

Détection indirecte de changements en imagerie satellitaire haute résolution

Vincent VIDAL^{1,2}, Lionel MOISAN¹

¹MAP5, Université Paris Descartes
45 rue des Saints-Pères, 75006 Paris, France

²Earthcube
13 rue de Calais, 75009 Paris, France

Vincent.Vidal@earthcube.eu, Lionel.Moisan@parisdescartes.fr

Résumé – Détecter des changements entre deux images satellite requiert en général d’être robuste à différents types de variations, par exemple de couleur, d’éclairement, de texture, voir même de résolution. Nous présentons ici une méthode de détection qui, en s’affranchissant des comparaisons directes entre les deux images, peut s’appliquer à des images radiométriquement différentes, ainsi qu’à des séries temporelles.

Abstract – Detecting changes between two satellite images generally requires some robustness to different types of variations, such as color, illumination, texture or even resolution. We here present a detection method that avoids direct comparisons between the two images, and can thus can be applied to radiometrically different images as well as to time series.

1 Introduction

Au cours des dernières décennies, le nombre de satellites et leur résolution spatiale ont connu une très forte augmentation, rendant l’utilisation des images satellite de plus en plus incontournable dans de nombreuses problématiques de surveillance et de détection. Le problème auquel on s’intéresse ici consiste à détecter divers changements entre au moins deux images satellite d’une même localisation, tout en étant robuste à d’autres différences jugées non pertinentes, à commencer par le bruit. De nombreuses approches ont été proposées pour la détection de changements, comme par exemple la « Change Vector Analysis », l’utilisation de champs de Markov ou l’analyse en composantes principales (voir [1] et ses références). Toutefois, la majorité d’entre elles ne prennent pas en compte deux difficultés récurrentes en imagerie satellitaire. La première difficulté réside dans les variations globales pouvant exister entre les deux images. Elles peuvent s’exprimer sous de multiples formes, comme des variations d’intensité ou de contraste (très présentes en imagerie thermique), de couleur, de texture (par exemple au niveau de la végétation), d’illumination et d’ombres (dues à angles solaires différents), ainsi que des variations de netteté, de résolution, de dynamique ou de longueur d’onde (cas typiquement rencontré lorsque les deux images sont prises par deux satellites différents). Pour comparer de telles paires d’images, que l’on qualifiera de « radiométriquement différentes », certaines méthodes proposent de remplacer la comparaison des intensités par la comparaison de quantités plus invariantes au contraste. C’est notamment le cas de la méthode de Gueguen et al. [2], qui repose sur l’information mutuelle.

Ces méthodes ne parviennent toutefois pas à garantir un haut niveau d’invariance (à la texture par exemple), et restent souvent très sensibles aux décalages spatiaux pouvant exister entre les deux images. Ceux-ci peuvent être dus au recalage imparfait des deux images, ou aux effets de parallaxe causés par des angles de vue différents. Cette incohérence spatiale étant d’autant plus présente que la résolution est forte, certaines méthodes se basent sur des descripteurs à plus grande échelle, comme les SIFT, pour y être robuste [3]. Notons également que certaines méthodes d’apprentissage automatique ont également été proposées pour la détection de changements, comme des arbres de décision ou des réseaux de neurones [4]. Elles nécessitent cependant de disposer d’un vaste ensemble de données d’apprentissage, ce qui constitue une limitation importante, notamment lorsqu’il s’agit de changements « rares », comme les catastrophes naturelles. Nous nous intéressons ici à la détection de changements dans des paires d’images radiométriquement différentes, pour lesquelles il n’est potentiellement pas envisageable de comparer directement (par exemple en mesurant des écarts d’intensités) les deux images. Ceci comprend notamment le cas où les deux images sont obtenues par des satellites qui ne couvrent pas les mêmes longueurs d’onde, et qui peuvent même avoir des nombres de canaux différents. Pour comparer de telles images, nous proposons, plutôt que de comparer directement les images entre elles, de comparer chaque image à une reconstruction de l’autre image obtenue en plaquant sa structure de similarité interne (obtenue en comparant les patches) sur la dynamique de la première. Une telle reconstruction, inspirée des travaux de Zontak et al. [5] et de ceux de Davy et al. [6], permet ainsi d’effectuer une comparaison croisée de deux ima-

ges potentiellement très différentes. Dans la Section 2, nous décrivons cette méthode de reconstruction ainsi que le modèle de détection de changements associé, formulé dans le cadre statistique de la détection « a contrario » [7]. Dans la Section 3, nous présentons des résultats obtenus sur une paire d'images, ainsi que sur une série temporelle où des changements très peu saillants (mais par ailleurs vérifiés à l'aide d'autres données) sont correctement détectés.

2 Modélisation

Dans toute la suite, nous considérons deux images observées, u et v , définies sur un domaine $\mathcal{D} \subset \mathbb{Z}^2$ et possédant respectivement c_u et c_v canaux. Nous noterons aussi V l'image aléatoire modélisant v .

2.1 Modèle de reconstruction croisée

Comme nous l'avons mentionné précédemment, nous nous plaçons a priori dans le cas où les deux images u et v présentent des différences radiométriques telles qu'il ne soit pas envisageable de les comparer directement. Pour contourner cette difficulté, nous proposons de comparer v à une image \tilde{v} obtenue en reconstruisant l'image u à partir des intensités de v et de la structure de similarité interne des patches de u . L'image \tilde{v} exprime ainsi la géométrie de u , mais dans l'espace radiométrique de v . La comparaison des images v et \tilde{v} peut alors être effectuée plus directement, soit pixel à pixel, soit de manière à incorporer une certaine robustesse aux légers décalages spatiaux (dus par exemple à des effets de parallaxe ou à de légers défauts de recalage).

Notations. Dans toute la suite, nous supposons fixée une forme de patch $P_0 \subset \mathbb{Z}^2$ (dans la pratique, un carré centré en 0), et nous noterons $[p]$ le domaine (patch) de forme P_0 centré au pixel $p \in \mathcal{D}$, c'est-à-dire $[p] = \{p+q, q \in P_0\}$. Nous noterons également $\mathcal{N}_p(u)$ l'ensemble des K plus proches voisins du patch $[p]$ dans l'image u , au sens de la norme L^2 et en excluant les patches trop proches de $[p]$ (pour les figures 1 et 2 nous avons utilisé une distance seuil de 6 pixels). De plus, pour éviter les problèmes de bords, nous nous restreindrons en pratique à $p \in \mathcal{D}'$, avec

$$\mathcal{D}' = \{p \in \mathcal{D}, [p] \subset \mathcal{D}\}. \quad (1)$$

Pour finir, nous conviendrons de noter $V_{[p]}$ le vecteur de $\mathbb{R}^{c_v|P_0|}$ correspondant à la restriction de V au patch $[p]$, et $(V_{[p]})_\tau$ la valeur au pixel $\tau \in P_0$ de cette restriction. Si les images u et v ont des résolutions différentes, on peut choisir des tailles de patch distinctes pour chaque image, puisque les comparaisons sont toujours faites au sein d'une même image.

Reconstruction d'image. La transformation de u en \tilde{v} est définie en reconstruisant tout d'abord chaque patch $[p]$ de v , pour

$p \in \mathcal{D}'$, à l'aide du nouveau patch

$$\overline{v_{[p]}} = \sum_{[q] \in \mathcal{N}_p(u)} \tilde{w}_{p,q}(u) v_{[q]}, \quad (2)$$

où $\tilde{w}_{p,q}(u)$ quantifie la similarité entre $u_{[p]}$ et $u_{[q]}$,

$$\tilde{w}_{p,q}(u) = \frac{w_{p,q}(u)}{\sum_{[q'] \in \mathcal{N}_p(u)} w_{p,q'}(u)} \text{ avec } w_{p,q}(u) = e^{-\frac{\|u_{[p]} - u_{[q]}\|_2^2}{h^2}}, \quad (3)$$

h étant une constante fixée à partir du niveau de bruit estimé dans u . On reconstruit alors l'image entière \tilde{v} par moyennage des patches reconstruits, soit

$$\forall p \in \mathcal{D}', \quad \tilde{v}_p = \frac{1}{|P_0|} \sum_{\tau \in P_0} (\overline{v_{[p-\tau]}})_\tau, \quad (4)$$

ce que l'on peut écrire directement

$$\forall p \in \mathcal{D}', \quad \tilde{v}_p = \frac{1}{|P_0|} \sum_{\tau \in P_0} \sum_{[q] \in \mathcal{N}_{p-\tau}(u)} \tilde{w}_{p-\tau,q}(u) v_{q+\tau}. \quad (5)$$

Erreur de reconstruction. Analysons, dans le cas d'une image aléatoire V , l'erreur de reconstruction $E = V - \tilde{V}$, c'est-à-dire, pour un pixel $p \in \mathcal{D}'$,

$$E_p = \frac{1}{|P_0|} \sum_{\tau \in P_0} \sum_{[q] \in \mathcal{N}_{p-\tau}(u)} \tilde{w}_{p-\tau,q}(u) (V_p - V_{q+\tau}). \quad (6)$$

Si l'on suppose que, dans le cas où aucun changement n'est présent, tout patch et ses voisins sont indépendants et suivent une même distribution gaussienne de variance $\sigma^2 I_{c_v|P_0|}$, on a alors $E_p \sim \mathcal{N}(0, \sigma^2 \sigma_p^2 I_{c_v})$ pour tout $p \in \mathcal{D}'$, avec

$$\sigma_p^2 = 2 \sum_{\tau \in P_0} \sum_{[q] \in \mathcal{N}_{p-\tau}(u)} \tilde{w}_{p-\tau,q}^2(u). \quad (7)$$

En notant $e_p = v - \tilde{v}$ l'erreur de reconstruction effectivement calculée, on se propose de détecter les changements comme les pixels $p \in \mathcal{D}'$ vérifiant $F(\|e_p\|_2) \leq \varepsilon$, pour une fonction $F : \mathbb{R} \rightarrow \mathbb{R}$ et un seuil ε convenablement choisis.

Nombre de fausses alarmes. Plaçons nous dans le cadre « a contrario » introduit dans [7]. Il s'agit de chercher une fonction F vérifiant la propriété suivante : en seuillant $F(\|e_p\|_2)$ au niveau ε , et en supposant qu'il n'existe pas de changement entre les deux images, le nombre de fausses détections devra être en moyenne au plus égal à ε , c'est-à-dire

$$\forall \varepsilon > 0, \quad \mathbb{E} \left[\text{Card} \{p \in \mathcal{D}', F(\|E_p\|_2) \leq \varepsilon\} \right] \leq \varepsilon. \quad (8)$$

Desolneux et al. [7] montrent qu'il suffit de considérer

$$F(t) = |\mathcal{D}'| \cdot \mathbb{P} \left(\|E_p\|_2 \geq t \right), \quad (9)$$

ce qui, après calcul, conduit à

$$F(t) = \frac{|\mathcal{D}'|}{\Gamma\left(\frac{c_v}{2}\right)} \Gamma\left(\frac{c_v}{2}, \frac{t^2}{2\sigma^2\sigma_p^2}\right), \quad (10)$$

où $\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt$ est la fonction gamma incomplète. En pratique, pour des raisons de représentabilité numérique, on calcule plutôt $\log F(t)$, ce qui peut se faire de manière robuste (sans *overflow*) à l'aide de l'algorithme proposé dans [8].

Pondération du patch central. Il est important de noter que pour fonctionner, ce modèle demande à chaque patch d’avoir au moins un voisin proche dans u , sous peine d’être automatiquement mal reconstruit. Pour éviter de détecter comme changement toutes les anomalies 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 sera alors reconstruit avec

$$\forall p \in \mathcal{D}, \quad \overline{v}_{[p]} = \frac{\lambda v_{[p]} + \sum_{[q] \in \mathcal{N}_p(u)} w_{p,q}(u) v_{[q]}}{\lambda + \sum_{[q] \in \mathcal{N}_p(u)} w_{p,q}(u)}, \quad (11)$$

le moyennage des patches reconstruits restant inchangé. Dans la pratique, nous avons fixé ce paramètre à 0,1, il serait toutefois intéressant d’étudier la pertinence d’adapter ce choix aux valeurs de K et h^2 .

3 Résultats et discussions

Pour détecter les changements entre deux images, nous avons appliqué l’algorithme précédent deux fois, en reconstruisant une image à partir des valeurs de similarité de l’autre puis en inversant les deux images. Nous avons alors pris le minimum des deux valeurs données par l’équation (10) et seuillé ces dernières avec un seuil $\varepsilon = 10^{-30}$.

Nous avons appliqué cet algorithme à une paire d’images radiométriquement différentes, présentées figure 1. Ici, les variations d’ombres provoquent la majorité des fausses détections. Notons que les ombres constituent un problème récurrent en imagerie satellitaire [9].

Applications aux séries temporelles Il est intéressant de remarquer que, dans le modèle proposé, les deux images u et v peuvent avoir des résolutions ainsi que des nombres de canaux différents. Par conséquent, en considérant des images successives d’une même scène comme différents canaux d’une même image, nous obtenons un algorithme permettant de détecter des changements au sein d’une série temporelle de plus de deux images. Nous montrons à la figure 2 des résultats obtenus pour une série de 20 images acquises dans l’infrarouge thermique. Il faut noter que dans le cas d’une série temporelle, il n’est plus réellement possible de déterminer manuellement l’ensemble des changements (à l’exception de quelques changements évidents, comme les nuages). De ce fait, pour évaluer la performance de l’algorithme, nous avons évalué les détections sans chercher les changements qui n’auraient pas été détectés. Notons que la non-détection de certains nuages s’explique par le manque de voisins pertinents des zones correspondantes. C’est par exemple le cas lorsqu’il existe un changement dans les images de référence, ce qui empêche alors toute similarité possible.

Remarquons par ailleurs qu’en concaténant l’information de plusieurs images, on estime la structure de similarité de la scène avec plus de précision. Il devient alors possible de diminuer

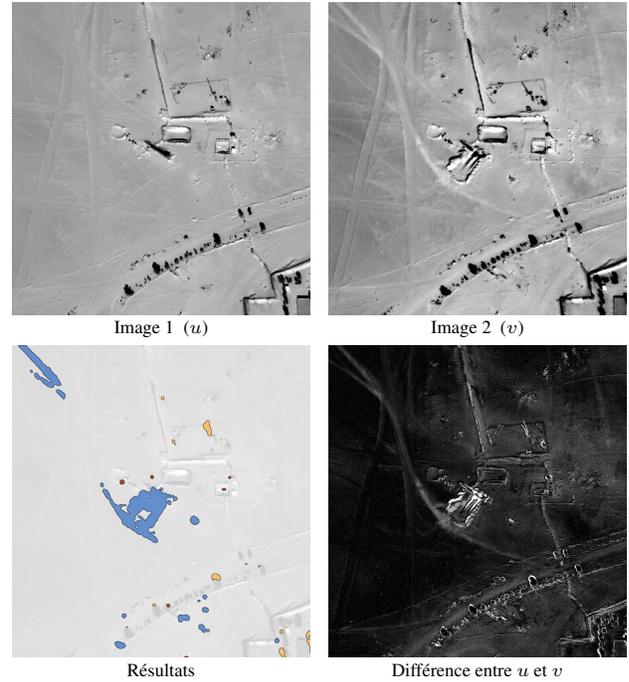


FIGURE 1 – Résultats obtenus sur une paire d’images (acquises par deux satellites différents, WorldView-2 et WorldView-3) de taille 1013×1062 et présentant des caractéristiques radiométriques différentes. L’algorithme a été utilisé avec 5 patches voisins de taille 9×9 , $\lambda = 0,1$ et $h = 100$. Les patches voisins sont ici calculés de manière exacte. **Haut** : Paire d’images utilisée. **Bas gauche** : Masque de détection obtenu. Les détections sont représentées par des taches sur une version éclaircie de la première image. Les taches bleues correspondent aux vrais positifs, les taches rouges à des fausses détections et les taches jaunes à des fausses détections imputables aux ombres. On détecte bien ici les apparitions de bâtiments, des décombres, ainsi que de la route. La majorité des fausses détections correspondent à des déplacements d’ombres. **Bas droite** : Valeur absolue de la différence des deux images, un pixel étant d’autant plus clair que la différence est élevée. Cette image montre bien la difficulté que présenterait la détection de changements par comparaison directe des deux images.

considérablement la taille des patches utilisés : dans l’expérience de la figure 2 par exemple, les résultats sont quasiment identiques lorsque l’on utilise des “patches” réduits à un seul pixel (contre 3×3 au départ).

Prise en compte des décalages locaux. Le modèle présenté jusqu’à présent est sensible aux décalages locaux, le processus de reconstruction supposant qu’un patch corresponde au même emplacement dans les deux images. Il est toutefois possible de modifier légèrement le modèle pour limiter les effets de ces décalages. Notons T la fonction de translation des pixels de u vers ceux de v . Pour reconstruire un patch $v_{[p]}$, il faudrait théoriquement utiliser les patches $[T(q)]$ pour $[q] \in \mathcal{N}_{T^{-1}(p)}(u)$.

En supposant alors que l’amplitude de la translation locale

des patches est bornée par une constante t_{\max} et évolue suffisamment lentement à l'échelle du patch pour éviter tout problème d'occlusion et de déformation, on peut considérer que chaque patch $[q] \in \mathcal{N}_{T-1(p)}(u)$ correspond à une version translattée d'au plus t_{\max} d'un patch de $\mathcal{N}_p(u)$. Il existe alors une incertitude d'au plus $2t_{\max}$ sur la position des patches de $\mathcal{N}_p(u)$, et l'on peut en tenir compte en définissant

$$\overline{v_{[p]}} = \sum_{[q] \in \mathcal{N}_p(u)} \tilde{w}_{p,q}(u) v_{[q+t_q^*]}, \quad (12)$$

où

$$t_q^* = \arg \min_{\|t\| \leq 2t_{\max}} \|v_{[p]} - v_{[q+t]}\|_2 \quad (13)$$

et les $\tilde{w}_{p,q}$ sont définis dans l'équation (3). Notons qu'avec cette modification, nous perdons la garantie théorique du nombre de fausses alarmes donnée par l'équation (10) (en pratique, cela n'est pas réellement gênant). Par ailleurs, cette méthode ne permet de limiter que les effets des décalages entre u et v . Dans le cas d'une série temporelle, il est nécessaire que toutes les images de référence, constituant les canaux de u , soient parfaitement recalées.

Références

- [1] 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*, vol. 80, pp. 91–106, 2013.
- [2] Lionel Gueguen, Pierre Soille, and Martino Pesaresi, "Change detection based on information measure," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4503–4515, 2011.
- [3] 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*. IEEE, 2014, pp. 1281–1284.
- [4] Rodrigo Caye Daudt, Bertr Le Saux, and Alexandre Boulch, "Fully convolutional siamese networks for change detection," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 4063–4067.
- [5] Maria Zontak and Israel Cohen, "Defect detection in patterned wafers using anisotropic kernels," *Machine Vision and Applications*, vol. 21, no. 2, pp. 129–141, 2010.
- [6] 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)*. IEEE, 2018, pp. 1058–1062.
- [7] Agnes Desolneux, Lionel Moisan, and Jean-Michel Morel, *From gestalt theory to image analysis : a probabilistic approach*, vol. 34, Springer Science & Business Media, 2007.

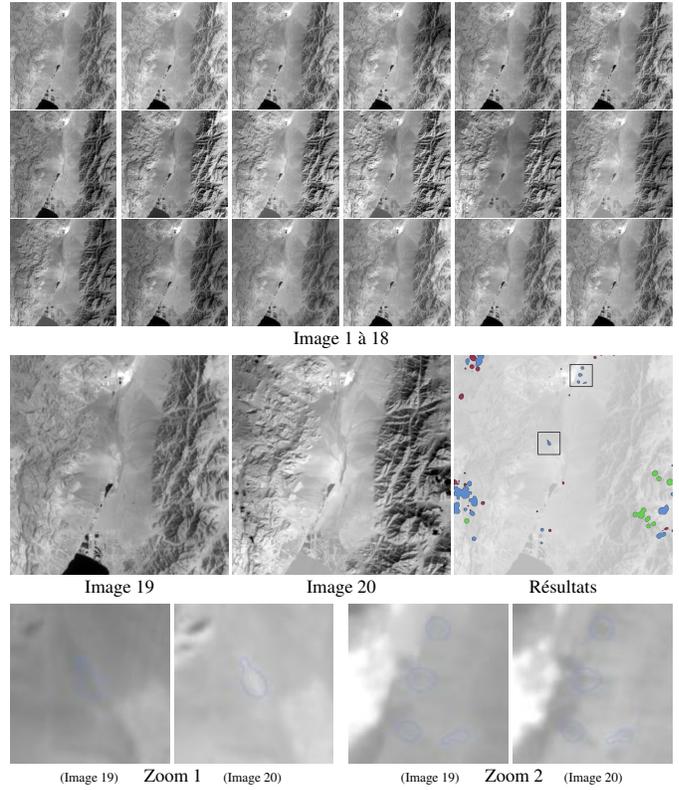


FIGURE 2 – Résultats obtenus sur une série de 20 images Landsat en infrarouge thermique de taille 1000×1000 . Les 19 premières images (formant u) servent à estimer la structure de voisinage, utilisée ensuite sur la vingtième image (v) pour détecter les changements sur cette dernière en utilisant des patches de taille 3×3 , les 5 patches plus proches voisins, $\lambda = 0,1$ et $h = 5000$. **Haut** : Les 18 premières images. **Milieu** : De gauche à droite, les images 19, 20 et le masque de détection. Les taches bleues correspondent aux vrais positifs, les taches rouges à des fausses détections et les taches vertes aux nuages n'ayant pas été détectés. Les cadres correspondent aux emplacements des zooms. On détecte bien ici la majorité des nuages ainsi qu'une tache suspecte (zoom 1) qui s'avère être une fuite de pétrole (documentée par d'autres sources) ainsi qu'un autre changement (zoom 2), identifié par ailleurs comme une zone de terrassement dans d'autres images (prises dans le domaine visible). Les nuages non détectés sont indiqués en rouge. **Bas** : Zooms de deux détections, prises dans les images 19 et 20. On a fait figurer en bleu transparent les silhouettes des détections.

- [8] Rémy Abergel and Lionel Moisan, "Fast and accurate evaluation of a generalized incomplete gamma function," preprint MAP5, 2016.
- [9] 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*, vol. 23, no. 4, pp. 403–420, 2013.