

## **Classification semi-automatique du terrain en zone rurale par télédétection à haute résolution**

**Roger Trias Sanz**

De nombreux travaux d'analyse d'image ont été et continuent d'être menés pour tenter de déterminer automatiquement l'occupation du sol en milieu rural, mais aucun n'a encore produit de résultats assez fiables pour être exploitables dans une chaîne de production industrielle.

L'Institut géographique national français (IGN) s'intéresse à la classification automatique du terrain pour accélérer la production de cartes topographiques à grande échelle. Le contexte à l'IGN est, cependant, différent de celui de la plupart des recherches dans la classification automatique: D'abord, on dispose d'images numériques à très haute résolution spatiale (50cm par pixel), mais ces images sont à faible résolution spectrale (canaux rouge, vert, bleu et, dans certains cas, proche-infrarouge), ce qui rend impossible l'utilisation de techniques classiques de classification de données hyperspectrales. De plus, on dispose des données cadastrales, qu'on peut utiliser pour obtenir une information grossière de la position des champs. Finalement, l'IGN n'est pas particulièrement intéressé à obtenir des classifications automatiques de qualité moyenne sur l'ensemble du territoire, classifications qui devraient être vérifiées, à grand coût en temps, par un photo-interprète. Par contre, l'IGN voudrait obtenir une classification de très haute qualité, même si c'est sur seulement une partie du territoire, car cette classification n'aurait pas à être vérifiée manuellement, et les photo-interprètes pourraient concentrer leur temps à classifier la partie restante.

Dans cette thèse, je présente une chaîne d'analyse d'image qui, à partir d'images numériques à haute résolution et à trois ou quatre canaux (50 cm, couleur et, dans certains cas, proche infrarouge), mais aussi en m'appuyant sur le parcellaire cadastral, rend une segmentation des images en parcelles agraires (champs, forêts, vignes, ...), et une classification de celles-ci, avec une très haute fiabilité, et attribue à chaque segment classifié une mesure qui indique la confiance que le système a en cette classification.

Une phase initiale de segmentation hiérarchique de l'image, qui utilise un espace de couleur, des paramètres de texture, et des critères de forme adaptés à la segmentation de parcelles agraires, permet de recalculer le cadastre sur l'image, produisant des régions grandes et en général homogènes. Ce recalage permet aussi d'utiliser le système pour la mise à jour de classifications anciennes. Ensuite, chacune de ces régions de cadastre recalées – ou, si les données cadastrales ne sont pas disponibles, des petites régions issues d'une segmentation par ligne de partage des eaux – est classifiée au moyen de nouveaux algorithmes probabilistes de classification par régions qui, à la différence des algorithmes classiques par pixels, ne produisent pas du bruit poivre-et-sel, et qui génèrent aussi une mesure de confiance pour chaque région classifiée. Les régions classifiées avec une trop faible confiance peuvent ensuite, selon les besoins de l'application, être rejetées et classifiées manuellement par des photo-interprètes. Ces algorithmes doivent être entraînés auparavant à partir d'une vérité terrain définie manuellement.

À la fin on obtient une segmentation de l'image en parcelles agraires homogènes, une classification de celles-ci, et des indicateurs de confiance sur chaque partie de la segmentation, ce qui permet à un photo-interprète de réaliser les corrections nécessaires et de concentrer son temps limité sur les zones qui plus vraisemblablement contiennent des erreurs.

## ***Semi-automatic rural land cover classification from high-resolution remote sensing images***

***Roger Trias Sanz***

*Much research has been done on the subject of automatically determining land use and land cover in rural areas from images. However, error rates have always been too high for industrial application.*

*The French National Mapping Agency (Institut géographique national, IGN-F), is interested in automatic land cover classification for the purpose of speeding up production of high-resolution topographic maps. The context at IGN is, however, slightly different from that at most other research into automatic land cover classification: First, very high spatial resolution digital images are available (50 cm per pixel), but these images have a low spectral resolution (red, green, blue, and in some cases near-infrared channels), which precludes the use of standard hyperspectral classification techniques. Second, cadastre data is available; this data gives a very rough indication of field position. Finally, IGN's interest stems from practical goals: to reduce the amount of time spent by human photo-interpreters doing manual classifications. Therefore, IGN is not especially interested in obtaining a medium-quality land cover classification of the whole region of interest but would rather have a very high quality classification of only a portion of the terrain; the first one would have to be double-checked in full by a photo-interpreter, whereas the second one would not need further human verification, so that photo-interpreters would be able to concentrate on the remaining areas.*

*In this Ph.D. thesis, I present a complete image analysis system which, from high-resolution 3 or 4-channel digital images (50 cm, colour and optionally near infrared), and using the cadastre database, segments the images into agriculturally-homogeneous regions (fields, forests, vines, and so on), and classifies these regions, labelling each classified region with a confidence measure which indicates the system's confidence in each classification, and which can be used to filter out regions that are more likely to have been incorrectly classified.*

*The process starts with a hierarchical segmentation, using a colour space, texture parameters, and shape criteria adapted to the problem of segmenting agricultural regions. This segmentation is used to register the cadastre onto the image, giving large, usually homogeneous regions. Through this registration, the system can also be used to update older classifications. Then, each of these registered cadastre regions – or, if cadastre data is not available, small regions obtained by watershed segmentation – is classified using novel probabilistic per-region classification algorithms which, unlike traditional per-pixel algorithms, do not produce salt-and-pepper noise, and which also output a classification confidence measure for each classified region. These classification algorithms are supervised, and need to be trained beforehand with a ground truth defined by a photo-interpreter.*

*As an end product we get an image segmentation into classified agriculturally-homogeneous regions, and confidence measures for each part of the segmentation, which a human photo-interpreter can use to correct these results or to concentrate limited available time into the most likely errors.*