Multiresolution Remote Sensing Image Clustering

Cédric Wemmert, Anne Puissant, Germain Forestier, Pierre Gançarski

***Abstract*—With the multiplication of satellite images with complementary spatial and spectral resolution, a major issue in the classification process is the simultaneous use of several images. In this context, the objective of this letter is to propose a new method which uses information contained in both spatial resolution. The main idea is that on one hand, the semantics level associated with an image depends on its spatial resolution, and on the other hand, information given by these images are complementary. The goal of this multiresolution image method is to automatically build a classification using knowledge extracted from both images, by unsupervised way and without pre-processing image fusion. The method is tested by using a Quickbird (2.8m) and a SPOT-4 (20m)images on the urban area of Strasbourg (France). The experiments have shown that the results are better than a classical unsupervised classification on each image, and comparable to a supervised region-based classification on the HSR image.**

***Index Terms*—multiresolution, clustering, remote sensing image analysis, high spatial resolution.**

I. Introduction

With the recent development of remote sensing technology, a large amount of satellite images is available with a complementary spatial, spectral and time resolution. These data are valuable for mapping urban areas at different scales. There is a wide range of object nomenclatures such as the Corine Land Cover nomenclature defined for Landsat images (30m spatial resolution), the SPOT Thema nomenclature defined for Spot images (5m to 20m) or the French national land cover database BDCarto®IGN (defined for aerial photographs and SPOT images). These existing products enable to map urban areas respectively from 1:100,000 (Corine Land Cover nomenclature) to 1:50,000 and 1:25,000 (SPOT Thema and BDCarto nomenclatures). The production of these maps is often based on manual image interpretation or semi-automatic techniques (combined with per-pixel algorithm). With High Spatial Resolution (HSR - 1 to 5m) satellite images, it is possible to extract urban objects (e.g. house, garden and road) using object-oriented approach based on a segmentation step [1]. This allows to map individual objects at scale from 1:10,000 to 1:5,000.

In the domain of urban planning and management, some users also need to map the territory at the scale of the urban blocks (which can be defined as a minimal cycle closed by communication way) corresponding to a scale near of 1:10,000. In this case, there is no existing available land cover/use product. Medium Spatial Resolution (MSR from 30 to 10m spatial resolution) satellite images have a too large spatial resolution and HSR images have a too fine spatial resolution to map urban blocks. In this context, a question is how to reconstruct urban blocks from individual objects and how to help end-user to extract automatically these blocks ?

This problem can be treated as a classification problem where images with different spatial resolutions can be used simultaneously in an unsupervised way. To address this problem, the question of the number of clusters on each image is not straightforward. In MSR images, urban areas can be classified into 5 to 7 clusters associated to land cover classes referring to the identification of 'urban zone' (Table I, left column). In HSR images, the number of clusters is higher (10 to 15 classes) referring to materials of each urban objects (Table I, right column). For instance, buildings can be differentiated by the materials and the roofs color (Table I, right column). To be able to offer to end-users a 1:10,000 mapping of urban areas, the number of semantic classes must range between 7 and 9 classes (for instance, Table I, middle column). However, these semantic classes cannot be directly obtained by a classification process from unique MSR or HSR images.

In this context, the objective of this paper is to propose a new method which simultaneously uses the information contained in both MSR and HSR images. This multiresolution image analysis takes into account the different numbers of clusters obtained on each image at very different spatial resolutions.

The sequel of this article is structured in three sections. The multiresolution remote sensing images analysis is briefly introduced and the new multiresolution clustering process is detailed in Section II. Some experiments performed on a HSR and MSR multispectral images on the urban area of Strasbourg (France) are presented. These images are issued from (1) different sensors (Spot and Quickbird) and (2) a degradation process from the Quickbird image (Section III). Some conclusions are then drawn in Section IV.

II. Multiresolution remote sensing images analysis

In [2], we proposed a framework of collaborative clustering, called Samarah which enables several clustering methods (referred here as classifiers) to collaborate, in order to produce an unique solution from a set of images. These images can have different resolutions. Each classifier deals with an image from this set. All the classifiers work together to end up at an agreement on their clustering. Each classifier modifies its results according to all the other clusterings until all the clusterings proposed by the classifiers are strongly similar. Thus, they can be more easily unified, for example, through a voting algorithm. This unified result represents a consensus among all the knowledge extracted from the different sources. Furthermore, the voting algorithm highlights the agreement

The authors are with the University of Strasbourg, 67081 Strasbourg Cedex, France (e-mail: wemmert@unistra.fr; anne.puissant@unistra.fr; forestier@unistra.fr; gancarski@unistra.fr)

TABLE I
TYPOLOGIES USED TO MAP URBAN AREA ON MSR AND HSR IMAGES AND TYPOLOGY USED BY END-USERS AT 1:10,000

1:100,000-1:25,000 MSR imagery (30 to 5 m) 'area level'	**1:10,000** Semantic class based on urban blocks 'block level'	**1:5,000** HSR imagery (3 to 1 m) 'object level'
• High-density urban fabric • Low-density urban fabric • Industrial areas • Forest zones • Agricultural zones • Water surfaces • Bare soil	• Continuous urban blocks • Discontinuous urban blocks - Individual urban blocks - Collective urban blocks • Industrial urban blocks • Urban vegetation • Forest • Agricultural zones • Water surfaces • Road	• Building/roofs: - red tile roof - light gray residential roof - light commercial roof • Vegetation: - green vegetation - non-photosynthetic vegetation • Transportation areas: - streets - parking lots • Water surfaces: - river - natural water bodies • Bare soil • Shadow

and the disagreement between the clustering methods. These two informations, as well as the result produced by each clustering method, lead to a better understanding of the scene by the expert.

This method was experimented on two study cases where all the classifiers generated the same number of clusters in order to produce a unique result. The first experiment[1] was the analysis of SPOT-5 images (panchromatic - 5m and multispectral - 10m with four spectral bands - red, green, blue and near infra-red) of the urban area of Strasbourg (France). The second experiment[2] was performed on four images of a coastal zone (Normandy Coast, Northwest of France). This area is especially interesting since it is periodically affected by natural and anthropic phenomena. Four images issued from three different satellites and with different spatial and spectral resolutions were used (SPOT-4 multispectral at 20m and panchromatic at 10m, SPOT-5 panchromatic at 2.5m and ASTER multispectral at 15m).

Results of these experiments were better than those produced from panchromatic band only, multispectral bands only or pan-sharpened images. However, the geographic objects to identify depends on the spatial resolutions, especially in the context of urban images classification. So another method is presented here that enables to use simultaneously two images with very different spatial resolutions, and for which each classifier does not necessarily search the same number of clusters.

A. Multi-source clustering

A first way to classify multi-source objects is to combine all the descriptions of the objects associated to the different sources. Each object has a new description composed of (all)[3] the attributes (also called features or characteristics) of all the sources [3, 4]. For instance, all the radiometric information given by the different sensors are used to describe the objects. Unfortunately, due to the *curse of dimensionality* [5], most of the classical distance-based algorithms are not sufficient to analyze objects having many attributes: the distances between these objects are not enough different to correctly determine the nearest objects. In addition, with the increase of the spectral dimensionality, some problems appear like the Hughes phenomena [6] which shows that classifier performance actually decreases with increasing data dimensionality objects.

A second way [7]–[9] to classify multi-source objects is to independently classify each data sets. Then a new description of each object is built, using the labels of the cluster to which the object belongs in each classification. Finally, a new clustering is done using the new description of the objects. The aim of the first clustering is to reduce the data space for the final clustering, making it more effective. This approach is similar to the *stacking* method [10] which is known as outperforming the methods based on the direct combination of attributes. Thus, the method we propose, is based on this second approach that we have adapted to the analysis of remote sensing images.

B. Multi-source remote sensing image clustering

1) Notations: An image $\mathcal{I}$ can be viewed as a function

$$\begin{array}{rl} \mathcal{I}: & E \subset \mathbb{Z}^2 \rightarrow \mathbb{Z}^b \\ & p \rightarrow \mathcal{I}(p) \end{array} \quad (1)$$

where $\mathcal{I}(p) = \langle I_1(p), \ldots, I_a(p) \ldots, I_b(p) \rangle$ with $b \in \mathbb{N}^*$ the number of spectral bands of the image and $I_a(p)$ the radiometric information associated to the pixel on the a-th band.

A clustering image $\mathcal{C}$ from a clustering of the image $\mathcal{I}$ is defined as

$$\begin{array}{rl} \mathcal{C}: & E \subset \mathbb{Z}^2 \rightarrow [1, K], K \in \mathbb{Z} \\ & p \rightarrow \mathcal{C}(p) \end{array} \quad (2)$$

where $\mathcal{C}(p)$ is the cluster label associated to the pixel p by the clustering algorithm (e.g., the number of the cluster to which the pixel belongs) and K the number of clusters.

From such an image, a connected component O_i (also called region or object) is defined as

$$O_i = \{p, q \in \mathcal{I} : \mathcal{C}(p) = \mathcal{C}(q) \wedge \textit{connected}(p, q) = 1\} \quad (3)$$

where *connected* is the classical 8-connectivity function in $\mathcal{I}$ (let $p(x, y)$ and $q(x', y')$ two pixels in the image,

[1]FoDoMuSt project: `http://fodomust.u-strasbg.fr`
[2]ECOSGIL project (JC05-50539): `http://ecosgil.u-strasbg.fr`
[3]Depending on the redundancy of these attributes

connected$(p,q) = 1$ if $max(|x-x'|, |y-y'|) = 1$). Note that the number of regions depends on the clustering and cannot be a priori defined. Let N_r be the number of regions.

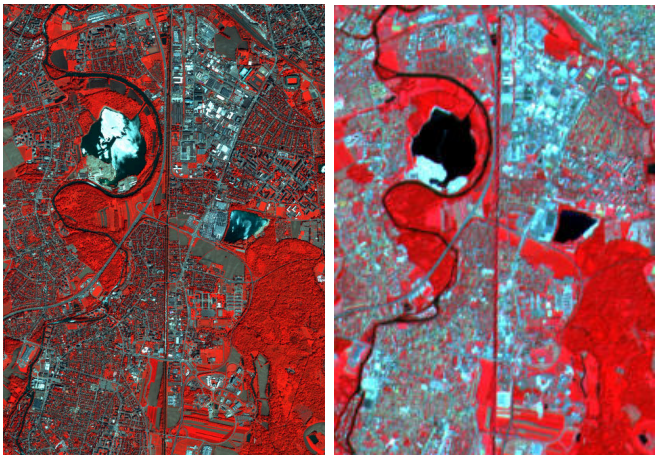
The region image $\mathcal{R}$ built from a clustering result image $\mathcal{C}$ is defined as

$$\mathcal{R}(\mathcal{C}) = \{O_n, \forall n \in [1, N_r]\} \quad (4)$$

In the following, we only focus on the case where we have two images $\mathcal{I}^n$ and $\mathcal{I}^{n'}$ having the resolutions r^n and $r^{n'}$ with $r_n \geqslant r_{n'}$.
Let $\lambda_{n,n'}$ be a correspondence function, associating one pixel from $\mathcal{I}^n$ to its corresponding pixel in $\mathcal{I}^{n'}$. $\lambda_{n,n'}$ can be easily defined using the georeferencing of the two images.
Let $\mathcal{C}^n$ (resp. $\mathcal{C}^{n'}$) be the clustering image associated to $\mathcal{I}^n$ (resp. $\mathcal{I}^{n'}$), k_n (resp. $k_{n'}$) be the number of clusters in $\mathcal{C}^n$ (resp. $\mathcal{C}^{n'}$) and $\mathcal{R}^n$ (resp. $\mathcal{R}^{n'}$) be the region image associated to $\mathcal{C}^n$ (resp. $\mathcal{C}^{n'}$).

2) The method: Our method is object-oriented. It consists in four sequential steps. First, a per-pixel clustering is done on both the images. Then, for each image, regions are built and characterized using the per-pixel clusterings. Then, these characterized regions are clustered.
The four steps of our approach are:

- Step 1 - *Initial clusterings*: both images are independently classified to obtain clustering images (Eq. 2).
- Step 2 - *Regions building*: both corresponding region images are built (Eq. 4).
- Step 3 - *Regions characterization*: each region $O_i^n \in \mathcal{R}^n$ from an image is characterized according to its clusters composition in the clustering image $\mathcal{C}^{n'}$ (resp. $O_i^{n'}$, $\mathcal{R}^{n'}$ and $\mathcal{C}^n$): for each region $O_i^n \in \mathcal{R}^n$, we calculate a composition histogram representing the distribution of the labels associated to the pixels of $\mathcal{I}^{n'}$ corresponding to the pixels of O_i^n using the correspondence function $\lambda_{n,n'}$.
  The composition histogram $H_{n'}(O_i^n)$ associated to a region O_i^n according to clustering image $\mathcal{C}^{n'}$ is defined by:
  $$H_{n'}(O_i^n) = \langle h_{i,1}^{n'}, \ldots, h_{i,k_{n'}}^{n'} \rangle, O_i^n \in \mathcal{R}^n \quad (5)$$
  where $k_{n'}$ is the number of clusters in $\mathcal{C}^{n'}$ and
  $$h_{i,j}^{n'} = \mid \{q = \lambda_{n,n'}(p) : \mathcal{C}^{n'}(q) = j, \ \forall p \in O_i^n\} \mid \quad (6)$$
  .
- Step 4 - *Object-oriented classification*: for each region image $\mathcal{R}^n$ (resp. $\mathcal{R}^{n'}$), a clustering algorithm is independently applied on all its objects (i.e. characterized regions) using the composition histograms. Let $C_{n'}(O_i^n)$ the class label associated to the object O_i^n characterized using a clustering image $\mathcal{C}^{n'}$ (resp. $C_n(O_i^{n'})$ and $\mathcal{C}^n$) Then, the final clustering image $\mathcal{F}_{n'}(n)$ (resp. $\mathcal{F}_n(n')$) is defined as
  $$\begin{array}{rl} \mathcal{F}: & E \subset \mathbb{Z}^2 \rightarrow [1, K_n] \\ & p \rightarrow C_{n'}(O_i^n), p \in O_i^n \end{array} \quad (7)$$
  where K_n is the number of clusters expected in the final clustering of the image $\mathcal{I}^n$ (resp. $K_{n'}$ and $\mathcal{I}^{n'}$).



(a) HSR - multispectral Quickbird image at 2.8 m (b) MSR - multispectral SPOT-4 image at 20m

Fig. 1. Extract of the urban area of Strasbourg (France)

Note that K_n can be different of k_n because the first classification is pixel-oriented while the second one is object-oriented. Commonly, $k_n > K_n$.

III. Results

A. Experiments

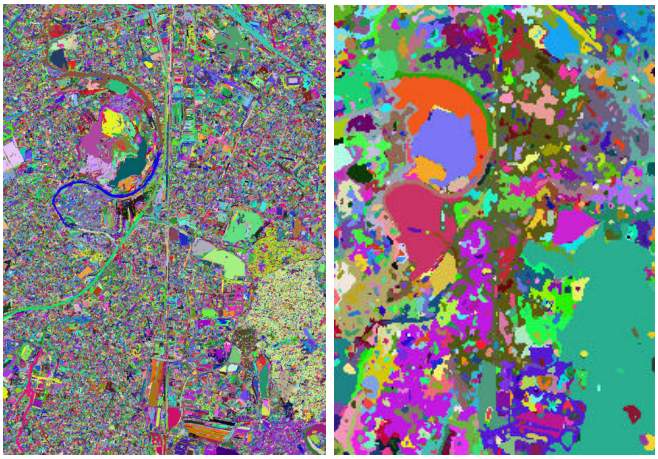
Some experiments are performed on two multispectral images with different spatial resolutions (2.8m and 20 m) on the urban area of Strasbourg (France), acquired: (1) from two different sensors (Quickbird[4] and Spot-4[5], respectively in May and July 2001) and (2) from a Quickbird image and a resampled image at 20m spatial resolution. The Quickbird multispectral image is available in four spectral bands (blue, green, red and near-infrared bands). The multispectral SPOT-4 image has three spectral bands (green, red, near-infrared).

These images (Fig.1(a) and Fig.1(b)) present an extract of the urban area of Strasbourg (France) which is a typical suburban area with some water surfaces (in the center), forest area in the South, industrial areas, agricultural zones with different spectral responses due to the seasons (bare soil in the HSR image - May can appear in red on the MSR image - July) and some individual or collective housing blocks (in red, black and white textured on the MSR image, in red, blue and white textured in the HSR image).

The experiments have consisted to perform the four steps described in Section II-B2.

- Step 1 - Initial classifications: in all experiments, each image is classified using the classical K-means algorithm [11] with a number of classes depending on the spatial resolution. Note that any algorithm which can deal with numeric data would be used. For MSR image, a lot of previous works shown that urban areas can be classified in 6 classes. For HSR image, the number of clusters depends on the materials of the urban objects (cf. above). In order

[4]©DigitalGlobe Inc.
[5]©CNES (Isis program)



(a) HSR image (b) MSR image

The colors affected to the regions have no semantic and have been randomly choosen

Fig. 2. Regions built from the initial clusterings (Step 2)

to find the best number of cluster according to the study zone, three experiments with respectively 10, 15 and 20 classes are carried out.
These experiments have shown that:

- with 10 classes, the regions are too large and there are not enough regions to classify;
- with 20 classes, the regions are too small and they are too close of the pixels (each region contains only 3 to 6 pixels).

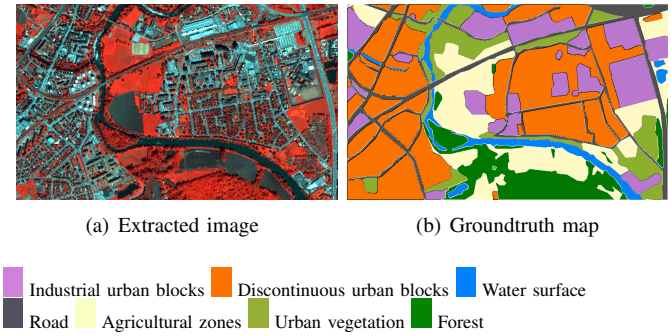
The best result is obtained with 15 classes.

- Step 2 - Regions building: from both classified images, the regions are built, integrating into a same region the connected pixels having the same class label. Fig. 2 shows the region maps.
- Step 3 - Regions characterization: All the composition histograms have been computed.
- Step 4 - Objects classification: After having computed another K-means algorithm on the region images, we obtain the final clusterings in the both spatial resolutions.

The first three steps are applied once and the fourth step is tested for 7, 8 and 9 clusters in order to find the best result, by taking into account the number of expected land use classes, based on the 'block level'. Results are presented here on an extract of the studied zone (North West part of the Fig. 1) with 7, 8 and 9 clusters (Fig.4).

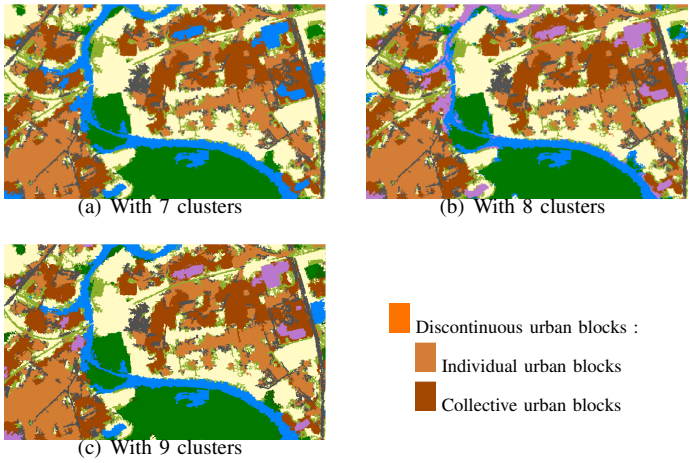
B. Results assessment and discussion

Results are assessed by a comparison with a groundtruth map from a landcover/use database (BDOCS 2000 CIGAL 2003) used for a 1:10,000 mapping. This groundtruth map contains 8 thematic classes at the urban blocks semantic level (see Tab. I). Only 7 thematic classes are present on the extract shown on Fig. 3(b). Fig. 4 shows respectively the results with 7, 8 and 9 clusters on this extract. The 7 classes found on the first image (Fig. 4(a)) do not exactly match those from the groundtruth map. Indeed, the industrial blocks are in the



(a) Extracted image (b) Groundtruth map

Fig. 3. Subset of the studied zone with the groundtruth map (BDOCS 2000 CIGAL 2003)



(a) With 7 clusters (b) With 8 clusters

(c) With 9 clusters

Fig. 4. Results of the proposed multiresolution method

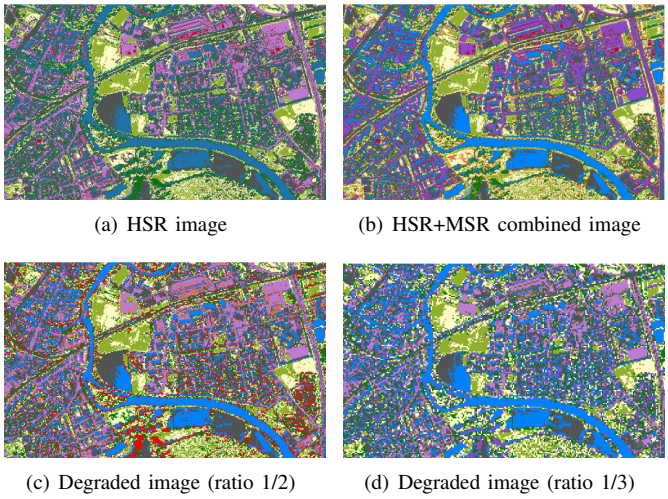
same cluster as the water surfaces, and there are 2 clusters representing the discontinuous urban blocks (individual in orange and collective in red). On the 8 clusters image (Fig. 4(b)), the industrial blocks appear in the 8th cluster (in purple). Finally, a new class of vegetation is discovered in the 9 clusters result (Fig. 4(c)).

We choose to calculate the Kappa index to evaluate quantitatively the quality of these first results by comparison with the groundtruth (Table II). The Kappa is a measure of classification accuracy which can be used as an indicator of the agreement between two classifications. It evaluates the percentage of correct values which are due to "true" agreements versus "chance" agreement. It is defined as:

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \tag{8}$$

where Pr(a) is the relative observed agreement and Pr(e) is the hypothetical probability of chance agreement. A Kappa value of 1 indicates a perfect agreement. The value of the Kappa decreases as the classification are in disagreement. A value between 1.00 et 0.81 reflects a perfect agreement, a value between 0.80 and 0.60 indicates a good agreement and so on [12].

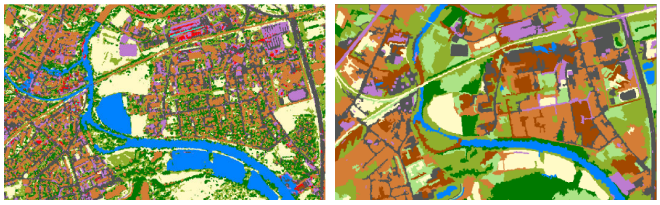
For comparison purpose, we also evaluated our results by comparing them with clusterings carried out with different



(a) HSR image (b) HSR+MSR combined image

(c) Degraded image (ratio 1/2) (d) Degraded image (ratio 1/3)

Fig. 5. Per-pixel clustering using K-means with 9 clusters



(a) K-means (b) Supervised object-oriented method

Fig. 6. Regions classification on HSR image, with 9 clusters

approaches:

- K-means clustering with 7, 8 and 9 clusters at a pixel level (Fig. 5(a));
- K-means clustering on an image built by direct fusion of the two images HSR+MSR (to each pixel is associated all the radiometric information from HSR and MSR images) (Fig. 5(b));
- K-means clustering on two degraded images of the HSR image with a resolution ratio of 1/2 and of 1/3 (Fig. 5(c) and 5(d));
- K-means clustering with 7, 8 and 9 clusters at an object level using only spectral information (the objects were created from a clustering with 15 clusters) (Fig. 6(a));
- Supervised object-oriented approach with Definiens Professional software, using only spectral information [6] (Fig. 6(b)).

Fig. 5 illustrates results obtained by the per-pixel classification with 9 clusters expected. Fig. 6 shows results obtained by the object-oriented classifications with 9 clusters expected.

The Kappa index was computed using the groundtruth map. It was computed for all the different results (II). The analysis of this global quality indicator shows that the results of the multiresolution proposed method are comparable to the supervised object-oriented method, and outperforms the other ones.

[6] http://www.definiens.com/

TABLE II
KAPPA VALUES OF THE DIFFERENT EXPERIMENTS

	7	8	9
Multiresolution proposed method	**0.73828**	**0.74259**	**0.74501**
K-means on HSR image	0.70805	0.70412	0.71832
K-means on HSR+MSR combined image	0.68772	0.70412	0.71831
K-means on degraded image (1/2)	0.68729	0.70396	0.71645
K-means on degraded image (1/3)	0.68569	0.70536	0.71620
K-means region classification	0.67864	0.68051	0.69002
Supervised object-oriented method	0.70604	0.72811	0.73843

IV. CONCLUSION

One of the challenging issues in remote sensing image information mining is the multiple uses of the acquired image data. Indeed, satellite images are now easier to acquire and consequently, a large amount of heterogeneous images is now available. If a classification on a MSR image gives not enough information, a classification on a HSR brings too much heterogeneous information (salt-pepper effect). To obtain aggregate information, users have to be apply a post-classification process (class merging). In order to help users to obtain directly this aggregate information for a land cover mapping at 1:10,000 for instance, we have presented here a new method which automatically combine information from two satellite images with very different spatial resolutions. This method offers the ability to discover new knowledge from these two images. These first experiments on the urban area of Strasbourg have shown very interesting results. In the future we wish to integrate the method into the SAMARAH framework of collaborative clustering. We also planed to extend the method in order to make it able to simultaneously deal with more than two images.

ACKNOWLEDGMENT

This work is supported by the french Centre National d'Etudes Spatiales (CNES Contract 70904/00).

REFERENCES

[1] U. C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen, "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for gis-ready information," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, no. 3-4, pp. 239 – 258, 2004, integration of Geodata and Imagery for Automated Refinement and Update of Spatial Databases.

[2] G. Forestier, C. Wemmert, and P. Gançarski, "Multi-source images analysis using collaborative clustering," *EURASIP Journal on Advances in Signal Processing - Special issue on Machine Learning in Image Processing*, vol. 2008, p. 11, 2008.

[3] Y. Chibani, "Selective synthetic aperture radar and panchromatic image fusion by using the a trous wavelet decomposition," *EURASIP Journal on Applied Signal Processing*, no. 14, pp. 2207–2214, 2005.

[4] Y.-L. Chang, L.-S. Liang, C.-C. Han, J.-P. Fang, W.-Y. Liang, and K.-S. Chen, "Multisource data fusion for landslide classification using generalized positive boolean functions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, 2007.

[5] R. Bellman, *Adaptive Control Processes*, N. Princeton University Press, Princeton, Ed., 1961.

[6] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Informations Theory*, vol. 14, no. 1, pp. 55–63, 1968.

[7] J. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1367–1377, 1999.

[8] L. Bruzzone, R. Cossu, and G. Vernazza, "Combining parametric and non-parametric algorithms for a partially unsupervised classification of multitemporal remote-sensing images," *Image Fusion*, vol. 3, pp. 289–297, 2002.

[9] M. Fauvel, J. Chanussot, and J. Benediktsson, "Decision fusion for the classification of urban remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2828–2838, 2006.

[10] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, July 2004.

[11] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, L. M. L. Cam and J. Neyman, Eds., vol. 1. University of California Press, 1967, pp. 281–297.

[12] R. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sensing of Environnnment*, vol. 37, pp. 35–46, 1991.