



HAL
open science

Détection de bâtiments à partir d'une image satellitaire par combinaison d'approches ascendante et descendante

Mohamed Mahmoud Sidi Youssef, Clément Mallet, Nesrine Chehata, Arnaud Le Bris, Adrien Gressin

► **To cite this version:**

Mohamed Mahmoud Sidi Youssef, Clément Mallet, Nesrine Chehata, Arnaud Le Bris, Adrien Gressin. Détection de bâtiments à partir d'une image satellitaire par combinaison d'approches ascendante et descendante. *Reconnaissance de Formes et Intelligence Artificielle (RFIA) 2014*, Jun 2014, France. hal-00989026

HAL Id: hal-00989026

<https://hal.science/hal-00989026>

Submitted on 9 May 2014

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

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

Détection de bâtiments à partir d'une image satellitaire par combinaison d'approches ascendante et descendante

Mohamed Mahmoud Sidi Youssef¹
Nesrine Chehata³

Clément Mallet²
Arnaud Le Bris² Adrien Gressin²

¹Sup'Com Ecole Supérieure des Communications de Tunis, Tunisie

²IGN/Laboratoire MATIS, Université Paris Est, France

³Laboratoire G&E (EA 4592), IPB / Université de Bordeaux, France
prenom.nom@ign.fr

Résumé

La détection de bâtiments à partir d'imagerie géospatiale a été un sujet de recherche très populaire ces vingt dernières années et en particulier avec l'émergence des satellites à très haute résolution. Les méthodes existantes présentent diverses limitations les empêchant d'être efficaces à grandes échelles spatiales et temporelles : elles dépendent du contexte, d'un paramétrage fastidieux ou nécessitent plusieurs sources de données. Dans ce travail, nous proposons une méthode automatique qui répond en partie à ces problèmes en combinant les avantages des méthodes descendantes et ascendantes d'extraction de bâtiments, i.e., à la fois des algorithmes de classification et de reconnaissance des formes. Cela permet de détecter correctement les objets grâce à des connaissances géométriques a priori tout en délimitant de façon précise leurs frontières et en préservant leur forme. La méthode est évaluée sur une zone semi-urbaine de plus de 230 bâtiments en utilisant une image fusionnée PAN+XS Pléiades de 0,5 m de résolution.

Mots Clef

Segmentation, classification, processus ponctuels marqués, bâtiments, imagerie satellitaire à très haute résolution.

Abstract

Building detection from geospatial optical images has been a popular topic of research for the last twenty years and in particular with the emergence of very high resolution satellites. Existing methods exhibit various flaws and prevent them from being efficient at large scales of space and time : they are context-dependent, require a tedious parameter tuning or several data sources. In this paper, we propose a fully automatic method that alleviates some of these issues by combining the strengths of bottom-up and top-down approaches, i.e., of both classification and pattern recognition algorithms. This allows to correctly detect the objects by geometric prior knowledge while finely delineating their borders and preserving their shapes. The method is evaluated over a complex area of more than 230 buildings using a 0.5 m multispectral pansharpened Pleiades image.

Keywords

Segmentation, classification, Marked Point Process, building, very high resolution imagery.

1 Introduction

La détection de bâtiments à partir d'images satellitaires à très haute résolution spatiale (THR) est un enjeu important pour de nombreuses applications principalement liées à la mise en place des politiques publiques. Durant les deux dernières décennies, elle a fait l'objet de nombreuses recherches utilisant différentes sources de données telles que les images optiques multispectrales ou hyperspectrales, les données SAR ou encore LIDAR, ainsi que des connaissances externes : bases de données topographiques et cadastre. La source de données la plus simple et la plus utilisée reste les images optiques monoculaires [1] même si de nombreux travaux introduisent l'information d'altitude [2, 3] ou SAR [4] en complément. Cependant, de telles données ne sont pas toujours disponibles et le fait d'utiliser différentes sources rend ces méthodes moins facilement applicables à grande échelle. Par conséquent, notre étude se focalise sur le traitement d'une image satellitaire optique THR monoculaire. Cela offre un bon compromis entre résolutions spatiale (forte), spectrale (faible) et temporelle (moyenne).

Les travaux antérieurs se regroupent en trois catégories. (1) Les méthodes ascendantes (*Bottom-Up*) se fondent sur l'extraction d'information bas-niveau. Les images optiques submétriques fournissent une description détaillée des scènes urbaines d'un point de vue spatial et spectral, permettant l'utilisation simple d'algorithmes de classification supervisée [5, 6]. Cependant, sans modélisation sous-jacente des bâtiments, de telles classifications peuvent être très bruitées en raison de fortes hétérogénéité et complexité des scènes urbaines. (2) Les méthodes descendantes (*Top-Down*) se focalisent directement sur la détection d'objets, par reconnaissance de formes et par utilisation de connaissances a priori sur les arrangements spatiaux et les interactions entre objets [7, 8]. Cependant, elles sont souvent

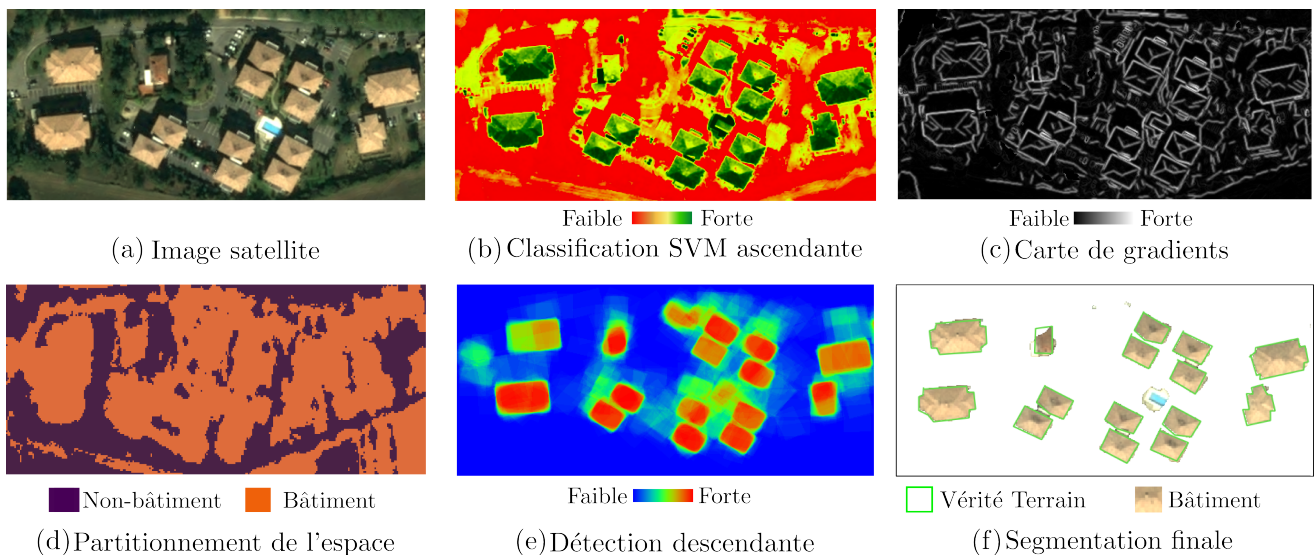


FIGURE 1 – Aperçu général de l’approche (*Zone 1*). Voir le texte pour plus de détails.

complexes, hautement paramétrées et avec des temps de calcul rédhibitoires pour des grandes zones. (3) Les méthodes mixtes visent à bénéficier des avantages des deux approches. En général, celles fondées sur l’optimisation par graphes sont choisies de manière à propager l’information locale obtenue par classification à de grandes zones par un raisonnement spatial structuré [1]. Elles sont efficaces, mais la recherche de la configuration optimale nécessite une optimisation globale, difficile à obtenir sur de grandes échelles. Par conséquent, une autre solution est d’incorporer une étape de reconnaissance de formes dans une segmentation d’images [9] ou d’alterner reconnaissance ascendante et détection descendante pour optimiser les deux approches [10, 11, 12]. En télédétection, seuls [13] ont fusionné des méthodes de classification et de reconnaissance de formes pour la détection de bâtiments.

Dans cet article, nous proposons une approche combinant des approches ascendante et descendante pour la détection des bâtiments. Les principales contributions de notre méthode sont les suivantes : une méthode entièrement automatique ne nécessitant qu’une image optique monoculaire, une approche mixte qui combine les avantages des approches ascendante et descendante, et un algorithme indépendant du contexte qui peut être appliqué à de grandes échelles et permet de gérer des scènes avec différentes tailles et configurations de bâtiments.

2 Méthodologie

L’approche proposée est illustrée sur la figure 1, et se décompose en trois étapes principales.

1. Approche ascendante : elle exploite les informations bas-niveau de l’image (section 3). D’une part, une classification supervisée par pixel est effectuée et permet de calculer une carte de probabilité de bâtiments (Figure 1b).

D’autre part, des primitives simples (lignes et régions) sont extraites de manière à contraindre, améliorer et accélérer la détection descendante de motifs rectangulaires (Figures 1c et d).

2. Approche descendante : elle vise également à fournir une carte de probabilité de bâtiments. Nous proposons d’utiliser une modélisation par Processus Ponctuels Marqués (PPM) qui permet de prendre en compte les contraintes géométriques et les interactions spatiales directement au niveau des objets. La carte de probabilité de bâtiments est obtenue par la fusion des extractions de rectangles par accumulation de PPM (section 4) à différents paramétrages (Figure 1e). En effet, puisque les bâtiments peuvent avoir différentes tailles et formes, et sachant que les motifs rectangulaires ne permettent de trouver que l’emplacement grossier des bâtiments, plusieurs hypothèses sont formulées et permettent de mieux délimiter les zones de bâti.

3. Etape de décision : elle fusionne la classification et le partitionnement ascendants avec la détection descendante de motifs rectangulaires. Une segmentation binaire *bâtiment/non-bâtiment* est calculée en utilisant une formulation Markovienne, optimisée par une coupure de graphe (Figure 1f).

Notre approche est très similaire au travail décrit dans [13] avec certaines différences : les étapes de classification et d’extraction de rectangles y sont effectuées alternativement jusqu’à convergence de la classification. En pratique, seuls quelques pixels changent de label après quelques itérations. Dans notre article, nous considérons que l’étape ascendante est suffisamment fiable pour effectuer une unique étape d’affinage par approche descendante. En outre, la succession des étapes d’optimisation entre les deux approches est remplacée par l’accumulation de plusieurs cartes de motifs rectangulaires binaires.

3 Approche ascendante

3.1 Classification supervisée

La première étape consiste, pour chaque pixel de l'image, à calculer la probabilité d'appartenance à la classe *bâtiment* (Figure 1b). Aucune analyse orientée-objet n'est réalisée. Le but de cette étape est de simplement mettre en évidence les contours de bâtiments qui ne peuvent pas être détectés de façon dense par l'approche descendante. Aucune régularisation spatiale n'est nécessaire à cette étape, elle sera effectuée lors de l'approche mixte par régularisation Markovienne.

Une approche supervisée est adoptée afin de mieux saisir la variabilité des matériaux de toiture sur l'image, même si une unique classe *bâtiment* est conservée. Une base de données 2D existante de bâtiments est utilisée à cette fin. L'ensemble d'apprentissage est obtenu automatiquement en sélectionnant aléatoirement 1000 échantillons de la classe *bâtiment* dans la base de données. Enfin, un classifieur SVM, avec un noyau gaussien est utilisé et permet de générer une carte de probabilité de bâtiments [14]. Les quatre bandes spectrales de l'image satellitaire (*bleu, vert, rouge, proche infra-rouge*) sont utilisées comme attributs et paraissent être suffisantes pour se focaliser sur les régions de bâti les plus probables (Figure 1b).

3.2 Détection de contours

L'extraction descendante de motifs rectangulaires est guidée par l'information de gradients dans l'image (Section 4). À cet effet, une carte de gradient améliorée G est calculée, en combinant une image de gradients avec un détecteur de segments, à savoir LSD [15]. Nous supposons que les bâtiments ont de faibles courbures et qu'ils peuvent être représentés par des rectangles : leurs bords peuvent être décomposés en segments. L'algorithme LSD permet d'avoir des segments non exhaustifs mais robustes et fiables, avec des temps de calcul réduits et une précision sub-pixellaire. La valeur de gradient améliorée correspond à la valeur moyenne du module de gradient (rééchantillonnée entre 0 et 1) et la valeur LSD binaire. La figure 1c montre que les bords de bâtiments sont correctement détectés et avec une grande précision. Cette approche peut conduire toutefois à de faux positifs situés à l'intérieur des toits, de la végétation, des zones d'ombre etc. Cependant, ils peuvent être éliminés par l'étape suivante d'extraction de rectangles (Section 4.1).

3.3 Partitionnement de l'image

L'extraction descendante de rectangles nécessite également l'exploration exhaustive de l'espace de configurations de rectangles afin de trouver celles qui correspondent le mieux aux informations bas-niveau extraites de l'image. Le processus peut être accéléré en limitant l'espace de recherche aux zones les plus pertinentes de l'image [16, 17], ici les zones de bâti. Ainsi, la carte de probabilité de la classe *bâtiment* (Section 3.1) est utilisée. Elle est lissée et binarisée en utilisant une segmentation de l'image satellitaire

originale [18]. L'image \mathcal{I} est donc segmentée en n régions R_i , et ensuite partagée en deux sous-ensembles disjoints : $\mathcal{I} = \mathcal{S}_b \cup \mathcal{S}_{nb}$, avec \mathcal{S}_b et \mathcal{S}_{nb} l'ensemble des pixels *bâtiment* et *non-bâtiment*, de tailles n_b et $n - n_b$ respectivement (zones orange et violette sur la Figure 1d). On a :

$$\begin{cases} \mathcal{S}_b &= \bigcup_{i=1..n_b} R_i / \overline{p_{R_i}} > \tau \\ \mathcal{S}_{nb} &= \bigcup_{j=1..n-n_b} R_j / \overline{p_{R_j}} \leq \tau \end{cases} \quad (1)$$

$\overline{p_{R_i}}$ est la probabilité moyenne de la classe *bâtiment* pour R_i et τ le seuil de probabilité à partir duquel une région sera utilisée pour l'extraction de rectangles. En pratique, τ est choisi suffisamment bas (1) pour être sélectionné de façon automatique et (2) pour éviter d'éliminer des régions potentiellement candidates. Par conséquent, τ est fixé de façon à éliminer les zones de végétation et les régions du sol présentant des réponses spectrales similaires. Pour cela, on utilise l'indice de végétation de différence normalisée (NDVI). Un sous-ensemble \mathcal{S}_v de 100 pixels de végétation est sélectionné aléatoirement parmi les 1000 pixels présentant les plus fortes valeurs de NDVI. On a alors : $\tau = \overline{p_{\mathcal{S}_v}}$. La figure 1d montre une carte du partitionnement binaire de l'image satisfaisante. Les zones de bâti sont conservées avec une focalisation grossière, réduisant considérablement les temps de calcul (jusqu'à 40% d'une image peut être supprimée).

4 Détection descendante

Nous supposons que chaque empreinte de bâtiment peut être approchée par un rectangle ou l'union de plusieurs rectangles qui se chevauchent légèrement. Nous présentons d'abord l'extraction des rectangles et ensuite la segmentation finale des bâtiments.

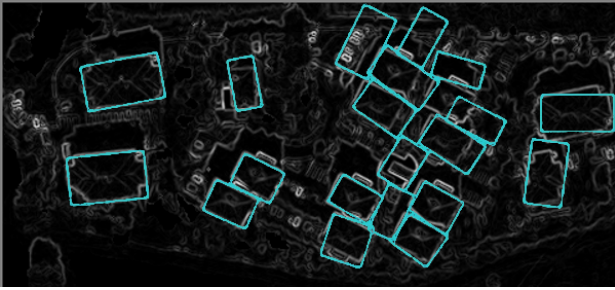
4.1 Extraction de rectangles

Le cadre stochastique des Processus Ponctuels Marqués (PPM) est utilisé pour modéliser l'image par un ensemble de rectangles [19]. Les PPM ont en effet montré toute leur pertinence et leur compétitivité pour l'extraction d'objets paramétriques dans de nombreux problèmes de vision et en particulier pour l'extraction de bâtiments sur des images satellitaires [20]. Les Processus Ponctuels Markoviens sont des modèles probabilistes introduits par [21] afin d'étendre les champs de Markov traditionnels grâce à un formalisme orienté-objet. Ces outils mathématiques exploitent des variables aléatoires dont les réalisations sont des configurations d'objets paramétriques, chaque objet étant assigné à un point positionné dans la scène [17]. Le nombre d'objets est lui-même une variable aléatoire. L'une des forces de ce formalisme est la possibilité d'introduire des interactions spatiales plus ou moins complexes entre les objets tout en imposant des contraintes de régularisation globale sur la scène d'intérêt. Un processus ponctuel est défini par des objets paramétriques (ici des rectangles qui créent une configuration x), une énergie (qui calcule la cohérence entre la configuration x et l'image observée, et prend en compte les interactions entre les rectangles $r \in x$)

et un échantillonneur (qui trouve la configuration minimisant l'énergie). Cette minimisation est complexe puisque le nombre d'objets dans la configuration est inconnu, et que l'énergie n'est pas convexe. Un échantillonneur stochastique est utilisé pour trouver le minimum global de U [19]. En particulier, il s'agit de l'algorithme de Métropolis-Hastings-Green [22] qui appartient à la famille des méthodes de Monte-Carlo par Chaîne de Markov (MCMC). L'énergie U mesure la qualité d'une configuration x et est composée d'un terme d'attache aux données U_d et d'un terme de régularisation (a priori) U_p telle que :

$$U(x) = \sum_{r \in x} U_d(r) + \beta \sum_{\substack{r_i, r_j \in x \\ i \neq j}} U_p(r_i, r_j), \quad (2)$$

avec $\beta \in \mathbf{R}^+$, le coefficient de pondération.



(a)



(b)

FIGURE 2 – Extraction de rectangles par Processus Ponctuels Marqués. (a) Pour un paramétrage donné (cyan – $\tau_G=60$, $C=2300$), en superposition sur la carte de gradient améliorée G . Des faux positifs apparaissent sur les routes puisque leurs bords correspondent aux bords des bâtiments de part et d'autre de la route ou bien sur des files de voitures puisqu'elles renvoient des forts gradients sur les bords. (b) Quand le modèle est paramétré pour des bâtiments larges, un fort taux de fausses détections existe et de grands rectangles couvrent des ensembles de bâtiments.

Attache aux données. Elle est fondée sur le fait que les bords d'un rectangle doivent correspondre à des pixels de forte valeur de gradient. Par conséquent, nous utilisons la carte de gradient améliorée G , sans formuler aucune hypothèse sur la couleur des toits ou l'ombre du bâtiment, afin

de conserver la généralité de l'approche. L'énergie d'attache aux données est définie par :

$$U_d(r) = C - U_G(r). \quad (3)$$

U_G est le terme d'énergie fondé sur le gradient, et C un paramètre de pénalité pour l'ajout de nouveaux rectangles (on favorise les configurations avec un nombre minimal de rectangles). C est une constante pour une zone donnée et est fixée en fonction des valeurs de U_G .

$$U_G(r) = \sum_{m=1}^4 \frac{l_{\tau_G}(r_m)}{l(r_m)} \times \prod_{m=1}^4 G(r_m). \quad (4)$$

r_m est un côté du rectangle r , $l(r_m)$ sa longueur, et $G(r_m)$ sa valeur moyenne de gradient. $l_{\tau_G}(r_m)$ est la longueur maximale de r_m où les valeurs du gradient sont supérieures à un seuil donné $\tau_G \in [50, 100]$.

Le premier terme de l'équation assure que les bords du rectangle ont continuellement de fortes valeurs de gradient, tandis que le second vise à favoriser les valeurs les plus élevées de G sur tous les bords du rectangle. Cette formulation permet de ne pas être pénalisé par des occultations par la végétation et les zones d'ombre. Une faible valeur de τ_G est suffisante pour se concentrer sur les empreintes de bâtiments. En pratique, une évaluation de deux bords orthogonaux des rectangles doit être suffisante pour une détection appropriée, mais pourrait conduire à de fausses détections sur les marquages routiers par exemple.

Régularisation. L'énergie U_p est utilisée pour favoriser des configurations souhaitées entre rectangles voisins et pénaliser les autres. Les objets sont considérés comme voisins s'ils s'intersectent. Un seul bâtiment peut être représenté par un seul rectangle (de voisinage vide) tandis que pour des blocs de bâtiments, l'arrangement des rectangles nécessaires pour les représenter se compose de rectangles contigus ou qui se chevauchent légèrement. Pour obtenir les configurations souhaitées, l'énergie a un seul terme de répulsion. Pour une paire d'objets voisins r_i et r_j , elle correspond tout simplement à leur surface d'intersection S_{\cap} [3, 20], ce qui évite une redondance de détection pour un même bâtiment :

$$U_p(r_i, r_j) = S_{\cap}(r_i, r_j). \quad (5)$$

Optimisation. La recherche de la configuration de rectangles qui minimise l'énergie non convexe U dans un espace de dimension variable se fait à l'aide d'un échantillonneur par saut réversible de Monte-Carlo par Chaîne de Markov (RJCMC) couplé à une relaxation stochastique. Un des avantages de cet algorithme itératif est qu'il ne dépend pas de l'état initial. Plus de détails peuvent être trouvés dans [19]. Les performances en temps de calcul et en stabilité de la convergence (dans le cas de grandes scènes) peuvent être limitées, c'est pourquoi nous utilisons le partitionnement grossier de l'image pour interdire la création de rectangles dans les zones *Non-bâtiment*. On fixe également un a priori sur les tailles maximales des bâtiments pour limiter l'exploration de l'espace des rectangles de l'image.



FIGURE 3 – Multiples détections par PPM sur la même zone d’intérêt (une couleur par réalisation) : des fausses détections persistent mais sont compensées par la fusion des différents résultats de détection de rectangles (Figure 1e).

Carte de probabilité de bâtiments. Le modèle proposé est efficace pour la détection des parties principales du bâtiment, mais souffre de deux limitations : la simplicité du modèle (de fausses détections existent) et la sensibilité au paramétrage (inadaptation quand les objets recherchés sont de taille variable). De plus, puisque le résultat d’une unique optimisation ne fournit qu’une carte binaire de bâtiments, on ne peut l’utiliser directement comme carte de probabilité de bâtiments pour la classification finale. Pour lever ces deux limitations, différents paramétrages réalistes sont proposés et leurs résultats sont fusionnés. Les intervalles de variation des différents paramètres sont fixés de façon expérimentale. Ensuite, les intervalles sont échantillonnés de façon uniforme selon l’impact de chaque paramètre en deux, trois ou quatre sous-intervalles pour respectivement τ_G , C et γ (le paramètre d’élongation du rectangle, cf. [3]). Ceci permet d’obtenir un ensemble de 24 configurations distinctes, noté \mathcal{R} . Enfin, chaque pixel p de l’image \mathcal{I} est associé à une probabilité P_D calculée à partir des rectangles r contenant ce pixel.

$$\forall p \in \mathcal{I}, P_D(p) = \frac{1}{\text{card } \mathcal{R}} \sum_{r \in \mathcal{R}} 1_{\{p \in r\}}, \quad (6)$$

où $1_{\{\cdot\}}$ désigne la fonction indicatrice. La figure 3 montre des résultats sur une zone résidentielle. On observe que l’extraction de rectangles n’est pas parfaite en termes de nombre d’objets et de précision de localisation. Toutefois, malgré la simplicité du modèle, on remarque que la plupart des bâtiments sont détectés et que le processus n’est pas perturbé par les zones d’ombre. La carte de probabilité donnée par l’équation 6 est illustrée sur la figure 1e. On peut voir que l’accumulation des multiples extractions de rectangles fournit une carte fiable de probabilité de bâtiments. Les faibles valeurs de confiance correspondent à des fausses détections ou à des bords de bâtiments, alors que les valeurs les plus élevées correspondent aux toits de bâtiments.

4.2 Segmentation finale

La décision finale est prise au niveau pixellaire et de façon non-supervisée par minimisation d’énergie. Cette énergie

globale d’une coupure \mathcal{C} est composée d’un terme d’attache aux données E_d , dérivé de l’approche de classification ascendante, et une énergie de régularisation notée E_p et définie par la carte de probabilité obtenue par l’approche descendante. On a alors :

$$E(\mathcal{C}) = \sum_{p \in \mathcal{I}} E_d(p) + \lambda \sum_{(p,q) \in \mathcal{I}} E_p(p,q), \quad (7)$$

avec p, q des pixels de l’image \mathcal{I} , q appartenant au voisinage de p , $\lambda > 0$ un paramètre de pondération entre les deux termes d’énergies ascendante et descendante.

L’énergie est minimisée par une coupure de graphe [23, 24] qui renvoie une solution exacte. Le terme d’attache aux données E_d , défini comme suit :

$$E_d(p) = \log(1 + k \times P_{\text{bât}}(p)), \quad (8)$$

avec $P_{\text{bât}}$ la probabilité (ascendante, par SVM) d’appartenance à la classe *bâtiment*, et k un paramètre de confiance dans le processus de labellisation (ici $k=1000$).

E_p vise à favoriser des pixels voisins ayant des probabilités *bâtiment* similaires calculées par l’approche descendante (P_D), et défavorise fortement les paires ayant des valeurs très différentes. Il est défini comme suit :

$$E_p(p,q) = \exp\left(-\frac{\max(P_D) - |P_D(p) - P_D(q)|}{\sigma}\right), \quad (9)$$

avec $\max(P_D)$ la probabilité descendante maximale, σ le facteur de normalisation permettant d’avoir des valeurs E_p similaires quelque soit la performance de l’étape d’extraction multiple de rectangles.

5 Résultats et discussions

5.1 Données

Notre méthode a été évaluée sur une zone mixte urbaine-rurale au sud de Toulouse. Elle couvre 1,5 km² et présente un terrain en pente. Une grande variabilité de bâtiments est présente : de larges bâtiments (une école, un hôpital), 230 petites maisons individuelles, avec des matériaux de toits parfois très différents [5]. Une zone de végétation est également présente avec de la forêt et des terrains cultivés. Une image Pléiades orthorectifiée fusionnée PAN+XS de 0,5 m de résolution a été utilisée.

5.2 Résultats

Pour une analyse détaillée, 3 sous-zones d’intérêt ont été sélectionnées. Le temps de calcul pour chaque zone est d’environ 50 minutes et est principalement dû à l’extraction multiple de rectangles par PPM. Chaque paramétrage de PPM correspond à 15 millions d’itérations et nécessite approximativement 2 minutes. La figure 4 montre des résultats de détection de bâtiments sur les deux sous-zones non présentées sur la Figure 1.

Puisque la classification SVM a été réalisée uniquement à partir des quatre bandes spectrales, on peut voir que la classification ascendante est très sensible aux matériaux

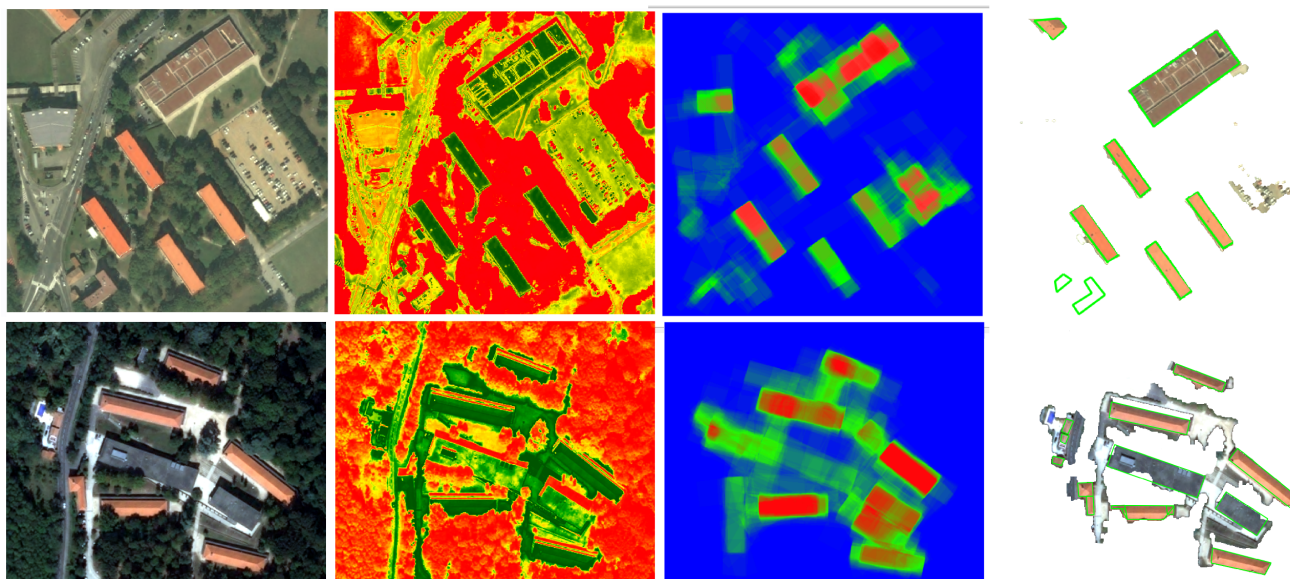


FIGURE 4 – Résultats sur des zones d’intérêt. **De haut en bas** : Zone 2 et Zone 3. **De gauche à droite** : Image satellitaire – carte de probabilité par classification ascendante – carte de probabilité par PPM multiples – Segmentation finale, comparée à la Vérité Terrain (voir Figure 1 pour la légende).

et à l’orientation des toits conduisant à différentes probabilités pour un même toit. De plus, certaines confusions apparaissent entre les toits et les routes quand ils ont des réponses spectrales similaires. Par contre, on peut également voir que la carte de probabilité par PPM est robuste aux zones d’ombre et met bien en évidence les bâtiments même si leur localisation reste grossière (*Zones 1 et 2*). La plupart des bâtiments est détectée, même ceux se trouvant dans les zones boisées. Cependant, certains faux positifs demeurent. Ils correspondent aux zones non bâties où les probabilités ascendante et descendante des *bâtiments* sont fortes (*Zone 3*) : il s’agit des zones au sol avec un contraste élevé telles que les piscines, les files de véhicules dans les parkings ou avec une réflectance très discriminante (*Zones 1, 2 et 3*, respectivement). Toutefois, la combinaison des approches permet de minimiser les erreurs individuelles et l’utilisation de modèles plus avancés en particulier pour la classification devrait résoudre ces problèmes.

L’approche délimite avec précision la plupart des bâtiments. Une analyse quantitative a été effectuée pour évaluer la qualité de l’approche proposée [25]. La classification binaire finale a été comparée à une carte de vérité terrain extraite de la BDTOP© IGN pouvant être parfois imprécise. Des taux de rappel et de précision ont été calculés via une évaluation par pixel (Tableau 1). Ils confirment les impressions visuelles décrites ci-dessus.

6 Conclusion

Nous avons proposé une chaîne de traitements simple et efficace pour combiner des approches ascendante et descendante pour la détection de bâtiments à partir d’une image optique THR monoculaire. Pour chaque étape, des

Zone	Rappel (%)	Précision (%)
1	89	81,5
2	90,5	80,5
3	94	67

TABLE 1 – Qualité de l’extraction de bâtiments par notre méthode, comparée à une base de données topographique de référence.

méthodes de l’état-de-l’art ont été adoptées. L’intérêt de notre méthode est de nécessiter peu de paramètres et peu de connaissances a priori ce qui la rend facilement applicable à grande échelle. En outre, la contribution principale de cet article consiste en la fusion des résultats de multiples processus ponctuels marqués, on s’affranchit du problème délicat du paramétrage des modèles stochastiques. Enfin, ce travail est fondé sur des logiciels libres [26, 27] ce qui rend le travail facilement reproductible.

Les résultats sont satisfaisants tant en termes de détection que de délimitation des bâtiments. D’autres tests de la méthode devront être réalisés sur différents types de paysage et d’autres capteurs.

La méthode est actuellement en cours de comparaison avec les approches développées par [1, 13, 28] afin de mieux quantifier l’apport des différentes techniques employées et en particulier de la pertinence de la combinaison entre méthodes ascendantes et descendantes.

Références

- [1] A. O. Ok, “Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts,” *ISPRS Journal of Photo-*

- grammetry and Remote Sensing*, vol. 86, pp. 21–40, 2013.
- [2] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny, “Structural approach for building reconstruction from a single DSM,” *PAMI*, vol. 32, no. 1, pp. 135–147, 2010.
- [3] M. Brédif, O. Tournaire, B. Vallet, and N. Champion, “Extracting polygonal building footprints from Digital Surface Models : A fully-automatic global optimization framework,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 77, pp. 57–65, 2013.
- [4] J. Wegner, R. Hänsch, A. Thiele, and U. Soergel, “Building detection from one orthophoto and high-resolution insar data using conditional random fields,” *JSTARS*, vol. 4, no. 1, pp. 83–91, 2011.
- [5] A. Le Bris and N. Chehata, “Change detection in a topographic building database using submetric satellite images,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVIII-3/W22, pp. 25–30, 2011.
- [6] C. Senaras, M. Ozay, and F. Yarman Vural, “Building detection with decision fusion,” *JSTARS*, vol. 6, no. 3, pp. 1295–1304, 2013.
- [7] K. Karantzalos and N. Paragios, “Recognition-driven two-dimensional competing priors toward automatic and accurate building detection,” *TGRS*, vol. 47, no. 1, pp. 133–144, 2009.
- [8] C. Benedek, X. Descombes, and J. Zerubia, “Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics,” *PAMI*, vol. 34, no. 1, pp. 33–50, 2012.
- [9] D. Grosgeorge, C. Petitjean, and S. Ruan, “Segmentation d’images par coupe de graphe avec a priori de forme,” in *RFIA*, 2012.
- [10] H. Harzallah, F. Jurie, and C. Schmid, “Combining efficient object localization and image classification,” in *ICCV*, 2009.
- [11] L. Ladicky, P. Sturgess, K. Alahari, C. Russell, and P. Torr, “What, Where & How Many ? Combining object detectors and CRFs,” in *ECCV*, 2010.
- [12] M. Maire, S. Yu, and P. Perona, “Object detection and segmentation from joint embedding of parts and pixels,” in *ICCV*, 2011.
- [13] D. Chai, W. Förstner, and M. Ying Yang, “Combine Markov Random Fields and Marked Point Processes to extract building from remotely sensed images,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-3, pp. 365–370, 2012.
- [14] T.-F. Wu, C.-J. Lin, and R. Weng, “Probability estimates for multi-class classification by pairwise coupling,” *JMLR*, vol. 5, pp. 975–1005, 2004.
- [15] R. G. von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, “LSD : A fast line segment detector with a false detection control,” *PAMI*, vol. 32, no. 4, pp. 722–732, 2010.
- [16] Z. Tu and S. Zhu, “Image segmentation by Data-Driven Markov Chain Monte Carlo,” *PAMI*, vol. 24, no. 5, pp. 657–673, 2002.
- [17] Y. Verdié and F. Lafarge, “Detecting parametric objects in large scenes by Monte Carlo sampling,” *IJCV*, vol. 106, no. 1, pp. 57–75, 2014.
- [18] L. Guigues, J. Cocquerez, and H. L. Men, “Scale-sets image analysis,” *IJCV*, vol. 68, no. 3, pp. 289–317, 2006.
- [19] X. Descombes, Ed., *Applications de la géométrie stochastique à l’analyse d’images*. Lavoisier, 2011.
- [20] M. Ortner, X. Descombes, and J. Zerubia, “Building outline extraction from digital elevation models using marked point processes,” *IJCV*, vol. 72, no. 2, pp. 107–132, 2007.
- [21] A. Baddeley and M. V. Lieshout, “Stochastic geometry models in high-level vision,” *Statistics and Images*, vol. 1, no. 2, pp. 233–258, 1993.
- [22] P. Green, “Reversible Jump Markov Chain Monte-Carlo computation and Bayesian model determination,” *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.
- [23] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” in *ICCV*, 1999, pp. 377–384.
- [24] V. Kolmogorov and R. Zabih, “What energy functions can be minimized via graph cuts,” *PAMI*, vol. 26, no. 2, pp. 147–159, 2004.
- [25] M. Rutzinger, F. Rottensteiner, and N. Pfeifer, “A comparison of evaluation techniques for building extraction from airborne laser scanning,” *JSTARS*, vol. 2, no. 1, pp. 11–20, 2009.
- [26] M. Brédif and O. Tournaire, “Librjcmc : An open-source generic c++ library for stochastic optimization,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXIX-B3, pp. 259–264, 2012. [Online]. Available : <http://librjcmc.ign.fr>
- [27] E. Christophe and J. Inglada, “Open source remote sensing : Increasing the usability of cutting-edge algorithms,” *IEEE Geoscience and Remote Sensing Newsletter*, vol. 35, no. 5, pp. 9–15, 2009.
- [28] J. Niemeyer, F. Rottensteiner, and U. Soergel, “Contextual classification of lidar data and building object detection in urban areas,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 87, pp. 152–165, 2013.