

# 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/>