Multi-resolution region-based clustering for urban analysis

Camille Kurtz, Nicolas Passat, Pierre Gançarski, Anne Puissant

▶ **To cite this version:**

Camille Kurtz, Nicolas Passat, Pierre Gançarski, Anne Puissant. Multi-resolution region-based clustering for urban analysis. International Journal of Remote Sensing, 2010, 31 (22), pp.5941-5973. 10.1080/01431161.2010.512312 . hal-01694411

HAL Id: hal-01694411
https://hal.univ-reims.fr/hal-01694411v1

Submitted on 3 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vol. , No. , , 1–32

Research Article

Multiresolution region-based clustering for urban analysis

C. KURTZ*†, N. PASSAT†, P. GANÇARSKI† and A. PUISSANT‡

† Université de Strasbourg, LSIIT - UMR 7005 CNRS, Strasbourg, France
‡ Université de Strasbourg, LIVE - ERL 7230 CNRS, Strasbourg, France

(*September 2010*)

In the domain of urban planning and management, it may be necessary to map the territory at different scales corresponding each to a semantic level. Three semantic levels are identified: (1) the *object level*, for mapping urban elements (buildings, etc.), (2) the *block level*, for mapping homogeneous patterns of urban elements, and (3) the *area level*, for mapping urban fabrics defined as sets of homogeneous patterns. Some of these levels are directly linked to specific satellite images presenting *ad hoc* resolutions (namely Medium Spatial Resolution –MSR– images for the area level, and High Spatial Resolution –HSR– images for the object level); in such cases, a straightforward mapping can be performed by clustering the data. By opposition, classical clustering techniques do not enable to extract directly the intermediate semantic level. The purpose of this article is to propose a methodology enabling to generate a clustering at this level. The proposed approach is, in particular, based on the segmentation and unsupervised, region-based and joined clustering of two images representing a same scene at MSR and HSR. The method has been applied to different and heterogeneous datasets composed of HSR images at 2.5 m and MSR images at 10 m and 20 m. Qualitative validations by an expert, and quantitative ones by comparison to other existing methods, tend to emphasise the soundness and efficiency of this methodology, thus justifying further developments.

Keywords: Clustering, multiresolution analysis, object-based, urban mapping.

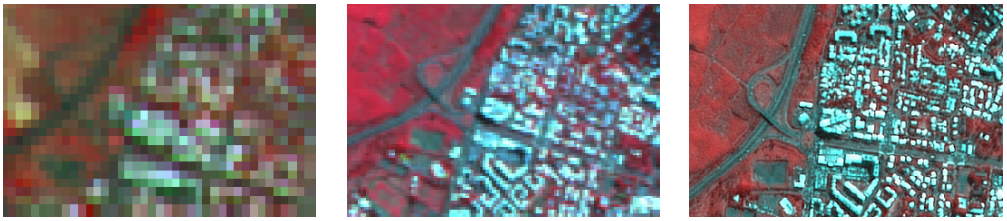
1. Introduction

1.1 *Context*

In the domain of urban planning and management, it may be necessary to map the territory at different scales corresponding each to a semantic level. To map urban areas from 1:100,000 to 1:25,000 –enabling, for instance, to specify the density of an urban fabric, etc. (Table 1, left column)– images at a medium spatial resolution (MSR - 30 to 5 meters, Fig. 1(a, b)) are available. For the mapping of areas at a scale of 1:5,000 enabling to deal with urban objects –*e.g.*, individual houses, gardens, roads, etc. (Table 1, right column)– with their material (*e.g.*, houses with orange tile roof), images at high spatial resolution (HSR - 3 to 1 meter(s), Fig. 1(c)) have been proposed since the end of the 90's.

For the intermediate scale of 1:10,000 enabling to analyse urban blocks, which can be defined by the minimal cycles closed by communication ways (Fig. 2), there does not exist any land cover/land use product. For this analysis, corresponding to a semantic level called *block level* (Table 1, centre column), MSR images have a too coarse spatial resolution (Fig. 2(a)) while HSR ones have a too fine spatial resolution (Fig. 2(b)). Consequently, the classes induced by these urban blocks cannot be obtained by a straightforward classification process from either or both

*Corresponding author. Email: ckurtz@unistra.fr



(a) 1 pixel = 20 m × 20 m *(b)* 1 pixel = 10 m × 10 m *(c)* 1 pixel = 2.5 m × 2.5 m

Figure 1. Satellite images representing the same geographical area (960 m × 650 m) with different resolutions. (a,b) MSR images. (c) HSR image. © CNES 2009 – Kalideos, Isle-Réunion database.

Table 1. Typologies and levels used by end-users to map urban areas at different scales.

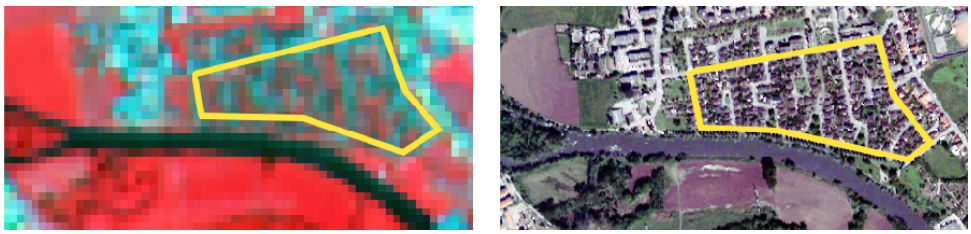
1:100, 000-1:25, 000 – **Urban areas level**	**1:10, 000** – **Urban blocks level**	**1:5, 000** – **Urban objects level**
• High-density urban fabric • Low-density urban fabric • Industrial areas • Forest zones • Agricultural zones • Water surfaces • Bare soil	• Continuous urban blocks • Discontinuous urban blocks - Individual urban blocks - Collective urban blocks • Industrial urban blocks • Urban vegetation • Forest • Agricultural zones • Water surfaces • Roads	• Building/roofs: - red tile roofs - light grey residential roofs - light commercial roofs • Vegetation: - green vegetation - non-photosynthetic vegetation • Transportation areas: - streets - parking lots • Water surfaces: - rivers - natural water bodies • Bare soil • Shadows

MSR or HSR images. Then, it is necessary to develop new methodological tools to consider urban areas at this intermediate semantic level by using the available data (the MSR and HSR ones) in an original fashion. In particular, their complete analysis (*i.e.*, both *segmentation and classification*) can be complementary and possibly provide information related to the semantic level of urban blocks.

1.2 *Multiresolution image analysis: a state of the art*

Multi-image per-pixel approaches In the context of multiresolution approaches (consisting of simultaneous MSR/HSR analysis), different methods have been proposed. A first approach (fusion approach) consists in combining all the descriptions of the objects associated to the different resolutions into a unique image at the highest resolution (Chibani 2005, Chang *et al.* 2007). However, due to the *curse of dimensionality* (Bellman 1961), most of the classical distance-based algorithms are not sufficient to correctly analyse objects having a large number of attributes: the distances between these objects are not sufficiently different to correctly determine the nearest ones. In addition, with the increase of the spectral dimensionality, some problems can appear, like the Hughes phenomenon (Hughes 1968), characterised by the fact that classifier performances decrease when the data dimensionality increases.

An alternative solution aims at finding a consensus of classifications of the images. In (Forestier *et al.* 2008), a framework is proposed to produce a unified result which represents a consensus among unsupervised classifications of different images. However, it requires to generate the same number of classes for all the images, which is generally not relevant for MSR and HSR ones. In (Wemmert *et al.* 2009), some of the authors describe an approach that uses simultaneously two images at different spatial resolutions, and for which each classification does not necessarily have the same number of clusters. This method consists in performing a per-pixel clustering on both images. For each of them, regions are built. Then, regions from the image at the highest resolution are characterised using the clustering of the image at the



(a) Urban blocks on a MSR image. *(b)* Urban blocks on an HSR image.

Figure 2. Urban blocks represented at different resolutions. The sensed area covers a surface of 1, 200 m × 600 m.

lowest resolution (each region of a clustered image is characterised according to its clusters composition in the other one). These regions are finally clustered using these compositions.

This method has produced promising results. Nevertheless, it directly (and only) works on per-pixel classifications, which is a serious weakness for dealing with the issue of the semantic gap (*i.e.*, the lack of concordance between low-level information automatically extracted from the images and high-level information analysed by experts (Smeulders *et al.* 2000)). Moreover, with HSR images the definition of a pure "urban" spectral class is necessarily done by incorporating pixels of other non-urban classes (Mesev *et al.* 2000). This spectral heterogeneity induces the well known salt-and-pepper effect during the application of traditional per-pixel approaches. Furthermore, the limitation to spectral information during these processes disadvantages the recognition of clusters corresponding to the semantic classes (Caprioli and Tarantino 2003, Marangoz *et al.* 2004).

Mono-image region-based approaches To reduce the problems related to per-pixel approaches, new methods using *object-based* (also called *region-based*) strategies, are being developed (Herold *et al.* 2002, Benz *et al.* 2004, Baatz *et al.* 2008). Such methods perform a segmentation pre-processing step to partition the HSR image into homogeneous regions (also called *objects*). Then, in a second step, these regions are gathered according to elementary characteristics like spectral and geometrical properties and, possibly, spatial relationships (contextual texture, topological relations) in order to perform an (unsupervised) region-based classification (Herold *et al.* 2003, Carleer and Wolff 2006). An overview of object-based methods dealing with remote sensing images is presented in (Blaschke 2010).

In the context of urban areas analysis, a region-based method extracting simultaneously information at the object and area levels has been proposed in (Jacquin *et al.* 2008). In this multiscale framework, a segmentation process is applied iteratively on an HSR image to produce partitions with different levels of details. Then, according to its level of details, a partition is used to extract urban areas or urban objects. The same kind of framework has been developed in (Corbane *et al.* 2008) to enhance the mapping of hydrological soil surface. In a different way, a hybrid multilevel pixel/region-based method dealing only with HSR images has been proposed in (Bruzzone and Carlin 2006). The aim of this method is to classify each pixel of the image by merging the spectral and the spatial context information. This spatial context of the pixel is obtained by studying simultaneously several levels of segmentation of the HSR image. For each resulting region of each segmentation, spatial features are computed. Then, for each pixel of the considered image, a vector is defined. Such vector combines the radiometric values of the pixel and its spatial features associated to its corresponding regions at the different levels of segmentation.

Region-based approaches provide promising tools to analyse urbanised territory from HSR images in terms of simple urban objects (individual buildings, road segments, road intersections, cars, etc.). However, these methods are not adapted to extract directly composite objects, like urban blocks (by considering that urban blocks are composed by sets of simple urban objects in HSR imagery). A major issue to deal with composite objects through HSR images consists in grouping several individual objects in order to construct more complex ones corresponding to a higher semantic level. In (Barnsley and Barr 1997) a graph-based, structural pattern recognition system that might be used to infer broad categories of urban land use from HSR images is presented. This system has been considered to analyse discrete land cover parcels by taking into account the structural properties and the relations between simple objects. Such relations are modelled through a graph of linked regions (called XRAG – eXtended Relational Attribute Graph). Preliminary tests performed using this framework on land cover parcels generated from digital (vector) map data have suggested that certain categories of urban semantic classes can be distinguished in terms of their structural compositions. In a more recent work (Guo *et al.* 2008), the authors have proposed an approach to classify complex objects. First, HSR images are segmented by a method which preserves the semantics of real-world objects. The resulting regions are then classified with a fuzzy classification process. Finally, these clustered regions are gathered by using hyperclique patterns to form more complex objects (Industrial buildings, Baseball field, etc.). Another method, similar to the last one but using spatial relationships and hierarchical segmentation, can be found in (Akcay and Aksoy 2008). In this article, a framework which aims at detecting complex objects in HSR images by combining spectral with structural information (exploited by using hierarchical image segmentation) is presented. Given the observation that different structures appear more clearly at different scales in different spectral bands, a new algorithm is described. This method is dedicated to unsupervised grouping of candidate segments belonging to multiple hierarchical segmentations. The idea is to find coherent sets of segments which can correspond to more complex objects. Then, the automatic labelling of the segments is done by computing the similarity of its feature distribution to the distribution of the learnt object models. However, one of the weaknesses of these methods is to omit the information contained in other images (of the studied area) at different spatial resolutions.

1.3 *Purpose*

To conclude on this synthetic state of the art, there exist two main categories of methods dealing with multiresolution analysis in the context of the mapping of urban blocks. The first one consists in simultaneously extracting information from both MSR and HSR images in a per-pixel fashion. The second one, mainly introduced to fix the problems induced by per-pixel strategies (the semantic gap and the limitation to spectral information during the clustering process), is based on a region-based methodology. However it led to techniques enabling to only process HSR images, without any use of other images with different spatial resolutions (by opposition to the per-pixel approaches).

Based on these considerations, the purpose of the work presented in this article is to propose a method which combines the advantages offered by the (per-pixel) multi-image analysis and the efficiency of the (mono-image) region-based frameworks for a *block level* analysis in the context of the mapping of urban areas. In this method, the spatial context of the urban objects and the semantic relationships of these last ones between the available resolutions are used to enhance the

simultaneous analysis of both MSR and HSR images. It is inspired from the work proposed in (Wemmert *et al.* 2009). However, it is based on a quite different way to perform multiresolution analysis. Indeed, in (Wemmert *et al.* 2009) the analysis is carried out (on a per-pixel fashion) by studying the composition of the highest resolution data in terms of clusters in the lowest resolution one, while in the method developed hereafter, the opposite strategy aims at studying the composition of the lowest resolution regions in terms of clusters in the image at the highest resolution. This new way to perform multiresolution analysis permits, in particular, to perform the refinement of final HSR clusters into more specific subclusters matching with hidden land cover classes.

The sequel of this article is organised as follows. In Section 2, the proposed methodology is fully described in both "visual" and "formal" fashions. Section 3 gathers experiments enabling to assess the efficiency of the proposed approach in the context of urban analysis, in particular by comparison with other methods. Finally, conclusions and perspectives will be found in Section 4.

2. Methodology

The methodology described in this section is devoted to the simultaneous analysis of two images of the same scene (namely an urban area) generated at distinct resolutions (see flowchart in Fig. 3). In the standard case, these two resolutions are the MSR and the HSR ones.

- **Step 1:** The two images are first independently segmented by a flat-zone merging process (Step 1 in Fig. 3; Subsection 2.2).
- **Step 2:** The HSR segmented image is clustered using the pixel radiometric average of the regions (right side of Step 2 in Fig. 3; Subsection 2.3). Then the composition of each region from the MSR image in terms of clusters of the HSR one is computed (left side of Step 2 in Fig. 3; Subsection 2.3).
- **Step 3:** Based on the composition of these regions, a clustering of the MSR segmented image is performed (Step 3 in Fig. 3; Subsection 2.4): by opposition to a "classical" clustering, this one aims at defining "semantic" clusters and no longer "radiometric" ones.
- **Step 4:** The global composition of these clusters, in terms of clusters of the highest resolution image, is then computed (right upon side of Step 4 in Fig. 3; Subsection 2.5). Thus, these classes have an intermediate level which can correspond to the semantic block level. Finally, regions of the HSR image are embedded in the MSR image data space in order to assign them to a class of the intermediate level (see ①, ② and ③ of Step 4 in Fig. 3; Subsection 2.5).

The main idea of this method is to fuse the information provided by the analysis of the high spatial resolution regions with the low resolution semantic clustering to obtain a final clustering result corresponding to the urban blocks level. Moreover, another advantage offered by multiresolution analysis is **cluster refinement**. The idea is that the information obtained by analysing simultaneously two different resolutions (HSR and MSR) can be used to improve HSR classification results (at the urban objects semantic level) by discovering hidden HSR subclusters (Subsection 2.6).

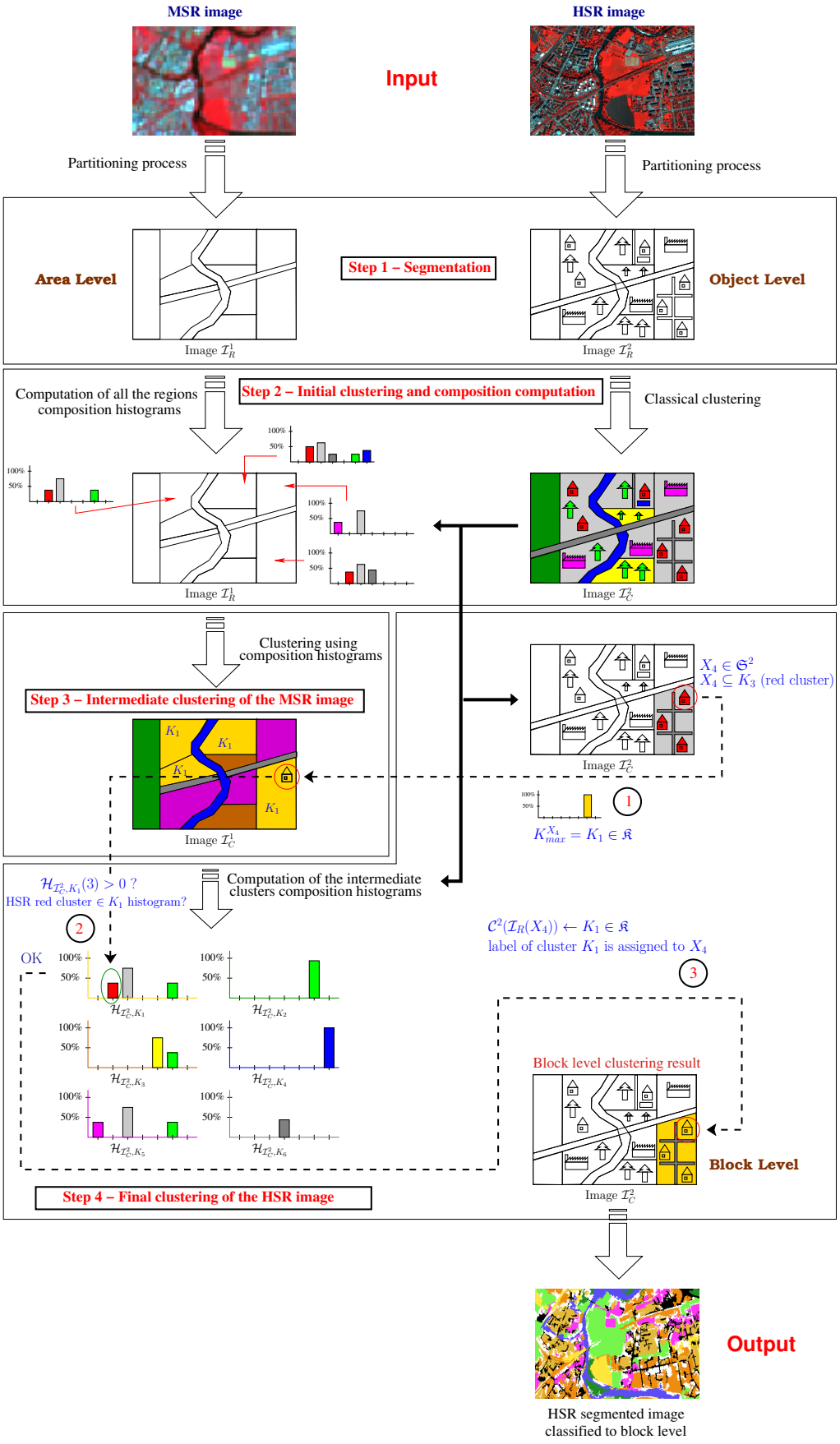


Figure 3. The proposed multiresolution method: inputs are composed of MSR and HSR images and the method provides as output a clustering of the HSR segmented image at the urban blocks semantic level.

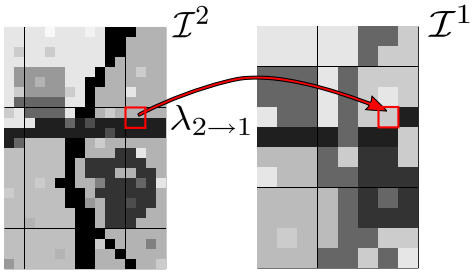


Figure 4. A correspondence map $\lambda_{2\to1}$ between two image functions $\mathcal{I}^2$ and $\mathcal{I}^1$. In this example, $\alpha = 2$, and a pixel in $\mathcal{I}^1$ then corresponds to 4 pixels in $\mathcal{I}^2$.

2.1 *Input/output*

Let $E = [\![0, d_x - 1]\!] \times [\![0, d_y - 1]\!] \subset \mathrm{N}^2$, E is the discrete partition (broadly speaking, the "square grid") of the scene being observed. Let $V_b = [\![0, v_m - 1]\!] \subset \mathrm{N}$, V_b is a discrete sampling of the intensities observed for the spectral band considered for the observation. A (mono-value) image $\mathcal{I}_b$ is defined as a function

$$\left| \begin{array}{rccc} \mathcal{I}_b : & E & \to & V_b \\ & \mathbf{x} = (x, y) & \mapsto & \mathcal{I}(\mathbf{x}) = v \end{array} \right.$$

which to each point (*i.e.*, to each pixel) $\mathbf{x}$ of the observed scene, associates a spectral intensity v. Now, let $V = \prod_{b=1}^{s} V_b$ with $V_b = [\![0, v_{b,m} - 1]\!] \subset \mathrm{N}$ for all $b \in [\![1, s]\!]$, V is an agglomeration of several spectral bands. A (multivalue) image $\mathcal{I}$ is defined as a function

$$\left| \begin{array}{rccc} \mathcal{I} : & E & \to & V \\ & \mathbf{x} = (x, y) & \mapsto & \mathcal{I}(\mathbf{x}) = \mathbf{v} = \prod_{b=1}^{s} \mathcal{I}_b(\mathbf{x}) \end{array} \right.$$

where $\mathcal{I}_b : E \to V_b$ is a monovalue image for all $b \in [\![1, s]\!]$. Broadly speaking, a multivalue image is the agglomeration of several mono-value images.

The method takes as input two multivalue images: a MSR image $\mathcal{I}^1 : E^1 \to V^1$ and an HSR image $\mathcal{I}^2 : E^2 \to V^2$ of the same scene (with $E^1 = [\![0, d_x^1 - 1]\!] \times [\![0, d_y^1 - 1]\!]$, $E^2 = [\![0, d_x^2 - 1]\!] \times [\![0, d_y^2 - 1]\!]$). We recall that –for a same scene– the higher the value of $d_x^\star$, $d_y^\star$ is, the higher the resolution of the image is. In particular, in the current context, we necessarily have $d_x^1, d_y^1 < d_x^2, d_y^2$.

We set $\alpha = d_x^2/d_x^1 = d_y^2/d_y^1 \in \mathrm{N}^*$. The coefficient α characterises the "difference" of resolution between the two (MSR and HSR) images. Note that a point $\mathbf{x} \in E^1$ then "physically" corresponds to a set composed of $\alpha \times \alpha$ points in E^2 (indeed the point $\mathbf{x} \in E^1$ and the $\alpha \times \alpha$ points of E^2 match the same region). In order to model the correspondence between the points of two such images $\mathcal{I}^1$ and $\mathcal{I}^2$, we define the two correspondence maps

$$\left| \begin{array}{rccc} \lambda_{2\to1} : & E^2 & \to & E^1 \\ & \mathbf{x} = (x, y) & \mapsto & (x/\alpha, y/\alpha) \end{array} \right. \quad \text{and} \quad \left| \begin{array}{rccc} \lambda_{1\to2} : & E^1 & \to & \mathcal{P}(E^2) \\ & \mathbf{x} = (x, y) & \mapsto & \alpha \times (x, y) + [\![0, \alpha - 1]\!]^2 \end{array} \right.$$

where $\mathcal{P}(E^2) = \{X \subseteq E^2\}$ is the set of all the subsets of E^2. The map $\lambda_{2\to1}$ indicates in which pixel of the MSR image a point of HSR one lies, while the map $\lambda_{1\to2}$ indicates which set of $\alpha \times \alpha$ pixels of the HSR image corresponds to a (single) pixel of the MSR one. Figure 4 exemplifies the correspondence map $\lambda_{2\to1}$.

The method provides as output a clustering of the scene at an intermediate semantic level (*i.e.*, a level corresponding to a resolution between the ones of $\mathcal{I}^1$ and $\mathcal{I}^2$) which corresponds, in the considered applicative context, to the block level. This clustering is modelled by a label image $\mathcal{R} : E^2 \to [\![1, k]\!] \cup \{\perp\}$ which,

to each point $\mathbf{x}$ of the scene (at the highest resolution), associates a class value $\mathcal{R}(\mathbf{x})$ among the k possible ones, or possibly the undetermined value $\perp$ (in the case where no intermediate semantics has been assigned to this point).

2.2 *Step 1 - Segmentation*

A segmentation of an image $\mathcal{I} : E \to V$ is a partition $\mathfrak{S} = \{S_i\}_{i=1}^{n}$ of E; broadly speaking, the scene visualised in $\mathcal{I}$ is "decomposed" into n distinct parts S_i, which are supposed to present specific semantic properties. In the context of this work, we can assume that each part S_i is connected w.r.t. a chosen adjacency; we will denote S_i as a *region*. To any segmented image $\mathcal{I}$, we then associate a region image $\mathcal{I}_R : E \to [\![1, n]\!]$ defined such that for all $\mathbf{x} \in E$, $\mathbf{x} \in S_{\mathcal{I}_R(\mathbf{x})}$; $\mathcal{I}_R$ can be seen as a synthetic image (with the false colours $1, 2, \ldots, n$) induced by the segmentation of the image $\mathcal{I}$.

Two main approaches are generally considered for the segmentation of satellite images: watershed segmentation (Vincent and Soille 1991) and "region-growing" segmentation (Cross *et al.* 1988). Recent studies (Meinel and Neubert 2004, Carleer *et al.* 2005) devoted to the comparison of these two kinds of segmentation techniques (especially in the case of HSR images) have emphasised the fact that region-growing approaches tend to outperform watershed ones.

Based on these studies (corroborated by preliminary experiments), it has been chosen to use a region-growing method to perform segmentation of both images. Such a method is initialised with a trivial partition of the image, and iteratively merges elements of this partition, chosen as the pairs of adjacent regions minimising a given evaluation function f.

The initial partition $\mathfrak{S}^{init} = \{S_i\}_{i=1}^{n^{init}}$ considered here is composed of the flat zones of the image. This choice is computationally less expensive than a partition composed by all the singleton sets of the image pixels. Moreover it guarantees that the constant regions are preserved in the segmentation result.

The merging process joins regions by considering the Region Adjacency Graph (RAG) formed by the image flat zones. The one used in this work (Baatz and Schape 2000) is guided by two main criteria: colour and shape. They enable to optimise the regions spectral homogeneity and spatial complexity. The ratio between both criteria depends on the desired output.

The evaluation function f used here is called heterogeneity function. It corresponds to the increase of heterogeneity (*i.e.*, the "difference") between the region $X_{1,2}$ susceptible to be formed and the two adjacent regions (X_1 and X_2) candidate to this fusion

$$f = w_{colour} \times \Delta h_{colour} + w_{shape} \times \Delta h_{shape} \tag{1}$$

where w_{colour}, w_{shape} are weights (which verify $w_{colour}, w_{shape} \geq 0$ and $w_{colour} + w_{shape} = 1$). The spectral heterogeneity Δh_{colour} corresponds to the increase of variance between the one of the region $X_{1,2}$ and the sum of the ones of the regions X_1 and X_2. The shape heterogeneity Δh_{shape} is a value that describes the improvement of the shape w.r.t. smoothness and compactness of an object shape (see (Baatz and Schape 2000) for more details related to the definition of f).

During the iterative merging process, f is computed for each couple of adjacent regions of $\mathfrak{S}^{current}$. Then, the couples of regions which minimise the evaluation function f are merged into (larger) ones.

Prior to the potential fusion of two adjacent regions, the resulting increase of heterogeneity f is computed. If it exceeds a given threshold τ (called "scale param-

eter") determined by the user, then no further fusion occurs and the segmentation ends.

This segmentation process is applied independently to both input images $\mathcal{I}^1$ and $\mathcal{I}^2$ to obtain the segmented images $\mathcal{I}^1_R$ and $\mathcal{I}^2_R$ (Step 1 in Fig. 3). The partition corresponding to $\mathcal{I}^1_R$ (resp. $\mathcal{I}^2_R$) is called $\mathfrak{S}^1$ (resp. $\mathfrak{S}^2$) and the number of elements in this set is denoted by n^1 (resp. n^2).

2.3 *Step 2 - Initial clustering and composition computation*

Initial clustering of the HSR segmented image Let $\mathfrak{S} = \{S_i\}_{i=1}^n$ be a segmentation of an image $\mathcal{I} : E \to V$, and $\mathcal{I}_R : E \to [\![1, n]\!]$ be the associated region image. A clustering of $\mathcal{I}$ into k classes is provided by the definition of a map $\mathcal{C} : [\![1, n]\!] \to [\![1, k]\!]$ which, to each one of the n regions S_i, associates one of the k classes $\mathcal{C}(i)$. A cluster K_i induced by such a clustering is then defined by $K_i = \bigcup_{j \in \mathcal{C}^{-1}(\{i\})} S_j$, *i.e.*, by gathering all the regions S_j which correspond to a same class. The set of the k clusters of $\mathcal{I}$ is noted $\mathfrak{K} = \{K_i\}_{i=1}^k$. Similarly to the case of segmentation, to any clustered image $\mathcal{I}$, we associate a cluster image $\mathcal{I}_C : E \to [\![1, k]\!]$ defined such that for all $\mathbf{x} \in E$, $\mathbf{x} \in K_{\mathcal{I}_C(\mathbf{x})}$; $\mathcal{I}_C$ can be seen as a synthetic image (with the false colours $1, 2, \ldots, k$) induced by the clustering of the image $\mathcal{I}$. Note in particular that $\mathcal{I}_C$ is straightforwardly defined as the composition of the region image and its clustering, *i.e.*, $\mathcal{I}_C = \mathcal{C} \circ \mathcal{I}_R$.

In this step, a clustering is performed on the segmented (HSR) image $\mathcal{I}^2_R$, using the radiometric values of the pixels of E^2. This leads to the generation of a cluster image $\mathcal{I}^2_C : E^2 \to [\![1, k^2]\!]$ (right side of Step 2 in Fig. 3).

Regions composition computation of the MSR segmented image Let $\mathcal{I}_C : E \to [\![1, k]\!]$ be a cluster image. The *composition histogram* of $\mathcal{I}_C$, noted $\mathcal{H}_{\mathcal{I}_C}$ is defined by

$$\left| \begin{array}{rcl} \mathcal{H}_{\mathcal{I}_C} : [\![1, k]\!] & \to & \mathrm{N} \\ i & \mapsto & |\mathcal{I}_C^{-1}(\{i\})| \end{array} \right. \tag{2}$$

Note that a composition histogram is nothing but a classical histogram, except that the "real" values of $\mathcal{I}$ are replaced here by "symbolic" ones (corresponding to its clusters).

The composition histogram of $\mathcal{I}_C$ associated to a subset $X \subseteq E$, noted $\mathcal{H}_{\mathcal{I}_C, X}$ is defined by

$$\left| \begin{array}{rcl} \mathcal{H}_{\mathcal{I}_C, X} : [\![1, k]\!] & \to & \mathrm{N} \\ i & \mapsto & |\mathcal{I}_C^{-1}(\{i\}) \cap X| \end{array} \right. \tag{3}$$

It corresponds to the composition histogram of $\mathcal{I}_C$, restricted to the set X; in particular, we obviously have $\mathcal{H}_{\mathcal{I}_C, E} = \mathcal{H}_{\mathcal{I}_C}$.

Once the HSR image $\mathcal{I}^2_R$ has been classified, it becomes possible to determine the composition of each region X of the MSR image $\mathcal{I}^1_R$ (*i.e.*, each $X \in \mathfrak{S}^1$) w.r.t. the cluster image $\mathcal{I}^2_C$. This composition is actually defined as the composition histogram $\mathcal{H}_{\mathcal{I}^2_C, \lambda_{1\to 2}(X)}$ (left side of Step 2 in Fig. 3) defined by

$$\left| \begin{array}{rcl} \mathcal{H}_{\mathcal{I}^2_C, \lambda_{1\to 2}(X)} : [\![1, k^2]\!] & \to & \mathrm{N} \\ i & \mapsto & |\bigcup_{\mathbf{x} \in X} \lambda_{1\to 2}(\mathbf{x}) \cap (\mathcal{I}^2_C)^{-1}(\{i\})| \end{array} \right. \tag{4}$$

Broadly speaking, this composition histogram associates to each label i of the clustered HSR image, the number of pixels which have the label i and which correspond to a pixel of the considered segmented region X of the MSR image.

2.4 *Step 3 - Intermediate clustering of the MSR image*

The previous step provides, for each segmented region $X \in \mathfrak{S}^1$ of the MSR image $\mathcal{I}^1$, its composition in terms of classes of the clustering of the HSR image $\mathcal{I}^2$, under the form of a composition histogram (left side of Step 2 in Fig. 3). Then, it becomes possible to compute a clustering of the regions of $\mathfrak{S}^1 = \{X_i^1\}_{i=1}^{n^1}$ based on the value of these regions in the space of composition histograms. This clustering $\mathcal{C} : [\![1, n^1]\!] \to [\![1, k]\!]$ enables to gather regions presenting similar characteristics w.r.t. the "objects" composing them. This leads, in particular, to the identification of local and frequent associations of structures identified in the HSR image, forming meta-structures at a coarser resolution.

This process provides a cluster image $\mathcal{I}_C^1 : E^1 \to [\![1, k]\!]$ associated to the MSR image $\mathcal{I}^1$ (Step 3 in Fig. 3) which is indirectly based on the radiometric values of $\mathcal{I}^1$ (thanks to the initial segmentation $\mathcal{I}_R^1$) and directly based on the implicit semantics of the HSR image $\mathcal{I}^2$ (thanks to its clustering $\mathcal{I}_C^2$, modelling the composition of the regions of the MSR image $\mathcal{I}^1$). This cluster image gathers information related to both MSR and HSR images, but at a medium level. Consequently, the resulting classes can potentially be close to the ones defined for block level analysis.

2.5 *Step 4 - Final clustering of the HSR image*

Similarly to the computation of the composition histogram for the regions of the segmentation $\mathfrak{S}^1$ of the MSR image $\mathcal{I}^1$, it is possible to compute the composition histogram for the k clusters obtained from the classification $\mathcal{C}$ described in the previous section, in terms of classes of the classification $\mathcal{I}_C^2$ of the HSR image $\mathcal{I}^2$. Such composition histograms provide information related to the way the classes of $\mathcal{C}$ are formed by the classes of $\mathcal{I}_C^2$ obtained at the highest resolution (right upon side of Step 4 in Fig. 3). Formally, these histograms are defined, for $i \in [\![1, k]\!]$ as $\mathcal{H}_{\mathcal{I}_C^2, \lambda_{1\to 2}((\mathcal{I}_C^1)^{-1}(\{i\}))}$, following the same definition as proposed in Eq. 4.

In order to simplify these histograms, and in particular to remove "semantic noise", a threshold Λ_t (defined by $\Lambda_t(v) = v$ if $v \geq t$ and 0 otherwise) is applied on them, thus removing the values corresponding to HSR classes having a non-relevant contribution. The final pruned composition histograms, noted $\mathcal{H}_{\mathcal{I}_C^2, i}$, are then defined as $\mathcal{H}_{\mathcal{I}_C^2, i} = \Lambda_t \circ \mathcal{H}_{\mathcal{I}_C^2, \lambda_{1\to 2}((\mathcal{I}_C^1)^{-1}(\{i\}))}$. Experimentally, it has been observed that choosing t close to the mean of the histogram values (*i.e.*, $|(\mathcal{I}_C^1)^{-1}(\{i\})|/k^2$), provided satisfying results.

Then, the regions of the HSR segmented image are embedded in the MSR image data space in order to assign them a class of the intermediate level (see ①, ② and ③ of Step 4 in Fig. 3). The idea consists in assigning to each region of the segmentation $\mathfrak{S}^2$ of the HSR image $\mathcal{I}^2$, a label of a cluster of $\mathfrak{K}$ provided by the classification $\mathcal{C}$. To this end, for each segmented HSR region $X_i \in \mathfrak{S}^2$, we compute its composition in terms of MSR/HSR classes of $\mathcal{I}_C^1$ (provided by the classification $\mathcal{C}$, see Step 3 in Fig. 3). These n^2 histograms are then actually defined, for $i \in [\![1, n^2]\!]$ as $\mathcal{H}_{\mathcal{I}_C^1, \lambda_{2\to 1}(X_i^2)}$ (see Eq. 3). From these composition histograms, we are then able to find, for each HSR segmented region $X_i \in \mathfrak{S}^2$, its main composition in terms of classes of $\mathcal{I}_C^1$ –the intermediate clustering of the MSR image (see ① of Step 4 in Fig. 3). In other words, for each segmented region X_i of the HSR image, we search what is the class of the main cluster of $\mathfrak{K}$ which composes X_i in $\mathcal{I}_C^1$. We denote by

$K_{max}^{X_i} \in \mathfrak{K}$ this cluster. It is defined, for each region X_i as

$$K_{max}^{X_i} = K_{t_{max}}^{X_i}, \text{ where } t_{max} = \underset{t\in[1,k]}{\arg\max}\{\mathcal{H}_{\mathcal{I}_C^1,\lambda_{2\to1}(X_i)}(t)\} \tag{5}$$

Here, t_{max} is the main class which composes X_i in $\mathcal{I}_C^1$. We call p its majority percentage (we have, in particular, $p = \mathcal{H}_{\mathcal{I}_C^1,\lambda_{2\to1}(X_i)}(t_{max})/|X_i|$). Given a "majority threshold" $s_{maj} \in [0,1]$, two cases can occur:

- $p < s_{maj}$ — X_i is not **embeddable**: this means that the class t_{max} is not "sufficiently majoritary", *i.e.*, that X_i is not correctly correlated (from a spatial point of view) to a semantic area at the MSR resolution. We consider that this is due to a problem in the segmentation process (in the MSR and/or the HSR image). We then assign the class label $\perp$ to X_i.
- $p \geq s_{maj}$ — X_i is **embeddable**: from the composition histogram $\mathcal{H}_{\mathcal{I}_C^2,K_{t_{max}}}$ of the cluster $K_{t_{max}}$ (which indicates what are the classes at the HSR resolution which mainly compose the cluster $K_{t_{max}}$), we get $v = \mathcal{H}_{\mathcal{I}_C^2,K_{t_{max}}}(\mathcal{C}^2(i))$, namely the contribution, in this composition histogram, of the class corresponding to the region X_i at the HSR resolution. Then, two cases can occur:
  - ⋆ $v = 0$ — X_i is **unclassifiable** at an intermediate semantic level: this means that the class of the HSR region X_i is not coherent with the HSR classes which should (mainly, w.r.t. the threshold Λ) compose the "MSR" cluster $K_{max}^{X_i}$. We consider that this is due to a problem in the classification process. We then assign the class label $\perp$ to X_i.
  - ⋆ $v > 0$ — X_i is **classifiable** at an intermediate semantic level (see ② of Step 4 in Fig. 3). We then assign the class label of $K_{t_{max}}^{X_i}$ to X_i (see ③ of Step 4 in Fig. 3).

A possible supplementary step consists in forcing **unclassifiable** regions to belong to a class label of the intermediate level by considering their neighbour regions. First, for each unclassifiable region $X_i \in \mathfrak{S}^2$, we get the label j of its class in $\mathcal{I}_C^2$ (the label of X_i in HSR clustering). Then, we search what is the main class label of all the neighbours of X_i in $\mathcal{C}$. Let $K_{max}^{X_i} \in \mathfrak{K}$ be the cluster corresponding to this class label. If $\mathcal{H}_{\mathcal{I}_C^2,K_{max}^{X_i}}(j) > 0$, then X_i takes the class label of $K_{max}^{X_i}$. Otherwise, X_i keeps the class label $\perp$.

2.6 *Addendum: Cluster refinement*

Another advantage offered by multiresolution analysis is the possibility to perform HSR cluster refinement. Indeed, the composition of MSR clusters in terms of clusters in HSR image, enable to split HSR clusters into more specific ones. For example, a *Vegetation cluster* extracted from the HSR image and taking part in the composition of different MSR clusters (*Industrial area*, *High-density urban fabric*) extracted from the MSR image could be split into two distinct HSR clusters (*Industrial vegetation cluster* and *Urban vegetation cluster*). The analysis of the MSR image gives contextual information during the analysis of the HSR one. Figure 5 illustrates this cluster refinement framework. The four steps necessary to perform cluster refinement are described bellow. Note that the described methodology is devoted to the simultaneous analysis of two images ($\mathcal{I}^1$ and $\mathcal{I}^2$) of the same urban scene generated at distinct resolutions.

- **Step A - Segmentation** The two images are first segmented (independently) by the previous method. Let be $\mathcal{I}_R^1$ and $\mathcal{I}_R^2$ the resulting segmented images (Step A in Fig. 5). This first step is similar to the Step 1 in Subsection 2.2.

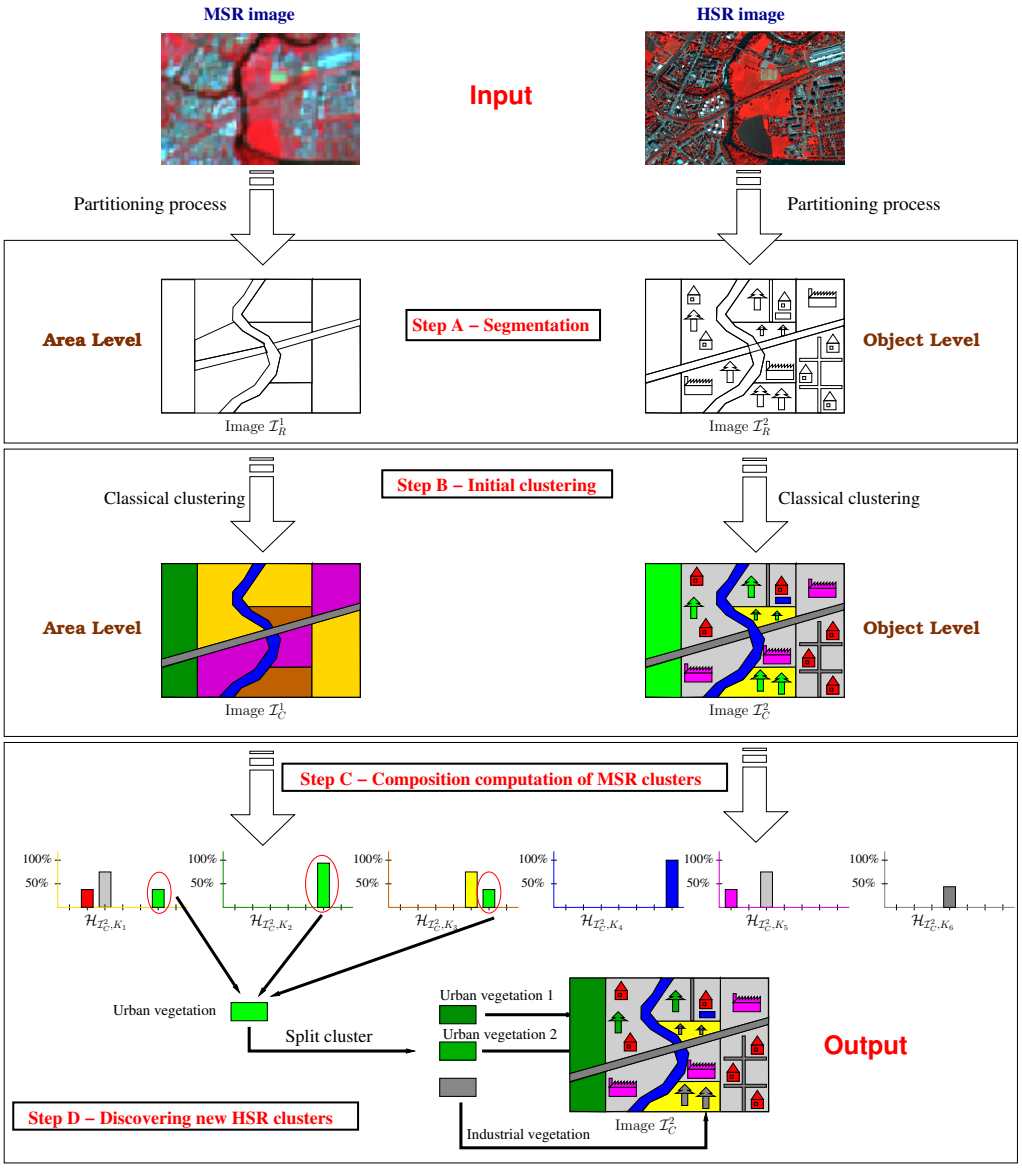


Figure 5. Cluster refinement: discovering hidden HSR subclusters. In this case, data inputs are composed of MSR and HSR images and the refinement method provides as output a clustering result of the HSR segmented image at the semantic level of urban objects.

- **Step B - Initial clustering** The region image $\mathcal{I}_R^1$ (resp. $\mathcal{I}_R^2$) is classified, based on the photometric values of the pixels of E^1 (resp. E^2), in its own semantic level (Urban Area level *vs.* Urban Object level) (see Step B in Fig. 5). The resulting cluster images are denoted by $\mathcal{I}_C^1$ and $\mathcal{I}_C^2$.
- **Step C - Composition computation of MSR clusters** Once $\mathcal{I}_R^1$ and $\mathcal{I}_R^2$ have been classified, it becomes possible to determine the composition of each cluster K of $\mathcal{I}_C^1$ w.r.t. the cluster image $\mathcal{I}_C^2$ (see Step C in Fig. 5).
- **Step D - Discovering new HSR clusters** Finally, based on the composition of these MSR clusters, HSR clusters could be split into specialised clusters: if an HSR cluster takes part in more than one MSR cluster histogram, this HSR cluster is split into different HSR subclusters (see Step D in Fig. 5). More formally, we split $K_i \in \mathfrak{K}_2$ if $\exists j_1 \neq j_2$ with $\mathcal{H}_{\mathcal{I}_C^2,\lambda_{1\to 2}((\mathcal{I}_C^1)^{-1}(\{j_1\}))}(i) > s$ and $\mathcal{H}_{\mathcal{I}_C^2,\lambda_{1\to 2}((\mathcal{I}_C^1)^{-1}(\{j_2\}))}(i) > s$ where s is a parameter set by the user. The split subclusters are called K_{ij_1} and K_{ij_2}. The number of resulting clusters in the HSR segmented image is denoted by k_{ref}^2.

This HSR refinement method provides as output a clustering result of the HSR segmented image at the semantic level of urban objects. This clustering is modelled by a label image $\mathcal{R} : E^2 \rightarrow [\![1, k_{ref}^2]\!]$ which, to each point $\mathbf{x}$ of the scene (at the highest resolution), associates a class value $\mathcal{R}(\mathbf{x})$ among the k_{ref}^2 possible ones.

3. Experiments and results

This section describes the experiments carried out with the proposed multiresolution framework in the context of clustering of urban patterns extracted from MSR and HSR images. Subsection 3.1 presents the data which were used to perform the method. Experiments and parametrisation are described in Subsection 3.2. The results of the multiresolution method devoted to analyse urban blocks are then presented and analysed in Subsection 3.3. Finally, some of the results obtained with the proposed cluster refinement method and their qualitative validations by an expert are showed in Subsection 3.4.

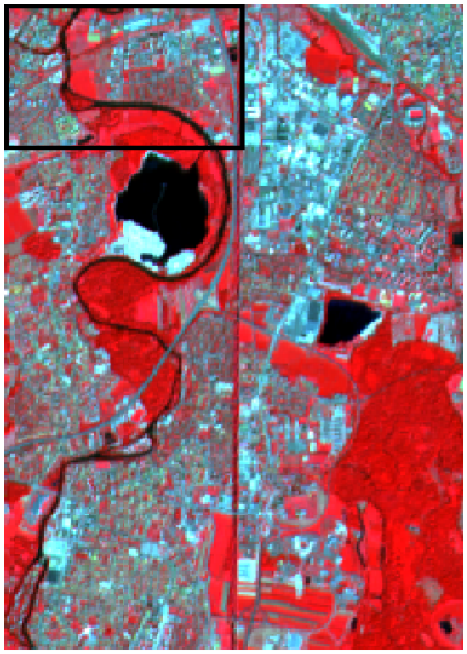
3.1 *Material*

Experiments have been performed on two sets of images called Strasbourg dataset and Toulouse dataset.

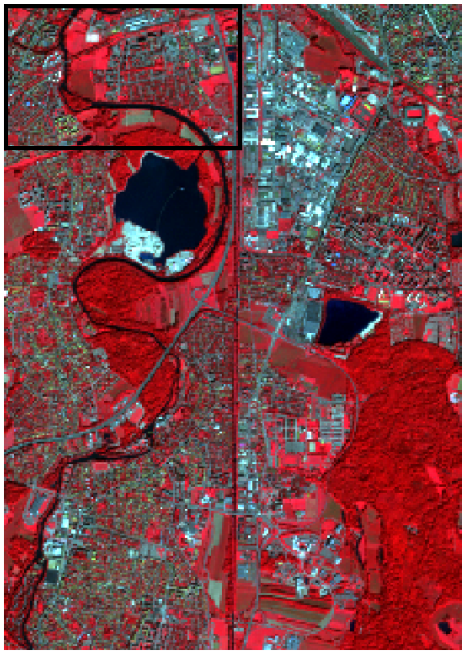
The Strasbourg dataset is composed by three multispectral images with different spatial resolutions (2.8 m, 10 m and 20 m) acquired by the QuickBird, Spot-5 and Spot-4 satellites (respectively in May 2001, August 2002 and July 2001). The Spot-5 and Spot-4 multispectral images (Fig. 6(a, b)) have three spectral bands (green, red, near-infrared). The QuickBird multispectral image (Fig. 6(c)) is available in four spectral bands (blue, green, red and near-infrared). All the data are georeferenced in the same local cartographic projection (Lambert I). These images present a part (1, 500 m $\times$ 2, 100 m) of the urban area of Strasbourg (France) which is a typical suburban area with water surfaces (in black, centre of the image), forest area (in red, bottom left of the image), industrial areas (in grey, upper right of the image), 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), agricultural zones with different spectral responses due to the seasons (bare soil at the end of spring on the HSR image can appear in red in summer on the MSR image).

The Toulouse dataset is composed by three multispectral images[1] with approximately the same spatial resolutions as above (2.5 m, 10 m and 20 m). The HSR image (2.5 m) was acquired by the Spot-5 satellite in September 2003 (Fig. 7(c)). This image is a result of a fusion between the panchromatic image at 2.5 m and XS bands at 5 m. The resulting HSR image has four spectral bands (green, blue, red, near-infrared). The MSR images (20 m and 10 m) were simulated from the HSR one by a degradation process (Fig. 7(a, b)). This degradation process transforms HSR images into MSR ones by simulating the "physical" properties of such sensors. These images are georeferenced in the same local cartographic projection (Lambert III) and present a part (1, 600 m $\times$ 2, 100 m) corresponding to the South West of the city of Toulouse (France) which is also a typical suburban area.

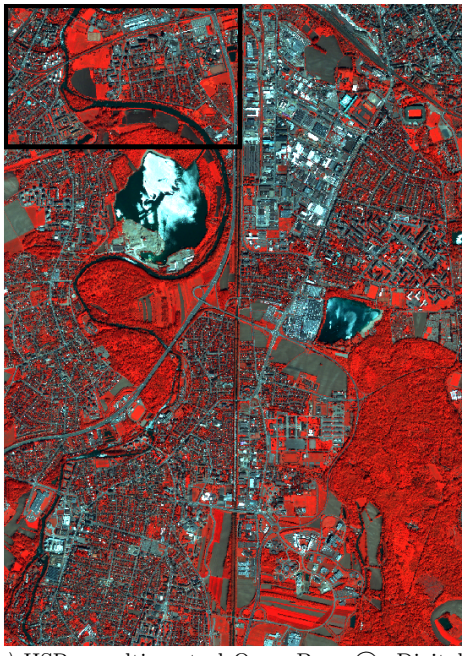
[1]The authors would like to acknowledge the support of the Centre National d'Études Spatiales (CNES) which provided the images of the Toulouse dataset. We are grateful to Jordi Inglada for his assistance in providing and processing these data.



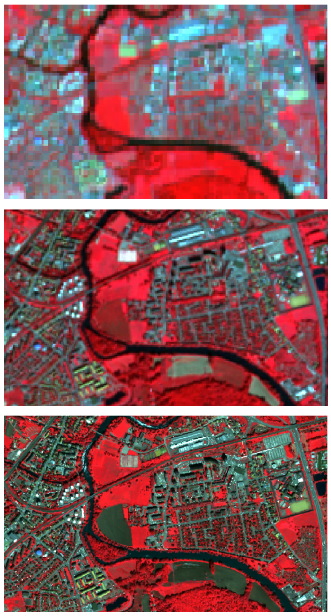
(a) MSR - multispectral SPOT-4 © CNES (Isis program). Image at 20 m.



(b) MSR - multispectral SPOT-5 © CNES (Isis program). Image at 10 m.



(c) HSR - multispectral QUICKBIRD © DigitalGlobe Inc. Image at 2.8 m.



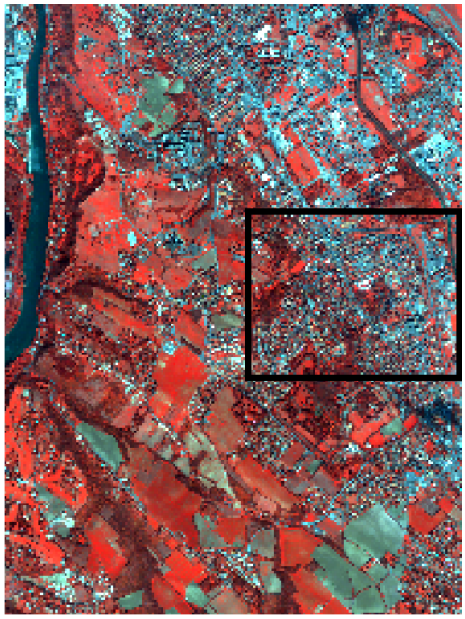
(d) Zoom (750 m × 525 m) on North West part of (a, b, c).

Figure 6. Data of the STRASBOURG dataset. These images present an extract (1, 500 m × 2, 100 m) of the urban area of Strasbourg (France).

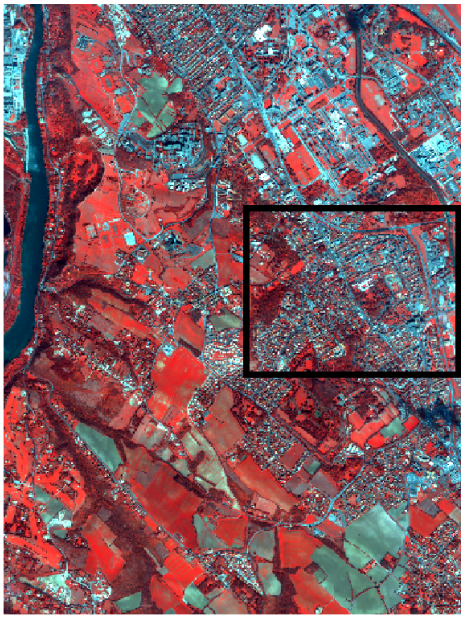
3.2 *Experiments and parametrisation*

To evaluate the multiresolution method (devoted to analyse urban blocks), we have carried out experiments with different configurations and parameters:

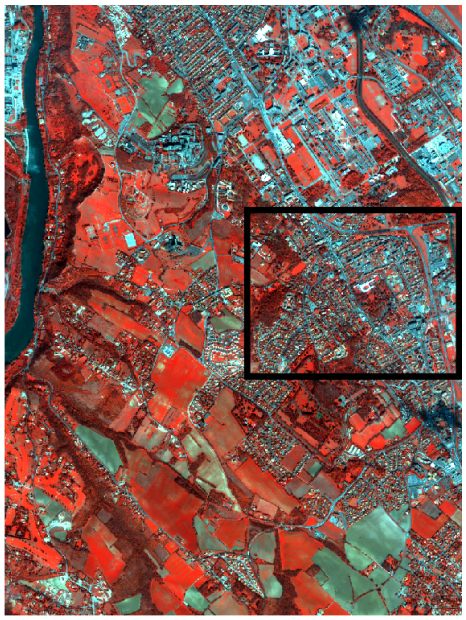
(i) to study the influence of the initial segmentations on the final result, we have run the segmentation algorithm with different scale parameters (Step 1 of the process, Subsection 3.3.2);
(ii) to study the influence of the number of clusters in both classifications, we



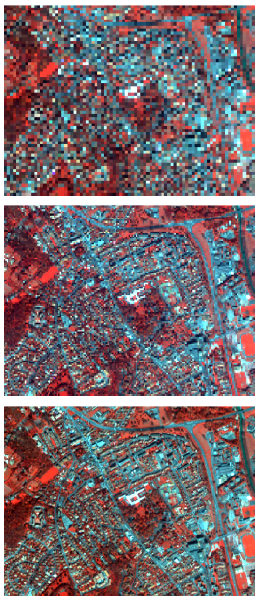
(a) MSR - multispectral simulated image degraded from the HSR SPOT-5 © CNES. Image at 20 m.



(b) MSR - multispectral simulated image degraded from the HSR SPOT-5 © CNES. Image at 10 m.



(c) HSR - multispectral THX SPOT-5 © CNES. Image at 2.5 m.



(d) Zoom (800 m × 550 m) on Centre East part of (a, b, c).

Figure 7. Data of the TOULOUSE dataset. These images present an extract (1, 600 m × 2, 100 m) of the urban area of Toulouse (France).

have run the algorithm with different numbers of clusters in the classifications (Steps 2 and 3 of the process, Subsection 3.3.3);

(iii) to study the influence of the spatial resolutions of MSR images on the final result, we have run the method with different initial configurations: (2.8 m/10 m) and (2.8 m/20 m) for the STRASBOURG dataset; (2.5 m/10 m) and (2.5 m/20 m) for the TOULOUSE dataset) (Subsection 3.3.4).

The four steps described in Section 2 (Subsection 2.2 to 2.5) have been performed as follows.

- **Step 1 - Segmentation** To find the best partition for each image according to the studied area, the segmentation process has been applied on both images with different parameters.

  The parameters w_{colour}, w_{shape} (Eq. 1) were respectively set to 0.75 and 0.25. This parameter configuration gives priority to spectral heterogeneity. For the HSR and MSR images, the scale parameter τ was successively set to distinct increasing values (see Table 2) in order to produce segmentations with different levels of details. In the sequel, the number of elements (*i.e.*, the regions to be classified) in the MSR (resp. HSR) partitions is referred as R_{MSR} (resp. R_{HSR}).
- **Step 2 - Initial clustering and composition computation** The initial classifications of the HSR segmented image were produced by using the classical K-MEANS algorithm. Note that any clustering algorithm which can deal with numerical data could also be used. For this kind of images, the number of clusters depends on the materials of the urban objects which can appear in the studied area. In agreement with the experts, we have carried out experiments with 15, 20, 22 and 25 classes (the number of classes used in each clustering is denoted by C_{HSR}). Then, for each experiment, the composition histograms of the regions from the MSR segmented image (in terms of clusters in HSR image) have been computed.
- **Step 3 - Intermediate clustering of the MSR image** During this step, the K-MEANS algorithm was applied on the MSR segmented images using the composition histograms. Experiments have shown that the method did not directly find all the appropriate clusters w.r.t. to the block level. To tackle this problem, the K-MEANS algorithm has been run with a higher number of clusters (9, 11, 13 and 15 clusters) noted C_{inter}. A post-processing step which consists in applying a hierarchical ascendant clustering algorithm can then be applied in order to reduce the number of classes (9 classes in the current application, see Table 1, centre column). This post-processing step is performed after Step 4 (Subsection 3.3.5).
- **Step 4 - Final clustering of the HSR image** The regions of the HSR image were embedded in the MSR image data space in order to assign them to a class of the intermediate level. The majority percentage s_{maj} was set to 75% (*i.e.*, an HSR region is embeddable in the intermediate level if 75% of the pixels of the region get the same intermediate label cluster during the projection).

These four steps were performed on the whole images presented in the previous subsection with the different sets of parameters described above. For a better visualisation, some of the results are only presented on an extract of the studied zone, corresponding to the North West part of Figure 6 for the STRASBOURG dataset and to the Centre East part of Figure 7 for the TOULOUSE dataset.

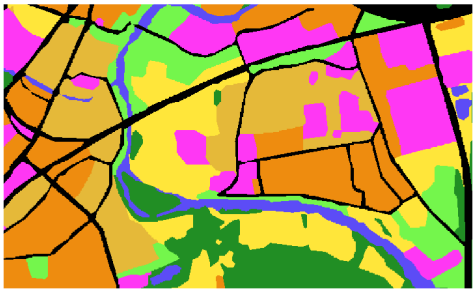
3.3 *Results analysis of urban blocks classification*

3.3.1 **Results evaluation.** Results produced by the method have been assessed by qualitative and quantitative comparisons with an extract (corresponding to the North West part for Strasbourg and the Centre East par for Toulouse) of a groundtruth map from a land-cover/land-use database used for a 1:10,000 mapping. These extracts of maps (Fig. 8(a, b)) contain 8 thematic classes at the semantic block level.

We have computed the Kappa index (Tables 3 and 4), which is a measure of

Table 2. Segmentations parameters and number of regions obtained for the two datasets STRASBOURG and TOULOUSE. The threshold τ is the segmentation parameter. The symbol R_{MSR} (resp. R_{HSR}) corresponds to the number of regions in the resulting partitions of the MSR (resp. HSR) image.

STRASBOURG dataset					
HSR (2.8 m)	τ	20	25	30	40
	R_{HSR}	25,594	19,752	15,105	9,314
MSR (10 m)	τ	10	13	17	20
	R_{MSR}	8,785	5,775	3,744	2,854
MSR (20 m)	τ	12	15	20	25
	R_{MSR}	2,255	1,688	1,067	768
TOULOUSE dataset					
HSR (2.5 m)	τ	20	25	30	40
	R_{HSR}	24,897	19,006	13,115	9,727
MSR (10 m)	τ	10	13	17	20
	R_{MSR}	7,863	4,980	2,482	1,515
MSR (20 m)	τ	12	15	20	25
	R_{MSR}	3,084	1,454	728	437



(a) Groundtruth map of the STRASBOURG dataset (BDOCS 2000 CIGAL 2003).

(b) Groundtruth map of the TOULOUSE dataset.

Figure 8. Groundtruth maps of the studied areas at the semantic block level.

global classification accuracy (Congalton 1991). The Kappa takes value in $[0, 1]$ and decreases as the classification is in disagreement with the groundtruth map. For instance, a value between 1.00 and 0.81 reflects a "perfect" agreement, a value between 0.80 and 0.61 indicates a good agreement and so on.

To get an overview of the impact of initial segmentations and classifications on the final image clustering results, the parameters which have been involved to produce these results have been studied independently, as described below. Note that in all the presented results, the colours of the clusters were chosen to correspond to those defined for the groundtruth map.

3.3.2 **Impact of the segmentations.** To highlight the influence of the initial segmentations on the final result, the classification parameters C_{HSR} (number of classes in HSR initial clustering) and C_{inter} (number of classes in MSR intermediate clustering) have been set to 20 and 13 respectively. Only the initial segmentations have been modified (Table 3). These experiments have shown that:

STRASBOURG dataset:

- With partitions composed of 9,314 and 15,105 regions from HSR image and 733 and 1,067 regions from MSR image at 20 m (resp. 2,854 regions from

MSR image at 10 m), the classes obtained do not match with the groundtruth map. Indeed, too many regions from HSR image (resp. MSR images) are too large to correspond to urban objects (resp. urban blocks). Moreover there are probably not enough regions to be classified.

- With $25,594$ regions from HSR image and $2,255$ regions from MSR image at 20 m (resp. $5,557$ and $8,785$ regions from MSR image at 10 m), numerous regions are too small and correspond to parts of urban objects or urban blocks.
- With $19,752$ regions from HSR image and $1,688$ regions from MSR image at 20 m (resp. $3,744$ regions from MSR image at 10 m), the number and the sizes of the regions seem well fitted for the classification of the images. Indeed, the visual interpretation of regions confirmed that the extracted regions are close to the objects of interest. The resulting classes better match those from the groundtruth map defined for the urban block analysis. There are also fewer problems of over/under-segmentation.

For these reasons, the segmentation from the MSR image at 20 m with $1,688$ regions (resp. $3,744$ regions for the MSR image at 10 m) and the segmentation from the HSR image with $19,752$ regions are kept for the next steps.

Toulouse dataset:
In the same way, experiments carried out on the Toulouse dataset have shown that the segmentation from MSR image at 20 m with $1,454$ regions (resp. $2,482$ regions for the MSR image at 10 m) and the segmentation from HSR with $13,115$ regions seem well fitted for the classification of the images. Thus, these segmentations will be kept for the next steps.

To summarise, these first experiments have shown that the quality of the clustering results is directly linked to the choice of the segmentation parameters. In order to obtain suitable results of classification it is necessary to choose good sets of parameters. This study have shown that in the two datasets the best configurations of parameters are quite similar. For the segmentation of an HSR image a good value of τ takes range between 25 and 30. For the segmentation of an MSR image at 20 m, the parameter τ could be set to 15 whereas for the segmentation of an MSR image at 10 m the threshold τ could be set to 17. Consequently, the results obtained in this study could help the end-user to choose the optimal values of segmentation parameters.

3.3.3 **Impact of the initial classifications.** To study the influence of the initial classifications on the clustering result, we have set the C_{HSR} and C_{inter} parameters to different values. For each experiment on the two partitions obtained previously (Subsection 3.3.2), the Kappa index has been computed (Table 4). From this table, one can see that both parameters have an actual influence:

Strasbourg dataset:

- C_{HSR}: If the number of HSR clusters is too small ($C_{HSR} = 15$), some HSR clusters seem to be irrelevant (probably due to the low number of clusters). This leads to the construction of incorrect intermediate clusters. For instance, some Agricultural zones belong to the same cluster as Water surfaces (Fig. 9(a, e)). However, if the number of HSR clusters is too high ($C_{HSR} = 25$), local and frequent associations of HSR structures (Step 3 of the method) are difficult to identify (probably since an insufficient number of MSR regions have a similar

Table 3. **Impact of the initial segmentations** of HSR and MSR images on the final clustering result. The Kappa values for the different experiments are presented on the STRASBOURG and the TOULOUSE datasets. The best values obtained are depicted in red.

STRASBOURG dataset					
MSR Image \ **HSR Image**		**Image at 2.8 m**			
		$R_{HSR} = 9,314$	$R_{HSR} = 15,105$	$R_{HSR} = 19,752$	$R_{HSR} = 25,594$
Image at 20 m	$R_{MSR} = 733$	0.7981	0.7939	0.7897	0.7762
	$R_{MSR} = 1,067$	0.8004	0.8054	0.8044	0.7895
	$R_{MSR} = 1,688$	0.8105	0.8097	**0.8118**	0.8081
	$R_{MSR} = 2,255$	0.7905	0.7956	0.8023	0.7932
Image at 10 m	$R_{MSR} = 2,854$	0.7673	0.7767	0.7845	0.7627
	$R_{MSR} = 3,744$	0.7726	0.7824	**0.7904**	0.7786
	$R_{MSR} = 5,775$	0.7710	0.7813	0.7883	0.7797
	$R_{MSR} = 8,785$	0.7643	0.7744	0.7797	0.7619

TOULOUSE dataset					
MSR Image \ **HSR Image**		**Image at 2.5 m**			
		$R_{HSR} = 9,727$	$R_{HSR} = 13,115$	$R_{HSR} = 19,006$	$R_{HSR} = 24,897$
Image at 20 m	$R_{MSR} = 437$	0.7345	0.7421	0.7413	0.7384
	$R_{MSR} = 728$	0.7462	0.7513	0.7491	0.7433
	$R_{MSR} = 1,454$	0.7613	**0.7689**	0.7656	0.7598
	$R_{MSR} = 3,084$	0.7512	0.7603	0.7564	0.7519
Image at 10 m	$R_{MSR} = 1,515$	0.7232	0.7298	0.7224	0.7186
	$R_{MSR} = 2,482$	0.7326	**0.7359**	0.7351	0.7313
	$R_{MSR} = 4,980$	0.7263	0.7285	0.7287	0.7231
	$R_{MSR} = 7,863$	0.7114	0.7196	0.7167	0.7138

composition in terms of HSR clusters) (Fig. 9(h)). The best results have been obtained for $C_{HSR} = 22$ (Fig. 9(c, d, g)) and $C_{HSR} = 20$ (Fig. 9(b, f)).

- C_{inter}: If the number of intermediate clusters is too low ($C_{inter} = 9$), some final clusters do not match with the semantic classes (probably because these clusters gather too many different HSR objects) (Fig. 9(e)). With $C_{inter} = 15$, the number of intermediate clusters is probably too high. The clusters seem to be too "specialised" and do not match with the groundtruth map. The best results were obtained with $C_{inter} = 13$ for the MSR image at 20 m (Fig. 9(c)) and with $C_{inter} = 11$ for the MSR image at 10 m (Fig. 9(g)).

TOULOUSE dataset:
In the same way, experiments carried out on the TOULOUSE dataset have shown that the C_{HSR} parameter have a direct influence on the final clustering results. If C_{HSR} is too low, many clusters from the HSR classified image seem to be inappropriate to classify the scene to the urban objects level (probably due to the low number of clusters). For example, some Forest zones belong to the same cluster as housing surfaces (Fig. 10(a, e)). When the number of HSR clusters is too high ($C_{HSR} = 25$), the same problem as in the STRASBOURG dataset appears (Fig. 10(h)). The best results have been obtained for $C_{HSR} = 22$ (Fig. 10(c, d, g)) and $C_{HSR} = 20$ (Fig. 10(b, f)). Concerning the impact of the parameter C_{inter}, the same kinds of observations (as in the STRASBOURG dataset) can be made. When the number of intermediate clusters is too low ($C_{inter} = 9$), some of the final clus-

ters of the clustering results do not match with the semantic classes. For instance, the collective housing blocks are sometimes in the same cluster as the individual housing blocks (Fig. 10(a, e)). With $C_{inter} = 15$, the number of intermediate clusters is probably too high. The same problem as in the STRASBOURG dataset appears. The best results were obtained with $C_{inter} = 13$ for the MSR image at 20 m (Fig. 10(c)) and with $C_{inter} = 11$ for the MSR image at 10 m (Fig. 10(f)).

To summarise, this study has demonstrated that the relevance of the block level clustering results is directly associated to the choice of the classification parameters. Moreover, this last one has shown that the optimal classification parameters are quite similar for the two studied datasets. Indeed, the best classification parameters were found when the number of clusters in the HSR image took range between 20 and 22 and when the number of clusters in the intermediate clustering result was set to 13 with the MSR image at 20 m (see Fig. 9(c) and Fig. 10(c)) and set to 11 with the MSR image at 10 m (see Fig. 9(g) and Fig. 10(f)). With these configurations, urban areas could be classified at an intermediate level whose resulting clusters correctly match with those defined by experts for an analysis of the urban blocks. To conclude on this study, the best number of clusters in the HSR image can be found by the expert knowledge (by analysing the number of materials composing the urban objects which can appear in the studied area) whereas the optimal number of clusters in the intermediate clustering can not be discovered directly. The end-user has to try different values of this parameter. However, the set of possible suitable values for C_{inter} has been significantly reduced (5 values –from 11 to 15).

3.3.4 **Impact of the spatial resolution of the MSR image.** To study the influence of the spatial resolutions of MSR images on the final result, we have applied the method on the two datasets with different spatial configurations:

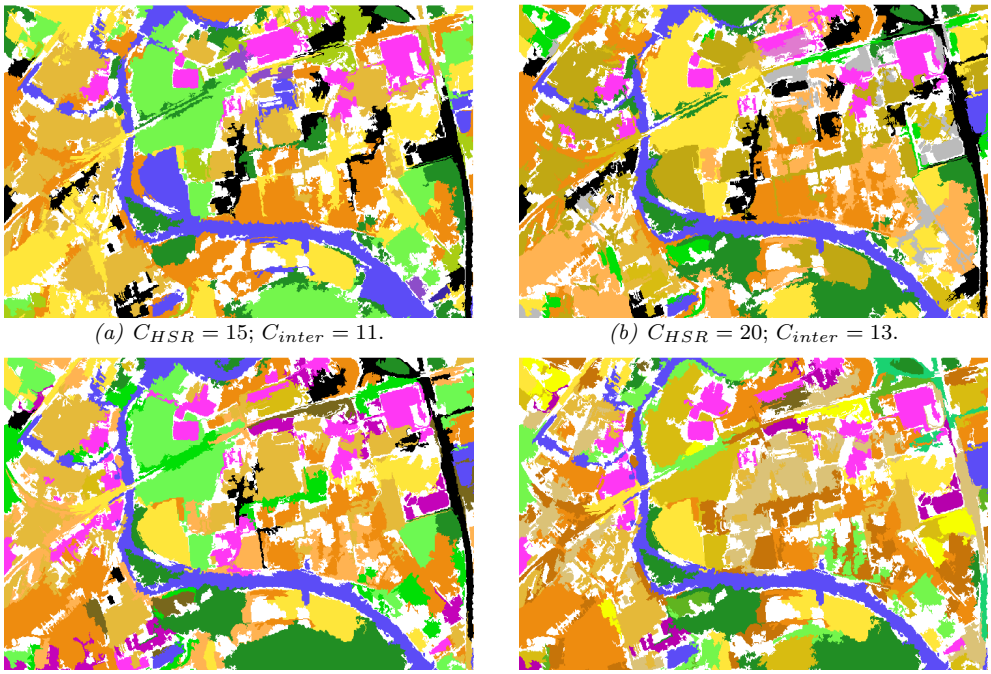
- (2.8 m/10 m) and (2.8 m/20 m) for the STRASBOURG dataset;
- (2.5 m/10 m) and (2.5 m/20 m) for the TOULOUSE dataset.

The Kappa index has been computed for each experiment with the two configurations (see Tables 3, 4) and the numerical results shown higher scores when the MSR image at 20 m is used. Indeed, most of the experiments carried out on the STRASBOURG dataset (resp. TOULOUSE dataset) have shown that the configuration (2.8 m/20 m) (resp. (2.5 m/20 m)) provides better results than the configuration (2.8 m/10 m) (resp. (2.5 m/10 m)).

Moreover, from a visual comparison between resulting cluster images (see Fig. 9(c, g) and Fig. 10(c, f)), we can see that final clusters obtained with the spatial configuration (2.8 m[or 2.5 m]/20 m) match with more semantic classes than those obtained with the spatial configuration (2.8 m[or 2.5 m]/10 m). The gap between these results is probably due to the possibility to extract the urban objects from the MSR image at 10 m (the urban blocks of the studied areas are too heterogeneous to be extracted from this image). The spatial resolutions of an HSR image at 2.8 m (or 2.5 m) and an MSR image at 10 m are too close for using our method to classify urban blocks on the studied datasets.

Concerning not embeddable regions (white clusters on Fig. 9 and 10 –label $\perp$ in Step 4 of the method in Subsection 2.5), experiments have shown that this problem is more perceptible when the spatial configuration (2.8 m[or 2.5 m]/10 m) is used. Indeed, experiments have shown that 22% of the regions of the HSR image are unclassifiable in the final result when the image at 10 m is used whereas only 15% of the regions of the HSR image are unclassifiable when the image at 20 m is used.

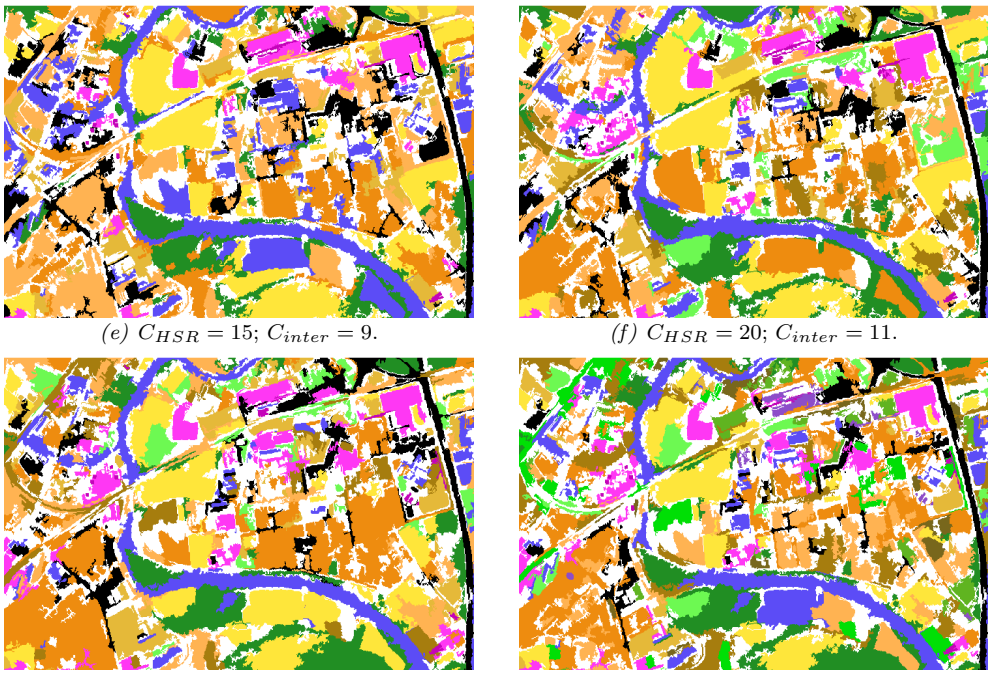
(HSR image at 2.8 m / MSR image at 20 m)



(a) $C_{HSR} = 15$; $C_{inter} = 11$. *(b)* $C_{HSR} = 20$; $C_{inter} = 13$.

(c) $C_{HSR} = 22$; $C_{inter} = 13$. *(d)* $C_{HSR} = 22$; $C_{inter} = 15$.

(HSR image at 2.8 m / MSR image at 10 m)

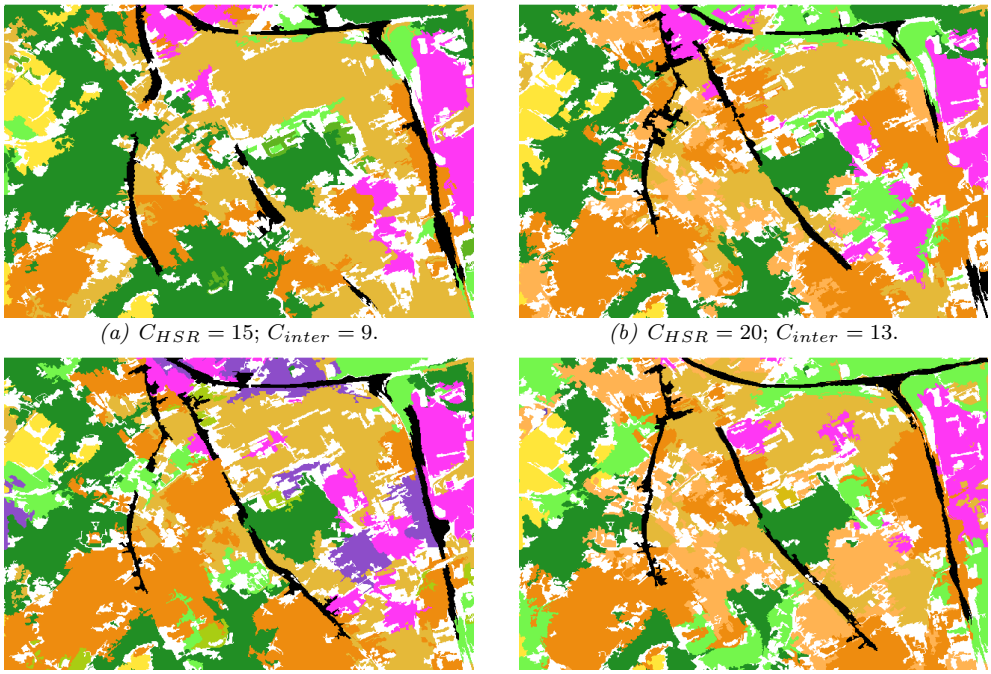


(e) $C_{HSR} = 15$; $C_{inter} = 9$. *(f)* $C_{HSR} = 20$; $C_{inter} = 11$.

(g) $C_{HSR} = 22$; $C_{inter} = 11$. *(h)* $C_{HSR} = 25$; $C_{inter} = 15$.

Figure 9. Block level classification results on the STRASBOURG dataset. Note that the white cluster gathers not embeddable regions (label $\perp$ – Step 4 of the method in Subsection 2.5).

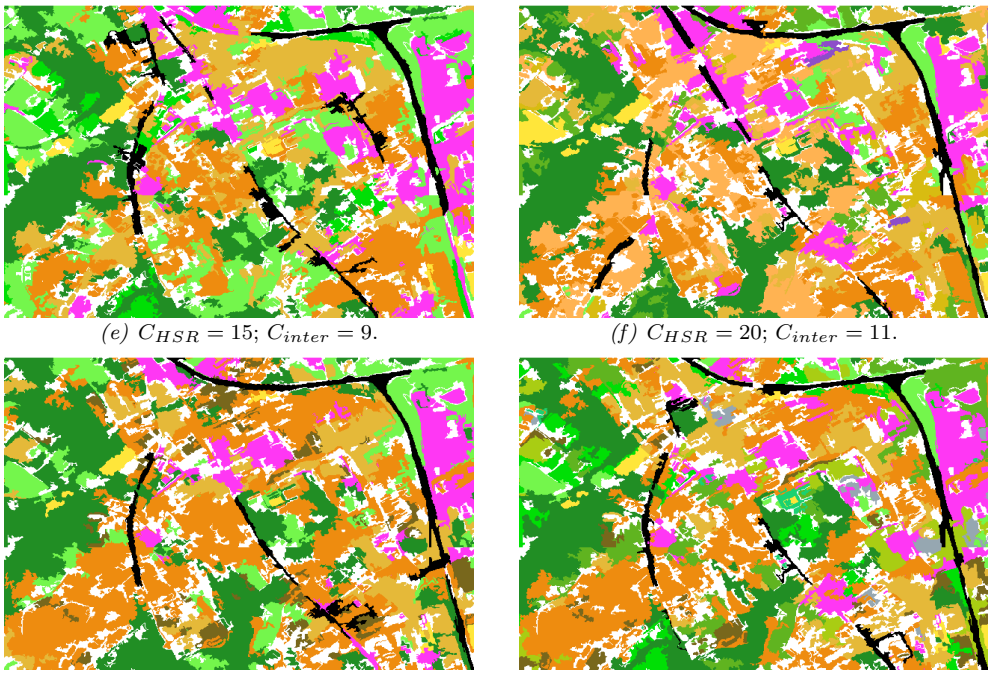
(HSR image at 2.5 m / MSR image at 20 m)



(a) $C_{HSR} = 15$; $C_{inter} = 9$. *(b)* $C_{HSR} = 20$; $C_{inter} = 13$.

(c) $C_{HSR} = 22$; $C_{inter} = 13$. *(d)* $C_{HSR} = 22$; $C_{inter} = 15$.

(HSR image at 2.5 m / MSR image at 10 m)



(e) $C_{HSR} = 15$; $C_{inter} = 9$. *(f)* $C_{HSR} = 20$; $C_{inter} = 11$.

(g) $C_{HSR} = 22$; $C_{inter} = 11$. *(h)* $C_{HSR} = 25$; $C_{inter} = 15$.

Figure 10. Block level classification results on the TOULOUSE dataset. Note that the white cluster gathers not embeddable regions (label $\perp$ – Step 4 of the method in Subsection 2.5).

Table 4. **Impact of the initial classifications** of HSR and MSR images on the final clustering result. The Kappa values for the different experiments are presented on the STRASBOURG and the TOULOUSE datasets. The best values obtained are depicted in red.

STRASBOURG dataset					
MSR Image \ **HSR Image**		**Image at 2.8 m**			
		$C_{HSR} = 15$	$C_{HSR} = 20$	$C_{HSR} = 22$	$C_{HSR} = 25$
Image at 20 m	$C_{inter} = 9$	0.7717	0.7825	0.7839	0.7648
	$C_{inter} = 11$	0.7921	0.7957	0.7972	0.7787
	$C_{inter} = 13$	0.8079	0.8080	**0.8203**	0.7913
	$C_{inter} = 15$	0.8047	0.7972	0.8171	0.7894
Image at 10 m	$C_{inter} = 9$	0.7813	0.8003	0.8018	0.7823
	$C_{inter} = 11$	0.7975	0.8094	**0.8105**	0.7897
	$C_{inter} = 13$	0.7924	0.7981	0.7969	0.7742
	$C_{inter} = 15$	0.7893	0.7904	0.7914	0.7698

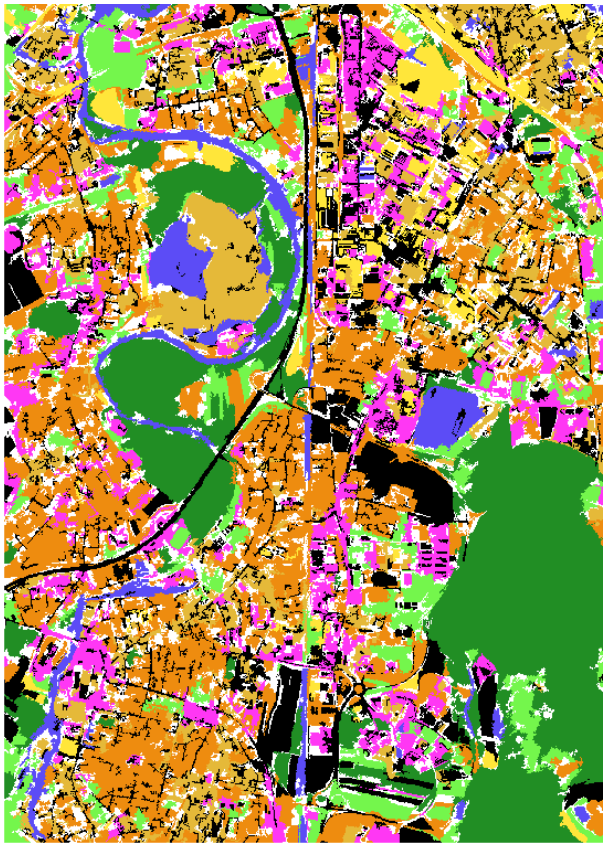
TOULOUSE dataset					
MSR Image \ **HSR Image**		**Image at 2.5 m**			
		$C_{HSR} = 15$	$C_{HSR} = 20$	$C_{HSR} = 22$	$C_{HSR} = 25$
Image at 20 m	$C_{inter} = 9$	0.7669	0.7687	0.7653	0.7591
	$C_{inter} = 11$	0.7794	0.7853	0.7797	0.7734
	$C_{inter} = 13$	0.7833	0.7847	**0.7859**	0.7820
	$C_{inter} = 15$	0.7699	0.7744	0.7638	0.7621
Image at 10 m	$C_{inter} = 9$	0.7520	0.7598	0.7564	0.7427
	$C_{inter} = 11$	0.7641	**0.7682**	0.7666	0.7612
	$C_{inter} = 13$	0.7638	0.7642	0.7613	0.7594
	$C_{inter} = 15$	0.7554	0.7569	0.7552	0.7489

This problem of miscorrespondence between HSR regions and MSR ones tends to increase when the two input images have spatial resolutions that become too close. Broadly speaking, larger are MSR regions, easier is to embed HSR regions into MSR ones.

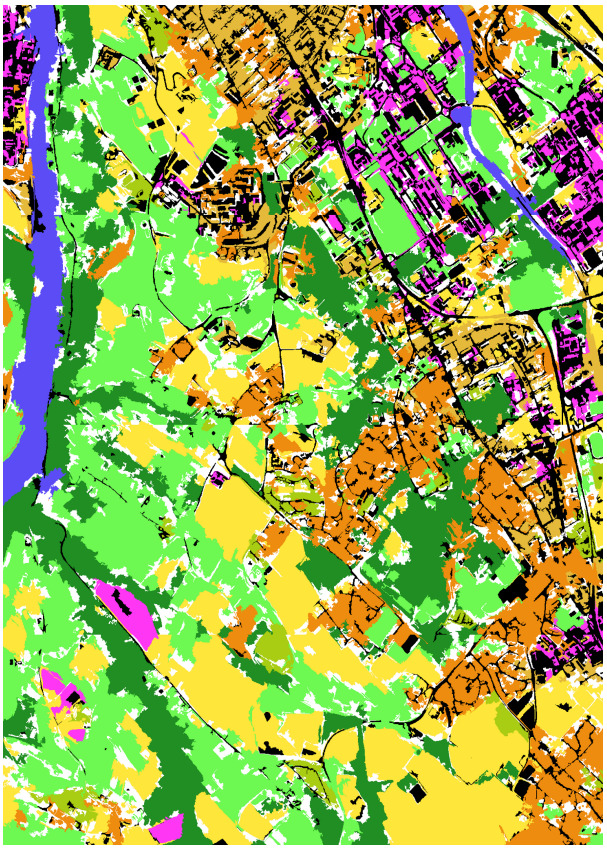
3.3.5 **Post-processing step.** As introduced previously, a post-processing step consisting in a hierarchical ascendant clustering algorithm (using the Euclidean distance (Cha and Srihari 2002)) was applied on the final result to reduce the number of clusters (from 13 (or 11) to 9 clusters).

The visual comparisons between the results obtained and the groundtruth maps show that urban classes (industrial blocks and housing blocks) are well identified. More especially, the distinction between individual housing and collective blocks is visible even if some confusions can sometimes appear for instance with the water surface. This is due to the reflection of this surface in the HSR image (in white in Fig. 6(c)). For industrial blocks, the extraction is reliable for small industrial surfaces. However, for large surfaces (in the North and in the South of the STRASBOURG dataset images), industrial blocks are not recognised due to their heterogeneity. Only some surfaces with a high reflectance are identified and other ones are confused with agricultural zones due to their similar reflectance. Other surfaces such as the forest and urban vegetation are well identified with few confusions.

The road class (in black in Fig. 9 and 10) is not directly extracted by the mul-

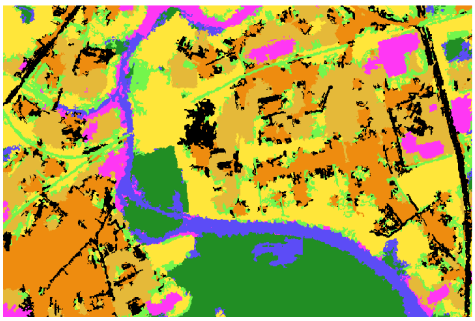


(a) Block level clustering result on the STRASBOURG dataset.



(b) Block level clustering result on the TOULOUSE dataset.

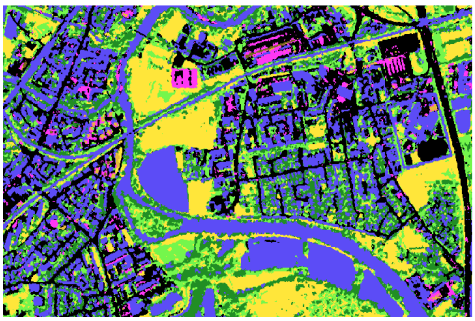
Figure 11. Results of block level clustering after post-processing step. These results were carried out on the configurations (2.8 m/20 m) for the STRASBOURG dataset and (2.5 m/20 m) for the TOULOUSE dataset.



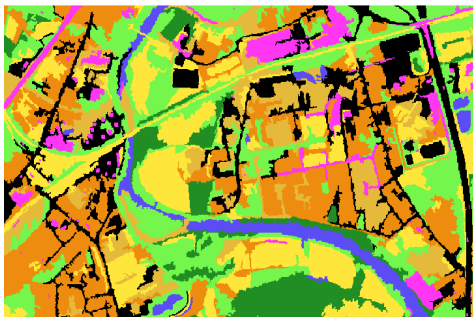
(a) Multiresolution Pixel Method (MPM) carried out on (MSR (20 m) ⊛ HSR (2.8 m)).

(b) K-MEANS carried out on (MSR (20 m) ⊕ HSR (2.8 m)) combined image.

Figure 12. Per-pixel clustering results from different methods with 8 clusters on the STRASBOURG dataset.



(a) K-MEANS directly carried out on HSR regions image.



(b) Definiens Professional on HSR image.

Figure 13. Regions clustering results from different methods with 8 clusters on the STRASBOURG dataset

tiresolution method since roads are often enclosed in urban blocks and then do not create a single intermediate cluster. In fact, the cluster corresponding to roads is divided into all the histograms. As the road cluster appears very well on the initial HSR clustering, due to their high elongation, this cluster has been integrated in the final clustering. The addition of this class enables to improve the comparison with the groundtruth. Note that this class contains both roads cluster and agricultural areas due to the similar reflectance on the HSR image.

This post-processing step was applied on the best results of the STRASBOURG and TOULOUSE datasets with the spatial configuration (2.8 m[or 2.5 m]/20 m). Results are presented on the whole images (Fig. 11(a, b)). In these two examples, 13 clusters were merged into 8 clusters by the post-processing step. Road clusters (in black) were extracted from initial HSR clustering and superimposed for a better visualisation. Note that the white cluster gathers not embeddable regions (label $\perp$ – Step 4 of the method in Subsection 2.5).

3.3.6 **Comparative study.** The proposed method has also been assessed by comparisons with other per-pixel/region-based approaches:

- **Per-pixel clustering**: Previous results were obtained by using the Multiresolution Pixel Method (MPM) described in (Wemmert *et al.* 2009) carried out on the HSR (at 2.8 m) and MSR (at 10 m and 20 m) images of the STRASBOURG and TOULOUSE datasets with 8 clusters. Figure 12(a) presents a result obtained by using the MPM method on the HSR (at 2.8 m) and MSR (at 20 m) images of the STRASBOURG dataset with 8 clusters. Note that the symbol (MSR ⊛ HSR) means that the method was applied simultaneously on the MSR and HSR images. Others results were obtained with the K-MEANS

Table 5. Kappa values of the different experiments with 8 clusters. The symbol (MSR ⊛ HSR) means that the method was applied simultaneously on the MSR and HSR images whereas the symbol (MSR ⊕ HSR) means that the method was applied on a fusion of the MSR and HSR images.

			Dataset	
	Method	**Configuration**	STRASBOURG	TOULOUSE
Pixels	MPM	MSR (10 m) ⊛ HSR	0.7617	0.7312
	MPM	MSR (20 m) ⊛ HSR	0.7826	0.7433
	K-MEANS	MSR (10 m) ⊕ HSR	0.7201	0.7094
	K-MEANS	MSR (20 m) ⊕ HSR	0.7219	0.7123
	K-MEANS	HSR	0.7021	0.6957
Regions	MRM	MSR (10 m) ⊛ HSR	**0.8105**	**0.7682**
	MRM	MSR (20 m) ⊛ HSR	**0.8203**	**0.7859**
	K-MEANS	HSR	0.7445	0.7215
	Definiens	HSR	0.7325	0.7183

algorithm carried out on an image built by direct fusion of the two images (to each pixel is associated all the radiometric information from HSR and MSR images) with 8 clusters. This method was tested for the two datasets. Figure 12(b) shows a result obtained by using this method on the fusion of the HSR and MSR (at 20 m) images of the STRASBOURG dataset. Note that the symbol (MSR ⊕ HSR) means that the method was applied on a fusion of the MSR and HSR images. The last "pixel" results were obtained by applying the K-MEANS algorithm directly on the HSR images of the two datasets with 8 clusters.

- **Region-based clustering**: Some results were obtained by applying the K-MEANS algorithm with 8 clusters on the HSR images of the two datasets at the object level using only spectral information – the objects were initially created by a clustering with 15 clusters. Figure 13(a) presents a result obtained by this method on the HSR image of the STRASBOURG dataset. Others results were produced by an object-oriented approach (Benz *et al.* 2004) with Definiens Professional software, using only spectral information. Figure 13(b) shows a result produced by applying this method on the HSR image of Strasbourg.

Table 5 presents the Kappa values for these experiments and shows that the results obtained with the proposed Multiresolution Region-based Method (MRM) are better –with an adequate choice of parameters– than those obtained with the Multiresolution Pixel-based Method (MPM) and outperforms the other ones (region-based and per-pixel ones).

A visual comparison between the results depicted in Figures 12 and 13 and those obtained with the proposed method tends to confirm the quantitative results of Table 5. For instance, they are fewer confusions between (i) water surfaces and industrial blocks (Fig. 12(a)) or roads and industrial blocks (Fig. 12(b)) and, (ii) agricultural areas and urban blocks (Fig. 12(a)). Moreover, compared to Figure 12(b) some clusters are more homogeneous (urban blocks, forest surfaces).

3.3.7 **Discussion about parametrisation.** The multiresolution method proposed in this article provides relevant results on the classification of urban blocks. However, some parameters are needed to effectively run the proposed algorithm. They can be classified into two categories: parameters which are devoted to "remain" parameters, since they naturally depend on the expert knowledge, and parameters which could/should be (as much as possible) automatically determined.

The first parameter which could be easily determined by the expert is the consid-

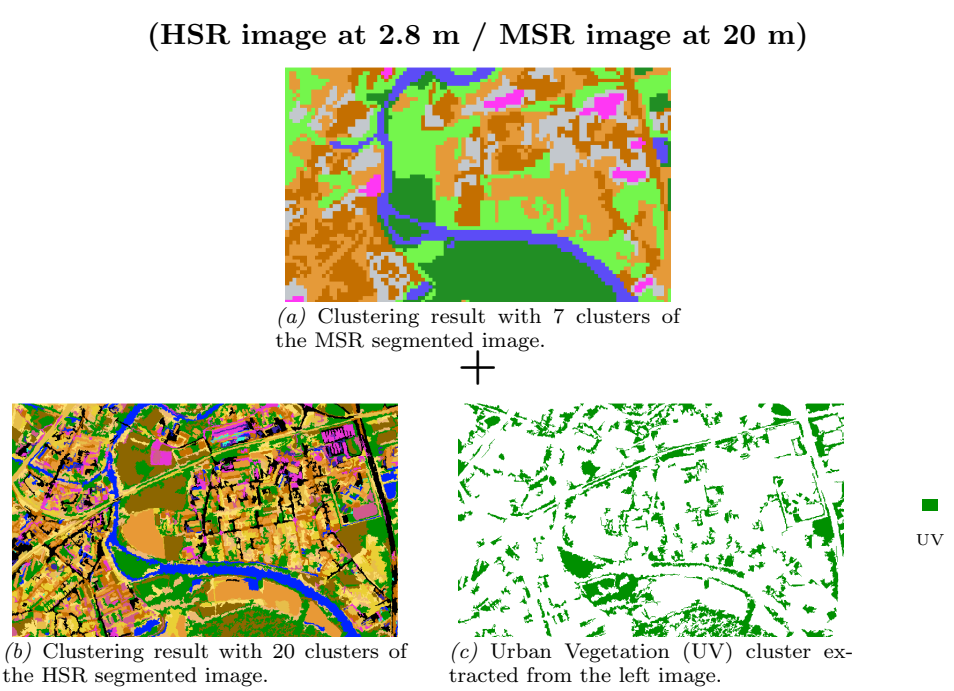
ered spatial configuration. Indeed, several configurations of HSR and MSR images could be used to classify urban blocks and to apply the cluster refinement method. Experiments have shown that HSR image at 2.8 m or at 2.5 m could be combined with MSR image at 20 m or at 10 m to obtain suitable clustering results. The choice of the best configuration depends on the studied area (Subsection 3.3.4). If the urban blocks are wide (composed by several houses), results will be of higher quality with a configuration using an MSR image at 20 m whereas if the urban blocks are narrow, an MSR image at 10 m could be more adapted.

The segmentation parameters (the different values of τ), which impact on the level of details in the resulting partitions of the HSR and MSR images, have to be defined manually by the end-user. For the HSR image (resp. MSR image), the expert has to choose parameter values to obtain partitioning results which match with the urban objects (resp. the urban blocks) of the studied ground scene. However, the studies carried out in this article could help the experts to find the optimal segmentation parameters. Indeed, the range of the possible suitable values for τ has been significantly reduced (Subsection 3.3.2). Moreover, some possible solutions to lead to the automation of this process are, in particular, discussed in the Section 4.

Concerning the number of classes in the HSR segmented images, the classification parameter C_{HSR} could be determined by the expert knowledge. Indeed, for this kind of images the number of clusters depends on the materials of the urban objects which can appear in the studied area (see Table 1, left column). For the classification parameter C_{inter} (which impacts on the number of clusters used in the MSR segmented image to classify urban blocks), experiments have shown the difficulty of finding best parametric values. Indeed, the choice of this parameter depends on the studied data (see Table 4). To obtain the best urban blocks clustering results with the proposed method, end-users have to try different values of this parameter. However, the studies carried out in this article could help the experts to find the optimal classification parameters (Subsection 3.3.3). As for the segmentation parameter τ, the set of the possible suitable values for C_{inter} has been significantly reduced (5 values –from 11 to 15). In further works, we plan to automate this process by running iteratively the proposed method with different values of this parameter. After each iteration, the clustering result could be evaluated by using some urban blocks examples provided by the expert (extracted manually from the studied images).

3.4 *Results analysis of cluster refinement*

To validate cluster refinement method, we have decided to study the refinement of the HSR vegetation cluster by applying the method on the Strasbourg dataset (with the HSR image at 2.8 m and the MSR image at 20 m). To this end, the two images were segmented; the partitions used for these images were those created for the block-level clustering method ($C_{HSR} = 19,752$ and $C_{MSR} = 1,688$, see Subsection 3.3.2). Then, each segmented image was classified in its own semantic level using the K-means algorithm with a variable number of clusters; in order to study the influence of the number of clusters in the HSR and the MSR segmented images (denoted by C_{HSR} and C_{MSR}), we have carried out experiments with 15, 17 and 20 clusters for the HSR image and with 5, 7 and 9 clusters for the MSR image. These values correspond to those used to classify an HSR region image (resp. an MSR image) to the urban object level (resp. the urban area level) (see Table 1, left and right columns). The next step was to select (on the clustering results of the HSR segmented image) the clusters corresponding to the vegetation class. Finally,



(a) Clustering result with 7 clusters of the MSR segmented image.

(b) Clustering result with 20 clusters of the HSR segmented image.

(c) Urban Vegetation (UV) cluster extracted from the left image.

(d) Clustering result of the refinement method on the HSR segmented image.

(e) Urban Vegetation clusters (UV1, UV2, UV3) extracted from the left image.

Figure 14. Refinement method carried out on the STRASBOURG dataset with the configuration (2.8 m/20 m). The "Urban Vegetation" cluster (UV) extracted from the HSR image was split into 3 new HSR "Urban Vegetation" clusters (UV1, UV2, UV3).

the last step was to split the HSR vegetation clusters by using the clustering results of the MSR segmented images.

Due to the unavailability of a groundtruth map for this kind of land cover, all the results were assessed visually by the experts. Experiments have shown that:

- When the number of HSR clusters is too high ($C_{HSR} = 23$), the HSR vegetation resulting clusters are too small to be split in subclusters matching new concepts (probably due to a too small number of regions in each cluster). However, if C_{HSR} is too low ($C_{HSR} = 15$), it becomes difficult to extract an HSR cluster corresponding to the vegetation since this cluster generally does not exist (regions corresponding to pieces of vegetation could be in the same cluster as regions corresponding to others kinds of land cover).
- When the number of MSR clusters is too high ($C_{MSR} = 15$), the refinement method splits the original HSR vegetation cluster into a high number of vegetation subclusters. These subclusters are often too numerous. Most of them are too specialised to be correctly interpreted by the expert. If the number of MSR clusters is too low ($C_{MSR} = 5$), the resulting HSR subclusters do not match with any land cover classes. Indeed, if C_{MSR} is too low the cluster-

ing results of the MSR segmented image are formed by clusters which gather different kinds of region (vegetation, roof, etc.).

- Best results were found when the MSR segmented image was classified with 7 clusters and the HSR segmented image was classified with 20 clusters (see Fig. 14). In this experiment the HSR urban vegetation cluster (UV, depicted in dark green in the Fig. 14(b, c)) was split into three subclusters (UV1, UV2, UV3, depicted in three shades of green in the Fig. 14(d, e)). These new clusters were categorised by the experts as follows: (UV1) Forest cluster, (UV2) Homogeneous urban vegetation cluster –for instance, large grass areas in collective housing blocks– and (UV3) Heterogeneous urban vegetation cluster –for instance, the gardens in the individual housing blocks.

Experiments carried out have shown the efficiency of this new approach which actually extracts new kinds of hidden information. The HSR cluster refinement method permits to improve the result of HSR regions classification. The new clusters discovered by this method could not be discovered directly by only using the radiometric values of the pixels of the HSR image. Only a simultaneous multiresolution analysis of HSR and MSR images could help the end-users to extract hidden clusters corresponding to sub-thematic land covers classes.

4. Conclusion and perspectives

The multiresolution method proposed in this article enables to generate clusterings at a block level, by taking advantage of information provided by images at higher and lower resolutions. Moreover this methodology combines the possibilities offered by the (per-pixel) multi-image analysis and the efficiency of the (mono-image) region-based frameworks for a *block level* analysis in the context of the mapping of urban areas.

The proposed method is slightly inspired from previous works by some of the authors (Wemmert *et al.* 2009), but it differs from them since it is no longer based on a per-pixel approach, but on a region-based one (in this method, the spatial context of the urban objects and the semantic relationships of these last ones between the available resolutions are used to enhance the simultaneous analysis of both MSR and HSR images). Moreover, it is based on a quite different way than (Wemmert *et al.* 2009) to perform multiresolution analysis. Indeed, in the previous approach the analysis was carried out by studying the composition of the highest resolution data in terms of clusters in the lowest resolution one, while in the method presented in this article, the opposite strategy aims at studying the composition of the lowest resolution regions in terms of clusters in the image at the highest resolution. This new way to perform multiresolution analysis enables, in particular, the refinement of final HSR clusters into more specific subclusters matching with hidden land cover classes.

Experimental results tend to prove that this new framework actually generates better results that the former one, and also outperform classical (K-MEANS) approaches, as illustrated by quantitative and qualitative validations in the context of urban area analysis. In particular, it has been used to help experts to obtain information from MSR and HSR satellite images for a land cover mapping at 1:10,000, thus enabling to discover new knowledge from these kinds of data. Moreover, experiments have shown that the images could arise from different satellites (in the STRASBOURG dataset the HSR image was acquired by QUICKBIRD whereas the MSR ones were acquired by SPOT-4) or from a single sensor (in the TOULOUSE dataset the MSR images were degraded from a SPOT-5 HSR image).

These results are promising and then justify further developments, which may lead to improvements of the method. The main issue, to be considered in priority, is to automate the method by determining iteratively the most suitable values of parameters. Indeed, the method actually requires the setting of some parameters (related to the segmentation and clustering). While some of them can hardly be set automatically (*e.g.*, the number of classes, which depends on the kind of applications), other ones could be computed, for instance, the ones (related to the heterogeneity function f) enabling to stop the segmentation process. These improvements will be devoted to make the current method more ergonomic for their potential users.

In the long term, some supplementary improvements will also be considered. Since it can be important to obtain complete clustering results, some solutions will be proposed to deal with this issue, namely the non-hierarchical structure between the MSR and HSR segmentations. Some possible ways to tackle this problem could be: (1) to produce a segmentation of the HSR image as a sub-partition of the MSR one, or (2) to develop an iterative segmentation/classification process using all the available multiresolution information in order to progressively improve the segmentation results thanks to the classification ones. Moreover, it is planned to integrate the method into the SAMARAH framework of collaborative clustering (Forestier *et al.* 2008) in order to extend the method to make it able to simultaneously deal with more than two images.

References

AKCAY, H. and AKSOY, S., 2008, Automatic Detection of Geospatial Objects Using Multiple Hierarchical Segmentations. *IEEE Transactions on Geoscience and Remote Sensing*, **46**, pp. 2097–2111.

BAATZ, M., HOFFMANN, C. and WILLHAUCK, G., 2008, Progressing from object-based to object-oriented image analysis. In *Object-Based Image Analysis*, pp. 29–42 (Thomas Blaschke and Stefan Lang and Geoffrey J. Hay).

BAATZ, M. and SCHAPE, A., 2000, Multiresolution segmentation–An optimization approach for high quality multi-scale image segmentation. In *Proceedings of the Angewandte Geographische Informationsverarbeitung Symposium*, pp. 12–23.

BARNSLEY, M. and BARR, S., 1997, Distinguishing urban land-use categories in fine spatial resolution land-cover data using a graph-based, structural pattern recognition system. *Computers, Environment and Urban Systems*, **21**, pp. 209 – 225 Remote Sensing of Urban Systems.

BELLMAN, R., 1961, *Adaptive Control Processes* (Princeton University Press).

BENZ, U., HOFMANN, P., WILLHAUCK, G., LINGENFELDER, I. and HEYNEN, M., 2004, Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, **58**, pp. 239–258.

BLASCHKE, T., 2010, Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, **65**, pp. 2 – 16.

BRUZZONE, L. and CARLIN, V., 2006, A multilevel context-based system for classification of very high spatial resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, pp. 2587–2600.

CAPRIOLI, M. and TARANTINO, E., 2003, Urban features recognition from VHR satellite data with an object-oriented approach. In *Proceedings of the Commission IV Joint Workshop, Challenges in Geospatial Analysis, Integration and Visualization II*, pp. 8–9.

CARLEER, A., DEBEIR, O. and WOLFF, E., 2005, Assessment of very high spa-

tial resolution satellite image segmentations. *Photogrammetric Engineering and Remote Sensing*, **71**, pp. 1285–1294.

Carleer, A. and Wolff, E., 2006, Urban land cover multilevel region-based classification of VHR data by selecting relevant features. *International Journal of Remote Sensing*, **27**, pp. 1035–1051.

Cha, S. and Srihari, S., 2002, On measuring the distance between histograms. *Pattern Recognition*, **35**, pp. 1355–1370.

Chang, Y.L., Liang, L.S., Han, C.C., Fang, J.P., Liang, W.Y. and Chen, K.S., 2007, Multisource data fusion for landslide classification using generalized positive Boolean functions. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, pp. 1697–1708.

Chibani, Y., 2005, Selective synthetic aperture radar and panchromatic image fusion by using the à trous wavelet decomposition. *EURASIP Journal on Applied Signal Processing*, **2005**, pp. 2207–2214.

Congalton, R., 1991, A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, **37**, pp. 35–46.

Corbane, C., Raclot, D., Jacob, F., Albergel, J. and Andrieux, P., 2008, Remote sensing of soil surface characteristics from a multiscale classification approach. *CATENA*, **75**, pp. 308 – 318.

Cross, A., Mason, D. and Dury, S., 1988, Segmentation of remotely-sensed images by a split-and-merge process. *International Journal of Remote Sensing*, **9**, pp. 1329–1345.

Forestier, G., Wemmert, C. and Gançarski, P., 2008, Multi-source images analysis using collaborative clustering. *EURASIP Journal on Advances in Signal Processing*, **2008**, pp. 1–11.

Guo, D., Xiong, H. and Atluri, V., 2008, Object discovery in high-resolution remote sensing images: a semantic perspective. *Knowledge and Information Systems*, **19**, pp. 211–233.

Herold, M., Liu, X. and Clarke, K., 2003, Spatial metrics and image texture for mapping urban land use. *Photogrammetric Engineering and Remote Sensing*, **69**, pp. 991–1001.

Herold, M., Scepan, J., Muller, A. and Gunter, S., 2002, Object-oriented Mapping and Analysis of Urban Landuse/Cover using Ikonos Data. In *Proceedings of the Earsel Symposium Geoinformation for European-Wide Integration*, pp. 531–538.

Hughes, G., 1968, On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, **14**, pp. 55–63.

Jacquin, A., Misakova, L. and Gay, M., 2008, A hybrid object-based classification approach for mapping urban sprawl in periurban environment. *Landscape and Urban Planning*, **84**, pp. 152 – 165.

Marangoz, A., Oruc, M. and Buyuksalih, G., 2004, Object-oriented image analysis and semantic network for extracting the roads and buildings from ikonos pan-sharpened images. In *Proceedings of the ISRPS Annual Conference*, pp. 19–23.

Meinel, G. and Neubert, M., 2004, A comparison of segmentation programs for high resolution remote sensing data. *International Archives of Photogrammetry and Remote Sensing*, **XXXV**, pp. 1097–1105.

Mesev, V., Gorte, B. and Longley, P., 2000, Modified maximum-likelihood classification algorithms and their application to urban remote sensing. In *Remote Sensing and Urban Analysis*, Vol. 1, pp. 71–89 (Taylor & Francis).

Smeulders, A., Worring, M., Santini, S., Gupta, A. and Jain, R., 2000, Content-based Image Retrieval at the End of the Early Years. *IEEE Transactions*

on Pattern Analysis and Machine Intelligence, **22**, pp. 1349–1380.

Vincent, L. and Soille, P., 1991, Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, pp. 583–598.

Wemmert, C., Puissant, A., Forestier, G. and Gançarski, P., 2009, Multiresolution Remote Sensing Image Clustering. *IEEE Geoscience and Remote Sensing Letters*, **6**, pp. 533–537.