UPDATING THE NEW FRENCH NATIONAL LAND COVER DATABASE

Adrien Gressin[1,2], Clément Mallet[1], Nicole Vincent[2], Nicolas Paparoditis[1]

[1] IGN/MATIS lab., Université Paris Est, France
[2] LIPADE - SIP, Université Paris-Descartes, France

ABSTRACT

Land-Cover databases (LC-DB) are very useful for environmental purposes, but need to be regularly updated to provide robust and instructive spatial indicators. Moreover, very high resolution satellite images allow to cover large areas regularly. Thus, automatic methods have to been developed to tackle this issue. In this paper, a hierarchical inspection method is proposed to both update and extend LC-DB using satellite image. This framework is successfully applied on the French National LC-DB using a single VHR satellite image.

Index Terms— Remote sensing, change detection, land cover, satellite imagery.

1. INTRODUCTION

1.1. Motivation

The French National Mapping Agency (IGN) is responsible for the generation of the large scale land-cover database (LC-DB) for the full French territory. It will be the official basis for setting up many environmental and agricultural public policies. This database is currently created by merging various existing topographic DB or LC-DB at different levels (both local and regional), introducing many flaws (limited geometric accuracy, weak completeness, heterogeneity). Moreover, end-users of such LC-DB require data as recent as possible (at least every year), which means it should be (1) exhaustive and corrected of geometric inaccuracies and (2) to be updated with automatic algorithms to be consistent with operational needs. For that purpose, Very High Resolution (VHR) optical satellite images are particularly suited, since they offer a suitable trade-off between spatial, spectral and temporal resolutions This study aims to provide a classification framework for both updating and enriching the newly established French National LC-DB using a single monoscopic satellite image.

1.2. Related work and contributions

Change detection is an important research topic in the remote sensing domain. Several approaches have been proposed using various input data : changes can be detected between diachronic images at pixel or object levels [1], or between one 2D database and one or several more recent image(s). Existing methods usually focus on one or a few numbers of specific classes (mainly buildings and roads). For instance, in [2], building databases are updated by merging SAR images and high resolution optical images. To cope with the issue of mono-class change detection, [3] propose to merge several methods designed for specific classes in a global semi-automatic workflow for change detection between an up-to-date image and a LC-DB. Nevertheless, it cannot be suited for a more general change detection problem as ours since on-purpose methods still need to be designed. Moreover, for efficient discrimination, most of those approaches are based on the computation of a high number of image features [4]. Therefore, feature selection becomes a key issue: e.g., [5] suggest a multiple-kernel method to both select the relevant image features and classify the data. Finally, a hierarchical inspection method has already been designed and applied on various simulated data [6]. In this paper, several improvements are proposed to increase the versatility and the scalability of the method, allowing us to apply it on a real LC-DB.

2. METHODOLOGY

2.1. A theoretical common workflow for updating and completing topographic databases

The hierarchical structure of geographical DB allows to focus on three inspection levels: (1) the object level, (2) the class level and (3) the DB level. The first inspection level is used

to take into account the various appearances of each object even of a single class. It is composed of two steps: (a) a selection of two subsets of pixels *inside* / *outside* an object in order to best discriminate it from the rest of the image and (b) a classification of the whole image into two classes (*inside* / *outside*). The subset selection is based on the maximization of the recall rate of a binary classification of the image pixels. This allows to retrieve for each object of each class of the DB, a confidence map that describes the probability of a pixel to belong to the same class as that of the object. The second inspection level allows to take into account the various appearances of the objects of one specific class. At this step, all the classifications of the object of one class are merged into a unique confidence map. This value is considered as the probability of each pixel to belong to the current class. Then, a labelling decision is taken at the DB level by integrating the probability maps of each class, each pixel is labelled by a label of the initial DB. Finally, a confidence measure derived from the belonging map of each class is associated with the classification map.

Therefore, the new classification and the confidence measure are used in two different ways depending on whether the initial DB has to be updated or completed. On the one hand, the new label of the pixel from the final decision is compared to the original one, to obtain a binary change / no-change label. This binary label is weighted with the confidence measure to finally have a change probability measure. On the other hand, for pixels without initial labelling, the labelling and the confidence measure previously obtained are used to complete the unlabelled areas of the initial DB. More details can be found in [6].

2.2. Adaptation to large scale databases

Such adaptation is necessary in order to tackle the scalability, the versatility and the computing time issues. Since an increasing number of various classes are introduced in the LC DB (*cf.* Sect. 3.1), and since, the most relevant features for a specific class cannot be known beforehand, a large number of features computed from the different channels of the image is, therefore, required. Those features highly influence the performance of the workflow. In this experiment, the number of features is increased. Nearly two hundred features are derived from the original image bands, for instance by introducing

morphological attribute profiles [4]. Therefore, a selection step has been carried out, prior to the hierarchical inspection. For each class of the DB, the best features are selected using the feature importance computed thanks to the Random Forest method [7]. The best number of features, in order to find a good trade-off between computational time and classification accuracy, have to be determined. In our previous work, the classification was based on the Support Vector Machine (SVM) method, chosen for its high generalisation ability and its capacity to deal with a large number of features. In this paper, Random Forest is introduced in order to speed up the classification. Finally, the numbers of objects per class to be inspected become too large to consider inspecting them all. Thus, this step is tailored by selecting a limited number of objects. The n largest objects of each class have been chosen.

3. RESULTS AND DISCUSSIONS

3.1. Datasets : Satellite images and Land-Cover data

On the one hand, our dataset is composed of one very high resolution Pleiades image, acquired in August 2012 (Fig. 1a). The image covers a surface of 760 km^2 in the Southern-West part of France, with a resolution of 0.5 m in panchromatic mode and 2 m in color (red, green, blue, and near infrared). On the other hand, the French National LC DB have a hierarchical nomenclature, consisting of a four levels, has been developed (Fig. 2a). It is both compatible with the Corine Land Cover and the regional LC-DB nomenclatures. The lowest level is composed of fifteen classes (e.g., built-up areas, impermeable unbuilt areas, deciduous forest, coniferous forest or herbaceous areas), but only ten classes are represented in our study area (Sect. 3.3). However, the proposed method has been first applied on a simplified DB in order to assess its performance (Sect. 3.2). This DB is composed of two classes, namely *Field* and *Forest* on a periurban subarea of the Pleiades image. The simplified LC DB is shown on Fig. 1b. This DB has a few numbers of classes. Nevertheless, they are composed of objects of large spatial extent and of a wide range of appearances, which remains challenging.

3.2. Per-class intermediate results

The previous workflow has been applied on the simplified DB and results are shown in Fig. 1. These results are based on

(a)   (b)

(c)   (d)

(e)   (f)
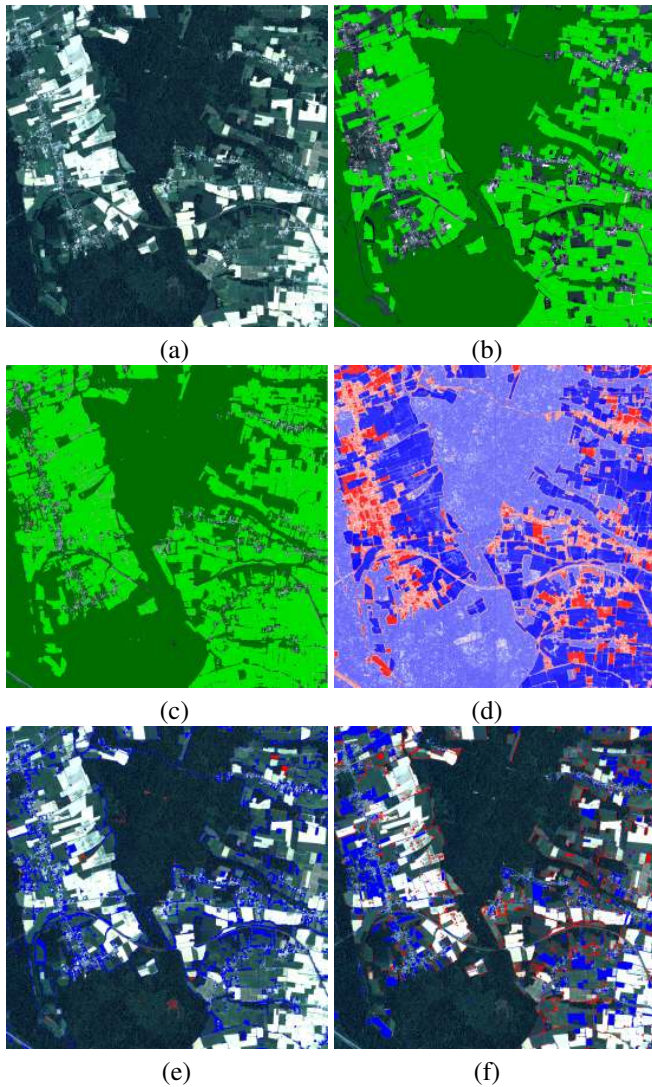
Fig. 1. Preliminary results: (a) Image, (b) the LC-DB composed of two classes (fields in light green and forest in dark green) and (c) the resulting classification (using the same color), (d) the probability change map (no-changes in blue, changes in red and confused areas in white), and the difference (before / after) for (e) the *Forest* class and (f) the *Field* class (appearance in blue and disappearance in red).

about 30 features (12 spectral features, and about 20 textural features) without selection step. It can be noticed that unlabelled areas of the DB have now a class with the method (Fig. 1c). On the probability change map (Fig. 1d) extended red areas correspond to real changes, here new fields and forest growth or decrease. White areas match with confusing areas, mainly objects that are not in the initial DB, such as building and roads (these issues may be easily solved by introducing more classes in the DB). Moreover, a class-by-class

(a)

(b)

(c)
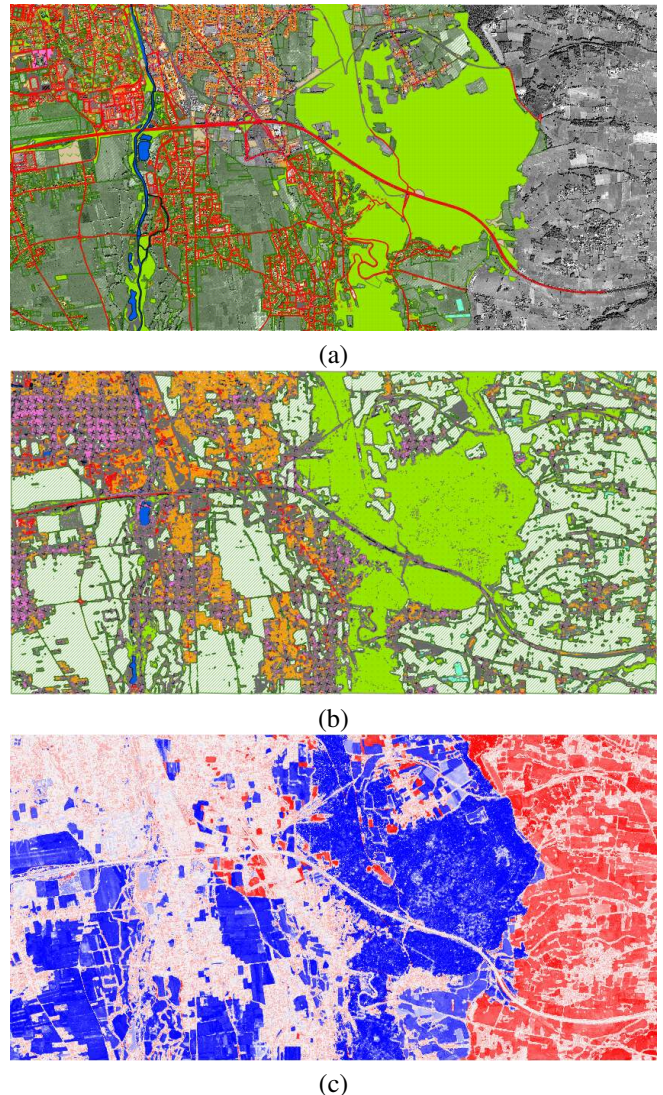
░░	Building area
░░	Road and parking area
▓▓	Mineral materials area
██	Water surfaces
░░	Deciduous forest
×××	Mixed forest
▨▨	Other forest
⌄⌄⌄	Shrubby formations
≈≈≈	Other woody formations
░░	Herbaceous formations

Map legend for (a) and (b)

Fig. 2. Results on real land-cover databases: (a) Land-cover DB and image, (b) the classification and (c) the probability change map (no-changes in blue, changes in red and confused areas in white), the red area on the right correspond to an unlabelled area of the initial DB and the white one, to urban area where confidence measure is low due to confusion in the classification.

study can be performed. Indeed, Fig. 1e shows the difference between the initial *Forest* class and the results of the classification. Appearance is displayed in blue and disappearance in red. Same result is shown on Fig. 1f for the *Fields* class. Those two figures show that a fine change detection can be carried out. For instance, hedges and copses, not in the *Forest* class of the initial DB, are correctly identified in the new image. For the *Field* class, many borders are labelled as disappeared, which is due to the coarse delineation of those objects during the creation of DB by photo-interpretation.

3.3. Results on real land-cover database

Finally, the proposed workflow has been applied on the French National LC-DB (Fig. 2). The number of features has been increased to reach approximately 150. This feature set is composed of about 20 spectral features, 100 textural features and 30 morphological features. Then, the 20 best features were selected for each class, using the Random Forest feature importance, and only the 10 largest objects of each class are inspected. Those values have been arbitrary fixed for now. The classification covered the whole area, including areas not labelled by the initial DB (Fig. 2b). The overall accuracy of the classification is about 70%. The classification is visually of good quality, specially for deciduous forest, herbaceous formations, building and roads classes. For the two first classes, the exactness is greater than 95% and the recall is about 75%, and for the buildings and roads classes the exactness is about 60% and the recall is about 45%. However, other classes give lower results both in term of exactness an of recall. Those classes are generally badly defined in the DB nomenclature (as mixed forest) and/or represented by a low number of objects, that are frequently of small size. Thus, those classes introduced confusion in the classification explaining the lowest result of the buildings and roads classes. The resulting change map is shown on Fig. 2c. One can notice that the large red area on the east side correspond to the unlabelled area, and that the white areas corresponds to low confidence areas. The latter ones are due to confusion in the classification seen previously, and correspond generally to urban areas. However, the change map allows to focus on real change areas.

4. CONCLUSION AND PERSPECTIVE

The work presented in this paper is based on a simple and robust change detection process between a land-cover database and a newer image [6]. The method has been first successfully applied to update a simplified land cover database using a single very high resolution satellite image. Then, the same theoretical method has been adapted in order to be applied on a general large scale land-cover database. For this purpose, several improvements have been introduced in terms of versatility and scalability. Finally, the adapted method is successfully applied on a full land cover database composed of 10 classes on an extended area covered by a single very high resolution satellite image. Our future works will focus on (1) scalability improvement by tuning the inspection method with spatial rules, and (2) versatility improvement by introducing information either from multi-sensor or from multi-temporal data.

5. REFERENCES

[1] A. Lefebvre, T. Corpetti, and L. Hubert-Moy, "Object-oriented approach and texture analysis for change detection in very high resolution images," in *IGARSS*, 2008, pp. 663–666.

[2] V. Poulain, J. Inglada, M. Spigai, J-Y Tourneret, and P. Marthon, "Fusion of high resolution optical and SAR images with vector data bases for change detection," in *IGARSS*, 2009, pp. 956–959.

[3] P. Helmholz, C. Becker, U. Breitkopf, T. Buschenfeld, A. Busch, C. Braun, D. Grunreich, S. Muller, J. Ostermann, M. Pahl, F. Rottensteiner, K. Vogt, M. Ziems, C. Heipke, and Others, "Semi-automatic Quality Control of Topographic Data Sets," *PERS*, vol. 78, no. 9, pp. 959–972, 2012.

[4] M. Dalla Mura, J.A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological Attribute Profiles for the Analysis of Very High Resolution Images," *IEEE TGRS*, vol. 48, no. 10, pp. 3747–3762, 2010.

[5] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning Relevant Image Features With Multiple-Kernel Classification," *IEEE TGRS*, vol. 48, no. 10, pp. 3780–3791, 2010.

[6] A. Gressin, N. Vincent, C. Mallet, and N. Paparoditis, "Semantic approach in image change detection," in *ACIVS*, Poznan, Poland, 2013.

[7] L. Breiman, "Random forests," *Machine learning*, pp. 1–35, 2001.