

FUSION TARDIVE D'IMAGES SPOT 6/7 ET DE DONNÉES MULTITEMPORELLES SENTINEL-2 POUR LA DÉTECTION DE LA TACHE URBAINE

Cyril Wendl^{1,2}, Arnaud Le Bris¹, Nesrine Chehata^{1,3}, Anne Puissant⁴, Tristan Postadjian¹

1: Univ. Paris-Est, LASTIG MATIS, IGN, ENSG, 73 Avenue de Paris, F-94160 Saint-Mandé, France

2: Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Suisse

3: EA 4592 G&E Bordeaux INP-Université Bordeaux Montaigne, F-33607 Pessac, France

4: CNRS UMR 7362 LIVE-Université de Strasbourg, 3 rue de l'Argonne, F-67000 Strasbourg, France

Résumé

La fusion d'images multispectrales à très haute résolution spatiale (THR) avec des séries temporelles d'images moins résolues spatialement mais comportant plus de bandes spectrales permet d'améliorer la classification de l'occupation du sol. Elle permet en effet de tirer le meilleur parti des points forts, respectivement, géométriques et sémantiques de ces deux sources. Le travail proposé ici s'intéresse à un processus d'extraction automatique de la tache urbaine fondé sur la fusion tardive de classifications obtenues respectivement à partir d'images satellitaires Sentinel-2 et SPOT 6/7. Ces deux sources sont d'abord analysées indépendamment selon 5 classes, respectivement par Forêt Aléatoire et réseaux de neurones convolutifs. Les résultats sont alors fusionnés afin d'extraire les bâtiments le plus finement possible. Cette étape de fusion inclut une fusion au niveau pixellaire, suivie d'une étape de régularisation spatiale intégrant un terme lié au contraste de l'image. Le résultat obtenu connaît ensuite une seconde fusion afin d'en déduire la tache urbaine en elle-même : une mesure *a priori* de zone urbaine est calculée à partir des objets bâtiments détectés au préalable, puis fusionnée avec une classification binaire dérivée de la classification originale des données Sentinel-2. Les résultats montrent bien la complémentarité des deux sources de données ainsi que la pertinence de l'adoption d'une stratégie de fusion tardive.

Mots clés : Multispectral, Fusion au niveau décision, Fusion tardive, Régularisation, Classification, Tache artificialisée, Tache urbaine, Segmentation, Sentinel-2, SPOT 6/7

Abstract

Fusion of very high spatial resolution multispectral (VHR) images with time series of lower spatial resolution images with more spectral bands can improve land cover classification, combining geometric and semantic advantages of both sources, respectively. This study presents a strategy to extract the urban footprint using decision-level fusion of individual classifications on Sentinel-2 and SPOT 6/7 satellite images. First, both sources are classified separately in five classes, using state-of-the-art supervised Random Forest classification and Convolutional Neural Networks. The two results are then merged in order to extract individual buildings with the highest spatial precision conceivable. Secondly, detected buildings are merged again with the Sentinel-2 classification so as to extract the artificialized areas; the building labels from the regularization step are dilated in order to connect the building objects; a binary classification is derived from the original Sentinel-2 classification before being reintroduced in a fusion and contrast sensitive regularization process so as to eventually retrieve the urban footprint. Results show well the complementary between both data sources as well as the relevance of the late fusion strategy.

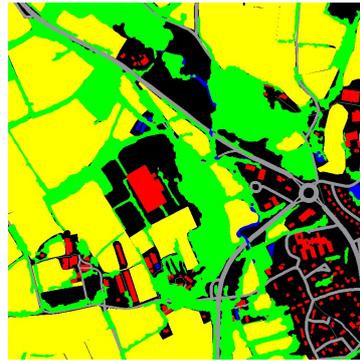
Keywords : Multispectral, Decision fusion, Late fusion, Regularization, Classification, Urban area, Artificialized area, Segmentation

1. Introduction

La connaissance de l'occupation du sol (OCS) est nécessaire au suivi de phénomènes tels que l'extension des zones urbaines et l'imperméabilisation des sols, ainsi que pour prédire les évolutions à venir dans ces cas de figure (Wemmert et al., 2009; Kurtz et al., 2012a,b; Lefebvre et al., 2016a). La télédétection apparaît comme un moyen d'observer l'évolution de ces phénomènes de manière dense et régulière au cours du temps. Aussi, des

approches de classification supervisée exploitant l'imagerie satellitaire ont été étudiées afin d'automatiser le processus de classification de l'occupation des sols (Li et al., 2016; Pesaresi et al., 2016a; Inglada et al., 2017).

Les milieux urbain et périurbain sont connus pour être des milieux complexes constitués de surfaces hétérogènes (surfaces imperméables, arbre, herbe, sol nu, eau, etc) (Ridd, 1995; Weng, 2012). La notion de "surfaces artificialisées" permet ainsi d'appréhender ces es-



■ Bâtiment
 ■ Route
 ■ Forêt
 ■ Eau
 ■ Autre végétation et cultures
 ■ Absence de données

FIGURE 1: Gauche : Image SPOT 6/7. Droite : données vecteur de référence.

paces en tenant compte à la fois de leur composante “bâtie”, mais aussi de leur composante “imperméable” c’est-à-dire les éléments construits tels que les routes, parkings, qui les constituent. Une zone artificialisée peut donc être comprise comme une zone imperméabilisée de manière irréversible par la construction de bâtiments et de routes, incluant de petites structures perméables enclavées comme des jardins, et des espaces verts publics (Puissant et al., 2014). L’agrégation des bâtiments, des routes et de ces autres petites structures enclavées constitue donc la tache artificialisée. Il n’existe pas de définition unique de la tache urbaine. Elle correspond toutefois généralement à une simplification de la tache artificialisée : on en exclut les réseaux routiers en dehors des zones bâties. Ce travail vise donc à la détecter automatiquement à partir de données de télédétection, en s’efforçant d’obtenir un résultat suivant les contours de la ville tels qu’ils apparaissent dans les images et sont discriminés dans un processus d’interprétation visuelle. On souhaite également retrouver les zones bâties isolées (hameaux), pas uniquement les grandes zones urbaines. Pour ce faire, on utilisera les données mise à disposition gratuitement et sur l’ensemble du territoire français dans le cadre du pôle de données Théia¹ et de l’Equipex GEOSUD²

L’imagerie satellitaire à très haute résolution spatiale, comme par exemple celle des capteurs SPOT 6/7, permet de cartographier précisément le contour des objets couvrant la surface terrestre et de caractériser leur texture. Toutefois, son information spectrale n’est pas toujours suffisante pour bien discriminer certaines classes d’OCS (essences végétales, cultures). A l’inverse, certains capteurs, comme Sentinel-2, offrent une plus grande richesse spectrale et temporelle, et donc un accès à une meilleure information sémantique. Généralement, cela se fait au prix d’une moins bonne caractérisation géométrique des objets en présence. La fusion de différentes sources vise alors à combiner les avantages de ces deux types de capteurs pour réduire à la fois les incertitudes

spatiales et sémantiques (Fauvel et al., 2006; Hervieu et al., 2016; Ouerghemmi et al., 2017). Ainsi, ce travail s’intéresse à l’utilisation conjointe de données SPOT 6/7 et Sentinel-2 pour la détection des zones urbaines. Un processus de fusion tardive a été adopté : on considère en effet que les images originales ont pu être classifiées auparavant et indépendamment par des méthodes spécifiques.

Les données SPOT 6/7 et Sentinel-2 sont donc d’abord classées indépendamment suivant la nomenclature à 5 classes suivante : “bâtiment”, “route”, “eau”, “forêt”, “culture et autre végétation”. L’approche mise en œuvre vise alors dans un premier temps à fournir une classification la plus précise possible spatialement des bâtiments (et du réseau routier) par fusion tardive des classifications issues de SPOT 6/7 et Sentinel-2 séparément. Une carte des zones urbaines est ensuite dérivée de ce résultat : une mesure *a priori* de se trouver en zone urbaine est dérivée des objets bâtiments puis fusionnée avec une classification binaire dérivée de la classification originale de la série temporelle Sentinel-2.

1.1. État de l’art

La cartographie et le suivi de cette urbanisation est réalisée depuis de nombreuses années à l’aide de la télédétection satellitaire, en particulier à travers la cartographie notamment des surfaces dites “artificialisées”, avec des approches différentes selon les capteurs.

Une part importante des travaux à large et moyenne échelles s’est focalisée sur l’exploitation de capteurs RADAR, polarimétriques ou non. La forte répétitivité des données et leur emprise globale ont permis l’extraction de tâches urbaines sur de nombreuses villes autour du globe, en partie sur les zones à fort développement urbain ces 30 dernières années (Europe, Chine, Brésil, Inde avant tout) (Fanelli et Santoro, 2003; Chen et al., 2013; Deng et Wang, 2014; Ban et al., 2015; Deng et al., 2015; Xiang et al., 2016; Corbane et al., 2017).

Depuis le début des années 70, de nombreux travaux de recherche se sont focalisés sur l’extraction des sur-

1. <https://www.theia-land.fr/>

2. <http://ids.equipex-geosud.fr/>

faces dites “impermeables” appelées aussi “tache artificialisée” ou tache urbaine (*urban footprint* en anglais) le plus souvent à partir d’images à moyenne et haute résolution spatiale de type Landsat ou SPOT (20 à 30 m) (Wu, 2004; Xian et Crane, 2005; Lu et Weng, 2006; Weng, 2012). Certains travaux ont exploité la notion de texture à partir des images panchromatiques à haute résolution spatiale de type SPOT (Gong et Howarth, 1990; Pesaresi et al., 2008; Pesaresi et Gerhardinger, 2011). Ces travaux ont souvent eu tendance à surestimer ces surfaces et, malgré leur approche mondiale, les résultats ne peuvent pas être exploités à une échelle plus locale d’analyse du territoire.

Plus récemment, de nombreuses études ont porté sur la détection des surfaces bâties à partir de données satellitaires optiques à très haute résolution spatiale (Pesaresi et al., 2008; Xu, 2008; Pesaresi et Ehrlich, 2009; Pesaresi et Gerhardinger, 2011; Pesaresi et al., 2016c). La littérature est extrêmement fournie sur le sujet. L’information de texture et de structure apportée par ce type d’images a alors montré son importance pour la détection de la tache urbaine.

Certaines approches exploitent directement la structure caractéristique des zones construites à l’aide d’indices spécifiques. Ainsi, différents indices calculés à partir de descripteurs texturaux ou fondés sur les segments de l’image ont été définis afin de rendre compte de la structure spécifiques des zones urbaines. Par exemple, un indice nommé panTex a été proposé par (Pesaresi et al., 2008). Il se calcule à partir de mesures de contraste issues d’une matrice de co-occurrence des valeurs en niveau de gris des pixels de l’image et permet une analyse invariante aux rotations de la texture de l’image. Cet indice a fait l’objet d’expériences sur 54 scènes Ikonos et Quickbird sur des villes de différentes parties du Monde (Pesaresi et Gerhardinger, 2011) et a montré sa robustesse à différents types de construction et de conditions d’acquisition des images. Les coins détectés dans les image peuvent servir d’indicateurs du type de structure (artificielle ou naturelle) en présence (Martinez-Fonte et al., 2005). Ainsi, un autre indice de présence de zones bâties est proposé dans (Gueguen et al., 2012) et se fonde sur la densité de coins détectés dans l’image par un opérateur multi-échelles. Cet indice est dédié à des images à très haute résolution spatiale. Il s’avère toutefois que cette densité de coins est très corrélée à l’indice Pantex. Une approche spécifique est mise en œuvre dans (Sirmacek et Ünsalan, 2009). Elle exploite les points détectés par l’algorithme SIFT pour détecter des constructions pour différentes conditions d’acquisition des images et en déduit ensuite une tache urbaine par une méthode reposant sur une modélisation à l’aide de graphes, plus précisément des appariements multiples de sous-graphes. Enfin, la méthode de segmentation des zones urbaines présentée dans (Weizman et Goldberger, 2009) se fonde sur des mots visuels : elle s’appuie sur un dictionnaire construit à partir d’images représentatives des zones urbaines. La comparaison de l’efficacité de trois descrip-

teurs (densité de segments, indice PanTex et indice textural de Gabor) pour la détection des zones urbaines à partir d’images THR effectuée dans (May et Inglada, 2009) montre que l’indice de Gabor s’avère être le plus robuste du fait de sa sensibilité aux alignements.

En dehors de ces approches spécifiques, les zones urbaines peuvent aussi être détectées par des approches de classification standards, comme par exemple dans (Dupuy et al., 2012). Benediktsson et al. (2003) appliquent à l’image des opérations de morphologie mathématique pour extraire des informations de structure utiles pour la détection des zones urbaines. Par ailleurs, il a été montré qu’une utilisation conjointe d’informations radiométriques et texturales améliorerait les résultats (Zhang et al., 2003; Puissant et al., 2005). Dans (Martino et al., 2004), des attributs spectraux et texturaux sont ainsi exploités à l’aide d’un algorithme de clusterisation hiérarchique et un processus de classification supervisée. Des indices fondés sur les lignes droites sont utilisés par Ünsalan et Boyer (2004) pour une classification entre zones urbaines ou non-urbaines. Cette approche est approfondie dans (Ünsalan et Boyer, 2004) en combinant ce type d’information avec des attributs spectraux.

D’autres travaux exploitent des informations contextuelles ou spatiales. Il peut alors s’agir d’informations issues d’une segmentation (approches “objets”) ou d’une analyse multi-échelles. Par exemple, l’approche de Peijun et al. (2007) exploite des caractéristiques morphologiques multi-échelles pour détecter les zones urbaines. Une approche de type *Conditional Random Fields* peut également être mise en œuvre pour apprendre les dépendances au sein de l’image, et combiner des informations bas et haut niveau sur la texture/structure, comme dans (Zhong et Wang, 2007). Une approche objet différente est proposée dans (Kurtz et al., 2012a) : il s’agit d’exploiter l’information de différentes images optiques multi-résolutions au sein d’un processus hiérarchique non supervisé qui extrait progressivement des régions d’intérêt du plus bas au plus haut niveau de résolution spatiale et identifie finalement différents types urbains.

Enfin, les dernières années ont été marquées par l’avènement de l’apprentissage profond, et en particulier les réseaux de neurones convolutifs (*Convolutional Neural Networks* ou CNNs) (Maggiori et al., 2017; Postadjan et al., 2017; Volpi et Tuia, 2017). Ces approches ont en effet montré leur supériorité par rapport aux approches de classification traditionnelles. En effet, contrairement à ces méthodes nécessitant l’extraction d’attributs pertinents en amont du classifieur, les CNNs apprennent directement les attributs (spectraux et texturaux) optimaux (filtres convolutifs) pour le problème de classification à effectuer ainsi que la manière de les exploiter (i.e., le classifieur, processus *end-to-end*). Par ailleurs, de manière implicite, les attributs appris par les CNNs prennent en compte une notion de voisinage et effectuent donc une analyse multi-échelles de l’image. En contrepartie, ces approches d’apprentissage profonds présentent un be-

soin massif de données d'apprentissage. C'est d'ailleurs partiellement grâce aux très grandes bases de données géographiques désormais librement disponibles que ces approches sont revenues sur le devant de la scène et peuvent être envisagées dans des solutions opérationnelles.

L'apparition des données issues de la constellation Sentinel, librement disponibles, a entraîné de nouveaux travaux sur la détection des taches urbaines. En particulier, les images optiques Sentinel-2 présentent une richesse spectrale et temporelle, et demeurent dans le domaine de la haute résolution spatiale pour certaines bandes spectrales. Aussi, certains travaux exploitent à la fois leur radiométrie et leur texture pour détecter les zones urbaines (Pesaresi et al., 2016b; Sabo et al., 2017). D'autres se focalisent sur leur richesse temporelle et spectrale. Ainsi, dans (Inglada et al., 2017), des séries temporelles Sentinel-2 sont analysées conjointement à l'aide d'un classifieur Forêt Aléatoire pour générer des cartes d'occupation du sol annuelles (environ 20 classes sont discriminées, cela varie selon les cas d'utilisation). Par ailleurs, une approche comme celle de (Lefebvre et al., 2016b) peut également être appliquée à ce type de données : chaque date de la série temporelle est classée indépendamment, puis les résultats des différentes dates fusionnés par un processus de type Dempster-Shafer.

Le processus proposé dans ce papier vise à exploiter à la fois l'information des séries temporelles spectralement riches de Sentinel-2 et l'information apportée par la très haute résolution spatiale de SPOT 6/7. Une approche de fusion tardive a été retenue de manière à pouvoir exploiter chaque source par un processus qui lui est le plus adapté. Cela permet en outre l'intégration de résultats issus de services existants et potentiellement indépendants. L'état de l'art précédent nous enseigne en effet que les approches d'apprentissage profond sont optimales pour le traitement des données THR, et que leur résolution nous incite à plutôt y détecter des éléments urbains (bâtiments) et à en déduire les zones urbaines. Pour ce qui concerne les données Sentinel-2, il apparaît intéressant d'exploiter plutôt leur spécificité de série temporelle. Le choix est fait de ne pas exploiter leur information texturale, moins riche que celle des données THR.

2. Zone d'étude et données

Les données utilisées sont d'une part des données de télédétection (images satellitaires Sentinel-2 et SPOT 6/7), et d'autre part les données de référence permettant d'entraîner les différents classifieurs.

2.1. Données de télédétection

2.1.1. Sentinel-2

Sentinel-2 présente le double avantage d'avoir une configuration spectrale plus riche que les capteurs multispectraux classiques et d'acquérir des séries tempo-

relles régulières (5 jours). Seules les 10 bandes dont la résolution spatiale est de 10 ou 20 m ont été utilisées. Toutes ont été ré-échantillonnées à la résolution de 10 m. Seules 6 dates (15 août 2016, 25 janvier 2017, 16 mars 2017, 12 avril 2017 et 25 mai 2017) ont été sélectionnées parmi toutes celles d'une série temporelle d'une durée d'un an. Ces 6 dates ont été retenues pour leur ennuagement très faible et de manière à correspondre dans la mesure du possible à différentes périodes de l'année, et donc à différents aspects des classes d'occupation du sol.

2.1.2. SPOT 6/7

Une image acquise à une seule date (16/04) a été utilisée. Les 4 bandes spectrales (rouge-vert-bleu-proche infrarouge) ont été préalablement fusionnées par "pan-sharpening" à une résolution spatiale de 1,5 m.

2.2. Données de référence

Le processus mis en œuvre s'appuie sur une classification à 5 classes dont la nomenclature est la suivante : "bâtiment", "route", "forêt", "culture et autre végétation", et "surface d'eau". Une carte de vérité terrain à 5 classes (Figure 1) a donc été générée à partir de différentes bases de données géographiques nationales de référence constituant une description de l'occupation du sol topographique (BD Topo[®] de l'IGN³), forestière (BD Forêt[®] de l'IGN⁴) et agricole (Référentiel Parcellaire Graphique⁵). Cette vérité a été utilisée à la fois pour l'apprentissage et l'évaluation, le nombre de pixels utilisés pour l'apprentissage étant faible par rapport à l'intégralité de la carte.

2.3. Zone d'étude

Les expériences ont été effectuées sur une zone test couvrant 648 km² du département du Finistère, dans le Nord-Ouest de la France.

3. Processus mis en œuvre

Le processus proposé (résumé par la figure 2) comporte 3 étapes.

1. La première étape consiste à classer indépendamment l'image SPOT 6/7 et la série temporelle Sentinel-2 suivant la nomenclature à 5 classes décrite précédemment (section 2.2). Pour chaque pixel, une mesure d'appartenance à chacune des classes est fournie à l'issue de cette étape (score de confiance).
2. Lors de la seconde étape, ces classifications à 5 classes sont fusionnées afin d'obtenir en particulier la meilleure détection possible des objets "bâtiments".

3. <http://professionnels.ign.fr/bdtopo>

4. <http://professionnels.ign.fr/bdforet>

5. <http://professionnels.ign.fr/rpg>

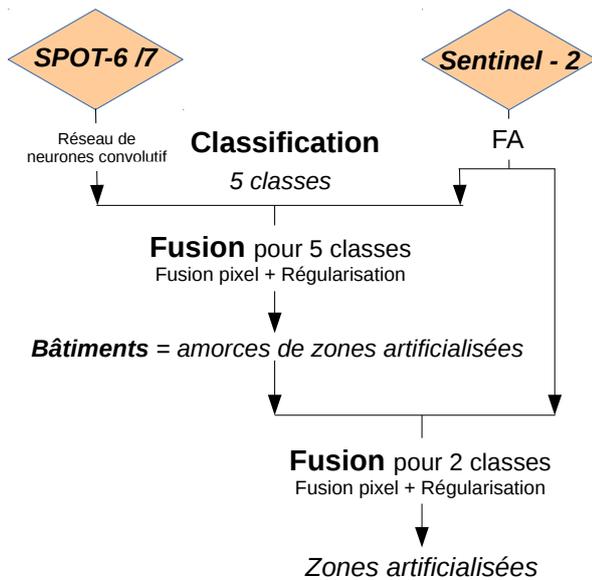


FIGURE 2: Processus proposé.

3. Les objets bâtiments détectés sont alors considérés comme les amorces de la tâche urbaine : ils servent à définir une mesure *a priori* de se trouver en zone urbaine. Cette mesure est ensuite fusionnée avec une classification binaire "urbain"/"non urbain" dérivée de la classification Sentinel-2 originale.

Ce processus comporte donc deux fusions au niveau décision, appliquées d'abord à la classification selon la nomenclature à 5 classes, puis à la classification binaire "urbain"/"non urbain". Ces fusions sont toutes deux effectuées suivant la méthode décrite précisément dans (Ouerghemmi et al., 2017). Cette méthode vise à prendre en compte à la fois les incertitudes sémantiques et spatiales, et se déroule en deux temps. Les mesures d'appartenance aux différentes classes sont d'abord fusionnées au niveau pixelaire par application de règles classiques de fusion de décision. Le résultat de cette fusion subit ensuite une régularisation spatiale visant à lisser le résultat obtenu et en éliminer le bruit résiduel, tout en utilisant l'information de contraste de l'image originale afin de contraindre le résultat à suivre au mieux les contours des objets dans l'image et à obtenir la meilleure précision géométrique possible.

Différentes règles de fusion au niveau pixelaire (cf. section 3.2), ainsi que différentes options de régularisation ont été testées. La combinaison donnant les meilleurs résultats en termes de précision quantitative, de qualité visuelle et d'absence d'artefacts a été retenue.

3.1. Classifications initiales

Les deux sources de données sont classées individuellement avant la fusion, par deux méthodes différentes mais générant toutes deux des mesures d'appartenance aux classes.

La série temporelle Sentinel-2 est classée à l'aide d'un classifieur "Forêt Aléatoire" (FA) (Breiman, 2001), entraîné à partir de 50000 échantillons par classe. Le choix s'est porté sur ce classifieur en partie car il a été utilisé avec succès pour la production des cartes d'occupation du sol du Centre d'Expertise Scientifique OSO⁶ également à partir de séries temporelles Sentinel-2 (Inglada et al., 2017). Les FA sont également adaptées à la classification de données avec un niveau de bruit faible mais non négligeable. On espère par ailleurs à l'avenir utiliser directement ce produit, lorsque des mesures d'appartenance aux classes l'accompagneront, plutôt que de classer les images Sentinel-2 ; aussi souhaitait-on rester proche de son processus de production.

L'image SPOT 6/7 est classée à l'aide d'une méthode d'apprentissage profond : un réseau de neurones convolutifs (Postadjian et al., 2017). Ce type de méthode a en effet montré une grande aptitude à exploiter efficacement l'information texturale et contextuelle de l'image THR. On se trouve par ailleurs dans un contexte où ce type de méthode est éligible du fait de l'importante masse de données d'apprentissage disponible. Il a également été démontré que les classes de routes et de bâtiments pouvaient être discriminées sans nécessité l'introduction d'information altimétrique (Modèle Numérique de Surface par exemple).

3.2. Opérateur de fusion-régularisation

Les deux fusions mises en œuvre au sein du processus proposé sont effectuées suivant la méthode de (Ouerghemmi et al., 2017). Elle consiste en une fusion au niveau pixelaire suivie d'une régularisation spatiale.

3.2.1. Opérateur de fusion par pixel

Plusieurs opérateurs de fusion ont été comparés. Ainsi, différentes règles de fusion de type logique floue (Dubois et Prade, 1997; Zadeh, 1965), bayésiennes ou évidentielles (avec la méthode de Dempster-Shafer (Shafer, 1976)) ont été testées. Ces règles sont présentées plus en détail dans (Ouerghemmi et al., 2017).

Outre ces opérateurs de fusion classiques, des méthodes supervisées ont également été testées. Ces approches consistent à apprendre à partir de données de référence un modèle pour fusionner de manière optimale les deux sources : un classifieur est entraîné à classer des vecteurs d'attributs correspondant à la concaténation au niveau pixel des mesures d'appartenance aux classes selon les deux sources. Trois classifieurs ont été comparés pour cette fusion supervisée : les Forêts Aléatoires (FA), les Séparateurs à Vaste Marge (SVM) avec noyau gaussien (*SVM-gauss*) ou linéaire (*SVM-lin*). Les classifieurs FA (100 arbres) et *SVM-lin* ont été entraînés à partir de 10000 échantillons par classes. Le classifieur *SVM-gauss* a été entraîné à partir d'un ensemble d'apprentissage plus restreint, à savoir 500 échantillons par

6. <https://www.theia-land.fr/fr/ces-occupation-sols>

classe pour des raisons pratiques de temps d'apprentissage plus longs. Ses hyperparamètres ont été optimisés par validation croisée.

3.2.2. Régularisation

Le résultat de la fusion au niveau pixel contient encore du bruit. Aussi, une régularisation spatiale globale (Schindler, 2012) lui est appliquée, afin de le lisser, mais aussi de le contraindre à suivre les contours des objets dans l'image. Tout comme dans (Ouerghemmi et al., 2017), cette étape de régularisation globale consiste à trouver la meilleure configuration de labels C minimisant une énergie E définie sur l'ensemble de l'image :

$$E(P_{fusion}, C) = \sum_{x \in I_{MS}} E_{data}(C(x)) + \lambda \sum_{\substack{x, y \in N \\ x \neq y}} E_{reg}(C(x), C(y)),$$

avec : $E_{data}(C(x)) = 1 - (P_{fusion}(C(x)))$,

$E_{reg}(C(x), C(y)) = 0$ si $C(x) = C(y)$ et

$E_{reg}(C(x), C(y)) = (1 - \gamma) + \gamma V(x, y, \varepsilon)$ sinon.

E_{data} est un terme d'attache à la donnée (c'est-à-dire ici au résultat de la fusion au niveau pixel), tandis qu' E_{reg} mesure la régularisation spatiale. E_{reg} se fonde sur un modèle classique de Potts (Schindler, 2012), privilégiant la classification la plus lisse possible, mais intégrant aussi une information de contraste de l'image. Cela permet de tenir compte du fait que les changements de classes situés dans des zones de fort contraste dans l'image sont plus compréhensibles et doivent donc être plus faiblement pénalisés. Une pondération $\gamma = 0$ correspond à un modèle de Potts pur tandis que $\gamma = 1$ donne tout le poids au terme de contraste. Le terme de contraste $V(x, y, \varepsilon)$ est celui utilisé dans (Ouerghemmi et al., 2017; Hervieu et al., 2016; Rother et al., 2004) :

$$V(x, y, \varepsilon) = \frac{1}{dim} \sum_{i \in [0, dim]} V_i(x, y)^\varepsilon$$

avec $V_i(x, y) = \exp\left(\frac{-(I_i(x) - I_i(y))^2}{2(I_i(x) - I_i(y))^2}\right)$, dim désignant la dimension de l'image I_{MS} et $I_i(x)$ l'intensité du pixel x dans le canal i de l'image.

En pratique, ce terme de contraste est calculé sur une image lissée par un filtre gaussien (d'écart-type 2) afin d'obtenir des contours plus lisses.

Ce problème de régularisation spatiale peut s'exprimer sous la forme d'un problème graphique résolu par coupe de graphe avec α -expansion (Kolmogorov et Zabih, 2004; Schindler, 2012). Dans le cas présent, c'est l'algorithme d'optimisation QPBO (*Quadratic Pseudo-Boolean Optimization*) (Rother et al., 2007) qui a été mis en œuvre pour sa flexibilité.

3.3. Classification binaire

La seconde étape de l'algorithme concerne une classification binaire "urbain/non urbain". Elle implique donc de dériver des mesures d'appartenance à ces deux classes à partir des résultats des étapes précédentes. Les objets bâtiments détectés à l'issue de la fusion des classifications à 5 classes sont considérés comme les

amorces des zones urbaines et utilisés pour définir une mesure *a priori* de s'y trouver : la probabilité d'être en zone urbaine est définie par une fonction linéairement décroissante de 1 à 0 et appliquée au voisinage de tous ces bâtiments, avec une probabilité de 1 de se trouver en zone urbaine au niveau des bâtiments et une probabilité nulle au delà d'une distance de 200 mètres.

Les mesures d'appartenance aux 5 classes de la classification des données Sentinel-2 sont quant à elles converties en une classification binaire, avec deux classes "surface urbaine" (u) et "surface non-urbaine" ($\neg u$). Les mesures d'appartenance $P(u)$ et $P(\neg u)$ à ces deux classes sont calculées comme suit à partir de $P(b)$ et $P(r)$, mesures d'appartenance respectives aux classes "bâtiments" et "routes" de la classification à 5 classes originales : $P(u) = P(b) + P(r)$ et $P(\neg u) = 1 - P(u)$.

4. Résultats et discussion

4.1. Classification à 5 classes

Les figures 3 et 4 présentent des exemples de résultats sur une sous-zone de 0,64 km². Une évaluation quantitative de cette classification à 5 classes a été réalisée, à partir de 5 dalles de taille 3000 × 3000 m (soit une surface totale de 45 km²), réparties sur la totalité de la zone d'étude de 648 km². Le résultat de la fusion a été comparé à la donnée de référence, issue des bases de données géographiques (Figure 1). Les mesures de qualité obtenues sont présentées dans la table 1 pour les classifications individuelles initiales, la fusion au niveau pixel et la fusion régularisée.

4.1.1. Classifications initiales

Les classifications initiales confirment bien le fait que la classification de l'image SPOT 6/7 par réseau de neurones profond a tendance à préserver les petits objets (bâtiments et routes), tandis que la classification de données Sentinel-2 par FA se comporte dans l'ensemble bien mais tend à confondre bâtiments et routes du fait de sa résolution spatiale plus grossière. Les résultats des classifications initiales des images SPOT 6/7 et Sentinel-2 sont visibles sur la figure 3. La zone présentée a été choisie afin d'illustrer les problèmes de ces deux classifications, et les améliorations apportées par la fusion. Signalons toutefois que les classifications initiales montrent de meilleurs résultats sur la plupart des autres zones.

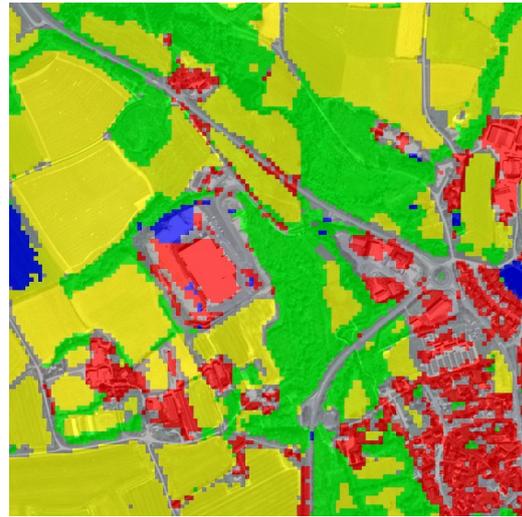
4.1.2. Fusion à 5 classes

Fusion au niveau pixel. Cette première fusion a été effectuée à la résolution spatiale (1,5 m) de l'image SPOT 6/7. On cherche en effet plus particulièrement à obtenir la meilleure détection possible des objets bâtiments, utiles pour les étapes suivantes.

Parmi les opérateurs de fusion classiques de (Ouerghemmi et al., 2017), la règle floue du *Min* est une de celles qui fournissent les meilleurs résultats, avec un bon suivi des contours des objets tout en produisant le moins de confusions entre classes, ainsi qu'on peut le constater d'après



(1) Classification de SPOT 6/7 par réseau de neurones.



(2) Classification de Sentinel-2 par Forêt Aléatoire.

FIGURE 3: Résultats indépendants de classifications pour les deux capteurs traités. L'image SPOT 6/7 apparaît en transparence comme fond de la classifications.

la figure 4 et la table 1.

Sur la sous-zone d'étude de la figure 4, toutes les règles de fusion parviennent à traiter la zone bâtie sur-détectée dans l'image SPOT 6/7 dans la partie supérieure gauche de l'image, préférant alors le résultat de la classification Sentinel-2. La zone industrielle située au milieu de l'image est confondue avec de l'eau dans tous les résultats. Une amélioration est toutefois observée à l'issue de la fusion puisque la partie sud-ouest est alors correctement classée comme de la végétation. Cette confusion avec l'eau trouve son explication dans le fait que la donnée d'entraînement pour les surfaces d'eau comportait des échantillons correspondant dans l'image à des zones asséchées très similaires aux autres classes.

La fusion supervisée impliquant un classifieur (*FA*, *SVM-lin* et *SVM-gauss*) avait initialement tendance à produire des "patches" de bâtiments assez grossiers plutôt que de distinguer des objets bâtiments individualisés. Ceci était dû à l'absence de données d'apprentissage dans les zones comprises entre les bâtiments et les routes sur la carte de référence utilisée pour l'apprentissage. Le fait de rajouter une sixième classe tampon "voisinage de bâtiment" définie comme étant la zone non labellisée située à moins de 10 m autour des objets bâtiments permet à l'issue de la fusion d'affiner les contours des bâtiments et d'atteindre un niveau de détail plus fin que pour la règle de fusion floue *Min* et la règle de Bayes. La fusion supervisée et l'ajout d'une classe additionnelle "voisinage de bâtiment" permettent donc de conserver plus de détails sur les bâtiments individuels, mais peut aussi avoir pour conséquence des confusions entre les bâtiments et cette classe supplémentaire, notamment au centre de grands bâtiments (ainsi qu'on peut le constater sur la figure 4 au niveau de la zone industrielle) ou de supprimer de petites parties de bâtiments. Ceci reste toutefois une exception et ne constitue pas un véritable

problème ici puisque le but visé reste la détection de la tache urbaine. Les mêmes observations ont été faites sur les autres zones test, même celles qui n'avaient pas contribué à fournir des données d'apprentissage pour le modèle de fusion supervisée. Les meilleurs résultats (qualitativement et quantitativement) sont obtenus par le classifieur SVM à noyau gaussien (*SVM-gauss*). Par conséquent, c'est le résultat de la fusion supervisée par *SVM-gauss* avec ajout d'une classe "voisinage de bâtiment" qui sera utilisé dans la suite.

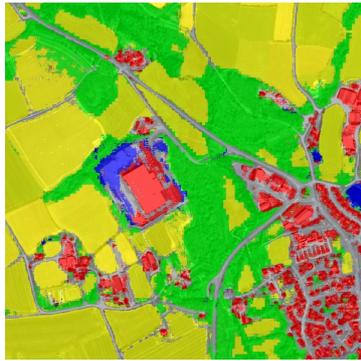
Précisons toutefois ici que lors des évaluations quantitatives globales de la qualité des résultats de la fusion par classification, la classe supplémentaire de "voisinage de bâtiment" a été ignorée.

Régularisation. Les paramètres retenus pour la régularisation sont les suivants : $\lambda = 10$, $\gamma = 0,7$, $\varepsilon = 50$. Ils ont été définis à la fois sur la base de résultats quantitatifs obtenus par validation croisée, mais aussi de manière qualitative selon l'interprétation visuelle du résultat de la régularisation. Le terme de sensibilité au contraste de l'image a été calculé à partir de l'image SPOT 6/7, car cette étape vise ici spécifiquement à obtenir une meilleure détection des objets bâtiments, c'est-à-dire une information plutôt apportée par ce capteur.

Un exemple de résultat est présenté sur la figure 43. La régularisation a éliminé de manière satisfaisante le bruit résiduel, créant un résultat visuellement légèrement meilleur, même si les taux de bonne classification obtenus restent assez similaires à ceux de la fusion par pixel (cf. table 1).

4.2. Classification binaire : détection de la tache urbaine

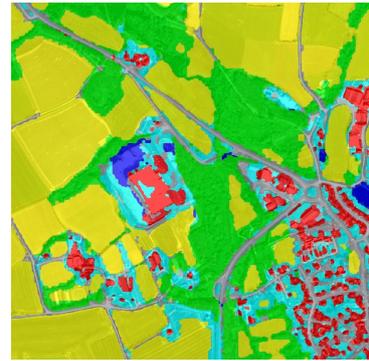
La figure 5 présente les cartes des zones urbaines avant fusion, telles qu'elles peuvent être dérivées direc-



(1) Règle floue du "Min", avant régularisation.



(2) *SVM-gauss*, avec classe supplémentaire, avant régularisation.



(3) *SVM-gauss* avec classe supplémentaire, après régularisation.

■ Bâtiment ■ Route ■ Forêt ■ Eau ■ Autre végétation et cultures ■ Voisinage des bâtiments

FIGURE 4: Résultats de fusion.

Classification	Kappa	PG	PM	Fm	Fb
Sentinel-2	72,0	83,7	81,5	64,6	52,2
SPOT 6/7	73,3	85,2	70,8	63,4	62,5
Floue Min	79,1	88,4	84,7	76,7	73,8
Floue Compromis	78,8	88,2	84,5	76,2	72,5
Bayésien (produit)	79,1	88,5	85,0	76,7	73,8
Dempster-Shafer	79,1	88,4	85,1	76,5	73,7
FA	81,8	90,0	90,1	81,2	81,6
SVM-lin	80,5	89,3	88,6	77,9	80,4
SVM-gauss	81,0	89,6	89,1	79,1	83,4
Après Régularisation					
Floue Min	75,1	85,8	82,3	73,9	73,8
SVM-gauss	81,4	89,8	89,3	79,8	83,9

TABLE 1: Taux de bonne classification (en %) obtenus pour la première partie de la méthode (classification à 5 classes). PG = Précision Globale, PM = Précision Moyenne, Fm = F-Score moyen, Fb = F-Score pour la classe des bâtiments. Pour une description des opérateurs de fusion, voir (Quergemmi et al., 2017).

tement des probabilités calculées à partir des objets bâtiments détectés à l'étape précédente (pour l'opérateur de fusion *SVM-gauss* avec classe supplémentaire) d'une part, et de la classification Sentinel-2 d'autre part. Toutes deux tendent à sur-détecter les zones urbaines. La première le fait de manière artificielle du fait de la zone tampon de taille constante créée autour des objets bâtiments par le calcul de la mesure *a priori* d'être en zone urbaine. La seconde intègre notamment le réseau routier.

Ces deux sources d'information sont d'abord fusionnées pixel par pixel par un opérateur de fusion. Les approches de fusion supervisées ne sont pas éligibles dans ce cas, du fait de l'absence de données d'apprentissage pour les classes "urbain"/"non urbain". On doit donc se limiter aux règles de fusion classiques. Plusieurs d'entre elles donnent des résultats visuellement assez similaires. C'est finalement la règle floue du "Min" qui a été retenue. La régularisation globale est alors appliquée pour le même jeu de paramètre $(\lambda, \gamma, \epsilon)$ que dans le cas précédent à 5

classes. La figure 5 présente des exemples de résultats obtenus. Les routes ainsi que quelques erreurs ont été éliminées par rapport à la détection à partir de données Sentinel-2. Les contours de la tache urbaine sont plus cohérents avec ce que l'on observe sur l'image en comparaison de la tache urbaine générée à partir de la dilatation des objets bâtiments issus de la première fusion.

En l'absence d'une donnée de référence pour la tache urbaine, l'évaluation quantitative des résultats n'était pas possible. Les résultats ont toutefois été comparés (Figure 6) à d'autres produits correspondant à cette thématique : - un masque issu de la dilatation pour un rayon de 20 m des objets bâtiments de la BD Topo®. Le résultat de la détection automatique est généralement assez semblable à ce masque, les principales différences apparaissant sur les frontières du masque issu de la BD Topo. Ceci s'explique par le fait que la BD Topo a été dilatée avec un rayon fixe tandis que la détection automatique tend à mieux suivre les contours des objets dans l'image.

- un masque calculé à partir de la carte d'occupation du sol du CES OSO, par agrégation des classes "urbain diffus", "urbain dense" et "zones industrielles et commerciales". Par rapport au résultat de la détection automatique par fusion SPOT 6/7 et Sentinel-2, ce produit comporte un certain nombre de trous au sein des zones urbaines. Ceci est en partie dû à la résolution spatiale des données (pas de THR pour le produit OSO) mais aussi à une sélection différentes des classes considérées comme artificialisée ou urbaine.

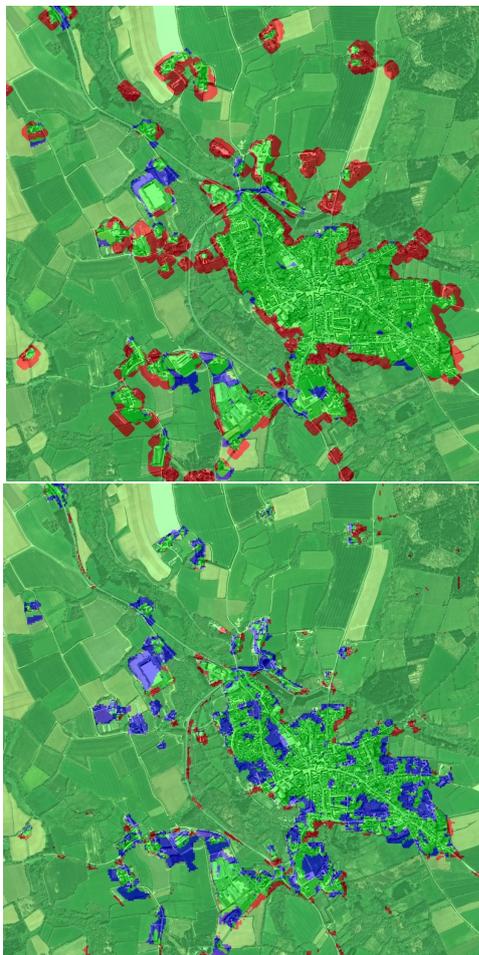
5. Conclusion

Un processus de détection de la tache urbaine à partir de données de télédétection complémentaires et librement disponibles sur l'intégralité du territoire français a été mis en œuvre. Des données Sentinel-2 et SPOT 6/7 sont d'abord classées indépendamment suivant une nomenclature constituée de 5 classes topographiques. Une fusion de décision associée à une régularisation spatiale est alors appliquée à ces résultats afin d'en tirer le meilleur parti et de préserver un bon niveau de détail



■ Tache urbaine ■ Tache non-urbaine

FIGURE 5: Tache urbaine : de gauche à droite, masque dilaté des bâtiments détectés à l'étape précédente, classification binaire Sentinel-2 et résultat de la fusion.



■ Accord ■ Sous-détection ■ Sur-détection

FIGURE 6: Évaluation de la détection de la tache urbaine. De haut en bas, comparaison, avec un masque dilaté de bâtiments de la BD Topo®, la fusion des zones urbaines du produit OSO.

géométrique tout en réduisant les confusions. Les règles de fusion classiques permettent de bien éliminer les erreurs mais conservent souvent des confusions entre bâtiments et routes, conduisant à un résultat assez grossier avec des patches de bâtiments connectés. À l'inverse, les méthodes de fusion supervisées conduisent à une meilleure détection des bâtiments si l'on prend en compte lors de la fusion une classe supplémentaire "voisinage de bâtiment" générée artificiellement à partir du jeu de données d'apprentissage.

La tache urbaine est ensuite obtenue à l'issue d'une seconde fusion associée à une régularisation spatiale. Cette fois, la fusion porte sur les probabilités de se trouver en zone urbaine selon la classification Sentinel-2 et selon une mesure *a priori* dérivée des bâtiments détectés à l'étape précédente.

Des améliorations pourraient vraisemblablement être obtenues en modifiant cette mesure *a priori* de se trouver en zone urbaine, notamment de manière à la faire décroître plus rapidement pour les objets bâtiments les plus incertains.

Par ailleurs, les résultats prometteurs de la fusion supervisée à l'aide de classifieurs classiques justifieraient l'usage pour la fusion de méthodes plus avancées comme les méthodes d'apprentissage profond. Leur usage est possible ici pour la première étape (travail à 5 classes), mais pas pour la détection de la tache urbaine par manque de données d'apprentissage.

Remerciement

Ce travail a bénéficié de fonds publics reçus dans le cadre de GEOSUD, un projet (ANR-10-EQPX-20) du programme "Investissements d'Avenir" de l'Agence Nationale pour la Recherche (ANR). Il a été en partie financé par le projet TOSCA "Artificialisation - Urbanisation" du CNES, sous l'égide du Centre d'Expertise Scientifique du même nom du pôle de données Théia.

Références

- Ban, Y., Jacob, A., Gamba, P., 2015. Spaceborne SAR data for global urban mapping at 30 m resolution using a robust urban extractor. *ISPRS Journal of Photogrammetry and Remote Sensing* 103, 28–37.
- Benediktsson, J. A., Pesaresi, M., Arnason, K., 2003. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing* 41, 1940–1949.
- Breiman, L., 2001. Random forests. *Machine Learning* 45 (1), 5–32.
- Chen, Z., Zhang, Y., Guindon, B., Esch, T., Roth, A., Shang, J., 2013. Urban land use mapping using high resolution SAR data based on density analysis and contextual information. *Canadian Journal of Remote Sensing* 38 (6), 738–749.
- Corbane, C., Lemoine, G., Pesaresi, M., Kemper, T., Sabo, F., Ferri, S., Syrris, V., 2017. Enhanced automatic detection of human settlements using Sentinel-1 interferometric coherence. *International Journal of Remote Sensing* 39 (3), 842–853.
- Deng, L., Wang, C., 2014. Improved building extraction with integrated decomposition of time-frequency and entropy-alpha using polarimetric SAR data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (10), 4058–4068.
- Deng, L., Yan, Y.-N., Sun, C., 2015. Use of sub-aperture decomposition for supervised PolSAR classification in urban area. *Remote Sensing* 7, 1380–1396.
- Dubois, D., Prade, H., 1997. Possibility theory and data fusion in poorly informed environment. *Control Engineering Practice* 2, 811–823.
- Dupuy, S., Barbe, E., Balestrat, M., 2012. An object-based image analysis method for monitoring land conversion by artificial sprawl use of rapideye and irs data. *Remote sensing* 4 (2), 404–423.
- Fanelli, A., Santoro, M., 2003. Characterization of urban areas using principal component analysis from multitemporal ERS coherence imagery. *Proc. SPIE 4886, Remote Sensing for Environmental Monitoring, GIS Applications, and Geology II*.
- Fauvel, M., Chanussot, J., Benediktsson, J., 2006. Decision fusion for the classification of urban remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 44, 2828 – 2838.
- Gong, P., Howarth, P., 1990. The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. *Photogrammetric Engineering and Remote Sensing* 56 (1), 67–73.
- Gueguen, L., Soille, P., Pesaresi, M., 2012. A new built-up presence index based on density of corners. Dans : *IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 22-27 July 2012, Munich, Germany. pp. 5398–5401.
- Hervieu, A., Le Bris, A., Mallet, C., 2016. Fusion of hyperspectral and VHR multispectral image classifications in urban areas. Dans : *ISPRS Annals of the Photogrammetric, Remote Sensing and Spatial Information Sciences*. Vol. III-3. pp. 457–464.
- Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., Rodes, I., 2017. Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sensing* 9 (1), 95.
- Kolmogorov, V., Zabih, R., 2004. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (2), 65–81.
- Kurtz, C., Passat, N., Gançarski, P., Puissant, A., 2012a. Extraction of complex patterns from multiresolution remote sensing images : A hierarchical top-down methodology. *Pattern Recognition* 45 (2), 685 – 706.
- Kurtz, C., Passat, N., Gançarski, P., Puissant, A., 2012b. A histogram semantic-based distance for multiresolution image classification. Dans : *IEEE International Conference on Image Processing*, IEEE, 30 September-2 October, Orlando, USA.
- Lefebvre, A., Sannier, C., Corpetti, T., 2016a. Monitoring urban areas with Sentinel-2A data : Application to the update of the Copernicus high resolution layer imperviousness degree. *Remote Sensing* 8 (7).
- Lefebvre, A., Sannier, C., Corpetti, T., 2016b. Monitoring urban areas with Sentinel-2A data : Application to the update of the Copernicus high resolution layer imperviousness degree. *Remote Sensing* 8 (7), 606.
- Li, M., Stein, A., Bijker, W., Zhan, Q., décembre 2016. Urban land use extraction from Very High Resolution remote sensing imagery using a Bayesian network. *ISPRS Journal of Photogrammetry and Remote Sensing* 122, 192–205.
- Lu, D., Weng, Q., 2006. Spectral mixture analysis of ASTER images for examining the relationship between urban thermal features and biophysical descriptors in Indianapolis, Indiana, USA. *Remote Sensing of Environment* 104 (2), 157–167.
- Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geosciences and Remote Sensing* 55 (2), 645–657.
- Martinez-Fonte, L., Gautama, S., Philips, W., Goeman, W., 2005. Evaluating corner detectors for the extraction of man-made structures in urban areas. Dans : *International Geoscience and Remote Sensing Symposium*, IEEE, 25-29 July, Seoul, South Korea.
- Martino, M. D., Macchiavello, G., Serpico, S. B., 2004. Partially supervised classification of optical high spatial resolution images in urban environment using spectral and textural information. Dans : *IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 20-24 September, Anchorage, USA.
- May, S., Inglada, J., 2009. Urban area detection and segmentation using OTB. Dans : *IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 13-17 July, Cap Town, South Africa.
- Ouerghemmi, W., Le Bris, A., Chehata, N., Mallet, C., 2017. A two-step decision fusion strategy : Application to hyperspectral and multispectral images for urban classification. *International Archives of the Photogrammetric, Remote Sensing and Spatial Information Sciences*, XLII-1-W1, 167–174.
- Peijun, L., Hongtao, H., Jiancong, G., 2007. Segmentation of high-resolution multispectral image based on extended morphological profiles. Dans : *IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 23-27 July, Barcelona, Spain.
- Pesaresi, M., Corbane, C., Julea, A., Florczyk, A., Syrris, V., Soille, P., 2016a. Assessment of the added-value of sentinel-2 for detecting built-up areas. *Remote Sensing* 8 (4).
- Pesaresi, M., Corbane, C., Julea, A., Florczyk, A. and Syrris, V., Soille, P., 2016b. Assessment of the added value of Sentinel-2 for detecting built-up areas. *Remote Sensing* 8 (4), 299.
- Pesaresi, M., Ehrlich, D., 2009. *Global Mapping of Urban Settlements*. London U.K. Taylor and Francis, Ch. A methodology to quantify built-up structures from optical VHR imagery, pp. 27–58.
- Pesaresi, M., Gerhardinger, A., 2011. Improved textural built-up presence index for automatic recognition of human settlements in arid regions with scattered vegetation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4 (1), 16–26.
- Pesaresi, M., Gerhardinger, A., Kayitakire, F., 2008. A robust built-up area presence index by anisotropic rotation-invariant textural measure. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 1, 180–192.
- Pesaresi, M., Syrris, V., Julea, A., 2016c. A new method for

- Earth observation data analytics based on Symbolic Machine Learning. *Remote Sensing* 8 (5), 399.
- Postadjian, T., Le Bris, A., Sahbi, H., Mallet, C., 2017. Investigating the potential of deep neural networks for large-scale classification of very high resolution satellite images. *ISPRS Annals of the Photogrammetric, Remote Sensing and Spatial Information Sciences IV-1-W1*, 183–190.
- Puissant, A., Hirsch, J., Weber, C., 2005. The utility of texture analysis to improve per-pixel classifications for high to very high spatial resolution imagery. *International Journal of Remote Sensing* 26, 733–745.
- Puissant, A., Rougier, S., Stumpf, A., 2014. Object-oriented mapping of urban trees using Random Forest classifiers. *International Journal of Applied Earth Observation and Geoinformation* 26, 235–245.
- Ridd, M. K., 1995. Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing : comparative anatomy for cities. *International Journal of Remote Sensing* 16, 2165–2185.
- Rother, C., Kolmogorov, V., Blake, A., 2004. "GrabCut" : Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics* 23 (3), 309–314.
- Rother, C., Kolmogorov, V., Lempitsky, V., Szummer, M., 2007. Optimizing binary MRFs via extended roof duality. Dans : *Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Sabo, F., Corbane, C., Ferri, S., 2017. Inter-sensor comparison of built-up derived from Landsat, Sentinel-1, Sentinel-2 and SPOT5/SPOT6 over selected cities. *Rapport technique, JRC*.
- Schindler, K., novembre 2012. An Overview and Comparison of Smooth Labeling Methods for Land-Cover Classification. *IEEE Transactions on Geoscience and Remote Sensing* 50 (11), 4534–4545.
- Shafer, G., 1976. *A Mathematical Theory of Evidence*. Princeton University Press.
- Sirmaçek, B., Ünsalan, C., 2009. Urban-area and building detection using SIFT keypoints and graph theory. *IEEE Transactions on Geoscience and Remote Sensing* 47 (4), 1156–1167.
- Ünsalan, C., Boyer, K., 2004. Classifying land development in high-resolution panchromatic satellite images using straight-line statistics. *IEEE Transactions on Geoscience and Remote Sensing* 42 (4), 907 – 919.
- Unsalan, C., Boyer, K., 2004. Classifying land development in high-resolution satellite imagery using hybrid structural-multispectral features. *IEEE Transactions on Geoscience and Remote Sensing* 42 (12), 2840–2850.
- Volpi, M., Tuia, D., 2017. Dense semantic labeling of sub-decimeter resolution images with convolutional neural networks. *IEEE Transactions on Geosciences and Remote Sensing* 55 (2), 881–893.
- Weizman, L., Goldberger, J., 2009. Urban-area segmentation using visual words. *Geoscience and Remote Sensing Letters, IEEE* 6 (3), 388–392.
- Wemmert, C., Puissant, A., Forestier, G., Gancarski, P., 2009. Multiresolution Remote Sensing Image Clustering. *IEEE Geoscience and Remote Sensing Letters* 6 (3), 533–537.
- Weng, Q., 2012. Remote sensing of impervious surfaces in the urban areas : Requirements, methods, and trends. *Remote Sensing of Environment* 117, 34–49.
- Wu, C., 2004. Normalized spectral mixture analysis for monitoring urban composition using ETM+ imagery. *Remote Sensing of Environment* 93 (4), 480–492.
- Xian, G., Crane, M., 2005. Assessments of urban growth in the Tampa Bay watershed using remote sensing data. *Remote Sensing of Environment* 97 (2), 203–215.
- Xiang, D., Tang, T., Hu, C., Fan, Q., Su, Y., 2016. Built-up area extraction from PolSAR imagery with model-based decomposition and polarimetric coherence. *Remote Sensing* 8 (8), 685.
- Xu, H., 2008. A new index for delineating built-up land features in satellite imagery. *International Journal of Remote Sensing* 29 (14), 4269–4276.
- Zadeh, L. A., 1965. Fuzzy sets. *Information and Control* 8, 338–353.
- Zhang, Q., Wang, J., Gong, P., Shi, P., 2003. Urban spatial pattern analysis from SPOT panchromatic imagery using textural analysis. *International Journal of Remote Sensing* 24, 4137–4160.
- Zhong, P., Wang, R., 2007. A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images. *IEEE Transactions on Geoscience and Remote Sensing* 45 (12), 3978–3988.