;\leq\; 2\sigma^2. \tag{4.83}$$

La fonction gamma incomplète supérieure, $\Gamma(n, x)$, étant décroissante en fonction de x, on trouve

$$\Gamma\left(\frac{K_V}{2}, \frac{\|\widetilde{v}_p - v_p\|_2^2}{2\sigma_p^2} \right) \;\leq\; \Gamma\left(\frac{K_V}{2}, \frac{\|\widetilde{v}_p - v_p\|_2^2}{2 \times 2\sigma^2} \right),$$

et en utilisant la propriété 3.1.5, sur la majoration des NFA, on obtient bien le résultat annoncé. □

Encadrement de l'erreur de majoration La preuve du corollaire précédent donne en réalité un encadrement de la variance, et non seulement une majoration. L'existence d'une minoration est intéressante puisqu'elle permet de limiter, dans une certaine mesure, l'erreur de majoration. Puisque σ_p^2 est encadrée par σ^2 et $2\sigma^2$, on est assuré de ne pas surestimer la variance de plus du double de sa valeur réelle, lorsque l'on utilise l'expression donnée dans le corollaire 4.3.2.

Asymétrie du détecteur et symétrisation De la même manière que précédemment, on dispose ici d'un modèle asymétrique, donnant aux deux images deux rôles différents, une étant utilisée pour extraire la structure de similarité, l'autre étant reconstruite. Il est toutefois possible de symétriser ce modèle en considérant

$$\mathrm{NFA}_{\mathrm{sym}}(p,u,v) \;=\; 2\min\Big(\mathrm{NFA}(p,u,v),\mathrm{NFA}(p,v,u)\Big)\,, \tag{4.84}$$

qui est bien un nombre de fausses alarmes, d'après la propriété 3.1.7.

4.3.2.4 Extensions du modèle par biais de reconstruction

Pondération du patch central Pour que ce modèle puisse fonctionner convenablement, il est nécessaire que les structures présentes dans chaque patch apparaissent en plusieurs endroits dans l'image. Sans cela, certains patchs, que l'on pourrait qualifier d'anormaux, n'auraient aucun voisin proche et seraient donc automatiquement mal reconstruits. Pour éviter de détecter comme changement toutes les structures anormales de u, il est possible de biaiser la reconstruction, en rajoutant chaque patch comme son propre voisin, pondéré par une constante $\lambda > 0$. Chaque patch est alors reconstruit avec, pour tout $p \in \mathcal{D}$,

$$\widetilde{v}_p^{\lambda} \;=\; \frac{1}{|P_0|}\sum_{\tau\in P_0}\Big(\overline{v_{[p-\tau]}}^{\lambda}\Big)_{\tau} \qquad \text{où} \qquad \overline{v_{[p]}}^{\lambda} \;=\; \frac{\lambda v_{[p]} + \sum\limits_{[q]\in\mathcal{N}_p(u)} w_{p,q}(u)v_{[q]}}{\lambda + \sum\limits_{[q]\in\mathcal{N}_p(u)} w_{p,q}(u)}. \tag{4.85}$$

Proposition 4.3.3. *En considérant $\widetilde{v}^{\lambda}$, l'image reconstruite à partir de v et de la structure de voisinage de u, en utilisant la pondération du patch central avec un biais $\lambda > 0$, comme explicité dans l'équation (4.85), alors la fonction définie pour tout $p \in \mathcal{D}$ par*

$$\mathrm{NFA}(p,u,v) \;=\; \frac{|\mathcal{D}|}{\Gamma\Big(\frac{K_V}{2}\Big)}\Gamma\left(\frac{K_V}{2},\frac{\left\|\widetilde{v}_p^{\lambda}-v_p\right\|_2^2}{4\sigma^2}\right), \tag{4.86}$$

est un nombre de fausses alarmes pour le modèle naïf donné dans l'équation (4.71).

Démonstration. En posant, pour tout $p,q \in \mathcal{D}$,

$$\widetilde{w}_{p,q}^{\lambda}(u) \;=\; \frac{w_{p,q}(u)}{\lambda + \sum\limits_{[q']\in\mathcal{N}_p^K(u)} w_{p,q'}(u)} \qquad \text{et} \qquad \widetilde{\lambda}_p \;=\; \frac{\lambda}{\lambda + \sum\limits_{[q']\in\mathcal{N}_p^K(u)} w_{p,q'}(u)}, \tag{4.87}$$

on peut écrire

$$\widetilde{V}_p^\lambda \;=\; \frac{1}{|P_0|}\sum_{\tau\in P_0}\left(\widetilde{\lambda}\,V_p + \sum_{[q]\in\mathcal{N}_{p-\tau}^K(u)} \widetilde{w}_{p-\tau,q}^\lambda(u)\,V_{q+\tau}\right). \tag{4.88}$$

On trouve alors des expressions très similaires à celles données dans les équations (4.76) et (4.78) :

$$\begin{aligned} V_p - \widetilde{V}_p^\lambda \;&=\; \frac{1}{|P_0|}\sum_{\tau\in P_0}\sum_{[q]\in\mathcal{N}_{p-\tau}^K(u)} \widetilde{w}_{p-\tau,q}^\lambda(u)\,(V_p - V_{q+\tau}) \\ &=\; \frac{1}{|P_0|}\sum_{\tau\in P_0}\sum_{[q]\in\mathcal{N}_{p-\tau}^K(u)} \widetilde{w}_{p-\tau,q}^\lambda(u)\,(\varepsilon_p - \varepsilon_{q+\tau}) \\ &=\; \gamma_p\varepsilon_p - \frac{1}{|P_0|^2}\sum_{s\in\mathcal{D}}\left(\sum_{\tau\in P_0}\sum_{[q]\in\mathcal{N}_{p-\tau}^K(u)} \widetilde{w}_{p-\tau,q}^\lambda(u)\mathbb{1}_{\{q+\tau=s\}}\right)\varepsilon_s, \end{aligned} \tag{4.89}$$

où

$$\gamma_p \;=\; \frac{1}{|P_0|}\sum_{\tau\in P_0}\sum_{[q]\in\mathcal{N}_{p-\tau}^K(u)} \widetilde{w}_{p-\tau,q}^\lambda(u). \tag{4.90}$$

La seule différence avec le cas précédent est que le coefficient γ_p n'est plus égal à 1. On en conclu donc, comme précédemment, que l'erreur de reconstruction suit une distribution gaussienne centrée et de variance σ_p^2 avec

$$\sigma_p^2 \;=\; \gamma_p^2\sigma^2 + \frac{1}{|P_0|^2}\sum_{s\in\mathcal{D}}\left(\sum_{\tau\in P_0}\sum_{[q]\in\mathcal{N}_{p-\tau}^K(u)} \widetilde{w}_{p-\tau,q}^\lambda(u)\mathbb{1}_{\{q+\tau=s\}}\right)^2\sigma^2. \tag{4.91}$$

Comme précédemment, on peut montrer l'encadrement suivant,

$$\gamma_p^2\sigma^2 \;\le\; \sigma_p^2 \;\le\; 2\gamma_p^2\sigma^2 \;\le\; 2\sigma^2, \tag{4.92}$$

puisque $\gamma_p \le 1$, et obtenir l'expression annoncée. □

Il est important de noter que, contrairement au cas non biaisé, il n'existe plus ici de minoration permettant de limiter l'erreur de majoration, la valeur $\gamma_p^2\sigma^2$ pouvant être arbitrairement proche de 0. Ainsi, si les propriétés de nombre de fausses alarmes restent vérifiés, on aura ici tendance à sous-estimer les changements au niveau de pixels présentant une valeur faible de γ_p. Toutefois, cette dernière étant d'autant plus proche de 0 que le patch $[p]$ est loin de ses voisins, une telle majoration aura alors tendance à amoindrir l'importance des changements au niveau des patchs « anormaux », ce qui peut sembler acceptable ici.

Prise en compte des décalages spatiaux des patchs voisins Comme nous l'avons signalé précédemment, la plupart des paires d'images satellites très haute résolution souffrent d'incohérences spatiales, provoquant des décalages spatiaux entre les deux images. Or, le modèle de détection de changements issus de la proposition 4.3.1 est sensible aux décalages spatiaux, le processus de reconstruction supposant qu'un patch corresponde au même emplacement dans les deux images.

Pour remédier à ce problème, il est possible de biaiser la reconstruction des patchs en prenant en compte une éventuelle déformation légère entre les deux images. En reprenant les notations précédentes, on définit la reconstruction robuste aux décalages spatiaux

$$\widetilde{v}_p^{\tau_m} = \frac{1}{|P_0|} \sum_{\tau \in P_0} \left(\overline{v_{[p-\tau]}}^{\tau_m} \right)_\tau \tag{4.93}$$

avec

$$\overline{v_{[p]}}^{\tau_m} = \sum_{[q] \in \mathcal{N}_p(u)} \widetilde{w}_{p,q}(u)\, v_{[q+t_q^*]} \qquad \text{où} \qquad t_q^* = \underset{\|t\| \leq 2\tau_m}{\operatorname{argmin}} \left\| v_{[p]} - v_{[q+t]} \right\|_2, \tag{4.94}$$

où τ_m est un paramètre entier représentant l'amplitude maximale du décalage spatial envisagé, les $\widetilde{w}_{p,q}$ restant inchangés,

$$\widetilde{w}_{p,q}(u) = \frac{w_{p,q}(u)}{\sum\limits_{[q'] \in \mathcal{N}_p^K(u)} w_{p,q'}(u)}. \tag{4.95}$$

Nous commenterons plus tard le $2\tau_m$ présent dans la définition de t_q^*.

Remarque 4.3.4. *L'image $\widetilde{v}^{\tau_m}$ reconstruite en utilisant la robustesse aux décalages spatiaux des patchs voisins ne permet pas d'assurer que la fonction*

$$F(p,u,v) = \frac{|\mathcal{D}|}{\Gamma\left(\frac{K_V}{2}\right)} \Gamma\left(\frac{K_V}{2}, \frac{\|\widetilde{v}_p^{\tau_m} - v_p\|_2^2}{4\sigma^2} \right) \tag{4.96}$$

est un nombre de fausses alarmes. En effet, si la robustesse utilisée biaise la reconstruction des patchs $\overline{V_{[p]}}^{\tau_m}$ vers $V_{[p]}$, avec pour tout $p \in \mathcal{D}$,

$$\left\| \overline{V_{[p]}}^{\tau_m} - V_{[p]} \right\|_2 \leq \left\| \overline{V_{[p]}} - V_{[p]} \right\|_2, \tag{4.97}$$

rien n'assure en réalité que l'image reconstruite $\widetilde{V}_p^{\tau_m}$ est plus proche de V_p que ne l'est $\widetilde{V}_p$.

Paramètre et amplitude du décalage Revenons sur le terme $2\tau_m$ présent dans l'équation (4.94). Pour l'expliquer plus facilement, considérons $T : \mathcal{D} \to \mathcal{D}$, un champ de translation s'appliquant existant entre u et v, de sorte que qu'un pixel p dans u correspond au pixel $p+T(p)$ dans v. Supposons que l'amplitude du champ T est bornée par une constante τ_m, dans le sens où $\|T(p)\| \leq \tau_m$ pour tout $p \in \mathcal{D}$. Considérons de plus S le champs de translation inverse, un pixel p dans v correspondant à $S(p)$ dans u.

On s'intéresse ici à la reconstruction d'un patch $v_{[p]}$, pour $p \in \mathcal{D}$. Pour retrouver les voisins de $[p]$ dans v, il faut tout d'abord considérer le point $S[p]$ correspondant à p dans u, puis retrouver, pour chacun de ses patchs voisins (dans u), leur nouvelle position dans v. On peut ainsi écrire, en notant $\mathcal{N}_p^K(v)$ la structure de voisinage qu'il faudrait utiliser dans v,

$$\mathcal{N}_p^K(v) \;=\; T\Big(\mathcal{N}_{S(p)}^K(u)\Big) \;=\; \Big\{T(q),\, q \in \mathcal{N}_{S(p)}^K(u)\Big\}. \tag{4.98}$$

Il apparaît une double incertitude d'au plus τ_m sur la position des patchs à utiliser pour reconstruire $v_{[p]}$.

Si l'on suppose de plus que les patchs voisins de $[S(p)]$ sont relativement proches de ceux de $[p]$, c'est à dire que

$$\forall [q] \in \mathcal{N}_{S(p)}^K(u), \quad \exists \tau \in [-\tau_m, \tau_m]^2, \quad [q+\tau] \in \mathcal{N}_p^K(u), \tag{4.99}$$

on est alors certains que les patchs de $\mathcal{N}_p^K(v)$ sont situés à une distance d'au plus $2\tau_m$ des patchs de $\mathcal{N}_p^K(u)$, ce qui explique le terme $2\tau_m$ utilisé dans l'équation (4.94).

Notons que si rien n'assure cette dernière supposition, en considérant des tailles de patchs grandes devant τ_m, on s'attend tout de même à ce qu'elle soit vérifiée pour la majeur partie des pixels.

Utilisation conjointe des deux biais Il est naturellement possible d'utiliser ces deux biais en même temps, pour limiter la détection des structures isolées ainsi que l'influence des décalages spatiaux entre les deux images étudiées. Nous avons représenté le processus de reconstruction obtenu dans la figure 4.12.

4.3.3 Résultats et limitations du modèle

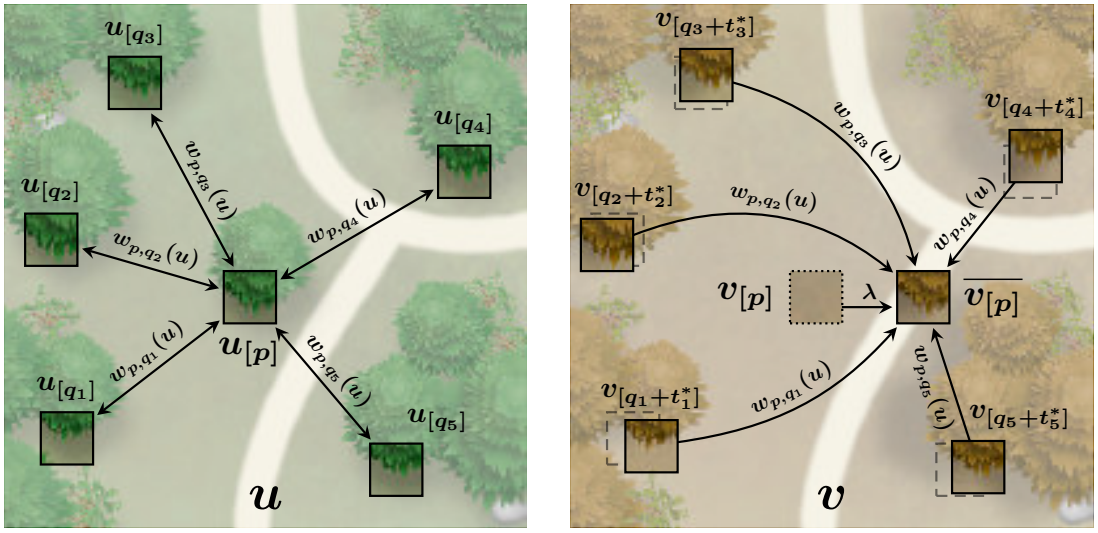
4.3.3.1 Résultats sur les exemples synthétiques

Retour sur les paramètres du modèles Avant toute chose, revenons rapidement sur les différents paramètres que possède le modèle, qui sont

- les descripteurs utilisés pour représenter les images, que l'on pourrait supposer être différents chacune des images, mais que nous prendrons ici égaux ;
- P_0, la forme des patchs utilisés ;
- K, le nombre de patchs voisins dans la structure de similarité ;
- h^2, la variance du noyau gaussien utilisé pour calculer les coefficients de similarité, présenté dans l'équation (4.61) ;
- τ_m, la translation maximale prise en compte entre les deux images ;
- λ, le coefficient du biais vers le patch central, utilisé pour limiter les détections des structures isolées.

Pour ne pas trop compliquer l'étude de ce modèle, nous ne considérons les images qu'en nuance de gris. Nous fixons la valeur de h^2 à 10^2 (les images synthétiques prenant des valeurs entières entre 0 et 255), un nombre $K = 5$ de plus proches voisins, et une forme des patchs P_0 carrée de taille 7×7.

Nous avons affiché, dans la figure 4.13, les résultats obtenus sans rajouter de biais (avec donc $\lambda = 0$ et $\tau_m = 0$).



Structure de similarité **Reconstruction biaisée du patch**

Figure 4.12 – Représentation du processus de reconstruction biaisée d'un patch $[p]$ à partir de la structure de similarité de u et des patchs de v. Ce schéma reprend celui de la figure 4.11 en y rajoutant les deux biais présentés dans la section 4.3.2.4. Ces deux biais correspondent d'une part à la pondération du patch central, le paramètre λ permettant de limiter la détection des structures isolées, et d'autre part à la prise en compte des décalages légers entre les deux images, au moyen du calcul des translations optimales de patch t_i^*.

Remarques générales et flou de reconstruction Tout d'abord, remarquons que l'on détecte bien toutes les apparitions et disparitions d'objets, et ce, à l'exception du camion sombre de la première image dont on ne détecte que l'ombre. Remarquons que, dans les images reconstruites $\widetilde{v}$ et $\widetilde{u}$, certains objets ou bords d'objets sont reconstruits flous. C'est principalement le cas du bateau dans la paire 1, des bords des nénuphars de la paire 2, des bords des palmiers de la paire 4, ainsi que des bords de la zone sombre de la paire 5. Ce flou peut s'expliquer de plusieurs manières :

— Dans le cas des nénuphars, ce flou s'explique simplement par le décalage local présent entre les deux images, la reconstruction moyennant des bords qui ne se correspondent pas.

— Dans le cas du bateau, il peut s'expliquer par un manque de patchs réellement similaires à ceux qui le composent. Il est alors impossible de reconstruire cette structure « unique » jusque dans les détails. C'est également le cas pour la paire 5, où certains champs présentent des formes sous-représentées et dont les bords sont donc mal reconstruits.

— Dans le cas des palmiers, en plus de la faible taille du domaine qui ne fournit certainement pas assez de patchs pour prendre en compte la grande variabilité des bords, le problème vient de la variation de texture de entre les deux images. Ce changement contredit en effet l'hypothèse de stabilité du champ de voisinage entre les deux images. On moyenne des patchs sur l'ensemble de la texture, ce

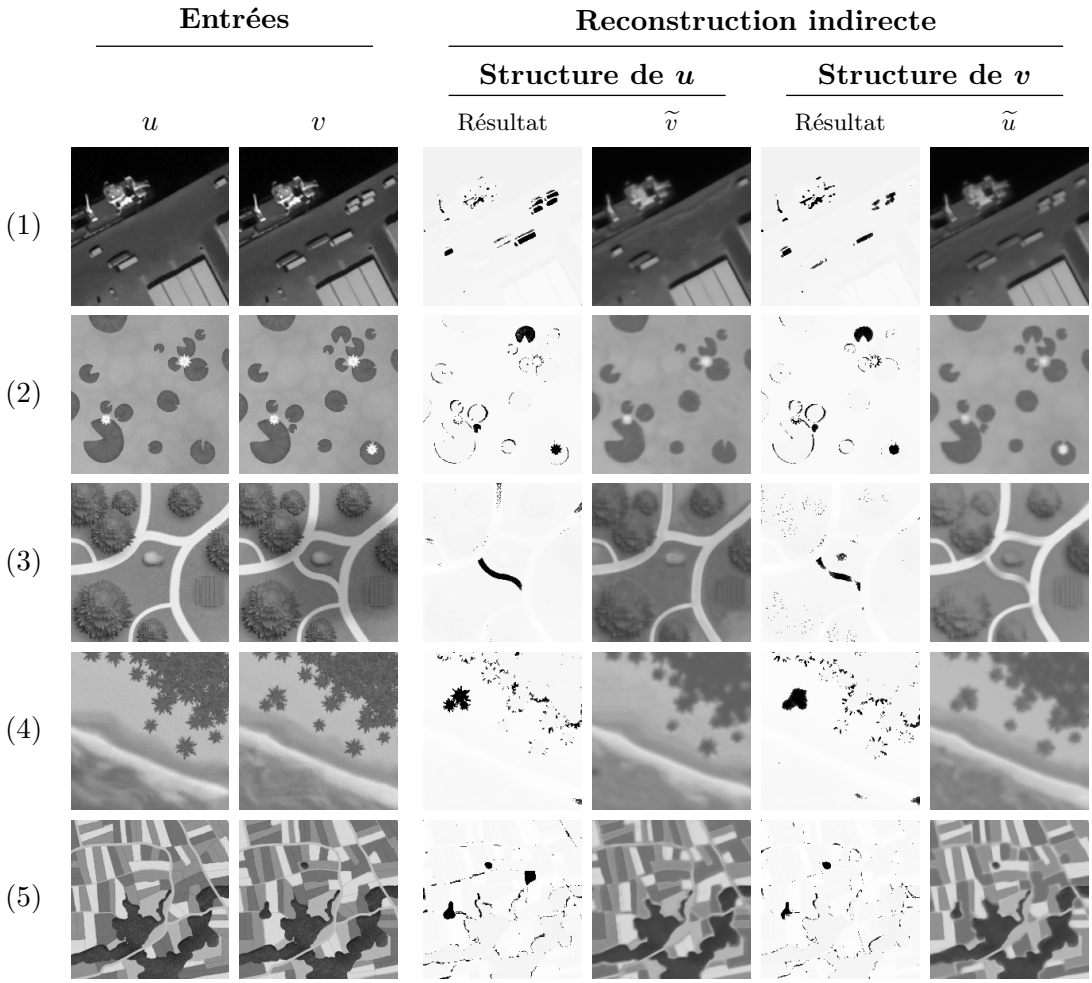


Figure 4.13 – Résultats du modèle de détection de changements par reconstruction indirecte sur les exemples synthétiques présentés dans la figure 2.9. Nous avons pris ici les images en nuance de gris, une variance du noyau gaussien $h^2 = 10^2$, des patchs de taille 7×7, ainsi qu'un nombre $K = 5$ de plus proches voisins pour chaque patch, sans rajouter de biais lors de la reconstruction (avec donc $\lambda = 0$ et $\tau_m = 0$). Chaque ligne correspond à un exemple synthétique et présente successivement la paire d'image donnée en entrée de l'algorithme, suivie des résultats obtenus en extrayant tout d'abord la structure de u et en reconstruisant v, puis inversement en reconstruisant u à partir de celle de v. Dans chaque cas, nous avons affiché le masque de détection ainsi que l'image reconstruite, $\widetilde{v}$ ou $\widetilde{u}$ en fonction de l'image de référence. Dans les images de résultat, nous avons représenté en noir les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$, le tout affiché sur une version éclaircie de la seconde image. On détecte bien ici toutes les apparitions et disparitions d'objets, à l'exception du champ évoluant de manière isolé lorsque l'on reconstruit u à partir de la structure de v (fait que l'on commente par la suite). On détecte cependant aussi les éléments « anormaux », comme le bateau dans la paire 1. Par ailleurs, on peut observer ici que les images reconstruites sont un peu floues, ce qui est particulièrement visible dans la paire 4.

qui aboutit à une reconstruction floue, principalement visible et détectée, à proximité des bords.

Ces différentes causes du flou correspondent aux raisons pour lesquelles nous proposons de biaiser la reconstruction, en rajoutant une pondération du patch central et une robustesse au décalage spatial.

Intérêt de la robustesse à la translation Intéressons-nous aux effets de la robustesse à la translation sur les résultats. Nous avons appliqué, dans la figure 4.14, ce modèle sur les exemples synthétiques, avec différentes valeurs de la translation maximale τ_m prise en compte. En augmentant la valeur de τ_m, on observe une disparition rapide des erreurs dues aux translations de la paire 2, ainsi que la disparition plus lente des erreurs dues aux variations de texture des bords de la paire 4, la robustesse à la translation permettant d'aller chercher des patchs plus proches au sein du nouveau tirage de la texture. D'un autre côté, les erreurs dues aux objets sous-représentés semblent, être conservées malgré une augmentation forte de la valeur de τ_m, comme nous pouvons l'observer dans les paires 1 et 5.

Notons qu'il est préférable de ne pas considérer une valeur de τ_m dépassant la largeur des patchs utilisés pour estimer la structure de voisinage. Puisque la notion de voisinage n'est évaluée que localement, sur un voisinage P_0, dès lors que l'on autorise une translation maximale des patchs voisins (égale à $2\tau_m$) supérieure à la largeur de P_0, on permet alors, lors de la reconstruction du patch central, d'utiliser des patchs n'intersectant pas le patch voisin initial et n'ayant a priori aucun lien avec ce dernier. Pour finir, notons que le temps de calcul devient vite très grand lorsque τ_m augmente. Dans la pratique, on se limitera ainsi à des valeurs de τ_m faible, inférieures à 2.

Intérêt de la pondération du patch central Examinons les effets de la pondération du patch central sur les résultats. Nous avons affiché ces derniers dans la figure 4.15 en faisant varier la valeur du biais λ. Comme attendu, et même pour des valeurs faibles de λ, le biais rajouté ici permet de limiter les détections des structures uniques ou sous-représentées dans l'image. Ainsi, dès la valeur $\lambda = 0{,}01$, on ne détecte plus les structures du bateau de la paire 1, ni les bords de la zone sombre de la paire 5.

Remarquons par ailleurs que ce biais masque les détections dues aux décalages ainsi qu'à la variation de texture dans les paires 2 et 4. Notons que ce masquage cesserait a priori dans des images plus grandes. En effet, la valeur relative du biais ne dépendant que de la structure de similarité de l'image de référence, c'est à cause de la légère sous-représentation des structures correspondant à ces translations et variation de texture qu'elles disparaissent rapidement, une sous-représentation qui n'aurait certainement pas existé si l'image avait été plus grande.

Notons pour finir que la valeur choisie pour le paramètre λ dépend fortement du nombre K de voisins considérés, ainsi que de la variance h^2 du noyau gaussien utilisé pour estimer les coefficients de similarités $w_{p,q}$. Dans la pratique, pour les valeurs $K = 5$ et $h^2 = 100$, nous prenons $\lambda = 10^{-1}$.

Le champ non détecté Dans la paire 5, en prenant comme référence l'image v, il peut sembler étonnant de ne pas détecter le champ présentant une évolution anormale,

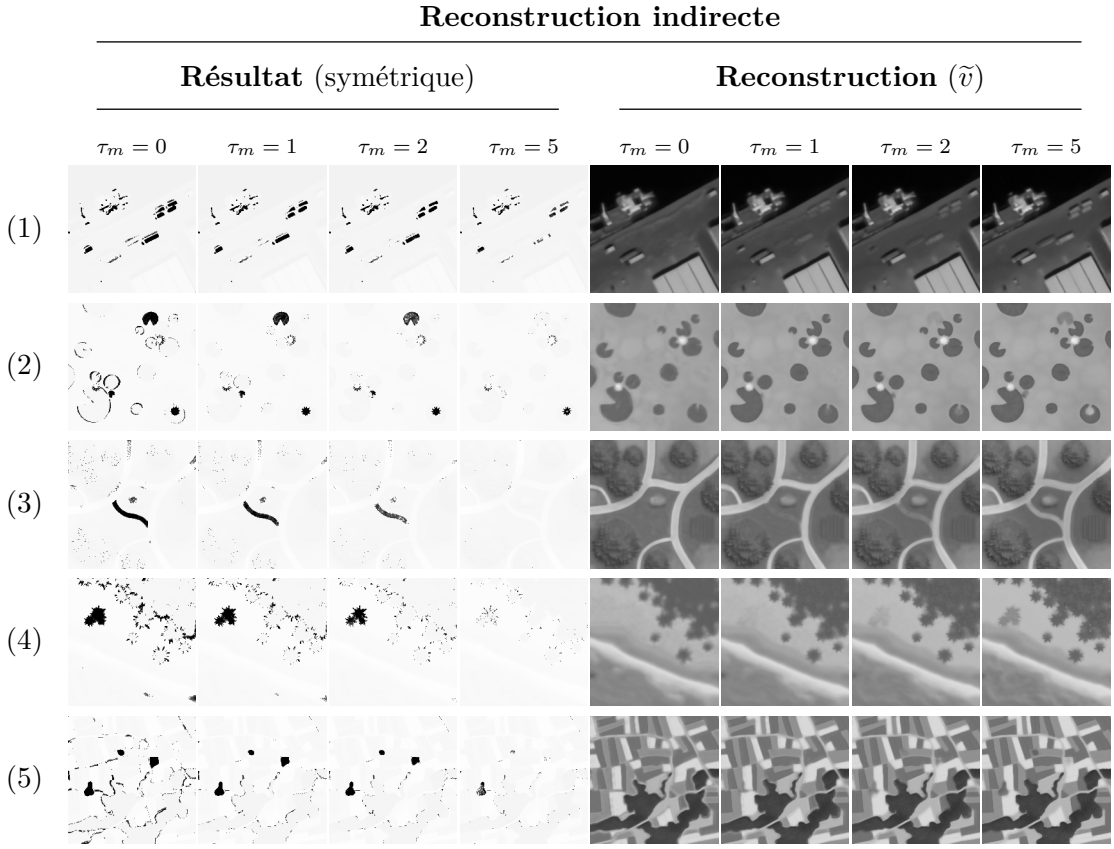


Figure 4.14 – Influence de la robustesse aux décalages locaux sur les résultats du modèle de détection de changements par reconstruction indirecte sur les exemples synthétiques présentés dans la figure 2.9. Nous reprenons ici les mêmes paramètres que ceux utilisés dans la figure 4.13, en faisant varier la valeur de τ_m. Chaque ligne correspond à un exemple synthétique pour différentes valeurs de la translation maximale τ_m prise en compte. On présente dans un premier temps le résultat d'une version symétrisée de l'algorithme, obtenue en prenant le maximum des NFA, considérant l'une ou l'autre des images comme référence. On présente ensuite la reconstruction obtenu en extrayant la structure de la première image, u, et en reconstruisant $\widetilde{v}$. Cette reconstruction permet de bien voir à quel point cette robustesse à la translation induit un biais non négligeable vers l'image v. La majorité des objets apparaissant dans la seconde image semblent en effet réapparaître en transparence dans l'image reconstruite lorsque τ_m augmente, et ce bien qu'étant absent dans la structure de u. Notons que la dernière valeur, $\tau_m = 5$, est délibérément prise trop grande vis-à-vis de la taille des patchs (voir section 4.3.3.1). Dans les images de résultat, nous avons représenté en noir les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$, le tout affiché sur une version éclaircie de la seconde image.

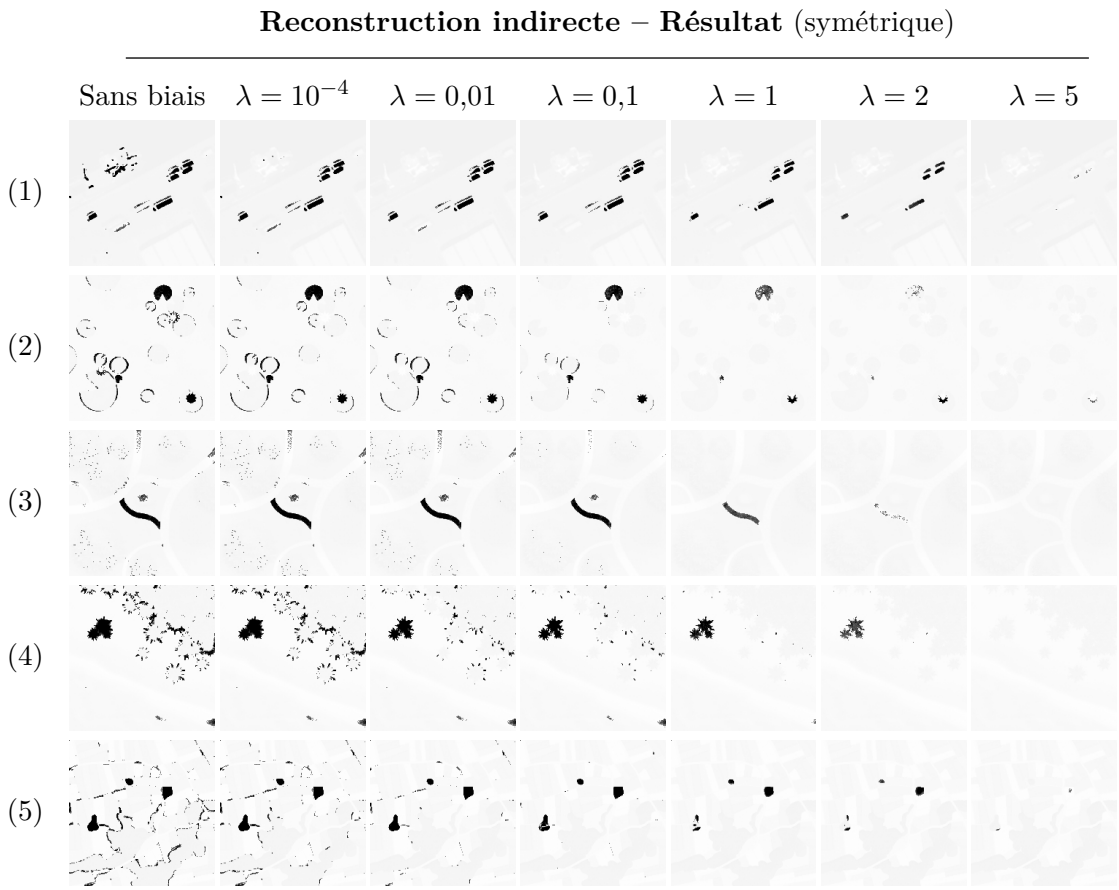


Figure 4.15 – Influence du biais λ sur les résultats du modèle de détection de changements par reconstruction indirecte sur les exemples synthétiques présentés dans la figure 2.9. Nous reprenons ici les mêmes paramètres que ceux utilisés dans la figure 4.13, en faisant varier la valeur de λ. Chaque ligne correspond à un exemple synthétique et présente les résultats d'une version symétrisée de l'algorithme, obtenue en prenant le maximum des NFA, considérant l'une ou l'autre des images comme référence, et ce pour différentes valeurs du biais λ. Il apparaît, comme attendu, que l'augmentation du paramètre λ permet ici de supprimer les détections au niveau des structures « uniques » ou sous-représentées dans l'image. Remarquons qu'il est normal perdre toutes les détections lorsque λ dépasse l'ordre de grandeur du nombre de voisins. Les $w_{p,q}$ de l'équation (4.85) étant tous inférieurs à 1, l'importance du biais serait alors supérieur à celle de la structure de similarité. Dans les images de résultat, nous avons représenté en noir les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$, le tout affiché sur une version éclaircie de la seconde image.

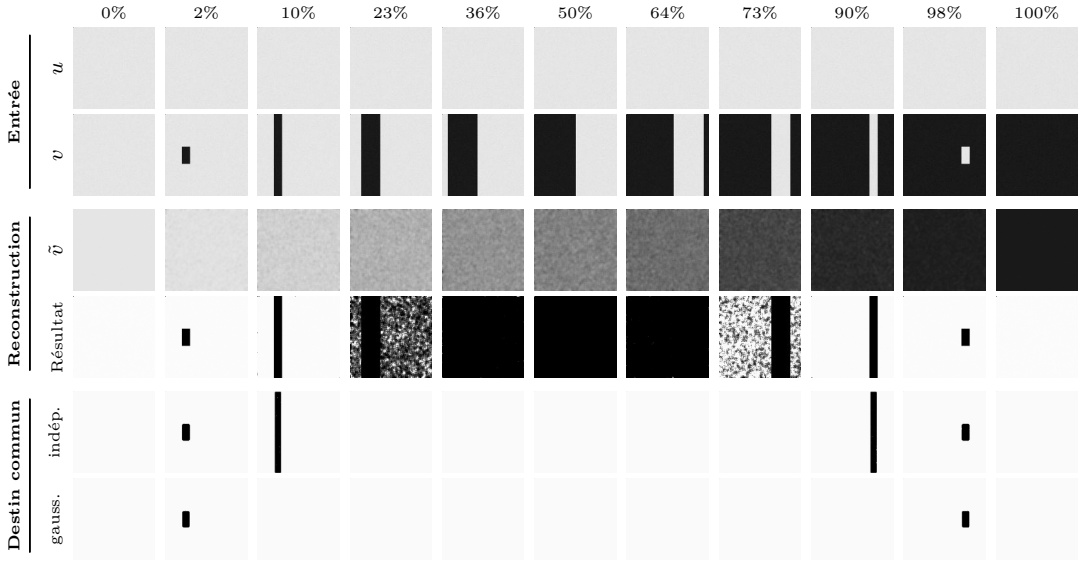


Figure 4.16 – Influence de la taille d'un changement sur les résultats des modèles de détection de changements par reconstruction indirecte et par destin commun, dans le cadre d'une apparition sur un fond homogène légèrement bruité. Chaque colonne correspond à une taille différente du changement, affichée en pourcentage de l'image, et présente successivement les deux images de la paire, le résultat du modèle par reconstruction indirecte, en montrant l'image reconstruite et le résultat final (utilisant les mêmes paramètres que dans la figure 4.13), ainsi que le résultat du modèle par destin commun, pour les deux modélisations, indépendante et gaussienne (utilisant les mêmes paramètres que dans la figure 4.8). Dans les images de résultat, nous avons représenté en noir les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$. Les deux modèles parviennent tout deux à détecter les changements locaux (de ratio proche de 0%) sans détecter les changements globaux (de ratio proche de 100%). Ils se distinguent toutefois pour les changements ni locaux, ni globaux, le modèle par reconstruction indirecte aboutissant à une détection complète de l'image, contrairement au modèle par destin commun qui ne détecte rien.

ce dernier apparaissant bien sur l'image reconstruite. L'algorithme reconstruit bien un champ différent (comme on peut le voir dans la figure 4.13), toutefois, la variation d'intensité existant entre l'image u et sa version reconstruite $\widetilde{u}$ s'avère ici trop faible par rapport au niveau de bruit utilisé dans les images et ne suffit pas à provoquer une détection.

4.3.3.2 Limitations du modèle

La plus grosse limitation de ce modèle peut être observée dans la figure 4.16. Il s'agit de paires d'images présentant un ensemble homogène de pixels, dont une partie prend une intensité différente dans la seconde image. Il peut sembler étrange que le modèle détecte l'intégralité de l'image lorsque le pourcentage de changement avoisine se rapproche de 50%.

Pour expliquer cela, notons i_1 et i_2 les deux intensités présentes dans la seconde image et $\alpha \in [0,1]$ la proportion des pixels ayant une intensité i_2. La première image n'étant constituée que de bruit, chaque emplacement de patch voisin peut-être considéré comme étant tiré uniformément dans tout le domaine. De ce fait, en reprenant l'écriture donnée dans l'équation (4.67),

$$\widetilde{V}_p \;=\; \frac{1}{|P_0|} \sum_{\tau \in P_0} \sum_{[q] \in \mathcal{N}^K_{p-\tau}(u)} \widetilde{w}_{p-\tau,q}(u)\, V_{q+\tau}, \tag{4.100}$$

on peut, en première approximation, supposer que les différents $V_{q+\tau}$ sont indépendants et suivent chacun une distribution binaire, valant i_1 avec une probabilité $1-\alpha$ et i_2 avec une probabilité α. En supposant de plus que les $w_{p,q}(u)$ dont tous égaux à 1, la distance entre deux patchs de u n'étant due qu'au bruit, on trouve directement que

$$\widetilde{V}_p \;=\; i_1 + \frac{i_2 - i_1}{K|P_0|} B_p, \tag{4.101}$$

où B_p est une variable aléatoire suivant une loi binomiale de paramètre $K|P_0|$ et de probabilité α. Ainsi, pour tout $p \in \mathcal{D}$ vérifiant $v_p = i_1$,

$$\mu(p,u,V) \;=\; \frac{|i_2 - i_1|}{K|P_0|} B_p, \tag{4.102}$$

qui est de moyenne $\alpha|i_2 - i_1|$ et de variance $\frac{|i_2-i_1|^2}{K|P_0|}\alpha(1-\alpha)$. Il apparaît que, si α est très proche de 0, le modèle ne rajoute qu'une faible erreur de reconstruction et ce avec une faible variance, le pixel p n'étant ainsi a priori pas détectés. Cela est bien visible dans les résultats donnés la figure 4.13, où l'on détecte bien les objets de taille réduite apparaissant sur un fond homogène, sans détecter le fond.

Toutefois, dès lors que α s'écarte se rapproche de 50%, la moyenne (et la variance) de ce bruit binomial augmente, ce qui entache la reconstruction et provoque alors une détection, si l'écart d'intensité $|i_2 - i_1|$ est suffisamment grand. Cet effet est aussi visible dans la paire 1 de la figure 4.13. Ici, lorsque l'on essaye de reconstruire $\widetilde{v}$, il apparaît tout le long du quai une trace plus claire que le reste du sol, trace correspondant aux quelques camions apparaissant dans la deuxième image. Dans ce cas-ci, cela n'aboutit pas une détection, le pourcentage α restant faible.

Ainsi, si le modèle parvient à détecter les changements locaux (n'impactant qu'un nombre réduit de pixels) tout en étant robustes au changement globaux, la présence de changements trop étendu pour être locaux, mais pas assez pour être globaux, peut poser problème.

4.3.3.3 Remarques

Utilisation pour des séries temporelles Le principal apport de ce modèle, vis-à-vis du modèle de détection de changements par destin commun, est qu'il est possible de l'appliquer à des images possédant un grand nombre de canaux sans augmenter trop fortement le temps de calcul. La reconstruction des images ne demandent de n'effectuer que des moyennes pondérées de patchs, ce qui est très rapide à faire. L'extraction

des structures de voisinages, elle, prend beaucoup plus de temps, toutefois, il existe de nombreux algorithmes permettant un calcul rapide, exacte ou approché, des plus K proches voisins. On peut par exemple citer l'algorithme « Fanng » de Harwood et Drummond [HD16] ou l'algorithme « PatchMatch » de Barnes et coauteurs [BSFG09].

Il est ainsi possible d'envisager des changements dans une série temporelle. En considérant des images successives d'une même scène comme différents canaux d'une même image « multitemporelle », il est alors possible d'envisager les changements dans une nouvelle image, en prenant comme référence l'ensemble des images prises précédentes.

Notons toutefois que, dans sa version actuelle, la distance L^2 est utilisée comme mesure de distance entre les patchs, ce qui nécessite d'avoir des images de référence parfaitement recalées entre elles.

Changement de dimension entre extraction et reconstruction En détaillant le modèle, nous nous sommes fixé une taille de patch, que nous avons utilisé dans l'ensemble du modèle. Toutefois, il est possible de reconstruire la seconde image à l'aide d'une taille de patch différente de celle utilisée pour extraire la structure de voisinage de la première image.

En notant d_S et d_R, les largeurs de patchs respectivement utilisées pour l'évaluation de la structure de voisinage et pour la reconstruction, on pourrait ainsi s'assurer que les versions translatés des patchs voisins, lors de la reconstruction, restent inclus dans le patch observé lors de l'évaluation des plus proches voisins, en imposant que $d_S \geq d_R + 4\tau_m$, la valeur $2\tau_m$ correspondant à l'amplitude maximale des translations considérées.

Comparaison non ponctuelle entre v et $\widetilde{v}$ Le processus de reconstruction peut ainsi être vu comme un moyen de se ramener de la comparaison de deux images radiométriquement différentes, u et v, à celle de deux images comparables, v et $\widetilde{v}$. En estimant la norme de la différence entre $\widetilde{v}$ et v, le modèle de détection de changement par reconstruction indirecte applique, après reconstruction, un modèle proche du la comparaison ponctuelle exposée dans la section 3.4. Il pourrait être intéressant de considérer d'autres méthodes de détection de changements dans la paire v et $\widetilde{v}$.

Fusion de détecteurs primaires 5

Table des matières

5.1 Retour sur les détecteurs primaires 168
5.2 Agrégation de détecteurs primaires 180
5.3 Évaluation des performances d'un détecteur 189
5.4 Détecteur agrégé sur un cas réel 197

CONTENU Dans ce chapitre, nous revenons brièvement sur les détecteurs primaires obtenus dans les chapitres précédents.

Nous introduisons le problème d'agrégation de détecteurs comme un problème de classification supervisé, en faisant une présentation rapide des principales méthodes existantes. Nous discutons alors brièvement des post-traitement et mesures de performances, avant de présenter les résultats obtenus sur un cas particulier.

5.1 Retour sur les détecteurs primaires

5.1.1 Choix des détecteurs primaires

5.1.1.1 Retour sur les modèles primaires

Introduction Dans les deux chapitres précédents, nous avons introduit différents descripteurs (voir section 3.3) ainsi que cinq modèles primaires. Chaque modèle peut alors être utilisé sur un (ou plusieurs) descripteurs d'images, pour donner un détecteur de changements dit « primaire », permettant de capturer une définition spécifique de ce qu'est un changement. Toutefois, comme cette définition correspond rarement au cas d'usage donné, nous proposons d'agréger ces détecteurs primaires de manière supervisée, dans le but de construire un détecteur combinant les différentes définitions de changements pour tenter de répondre à la tâche de détection de changements visée.

Avant de nous intéresser à la méthode d'agrégation, reprenons brièvement les cinq modèles que nous avons définis, à savoir :

— le modèle par comparaison ponctuelle, présenté section 3.4 ;
— le modèle par comparaison d'histogramme, présenté section 3.5 ;
— le modèle par modélisation locale du contexte, présenté section 4.1 ;
— le modèle par destin commun, présenté section 4.2 ;
— ainsi que le modèle par reconstruction indirecte, présenté section 4.3.

Nous rapellons ci-dessous les valeurs utilisées pour les différents paramètres. Nous avons affiché, dans la figure 5.1, les résultats qu'ils permettent d'obtenir sur les images synthétiques présentés dans la figure 2.9, en utilisant comme descripteur l'intensité en nuance de gris.

Modèle par comparaison ponctuelle Le modèle par comparaison ponctuelle est le premier modèle construit et consiste simplement à envisager les variations significatives de descripteur au niveau des différents pixels, de manière indépendante du contexte, qu'il soit local ou global. De ce fait, et comme on peut l'apercevoir dans la figure 5.1, ce modèle aura tendance à détecter presque tous les changements, pertinents ou pas. Notons que nous l'avons tout de même doté d'une robustesse aux faibles translations pour limiter les effets des incohérences spatiales, ces dernières apparaissant très souvent dans les images satellite.

En ce qui concerne les paramètres, nous avons choisi d'utiliser une translation maximale de 2 pixels, qui permet d'obtenir un bon compromis entre temps de calcul, robustesse aux décalages spatiaux et biais (voir section 3.4.4.1 pour plus de détails). Par ailleurs, nous avons utilisé une version symétrisée du modèle, consistant à appliquer le modèle d'une image vers l'autre, puis inversement, et à considérer alors le minimum des NFA obtenus.

Modèle par comparaison d'histogramme Le modèle par comparaison d'histogramme a été construit pour ne comparer que les distributions locales, sans prendre en compte une information précise de localisation, dans le but d'obtenir une bonne robustesse aux décalages modérés. Si l'on observe une baisse de la « finesse de détection », particulièrement visible dans la paire 4 de la figure 5.1, ce modèle ne détecte

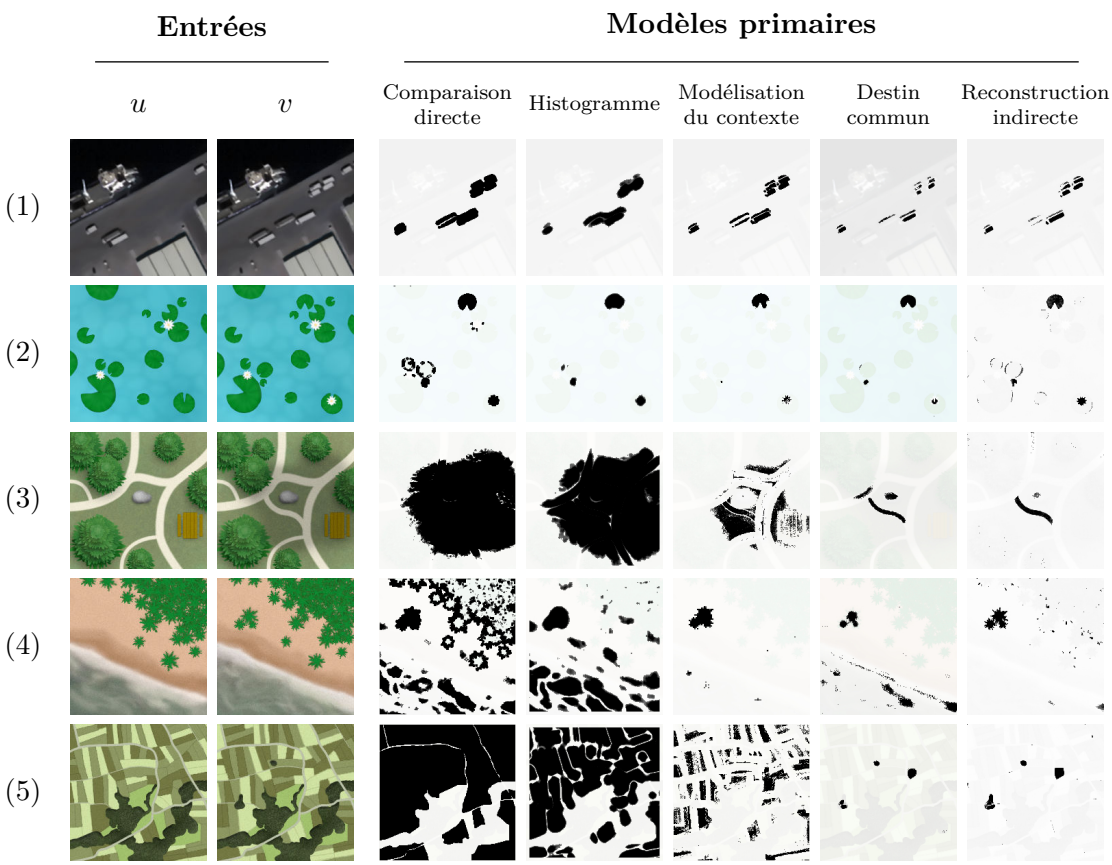


Figure 5.1 – Présentation synthétique des résultats des cinq modèles primaires sur les paires d'images synthétiques présentées dans la Figure 2.9. Pour chacun des modèles, nous présentons successivement le modèle par comparaison directe, par comparaison d'histogramme, par modélisation locale du contexte, par destin commun et par reconstruction indirecte. L'ensemble des paramètres utilisés est détaillé dans la section 5.1.1.1. Dans chacun des cas, nous avons utilisé la nuance de gris comme descripteurs. Dans chaque cas, nous avons représenté en noir les pixels p vérifiant $\mathrm{NFA}(p, u, v) \leq 10^{-10}$, le tout affiché sur une version éclaircie de la seconde image, v. On peut observer ici que le modèle par comparaison directe détecte à peu près tous les changements, pertinents ou pas, que le modèle par comparaison d'histogramme s'avère robuste aux translations de la paire 2, que le modèle par modélisation locale du contexte parvient à éviter presque tous les changements de texture de la paire 4, tandis que les modèles par destins communs et par reconstruction parviennent tout deux à ne détecter que les changements locaux dans les paires 3 et 5, en se distinguant sur les micro et macro-textures de la paire 4.

bien aucune des translations de la paire 2 de cette même figure.

En ce qui concerne les paramètres de ce modèle, il s'agit uniquement de fixer la taille du contexte sur lequel sera faite l'évaluation de la fonction de répartition. Nous avons choisi un rectangle de taille 11×11, qui nous semble être un bon compromis entre un nombre suffisant de points pour pouvoir bien estimer la distribution locale, et une perte limitée dans la résolution du résultat.

Modèle par modélisation locale du contexte Le modèle par modélisation locale du contexte a été conçu pour être robuste aux incohérences spatiales, à l'instar du modèle précédent, mais sans posséder les principales limitations de ce dernier. En confrontant les valeurs des pixels d'une image à la distribution locale d'une seconde image, il parvient alors, dans la figure 5.1, à conserver une certaine finesse du résultat, sans toutefois détecter les translations de la paire 2 ou les variations de texture de la paire 4.

En ce qui concerne les paramètres de ce modèle, comme annoncé dans la section 4.1, nous avons considéré une modélisation du contexte par un mélange de trois gaussiennes, en utilisant un contexte de 11×11 pixels, toujours en considérant une version symétrisée du modèle.

Modèle par destin commun Le modèle par destin commun vient apporter une première compréhension des changements globaux, en s'intéressant aux différents changements d'intensités des pixels vis-à-vis de leurs statistiques globales. On peut observer cela dans la figure 5.1, où ce modèle ne détecte ni le changement de contraste local de la paire 3, ni le changement global de la paire 5. Par ailleurs, il réussit à être robuste aux faibles translations (dans la mesure où elles se présentent en de multiples endroits), ainsi qu'aux variations de micro-textures, visible dans la paire 4.

En ce qui concerne les paramètres utilisés, nous avons choisi d'utiliser le modèle indépendant et une taille d'objets (i.e. E) égale à 7×7. Pour la largeur des « contextes en intensité » (i.e. F et G) nous avons utilisé 20% de l'amplitude globale des images.

Cependant, comme nous l'avons signalé dans la section 4.2.3.2, il s'agit là d'un paramètre important pouvant, s'il est mal choisi, dégrader significativement les résultats du modèle. Pour cette raison, pour les applications sur des images réelles, nous considérons quatre valeurs différentes de ce paramètres (10%, 20%, 30% et 40%) et nous fusionnons les nombres de fausses alarmes associés deux à deux (en utilisant le double du minimum des NFA) pour limiter le nombre de détecteurs associés à ce modèle.

Modèle par reconstruction indirecte Pour finir, le modèle par reconstruction indirecte propose une seconde approche des changements locaux. L'utilisation des K plus proches voisins seulement permet à ce modèle d'obtenir un temps de calcul relativement raisonnable en haute dimension. De la même manière que le modèle précédent, il parvient à détecter les changements locaux, tout en étant robustes aux changements globaux, comme on peut le voir dans les paires 3 et 5 de la figure 5.1.

Toutefois, contrairement au modèle de détection de changements par destin commun, qui envisageait les pixels sans considérer précisément leur spatialisation, ce modèle est capable d'apprécier les éventuels changements de forme d'un objet. On peut

apprécier cette différence dans la paire 4 de la figure 5.1, où les changements au niveau du bord de la micro-texture des arbres, sont en partie détectées.

En ce qui concerne les paramètres de ce modèle, nous avons considéré les 5 plus proches voisins de chaque patch, pris avec une taille de 7×7 pixels, tant pour l'évaluation de la structure de voisinage que pour la reconstruction. Pour la mesure de distance entre patch, nous avons choisi un noyau gaussien de variance $h^2 = 3\sigma^2$, avec σ^2 une estimation de la variance du bruit dans les images. De plus, comme annoncé dans la section 4.3.3.1, nous avons utilisé une robustesse aux translations en prenant $\tau_{\max} = 1$ ainsi qu'un biais vers le patch central pondéré par le paramètre $\lambda = 10^{-1}$.

5.1.1.2 Retour sur les descripteurs utilisés

Introduction Revenons brièvement sur les différents descripteurs que nous avons utilisés pour la construction des détecteurs « primaires ». Nous montrons, dans la figure 5.2, les différents descripteurs obtenus sur une même image, initialement présentée en nuance de gris. Comme mentionné précédemment, chaque descripteur est utilisé avec les différents modèles primaires pour constituer un ensemble de détecteurs « primaires ».

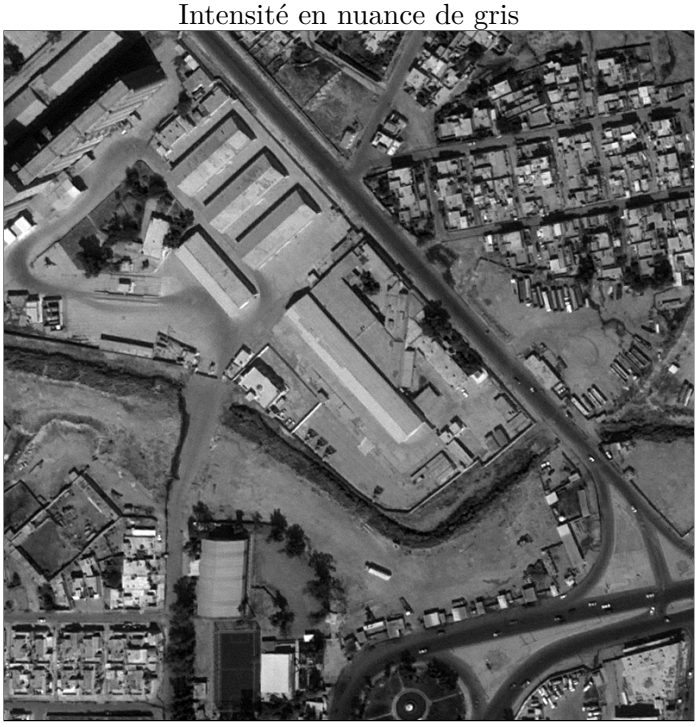
Intensité en nuance de gris Le premier descripteur utilisé est l'intensité en nuance de gris (voir section 3.3.2). Nous avons choisi la nuance de gris plutôt que les trois canaux RGB car nous pensons que le supplément d'information apporté par la couleur n'était pas suffisant pour justifier une hausse du temps de calcul des différents modèles.

Le gradient et norme du gradient Les deux descripteurs suivants, présentés dans la section 3.3.3.1, correspondent au gradient orienté, estimé par produit de convolution avec les filtres de Sobel, et à sa norme. Ces deux descripteurs sont intéressant de part le fait qu'ils capturent une certaine information géométrique de l'image à une échelle fine. Comme annoncé, nous normalisons le gradient de sorte de préserver la variance du bruit.

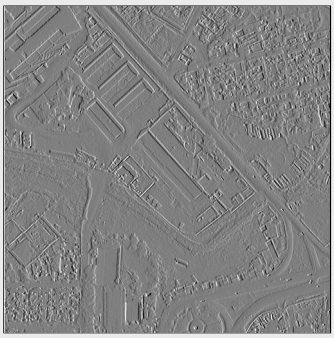
Descripteurs basés sur le tenseur de structure Nous avons utilisé quatre types de descripteurs basés sur le tenseur de structure, défini dans la section 3.3.3.2, à savoir le tenseur brut, ses deux valeurs propres, rangées dans l'ordre décroissant, sa cohérence et le descripteur de coin qui lui est associé, en utilisant la première expression donnée dans l'équation (3.48). Ces quatre descripteurs nous intéressent particulièrement puisqu'ils apportent une compréhension géométrique de l'image à une plus grande échelle que le gradient, en mettant en valeur l'intensité du gradient moyen, l'anisotropie du champ de gradient, ainsi que les structures de bords et de coins. Dans chaque cas, nous avons utilisé une fenêtre gaussienne d'écart type 10 pixels. Par ailleurs, nous avons considéré la version régularisée de la cohérence du tenseur de structure, dont l'expression est donnée dans l'équation (3.47), en prenant le coefficient α de l'ordre de la variance du bruit présent dans les images.

On peut remarquer, dans la figure 5.2, que le descripteur de « Cornerness » apparaît comme très proche de la seconde valeur propre. Cela s'explique par l'expression de la

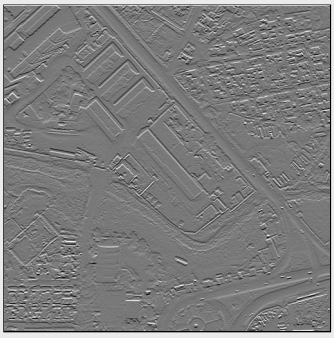
Intensité en nuance de gris



Gradient horizontal

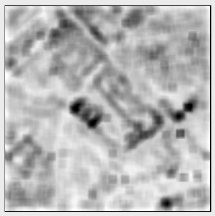


Gradient vertical

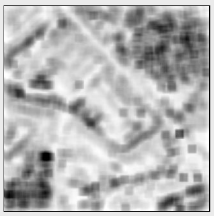


Ondelette H–1

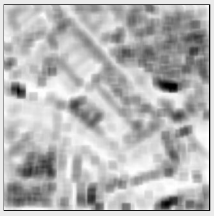
Ondelette D–1



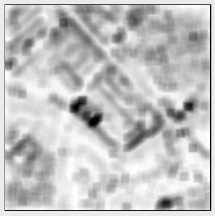
Ondelette V–1



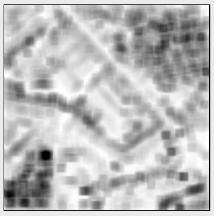
Ondelette H–2



Ondelette D–2

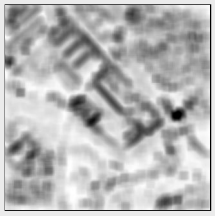


Ondelette V–2

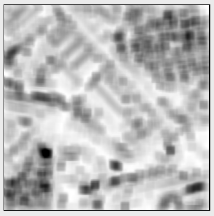


Ondelette H–3

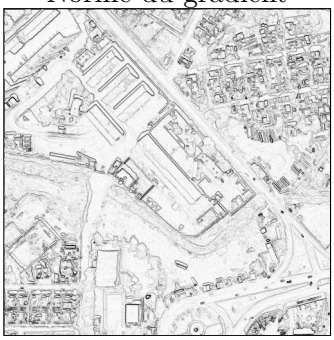
Ondelette D–3



Ondelette V–3



Norme du gradient



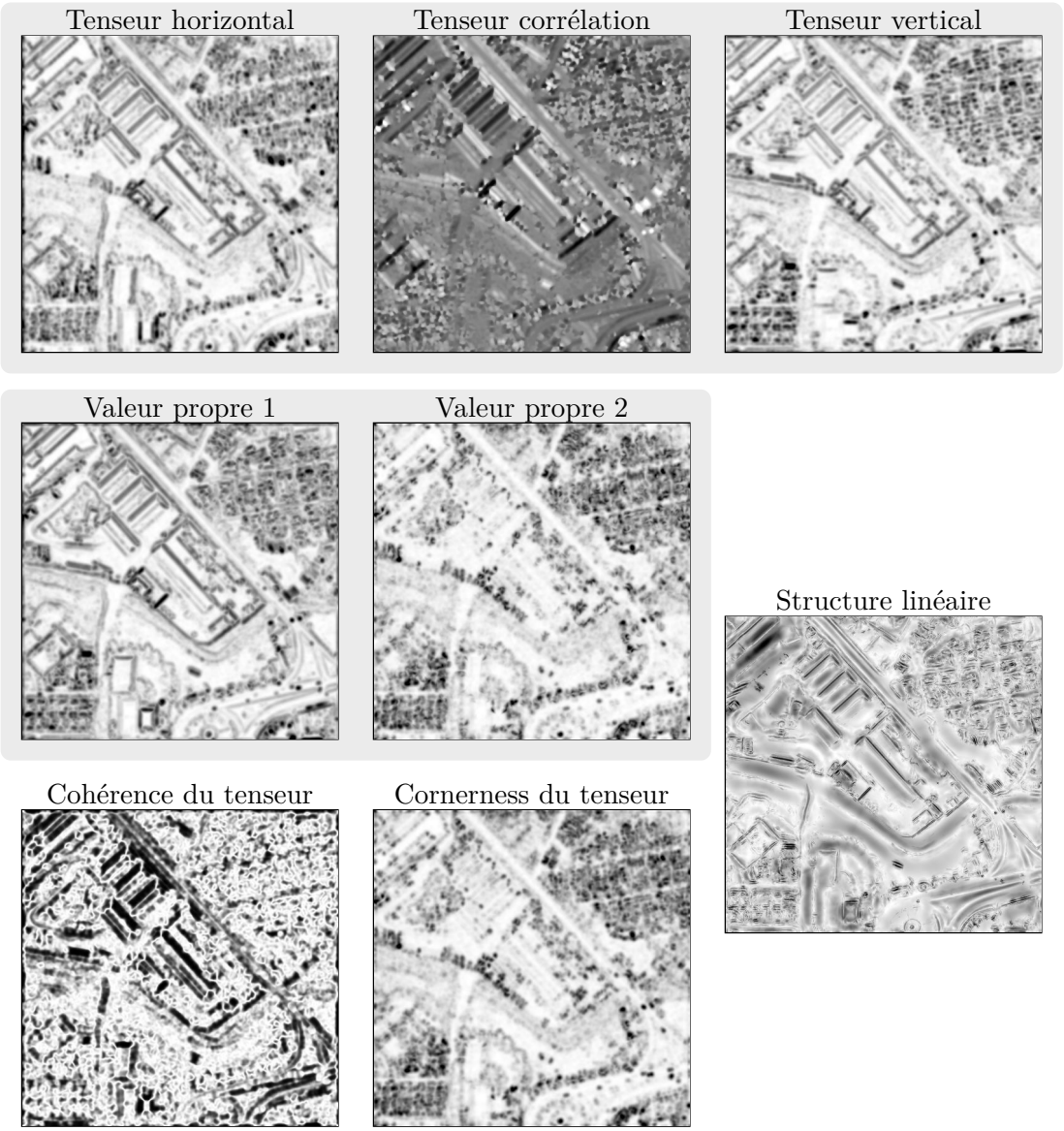


Figure 5.2 – Visualisation des descripteurs choisis appliqués à la première image de la paire d'images réelles présentée dans la figure 5.3. Nous avons ici représenté chaque descripteur utilisé lors de la construction des modèles primaires, en regroupant les différentes composantes par un rectangle grisé. Sur la page de gauche, on trouve successivement l'intensité en nuances de gris, le descripteur d'ondelette (où H/D/V–k représente une ondelette horizontale/diagonale/verticale à l'échelle k), le gradient orienté ainsi que sa norme. Sur la page de droite, on trouve successivement le tenseur de structure brut, ses deux valeurs propres, sa cohérence, sa « cornerness », ainsi que le descripteur de structures linéaires À l'exception de l'intensité en nuance de gris, chaque image est représentée en utilisant une teinte foncée pour des valeurs élevées et une teinte claire pour des valeurs faibles. Notons que, dans le cas du gradient horizontal, vertical ainsi que du coefficient de corrélation du s tenseur, la couleur grise correspond à une valeur nulle, ces descripteurs étant signés.

cornerness, donnée dans l'équation (3.48),

$$\text{Cornerness} = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}, \tag{5.1}$$

où λ_1 et λ_2 sont les deux valeurs propres, dont la valeur varie donc entre $\frac{1}{2}\lambda_2$ et λ_2.

Descripteur de structures linéaires Nous avons utilisé le descripteur de structures linéaires, détaillé dans la section 3.3.3.3. S'il peut sembler redondant vis-à-vis du descripteur de bords, apporté par la cohérence du tenseur de structure, nous l'avons jugé utile dans le contexte des images satellite, présentant très souvent des routes, murs et autres structures linéaires que l'on pourrait vouloir distinguer d'un bord quelconque.

En ce qui concerne les paramètres de ce modèle, nous avons considéré, comme filtre de base, la dérivée seconde d'une gaussienne anisotrope de facteur $\alpha = 2$, dont l'expression est donnée dans l'équation (3.50). Par ailleurs, nous considérons comme descripteur le maximum des réponses des filtres, pour une variance allant de 1 à 28.

Descripteur de texture Nous utilisons le descripteur de texture basé sur les coefficients d'ondelettes, défini dans la section 3.3.4.

Nous utilisons ce descripteur en moyennant « l'énergie d'ondelette » sur un carré de 32×32 pixels et pour les trois directions d'ondelettes sur les trois premières échelles (pour un total de 9 composantes), et en utilisant les ondelettes de Daubechies.

Résolution des descripteurs Il est intéressant d'observer les différentes échelles représentées par ces descripteurs. L'intensité en nuance de gris et les deux descripteurs associés au gradient présentent une information à l'échelle ou proche de la résolution de l'image. Les descripteurs associés au tenseur apportent une information à une échelle plus large (une dizaine de pixels), les images rendues apparaissant relativement « floues », tandis que le descripteur de texture fourni une information à grande échelle (plusieurs dizaines de pixels). Enfin, le descripteur de structure linéaire donne comme résultat une image mélangeant des éléments flous avec d'autres plus précis. Ce mélange de plusieurs résolutions différentes provient directement de sa définition, $D(I) = \max_\sigma D_\sigma(I)$, $D_\sigma(I)$ représentant la réponse des filtres gaussiens à l'échelle σ appliqués l'image I.

5.1.2 Détections primaires

Détecteurs utilisés Nous disposons de cinq modèles différents, dont un (le modèle par destin commun) possédant deux jeux de paramètres, ainsi que neuf descripteurs différents. En combinant chaque modèle avec chaque descripteur, nous obtenons donc un total de 54 détecteurs primaires.

Notons qu'il serait possible d'appliquer un modèle à la concaténation de plusieurs descripteurs, démultipliant ainsi le nombre de détecteurs primaires potentiels. Nous jugeons toutefois que l'apport de tels détecteurs serait négligeable devant la hausse de complexité que cela entraînerait. Nous nous sommes donc restreint aux couples formés par un descripteur et un modèle.

Toutefois, le descripteur de texture étant de grande dimension, le temps de calcul nécessaire pour appliquer les modèles « avancés » s'avère particulièrement important. Pour cette raison, nous n'avons pas utilisé les détecteurs primaires correspondant aux modèles par modélisation du contexte local, par destin commun ou par reconstruction indirecte, appliqués aux descripteurs de texture. Nous nous donc d'un total de 50 détecteurs primaires.

Détections primaires sur un cas réel Nous avons appliqués les 50 détecteurs primaires à une paire d'images réelles, présentée dans la figure 5.3, et affiché les résultats obtenus dans la figure 5.4. Ces résultats nous donnent l'occasion de faire quelques dernières remarques sur les différents modèles primaires utilisés.

Avant toute chose, précisons la manière dont sont visualisées ces images. Dans les résultats sur les images synthétiques présentées dans les figures précédentes, nous avions représenté en noir les pixels dont la valeur du nombre de fausses alarmes dépassait une valeur seuil (fixée à 10^{-10}). Puisque nous cherchons à agréger les détecteurs primaires, il peut être pertinent ici de représenter une plus large plage de valeurs. Ainsi, dans chaque image de résultat de cette figure, nous avons affiché en blanc les pixels de NFA inférieur à 10^{-10}, puis du gris au noir, les pixels dont la valeur du NFA est supérieure à 10^{-10}, d'autant plus sombre que leur NFA est élevé.

Incompatibilité d'échelle entre modèle et descripteur Tout d'abord, remarquons que les deux modèles, par comparaison d'histogramme et par destin commun, ont tendance à détecter presque l'intégralité de l'image lorsqu'ils sont utilisés sur des descripteurs basés sur le tenseur de structure ou sur le descripteur de texture. Ceci est particulièrement visible pour le détecteur par comparaison d'histogramme appliqué au descripteur d'ondelettes. Ceci s'explique par l'hypothèse d'indépendances des pixels que supposent ces deux modèles. En utilisant des descripteurs capturant l'information à une grande échelle, sans toutefois réduire la résolution pixellique des images, les valeurs des pixels sont fortement corrélés, ce qui remet complètement en cause l'hypothèse d'indépendance et engendre donc des détections indésirable. Il peut être utile de préciser que le modèle naïf, utilisé pour la construction des nombres de fausses alarmes, fournit un cadre pour les images et non un critère fixe, il n'est donc pas choquant qu'il ne soit pas systématiquement vérifié par les différentes images.

Il serait intéressant d'étudier la limite de « résolution de l'information », capturée par les descripteurs, à partir de laquelle un détecteur primaire construit avec ces deux modèles perdrait tout intérêt. Nous utilisons tout de même utilisé l'intégralité des détecteurs, y compris ceux qui pourraient sembler sans intérêt, en nous attendant malgré tout à ce que ces derniers n'aient qu'une influence négligeable dans le processus de fusion.

Notons que cela souligne à nouveau les différences entre les modèles par destin commun et par reconstruction indirecte, ce dernier ne souffrant pas du tout de cette forte corrélation.

Les deux paramètres du modèle par destin commun Arrêtons-nous quelques instants sur les deux versions du modèle par destin commun. En premier abord, on

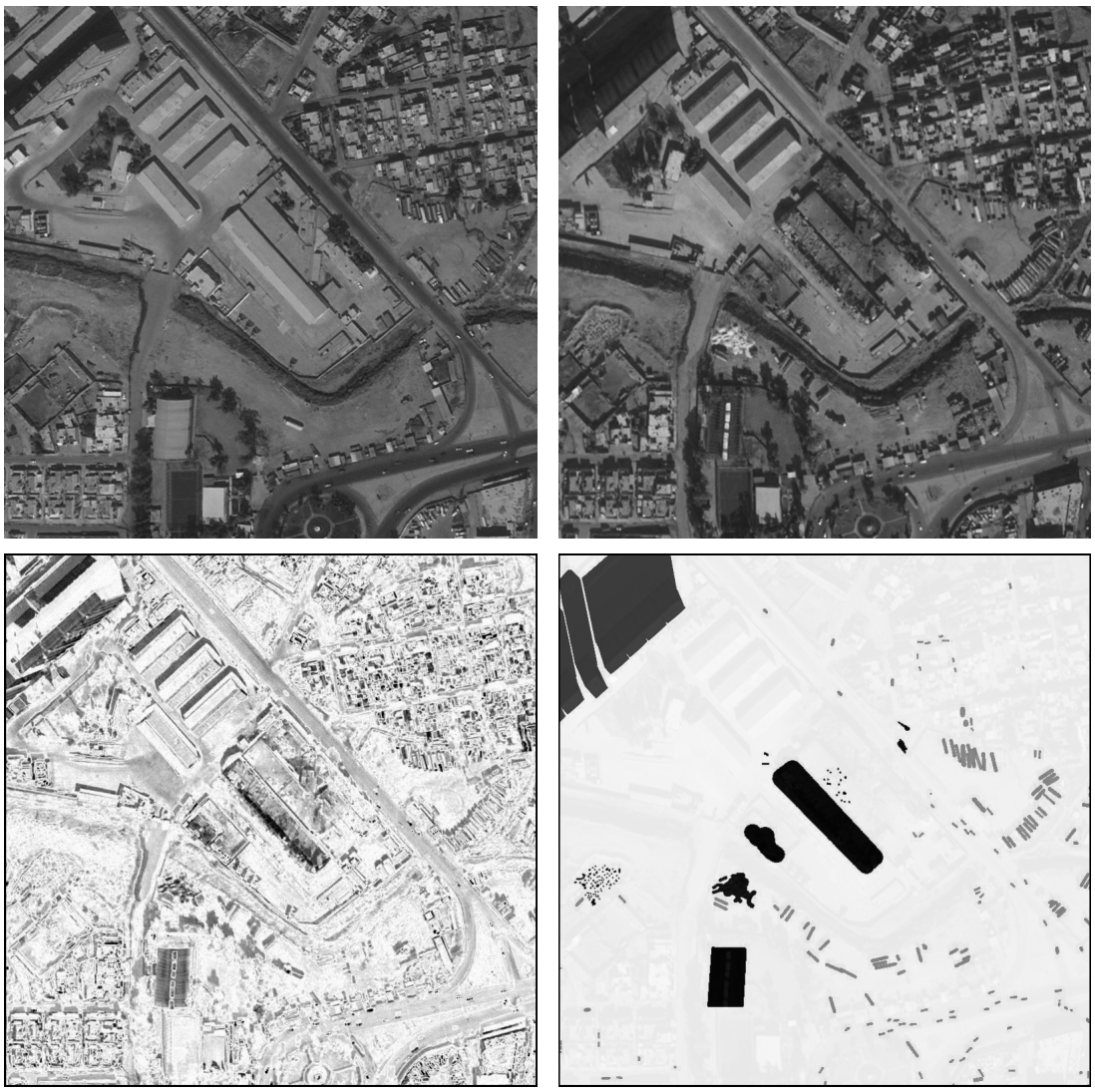
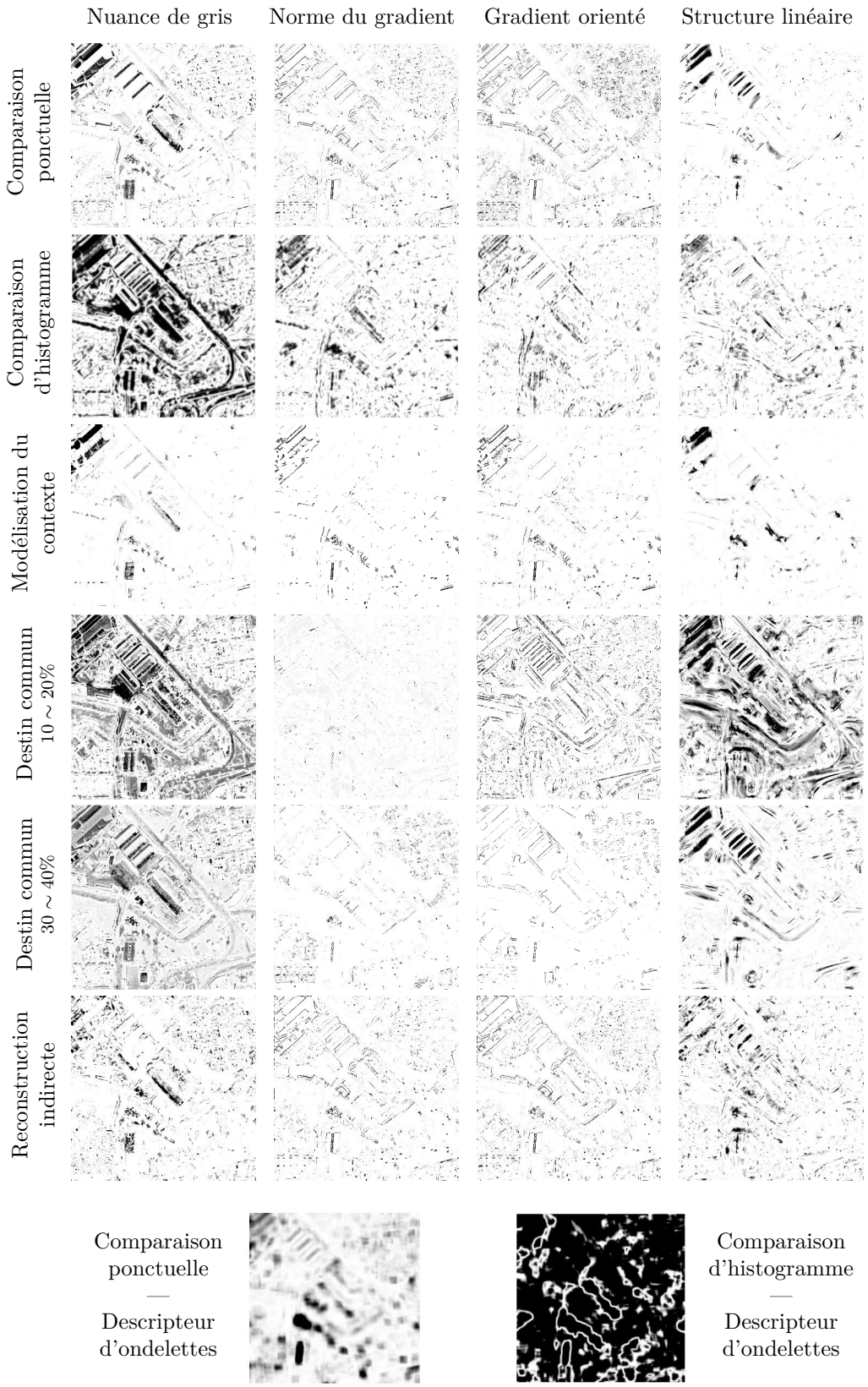


Figure 5.3 – Paire d'images acquises respectivement par les satellites « Pléiades » et « WorldView-2 ». On présente ici successivement les deux images, la valeur absolue de leur différence (d'autant plus foncée que la différence est importante), ainsi qu'une représentation schématique des différents changements présents dans la scène, en excluant ceux dus à la parallaxe et aux variations d'ombres, particulièrement importantes entre ces images et bien visibles dans l'image de différence (à l'exception du bâtiment dans le coin haut-gauche, tout de même indiqué ici). Dans la représentation schématique des changements, nous avons affiché en gris les changements correspondant aux apparitions et disparitions de véhicules (voiture, camions et bus), que ce soit sur la route, sur les parking ou autre, et en noir les autres changements, c'est à dire respectivement, de haut en bas, l'ombre et la parallaxe de l'immeuble, l'apparition de taches sur le sol, la reconstruction de deux bâtiments détruits, la disparition d'un ensemble d'objets blancs difficilement identifiables et l'apparition du toit d'un troisième bâtiment.

peut avoir l'impression que le modèle utilisant les paramètres 10% et 20%, que nous qualifierons ici de « première version », donne des valeurs de NFA toujours supérieures à celles du modèle utilisant 30% et 40%, nommé ici « deuxième version », en rajoutant de plus des détections apparemment sans importance. Ça n'est toutefois pas le cas lorsque l'on utilise la norme du gradient et la cohérence du tenseur de structure, et on peine à expliquer ce comportement. Pour cette raison, nous avons conservé ces deux modèles distincts.

Une orientation du gradient non informative Remarquons la très grande similitude entre les résultats obtenus avec le tenseur de structure brut et avec ses valeurs propres (à l'exception du modèle par modélisation du contexte), ainsi qu'entre ceux obtenus avec le gradient orienté et avec sa norme (à l'exception cette fois-ci de l'un des deux modèles par destin commun). Dans chacun des deux cas, les deux descripteurs se distinguent par leur prise en compte ou non de l'orientation du gradient. Ainsi, on pourrait en conclure que cette orientation n'est que peu pertinente ici. Toutefois, il ne s'agit que d'une image en particulier, et nous préférons continuer d'utiliser tous les descripteurs proposés.


Nuance de gris
Norme du gradient
Gradient orienté
Structure linéaire
Comparaison ponctuelle
Comparaison d'histogramme
Modélisation du contexte
Destin commun 10 ∼ 20%
Destin commun 30 ∼ 40%
Reconstruction indirecte
Comparaison ponctuelle — Descripteur d'ondelettes
Comparaison d'histogramme — Descripteur d'ondelettes

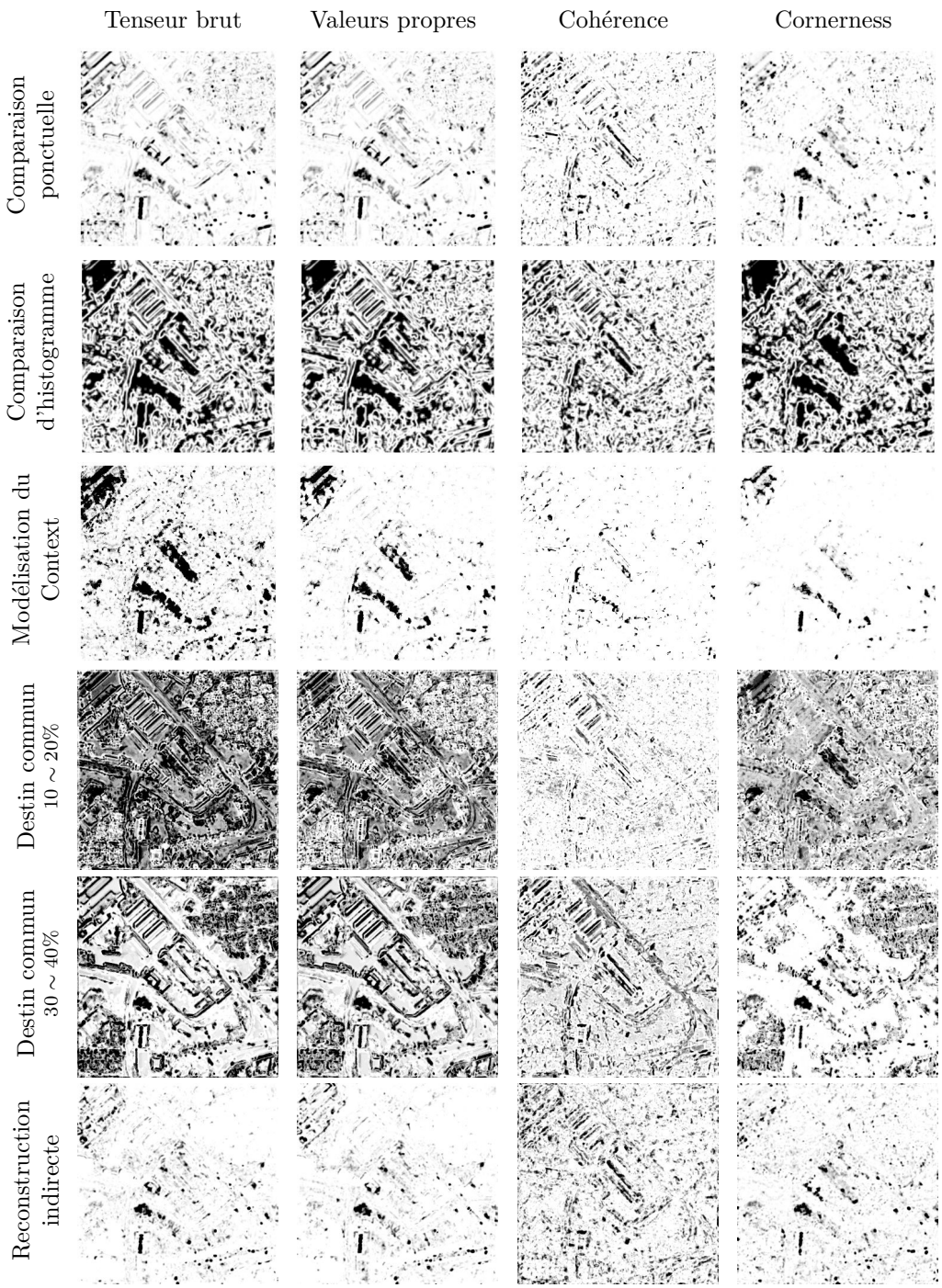


Figure 5.4 – Résultats des détecteurs primaires choisis (détaillés dans la section 5.1.2), appliqué à un cas réel présenté dans la figure 5.3. Chaque ligne correspond ici à un des modèles primaires listé dans la section 5.1.1.1, et chaque colonne à un des descripteurs listés dans la section 5.1.1.2, et ce à l'exception des deux modèles utilisant les descripteurs de texture qui sont isolés en bas de la page de gauche. Dans chaque image, la couleur d'un pixel est d'autant plus sombre que sa significativité est importante. On se référera à l'intégralité de la section 5.1.2 pour plus de détails.

5.2 Agrégation de détecteurs primaires

5.2.1 Problème de classification

5.2.1.1 Formulation du problème

Comme nous l'avons mentionné précédemment, chaque couple formé par un modèle primaire et un descripteur d'image permet de définir un détecteur dit « primaire ». Si séparément, ces détecteurs primaires conviennent rarement aux différents cas d'applications, en les agrégeant, nous pouvons construire un modèle de changement plus expressif et qui pourra alors s'adapter à des tâches plus spécifiques.

Données labelisées Il peut être difficile de caractériser précisément ce qu'est un changement dans le cadre d'une tâche spécifique. De ce fait, nous envisageons chaque tâche à effectuer au travers d'un utilisateur extérieur décidant au cas par cas de ce qui est ou n'est pas un changement pour cette tâche particulière. Formellement, on se donne un ensemble de M paires d'images (u^i, v^i), pour $i \in [\![1, M]\!]$ et $M \in \mathbb{N}^*$, chacune définie sur un domaine $\mathcal{D}^i \subset \mathbb{Z}^2$ et à valeur dans $\mathbb{R}^K$, $K \in \mathbb{N}^*$ étant le nombre de canaux supposé commun à toutes les images. On suppose, par ailleurs, que pour chaque paire, l'utilisateur fournit une « vérité terrain », noté w^i, correspondant aux emplacements des changements dans la paire en question et prenant la forme d'un masque de segmentation binaire, c'est-à-dire

$$w^i : \mathcal{D}^i \to \{-1, 1\} \tag{5.2}$$

avec pour tout $p \in \mathcal{D}^i$, $w^i_p = 1$ si et seulement si l'utilisateur juge le pixel p de la paire i comme appartenant à un changement.

Notation des détecteurs primaires On suppose alors fixé un nombre $n \in \mathbb{N}^*$ de détecteurs primaires, appliqués aux différentes paires d'images, dont les résultats seront agrégés, et que nous noterons par la suite F_j, pour tout $j \in [\![1, n]\!]$.

Notons la définition de F_j devrait formellement dépendre du domaine de l'image, avec, pour tout $j \in [\![1, n]\!]$,

$$F_j^D : \begin{cases} \mathbb{R}^{K\times|D|} \times \mathbb{R}^{K\times|D|} & \to & \mathbb{R}^{|D|} \\ (u, v) & \mapsto & F_j^D(u, v) \end{cases}, \tag{5.3}$$

avec $D \subset \mathbb{Z}^2$. Pour simplifier les notations, nous ne noterons pas cette dépendance. Par la suite, pour tout $j \in [\![1, n]\!]$, $i \in [\![1, M]\!]$ et $p \in \mathcal{D}^i$, nous noterons $F_j(u^i, v^i)_p$ la valeur du détecteur j sur la paire i au niveau du pixel p, ainsi que $F(u^i, v^i)_p$, leur concaténation, appartenant à $\mathbb{R}^n$.

Agrégation supervisée Il est alors question d'agréger les différents détecteurs de manière supervisée. Pour cela, nous effectuons une classification directe des différents pixels, sans tenir compte des corrélations spatiales existant dans le résultat obtenu.

Nous construisons l'ensemble d'apprentissage, à partir des différentes paires d'images labellisées, en isolant l'ensemble de tous les pixels, c'est-à-dire en définissant

$$\mathcal{E} \;=\; \left\{\left(F(u^i, v^i)_p, w_p^i\right), i \in [\![1, M]\!], p \in \mathcal{D}^i\right\} \;\subset\; \mathbb{R}^n \times \{-1, 1\}. \tag{5.4}$$

On retrouve alors le cadre classique des problèmes de classification.

Notons que, si, comme annoncé dans la section 3.2.1, nous n'avons pas ici considéré, dans les notations, les problèmes de définition au niveau des bords des différents domaines $\mathcal{D}^i$, il est nécessaire, dans la pratique, de ne pas considérer les pixels trop proches des bords.

5.2.1.2 Problème de classification

Agrégation ou classification Si l'appellation « agrégation des détecteurs primaires » semble nous orienter directement vers des méthodes d'agrégation de classifieurs, notons que les différents détecteurs primaires ne répondent pas à la même définition de changements. Aussi, les méthodes d'agrégation basiques, comme l'union, l'intersection ou encore le vote des différents masques de détection n'auraient a priori que peu de sens ici.

Nous abordons donc ce problème d'agrégation comme un problème de classification supervisée, en utilisant les résultats des différents détecteurs primaires comme ensemble de coordonnées associé à chaque pixel.

Cadre formel d'apprentissage Pour formaliser les choses, et en simplifiant les notations, nous disposons donc d'un ensemble d'apprentissage

$$\mathcal{E} \;=\; \left\{\left(x^i, y^i\right), i \in [\![1, N]\!]\right\} \;\subset\; \mathbb{R}^n \times \{-1, 1\}, \tag{5.5}$$

chaque x^i (appartenant à $\mathbb{R}^n$) correspondant aux n réponses des détecteurs primaires sur l'un des pixels d'une des paires d'images labellisées, N représentant le nombre total de pixels, c'est-à-dire

$$N \;=\; \sum_{i \in [\![1, M]\!]} |\mathcal{D}_i|. \tag{5.6}$$

L'apprentissage d'un classifieur consiste formellement à trouver une fonction $\widetilde{G}$ de $\mathbb{R}^n$ à valeur dans $\{-1, 1\}$, et correspondant « le plus possible » aux données d'apprentissage $\mathcal{E}$. Cependant, il est plus simple de chercher une fonction de régression G de $\mathbb{R}^n$ à valeur dans $\mathbb{R}$, que l'on seuille ensuite pour obtenir une classification binaire, avec

$$\widetilde{G}(x) \;=\; \begin{cases} 1 & \text{si } G(x) \geq 0 \\ -1 & \text{si } G(x) < 0. \end{cases} \tag{5.7}$$

Il s'agit alors de trouver G de sorte que $G(x^i)$ corresponde le plus possible à y^i, pour tout $(x^i, y^i) \in \mathcal{E}$. La correspondance est spécifiée au moyen d'une fonction d'objectif, $\mathcal{L}_{\mathcal{E}}$, qui à G associe une valeur réelle, d'autant plus petite que G correspond bien à $\mathcal{E}$.

Dans la grande majorité des cas, $\mathcal{L}$ est écrite à l'aide d'une fonction L de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}$, dite souvent « de perte », sous la forme

$$\mathcal{L}_{\mathcal{E}}(G) \;=\; \sum_{(x,y)\in\mathcal{E}} L(y, G(x)), \tag{5.8}$$

le terme $L(y, G(x))$ quantifiant alors la justesse de la prédiction $G(x)$ par rapport à la valeur attendue y, par exemple en considérant la fonction de « perte quadratique »

$$L(y, y') \;=\; (y - y')^2. \tag{5.9}$$

Notons que l'expression de la fonction d'objectif, donnée dans l'équation (5.8), peut être interprété comme une estimation de l'espérance $\mathbb{E}[\, L(Y, G(X))\,]$, où le couple de variables aléatoires (X, Y) correspond aux données et labels sur lesquels sont appliqués la fonction de classification, l'ensemble d'apprentissage $\mathcal{E}$ étant alors supposé approcher la distribution « naturelle » de (X, Y).

5.2.2 Méthodes de classification

Résumons brièvement les principales méthodes classiques existant pour résoudre le problème de classification. Dans toute la suite, nous avons utilisé les notations définies dans la section précédente, en particulier celles données dans les équations (5.5) et (5.8).

5.2.2.1 Classifieurs simples

Machine à vecteur de support (SVM) Inventé par Vapnik en 1963 et reprise, dans sa forme actuelle, par Cortes et coauteurs en 1995 [CV95], la machine à vecteur de support (SVM) consiste à essayer de séparer les deux classes d'objets, $y^i = 1$ et $y^i = -1$, par un hyperplan, en maximisant la distance minimale des points à celui-ci. Cela revient alors à écrire la fonction de classification G sous la forme

$$G(x) \;=\; \omega^{\mathsf{T}} x + b, \tag{5.10}$$

pour tout $x \in \mathbb{R}^n$, avec $\omega \in \mathbb{R}^n$ et $b \in \mathbb{R}$. Ce problème est exprimé comme la minimisation de

$$\|\omega\|_2^2 + \gamma \sum_{i=1}^{N} \xi_i \tag{5.11}$$

par rapport à ω, b et de nouvelles variables ξ_i (pour $i \in [\![1, N]\!]$), sous les contraintes

$$\begin{aligned} &\forall i \in [\![1, N]\!]\,, \quad y^i(\omega^{\mathsf{T}} x^i + b) \geq 1 - \xi_i, \\ &\forall i \in [\![1, N]\!]\,, \qquad\qquad \xi_i \in \mathbb{R}^+, \end{aligned} \tag{5.12}$$

où γ est une constante positive fixée. Dans le cas où il n'existe pas d'hyperplan séparant les données $y^i = 1$ et $y^i = -1$, une méthode, introduite par Boser et coauteur en 1992 [BGV92], permet de généraliser le problème à des cas non linéaires avec des méthodes à noyaux, cette méthode étant surnommée l'« astuce du noyau ».

Arbre de régression Une autre méthode classique consiste à partitionner l'espace $\mathbb{R}^n$ en plusieurs régions auxquelles sont attribuées une même valeur. La construction de ces régions se fait à l'aide d'un arbre de régression, constitué d'un ensemble arborescent de tests correspondant chacun à la comparaison d'une composante avec une valeur seuil donnée. En ayant partitionné $\mathbb{R}^n$ en m régions R_j, pour $j \in [\![1, m]\!]$, la fonction de classification s'écrit alors

$$G(x) \;=\; \sum_{j=1}^{m} c_j \mathbb{1}_{\{x \in R_j\}} \tag{5.13}$$

pour tout $x \in \mathbb{R}^n$, avec

$$c_j \;=\; \frac{\displaystyle\sum_{i=1}^{n} y^i \mathbb{1}_{\{x^i \in R_j\}}}{\displaystyle\sum_{i=1}^{n} \mathbb{1}_{\{x^i \in R_j\}}}, \tag{5.14}$$

cette forme étant alors qualifiée d'arbre de régression [BFSO84].

La construction d'un tel arbre se fait de manière gloutonne, comme illustré dans la figure 5.5, en commençant par trouver la meilleure séparation du plan en deux régions, puis récursivement en divisant une nouvelle région. Cette nouvelle séparation est choisie de sorte de minimiser une version pénalisée de la fonction d'objectif présentée dans l'équation (5.8) avec perte quadratique,

$$\sum_{j=1}^{m} \left(\sum_{x^i \in R_i} (y^i - c_j)^2 \right) \;+\; \alpha\, m \tag{5.15}$$

où m est le nombre de région, où les c_j sont définis comme dans l'équation (5.14) et où $\alpha > 0$ est le paramètre de régularisation. Sans le terme de régularisation (et en supposant les x^i disjoints), il serait possible de continuer de scinder chaque régions, en faisant baisser la fonction d'objectif, jusqu'à ce que tous les x^i soient isolés. Le terme αm permet d'éviter cette situation en stoppant le processus dès lors qu'une scission n'aboutit pas à une baisse suffisante de la fonction d'objectif. Le paramètre α représente donc le compromis fait entre la complexité de l'arbre et sa « précision ».

Le grand intérêt des arbres de régression est que leur construction se fait très rapidement et il est donc possible d'utiliser ces derniers comme support de méthodes plus complexes, comme celles présentées dans la section suivante.

5.2.2.2 Forêts aléatoires et gradient boosting

Algorithme d'agrégation « Bootstrap » L'algorithme d'agrégation « Bootstrap » (ou « Bagging ») est un algorithme introduit par Breiman [Bre96] permettant d'améliorer les performances d'un algorithme de classification (ou de regression), en moyennant les résultats de plusieurs instances de ce dernier, appelée classifieurs faibles, appris sur des sous-ensembles de l'ensemble d'apprentissage.

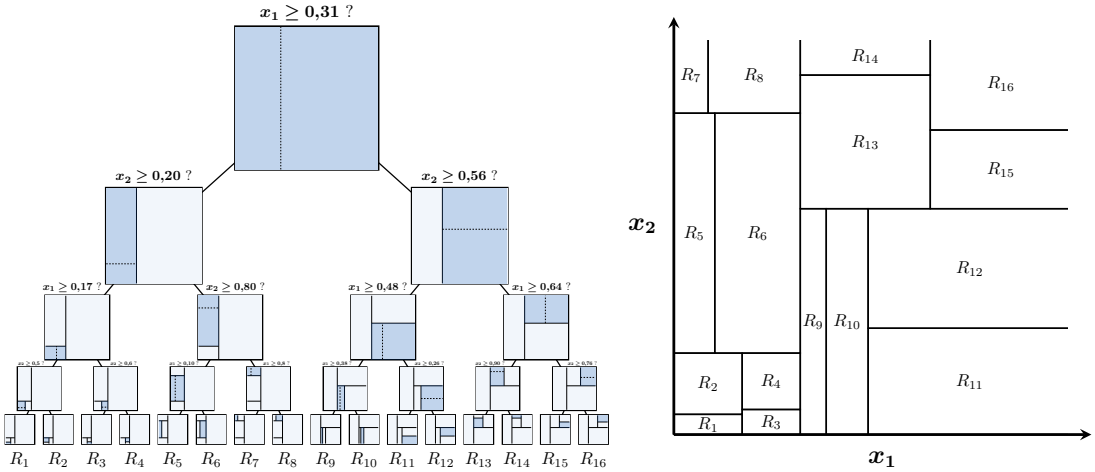


Figure 5.5 – Représentation schématique de la partition de l'espace en régions dans le cadre de la classification par un arbre de régression, pour des données à deux dimensions x_1 et x_2. À gauche, nous avons affiché une vision arborescente de l'arbre de régression en représentant à chaque noeud le découpage de la région correspondante, schématisant le processus de construction d'un arbre de régression. À droite, nous avons représenté la partition du plan par les différents rectangles caractérisant l'arbre de régression.

En notant h_E, le classifieur faible entraîné sur un sous-ensemble $E \subset \mathcal{E}$ (avec donc $h_E : \mathbb{R}^n \to \{-1, 1\}$), l'algorithme de *bagging* » considère le classifieur défini par

$$G(x) \;=\; \frac{1}{m} \sum_{i=1}^{m} h_{E_i}(x), \tag{5.16}$$

où E_i est un sous-ensemble aléatoire de $\mathcal{E}$. En considérant des sous-ensembles aléatoires E_i, on moyenne ainsi des classifieurs identiquement distribués, cherchant à obtenir un classifieur de même moyenne, mais de plus faible variance. De ce fait, le *bagging* est principalement utilisé avec des classifieurs faibles à très grande variance, comme les arbres de régression.

Forêt d'arbres de régression (Random Forest) L'algorithme de Forêt d'arbres de régression [Ho95] repose sur le constat suivant. En moyennant m classifieurs identiquement distribués et indépendants, de variance σ^2, la variance du classifieur obtenu est alors σ^2/m. Toutefois, si ces classifieurs sont corrélés, avec un coefficient de corrélation $\rho > 0$, cette variance s'écrit alors

$$\frac{\sigma^2}{m} \;+\; \rho\sigma^2 \left(1 - \frac{1}{m}\right), \tag{5.17}$$

le second terme venant ainsi réduire l'intérêt d'un algorithme d'agrégation *bootstrap*. L'algorithme de Forêt d'arbres de régression ou « Random Forest » reprend alors l'algorithme d'agrégation *bootstrap* pour des arbres de régression, en modifiant la manière ceux-ci sont construits à chaque étape, pour limiter leur corrélation, sans trop amoindrir leur efficacité.

Il s'agit, lors de la génération d'un arbre de régression, et à chaque étape de recherche de région à scinder, de se restreindre à un sous-ensemble de r dimensions tirées aléatoirement parmi les n dimension de l'espace, avec classiquement $r = \sqrt{n}$, et ce en continuant de minimiser une fonction d'énergie similaire à celle présentée dans l'équation (5.15).

Dans la pratique, cet algorithme donne de bons résultats pour un nombre très réduit de paramètres.

Algorithme de gradient boosting Les algorithmes de gradient boosting, introduit par Friedman [Fri01], et tout comme les algorithmes d'agrégation *bootstrap*, sont des méthodes cherchant à agréger un ensemble de « régresseurs faibles ». Seulement, contrairement à ces derniers, il ne s'agit plus ici de réduire la variance des régresseurs « faibles », mais au contraire, de les associer de sorte d'améliorer constamment une fonction d'objectif. En écrivant la fonction de regression sous la forme

$$G(x) \;=\; \sum_{t=1}^{m} \alpha_t h_t(x), \tag{5.18}$$

les différents h_t correspondant aux régresseurs faibles, construits successivement de manière à optimiser la fonction d'objectif

$$\mathcal{L}(G) \;=\; \sum_{t=1}^{N} L(y^t, G(x^t)), \tag{5.19}$$

où L est une fonction de perte choisie en amont.

L'algorithme du gradient boosting consiste alors à effectuer cette optimisation effectuant une descente de gradient, chaque fonction de régression h_t étant appris pour approcher au maximum le gradient de la fonction d'objectif. À un temps t fixé, et en ayant déjà construit $G_t = \sum_{u=1}^{t} \alpha_u h_u$, il s'agit alors d'apprendre un classifieur faible h_{t+1} sur les données (x^i, r_i), pour tout $i \in [\![1, N]\!]$, avec r_t^i la direction du gradient au point x^i,

$$r_t^i \;=\; \left[\frac{\partial L(y^i, f(x^i))}{\partial f(x^i)}\right]_{f=G_t(x^i)}, \tag{5.20}$$

et de mettre à jour le classifieur agrégé en posant $G_{t+1} = G_t + \alpha_{t+1} h_{t+1}$, avec

$$\alpha_{t+1} \;=\; \underset{\alpha}{\operatorname{argmin}}\; \mathcal{L}(G_t + \alpha h_{t+1})\,. \tag{5.21}$$

Gradient Tree Boosting et XGBoost Friedman, à l'origine du développement de l'algorithme de gradient boosting, présente aussi, dans [Fri01], l'algorithme de « Gradient Tree Boosting », adaptation de l'algorithme de gradient boosting dans le cas d'arbres de régression. En 2016, Chen et Guestrin [CG16] introduise l'algorithme « XGBoost », une nouvelle méthode d'approximation d'arbres de régression ainsi que différentes optimisations de l'algorithme de Friedman permettant d'accélérer fortement ce dernier.

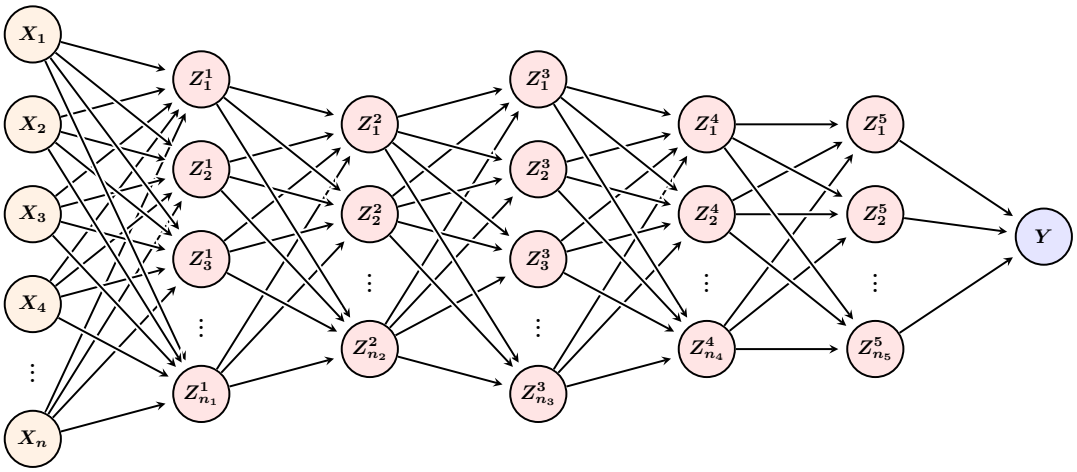


Figure 5.6 – Représentation schématique d'un réseau de neurones artificiels à six couches. Les variables X_1, …, X_n constituent les variables d'entrée, Y la variable de sortie et les Z_i^j les variables cachées, chaque flèche représentant les différentes dépendances de calcul.

5.2.2.3 Réseau de neurones

Perceptrons Inventé par Rosenblatt en 1958 [Ros58], un perceptron est une fonction de classification écrite sous la forme

$$G(x) \;=\; \sigma(\omega^{\mathsf{T}} x + b), \tag{5.22}$$

pour $x \in \mathbb{R}^n$, $\omega \in \mathbb{R}^n$ et $b \in \mathbb{R}$, les paramètres, et $\sigma : \mathbb{R} \to \mathbb{R}$, une fonction dite « d'activation » fixée en amont et correspondant souvent à la fonction de Heaviside ($\sigma(t) = \mathbb{1}_{\{t \geq 0\}}$, pour $t \in \mathbb{R}$) ou à la fonction sigmoïde,

$$\sigma(t) \;=\; \frac{1}{1 + e^{-t}}. \tag{5.23}$$

Pour apprendre les paramètres du modèle, il s'agit alors de minimiser la fonction d'objectif donnée dans l'équation (5.8), (associée à la perte quadratique), c'est-à-dire

$$\mathcal{L}(\omega, b) \;=\; \sum_{i=1}^{N} \left(y^i - \omega^{\mathsf{T}} x^i - b \right)^2, \tag{5.24}$$

en notant abusivement $\mathcal{L}$ comme une fonction de ω et b, cette minimisation étant souvent en effectuée à l'aide d'une descente de gradient.

Réseau de neurones artificiels La classification par réseau de neurones reprend en quelque sorte le concept d'agrégation de classifieurs faibles, présent dans les algorithmes d'agrégation *bootstrap* et de *bagging*, mais sans chercher à construire ces derniers de manière itérative. La fonction de classification s'écrit comme une composition de couches,

$$G(x) \;=\; g_m \circ g_{m-1} \circ \cdots \circ g_2 \circ g_1(x), \tag{5.25}$$

où chaque couche g_k, pour $k \in [\![1, m]\!]$, correspond à un ensemble de perceptrons, c'est-à-dire, en reprenant les notations de l'équation (5.22),

$$g_k \ : \ \begin{cases} \mathbb{R}^{n_{k-1}} & \rightarrow & \mathbb{R}^{n_k} \\ z^{k-1} & \mapsto & z^k = \Big(\sigma(\omega_{k,l}^{\mathsf{T}}\, z^{k-1} + b_{k,l})\Big)_{l \in [\![1, n_k]\!]} \end{cases}, \tag{5.26}$$

avec $m \in \mathbb{N}^*$, le nombre de couches, n_k, la dimension de la couche $k \in [\![1, m]\!]$ (avec $n_0 = n$ et $n_m = 1$), σ, la fonction d'activation et où les $w_{k,l}$ et $b_{k,l}$, pour $k \in [\![1, m]\!]$ et $l \in [\![1, n_k]\!]$, constituent l'ensemble des paramètres. Nous avons schématisé cette structure dans la figure 5.6.

Comme pour le perceptron, l'apprentissage des paramètres se fait en utilisant une descente de gradient pour optimiser l'énergie

$$\mathcal{L}(\omega, b) \ = \ \sum_{i=1}^{N} L\Big(y^i, G(x^i; \omega, b)\Big)\,, \tag{5.27}$$

en explicitant ici les paramètres de G et pour L une fonction de perte choisie en amont et supposée facilement différentiable. Notons le calcul du gradient est ici permis grâce à la méthode dite de « backpropagation » [Wer94].

Remarquons par ailleurs le grand nombre de paramètres qui, s'il permet une grande expressivité du modèle, exige de disposer d'un grand nombre de données d'entraînement. Il est alors nécessaire d'effectuer une descente de gradient stochastique, en substitution d'une descente standard, celle-ci consistant à utiliser, à chaque étape de la descente, le gradient d'une « sous-fonction » $\mathcal{L}_R(\omega, b)$,

$$\mathcal{L}_R(\omega, b) \ = \ \sum_{i \in R} L\Big(y^i, G(x^i; \omega, b)\Big)^2\,, \tag{5.28}$$

où R est un sous-ensemble réduit de $[\![1, N]\!]$ pouvant par exemple être choisi aléatoirement à chaque étape.

Réseaux de neurones convolutifs Dans le cadre du traitement d'images, il est possible d'utiliser d'autres fonctions que des perceptrons. Il s'agit d'une part de tirer profit de la structure particulière des images, notamment de l'invariance par translation (une même structure ayant souvent une même signification, peu importe l'endroit où elle apparaît) et le caractère multi-échelle des informations. D'autres part, les images correspondant à des données en très grande dimension (une image RGB de taille 256 par 256 pixels approchant la dimension 200 000), il est aussi nécessaire de pouvoir réduire le nombre de paramètres.

Un réseau de neurones convolutif [FM82, LBD+89] reprend alors la structure d'un réseau de neurones artificiel, en écrivant la fonction de classification comme une composition de différentes couches, chaque couche pouvant être

— une couche de convolution, correspondant à l'application de plusieurs noyaux de convolution sur la sortie de la couche précédente ;
— une couche d'activation, consistant à appliquer une fonction d'activation à chaque élément de la couche précédente ;

— une couche de « pooling », réduisant la taille de la couche précédente en fusionnant des blocs de pixels, souvent en considérant le maximum de leur valeur ;
— ou encore une couche entièrement connectée, correspondant à une couche de perceptron sans fonction d'activation.

De la même manière que pour les réseaux de neurones à base de perceptrons, le nombre de paramètres d'un modèle par réseau de neurones est souvent très grand et demande donc un nombre important de données pour la phrase d'apprentissage. Notons que la multiplication du nombre de couches utilisés démultiplie le nombre de paramètres. Si la fonction d'objectif,

$$\mathcal{L}_R(\Theta) \;=\; \sum_{i=1}^{N} L\Big(y^i, G(x^i;\Theta)\Big)\,, \tag{5.29}$$

peut s'adapter à de très nombreux problèmes, son optimisation s'avère particulièrement difficile, de part l'existence de nombreux minima.

5.2.3 Choix pratique

Il s'agit ici de résoudre un problème de classification à partir des résultats des différents détecteurs primaires, et ce en utilisant un nombre limité de données labellisées, et en demandant une certaine expressivité du modèle choisi pour qu'il puisse s'adapter au maximum de cas d'usage possibles. Résumons alors brièvement les méthodes envisagées.

— La première méthode est d'utiliser une machine à vecteur de support (SVM). Cette méthode est très efficace pour les problèmes séparables linéairement et ne possède que très peu de paramètres, ce qui lui permet d'être utilisable en présence de peu de données. Si l'utilisation de « l'astuce du noyau » lui permet d'être appliquée pour des problèmes non linéaires, il n'est cependant pas certain que ce modèle soit suffisamment expressif pour pouvoir être utilisé dans de multiple cas.
— La seconde méthode envisagée est d'utiliser un algorithme de gradient boosting reposant sur des arbres de régression, en particulier l'algorithme XGBoost. L'expressivité des arbres de régression permet ici d'obtenir un modèle pouvant s'adapter à de nombreux cas, sans nécessiter trop de données. En revanche, ce modèle est sensible aux erreurs dans les données d'apprentissages et peut demander un ajustement manuel des paramètres pour éviter un sur-apprentissage.
— Pour finir, nous avons envisagé d'utiliser des structures de réseaux de neurones. Toutefois, comme nous l'avons précisé précédemment, ces modèles nécessitent, pour pouvoir être bien entraîné, une quantité conséquente de données, dont on ne dispose a priori pas.

Ainsi, en raison de la faible quantité de données labellisées dont nous disposions pour apprendre le modèle agrégé et dans l'idée d'être le plus générique possible, nous avons choisi d'utiliser l'algorithme de gradient boosting XGBoost [CG16].

5.3 Évaluation des performances d'un détecteur

5.3.1 Algorithme de post-processing

Introduction Avant de nous intéresser aux résultats, arrêtons-nous quelques instants sur les principales méthodes de post-processing utilisées parfois pour améliorer les résultats obtenus. Ces méthodes sont utilisées pour rajouter certains aprioris sur le résultat voulu et permettent ainsi de prendre en compte d'éventuelles corrélations spatiales dans le masque de détection, corrélation dont nous n'avons pas tenu compte lors de la fusion des détecteurs primaires.

Notons que l'application de ces différents algorithmes dépend très fortement de l'application voulue, et il est donc préférable de les envisager au cas par cas. Par ailleurs, si ces méthodes sont susceptibles d'améliorer le masque de détection, elles risquent de rendre plus difficile l'interprétation des résultats obtenus. Pour cette raison, les résultats qui sont présentés ici n'utilisent aucun algorithme de post-processing.

Pour la suite, étant donné deux images u et v, définies sur un domaine pixellique $\mathcal{D} \subset \mathbb{Z}^2$, on notera $G : \mathcal{D} \to \mathbb{R}$ le résultat du détecteur, qu'il suffira alors de seuiller à une valeur $\lambda \in \mathbb{R}$ pour obtenir un masque de détection, en notant, pour tout $p \in \mathcal{D}$,

$$G_\lambda(p) \;=\; \mathbb{1}_{\{G(p) \geq \lambda\}} \tag{5.30}$$

le masque de détection associé au seuil λ. Notons que l'on suppose ici que $G(p)$ est d'autant plus grand que le pixel p est jugé comme correspondant à un changement.

Masquage de changements pris en compte séparément Comme discuté dans la section 2.4.2, il est souvent préférable de traiter certains changements à part. C'est par exemple le cas des apparitions de nuages, de zones de saturation, d'ombres, ou encore des changements dus à des effets de parallaxe importante (par exemple causés par la présence simultanée de bâtiments élevés et d'un changement de l'angle de vue entre les deux images).

Il s'agit alors de détecter ces changements, puis de les masquer dans l'image de résultat. En notant C, l'ensemble des pixels de $\mathcal{D}$ correspondant à ces changements, on peut définir une version masquée du résultat G_λ par

$$\tilde{G}_\lambda(p) \;=\; G_\lambda(p)\, \mathbb{1}_{\{p \notin C\}}, \tag{5.31}$$

pour tout seuil $\lambda \in \mathbb{R}$, ce qui contraint bien les détections à n'avoir lieu qu'en dehors de C.

Régularisation à l'aide de filtres morphologiques Comme nous l'avons dit précédemment, la classification pixel à pixel ne prend pas du tout en compte une éventuelle cohérence spatiale des changements, notamment en très haute résolution, où les changements correspondent le plus souvent à des objets don la largeur atteint plusieurs pixels. Il est alors possible de rajouter un certain a priori à l'aide des filtres morphologiques [Ser83, DMBBB08, SSV+13].

Les deux premiers filtres morphologiques sont l'érosion et la dilatation. Pour un ensemble $X \subset \mathbb{Z}^2$, et un réel $\alpha > 0$, on définit alors :

— la dilatation de X, noté $\delta_\alpha(X)$, comme l'ensemble des points situés à une distance d'au plus α d'un point de X, c'est-à-dire

$$\delta_\alpha(X) \;=\; \Big\{x + a, \quad x \in X,\, a \in \mathbb{Z}^2,\, \|a\| \leq \alpha\Big\}, \tag{5.32}$$

— et l'érosion de X, noté $\epsilon_\alpha(X)$, comme l'ensemble des points de X dont tous les points situés à une distance d'au plus α restent dans X, c'est-à-dire

$$\epsilon_\alpha(X) \;=\; \Big\{x \in X, \forall a \in \mathbb{Z}^2, \|a\| \leq \alpha \Rightarrow x + a \in X\Big\}. \tag{5.33}$$

On peut alors définir les opérateurs d'ouverture et de fermeture, noté respectivement γ_α et ϕ_α, par

$$\gamma_\alpha \;=\; \delta_\alpha \circ \epsilon_\alpha \qquad \text{et} \qquad \phi_\alpha \;=\; \epsilon_\alpha \circ \delta_\alpha. \tag{5.34}$$

Nous avons représenté, dans la figure 5.7, les effets de ces quatre opérateurs sur un exemple de masque binaire.

Comme on peut le voir, la dilatation et l'érosion ont un effet systématique d'agrandissement et de rétrécissement des différents changements. L'opérateur d'ouverture a ici pour effet de ne conserver que les structures suffisamment étendues, en masquant alors toutes les détections trop fines. Au contraire, l'opérateur de fermeture va combler les zones vides à l'intérieur et entre les différentes détections.

Notons qu'il est possible de définir ces filtres directement sur la fonction G, les opérateurs d'érosion et de dilatation correspondant alors à appliquer respectivement un minimum ou un maximum local :

$$\delta_\alpha(G)(p) \;=\; \max_{\substack{q \in \mathcal{D} \\ \|p-q\| \leq \alpha}} G(q) \qquad \text{et} \qquad \epsilon_\alpha(G)(p) \;=\; \min_{\substack{q \in \mathcal{D} \\ \|p-q\| \leq \alpha}} G(q). \tag{5.35}$$

5.3.2 Mesures de performance

Nous nous intéressons ici aux moyens d'évaluer les performances d'un algorithme de détection de changement. Remarquons toutefois qu'une telle mesure de performance n'a qu'un intérêt limité dans le cadre de l'aide à la photo-interprétation. Il est en effet très difficile d'apprécier la pertinence d'un changement sans la connaissance métier d'un photo-interprète et donner un score stricte à algorithme peut ainsi sembler peu pertinent. Toutefois, un score est souvent nécessaire pour pouvoir comparer les différents modèles de manière objective.

5.3.2.1 Mesures pixelliques

Formalisation du problème Pour évaluer les performances, on suppose que l'on dispose d'un ensemble de test, formalisé de manière similaire à l'ensemble d'entraînement donné dans l'équation (5.5),

$$\mathcal{E}_{\text{test}} \;=\; \Big\{\Big(x^i, y^i\Big), i \in [\![1, N]\!]\Big\} \;\subset\; \mathbb{R}^n \times \{-1, 1\}, \tag{5.36}$$

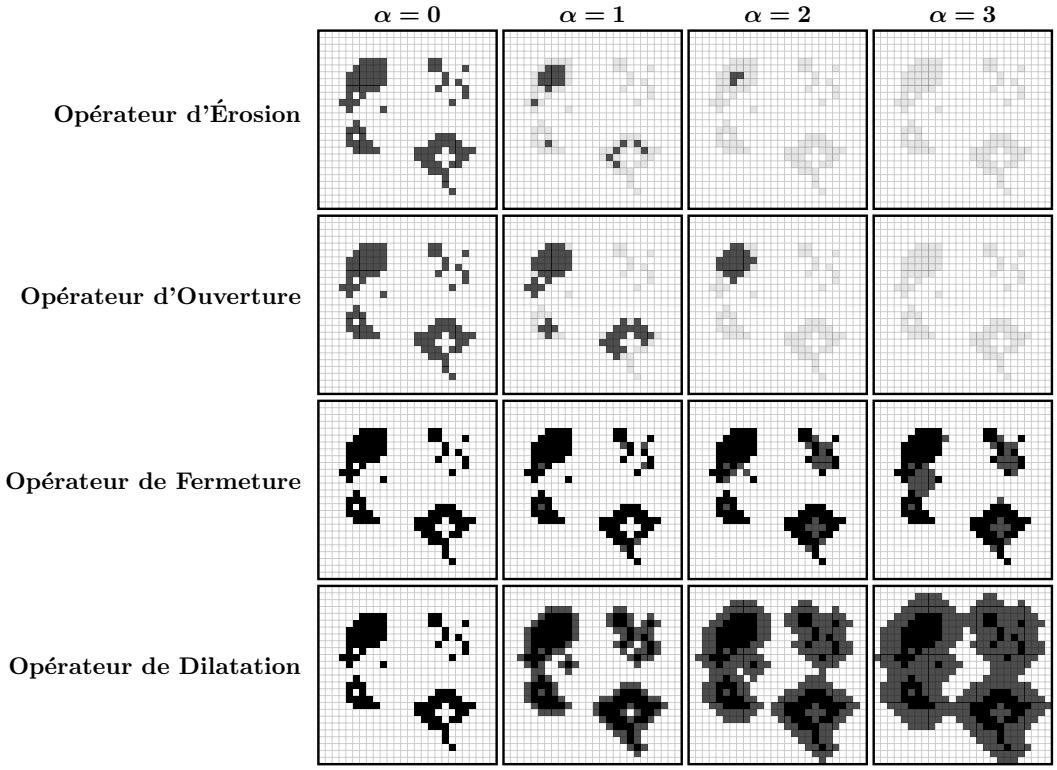


Figure 5.7 – Représentation des filtres morphologiques d'érosion, d'ouverture, de fermeture et de dilatation, appliquée à un même exemple en utilisant différentes valeur d'échelle α. Nous affichons ici en noir ou gris foncé les résultats des filtres et en gris clair ou en noir, la trace du masque initial. On voit ici que l'opérateur d'ouverture ne conserve que les structures suffisamment « grosses », tandis qu'appliquer l'opérateur de fermeture revient à combler les espaces enclavés dans une zone de détection.

pour $N \in \mathbb{N}$, la valeur $y^i = 1$ représentant un changement. On souhaite alors évaluer les performances d'une détection de changements, ce détecteur étant vu comme une fonction G, de $\mathbb{R}^n$ dans $\mathbb{R}$, $G(x)$ étant d'autant plus grand que le point $x \in \mathbb{R}$ est considéré comme un changement.

Trouver une mesure de performance revient alors à construire une fonction associant à tout détecteur G un score quantifiant à quel point $G(x^i)$ correspond bien à y^i, pour tout $(x^i, y^i) \in \mathcal{E}_{\text{test}}$.

Vrais/Faux Positifs/Négatifs Les mesures de performance classiques reposent sur quatre grandeurs, les nombres de vrais positifs (VP), faux positifs (FP), vrais négatifs (VN) et faux négatifs (FN), l'aspect positif ou négatif renvoyant à la décision du détecteur, tandis que l'aspect vrai ou faux correspond à la véracité de cette réponse,

ce que l'on peut formaliser de la manière suivante

$$\begin{aligned} \mathrm{VP}(\lambda) &= \sum_{y^i=1} \mathbb{1}_{\{G(x^i)\geq\lambda\}}, \qquad & \mathrm{FP}(\lambda) &= \sum_{y^i=-1} \mathbb{1}_{\{G(x^i)\geq\lambda\}}, \\ \mathrm{VN}(\lambda) &= \sum_{y^i=-1} \mathbb{1}_{\{G(x^i)<\lambda\}}, \qquad & \mathrm{FN}(\lambda) &= \sum_{y^i=1} \mathbb{1}_{\{G(x^i)<\lambda\}}, \end{aligned} \tag{5.37}$$

pour $\lambda \in \mathbb{R}$ la valeur de seuil choisie. Par la suite, nous ne noterons plus les dépendances en la valeur du seuil λ pour ne pas alourdir les notations.

Précision, rappel et spécificité À partir des grandeurs précédentes, on définit classiquement plusieurs mesures :

— le rappel, correspondant à la proportion des changements réels effectivement détectés, et qui s'exprime sous la forme

$$\text{Rappel} = \frac{\mathrm{VP}}{\mathrm{VP}+\mathrm{FN}} \; ; \tag{5.38}$$

— la précision, représentant à la proportion des changements annoncés correspondant réellement à des changements, à savoir

$$\text{Précision} = \frac{\mathrm{VP}}{\mathrm{VP}+\mathrm{FP}} \; ; \tag{5.39}$$

— et la spécificité, correspondant à la proportion de non-changements bien labellisées, c'est-à-dire

$$\text{Spécificité} = \frac{\mathrm{VN}}{\mathrm{VN}+\mathrm{FP}}. \tag{5.40}$$

Si la spécificité est rarement utilisée dans l'évaluation des performances de détecteurs d'objets, elle devient pertinente dans le cadre de la détection de changements pour l'aide à la photo-interprétation, où assurer qu'une partie de l'image ne possède aucun changement pertinent constitue une information utile

Notons que ces trois mesures ne sont pas pertinentes prises indépendamment. En effet, un détecteur détectant systématiquement tous les points aura nécessairement un rappel de 1, mais une précision et une spécificité très faible. Au contraire, un détecteur ne détectant presque rien aura certainement une précision et une spécificité élevées, mas un rappel très faible.

Mesures classiques de performance Il existe plusieurs mesures prenant en compte à la fois le rappel et la précision, comme le F1-score, le G-score ou encore la statistique J de Youden, définies respectivement par

$$\text{F1-score} = 2\,\frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}}, \tag{5.41}$$

$$\text{G-score} = \sqrt{\text{Précision} \times \text{Rappel}}, \tag{5.42}$$

et

$$\text{Statistique J de Youden} = \text{Précision} + \text{Rappel} - 1. \tag{5.43}$$

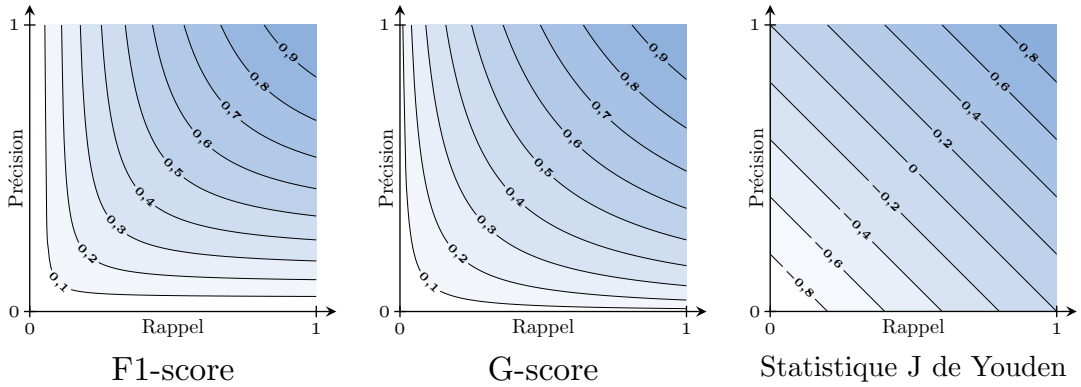


Figure 5.8 – Représentation des mesures de F1-score, G-score et Statistique J de Youden, en fonction des valeurs de précision et de rappel. Dans chaque cas, les courbes tracées représentent différentes lignes de niveau des mesures, c'est-à-dire des courbes passant par tous les points partageant une même valeur, une couleur foncée représentant des valeurs élevées.

Nous les avons représentées dans la figure 5.8.

Le désavantage de ces mesures est qu'elles demandent de fixer en amont une valeur de seuil, ce que l'on préfère ici éviter dans le cas de l'aide à la photo-interprétation, pour permettre à l'utilisateur de faire varier le niveau de fiabilité du détecteur.

Courbe de performances Un autre moyen classique d'évaluer les performances d'un algorithme est de faire varier la valeur de seuil et d'observer les courbes de performances obtenues. Les deux courbes les plus utilisées correspondent à la courbe de la précision en fonction du rappel, et la courbe du rappel en fonction de « $1-$spécificité », cette dernière étant appelée courbe ROC (pour « Receiver Operating Characteristic »).

L'intérêt d'une courbe est de pouvoir évaluer visuellement les performances pour différentes manières de seuiller. Dans le cas de la courbe précision-rappel, on peut en effet chercher à privilégier la précision, si l'on estime qu'il est préférable de ne pas détecter des changements que se tromper en en détectant un, à privilégier le rappel, si au contraire, on souhaite détecter le maximum de changements, quitte à avoir un grand nombre de faux positifs, ou encore chercher un compromis entre les deux.

Il est possible d'obtenir une mesure de performance quantitative à partir de ces courbes en considérant à l'aire sous la courbe. Dans le cas de la courbe précision-rappel, cette valeur est souvent appelée « Précision moyenne », dans le cas de la courbe ROC, il s'agit de « l'aire sous la courbe ROC » ou simplement « AUC ».

Par la suite, nous privilégions la représentation des performances au travers de la courbe précision-rappel.

5.3.2.2 Mesures objets

Introduction Si les mesures pixelliques sont parfaitement adaptés à de nombreux problèmes de classification, dans le cas d'un problème de classification dans des images,

elles ne prennent pas du tout en compte la structure spatiale du résultat. Si l'on cherche, par exemple, à détecter deux objets de même taille, une mesure pixellique donnera la même performance à un algorithme détectant un seul des objets et à un algorithme détectant chaque objet à moitié.

Pour remédier à ce problème, il est possible de regrouper les pixels détectés en différents « objets ». Formellement, il s'agit de construire une partition de l'ensemble

$$\left\{ i \in [\![1, N]\!] \, , G(x^i) \geq \lambda \right\} , \tag{5.44}$$

noté $\{O_j\}_{j \in [\![1,M]\!]}$, où $M \in \mathbb{N}$ est le nombre d'objets détectés, par exemple en effectuant une décomposition de l'ensemble en composante connexe. En décomposant de même l'ensemble des pixels à détecter,

$$\left\{ i \in [\![1, N]\!] \, , y(x^i) = 1 \right\} , \tag{5.45}$$

en un ensemble d'objets $\{T_k\}_{k \in [\![1,M']\!]}$, où $M' \in \mathbb{N}$ est le nombre d'objets à détecter, il s'agit alors d'évaluer les objets détectés vis-à-vis des objets à détecter.

Précision et rappel « objet » Une manière usuelle de procéder et de redéfinir les notions de précision et rappel à l'échelle des objets. Il suffit de définir une fonction s qui à deux objets A et B, sous ensembles de $[\![1, N]\!]$, représente à quel point A « détecte » bien B. La mesure dite de « Intersection sur Union », définie alors par

$$s(A, B) \;=\; \frac{\mathrm{Card}(A \cap B)}{\mathrm{Card}(A \cup B)}, \tag{5.46}$$

est souvent utilisée dans le cadre de la détection d'objets, cette mesure exigeant que les deux ensembles soient le plus égaux possible. On peut alors redéfinir la notion de précision et de rappel dans ce nouveau cadre. En considérant qu'un objet détecté O_j, où $j \in [\![1, M]\!]$, est un vrai positif lorsqu'il existe un objet à détecter T_k, où $k \in [\![1, M']\!]$, vérifiant $s(O_j, T_k) \geq C_0$, pour C_0 une constante donnée, on peut définir la « précision objet » par

$$\text{Precision} \;=\; \frac{1}{M} \,\mathrm{Card}\left(\left\{ j \in [\![1, M]\!] \, , \max_{k \in [\![1,M']\!]} s(O_j, T_k) \;\geq\; C_0 \right\} \right) . \tag{5.47}$$

De manière similaire, on peut définir le « rappel objet » par

$$\text{Rappel} \;=\; \frac{1}{M'} \,\mathrm{Card}\left(\left\{ k \in [\![1, M']\!] \, , \max_{j \in [\![1,M]\!]} s(T_k, O_j) \;\geq\; C_0 \right\} \right) . \tag{5.48}$$

Toutefois, cette recherche d'une égalité stricte entre les ensembles, imposée par la fonction s, pose problème dans le cas de la détection de changements. En effet, supposons que l'on cherche à détecter l'apparition de deux véhicules adjacentes, considérés comme un seul changement T, et que l'algorithme détecte un changement par véhicule, notés O_1 et O_2. Même si l'algorithme détecte bien l'ensemble des changements ($T = O_1 \cup O_2$), puisque $s(O_1, T) = s(O_2, T) \simeq \frac{1}{2}$, si C_0 est pris trop stricte, l'algorithme obtiendra une précision et un rappel nuls.

Pour éviter ce problème, nous prendrons une autre définition de la précision et du rappel. Nous considérons qu'un objet détecté O_j, où $j \in [\![1, M]\!]$, est un vrai positif lorsqu'il intersecte suffisamment l'union de tous les objets à détecter, et définissons ainsi

$$\text{Precision}' = \frac{1}{M} \text{Card}\left(\left\{ j \in [\![1, M]\!] \,,\, \frac{\text{Card}(O_j \cap (\cup_k T_k))}{\text{Card}(O_j)} \geq C_0 \right\}\right). \tag{5.49}$$

De manière similaire, on peut définir

$$\text{Rappel}' = \frac{1}{M'} \text{Card}\left(\left\{ k \in [\![1, M']\!] \,,\, \frac{\text{Card}(T_k \cap (\cup_j O_j))}{\text{Card}(T_k)} \geq C_0 \right\}\right). \tag{5.50}$$

Définis comme cela, l'exemple des deux véhicules ne posent plus de problème et abouti à une précision et un rappel de 1.

Remarquons qu'il n'est ici pas possible de définir une spécificité, puisqu'aucun objet correspondant à des non-changements n'a été défini. Remarquons de plus que le nombre d'objets détectés, M, dépend ici fortement de la valeur de seuil λ, notamment lorsque les objets sont construits comme les ensembles connexes du masque binaire de détection correspondant à un seuil. Pour des valeurs élevées de λ, M risque alors de diminuer fortement et de rendre moins pertinent la mesure de précision objet.

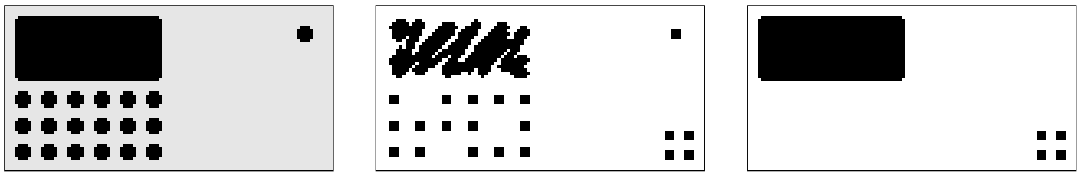
5.3.2.3 Limitation des mesures classique

Le cas de la détection de changements Nous nous intéressons ici à la pertinence des mesures pixelliques et objets dans le cadre de la détection de changements. Tout d'abord, il est important de différencier l'évaluation de la performance d'un algorithme de détection de changements avec celui d'une détection d'objet. En effet, dans le cadre de ce dernier, le résultat recherché peut naturellement se décomposer en un ensemble d'objets et il apparaît pertinent de chercher à faire correspondre un objet à détecter avec un groupement de pixels détectés. Pour la détection de changement, une telle décomposition n'est pas évidente, certains changements pouvant être décomposés en plusieurs sous-changements. Toutefois, les adaptations des mesure de précision et de rappel permettent en partie de prendre cela en compte.

Par ailleurs, les différents changements appréhendés peuvent présenter des tailles différentes, et utiliser une mesure pixellique aura tendance à augmenter l'importance donnée à la détection des gros changements, tandis qu'une mesure objet privilégiera celle des petits. L'importance relative que l'on souhaite donner aux différents changements n'étant pas nécessairement corrélée à leur taille, la pertinence des approches pixelliques ou objets dépend alors des cas d'usage.

On montre dans la figure 5.9 des exemples de résultats illustrant la difficulté de choisir une mesure adaptée au problème de détection de changements.

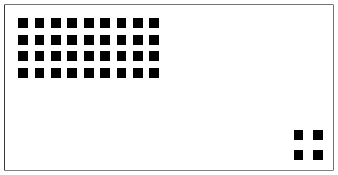
Prise en compte des incohérences spatiales Les éventuels décalages spatiaux présents entre les images étudiées peuvent aussi entraîner des problèmes de décalages spatiaux entre le masque de « vérité terrain », présentant les changements à détecter, et les masques de changements rendus par les algorithmes.



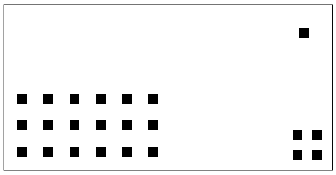
Changements à détecter

Résultat A

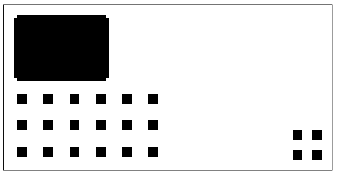
Résultat B



Résultat C



Résultat D



Résultat E

	Précision		Rappel		F1-score		G-score		Statistique J	
	pixel	objet	pixel	objet	pixel	objet	pixel	objet	pixel	objet
A	0.947	0.809	0.505	0.850	0.659	0.829	0.692	0.829	0.453	0.659
B	**0.961**	0.200	**0.691**	0.050	**0.804**	0.080	**0.815**	0.100	**0.653**	-0.750
C	0.900	**0.900**	0.250	0.050	0.391	0.094	0.474	0.212	0.150	-0.050
D	0.826	**0.826**	0.132	**0.950**	0.227	**0.883**	0.330	**0.885**	-0.041	**0.776**
E	**0.953**	**0.826**	**0.569**	**0.950**	**0.713**	**0.883**	**0.737**	**0.885**	**0.523**	**0.776**

Figure 5.9 – Mesures de performances pixelliques et objets obtenues à partir de cinq masques de résultats. On présente ici successivement l'image binaire représentant les changements que l'on souhaite détecter, ainsi que cinq autres images représentant cinq résultats possibles, le tout suivi d'un tableau présentant les mesures de précision, rappel, F1-score, G-score et Statistique J de Youden. Pour chaque mesure, les deux résultats ayant les performances les plus élevées sont indiquées en gras. Pour les mesures objets, la valeur seuil C_0 (voir équation (5.50)) est prise à 1/3. Les résultats A et B montrent le principal problème des approches pixelliques : A présente une précision et un rappel pixellique plus faible que le B, ce dernier ne détectant aucun des petits changements présents. Les résultats D et E révèlent un premier problème de l'approche objet, en donnant à ces deux résultats des mesures identiques, lorsque l'on pourrait préférer détecter le plus gros objet. Les résultats B et C montrent un second problème de cette approche : en détectant moins de pixels correspondant aux « vrais » changements, on peut séparer un des « objets » du résultat B en plusieurs parties, dans le résultat C, ce qui diminue le rappel et la précision pixellique mais augmente la précision objet.

Si les deux grilles de pixels correspondent à des référentiels géographiques différents, il faudrait théoriquement s'assurer que les changements annotés et les changements détectés soient tous les deux indiqués dans une même grille ou un même référentiel. Malheureusement, la transformation entre les deux images n'est souvent pas exactement connue et le plus souvent, les masques obtenus sont construit en mélangeant les référentiels des deux images. Il peut ainsi exister des décalages spatiaux entre les masques de résultats obtenus et la version annotée manuellement, ce qui peut impacter les mesures de performances pixelliques, ainsi que les mesures objets dans une moindre mesure.

Performance pour la photo-interprétation Comme nous l'avons déjà précisé, la motivation centrale de cette thèse sur la détection de changements est l'aide à la photo-interprétation. Le fait que le masque de changements obtenu ne soit pas utilisé directement, mais serve de support au travail d'un photo-interprète modifie alors les attentes que l'on a de la mesure de performances. Il est aussi plus important de détecter un changement isolé, qu'un changement situé à proximité d'autres changements détectés. De manière similaire, lorsque l'on détecte partiellement un changement, on peut préférer une répartition homogène sur la surface d'un changement plutôt que d'en détecter intégralement une petite partie.

Ces considérations ne sont toutefois pas prises en compte, ni par les mesures pixelliques, ni par les mesures « objet ». On peut voir cela dans la figure 5.9 : le résultat A peut apparaître comme le meilleur résultat, détectant à la fois le « gros » changement, partiellement mais dans son intégralité, le groupement de petits changements ainsi que le changement isolé, sans toutefois être considéré comme particulièrement bon, au vu des mesures considérées, devant les autres résultats présentés.

Mesure utilisée Il apparaît ainsi que les mesures classiques ne sont pas réellement pertinentes pour envisager les performances d'un algorithme de détection de changements. Toutefois, il peut tout de même être intéressant de comparer quantitativement les performances des différents algorithmes. Nous considérons donc par la suite deux mesures, la précision pixellique ainsi que le rappel objet, ces derniers subissant peu les problèmes exposés ici.

5.4 Détecteur agrégé sur un cas réel

5.4.1 Description des données

Introduction Pour tester l'algorithme de détection de changements par agrégation de détecteurs primaires, nous avons cherché à nous placer dans les conditions d'utilisation d'un photo-interprète cherchant à étudier une nouvelle image. Pour cela, nous avons sélectionné un petit ensemble de 13 images, de taille variant entre 500×500 et 800×800 pixels et pour chacune d'elles, nous avons réalisé un masque binaire correspondant aux différents changements que nous souhaitions détecter. Ces images, ainsi que leur masque de changements sont affichées dans la figure 5.10. Nous avons ensuite entraîné un détecteur agrégé sur ces données.

Sur les 13 paires d'images utilisées, notons que les 8 premières proviennent d'une même paire de macro-images, les 5 dernières provenant d'une seconde paire. De plus, chaque image possède ici une même résolution spatiale de 50 cm par pixel.

Choix des changements En ce qui concerne les changements que l'on cherche à détecter, nous avons décidé de détecter ici toutes les apparitions et disparitions d'objets isolés (véhicule, bâtiment, etc), sans détecter toutefois d'éventuels changements du sol, comme des apparitions de routes ou de taches.

Remarquons que les 13 images d'entraînement, présentées dans la figure 5.10, ne présentent que très peu de changements pertinents. Nous avons fait ce choix délibérément, pour correspondre au cas d'un utilisateur ne disposant que d'un temps limité. De plus, lors de la réalisation des masques de changement, nous n'avons pas cherché à segmenter de manière exacte, au pixel près, les différents changements pertinents (selon la définition que nous nous sommes donnée), et ce dans la même optique.

Nous avons toutefois fait attention à ce que les différentes images présentent bien certains changements globaux (ou non locaux) qui risque d'être présents lors de l'application du modèle, notamment :

— l'augmentation du contraste (par exemple dans les paires 3 et 8) ;
— l'éclaircissement ou assombrissement de l'ensemble de l'image (par exemple dans les paires 6 et 11) ;
— l'apparition de multiples routes et chemins (comme dans la paire 2) ;
— ou encore l'apparition de taches sur le sol (comme dans les paires 10 et 11).

5.4.2 Résultats obtenus

Images de test utilisées Nous avons appliqué le modèle agrégé de détection de changements, appris sur les données représentées dans la figure 5.10, sur différentes paires d'images réelles, présentant respectivement ainsi qu'une scène rurale relativement simple, où l'on trouve cependant de nombreuse apparition de tache sur le sol, une scène urbaine en présence de parallaxe et de changements au niveau des ombres (cette paire correspond à celle présentée dans le chapitre 2, figure 2.3.1), une scène de chantier où se trouvent de multiple changements non pertinents au niveau du sol, ainsi qu'une scène désertique où l'on trouve entre autres de grandes variations globales d'intensité et de contraste. Il est important de noter que, sur ces quatre paires d'images, seule la première est extraite d'une des paires de macro-images ayant servi pour l'apprentissage du modèle. Elles possèdent toutes, cependant, une même résolution de 50 cm par pixel.

Convention d'affichage des résultats Dans les quatre figures présentant les résultats obtenus, nous affichons successivement sur la première ligne les deux images étudiées, sur une deuxième ligne la valeur absolue de leur différence (d'autant plus foncée que la différence est importante) ainsi que les résultats obtenus. Pour faciliter leur interprétation, ces résultats sont représentés en noir sur une version éclaircis d'une segmentation manuelle des différents changements, dans laquelle les changements jugés pertinents sont indiquées à l'aide d'une couleur plus sombre et le plus souvent d'une auréole. Nous présentons enfin, dans la dernière ligne, deux agrandissement de la paire

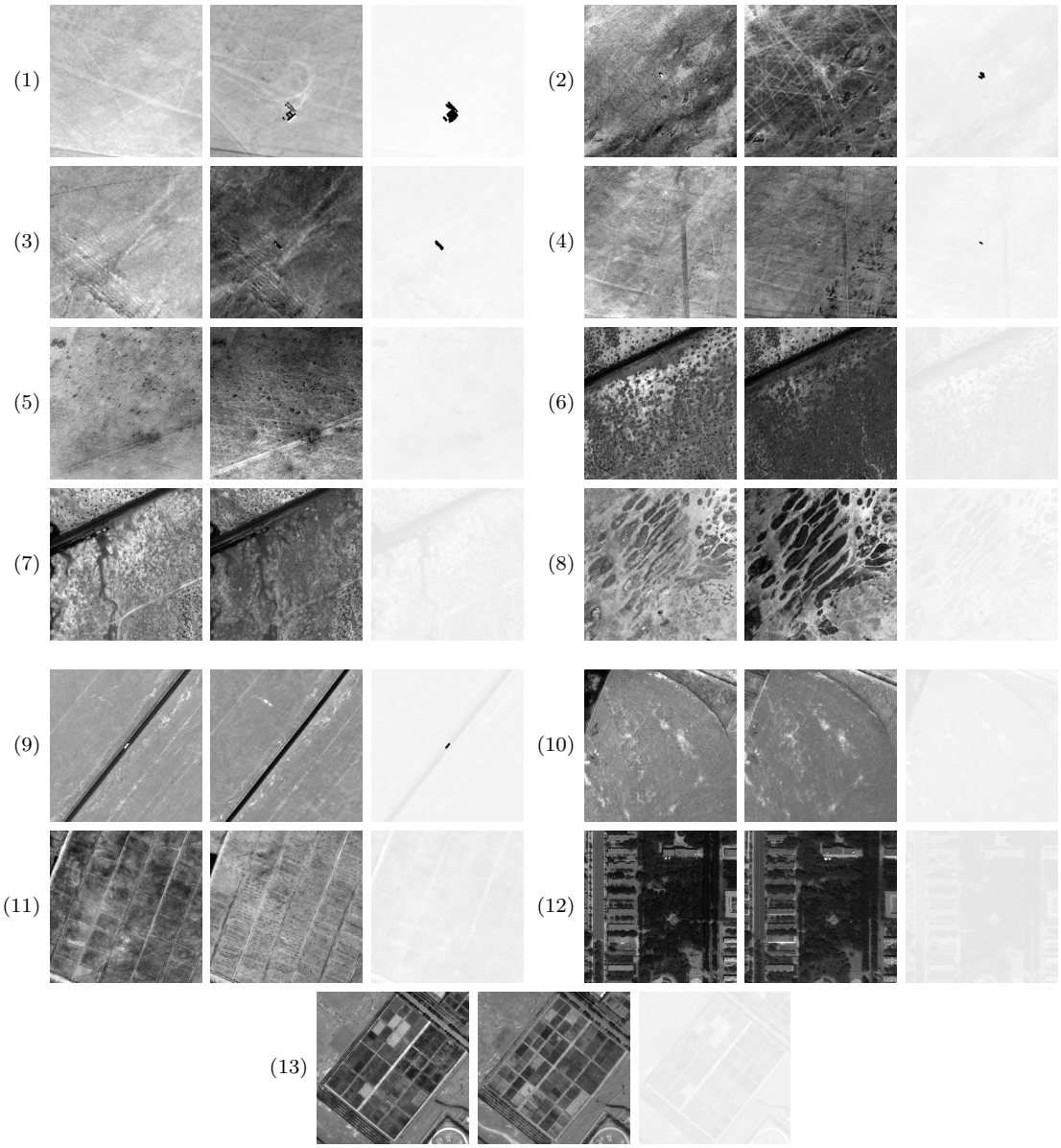


Figure 5.10 – Représentation des données utilisées pour l'apprentissage d'un modèle de changements spécifiques. Pour chacune des 13 paires d'images utilisées, nous avons représenté les deux images suivies du masque de changement construit pour l'occasion, les changements étant alors affiché en noir sur une version éclaircie de la première image de la paire. Il s'agit ici de scènes rurales présentant presque toutes des incohérences spatiales dues à une variation d'angle de vue du satellite entre les deux images, mais ne présentant aucun cas de occlusion importante, cette dernière nécessitant une prise en charge spécifique. Notons que le nombre de changements annotés est ici volontairement bas pour correspondre au faible temps que posséderait un utilisateur pour leur labellisation.

étudiée, présentant successivement les deux images suivies du résultat obtenu, affiché en noir sur une version éclaircie de la deuxième image.

Scène rurale La première paire d'images sur laquelle nous testons le modèle agrégé de détection de changements est présentée dans la figure 5.11, en même temps que les résultats obtenus. Il s'agit de la paire la plus facile. Nous résumons, dans le tableau ci-dessous, les différents changements présents, en précisant leur pertinence vis-à-vis de la tâche que l'on s'est fixée, ainsi que les résultats obtenus.

Changement	**Pertinence**	**Résultat du détecteur**
Apparitions de petits objets (avion, véhicules, petit bâtiment)	oui	Un changement manquant
Apparitions de taches de forme étrange	questionnable	Tout est détecté
Variations d'ombres	non	Plusieurs détections
Multiples apparitions de petites taches	non	Quelques détections
Apparitions de routes et chemins	non	Aucune détection

Les deux difficultés principales de cette scène sont les variations d'ombres ainsi que les multiples apparitions de taches, que nous avons jugées non pertinente lors de l'élaboration de l'ensemble d'apprentissage. On parvient ici à détecter l'intégralité des changements pertinents, à l'exception d'un, en limitant grandement les fausses détections liées aux taches. La majeur partie des fausses détections (c'est à dire des détections de changements non pertinent) sont dues à des variations d'ombres. Comme nous l'avons mentionné précédemment, les grandes variations d'ombres constituent un réel problème de la détection de changements, et il est souvent préférable de les traiter séparément.

Notons la pertinence questionnable de deux taches, qui pourraient correspondre à des apparitions d'objets. Ces apparitions illustrent bien à quel point il peut être difficile d'évaluer la pertinence des différents changements.

Scène urbaine Dans la figure 5.12, nous présentons les résultats obtenus sur la deuxième paire d'images de test. Cette paire d'image correspond à une zone urbaine et présente d'importantes variations d'ombres ainsi que des effets de parallaxe, particulièrement visible au niveau du bâtiment au nord-ouest de l'image, qui restent toutefois raisonnables dans le reste de l'image. Cette scène présente un très grand nombre de changements et il est difficile d'en faire la liste exhaustive. Nous avons toutefois donné, dans le tableau ci-dessous, un résumé des principaux changements en précisant leur pertinence et les résultats obtenus.

Changement	**Pertinence**	**Résultat du détecteur**
Incendie des bâtiments centraux	oui	Détection partielle d'un bâtiment
Apparition et modification d'objets (toit et « objets blancs »)	oui	Tout est détecté
Apparitions de véhicules hors des routes et parking	oui	Détection partielle
Apparitions et disparitions de véhicules sur les routes et parking	non	Quelques fausses détections
Variation d'ombre et parallaxe	non	Un cas de fausses détections

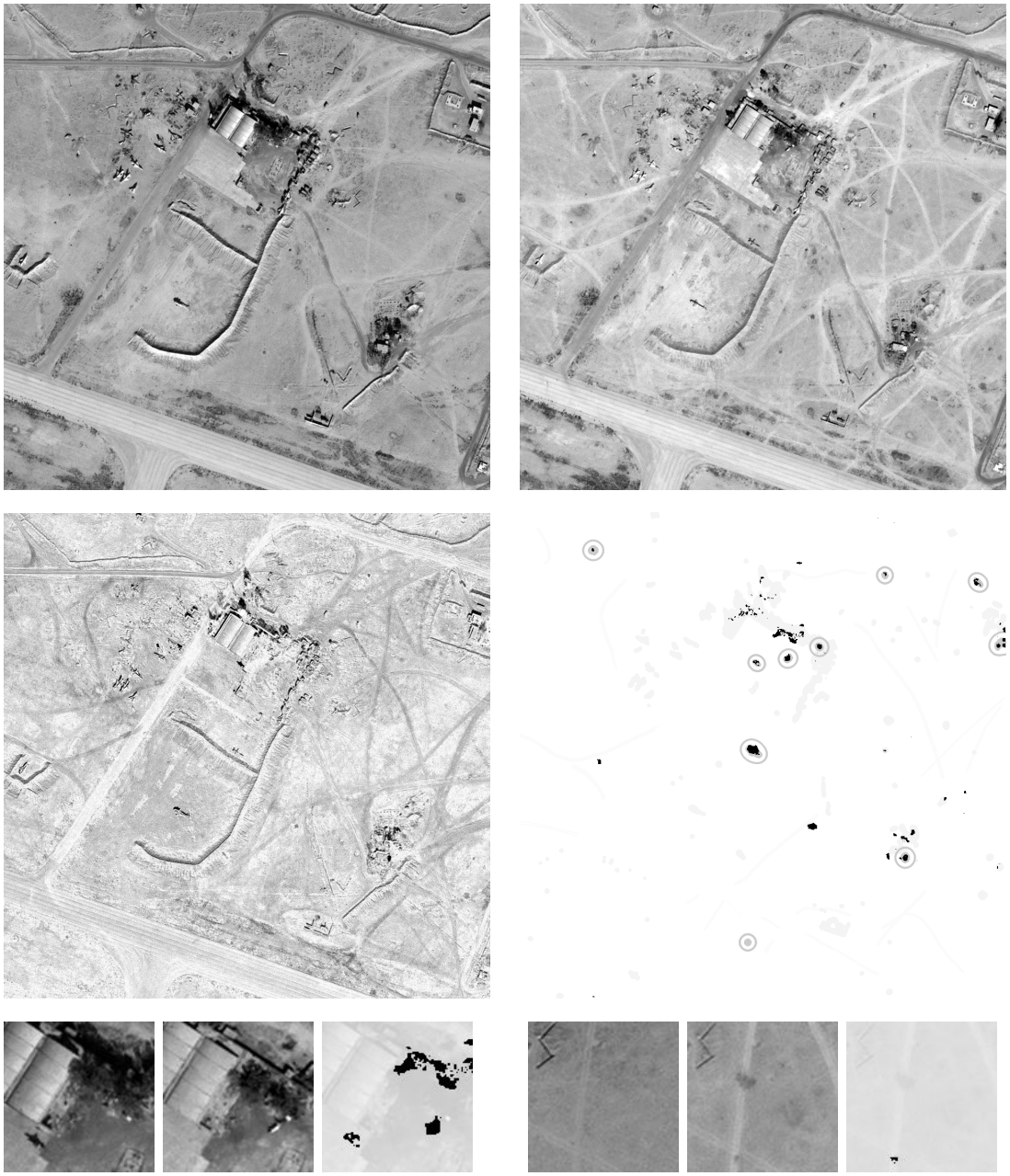


Figure 5.11 – Résultats obtenus sur une paire d'image réelle en utilisant le modèle agrégé de changements entraînée sur les données explicitées dans la section 5.4.1. Nous utilisons les conventions d'affichage données page 198. Le détecteur parvient ici à détecter presque tous les changements pertinents pour un nombre réduit de fausses détections, dont une partie conséquente es due à la variation des ombres entre les deux images. Dans le premier agrandissement, notons l'apparition d'une tache de forme étrange, dont il est difficile d'estimer la pertinence. Le deuxième agrandissement illustre les multiples apparitions de taches sur le sol.

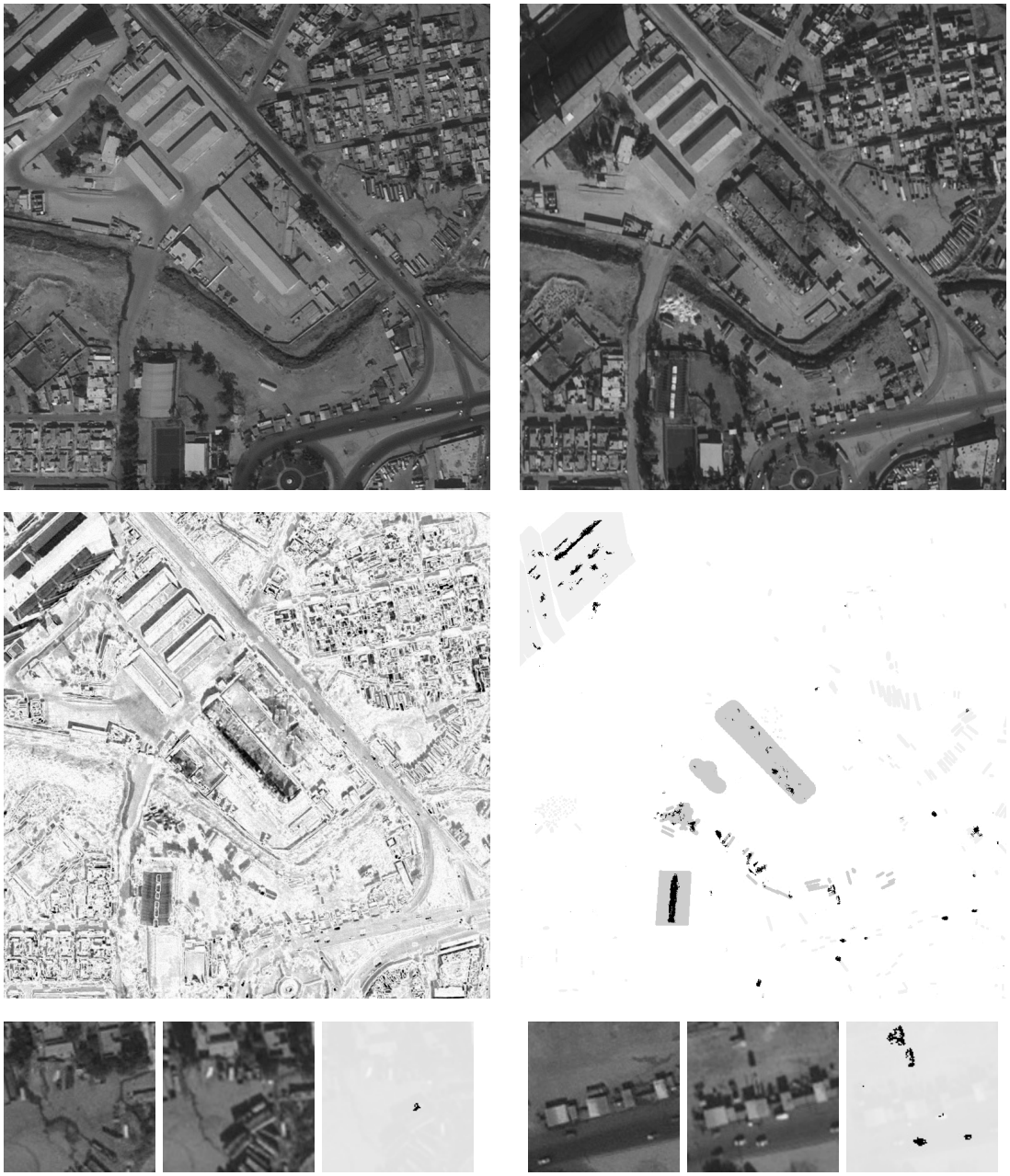


Figure 5.12 – Résultats obtenus sur une paire d'image réelle en utilisant le modèle agrégé de changements entraînée sur les données explicitées dans la section 5.4.1. Nous utilisons les conventions d'affichage données page 198. Le modèle parvient ici à être robuste aux variations d'ombres dans les quartiers résidentiels (coin haut-droit et bas-gauche des images), particulièrement visibles dans l'image de différence, sans détecter toutefois les apparitions de véhicules dans ces régions. S'il détecte la plupart des changements marqués comme pertinents, il faut cependant remarquer que ceux visibles uniquement par une variation d'intensité, comme l'incendie du « petit bâtiment » au centre de l'image, ne sont pas détectés.

Tout d'abord, remarquons que le modèle parvient à ne détecter aucun des changements issus des variations d'ombres dans les quartiers résidentiels (dans les deux coins supérieur droit et inférieur gauche du domaine). Il ne détecte cependant pas non plus les différentes apparitions qui s'y trouve. Par ailleurs, on réussit à détecter l'incendie du bâtiment au centre, l'apparition d'objets blancs ainsi que le rabattement du toit, marqué comme particulièrement pertinents dans le masque de segmentation manuel de la figure 5.12. Notons qu'il ne parvient pas à détecter l'incendie du bâtiment annexe, ce dernier n'étant visible qu'au travers d'un changement d'intensité à peu près constant sur tout le bâtiment. Ceci se retrouve aussi dans le coin haut-gauche, au niveau du changement correspondant à une parallaxe et une variation d'ombre très importante, où le modèle ne détecte des changements qu'au niveau des bords de l'immeuble. Ceci s'explique par la robustesse aux changements de contraste imposé par l'ensemble d'apprentissage.

Compte tenu de la complexité de la scène, il est difficile ici d'évaluer réellement les performances du détecteurs obtenu. Néanmoins, il nous semble particulièrement intéressant de réussir à souligner les changements notés comme pertinents en détectant aussi peu les nombreuses apparitions et disparitions de véhicules ainsi que les variations d'ombre.

Scène de chantier Les résultats obtenus sur la troisième paire sont présentés dans la figure 5.13. Comme précédemment, nous donnons, dans le tableau suivant, un résumé des principaux changements intervenant dans cette scène en précisant leur pertinence et les résultats obtenus.

Changement	**Pertinence**	**Résultat du détecteur**
Disparition de véhicules de chantier	oui	Quelques non-détections
Apparition et changement de couleur des routes	non	Aucune détection
Disparition de débris	non	Détections partiels
Modification du terrain	non	Aucune détection

Avant toute chose, notons que les débris ne sont pas considérés ici comme pertinents puisqu'ils ne sont pas isolés.

Les résultats obtenus ici sont relativement bon, puisque le détecteur agrégé s'avère complètement robuste à la modification du terrain et des routes. Il parvient à détecter presque tous les véhicules (au moins l'habitacle), en ne détectant qu'une faible proportion des débris.

Scène désertique Nous présentons la dernière paire étudiée dans la figure 5.14. Comme précédemment, nous donnons un tableau présentant un résumé des principaux changements intervenant dans cette scène.

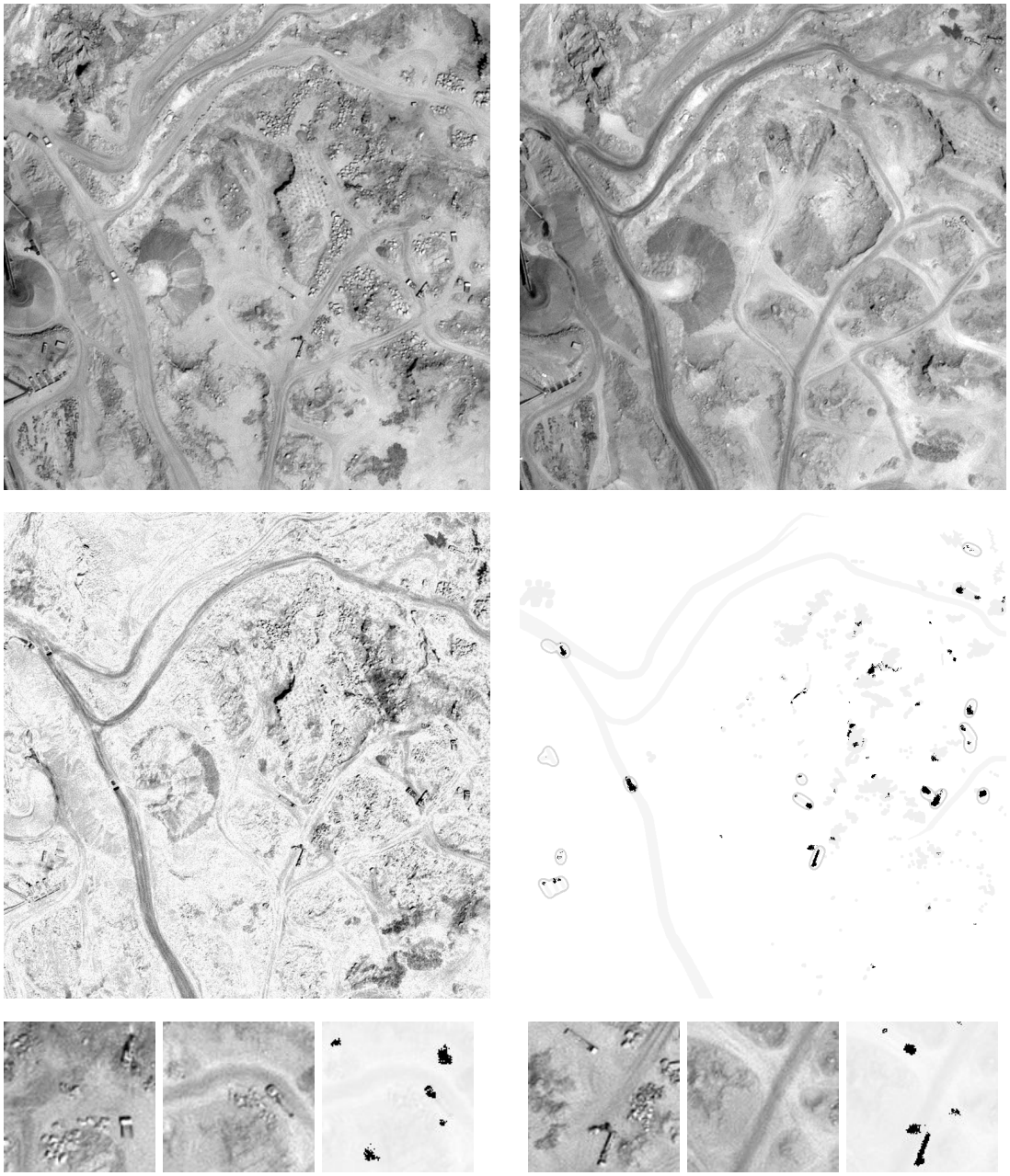


Figure 5.13 – Résultats obtenus sur une paire d'image réelle en utilisant le modèle agrégé de changements entraînée sur les données explicitées dans la section 5.4.1. Nous utilisons les conventions d'affichage données page 198. Cet exemple est particulièrement intéressant du fait qu'il contient beaucoup de changements tout en restant simple à décrire, la majorité des changements consistant en la disparition de gravats et la modification du terrain. Les agrandissements permettent ici de voir certaines confusions de l'algorithme sur la distinction entre véhicule et débris. Ici, l'algorithme parvient à détecter la majorité des disparitions de véhicules, tout en étant robuste à l'assombrissement des routes et en limitant les détections de disparitions de débris.

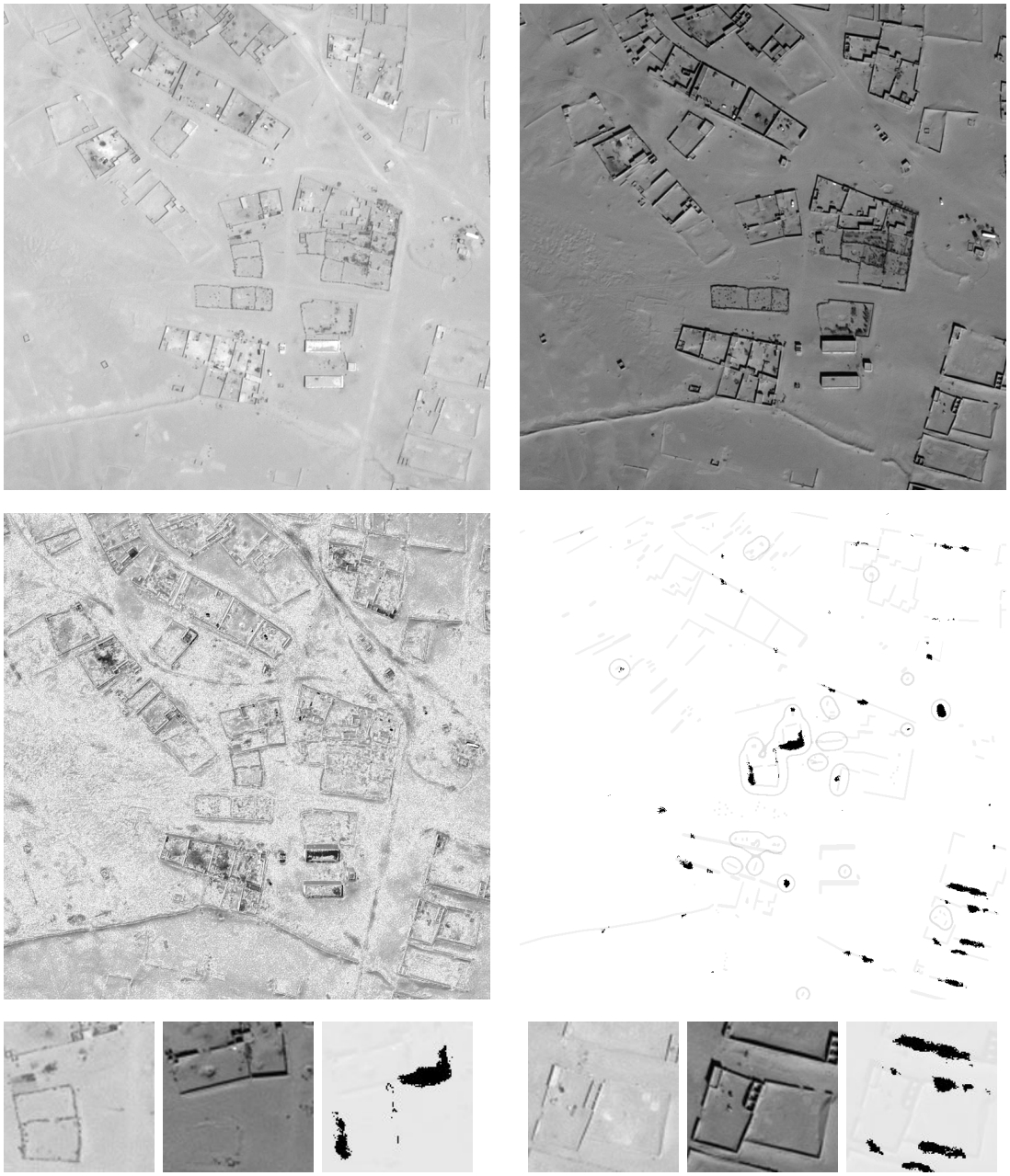


Figure 5.14 – Résultats obtenus sur une paire d'image réelle en utilisant le modèle agrégé de changements entraînée sur les données explicitées dans la section 5.4.1. Nous utilisons les conventions d'affichage données page 198. De part l'existence d'un fort changement d'intensité global entre les deux images, la différence est calculée sur une version recalée en intensité des deux images. Il s'agit d'un exemple particulièrement difficile, à cause de la grande variation de contraste et de l'apparition des ombres. Dans la première image, l'angle solaire est tel qu'une grande partie des murs est à peine visible, comme on peut le voir dans les agrandissements donnés. L'apparition d'une ombre dans la seconde provoque alors un grand nombre de fausses détection. Ici, le modèle agrégé parvient à détecter en partie la disparition et l'apparition de bâtiments au centre de l'image. Toutefois, il ne parvient à détecter que 5 des apparitions d'objets jugées pertinentes sur les 17 marquées et détecte une bonne partie des apparitions d'ombres.

Changement	Pertinence	Résultat du détecteur
Apparition et disparition de murets et habitations	oui	Détections partiels
Apparition de véhicules et objets	oui	Quelques détections
Apparition des ombres	non	Plusieurs détections
Variation du contraste	non	Aucune détection

Il s'agit d'un cas particulièrement difficile à cause de la variation d'angle solaire. Celui-ci étant presque vertical dans la première image, les différents éléments (murets et habitations) peuvent facilement se confondre avec le sol, tandis que leurs ombres dans la seconde image les font ressortir très clairement.

Tout d'abord, notons que le détecteur parvient ici à ne pas détecter le changement de contraste globale existant entre les deux paires. Cela peut s'expliquer par la présence de changement de contraste similaire dans les données d'apprentissage.

En ce qui concerne les résultats, le modèle ne détecte qu'un faible nombre d'apparitions et de disparitions de objets, pour un total de 5 sur la vingtaine de changements présents. Ceci peut éventuellement être expliquée par la position de ces derniers, trop des habitations.

Notons cependant qu'il s'agit d'une paire d'images particulièrement difficile et il n'est donc pas étonnant que le modèle présente ici des résultats limités.

5.4.3 Étude quantitative

Détecteurs agrégés « partiels » Pour pouvoir évaluer l'intérêt de l'agrégation des cinq modèles primaires développés dans les chapitres 3 et 4, nous avons comparé les performances du modèle agrégé « complet » (basé sur les 50 détecteurs primaires précisés en début de chapitre) avec les modèles primaires pris séparément. Toutefois, pour ne pas biaiser les performances d'un modèle en choisissant arbitrairement un descripteur, nous avons considéré, pour chaque modèle primaire, le détecteur obtenu en agrégeant tous les détecteurs primaires appariant ce modèle avec un des neufs descripteurs considérés.

Chacun de ces détecteurs agrégés « partiels » a été appris sur les mêmes données, présentées dans la figure 5.10, avec les mêmes paramètres d'agrégation. Nous avons alors mesuré les performances de ces 5 modèles agrégés partiels et du modèle agrégé « complet ».

Mesure de performance utilisée Comme précisé précédemment, il est difficile de trouver une mesure de performance qui soit réellement satisfaisante et représente réellement les attentes qu'on pourrait avoir dans le cadre de l'aide à la photo-interprétation. Nous avons donc visualisé les courbes représentant la précision de la détection d'un point de vu pixellique (donc la proportion de pixels détectés correspondant à un changement voulu par rapport à la quantité totale de pixels détectés) en fonction du « rappel objet », défini comme la proportion de changements effectivement détectés par rapport au nombre total de changements recherchés. Nous avons représenté ces courbes dans la figure 5.16.

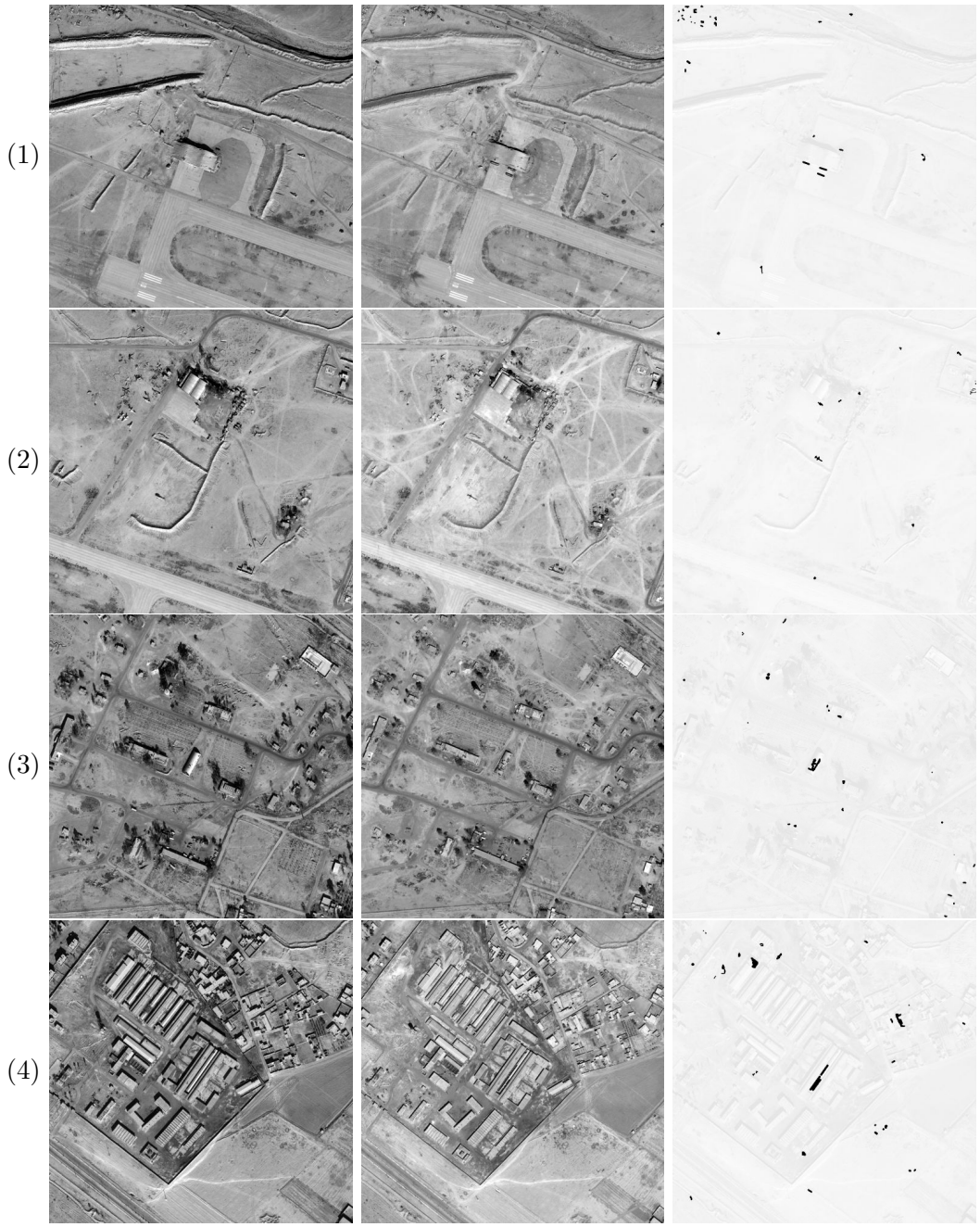


Figure 5.15 – Représentation des données utilisées pour tester les modèles de changements appris sur les données présentées dans la figure 5.10. Pour chacune des 4 paires d'images utilisées, nous avons représenté les deux images suivies du masque de changement construit pour l'occasion, les changements étant alors affichés en noir sur une version éclaircie de la première image de la paire.

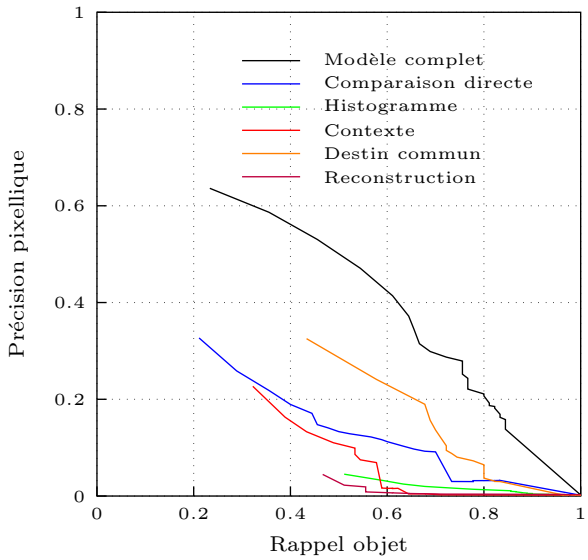


Figure 5.16 – Courbe de précision pixellique en fonction du rappel « objet », pour le modèle complet ainsi que pour les cinq modèles partiels, obtenus en agrégeant les détecteurs primaires correspondant à un unique modèle. Chaque modèle a été appris en utilisant les données présentées dans la figure 5.10, et les performances ont toutes étés estimées sur les données présentées dans la figure 5.15. Notons que chaque courbe possède ici un rappel minimal non nul, ce qui s'explique par une quantification des valeurs rendue par le modèle de classification utilisé (voir section 5.4.3). Comme attendu, les performances des modèles seuls apparaissent comme inférieures à celle du modèle complet.

Résultats obtenus Avant toute chose, remarquons que les différentes courbes de performance possèdent toutes une valeur de rappel minimale non nulle, ce qui peut sembler étrange. Cela est dû à la quantification des valeurs rendues par le modèle d'agrégation d'arbre de régression. À cause de celle-ci, de nombreux pixels partagent une même « valeur de détection », ainsi, le point de départ de la courbe correspond à la mesure de précision et rappel des pixels partageant la « valeur de détection » maximale.

Comme attendu, le modèle complet présente de meilleures performances que les différents modèles isolés.

Remarquons que, au moins pour ce cas d'usage, le modèle de changement pas destin commun s'avère particulièrement bon par rapport aux autres modèles Les modèles par comparaison d'histogrammes et par reconstruction indirecte présentent ici des résultats plutôt mauvais. En ce qui concerne le modèle par comparaison d'histogramme, cela peut être dû au caractère relativement grossier du masque de résultats obtenus, qui pourrait « déborder » et faire baisser la précision. Les performances du modèle par reconstruction indirecte sont plus difficilement compréhensibles.

Limitation de l'étude quantitative Il peut être important de discuter des limitation d'une telle étude quantitative. Tout d'abord, comme nous l'avons précisé dans

la section 5.3.2.3, les mesures de performance utilisées ne conviennent pas réellement au problème de la détection de changements. Le meilleur résultat selon une mesure de performance peut ne pas correspondre au meilleur résultat que l'on pourrait choisir manuellement.

Par ailleurs, l'étude quantitative est aussi limitée par les imprécisions sur les masques de changements réalisés manuellement et utilisés pour le test. En effet, comme nous l'avons vu dans les différentes paires étudiés, il peut être difficile de comprendre certains éléments présents dans les images (comme les taches de forme étrange, visible dans la figure 5.11) et ainsi d'évaluer la pertinence d'un changement sans la connaissance métier que possède un photo-interprète. Il est donc pas rare de faire des erreurs lors de la réalisation d'un masque de changements. Par ailleurs, il est aussi très difficile de repérer, de manière exhaustive, tous les changements pertinents entre deux images. Les masques utilisés sont donc souvent incomplets ou erronés, ce qui induit des erreur dans la mesure de performance.

Ainsi si cette étude permet de montrer, au moins pour la tâche considérée, l'intérêt d'un modèle de détection de changement agrégé, celui-ci permettant d'augmenter significativement les performances des différentes modèles. Il ne semble pas nécessairement pertinent de s'arrêter trop longtemps sur les valeurs précises des performances obtenues par les différents algorithmes de détection de changements.

Conclusion 6

Table des matières

6.1 Modèles primaires 211
6.2 Utilisation de l'apprentissage continue 212
6.3 Détection de changements dans des séries temporelles 212

6.1 Modèles primaires

Dans cette thèse, nous nous sommes intéressé au problème de la détection générique de changements pour l'aide à la photo-interprétation. Nous avons donc construit cinq modèles de détection de changements, chacun venant combler des limitations des modèles précédents et apportant une compréhension supplémentaire de la notion de changement.

Le premier modèle, par comparaison ponctuelle, s'inscrit dans les approches pixelliques, considérant les changements comme des variations significatives de l'intensité des pixels. Le second modèle, par comparaison d'histogramme, a trait à la comparaison des distributions locales des images, et permet ainsi une robustesse accrue aux décalages spatiaux entre les deux images, en perdant cependant une certaine résolution spatiale du résultat. Nous avons construit le troisième modèle, par modélisation du contexte local, en s'appuyant sur l'approche de modélisation du fond de l'image, proposée par l'algorithme de Reed-Xiaoli. Nous avons ainsi obtenu un modèle qui permet de tenir compte des décalages spatiaux tout en conservant une bonne résolution spatiale. Le quatrième modèle, par destin commun, s'inscrit dans une approche reposant sur la théorie de l'information et permet de comparer des images radiométriquement différentes, tout en offrant une robustesse aux changements globaux. Enfin, nous avons construit le cinquième modèle, par reconstruction indirecte, dans une optique de recalage d'image, pour pouvoir exprimer une des deux images dans le référentiel radiométrique de la seconde, permettant ainsi de les comparer simplement.

Ce faisant, nous avons effectué un petit tour d'horizon des méthodes de détection de changements. Nous n'avons évidemment pas considéré toutes les approches existant dans la littérature, de même que les cinq modèles que nous avons construit ne suffisent évidemment pas à capturer toute la variabilité des changements. Il existe plusieurs type de changements qui ne sont pas pris en compte et qui mettent en défaut ces cinq détecteurs. Comme nous l'avons vu dans les résultats de la figure 5.12, le cas d'une apparition simultanée d'un même objet (de véhicule ici) dans l'ensemble du domaine reste encore peu compris par les différents modèles construits ici.

Plus simplement, comme nous l'avions mentionné dans le chapitre 2, certains effets comme la saturation des pixels, les variations d'ombres ou encore la forte parallaxe et les effets d'occlusion ne sont pas réellement pris en compte. Si certains modèles parviennent à afficher une certaine robustesse, par exemple à une parallaxe légère, aucun des cinq modèles primaires ne permet réellement de comprendre un effet d'occlusion. On peut noter, par ailleurs, que ce problème est beaucoup étudié et plusieurs approchent présentent déjà de bon résultats, comme celle de Liu et coauteurs [LGT19], qui utilisent des descripteurs d'images similaires au descripteurs SIFT et détectent les changements au travers des défauts de correspondance. Des méthodes utilisant une modélisation 3D de la scène existent aussi, comme le montre Karantzalos [Kar15].

6.2 Utilisation de l'apprentissage continue

Pour agréger les différents modèles primaires et construire un détecteur de changements, spécifique à un cas d'usage particulier, nous avons utilisé un algorithme de *gradient boosting*, qui présente une grande capacité d'adaptation sans nécessiter un trop grand ensemble d'apprentissage. Les résultats prometteur obtenus sur le cas étudié dans le chapite 5 ont permis de montrer l'intérêt d'une telle agrégation, combinant les différentes approches de la notion de changements pour augmenter l'expressivité du modèle de détection de changement.

Toutefois, il serait plus pertinent d'inscrire cette fusion dans le cadre de d'un problème d'apprentissage continue, où l'on donne la possibilité au classifier de poursuivre continuellement son apprentissage. Dans ce cadre, on pourrait permettre à l'utilisateur de corriger le détecteur, ce dernier continuant de s'améliorer au fur et à mesure de son utilisation. Cette démarche peut sembler particulièrement intéressant puisqu'elle permet une interaction beaucoup plus naturelle avec un éventuel utilisateur. On éviterait par ailleurs l'étape de labellisation d'un ensemble d'images d'entraînement, étape qui peut être particulièrement pénible à effectuer.

6.3 Détection de changements dans des séries temporelles

Dans cette thèse, nous nous sommes restreint au problème de la détection de changements dans des paires d'images satellites. Toutefois, dans la pratique, il est rare de trouver un cas d'usage où l'on cherche réellement à comparer deux images satellites. Il s'agit le plus souvent d'étudier une nouvelle image au regard d'une situation antérieure, cette situation étant envisagée ici en considérant une seconde image capturée antérieurement. Dans ce cadre, rien n'empêche en réalité de considérer plus d'une image en tant que référence pour envisager les changements intervenant dans la nouvelle image, en envisageant ainsi le problème comme un problème de détection de changements dans des séries temporelles d'images. Notons que l'on s'intéresse ici à un cas particulier de séries temporelles d'images, puisqu'on envisage a priori les changements exclusivement dans la dernière image.

L'étude des séries temporelles d'images est un problème beaucoup étudié en télédétection, mais elle reste souvent cantonnée à des images de résolution moyenne, pour

lesquelles les effets des décalages spatiaux entre les images sont limités. Il est intéressant de remarquer que certains modèles primaires que nous avons construit peuvent s'adapter au cadre des séries temporelles. Le modèle de détection de changements par modélisation locale du contexte peut se généraliser en modélisant la distribution de référence non plus sur un contexte spatial (dans une seule image) mais sur un contexte spatio-temporel évaluée sur l'ensemble des images de référence. Les deux derniers modèles présentés, par destin commun et reconstruction indirecte, étant capable de comparer des images possédant un nombre différent de canaux, il est possible de fusionner les images de référence au sein d'un « images multi-temporelle » pour pouvoir appliquer ces modèles. Notons toutefois que la forte complexité temporelle du modèle par destin commun limiterait grandement son utilisation dans ce cadre. Toutefois, cela suppose que les images de référence soient parfaitement recalées.

Le fait de considérer plus d'une image de référence complexifie en réalité le problème de détection de changements, en démultipliant les types de changements à considérer. En effet, dans le cas d'une paire d'images, nous avons vu qu'il peut être nécessaire de considérer le contexte spatial (local ou global) d'un changement pour pouvoir évaluer sa pertinence. Dans le cas d'une série temporelle, il faudra aussi prendre en compte un contexte temporel, un objet pouvant par exemple présenter une intensité alternant entre plusieurs valeurs au cours du temps, voir un contexte spatio-temporel, cet objet pouvant en plus être translaté entre les images. Ces considérations peuvent être nécessaire lorsque l'on s'intéresse à des longues séries temporelles, par exemple pour prendre en compte d'éventuels effets saisonniers. Ce problème, s'insérant entre le traitement d'images et l'étude des séries temporelles, s'avère particulièrement intéressant mais peu étudié dans la littérature.

Annexes A

Table des matières

A.1 Propriété des nombres de fausses alarmes 215
A.2 Recalage d'intensités affine 218
A.3 Conditionnement gaussien 218

A.1 Propriété des nombres de fausses alarmes

Theorem 3.1.2 (Construction d'un NFA)**.** *Soit $X \in \mathcal{X}$ une variable aléatoire suivant une distribution H_0 et soit un ensemble de structures $\mathcal{I}$. Alors, pour toute fonction $\mu : \mathcal{I} \times \mathcal{X} \to \mathbb{R}$ et tout ensemble de réels positifs $(\eta_i)_{i\in\mathcal{I}}$, la fonction définie pour tout $x \in \mathcal{X}$ et $i \in \mathcal{I}$ par*

$$\mathrm{NFA}_{\mu,\eta}(i,x) \;=\; \eta_i \,.\, \mathbb{P}_{X\sim H_0}\big(\,\mu(i,X) \geq \mu(i,x)\,\big)\,, \tag{A.1}$$

est un nombre de fausses alarmes dès lors que

$$\sum_{i\in\mathcal{I}} \eta_i^{-1} \;\leq\; 1. \tag{A.2}$$

Par la suite, nous qualifierons la fonction μ de « mesure de structure » et les η_i, pour $i \in \mathcal{I}$, de coefficients de pondération.

Pour démontrer cela, considérons le lemme suivant.

Lemme A.1.1 (Sous uniformité de la p-valeur)**.** *Soit $X \in \mathbb{R}$ une variable aléatoire à valeur réelle, la fonction $F : x \in \mathbb{R} \mapsto \mathbb{P}(\,X \geq x\,)$ vérifie la propriété de sous-uniformité*

$$\forall \varepsilon > 0, \qquad \mathbb{P}(\,F(X) \leq \varepsilon\,) \leq \varepsilon. \tag{A.3}$$

Par ailleurs, si F est continue, il y a alors égalité dans l'équation précédente pour tout $\varepsilon \in [0,1]$.

Démonstration. Par définition, F est décroissante et continue à gauche. Considérons $\varepsilon \in [0,1[$ et posons

$$A_\varepsilon \;=\; \Big\{x \in \mathbb{R},\; F(x) \leq \varepsilon\Big\} \;=\; F^{-1}([0,\varepsilon]) \qquad \text{et} \qquad \alpha \;=\; \inf(A_\varepsilon).$$

On a, par définition, $\mathbb{P}(\,F(X) \leq \varepsilon\,) = \mathbb{P}(\,X \in A_\varepsilon\,)$. Puisque F est décroissante, tout y inférieur à α vérifie $F(y) > \varepsilon$, et donc nécessairement, $F(\alpha) = \varepsilon$ ou $F(\alpha) > \varepsilon$. Distinguons alors ces deux cas.

— Si $F(\alpha) = \varepsilon$, alors $A_\varepsilon = [\alpha, \infty[$ et on trouve directement

$$\mathbb{P}(\, X \in A_\varepsilon \,) \;=\; \mathbb{P}(\, X \geq \alpha \,) \;=\; F(\alpha) \;=\; \varepsilon.$$

— Si $F(\alpha) > \varepsilon$, alors $A_\varepsilon =]\alpha, \infty[$ et par décroissance et continuité à gauche de F,

$$\mathbb{P}(\, X \in A_\varepsilon \,) \;=\; \mathbb{P}(\, X > \alpha \,) \;=\; \lim_{\eta \to 0^+} \mathbb{P}(\, X \geq \alpha + \eta \,) \;=\; \lim_{\eta \to 0^+} F(\alpha + \eta).$$

Cette dernière limite ne peut pas être strictement supérieure à ε, il existerait sinon un $\eta_0 > 0$ tel que $F(\alpha + \eta_0) > \varepsilon$, ce qui contredit la définition de α. On a ainsi bien $\mathbb{P}(\, X \in A_\varepsilon \,) \leq \varepsilon$.

Pour prouver le dernier point, il suffit de remarqur que le second cas ne peut pas exister si F est continue. On obtient donc bien, dans ce cas, que pour tout $\varepsilon \in [0, 1[$, $\mathbb{P}(\, F(X) \leq \varepsilon \,) = \mathbb{P}(\, A_\varepsilon \,) = \varepsilon$. Par ailleurs, cela reste évidemment vrai pour $\varepsilon = 1$, puisque $\mathbb{P}(\, F(X) \leq 1 \,) = 1$. □

Lemme A.1.2. *Soit $X \in \mathcal{X}$ une variable aléatoire et $\mu : \mathcal{X} \to \mathbb{R}$, alors la fonction $F : x \in \mathcal{X} \mapsto \mathbb{P}(\, \mu(X) \geq \mu(x) \,)$ vérifie la propriété de sous-uniformité,*

$$\forall \varepsilon > 0, \qquad \mathbb{P}(\, F(X) \leq \varepsilon \,) \leq \varepsilon. \tag{A.4}$$

Par ailleurs, il suffit que $\mu(X)$ soit une variable à densité pour que l'on ait une égalité, pour tout $\varepsilon \in [0, 1]$.

Démonstration. Il suffit de poser $Y = \mu(X)$ et $G(y) = \mathbb{P}(\, Y \geq y \,)$, pour $y \in \mathbb{R}$. Pour $\varepsilon > 0$, le lemme précédent nous donne alors que $\mathbb{P}(\, G(Y) \leq \varepsilon \,) \leq \varepsilon$ et ainsi, on a bien

$$\mathbb{P}(\, F(X) \leq \varepsilon \,) \;=\; \mathbb{P}(\, G(\mu(X)) \leq \varepsilon \,) \;=\; \mathbb{P}(\, G(Y) \leq \varepsilon \,) \;\leq\; \varepsilon.$$

Par ailleurs, si $\mu(X)$ est une variable aléatoire à densité, nécessairement F est continue et le lemme précédent nous assure alors l'égalité ici. □

Démonstration du théorème 3.1.2. Soit $\varepsilon > 0$, on souhaite montrer ici que la fonction $\mathrm{NFA}_{\mu,\eta}$ est un nombre de fausses alarmes, c'est-à-dire que l'on peut majorer par ε l'expression

$$\begin{aligned}
\mathbb{E}\Big[\, \mathrm{Card}\, \{i, \mathrm{NFA}_{\mu,\eta}(i, X) \leq \varepsilon\} \,\Big] \;&=\; \mathbb{E}\Big[\sum_{i \in \mathcal{I}} \mathbb{1}_{\{\mathrm{NFA}_{\mu,\eta}(i,X) \leq \varepsilon\}} \Big] \\
&=\; \sum_{i \in \mathcal{I}} \mathbb{E}\Big[\, \mathbb{1}_{\{\mathrm{NFA}_{\mu,\eta}(i,X) \leq \varepsilon\}} \,\Big] \\
&=\; \sum_{i \in \mathcal{I}} \mathbb{P}(\, \mathrm{NFA}_{\mu,\eta}(i, X) \leq \varepsilon \,)\,.
\end{aligned}$$

Pour $i \in \mathcal{I}$ et en posant $F_i(X) = \mathbb{P}(\, \mu(i, X) \geq \mu(i, x) \,)$, le lemme A.1.2 nous donne alors la majoration

$$\mathbb{P}(\, \mathrm{NFA}_{\mu,\eta}(i, X) \leq \varepsilon \,) \;=\; \mathbb{P}\Big(\, F_i(X) \leq \varepsilon \eta_i^{-1} \,\Big) \;\leq\; \varepsilon\, \eta_i^{-1}$$

et ainsi

$$\begin{aligned}\mathbb{E}\Big[\operatorname{Card}\{i, \mathrm{NFA}_{\mu,\eta}(i,X) \leq \varepsilon\}\Big] &= \sum_{i\in\mathcal{I}} \mathbb{P}(\,\mathrm{NFA}_{\mu,\eta}(i,X) \leq \varepsilon\,) \\ &\leq \sum_{i\in\mathcal{I}} \varepsilon.\eta_i^{-1} \;\leq\; \varepsilon,\end{aligned}$$

et on retrouve bien la propriété annoncée. □

Propriété A.1.3 (Critère d'optimalité d'un NFA)**.** *En gardant les notations du théorème 3.1.2, si chaque coefficients de pondération η_i est égal à* $\operatorname{Card}(\mathcal{I})$ *et pour tout $i \in \mathcal{I}$, la variable aléatoire $\mu(i,X)$ possède une fonction de densité, alors la fonction* $\mathrm{NFA}_{\mu,\eta}$ *est un NFA optimal.*

De plus, sous ces conditions, $\mathrm{NFA}_{\mu,\eta}$ *correspond bien au nombre moyen de fausses alarmes, dans le sens où*

$$\forall \varepsilon \in [0, \operatorname{Card}(\mathcal{I})]\,, \qquad \mathbb{E}_{X\sim H_0}\Big[\operatorname{Card}\{i \in \mathcal{I}, \mathrm{NFA}_{\mu,\eta}(i,X) \leq \varepsilon\}\Big] \;=\; \varepsilon.$$

Démonstration. Soit $\varepsilon \in [0,N]$ avec $N = \operatorname{Card}(\mathcal{I})$. D'après le lemme A.1.2, comme, pour tout $i \in \mathcal{I}$, $\mu(i,X)$ est une variable à densité, et $\varepsilon\,\eta_i^{-1} = \varepsilon/N \leq 1$,

$$\mathbb{P}(\,\mathrm{NFA}_{\mu,\eta}(i,X) \leq \varepsilon\,) \;=\; \varepsilon\,\eta_i^{-1}.$$

On trouve ainsi

$$\begin{aligned}\mathbb{E}\Big[\operatorname{Card}\{i, \mathrm{NFA}_{\mu,\eta}(i,X) \leq \varepsilon\}\Big] &= \sum_{i\in\mathcal{I}} \mathbb{P}(\,\mathrm{NFA}_{\mu,\eta}(i,X) \leq \varepsilon\,) \\ &= \sum_{i\in\mathcal{I}} \varepsilon.\eta_i^{-1} \;=\; \varepsilon,\end{aligned}$$

et ce pour tout $\varepsilon \in [0,N]$. Considérons alors F, un nombre de fausses alarmes inférieur à $\mathrm{NFA}_{\mu,\eta}$. Puisque $F \leq \mathrm{NFA}_{\mu,\eta}$, nécessairement, pour tout $i \in \mathcal{I}$ et $\varepsilon \in [0,N]$,

$$\mathbb{P}(\,\mathrm{NFA}_{\mu,\eta}(i,X) \leq \varepsilon\,) \;\leq\; \mathbb{P}(\,F(i,X) \leq \varepsilon\,)\,, \tag{A.5}$$

ainsi,

$$\varepsilon \;=\; \sum_{i\in\mathcal{I}} \mathbb{P}(\,\mathrm{NFA}_{\mu,\eta}(i,X) \leq \varepsilon\,) \;\leq\; \sum_{i\in\mathcal{I}} \mathbb{P}(\,F(i,X) \leq \varepsilon\,) \;\leq\; \varepsilon, \tag{A.6}$$

la dernière inégalité découlant de la définition d'un nombre de fausses alarmes. Ainsi, il y a nécessairement égalité des sommes et, chaque terme étant positif et vérifiant l'inégalité de l'équation A.5, on en déduit que, pour tout $\varepsilon \in [0,N]$ et $i \in \mathcal{I}$,

$$\mathbb{P}(\,\mathrm{NFA}_{\mu,\eta}(i,X) \leq \varepsilon\,) \;=\; \mathbb{P}(\,F(i,X) \leq \varepsilon\,)\,. \tag{A.7}$$

Il suffit alors de remarquer que

$$F(i,x) \leq \mathrm{NFA}_{\mu,\eta}(i,x) \leq N$$

pour tout $i \in \mathcal{I}$ et $x \in \mathcal{X}$ pour en conclure que pour tout $i \in \mathcal{I}$, $F(i,X)$ et $\mathrm{NFA}_{\mu,\eta}(i,X)$ ont la même fonction de répartition et de ce fait, les variables aléatoires $\mathrm{NFA}_{\mu,\eta}(i,X)$ et $F(i,X)$ sont bien égales, sauf sur un ensemble de mesure nulle.

□

A.2 Recalage d'intensités affine

Soit u et v, deux images définies sur $\mathcal{D} \subset \mathbb{Z}^2$, à valeur dans $\mathbb{R}^n$, pour $n > 0$. Spécifiquement ici, nous verrons u et v comme des matrices de $\mathbb{R}^{n\times N}$, en notant $N = |\mathcal{D}|$ le nombre de pixels de $\mathcal{D}$, en notant aussi u_p et v_p les intensités de ces images au pixel $p \in \mathcal{D}$, vus comme des vecteurs de $\mathbb{R}^n$.

On cherche ici à trouver le recalage en intensité minimisant la distance quadratique entre u et la version recalée de v, c'est-à-dire trouver $a^* \in \mathbb{R}$ et $b^* \in \mathbb{R}^n$ vérifiant

$$(a^*, b^*) \;=\; \underset{(a,b)\in\mathbb{R}^2}{\operatorname{argmin}} \sum_{p\in\mathcal{D}} \|u_p - av_p - b\|_2^2 \;=\; \underset{(a,b)\in\mathbb{R}^2}{\operatorname{argmin}} \left\|u - av - b\mathbf{1}^\mathsf{T}\right\|_F^2, \tag{A.8}$$

où $\mathbf{1}$ est le vecteur colonne de $\mathbb{R}^{N\times 1}$ dont toutes les composantes sont égales à $\mathbf{1}$ et où $\|\cdot\|_F$ est la norme de Frobenius. En notant $f(a,b) = \|u - av - b\mathbf{1}^\mathsf{T}\|_F^2$, puisque

$$\begin{aligned} f(a+\alpha, b+\beta) &= \left\|u - av - b\mathbf{1}^\mathsf{T} - \alpha v - \beta\mathbf{1}^\mathsf{T}\right\|_F^2 \\ &= \mathrm{Tr}\Big(\left(u - av - b\mathbf{1}^\mathsf{T} - \alpha v - \beta\mathbf{1}^\mathsf{T}\right)^\mathsf{T}\left(u - av - b\mathbf{1}^\mathsf{T} - \alpha v - \beta\mathbf{1}^\mathsf{T}\right)\Big) \\ &= f(a,b) \;-\; 2\alpha\mathrm{Tr}\Big(v^\mathsf{T}u\Big) \;+\; 2\alpha a\,\|v\|_F^2 \;+\; 2\alpha\mathrm{Tr}\Big(v^\mathsf{T}b\mathbf{1}^\mathsf{T}\Big) \\ &\quad - 2\mathrm{Tr}\Big(\mathbf{1}\beta^\mathsf{T}u\Big) \;+\; 2a\mathrm{Tr}\Big(\mathbf{1}\beta^\mathsf{T}v\Big) \;+\; 2\mathrm{Tr}\Big(\mathbf{1}\beta^\mathsf{T}b\mathbf{1}^\mathsf{T}\Big) \;+\; o(\alpha, \|\beta\|), \end{aligned} \tag{A.9}$$

on trouve alors directement une expression du gradient de f, noté $\nabla f(a,b)$, à savoir

$$\nabla f(a,b) \;=\; 2\begin{pmatrix} a\,\|v\|_F^2 - \mathrm{Tr}\Big(v^\mathsf{T}u\Big) + (v\mathbf{1})^\mathsf{T}b \\ -u\mathbf{1} + a\,v\mathbf{1} + Nb \end{pmatrix}, \tag{A.10}$$

en notant que $\mathbf{1}^\mathsf{T}\mathbf{1} = N$. En notant alors $\overline{u} = \frac{1}{N}u\mathbf{1}$ et $\overline{v} = \frac{1}{N}v\mathbf{1}$, les vecteurs moyens associés à u et v, la condition d'annulation du gradient, $\nabla f(a^*, b^*)$ nous donne alors directement le système

$$\begin{cases} a^*\,\|v\|_F^2 \;+\; N\overline{v}^\mathsf{T}b^* \;=\; \mathrm{Tr}\Big(v^\mathsf{T}u\Big) \\ a^*\overline{v} \;+\; b^* \;=\; \overline{u}, \end{cases} \tag{A.11}$$

que l'on résout facilement par substitution, pour trouver

$$a^* \;=\; \frac{\mathrm{Tr}\Big(v^\mathsf{T}u\Big) - N\overline{v}^\mathsf{T}\overline{u}}{\|v\|_F^2 - N\,\|\overline{v}\|_2^2} \;=\; \frac{\displaystyle\sum_{p\in\mathcal{D}} v_p^\mathsf{T}u_p - |\mathcal{D}|\,\overline{v}^\mathsf{T}\overline{u}}{\|v\|_F^2 - |\mathcal{D}|\,\|\overline{v}\|_2^2}, \tag{A.12}$$

et donc

$$b^* \;=\; \overline{u} - a^*\overline{v}. \tag{A.13}$$

A.3 Conditionnement gaussien

Propriété A.3.1. *On considère deux variables aléatoires $X \in \mathbb{R}^n$ et $Y \in \mathbb{R}^m$ de sorte que la variable jointe $(X,Y) \in \mathbb{R}^{n+m}$ suive une loi gaussienne de moyenne μ et de matrice de covariance Σ, avec*

$$\mu = (\mu_x, \mu_y) \qquad \textit{et} \qquad \Sigma = \begin{pmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{pmatrix}. \tag{A.14}$$

Soit $y \in \mathbb{R}^n$, alors la variable X conditionnée par $Y = y$ suit alors une distribution gaussienne de paramètres

$$\mu_{|y} \;=\; \mu_x + \Sigma_{xy}\Sigma_y^{-1}(y-\mu_y) \qquad \textit{et} \qquad \Sigma_{|y} \;=\; \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{xy}{}^{\mathsf{T}}. \tag{A.15}$$

Démonstration. Pour tout $(x,y) \in \mathbb{R}^{n+m}$, notons $p(x,y)$, la fonction de densité de la variable jointe (X,Y),

$$p(x,y) \;=\; \frac{1}{\sqrt{(2\pi)^{n+m}\,|\Sigma|}} e^{-\frac{1}{2}\begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}^{\mathsf{T}} \Sigma^{-1} \begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}}, \tag{A.16}$$

ainsi que $p(y)$ la fonction de densité de Y,

$$p(y) \;=\; \frac{1}{\sqrt{(2\pi)^{m}\,|\Sigma_y|}} e^{-\frac{1}{2}(y-\mu_y)^{\mathsf{T}}\Sigma_y^{-1}(y-\mu_y)}. \tag{A.17}$$

Soit $y \in \mathbb{R}^m$, en notant $p_y(x)$, la densité de X conditionné par $Y = y$, pour $x \in \mathbb{R}^n$, on a

$$p_y(x) \;=\; \frac{p(x,y)}{p(y)} \;=\; \sqrt{\frac{|\Sigma_y|}{(2\pi)^n\,|\Sigma|}} e^{-\frac{1}{2}\begin{pmatrix} x' \\ y' \end{pmatrix}^{\mathsf{T}} \Sigma^{-1} \begin{pmatrix} x' \\ y' \end{pmatrix} + \frac{1}{2} y'^{\mathsf{T}}\Sigma_y^{-1}y'}, \tag{A.18}$$

avec $x' = x - \mu_x$ et $y' = y - \mu_y$. En inversant la matrice Σ en utilisant le complément de Schur, $S = \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}$,

$$\Sigma^{-1} \;=\; \begin{pmatrix} S^{-1} & -S^{-1}\Sigma_{xy}\Sigma_y^{-1} \\ \Sigma_y^{-1}\Sigma_{yx}S^{-1} & \Sigma_y^{-1} + \Sigma_y^{-1}\Sigma_{yx}S^{-1}\Sigma_{xy}\Sigma_y^{-1} \end{pmatrix}, \tag{A.19}$$

le terme dans l'exponentielle s'écrit (à un facteur $-\frac{1}{2}$ près)

$$\begin{aligned} \begin{pmatrix} x' \\ y' \end{pmatrix}^{\mathsf{T}} \Sigma^{-1} \begin{pmatrix} x' \\ y' \end{pmatrix} - y'^{\mathsf{T}}\Sigma_y^{-1}y' \;&=\; x'^{\mathsf{T}}S^{-1}x' \;-\; 2\,x'^{\mathsf{T}}S^{-1}\Sigma_{xy}\Sigma_y^{-1}y' \\ &\quad+\; y'^{\mathsf{T}}\Sigma_y^{-1}\Sigma_{yx}S^{-1}\Sigma_{xy}\Sigma_y^{-1}y' \\ &=\; \left(x' - \Sigma_{xy}\Sigma_y^{-1}y'\right)^{\mathsf{T}} S^{-1}(x' - \Sigma_{xy}\Sigma_y^{-1}y'). \end{aligned} \tag{A.20}$$

On retrouve ainsi bien l'expression d'une distribution gaussienne,

$$p_y(x) \;=\; \sqrt{\frac{|\Sigma_y|}{(2\pi)^n\,|\Sigma|}} e^{-\frac{1}{2}\left(x' - \Sigma_{xy}\Sigma_y^{-1}y'\right)^{\mathsf{T}} S^{-1}(x' - \Sigma_{xy}\Sigma_y^{-1}y')},$$

dont la moyenne et la covariance correspondent bien aux expressions annoncées. □

Corollaire A.3.2. *Soit deux variables aléatoires X et Y à valeur dans $\mathbb{R}$. Si la variable jointe (X,Y) suit une distribution gaussienne de moyenne et matrice de covariance*

$$\mu = (\mu_x, \mu_y) \qquad \textit{et} \qquad \Sigma = \begin{pmatrix} \sigma_x & \sigma_{xy} \\ \sigma_{xy} & \sigma_y \end{pmatrix}. \tag{A.21}$$

alors pour tout $x \in \mathbb{R}$, la variable Y conditionnée par $X = x$ suit une distribution gaussienne de paramètre $\mu_{|x}$ et $\sigma_{|x}$, avec

$$\mu_{|x} \;=\; \mu_y + \frac{\sigma_{xy}}{\sigma_x}(x - \mu_x) \qquad \textit{et} \qquad \sigma_{|x} \;=\; \sigma_y - \frac{\sigma_{xy}^2}{\sigma_x}. \tag{A.22}$$

Démonstration. Ce corollaire est un cas particulier de la proposition précédente. □

Bibliographie

[Agu14] Cecilia Aguerrebere. *On the generation of high dynamic range images : theory and practice from a statistical perspective.* PhD thesis, 2014.

[Alb09] Vito Alberga. Similarity measures of remotely sensed multi-sensor images for change detection applications. *Remote Sensing*, 1(3) :122–143, 2009.

[AM16] Rémy Abergel and Lionel Moisan. Fast and accurate evaluation of a generalized incomplete gamma function. 2016.

[AMHS16] M Aktar, MA Mamun, MA Hossain, and MSR Shuvo. Weighted normalized mutual information based change detection in remote sensing images. In *2016 19th International Conference on Computer and Information Technology (ICCIT)*, pages 257–260. IEEE, 2016.

[ASS+12] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11) :2274–2282, 2012.

[Att54] Fred Attneave. Some informational aspects of visual perception. *Psychological review*, 61(3) :183, 1954.

[B+98] Jeff A Bilmes et al. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute*, 4(510) :126, 1998.

[BB11] Francesca Bovolo and Lorenzo Bruzzone. An adaptive thresholding approach to multiple-change detection in multispectral images. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 233–236. IEEE, 2011.

[BBM08] Francesca Bovolo, Lorenzo Bruzzone, and Mattia Marconcini. A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7) :2070–2082, 2008.

[BCM05] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005.

[BDZ12] Csaba Benedek, Xavier Descombes, and Josiane Zerubia. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1) :33–50, 2012.

[BFSO84] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.

[BGV92] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152, 1992.

[BMB12] Francesca Bovolo, Silvia Marchesi, and Lorenzo Bruzzone. A framework for automatic and unsupervised detection of multiple changes in multitemporal images. *IEEE Transactions on Geoscience and Remote Sensing*, 50(6) :2196–2212, 2012.

[Boa04] Ronald J Boain. AB-Cs of sun-synchronous orbit mission design. 2004.

[BP00] Lorenzo Bruzzone and Diego F Prieto. Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote sensing*, 38(3) :1171–1182, 2000.

[Bre96] Leo Breiman. Bagging predictors. *Machine learning*, 24(2) :123–140, 1996.

[BS15] Alan S Belward and Jon O Skøien. Who launched what, when and why ; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103 :115–128, 2015.

[BSFG09] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. PatchMatch : A randomized correspondence algorithm for structural image editing. In *ACM Transactions on Graphics (ToG)*, volume 28, page 24. ACM, 2009.

[BSK+15] Csaba Benedek, Maha Shadaydeh, Zoltan Kato, Tamás Szirányi, and Josiane Zerubia. Multilayer Markov random field models for change detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 107 :22–37, 2015.

[BT15] RECOMMENDATION ITU-R BT. Parameter values for the HDTV standards for production and international programme exchange. 2015.

[BW05] Yanchen Bo and Jinfeng Wang. Combining multiple classifiers for thematic classification of remotely sensed data. *JOURNAL OF REMOTE SENSING-BEIJING-*, 9(5) :563, 2005.

[BY16] Yifang Ban and Osama Yousif. Change detection techniques : a review. In *Multitemporal Remote Sensing*, pages 19–43. Springer, 2016.

[Car00] Mark J Carlotto. Nonlinear background estimation and change detection for wide area search. *Optical Engineering*, 39(5) :1223–1229, 2000.

[CB94] Pol R Coppin and Marvin E Bauer. Processing of multitemporal Landsat TM imagery to optimize extraction of forest cover change

features. *IEEE Transactions on Geoscience and remote Sensing*, 32(4) :918–927, 1994.

[CCA11] Andrzej Cichocki, Sergio Cruces, and Shun-ichi Amari. Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization. *Entropy*, 13(1) :134–170, 2011.

[CDVLL98] Antonin Chambolle, Ronald A De Vore, Nam-Yong Lee, and Bradley J Lucier. Nonlinear wavelet image processing : variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing*, 7(3) :319–335, 1998.

[CG16] Tianqi Chen and Carlos Guestrin. Xgboost : A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[CHCW12] Gang Chen, Geoffrey J Hay, Luis MT Carvalho, and Michael A Wulder. Object-based change detection. *International Journal of Remote Sensing*, 33(14) :4434–4457, 2012.

[CJN+04] Pol Coppin, Inge Jonckheere, Kris Nackaerts, Bart Muys, and Eric Lambin. Digital Change Detection Methods in Ecosystem Monitoring : A Review. *International Journal of Remote Sensing - INT J REMOTE SENS*, 25 :1565–1596, 05 2004.

[CK93] Tianhorng Chang and C-CJ Kuo. Texture analysis and classification with tree-structured wavelet transform. *IEEE Transactions on image processing*, 2(4) :429–441, 1993.

[CLPT11] D Casciello, T Lacava, N Pergola, and V Tramutoli. Robust Satellite Techniques for oil spill detection and monitoring using AVHRR thermal infrared bands. *International journal of remote sensing*, 32(14) :4107–4129, 2011.

[CV95] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3) :273–297, 1995.

[CVGCMM+08] Gustavo Camps-Valls, Luis Gómez-Chova, Jordi Muñoz-Marí, José Luis Rojo-Álvarez, and Manel Martínez-Ramón. Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 46(6) :1822–1835, 2008.

[Dar05] Paul M Dare. Shadow analysis in high-resolution satellite imagery of urban areas. *Photogrammetric Engineering & Remote Sensing*, 71(2) :169–177, 2005.

[Dau92] Ingrid Daubechies. *Ten lectures on wavelets*. SIAM, 1992.

[DBD06] Baudouin Desclée, Patrick Bogaert, and Pierre Defourny. Forest change detection by statistical object-based method. *Remote Sensing of Environment*, 102(1-2) :1–11, 2006.

[DDG+14] Flora Dellinger, Julie Delon, Yann Gousseau, Julien Michel, and Florence Tupin. Change detection for high resolution satellite images, based on SIFT descriptors and an a contrario approach. In *2014 IEEE Geoscience and Remote Sensing Symposium*, pages 1281–1284. IEEE, 2014.

[DEMD18] Axel Davy, Thibaud Ehret, Jean-Michel Morel, and Mauricio Delbracio. Reducing anomaly detection in images to detection in noise. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 1058–1062. IEEE, 2018.

[DGP06] Johan Debayle, Yann Gavet, and Jean-Charles Pinoli. General adaptive neighborhood image restoration, enhancement and segmentation. In *International Conference Image Analysis and Recognition*, pages 29–40. Springer, 2006.

[DK99] X Long Dai and Slamak Khorram. Remotely sensed change detection based on artificial neural networks. *Photogrammetric engineering and remote sensing*, 65 :1187–1194, 1999.

[DLG+12] Peijun Du, Sicong Liu, Paolo Gamba, Kun Tan, and Junshi Xia. Fusion of difference images for change detection over urban areas. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(4) :1076–1086, 2012.

[DMBBB08] Mauro Dalla Mura, Jon Atli Benediktsson, Francesca Bovolo, and Lorenzo Bruzzone. An unsupervised technique based on morphological filters for change detection in very high resolution images. *IEEE Geoscience and Remote Sensing Letters*, 5(3) :433–437, 2008.

[DMM00] Agnès Desolneux, Lionel Moisan, and Jean-Michel Morel. Meaningful alignments. *International Journal of Computer Vision*, 40(1) :7–23, 2000.

[DMM07] Agnes Desolneux, Lionel Moisan, and Jean-Michel Morel. *From gestalt theory to image analysis : a probabilistic approach*, volume 34. Springer Science & Business Media, 2007.

[DWDQ08] JS Deng, K Wang, YH Deng, and GJ Qi. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. *International Journal of Remote Sensing*, 29(16) :4823–4838, 2008.

[EA96] H Erol and F Akdeniz. A multispectral classification algorithm for classifying parcels in an agricultural region. *International Journal of Remote Sensing*, 17(17) :3357–3371, 1996.

[EALW17] Arabi Mohammed El Amin, Qingjie Liu, and Yunhong Wang. Zoom out CNNs features for optical remote sensing change detection. In *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, pages 812–817. IEEE, 2017.

[EDMD18] Thibaud Ehret, Axel Davy, Jean-Michel Morel, and Mauricio Delbracio. Image Anomalies : a Review and Synthesis of Detection Methods. *arXiv preprint arXiv :1808.02564*, 2018.

[Egh79] Hassan J Eghbali. KS test for detecting changes from Landsat imagery data. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1) :17–23, 1979.

[FA91] William T. Freeman and Edward H Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (9) :891–906, 1991.

[FCL+10] Mariapia Faruolo, Irina Coviello, Teodosio Lacava, Nicola Pergola, and Valerio Tramutoli. On the potential of Robust Satellite Technique (RST) approach for flooded areas detection and monitoring using thermal infrared data. In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, pages 914–917. IEEE, 2010.

[Flo19] IEEE Standard for Floating-Point Arithmetic. *IEEE Std 754-2019 (Revision of IEEE 754-2008)*, pages 1–84, 2019.

[FM82] Kunihiko Fukushima and Sei Miyake. Neocognitron : A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets*, pages 267–285. Springer, 1982.

[Fri01] Jerome H Friedman. Greedy function approximation : a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[Gao96] Bo-Cai Gao. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, 58(3) :257–266, 1996.

[GCC+11] CSL Grimaldi, D Casciello, I Coviello, T Lacava, N Pergola, and V Tramutoli. An improved RST approach for timely alert and Near Real Time monitoring of oil spill disasters by using AVHRR data. *Natural Hazards and Earth System Sciences*, 11(5) :1281–1291, 2011.

[GDMDD09] Ruvimbo Gamanya, Philippe De Maeyer, and Morgan De Dapper. Object-oriented change detection for the city of Harare, Zimbabwe. *Expert Systems with Applications*, 36(1) :571–588, 2009.

[GJP+12] Nil Goyette, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, and Prakash Ishwar. Changedetection. net : A new change detection benchmark dataset. In *2012 IEEE computer society conference on computer vision and pattern recognition workshops*, pages 1–8. IEEE, 2012.

[GLR+78] D Galmier, R Lacot, R Richard, R Vyain, J-Y Scanvic, and G Weecksteen. Applications de la photo-interprétation et de la télédétection à la géologie de l'ingénieur. *Revue Française de Géotechnique*, (4) :73–84, 1978.

[GM84] Alexander Grossmann and Jean Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis*, 15(4) :723–736, 1984.

[GM08] Bénédicte Grosjean and Lionel Moisan. A-contrario Detectability of Spots in Textured Backgrounds. *Journal of Mathematical Imaging and Vision*, 33(3) :313–337, 2008.

[GSP11] Lionel Gueguen, Pierre Soille, and Martino Pesaresi. Change detection based on information measure. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11) :4503–4515, 2011.

[HCC+13] Masroor Hussain, Dongmei Chen, Angela Cheng, Hui Wei, and David Stanley. Change detection from remotely sensed images : From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80 :91–106, 2013.

[HD16] Ben Harwood and Tom Drummond. Fanng : Fast approximate nearest neighbour graphs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5713–5722, 2016.

[HN00] Stefan Huwer and Heinrich Niemann. Adaptive change detection for real-time surveillance applications. In *Proceedings Third IEEE International Workshop on Visual Surveillance*, pages 37–46. IEEE, 2000.

[Ho95] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.

[HPZ17] Tao Huang, Heng Peng, and Kun Zhang. Model selection for Gaussian mixture models. *Statistica Sinica*, pages 147–169, 2017.

[HS+88] Christopher G Harris, Mike Stephens, et al. A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Citeseer, 1988.

[HS18] Mahdi Hasanlou and Seyd Teymoor Seydi. Hyperspectral change detection : An experimental comparative study. *International journal of remote sensing*, 39(20) :7029–7083, 2018.

[HSD73] Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6) :610–621, 1973.

[HU75] Floyd M Henderson and Jack J Utano. Assessing general urban socioeconomic conditions with conventional air photography. *Photogrammetria*, 31(3) :81–89, 1975.

[HV16] Seokyong Hong and Ranga Raju Vatsavai. A Scalable Probabilistic Change Detection Algorithm for Very High Resolution (VHR) Satellite Imagery. In *Big Data (BigData Congress), 2016 IEEE International Congress on*, pages 275–282. IEEE, 2016.

[IJ05] Jungho Im and John R Jensen. A change detection model based on neighborhood correlation image analysis and decision tree classification. *Remote Sensing of Environment*, 99(3) :326–340, 2005.

[IM07] Jordi Inglada and Grégoire Mercier. A new statistical similarity measure for change detection in multitemporal SAR images and its exten-

sion to multiscale change analysis. *IEEE transactions on geoscience and remote sensing*, 45(5) :1432–1445, 2007.

[Ing07] Jordi Inglada. Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. *ISPRS journal of photogrammetry and remote sensing*, 62(3) :236–248, 2007.

[İÜ12] Murat İlsever and Cem Ünsalan. Pixel-based change detection methods. In *Two-Dimensional Change Detection Methods*, pages 7–21. Springer, 2012.

[Jah93] Bernd Jahne. *Spatio-temporal image processing : theory and scientific applications*. Springer-Verlag, 1993.

[JKC13] Xiuping Jia, Bor-Chen Kuo, and Melba M Crawford. Feature mining for hyperspectral image classification. *Proceedings of the IEEE*, 101(3) :676–697, 2013.

[Kan80] Gaetano Kanizsa. *Grammatica del vedere : saggi su percezione e gestalt*. Il mulino, 1980.

[Kar15] Konstantinos Karantzalos. Recent advances on 2D and 3D change detection in urban environments from remote sensing data. In *Computational Approaches for Urban Environments*, pages 237–272. Springer, 2015.

[KKH95] Stanley Q Kidder, RUSHWORTH M KIDDER, and Thomas H Vonder Haar. *Satellite meteorology : an introduction*. Gulf Professional Publishing, 1995.

[KL12] Joshua Kelcey and Arko Lucieer. Sensor correction of a 6-band multispectral imaging sensor for UAV remote sensing. *Remote Sensing*, 4(5) :1462–1493, 2012.

[KMB15] Nada Kadhim, Monjur Mourshed, and Michaela Bray. Shadow detection from very high resoluton satellite image using grabcut segmentation and ratio-band algorithms. International Society for Photogrammetry and Remote Sensing, 2015.

[KN05] Heesung Kwon and Nasser M Nasrabadi. Kernel RX-algorithm : A nonlinear anomaly detector for hyperspectral imagery. *IEEE transactions on Geoscience and Remote Sensing*, 43(2) :388–397, 2005.

[Kol33] Andrey Kolmogorov. Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.*, 4 :83–91, 1933.

[KWA11] Hans Knutsson, Carl-Fredrik Westin, and Mats Andersson. Representing local structure using tensors II. In *Scandinavian Conference on Image Analysis*, pages 545–556. Springer, 2011.

[LBD+89] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4) :541–551, 1989.

[Ler92] Brian G Leroux. Consistent estimation of a mixing distribution. *The Annals of Statistics*, pages 1350–1360, 1992.

[LGT19] Gang Liu, Yann Gousseau, and Florence Tupin. A contrario comparison of local descriptors for change detection in very high spatial resolution satellite images of urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6) :3904–3918, 2019.

[LHMS04] S Le Hégarat-Mascle and R Seltz. Automatic change detection by evidential fusion of change indices. *Remote Sensing of Environment*, 91(3) :390–404, 2004.

[LLZ+16] Liang Li, Xue Li, Yun Zhang, Lei Wang, and Guowei Ying. Change detection for high-resolution remote sensing imagery using object-oriented change vector analysis method. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 2873–2876. IEEE, 2016.

[LMS97] Donald T Lauer, Stanley A Morain, and Vincent V Salomonson. The Landsat program : Its origins, evolution, and impacts. *Photogrammetric Engineering and Remote Sensing*, 63(7) :831–838, 1997.

[LSD15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[Mal80] William A Malila. Change vector analysis : an approach for detecting forest changes with Landsat. In *LARS symposia*, page 385, 1980.

[Mal89] Stephane G Mallat. Multiresolution approximations and wavelet orthonormal bases of L2(R). *Transactions of the American mathematical society*, 315(1) :69–87, 1989.

[MBZ18] Lichao Mou, Lorenzo Bruzzone, and Xiao Xiang Zhu. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2) :924–935, 2018.

[MDP+06] Grégoire Mercier, Stéphane Derrode, Wojciech Pieczynski, J-M Nicolas, Annabele Joannic-Chardin, and Jordi Inglada. Copula-based stochastic kernels for abrupt change detection. In *2006 IEEE International Symposium on Geoscience and Remote Sensing*, pages 204–207. IEEE, 2006.

[Moo79] Gerald K Moore. What is a picture worth ? A history of remote sensing/Quelle est la valeur d'une image ? Un tour d'horizon de télédétection. *Hydrological Sciences Bulletin*, 24(4) :477–485, 1979.

[NCB04] Ramesh Neelamani, Hyeokho Choi, and Richard Baraniuk. ForWaRD : Fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Transactions on signal processing*, 52(2) :418–433, 2004.

[NCS98] Allan A Nielsen, Knut Conradsen, and James J Simpson. Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data : New approaches to change detection studies. *Remote Sensing of Environment*, 64(1) :1–19, 1998.

[PBC08] Gabriel Peyré, Sébastien Bougleux, and Laurent Cohen. Non-local regularization of inverse problems. In *European Conference on Computer Vision*, pages 57–68. Springer, 2008.

[PG81] John W Pratt and Jean D Gibbons. Kolmogorov-smirnov two-sample tests. In *Concepts of nonparametric theory*, pages 318–344. Springer, 1981.

[PMV00] Josien PW Pluim, JB Antoine Maintz, and Max A Viergever. Interpolation artefacts in mutual information-based image registration. *Computer vision and image understanding*, 77(2) :211–232, 2000.

[PNDS15] Otávio AB Penatti, Keiller Nogueira, and Jefersson A Dos Santos. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains ? In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 44–51, 2015.

[RAK+13] RA Alagu Raja, V Anand, A Senthil Kumar, Sandeep Maithani, and V Abhai Kumar. Wavelet based post classification change detection technique for urban growth monitoring. *Journal of the Indian Society of Remote Sensing*, 41(1) :35–43, 2013.

[RHSD74] JW Rouse, RH Haas, JA Schell, and DW Deering. Monitoring vegetation systems in the Great Plains with ERTS. *NASA special publication*, 351 :309, 1974.

[RKK09] Rudolf Richter, Tobias Kellenberger, and Hermann Kaufmann. Comparison of topographic correction methods. *Remote Sensing*, 1(3) :184–196, 2009.

[RMLHM10] Amandine Robin, Lionel Moisan, and Sylvie Le Hégarat-Mascle. An a-contrario approach for sub-pixel change detection in satellite imagery. *Transactions on Pattern Analysis and Machine Intelligence*, 32(11) :1977–1993, 2010.

[Ros58] Frank Rosenblatt. The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6) :386, 1958.

[Rub60] Harold Ruben. Probability content of regions under spherical normal distributions, I. *The Annals of Mathematical Statistics*, 31(3) :598–618, 1960.

[RVZ93] Eric JM Rignot and Jakob J Van Zyl. Change detection techniques for ERS-1 SAR data. *IEEE Transactions on Geoscience and Remote sensing*, 31(4) :896–906, 1993.

[RY90] Irving S Reed and Xiaoli Yu. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(10) :1760–1770, 1990.

[SB01] Sebastiano B Serpico and Lorenzo Bruzzone. A new search algorithm for feature selection in hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7) :1360–1367, 2001.

[SBB19] Sudipan Saha, Francesca Bovolo, and Lorenzo Bruzzone. Unsupervised deep change vector analysis for multiple-change detection in VHR images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6) :3677–3693, 2019.

[SBH+02] David WJ Stein, Scott G Beaven, Lawrence E Hoff, Edwin M Winter, Alan P Schaum, and Alan D Stocker. Anomaly detection from hyperspectral imagery. *IEEE signal processing magazine*, 19(1) :58–69, 2002.

[Ser83] Jean Serra. *Image analysis and mathematical morphology.* Academic Press, Inc., 1983.

[SF68] Irwin Sobel and Gary Feldman. A 3x3 isotropic gradient operator for image processing. *a talk at the Stanford Artificial Project in*, pages 271–272, 1968.

[SKZ14] Praveer Singh, Zoltan Kato, and Josiane Zerubia. A multilayer markovian model for change detection in aerial image pairs with large time differences. In *2014 22nd International Conference on Pattern Recognition*, pages 924–929. IEEE, 2014.

[SS04] Alan P Schaum and Alan Stocker. Hyperspectral change detection and supervised matched filtering based on covariance equalization. In *Algorithms and technologies for multispectral, hyperspectral, and ultraspectral imagery X*, volume 5425, pages 77–90. International Society for Optics and Photonics, 2004.

[SSV+13] Lucia Seebach, Peter Strobl, Peter Vogt, Wolfgang Mehl, and Jesús San-Miguel-Ayanz. Enhancing post-classification change detection through morphological post-processing–a sensitivity analysis. *International journal of remote sensing*, 34(20) :7145–7162, 2013.

[SUD10] Md Haidar Sharif, Sahin Uyaver, and Chabane Djeraba. Crowd behavior surveillance using bhattacharyya distance metric. In *International Symposium Computational Modeling of Objects Represented in Images*, pages 311–323. Springer, 2010.

[SYW+13] AmirReza Shahtahmassebi, Ning Yang, Ke Wang, Nathan Moore, and Zhangquan Shen. Review of shadow detection and de-shadowing methods in remote sensing. *Chinese Geographical Science*, 23(4) :403–420, 2013.

[TBU00] Philippe Thévenaz, Thierry Blu, and Michael Unser. Interpolation revisited [medical images application]. *IEEE Transactions on medical imaging*, 19(7) :739–758, 2000.

[TDBPP01] Valerio Tramutoli, Gerardo Di Bello, Nicola Pergola, and S Piscitelli. Robust satellite techniques for remote sensing of seismically active areas. *Annals Of Geophysics*, 2001.

[Tei86] PM Teillet. Image correction for radiometric effects in remote sensing. *International Journal of Remote Sensing*, 7(12) :1637–1651, 1986.

[TEK11] Daniel Tomowski, Manfred Ehlers, and Sascha Klonus. Colour and texture based change detection for urban disaster analysis. In *2011 Joint Urban Remote Sensing Event*, pages 329–332. IEEE, 2011.

[TFBM16] Frank Thonfeld, Hannes Feilhauer, Matthias Braun, and Gunter Menz. Robust Change Vector Analysis (RCVA) for multi-sensor very high resolution optical satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 50 :131–140, 2016.

[TP06] James Theiler and Simon Perkins. Proposed framework for anomalous change detection. In *ICML Workshop on Machine Learning Algorithms for Surveillance and Event Detection*, pages 7–14. Citeseer, 2006.

[Tsa03] Jeffrey Tsao. Interpolation artifacts in multimodality image registration based on maximization of mutual information. *IEEE transactions on medical imaging*, 22(7) :854–864, 2003.

[TZH11] Yuqi Tang, Liangpei Zhang, and Xin Huang. Object-oriented change detection based on the Kolmogorov–Smirnov test using high-resolution multispectral imagery. *International journal of remote sensing*, 32(20) :5719–5740, 2011.

[Uns07] Cemr Unsalan. Detecting changes in multispectral satellite images using time dependent angle vegetation indices. In *2007 3rd International Conference on Recent Advances in Space Technologies*, pages 345–348. IEEE, 2007.

[VG12] Ranga R Vatsavai and Jordan Graesser. Probabilistic change detection framework for analyzing settlement dynamics using very high-resolution satellite imagery. *Procedia Computer Science*, 9 :907–916, 2012.

[VT04] Harry L Van Trees. *Detection, estimation, and modulation theory, part I : detection, estimation, and linear modulation theory.* John Wiley & Sons, 2004.

[VTB+13] Michele Volpi, Devis Tuia, Francesca Bovolo, Mikhail Kanevski, and Lorenzo Bruzzone. Supervised change detection in VHR images using contextual information and support vector machines. *International Journal of Applied Earth Observation and Geoinformation*, 20 :77–85, 2013.

[VV99] EF Vermote and A Vermeulen. Atmospheric correction algorithm : spectral reflectances (MOD09). *ATBD version*, 4 :1–107, 1999.

[Wer94] Paul John Werbos. *The roots of backpropagation : from ordered derivatives to neural networks and political forecasting*, volume 1. John Wiley & Sons, 1994.

[WJS16] Melanie Weber, Jürgen Jost, and Emil Saucan. Forman-Ricci flow for change detection in large dynamic data sets. *Axioms*, 5(4) :26, 2016.

[Xu08] Hanqiu Xu. A new index for delineating built-up land features in satellite imagery. *International Journal of Remote Sensing*, 29(14) :4269–4276, 2008.

[YSLB05] Fei Yuan, Kali E Sawaya, Brian C Loeffelholz, and Marvin E Bauer. Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing. *Remote sensing of Environment*, 98(2-3) :317–328, 2005.

[ZC10] Maria Zontak and Israel Cohen. Defect detection in patterned wafers using anisotropic kernels. *Machine Vision and Applications*, 21(2) :129–141, 2010.

[ZFY+17] Yang Zhan, Kun Fu, Menglong Yan, Xian Sun, Hongqi Wang, and Xiaosong Qiu. Change detection based on deep siamese convolutional network for optical aerial images. *IEEE Geoscience and Remote Sensing Letters*, 14(10) :1845–1849, 2017.

[ZKAE16] Jin Zhou, Chiman Kwan, Bulent Ayhan, and Michael T Eismann. A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(11) :6497–6504, 2016.

[ZWGL17] Wei Zhao, Zhirui Wang, Maoguo Gong, and Jia Liu. Discriminative feature learning for unsupervised change detection in heterogeneous images based on a coupled neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12) :7066–7080, 2017.