A Multiple-Objects Recognition Method Based on Region Similarity Measures: Application to Roof Extraction from Orthophotoplans

Abdellatif El Idrissi*, Youssef El Merabet†, Yassine Ruichek‡, Raja Touahni †, Abderrahmane Sbihi*, Cyril Meurie§ and Ahmed Moussa*

*Communication and Information Technologies Laboratory,Abdelmalek Essaadi University, 90000 Tangier, Morocco
†LASTID Laboratory, Department of Physics, Faculty of Sciences, Ibn Tofail University, B.P 133, 14000 Kenitra, Morocco.
‡IRTES-SeT Laboratory, Technology University of Belfort-Montbeliard, 13 Ernest Thierry-Mieg Street, 90010, Belfort,France
§Univ Lille Nord de France, F-59000 Lille, IFSTTAR, LEOST, F59650, Villeneuve d'Ascq,France.

***Abstract*—In this paper, an efficient method for automatic and accurate detection of multiple objects from images using a region similarity measure is presented. This method involves the construction of two knowledge databases: The first one contains several distinctive textures of objects to be extracted. The second one is composed with textures representing background. Both databases are provided by some examples (training set) of images from which one wants to recognize objects. The proposed procedure starts by an initialization step during which the studied image is segmented into homogeneous regions. In order to separate the objects of interest from the image background, an evaluation of the similarity between the regions of the segmented image and those of the constructed knowledge databases is then performed. The proposed approach presents several advantages in terms of applicability, suitability and simplicity. Experimental results obtained from the method applied to extract building roofs from orthophotoplans prove its robustness and performance over popular methods like K Nearest Neighbours (KNN) and Support Vector Machine (SVM).**

***Keywords*—Object recognition; Region Similarity Measure; Texture; Feature extraction; Orthophotoplans**

I. Introduction

Nowadays, automatic object recognition has become a topic of growing interest for computer vision community. For instance, automatic extraction of man-made objects such as buildings and roads in urban areas has gained significant attention for photogrammetric researchers community over the last decade. This problem is usually considered when we talk about high-level image processing in order to produce numerical or symbolic information [1], [2]. In this context, several approaches have been proposed in the literature. First, one can cite interactive methods that need user interaction in order to extract desired targets or objects of interest from images. Generally, this category of methods has been introduced to alleviate the problems inherent to fully automatic segmentation which seems to never be perfect. These methods endeavour to divide an image into two segments: "object" and "background". The interactivity consists in imposing certain hard constraints for segmentation by pointing out certain pixels (seeds) that absolutely have to be part of the object and certain pixels that have to be part of the background.

Boykov and Jolly proposed an interactive graph cuts (IGC) for interactive image segmentation [3]. The segmentation is performed by the min-cut/max-flow algorithm. User scribbles extract color information that will be used thereafter as hard constraints. Rother et al. in [4] presented an iterative algorithm called GrabCut by simplifying user interaction. Their method combines image segmentation using graph cut and Gaussian Mixture Models (GMMs) based statistical models. A very useful segmentation benchmark, with a platform implementing important algorithms, has recently been proposed by McGuinness and Connor [5]. The authors compared many algorithms such as IGC [3], seeded region growing (SRG) [6], simple interactive object extraction (SIOX) [7] and binary partition tree (BPT) [8], in order to provide a good coverage of the various techniques currently available for foreground extraction, as stated in [5].

The SIOX [7] algorithm is also based on color information and has recently been integrated into the popular imaging program GIMP as "Foreground Selection Tool". The BPT [8] algorithm is based on hierarchical region segmentation, exploiting user interaction to split and merge regions in the tree. Bai and Sapiro [9] proposed a method based on fast kernel density estimation [10] for color statistics, improving geodesic distance-based approach described in [11].

Ning et al. [12] have recently proposed a novel maximal similarity based region merging (MSRM) mechanism for interactive image segmentation. The key idea of MSRM is to perform region merging between adjacent regions by exploiting an effective representation of color statistics basing on (quantized) color histograms computed from the regions. First, the input image is segmented using the mean shift segmentation algorithm. User must then indicate the location and region of the object to be extracted and background by using strokes as markers. Finally, a maximal-similarity based region merging mechanism is used in order to separate the object of interest from the background image while relying on

the help of markers introduced by the user. A similar algorithm also based on maximal similarity based region merging has been proposed in [13]. The difference is that this algorithm considers regions as seeds and takes the regions as growth units for region growing (i.e. merging of adjacent regions).

These methods give generally good results, but depending on the degree of user interaction. Thus, they become not suitable for images with high resolution containing too many objects of interest such as aerial and satellite images.

To address these issues, another category of methods, namely semi-automatic or automatic methods was developed. These methods are not only devoted to be applied on aerial or satellite images, but also on any kind of images ranging from simple single intensity images and color images, to laser and stereo images. A considerable number of methods from this category first tend to inspire from techniques introduced in pattern recognition and machine learning domains. In [14], Tso and Mather reported some classification methods used in remote sensing and which are pixel-based approaches like K Nearest Neighbours (KNN), maximum likelihood method and Support Vector Machines (SVM). Several variants or methods were developed for improving SVM method. Mountrakis and al. in [15] wrote a review of methods based on SVM in remote sensing field. They highlighted that SVM based methods are particularly considered in the remote sensing field due to their ability to generalize well even with limited training samples. It took place that SVM still outperformed best odd neural networks [15]. In the papers [16], [17], the authors show that neural networks can also be used for object recognition. Kinnunen and al. presented in [18] a method based on self-organization to deal with unsupervised object discovery. It is based on similar techniques that use bag of features approach and clustering to automatically classify image data. In their method, they replace clustering step by a self-organizing map.

Some authors tried to combine different methods. For instance, the authors of [19] combine KNN and SVM. Another method is what has been proposed by [19] using KNN, SVM and Geometric Moment Invariants (GMI). Introduced by Hu M.K. in [20] and used in several methods as in [21], [22], [23], GMI has been chosen to extract image features like rotation, scale and translation (RST)-invariant. Mathematical morphology has been also used to detect objects of interest. Soille and Pesaresi in [24], [25] developed a method to extract roads. It consists in two stages: a pre-processing one, in order to remove noise from the image, and a processing phase in which a structuring element is defined according to the shape of the object to be extracted. Roughly similar to ours, a recent method is presented by Ahmadi et al. in [26]. The authors adapted the active contour or Snakes model, originally introduced by Kass et al. [27], to automatically extract urban building boundaries. For that, the knowledge about the buildings is incorporated by the user into the system by introducing some pixel values of points inside building boundaries as training data. The system can then make a difference between buildings and background in the image.

Another class of methods consists in joining to classification algorithms prior information like height data or Light Detection and Ranging (LIDAR) data to detect objects of interest. Examples of this class of methods are the works of Halla and Brenner in [28], and Zhao and Trinder [29] who utilized height data and morphological operators for buildings extraction. Following this idea, Samadzadegan et al. proposed a novel approach for objects recognition, based on neuro-fuzzy modeling. They extract structural, textural and spectral information and integrate them in a fuzzy reasoning process to which learning capability of neural networks is introduced [30]. Zimmermann et al. produced Digital Surface Model (DSM) data from stereo images. In this model, multiple cues, colour segmentation, edge detection, texture segmentation and blob detection are combined. They then used the model to detect building roofs using slope and aspect operators [31]. Miliaresis and Kokkas developed in [32] a method for extracting a class of buildings using Digital Elevation Models (DEMs) derived from Lidar data. The method is based on geomorphometric segmentation principles with k-means to obtain a set of clusters formed by background and foreground objects represented on the basis of elevation and slope. Lafarge et al. presented in [33] an automatic building extraction method from DEM based on an object approach. They start by applying marked point process tools to realize a rough approximation of building footprints, which are then regularized by improving connection between neighboring elements and detecting roof height discontinuities.

Methods based on the notion of interest points that allow representing characteristics of targeted objects should also be mentioned here [34], [35], [36]. In [34], Lowe proposed to extract distinctive invariant features from images and used them as key points for matching different views of a sought object using a fast nearest-neighbour algorithm. Similarly, in [36], Berg et al. proposed an algorithm dealing with the problem of deformable shape matching by defining a cost function that measures similarity of corresponding geometric blur point descriptor and geometric distortion between corresponding feature points. Recently, in [37], Liu et al. reported a series of other methods that use the same idea (key points based) on discriminative parts. This means that an object may be represented by local parts which allow to distinguish it from others. These methods can be divided into two classes, the class with methods that integrate selection of discriminative parts with model construction, and the class with methods that separate the two processes [38], [39].

Regarding the related literature, a large number of the above methods present several shortcomings. The methods of the first category require numerous initializations and manual interaction which is very time-consuming when there are many object instances. Methods from the second category are most of the time context-dependent and are sensitive to noise. In the third category, and as stated in [26], the mentioned investigations have stressed to introduce height data in the context of aerial or satellite images to automatically extract buildings. This leads to high computational efforts and makes the approach requiring significant technological resources for data production and processing.

In this work, we propose a new method which is simple but copes with those drawbacks and robustly extract objects of interest. The relevance of the proposed technique could be expressed through the following advantages. First, the method allows automatic extraction of objects of interest and performs without any user interaction. Second, by this method, it becomes possible to accurately detect multiple objects in the same time from a given image. Third, one can achieve robust

results under various real-world conditions when considering for example, complex images for which both foreground and background regions have similar colors. Fourth, The method does not require height data or any prior information to recognize the difference between buildings and other background objects. Fifth, the method can be applied in several fields like medical image processing (e.g cancer cell recognition) and remote sensing image processing (e.g vegetation and buildings detection). In this paper, we are especially interested in extracting building roofs from orthophotoplans.

The remainder of the paper is organized as follows. In section II, we explain the proposed methodology and describe its main steps in details. Experiments and both qualitative and quantitative evaluations are presented in section III. A comparative analysis with other methods is also reported in this section. Section IV concludes the paper and addresses future works with the aim of enhancing the performance of the proposed method.

II. Proposed method

A. General description

The proposed method incorporates two major stages: off-line and on-line stages. In the off-line stage, two knowledge databases must be created in order to robustly avoid user interaction. The first one contains representative and distinctive textures of objects to be extracted. The other database is composed of textures picked up from objects that represent background in the image. As an illustrative example, for the application of building roof extraction from aerial images, the first knowledge database $\mathbf{B_{obj}}$ will be constructed with m distinctive textures of building roofs while the second one $\mathbf{B_{back}}$ will be constructed with n distinctive textures of other objects such as vegetation, road, forest, etc. These two databases are provided considering some examples of images. Having these two knowledge databases $\mathbf{B_{obj}}$ and $\mathbf{B_{back}}$ as reference, it is possible to automatically extract building roofs from any aerial image (orthophotoplan, in this study case). Figure 1 illustrates an example of knowledge databases used in this work. It should be noted here that more specific details either on used data or on how to construct those knowledge databases will be provided at the experimental result section (see section III). In the on-line stage, the object extraction process is performed. To do this, We begin by over-segment the original image into many small and homogeneous regions. This is called a low-level processing step. In this paper, we have used SRM algorithm [40] (cf. section II-B) as a tool of segmentation. Having a segmented image, the following task is a high-level processing step that consists in extracting features characterizing regions of both segmented image and constructed knowledge databases. In this work, RGB color histogram features (cf. section II-C)are used. The question that arises then is how can we measure the similarity between those regions. Several well-known goodness-of-fit statistical metrics using RGB color histogram features exist in the literature. In this work, the Bhattacharyya descriptor is adopted to accomplish this operation (cf. section II-D).

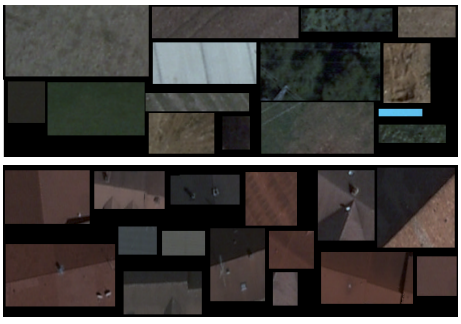


Fig. 1: Example of knowledge databases used in this work. From top to bottom: knowledge database $\mathbf{B_{back}}$ of background (vegetation, road, forest, etc) and knowledge database $\mathbf{B_{obj}}$ of building roofs (red and non-red rooftop buildings).

Once similarity measure is evaluated for all regions, each one of them can be classified as a part of an object of interest or rather as a part of the background of the image (see section II-E). Finally, object contours are delineated keeping only regions labelled as object of interest (building roof in this case).

Figure 2 resumes the general flowchart of the proposed building-detection method.

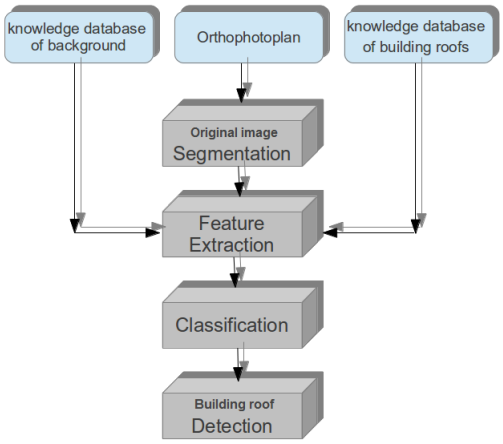


Fig. 2: General flowchart of the proposed building-detection method.

B. Initial segmentation using Statistical Region Merging

The low-level processing step consists in over-segmenting the input image into many small and homogeneous regions with the same properties. The goal of this initial segmentation is to avoid the under-segmentation problem and thus correctly extract all significant regions where boundaries coincide as closely as possible with the significant edges present in the image. Of course, there are many low level segmentation methods in the literature which can achieve that. One can cite Mean shift, Jseg unsupervised segmentation algorithm [41], watershed, Turbopixels [42], Statistical Region Merging (SRM) [40], etc. In this paper, authors have chosen SRM algorithm to obtain the initial segmentation of the input image. Particular advantages of using this algorithm for dealing with large images are that SRM dispenses dynamical maintenance of region adjacency graph (RAG), it allows defining a hierarchy of partitions and it runs in linear-time by using bucket

sorting algorithm while transversing the RAG. In addition, the SRM segmentation method not only considers spectral, shape and scale information, but also has the ability to cope with significant noise corruption and handle occlusions (Fig. 3).

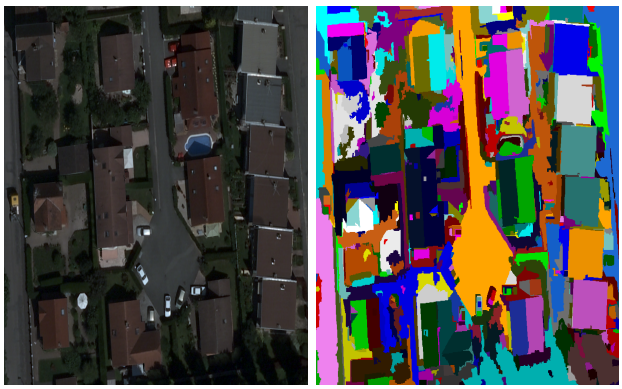


Fig. 3: Example of segmentation result using Statistical Region Merging (SRM) method. From left to right: Original image and its SRM segmentation result.

C. Region representation

In this stage of the method, we dispose of a segmented image obtained via the SRM algorithm. It is still a challenging problem to accurately extract the object contours from this image because only the segmented regions are calculated and no information estimation on their content, which is necessary for the extraction process, is yet available. The main goal consists in classifying each segmented region as target object or background. For this purpose, we need first to join the strategy adopted by many authors and which characterizes the regions using suitable descriptors.

It appears from the literature that there are several aspects that could be considered for representing a region such as edge [43], texture [44], shape, size or color. For the present purpose, the most appropriate information is color. In fact, region texture, which can be understood as repeatedly occurring local patterns in images and its arrangement rules, are unfortunately difficult to describe; Also, the same difficulties can be faced regarding shape and edge. Moreover, region size, although it can be measured simply by computing the number of pixels, it doesn't allow a unique distinction of objects of interest since they can have different sizes from an image to an other or simply they can have the same size as other objects belonging to the background of the image. Hence, color information which can be tackled using simply by computing its mean value or its histogram is an effective parameter to describe statistical information of object color distribution. Note that region histograms are local histograms and they reflect local features in images. Therefore, we exploit color histogram to represent all regions of the segmented image and those of the constructed knowledge databases.

In this purpose, each color channel is at first uniformly quantized into *l*=16 levels; afterwards, the color histogram of each region is calculated within the feature space of $l \times l \times l$ = 4096 bins. Obviously, quantization reduces the information regarding the content of regions and it is used as trade off when one wants to reduce processing time. The RGB color space is used in order to perform these computations.

Now that we have defined the feature adopted for characterizing the regions, the key issue is to determine similarity between regions of the segmented image and those of the constructed knowledge databases. For that, a similarity measure rule $\varrho(R,Q)$ between two regions R and Q should be defined basing on their color histograms.

D. Similarity measure rules

The most similarity measures commonly used are based on vector space model, i.e. taking image region features as points in the vector space, through the calculation of close degree of two points to measure the similarities between the image region features. Common similarity measures include Minkowski measure, histogram intersection method [45], second type distance [46], Bhattacharyya coefficient [47], and log-likelihood ratio statistic [48], etc. For regions R and Q, using the notation $\varrho(R,Q)$ for representing the similarity between regions R and Q, the larger ϱ is, the larger similarity between region R and Q we will get. Denote by $Hist_R^i$ the normalized histogram of a region R, the superscript i represents its i^{th} element. $z = l \times l \times l = 4096$ represents the feature space.

Examples of similarity measures are given as follows:

- Minkowski measure:

$$\varrho(R,Q) = \left(\sum_{i=1}^{z} |Hist_R^i - Hist_Q^i|^p \right)^{\frac{1}{p}}, \quad (1)$$

where $p = 1, 2 or \infty$;

- Euclidean distance:

$$\varrho(R,Q) = \sqrt{\sum_{i=1}^{z} (Hist_R^i - Hist_Q^i)^2}, \quad (2)$$

which is a Minkowski measure with p=2.

- Quadratic distance metric:

$$\varrho(R,Q) = \sqrt{\sum_{i=1}^{z} [(Hist_R^i - Hist_Q^i)^T A (Hist_R^i - Hist_Q^i)]}, \quad (3)$$

A is the bin-similarity matrix;

- Histogram intersection method:

$$\varrho(R,Q) = \frac{\sum_{i=1}^{z} min\{Hist_R^i, Hist_Q^i\}}{\sum_{i=1}^{z} Hist_R^i} \quad (4)$$

- Bhattacharyya coefficient:

$$\varrho(R,Q) = \sum_{i=1}^{z} \sqrt{Hist_R^i . Hist_Q^i} \quad (5)$$

In this work, authors adopted Bhattacharyya coefficient, which represents the cosine of angle between the unit vectors

$$(\sqrt{Hist_R^1}, ..., \sqrt{Hist_R^z})^T$$

and

$$(\sqrt{Hist_Q^1}, ..., \sqrt{Hist_Q^z})^T$$

This choice is due to its ability to simulate very well the similarity value of vector shape. The higher the Bhattacharyya coefficient between regions R and Q is, the higher the similarity between them is. That is to say their histograms are very similar and the angle between the two histogram vectors is very small. Certainly, two similar histograms do not necessarily involve that the two corresponding regions are perceptually similar. Nevertheless, coupling with the proposed classification process introduced in the next section II-E, Bhattacharyya similarity works well in the proposed approach.

It should be mentioned that a histogram is a global descriptor of a local region and it is robust to noise and small variations. Given that the Bhattacharyya coefficient is the inner product of two histogram vectors, this coefficient is thus robust to noise and small variations too. It has been used in [12], [13] for user interaction based image segmentation. Unlike theses methods, the proposed one aims for multiple extraction of objects of interest using two constructed knowledge databases without any need for user to provide markers input usually necessary for region merging process.

E. Classification process

At this stage of the method, we aim to determine which of the two classes (objects of interest or background) will be affected to the regions composing the initial SRM segmentation result, which we denote $\mathbf{M_{SRM}}$. For this end, candidate regions of $\mathbf{M_{SRM}}$ that have maximal similarity with the regions of the knowledge database $\mathbf{B_{obj}}$ of objects and those having maximal similarity with the regions of the knowledge database $\mathbf{B_{back}}$ of background are identified. Once all regions of $\mathbf{M_{SRM}}$ are classified, this leads directly to extracting the desired objects (e.g. building roofs). The proposed object extraction method can be summarized as in algorithm 1. As one can state, the similarity rule is very simple but it is efficient for the classification process. Note that the mean values of similarity moy^R_{obj} and moy^R_{back} are inversely proportional to the value of k, i.e. the higher the value of k is, the lower the mean value of similarity is. If this is the case, a dispersion of the mean values of similarity is obtained, which involves obtaining false classification result of the regions of $\mathbf{M_{SRM}}$. Besides, the k value has an important impact on the quality of results. For the purpose of keeping a significant similarity mean value, avoid the dispersion phenomenon and hence obtain good classification results, the two values moy^R_{obj} and moy^R_{back} are only calculated on the k first values of the sorted similarity vectors V^R_{obj} and V^R_{back} respectively. Although the similarity mean values moy^R_{obj} and moy^R_{back} are sensitive to outliers of k value, we empirically found that there is a range of values where the classification results remain stable. This optimum range is determined experimentally using a trial and error approach. Once this parameter is determined, it keeps the same value for test images. In this work, k is adjusted at 7.

Algorithm 1 Multiple objects extraction algorithm

Require: $I \leftarrow$ input image.
$\mathbf{B_{obj}} \leftarrow$ Knowledge database of objects of interest (building roofs).
$\mathbf{B_{back}} \leftarrow$ Knowledge database of background (vegetation, road, forest, etc)

(over)Segment I into regions through SRM algorithm in order to obtain the set $\mathbf{M_{SRM}}$ of segmented regions.
Calculate the RGB color histogram features for all regions of $\mathbf{M_{SRM}}$ and for those composing the two constructed knowledge databases $\mathbf{B_{obj}}$ and $\mathbf{B_{back}}$.
for each candidate region R $\in \mathbf{M_{SRM}}$ **do**
  Calculate the similarity vector $V^R_{obj} = \{\varrho(R,Q_i); (Q_i)_{i=1..m} \in \mathbf{B_{obj}}\}$ between R and $\mathbf{B_{obj}}$. $\varrho(R,Q_i)$ is the similarity between the region R and the region $Q_i \in \mathbf{B_{obj}}$.
  Calculate the similarity vector $V^R_{back} = \{\varrho(R,Q_j); (Q_j)_{j=1..n} \in \mathbf{B_{back}}\}$ between R and $\mathbf{B_{back}}$. $\varrho(R,Q_j)$ is the similarity between the region R and the region $Q_j \in \mathbf{B_{back}}$.
  Get the order of V^R_{obj} and V^R_{back} by decreasing sorting;
  Calculate $moy^R_{obj} = \frac{\sum_{i=1}^{k} \varrho(R,Q_i)}{k}$, $k \leq m$, the mean of the k first elements of V^R_{obj}.
  Calculate $moy^R_{back} = \frac{\sum_{j=1}^{k} \varrho(R,Q_j)}{k}$, $k \leq n$, the mean of the k first elements of V^R_{back}.
  if $(moy^R_{obj} \geq moy^R_{back})$ **then**
    The region R maximizes the similarity with $\mathbf{B_{obj}}$, it is then classified as a part of building roof.
  else
    The region R maximizes the similarity with $\mathbf{B_{back}}$, it is then classified as a part of background.
  end if
end for
The classification process is finished.
return The final segmentation map.

III. Experimental results

In this section, we are interesting in assessing the ability of the proposed building-extraction strategy to deal with multiple detection of building roofs from orthophotoplans. As pointed out in the introduction, the proposed building-extraction algorithm runs automatically without any user interaction. To avoid each time calculating region features from the two constructed knowledge databases and thus reduce the computation time, an alternative consists in calculating them once and for all and save them in a binary file. Thus, the process of similarity measure is performed using this binary file and no more the two knowledge databases.

A. Material description

1) Study area and knowledge databases: Data used in this research to evaluate the accuracy of the proposed algorithm are aerial images, particularly orthophotoplans. Several images were acquired for the region of Belfort city situated on the north-eastern of France in 2003, from a hot air balloon. Their spatial resolution is 16 cm/px. These images cover a wider area, where appear complex and multiple objects of different classes, various shadows, occlusions, multiple colors and textures and some terrain height variability. Namely, targeted objects, that are roofs of buildings, are often red and rarely non-red. In addition, they may differ according to their exposure to the sun and so they could change in terms of contrast and luminance. Therefore, these differences should be taken in consideration during the step of the construction of the knowledge databases.

This construction is explicitly performed by selecting a number of distinctive textures representing both the roofs and

TABLE I: Distinctive textures used to build object and background databases.

	Vegetation	Roads	Floors	Swimming Pools	Red roofs	Non-red roofs	Total
Object database	-	-	-	-	7	6	13
Background database	5	5	4	1	-	-	15

the background of the image. For this purpose, and as one can see on the second row of table I, a total of thirteen roof textures have been picked up from original images. Among them, seven are red and six others are gray or somewhat black. Thus, differences in contrast and brightness of objects of interest should be token into account. As for background database, we had taken a total of fifteen textures that belong to the background of the images. Five textures, related to vegetation and roads, are token for each category; four textures are selected from floors whereas only a single texture was kept to represent pools found on ortophotoplan images (cf. last row of tableI).

2) Test images: A set of six images is considered to evaluate the performance of the proposed roof extraction method. These images have been extracted from a huge original image like those described at the previous paragraph (cf. III-A1). To achieve that, the captured scenes should be varied in order to have a set of images that exhibit various conditions and increasing levels of difficulty. Having this in mind, the following criteria were selected: the number of roofs within the scene, their size and color, and finally the degree of discrepancies between roofs and the background. The first row of figure 4 shows four test images. The two other test images, which are enlarged, are shown in figure 5.

B. Accuracy assessment of the method

We begin by a qualitative evaluation of the proposed method using representative test images. Figure 4 illustrates the results of roof detection on the set of processed images. In the upper row of this figure, we show the original images; in the midst, the segmented images are given and in the lower row the corresponding building roof extraction where the final detected building boundaries drawn with red color are superimposed upon the original images. Basing on visual evaluation of the results, one can state that the developed approach demonstrates excellent accuracy in terms of building boundary extraction; this means that the majority of the building roofs present in the images are detected with good boundary delineation. Indeed, this method gives reliable results across complex environment composed of buildings presenting red and non-red rooftop, road areas, vegetation, etc. The images of figures 4.a, 4.b, 4.c and 4.d include several building rooftops and road areas with same color and texture, the proposed approach is able to successfully distinguish between them.

However, as one can see from the experimental results of figure 4, due to radiometric similarity between building roofs and image background, some false or imperfect detections can be generated. In fact, although we obtained notably accurate multiple detection of building roofs, the proposed method missed some part of buildings when the contrast between their rooftop and the background is low. Also, some vegetation areas are extracted as part of buildings because of their radiometric characteristics which are similar. In figure 5, some of building parts that have not been extracted are pointed out by yellow ellipses while some false detections are pointed out by green ellipses.

As for quantitative evaluation, we use measures widely employed in evaluating effectiveness. They constitute a useful and accepted tool in the object recognition field [49]. Within the orthophotoplans used in this work, 100 buildings were first manually delineated. Then, they are used as a reference building set to assess the accuracy of the automated building extraction. The extraction results and reference ones are compared pixel-by-pixel. Each pixel in the images is categorized as one of four possible outcomes:

1) True positive (TP): Both manual and automated methods label the pixel belonging to building.
2) True negative (TN): Both manual and automated methods label the pixel belonging to background.
3) False positive (FP): The automated method incorrectly labels the pixel as belonging to building.
4) False negative (FN): The automated method does not correctly label the pixel truly belonging to building.

To examine detection performance, the number of pixels that fall into each of the four categories TP, TN, FP, FN are determined, and the following measures are computed:

$$\text{Branching Factor (B.F)} = \frac{FP}{TP}$$

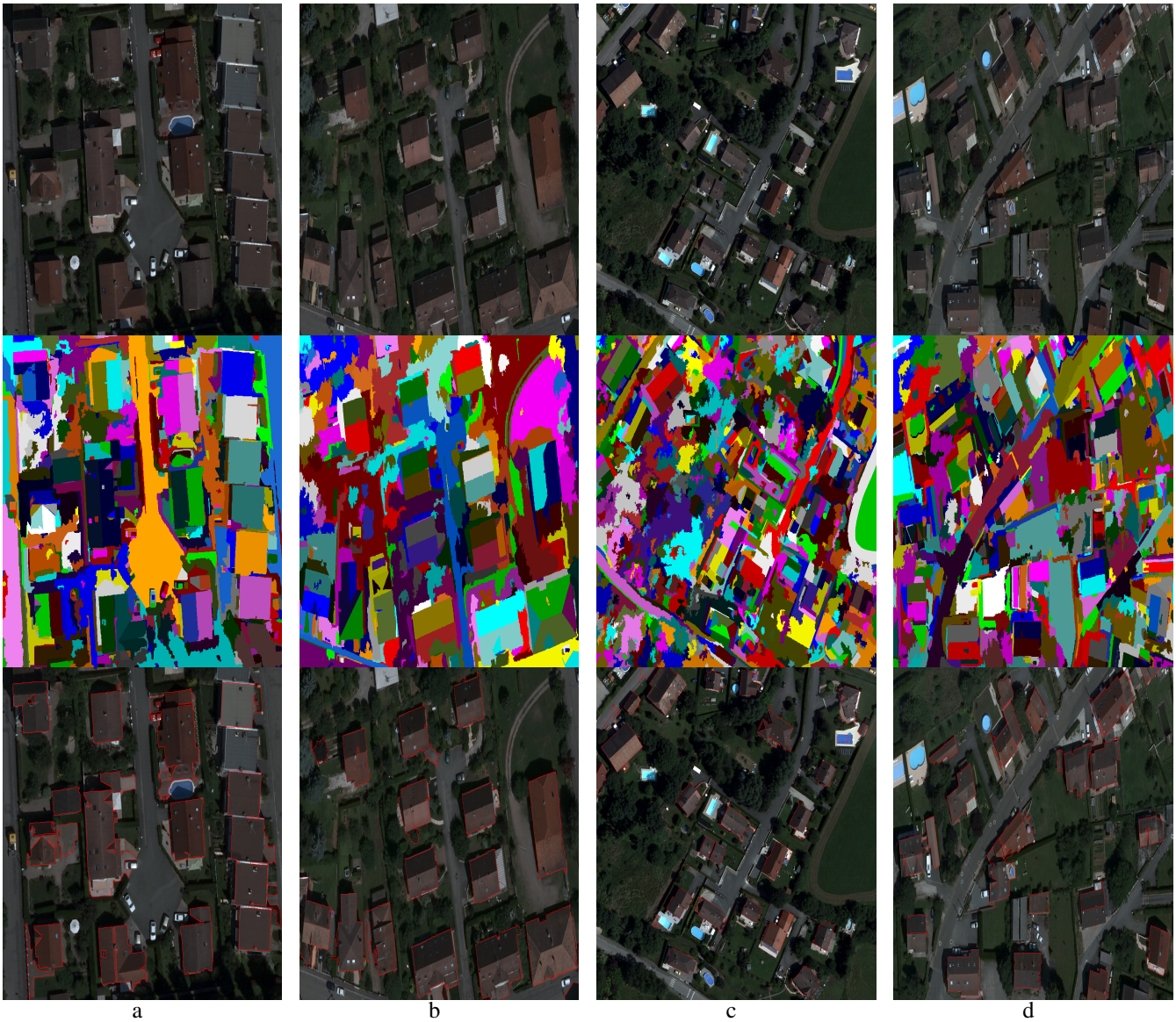
$$\text{Miss Factor (M.F)} = \frac{FN}{TP}$$

Detection Percentage (D.P) = $100.\text{TP}_{\overline{TP+TN}}$

Quality Percentage (Q.P) = $100.\text{TP}_{\overline{TP+FP+FN}}$

The interpretation of the above measures is as follows. The detection percentage denotes the percentage of building pixels correctly labelled by the automated process. The branching factor is a measure of the commission error where the method incorrectly labels background pixels as building. The more accurate the detection is, the closer the value is to zero. The miss factor measures the omission error where the method incorrectly labels building pixels as background. These quality metrics are closely related to the boundary delineation performance of the building extraction method. The quality percentage in turn, measures the absolute quality of the extraction and is the most stringent measure. To obtain 100% quality, the extraction algorithm must correctly label every building pixel ($FN = 0$) without mislabelling any background pixel ($FP = 0$).

The results of the quality assessment of the method for the images, illustrated in figures 4 and 5, are given in table II. The last row of the table gives the average values obtained with all the orthophotoplans used in this work. The



a Image 1 | b Image 2 | c Image 3 | d Image 4

Fig. 4: Automatic extraction of multiple building roofs from the set of processed images. From top to bottom: original images, SRM segmentation results, and corresponding building roof extraction.

values obtained on the set of the processed images confirm the claims mentioned above regarding the performance of the proposed approach. Effectively, the results show that the building-extraction approach is quite successful for extracting the buildings from orthophotoplans with the D.P and Q.P average values of 93.91% and 85.30%, respectively. In addition to this, the branching factor and the miss factor average values were found to be 0.111 and 0.067 , respectively.

Also, we have transcript these comparison results in terms of Receiver Operating Characteristics (ROC) graphs [50]. In machine learning, ROC graphs are used as a useful technique for visualizing and selecting classifiers based on their performance. ROC graphs are two-dimensional graphs in which **True Positive Rate (TPR)** (also called *recall* or *sensitivity*) is plotted on the Y axis and **False Positive Rate (FPR)** (also called *false alarm rate*) is plotted on the X axis. These measures are computed using the four outcomes mentioned above as

TABLE II: The quality assessment results of the building extraction.

Image measures	B.F	M.F	D.P	Q.P
Fig.4.a	0.1745	0.1116	90	77.75
Fig.4.b	0.14	0.0289	97.19	85.58
Fig.4.c	0.1804	0.0303	97.06	82.59
Fig.4.d	0.0871	0.1126	89.87	83.35
Fig.5.a	0.0184	0.0789	93.37	92.66
Fig.5.b	0.0710	0.0414	96.01	89.88
Avr./100 building roofs	**0.1118**	**0.0673**	**93.91**	**85.30**

a. Image 5

b. Image 6

Fig. 5: Examples of some building parts that have not been extracted (yello ellipses) and some false detections (green ellipses).

follows:

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

$$FPR = \frac{FP}{FP + TN} \tag{7}$$

As general rule in the context of discrete classifiers, the most important point in ROC space is the upper left corner, point of coordinate (0,1) which represents perfect classification. Informally, a point in ROC space is better than another if the first one is located into the north-west side of the second one(TPR is higher, FPR is lower, or both). Points above the diagonal dividing the ROC space represent good classification results (better than random), while points below the diagonal represent poor results (worse than random). In sum, the closer the ROC plot is located at the top-left border of ROC space, the more accurate are the results.

Figure 6 shows the accuracy of the proposed method applied on the test images. Basing on this ROC graph, all measures are on the top-left side, indicating hence perfect detection performance on this test set.

C. Comparative evaluation and discussion

To provide further evaluation of the performance of the developed method and considering that the proposed approach

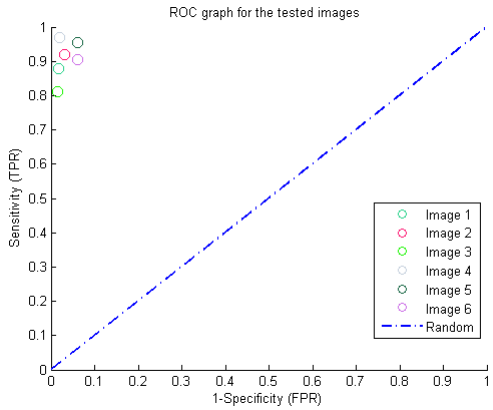


Fig. 6: The ROC graph of the test images using the proposed method.

is mainly based on a simple supervised classification technique, we propose in this section to compare it with two popular supervised classification methods of the literature, within the same framework of building detection problem. The classification algorithms we have selected for this comparison are Support Vector Machine (SVM) and K Nearest Neighbours (KNN). These supervised classification methods are well known and often used for image classification purposes [14]. The two methods have several control parameters. For KNN, k nearest neighbors and distance are the key words in this algorithm. The principal parameters for SVM algorithm are the type of SVM, the type of kernel function and the degree in the kernel function. For furthermore details concerning these settings, we refer to the paper written by Chang and al. [51]. In this comparison with KNN and SVM classifiers, training samples are given by the color histograms of regions of the two knowledge databases whose labels are known and the test samples are given by the color histograms of regions of the test segmented images.

To obtain a meaningful comparison, each algorithm must be tested considering many possible combinations of input parameters. In effect, for each classification method, we consider its performance and correctness, as measured by its success rate calculated by the measures previously detailed, as well as its stability with respect to changes in parameter settings and with respect to all tested images.

The parameter setting that we have used here was set empirically through manual checking of the recognition results and is reported on table III.

Figure 7 illustrates an example of building extraction results obtained under different possible combinations of input parameters, as mentioned in Table III, for both KNN and SVM algorithms. The visual analysis shows that KNN_0 (using euclidean distance and k=1) and SVM_2 (using ν-SVC as type of SVM, sigmoid as type of kernel function and a degree in kernel function equal to 4) give good detection results. Note that, despite several combinations of parameters, neither of the two tested algorithms outperforms the proposed method. In fact, the approach is quite successful for extracting the buildings from the images (the extraction result match

TABLE III: Experimental parameter values related to the used methods (KNN and SVM) for comparison.

Methods	Parameters	Used Values			
		KNN_0	KNN_1	KNN_2	KNN_3
KNN	K:	1	3	1	3
	Distance:	Euclidian	Euclidian	Cityblock	Cityblock
		Used Values			
		SVM_0	SVM_1	SVM_2	SVM_3
SVM	Type of SVM:	C-SVC	ν-SVC	ν-SVC	ν-SVR
	Type of kernel function:	Radial basis function	Linear	Sigmoid	Polynomial
	Degree in kernel function:	3	4	4	1

TABLE IV: Quality assessment obtained for all the methods (KNN and SVM under different parameters, and the proposed method).

(%)	KNN				SVM				Our method
	KNN_0	KNN_1	KNN_2	KNN_3	SVM_0	SVM_1	SVM_2	SVM_3	
Q.P	42.34	47.48	41.63	29.96	58.79	68.53	70.11	66.33	85.59
D.P	73.69	67.71	69.58	30.86	64.12	78.51	81.27	78.54	97.19
B.F	1.00	0.63	0.97	0.09	0.14	0.19	0.20	0.23	0.14
M.F	0.36	0.48	0.44	2.24	0.56	0.27	0.23	0.27	0.02

most closely the ground truth) whereas the KNN and SVM algorithms lead to many false positives on road and vegetation areas and false negatives within buildings accompanied by a loss of several parts of roofs. The performance evaluation of the tested classification methods has been summarized in table IV and figure 8 resumes it with a graphical representation. The measures used to assess the quality of detection for this comparative analysis are the same as those used within the experiments addressed above to evaluate the proposed method separately (cf. section III-B).

Table IV shows the quantitative indicators based results obtained for each method. For KNN, we note that the optimal result in terms of quality detection (QD) is obtained using the second configuration KNN_1 with a value of 47.5%, while for the SVM, quality detection indicator reaches a value of 70% by the configuration SVM_2. The proposed method shows higher performance reaching 85.6% as a value for quality detection indicator. In relation to the percentage of detection (DP), the optimal values are 73.7% and 81.27% respectively for KNN and SVM against 97.19% for our method. To get an idea on missed parts in detection results, we rely to the omission factor (MF) which reaches 0.36 and 0.23 for KNN (KNN_0) and SVM (SVM_2) respectively. This measure attains 0.02 for the proposed method showing thus its superiority to the other methods in terms of the percentage of pixels that are not classified as building.

These results confirm the efficiency of the decision rule that the proposed method uses. For instance, unlike KNN method, which classifies an item by a majority vote of its neighbours, (i.e. the test item is assigned to the class most common among its k nearest neighbours taken from the training samples), considering instead the maximal average over the k similarity measures of regions from both object and background knowledge databases allows this new method to outperform KNN method.

By considering figures 9 and 10, one can see how ROC graphs show the out-performance of the developed method comparatively to the variants of KNN and SVM classifiers.

As for the computation time, It should be noted that the proposed method requires in average 5 seconds for extracting objects of interest (building roofs in the current application) from images of about 1500 by 1000 pixels working on a machine of 2.75 MHZ(CPU) and a memory of 3 GO (RAM). Besides, it depends on the number of regions of the segmented image.

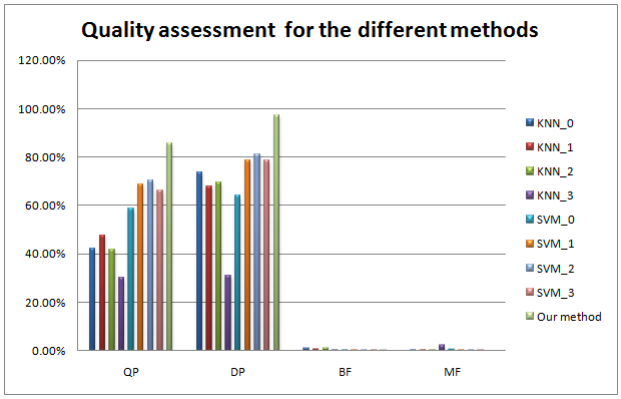


Fig. 8: Graphical representation of table IV.

IV. Conclusion and Future Works

In this paper, we have presented an efficient method for automatic and accurate multiple objects extraction from images. Unlike interactive methods, the proposed one requires no user interaction. The method involves two knowledge databases where the first one is constructed with several significant textures of objects to be extracted and the second one is composed with textures representing background. After an over-segmentation of the original image, the segmented regions are classified as objects or background using a region similarity measure and the constructed knowledge databases. The proposed method is evaluated for building roof extraction from orthophotoplans, which is a very challenging problem

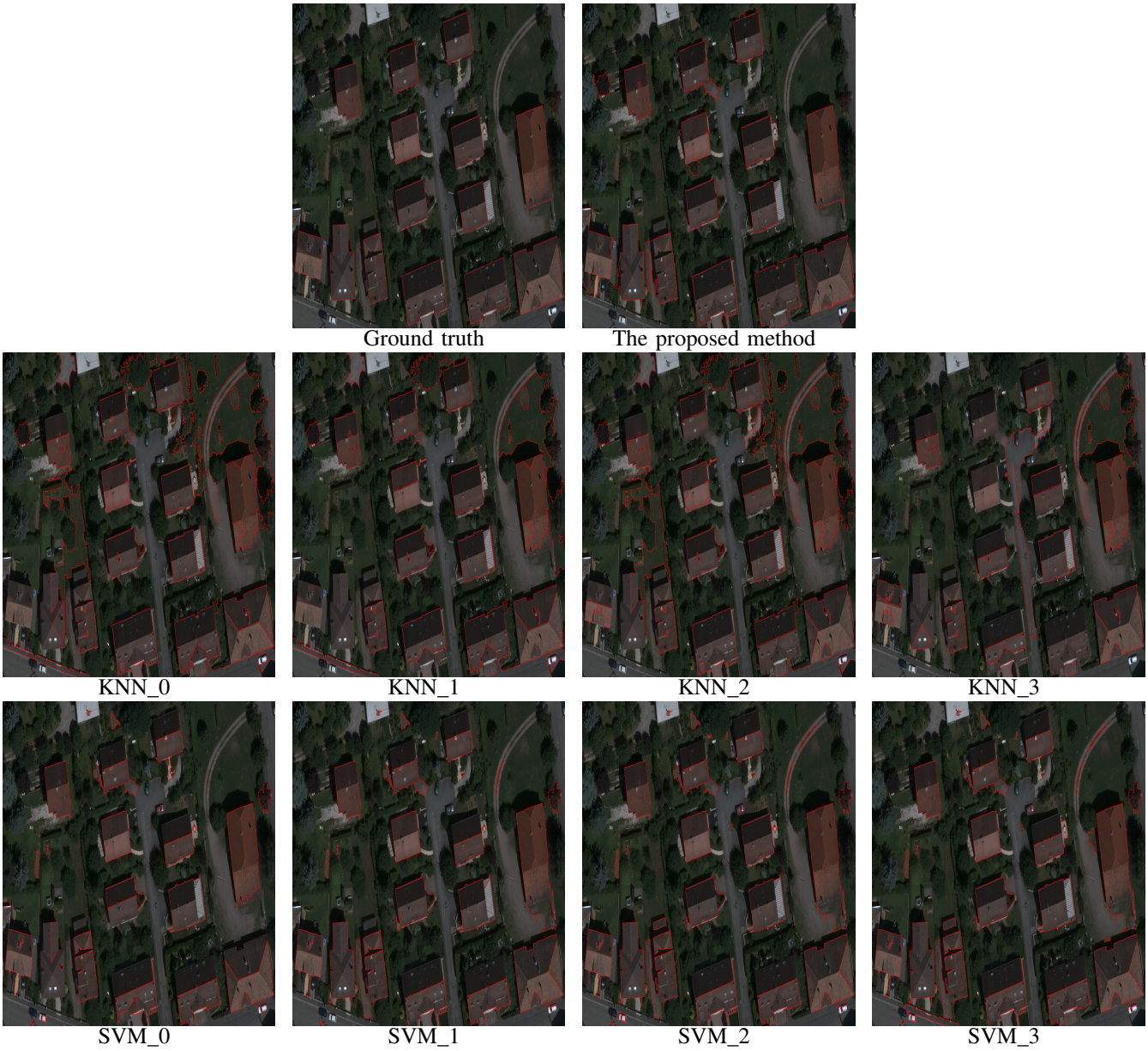


Fig. 7: Comparison between KNN, SVM and the proposed method. First row: the ground truth image and the roof extraction results using the proposed method; second row: extraction results by KNN method under different parameters (see table III); last row: extraction results by SVM method under different parameters (see table III).

because of the complexity of scenes with a large number of different objects (buildings, roads, vegetation, etc.). The evaluation consisted also of a comparison analysis between the proposed method and popular ones (KNN and SVM).

In order to improve the proposed method, there are several open questions that we still need to explore. First, the color histogram features are calculated using the RGB color space. The orthophotoplan images in our possession contain a certain heterogeneity in terms of lights, illumination changes, shadows, etc, what constitutes a breeding ground for false detections. To overcome these drawbacks and hence reduce the effect of illumination and limit the artefacts of the acquired image, studying and evaluating different color spaces and/or colorimetric invariants seems to be an interesting way forward [52], [53]. In addition, the proposed object-extraction method enables flexible feature descriptor integration. Thus, we propose to study the effect of other region characteristics on the quality of the results. One can cite the Local Binary Patterns (LBP) texture operator which is a powerful structural model of texture analysis [54]. Also, we think that it could be possible to estimate analytically the value of the parameter k involved in the similarity computation. To evaluate the genericity of the proposed method, we envisage to apply it for other image types (medical,...).

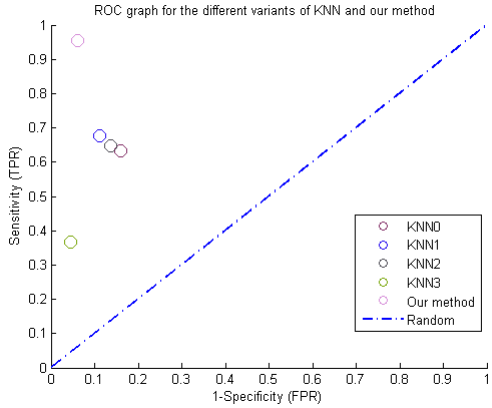


Fig. 9: The ROC graph comparing variants of KNN and the proposed method.

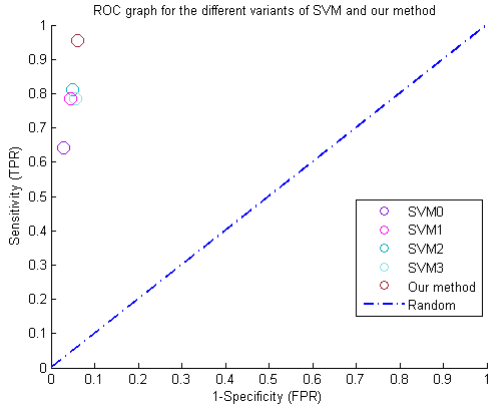


Fig. 10: The ROC graph comparing variants of SVM and the proposed method.

ACKNOWLEDGMENT

The authors would like to thank the "Communauté d'Agglomération Belfortaine" for making available the "orthophotoplan" images on which the tests have been performed.

REFERENCES

[1] Pueyo, O.; Patow, G. Structuring urban data. *The Visual Computer*, **2014**, *30, Issue 2*, 159-172

[2] Sadgal,M.; El Fazziki,A.; Ait Ouahman, A. *Aerial image processing and object recognition*, *The Visual Computer* **2005** , *21, Issue 1-2*, 118-123

[3] Boykov,Y.; Jolly, M. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. *IEEE Intl. Conf. on Comput. Vision* **2001**. *1*, 105-112

[4] Rother, C.; Kolmogorov, V.; Blake, A. GrabCut -Interactive Foreground Extraction using Iterated Graph Cuts. *ACM Transactions on Graphics* **2004**. *23(3)*, 309-314

[5] McGuinness, K.; O'Connor, N. E. A comparative evaluation of interactive segmentation algorithms. *Pattern Rec.* **2010**, *43*, 434-444

[6] Adams, R.; Bischof, L. Seeded region growing.*IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 641-647

[7] Friedland, G.; Jantz, K.; Rojas, R. SIOX: simple interactive object extraction in still images. *IEEE Intl. Symposium on Multimedia* **2005**. 253-259

[8] Adamek, T.: Using contour information and segmentation for object registration, modeling and retrieval. *PhD thesis*, Dublin City Univ **2006**

[9] Bai, X.; Sapiro,G. Geodesic matting: a framework for fast interactive image and video segmentation and matting.*Intl. J. Comput.* **2009**,*82*, 113-132

[10] Yang,C.; Duraiswami, R.; Gumerov, N. A.; Davis, L.: Improved fast gauss transform and efficient kernel density estimation. *IEEE Intl. Conf. on Comput. Vision* **2003**. *1*, 664-671

[11] Protiere, A.; Sapiro, G.: Interactive image segmentation via adaptive weighted distances. *IEEE Transactions on Image Processing* **2007**.*16*, 1046-1057

[12] Ninga, J.; Zhanga, L.; Wub, D. Z. C.: Interactive image segmentation by maximal similarity based region merging.*Pattern Recognition* **2010**.*43* , n 2. 445-456

[13] Chen, M.; Hou, Y.; Zeng, X.; Lu,Y.: A multi-label interactive image segmentation method based on region merging. *IEEE International Conference on Mechatronic Science, Electric Engineering and Computer* **2011**, Jilin, China. 2336-2340

[14] Tso, B.; Mather, P.M. : Classification Methods for Remotely Sensed Data, 2nd ed. ; *CRC Press: Boca Raton*, FL, USA **2009**

[15] Giorgos, M.; Jungho, I.; Caesar, O. : Support vector machines in remote sensing: A review.*ISPRS Journal of Photogrammetry and Remote Sensing* **2011**, *66*, 247-259

[16] Knerr, S.; Personnaz, L.; Dreyfus, G. : Single-layer learning revisited: a stepwise procedure for building and training a neural network. *In: Neurocomputing: Algorithms, Architectures and Applications. In: NATO ASI Series* , Springer, **1990**

[17] Bischof, H.; Schneider, W.; Pinz, A.J. : Multispectral classification of Landsat-images using neural networks. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 482-490

[18] Teemu, K.; J. Kristian, K.; Lasse, L.; Heikki, K. : Unsupervised object discovery via self-organisation. *Pattern Recognition Letters* **2012**,*33*, 2102-2112

[19] Muralidharan,R.; Chandrasekar,C. : Object Recognition using SVM-KNN based on Geometric Moment Invariant. *International Journal of Computer Trends and Technology* **2011**, 215-220

[20] Hu, M.K. : Visual pattern recognition by moments invariants. *IRE Trans. Information Theory* **1962**, *8*, 179-87

[21] Dudani, S.; Breeding, K.; McGhee, R.: Aircraft identification by moments invariants. *IEEE Transactions on Computers* **1977**, *26(1)*, 39-45

[22] Flusser, J.; Suk, T.: A moment based approach to registration of image with affine geometric distortion. *IEEE Transactions on Geoscience and Remote Sensing* **1994**, *32(2)*,382-387

[23] Flusser, J. On the independence of rotation moment invariants. *Pattern Recognition* **2000**, *33(9)*,1405-1410

[24] Soille, P.; Pesaresi, M. : Advances in mathematical morphology applied to geoscience and remote sensing. *IEEE Transactions on Geosciences and Remote Sensing* **2006**, *40*, 2042-2055

[25] Giada, S.; De Groeve, T.; Ehrlich, D.; Soille, P. Information extraction from very high resolution satellite imagery over Lukole refugee camp, Tanzania. *Int. J. Remote Sensing* **2002** , *24* , N 22, 4251-4266

[26] Salman, A.; , Valadan, M.J.Z.; Hamid, E.;Hamid, A. M.; Ali, M. : Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours. *International Journal of Applied Earth Observation and Geoinformation* **2010**, *12*, 150-157

[27] Kass, M.; Witkin, A.; Terzopoulos, D. Snakes: active contour models. *International Journal of Computer Vision* **1998**, *1*, 321-331

[28] Halla, Brenner, N.C. Extraction of building and trees in urban environment. *ISPRS Journal of Photogrammetry & Remote Sensing* **1999**, *54*, 130-137

[29] Zhao, B.; Trinder, J.C. Integrated-approach-based automatic building extraction. *International Archives of Photogrammetry and Remote Sensing XXXIII* **2000**, *Part B3*

[30] Samadzadegan, F.; Azizi, A.; Hahn, M.; Lucas, C. Automatic 3D object recognition and reconstruction based on neuro-fuzzy modeling. *ISPRS Journal of Photogrammetry & Remote Sensing* **2005**, *59*, 255-277

[31] Zimmermann, P. A new framework for automatic building detection

analyzing multiple cue data. *International Archives of Photogrammetry and Remote Sensing XXXIII* **2000**, *Part B3*

[32] Miliaresis, G.; Kokkas, N. : Segmentation and object-based classification for the extraction of the building class from LIDARDEMs. *Computers & Geosciences* **2007**, *33*, 1076-1087

[33] Lafarge, F.; Descombes, X.; Zerubia, J.; Pierrot-Deseilligny, M. Automatic building extraction from DEMs using an object approach and application to the 3D-city modeling. *ISPRS Journal of Photogrammetry & Remote Sensing* **2008**,*63*, 365-381

[34] Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* **2004**, *60(2)*, 91-110

[35] Berretti, S.; Ben Amor, B.; Daoudi,M.; Del-Bimbo, A. 3D facial expression recognition using SIFT descriptors of automatically detected keypoints. *The Visual Computer* **2011**, *27, Issue 11*, 1021-1036.

[36] Berg,A.;Berg, T.; Malik, J. Shape matching and object recognition using low distortion correspondence. *in: Proc. International Conference on Computer Vision and, Pattern Recognition* **2005**,26-33

[37] Ying-Ho, L.; Anthony, J.T.L.; Fu, C. : Object recognition using discriminative parts.*Computer Vision and Image Understanding* **2012**, *116*, 854-867

[38] Lazebnik, S.; Schmid, C.; Ponce, J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in: *Proc. International Conference on Computer Vision and, Pattern Recognition*, 2169-2178 **2006**

[39] Amores,J.; Sebe, N.; Radeva, P. Fast spatial pattern discovery integrating boosting with constellations of contextual descriptors. in: *Proc. International Conference on Computer Vision and Pattern Recognition* **2005**, *2*,769-774

[40] Nock,R.; Nielsen, F.: Statistical region merging. *IEEE Trans. Pattern Anal. Machine Intelligence* **2004**. *26*, n 11, 1452-1458

[41] Deng, Y.; Manjunath, B.S.: Unsupervised Segmentation of Color-Texture Regions in Images and video. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *23*, 800-810

[42] Levinshtein, A.; Stere, A.; Kutulakos, K. N.; Fleet, D. J.; Dickinson, S. J.; Siddiqi, K.: Turbo Pixels: Fast Super pixels Using Geometric Flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2009**. *31*, 2290-2297

[43] Birchfield,S.: Elliptical head tracking using intensity gradients and color histograms.*IEEE Conference on Computer Vision and Pattern Recognition* **1998**. 232-237

[44] Ojala, T.; Pietikainen, M.; Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Transactions on Pattern Analysis and Machine Intelligence* **2002**.*24*,971-987

[45] Swain, M.J.; Ballard, D.H.: Color indexing.*International Journal of Computer Vision* **2002**. *7*, 11-32

[46] Stricker, M.; Orengo, M.: Similarity of color images. *SPIE Storage and Retrieval for Image and Video Databases* **1995**. *3*, 381-392

[47] Kailath, T.: The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communications Technology* **1967** . *15*,52-60

[48] Fukunaga, K.: Introduction to Statistical Pattern Recognition. *Second ed., Academic Press* **1990**

[49] Shufelt, J. A. Performance evaluation and analysis of monocular building extraction from aerial imagery. *IEEE Trans. Pattern Anal. Machine Intell.* **1999**,*21*,311-326

[50] Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters*, **2006**, *27* , 861874.

[51] Chih-Chung, C.; Chih-Jen, L. : LIBSVM : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* **2011**. 2:27:1–27:27 . Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm

[52] El-Merabet, Y.; Meurie, C.; Ruichek, Y.; Sbihi, A.; Touahni, R.: Orthophotoplan segmentation based on regions merging for roof detection. *S&T/SPIE Electronic Imaging 2012 - Image Processing: Machine Vision Applications V*, SPIE 8300, Burlingame USA

[53] El-Merabet, Y.; Meurie, C.; Ruichek, Y.; Sbihi, A.; Touahni, R.: Orthophotoplans segmentation and colorimetric invariants for roof detection. *LNCS, 16th International Conference on Image Analysis and Processing (ICIAP)*, Ravenna Italy, 394-403 **2011**

[54] Abdou-Berbar, M. Three robust features extraction approaches for facial gender classification. *The Visual Computer*, **2014**, *30, Issue 1*, 19-31