

# Sur l'évaluation et l'élaboration d'un jeu de données de référence de bonne qualité en télédétection

Andrés Troya-Galvis\* Pierre Gançarski\* Isabelle Mougenot\*\* Laure Berti-Équille\*\*

\*ICube, Université de Strasbourg,  
300 bd Sébastien Brant - CS 10413 - F-67412 Illkirch Cedex  
{troyagalvis,gancarski}@unistra.fr,

\*\*UMR 228 Espace Dev (UM,UR,UG,UA,IRD), Maison de la Télédétection,  
500 rue JF Breton, 34093 Montpellier Cedex 5  
isabelle.mougenot@umontpellier.fr, laure.berti@ird.fr

**Résumé.** En analyse d'images de télédétection, les données de référence, venant étiqueter les objets des images, y jouent un rôle crucial mais sont parfois imprécises voire incertaines et en nombre limité. Dans cet article, nous présentons une méthodologie pour l'amélioration de données de référence pour la télédétection en trois étapes : réalignement des données, évaluation via crowdsourcing et création d'un jeu de données de référence de bonne qualité.

## 1 Introduction

L'analyse d'images de télédétection consiste à associer à chaque pixel d'une image optique captée par un satellite ou un engin aérien, une sémantique liée à un domaine précis, par exemple à l'analyse urbaine (Puissant et al., 2014). En analyse d'images à très haute résolution spatiale, les approches basées objets (Blaschke, 2010) sont de prédilection. Ces approches mettent en œuvre une étape de segmentation suivie d'une étape de classification. Les données de référence (i.e., exemples de segments étiquetés au préalable par un expert) jouent alors un rôle important dans ce processus d'analyse d'images pour l'entraînement de modèles de classification supervisée, mais aussi pour l'évaluation objective des résultats. De fait, un jeu de données de référence de bonne qualité se doit d'avoir deux propriétés essentielles : fournir une labellisation la plus complète et précise possible des objets d'intérêt dans l'image (qualité de la classification) ; et permettre d'aligner correctement les contours des segments de référence avec les objets de l'image (qualité de la segmentation).

Il est tentant d'employer des données issues de bases de données géographiques comme données de référence. En effet, ces données respectent généralement la première propriété. En revanche, la deuxième propriété n'est pas toujours satisfaite. En effet, les contours des polygones présentent souvent un décalage non régulier dû par rapport aux objets d'intérêt (Sublime et al., 2015).

Les données de référence exploitées dans notre travail ont été acquises par le SERTIT<sup>1</sup> au moyen d'une méthodologie d'intégration de données multi-sources. Elles consistent en 79 796

---

1. <http://sertit.u-strasbg.fr/>

polygones regroupés en 16 classes thématiques. Ces données semblent tout-à-fait utilisables comme données de référence en télédétection. Pour exemple, la figure 1a laisse apparaître une complétude des données ; les polygones correspondant aux bâtiments et aux routes semblent également très réguliers et bien délimités. Néanmoins, la figure 1b révèle que les polygones ne s'alignent pas correctement avec les objets dans l'image, et ce notamment, au niveau des bâtiments. De plus, ce désalignement n'est pas régulier sur toute l'image. Par conséquent, l'utilisation en l'état de ces données entraînerait des résultats biaisés et dont l'erreur et l'incertitude sont difficiles à estimer.

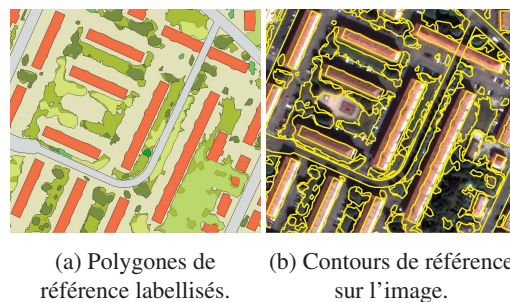


FIGURE 1: Exemple de données de référence issues de bases de données géographiques.

Dans l'objectif d'améliorer ce type de données et de les rendre exploitables comme données de référence dans un contexte d'analyse d'images de télédétection que ce soit en segmentation ou en classification, nous proposons une méthodologie qui comprend trois étapes : le réalignement des données ; l'évaluation des données réalignées à l'aide du crowdsourcing ; et la création d'un jeu de données de segments fiables.

## 2 Réalignement des données

En vue d'améliorer ces données de référence et ainsi obtenir des données plus adaptées à notre problématique, nous avons défini une procédure qui se sert d'une sur-segmentation pour corriger les décalage des données de référence. Cette procédure, illustrée par la figure 2, prend en entrée un ensemble de segments labellisés dont les contours s'alignent mal avec les objets d'intérêt et une sur-segmentation de l'image. Elle consiste alors à superposer les segments de

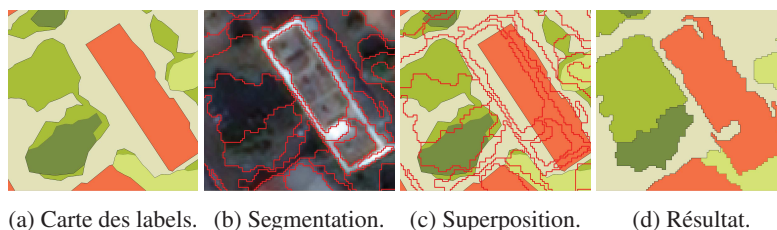


FIGURE 2: Illustration de la procédure d'amélioration des données de référence.

la sur-segmentation avec la carte de labels et à leur attribuer le label majoritairement présent. Finalement, tous les segments adjacents ayant le même label sont fusionnés entre eux. Cette procédure permet effectivement d'obtenir des données de référence améliorées. En effet, les frontières des segments de la sur-segmentation ont tendance à s'aligner avec les objets d'intérêt. Ainsi, l'attribution du label majoritaire permet d'obtenir des segments de bonne qualité et qui semblent correctement labellisés. Enfin, la procédure de fusion en post-traitement permet de réduire les erreurs de sur-segmentation.

### 3 Crowdsourcing pour évaluer la qualité des données de référence

Afin d'évaluer la qualité des données réalignées de façon objective, nous avons eu recours à « la sagesse des foules » dans une démarche de crowdsourcing. Le crowdsourcing consiste à diviser un problème complexe en sous-tâches facilement réalisables par une personne ayant un minimum d'entraînement ou d'expertise sur un domaine. Ces tâches sont alors soumises à un grand nombre de personnes, couramment appelées contributeurs (la « foule »), qui vont les traiter jusqu'à ce que le problème initial soit résolu. La pratique du crowdsourcing a gagné rapidement en popularité depuis que Howe (2006) a avancé ce terme pour décrire les activités d'externalisation ouverte mobilisant un grand nombre de personnes agissant de manière indépendante. Le crowdsourcing a ainsi été employé avec succès dans une large variété d'applications, notamment en télédétection (Barrington et al., 2012), mais aussi en marketing (Whitla, 2009), ou pour la découverte de nouveaux médicaments (Lessl et al., 2011), entre autres. Néanmoins, à notre connaissance, elle n'a jamais été explorée pour l'évaluation de la segmentation d'images de télédétection. Par conséquent, l'objectif et l'intérêt de cette expérience est double : d'une part, quantifier la qualité de la segmentation réalignée ainsi que de la labellisation qui lui est associée ; et d'autre part, estimer la pertinence du crowdsourcing pour le traitement des données issues de la segmentation d'images de télédétection.

Pour effectuer cette tâche de crowdsourcing nous avons décidé d'utiliser la plateforme CrowdFlower<sup>2</sup>. L'expérience a été effectuée en deux phases différentes que nous présentons et analysons ci-après.

#### 3.1 Première phase de crowdsourcing

La première phase de crowdsourcing a consisté à évaluer 10 000 segments choisis de manière aléatoire parmi les 26 872 segments de la segmentation réalignée. Pour chaque segment, un extrait d'image montrant le segment à évaluer est présentée (figure 3a). Une carte interactive Google Maps centrée sur le barycentre du segment à évaluer est également présentée pour faciliter la tâche d'interprétation (figure 3b). Les questions suivantes étaient alors posées aux participants :

1. le segment semble-t-il sur-segmenté (trop petit), sous-segmenté (trop large) ou de la bonne taille par rapport à l'objet géographique ;
2. quelle est la classe de l'objet géographique sous-jacent ? (choix unique parmi une liste prédéfinie)

---

2. <https://crowdflower.com>

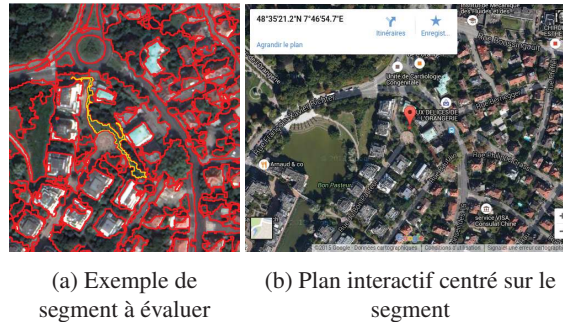


FIGURE 3: Exemple de segment soumis à évaluation lors du crowdsourcing.

Ces questions ont été soumises à 165 contributeurs d'origines différentes (États-Unis et Europe). Pour chacun des 10 000 segments, nous avons obtenu 3 avis. Nous présentons ci-après une analyse purement factuelle des données résultantes ainsi qu'une analyse d'un point de vue thématique.

**Analyse statistique des résultats :** En ce qui concerne l'évaluation de la segmentation, un total de 5 329 segments ont été jugés comme bien segmentés, 2 719 segments jugés comme sur-segmentés, et 681 segments ont été jugés comme sous-segmentés. Bien que le nombre de segments sous-segmentés soit moindre, ils représentent est plus importante que celle des segments sur-segmentés puisqu'il s'agit la plupart du temps de segments très larges. Il existe également 1 271 segments avec des avis en conflit. L'analyse des labels donnés par la foule a révélé un fait surprenant : il existe une grande variabilité entre les labels attribués par la foule et ceux d'origine, il n'y a pas de tendance évidente qui puisse expliquer ce fait, et une étude thématique plus poussée s'est avérée nécessaire pour expliquer ce phénomène et pour déterminer la pertinence de ces données labellisées.

**Analyse thématique des données :** Cette analyse a été réalisé avec l'aide d'un expert géographe et nous a permis d'identifier trois problèmes liés aux données d'origine ainsi qu'aux évaluations des contributeurs. 1) Les classes présentes dans les données de référence ne sont pas distribuées uniformément. En effet, il existe une prédominance des classes de végétation et de bâtis. 2) Plusieurs incohérences entre l'évaluation du segment en termes de segmentation et le label donné à celui-ci ont été détectées. Cela laisse penser que les instructions du crowdsourcing n'étaient pas clairement énoncées. 3) La plupart des segments évalués n'ont pas été labellisés de manière unanime par les contributeurs, et dans certains cas, les labels attribués étaient contradictoires.

### 3.2 Deuxième phase de crowdsourcing

Nous avons mis en place une deuxième campagne de crowdsourcing, en corrigeant les choix techniques et de présentation qui se sont avérés peu pertinents auparavant. Notamment, la description de la tâche a été simplifiée et améliorée grâce à l'ajout de plusieurs exemples explicites ; les labels proposés aux contributeurs ont été simplifiés et regroupés dans des classes plus génériques et plus courants ; le barycentre des du segment a été enlevé de la carte Google

map ; et les questions du questionnaire ont été retravaillées pour les rendre plus faciles à comprendre.

Pour ce crowdsourcing, deux sous-ensembles de segments différents ont été évalués. Le premier sous-ensemble correspond aux segments ayant obtenu des évaluations presque consensuelles lors de la première phase. Le deuxième sous-ensemble correspond aux segments ne présentant pas de consensus. Cette fois-ci, la variabilité des réponses a été nettement moins importante. Pour l'ensemble de segments presque consensuels, la seule classe où il reste des confusions considérables est « Zone artificielle ». Pour le sous-ensemble de segments non consensuels, on remarque la présence d'un nombre très important de segments avec le label « Inconnu ».

Les résultats de cette deuxième phase ont été beaucoup plus consensuels et cohérents que ceux de la première phase. La présentation des données et de la tâche est donc cruciale au bon déroulement d'un crowdsourcing. Il est également important de faire une description claire et concise de la tâche et de bien spécifier l'objectif de celle-ci, en évitant le vocabulaire trop spécialisé.

## 4 Création d'un jeu de données fiable

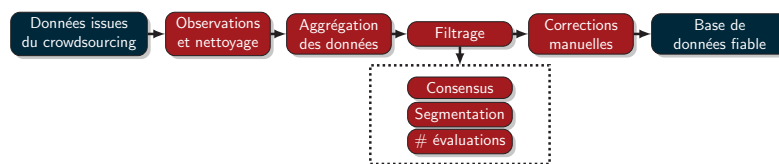


FIGURE 4: Méthodologie utilisée pour la production du jeu d'exemples.

Afin de nettoyer et d'agréger les données obtenues au moyen du crowdsourcing, nous avons mis en place la méthodologie présentée dans la figure 4. Tout d'abord, une étape d'analyse statistique et empirique des données permet d'éliminer les éléments trop contradictoires. Puis, les données sont agrégées et un indice de consensus est calculé. Ensuite, une étape de filtrage permet de supprimer des segments de mauvaise qualité selon plusieurs critères tels que leur degré de consensus. Finalement, une étape de correction manuelle permet de corriger les segments dont un label erroné a été attribué par la majorité des contributeurs.

## 5 Conclusion

Nous avons employé le crowdsourcing pour évaluer et améliorer un jeu de données de référence en télédétection. Nous avons procédé en deux phases : la première a servi à identifier et filtrer les segments de mauvaise qualité ainsi que les segments dont l'évaluation est particulièrement difficile ; la deuxième, s'est focalisée sur ces segments difficiles et a permis d'obtenir une évaluation plus pertinente de ceux-ci. Nous avons ainsi généré un jeu de données d'exemples dont la qualité est élevée (figure 5) et qui peut servir tant à l'apprentissage automatique de modèles de prédiction qu'à l'évaluation des résultats des méthodes d'analyse

Sur l'évaluation et l'amélioration d'une base d'objets géographiques



FIGURE 5: Extrait de la base de données produite.

d'images en télédétection. Un travail futur sera d'évaluer les bénéfices à utiliser ce jeu de données dans des applications réelles.

## Références

- Barrington, L., S. Ghosh, M. Greene, S. Har-Noy, J. Berger, S. Gill, A. Y.-M. Lin, et C. Huyck (2012). Crowdsourcing earthquake damage assessment using remote sensing imagery. *Annals of Geophysics* 54(6).
- Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS J Photogramm* 65, 2–16.
- Howe, J. (2006). The rise of crowdsourcing. *Wired magazine* 14(6), 1–4.
- Lessl, M., J. S. Bryans, D. Richards, et K. Asadullah (2011). Crowd sourcing in drug discovery. *Nature Reviews Drug Discovery* 10(4), 241–242.
- Puissant, A., S. Rougier, et A. Stumpf (2014). Object-oriented mapping of urban trees using random forest classifiers. *International Journal of Applied Earth Observation and Geoinformation* 26, 235 – 245.
- Sublime, J., A. Troya-Galvis, Y. Bennani, A. Cornuéjols, et P. Gañçarski (2015). Semantic rich icm algorithm for vhr satellite images segmentation. *14th IAPR International Conference on Machine Vision Applications (MVA)*.
- Whitla, P. (2009). Crowdsourcing and its application in marketing activities. *Contemporary Management Research* 5(1).

## Summary

In remote sensing image analysis, reference data play a crucial role but are often inaccurate and uncertain. In this article we present a methodology for improving remote sensing reference data in three steps: segment realignment, evaluation via crowdsourcing and creation of a good quality dataset.