ISSN 2395-8618

Object Classification using Hybrid Holistic Descriptors: Application to Building Detection in Aerial Orthophotos

Fadi Dornaika, Abdelmalik Moujahid, Alireza Bosaghzadeh, Youssef El Merabet, and Yassine Ruichek

***Abstract*—We present a framework for automatic and accurate multiple detection of objects of interest from images using hybrid image descriptors. The proposed framework combines a powerful segmentation algorithm with a hybrid descriptor. The hybrid descriptor is composed by color histograms and several Local Binary Patterns based descriptors. The proposed framework involves two main steps. The first one consists in segmenting the image into homogeneous regions. In the second step, in order to separate the objects of interest and the image background, the hybrid descriptor of each region is classified using machine learning tools and a gallery of training descriptors. To show its performance, the method is applied to extract building roofs from orthophotos. We provide evaluation performances over 100 buildings. The proposed approach presents several advantages in terms of applicability, suitability and simplicity. We also show that the use of hybrid descriptors lead to an enhanced performance.**

***Index Terms*—Automatic building detection and delineation, classification, supervised learning, image descriptors, orthophoto.**

I. Introduction

AUTOMATIC objects recognition has become a topic of growing interest for computer vision community. In the last two decades, machine vision techniques were more and more used in order to assist the whole process of Geographical Information Systems (GIS), cultural heritage preservation, risk management, and monitoring of urban regions. For instance, automatic extraction of man-made objects such as buildings and roads has gain significant attention over the last decade. Aerial data are very useful for the coverage of large areas such as cities and several aerial-based approaches are proposed for the extraction of buildings. More precisely, the data essentially employed as input to these approaches are either optical aerial images and derived Digital Surface Model (e.g., [1]) or aerial LiDAR 3D point clouds (e.g., [2]). It is well-known that segmenting buildings in aerial images is a challenging task. This problem is generally considered when we talk about high-level image processing in order to produce numerical or symbolic information. In this context, several methods have been proposed in the literature. Among the techniques most frequently used, one can cite semi-automatic 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. It consists to divide an image into two segments: "object" and "background." The interactivity consists in imposing certain hard constraints for segmentation by indicating certain pixels (seeds) that absolutely have to be part of the object and certain pixels that have to be part of the background. Rother et al. [3] presented an iterative algorithm called GrabCut by simplifying user interaction. Their method combines image segmentation using graph cut and GMMs (Gaussian Mixture Models) based statistical models (using the Orchard-Bouman clustering algorithm) of foreground and background structures in color space. A very useful segmentation benchmark, with a platform implementing important algorithms, has recently been proposed by McGuinness and Connor [4]. The authors compared important algorithms such as IGC [5], seeded region growing (SRG) [6], simple interactive object extraction (SIOX) [7]. The SIOX [7] algorithm is also based on color information and has recently been integrated into the popular imaging program GIMP as the "Foreground Selection Tool."

From the point of view of machine learning paradigms, it is desirable to keep the user interaction at the training phase only and to fully automate the detection and recognition at the test phase. In this paper, we propose an image-based approach for object detection and classification namely, detecting roof building in orthophotos. We use a Statistical Region Merging (SRM) regions to get an over-segmented image. The obtained regions are then described by holistic and hybrid descriptors for detection of roof building in orthophotos. First, an over-segmentation is applied on the orthophoto using the SRM algorithm. This over-segmentation is applied on both the training and test images. Second, holistic descriptors including color and Local Binary Patterns are fused in order to get the feature descriptor of a given region. Third, the SRM regions

Manuscript received on May 10, 2015, accepted for publication on June 5, 2015, published on June 15, 2015.

Fadi Dornaika is with the University of the Basque Country (UPV/EHU) and IKERBASQUE, Basque Foundation for Science, Spain (e-mail: fdornaika@hotmail.fr).

Abdelmalik Moujahid, Alireza Bosaghzadeh are with the University of the Basque Country (UPV/EHU), Spain (e-mail: jibmomoa@gmail.com, alireza.bossaghzadeh@gmail.com).

Youssef El Merabet is with Université Ibn Tofail, Kenitra, Morocco (e-mail: elmerabet113@gmail.com).

Yassine Ruichek is with IRTES-SeT, UTBM, Belfort, France (e-mail: yassine.ruichek@utbm.fr).

ISSN 2395-8618

in a test image are then classified using machine learning tools. We argue that the use of color and LBP descriptors will lead to better performance than relying on a single type of descriptors. We provide a performance study on classifiers whose role is to decide if any arbitrary region is a building or not. Furthermore, we provide performance evaluation at pixel level. This evaluation quantifies both the quality of the segmentation and the classification.

II. Proposed Method

The general flowchart of the proposed building-detection method is illustrated in Figure 1. It should be noticed that the training set is formed by a set of labeled regions together with their image descriptor.

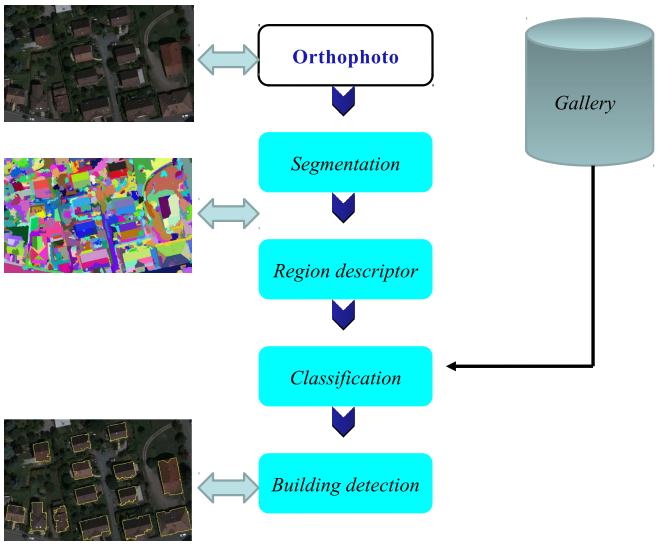


Fig. 1. General flowchart of the proposed building-detection method.

A. 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 characterizing the image. Of course, there are many low level segmentation methods in the literature which can do the job. One can cite Mean shift, Jseg unsupervised segmentation algorithm, watershed, Turbopixels, Statistical Region Merging (SRM), etc. In this work, we have used SRM algorithm to obtain the initial segmentation of the input image. Particular advantages of using this algorithm for dealing with large images are that it dispenses dynamical maintenance of a region adjacency graph, allows defining a hierarchy of partitions. In addition, the SRM segmentation method not only considers spectral, shape, scale information, but also has the ability to cope with significant noise corruption, handle occlusions.

B. Region Representation

In this stage of our method, we dispose of a segmented image obtained via the SRM algorithm. It is still a challenging problem to extract accurately the object contours from this image because only the segmented regions are calculated and no information estimation on their content necessary for the extraction process, is yet done. Our main goal consists in classifying each segmented region as target object or background. For this purpose, we need to characterize these regions using some suitable descriptors. It appears from the literature that there are several aspects that could be considered to represent a region such as the color, edge, texture, shape and size of the region. In our particular context, we believe that color and texture information are the most useful information.

1) Color Histograms: Color histograms were common image descriptors that can describe an object. Note that the region histograms are local histograms and they represent local features of images, and hence the regional color mean value or color histogram are effective parameters to describe statistical information of the object's color distribution. Therefore, we use the color histogram to represent all regions of the segmented image. First we uniformly quantize each color channel into $l = 16$ levels and then the color histogram of each region is calculated in the feature space of l6 x 16 x 16 = 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 calculate the color histogram. Obviously, other color space can be used. Figure 2 illustrates this process on two segmented regions, each belongs to a class.

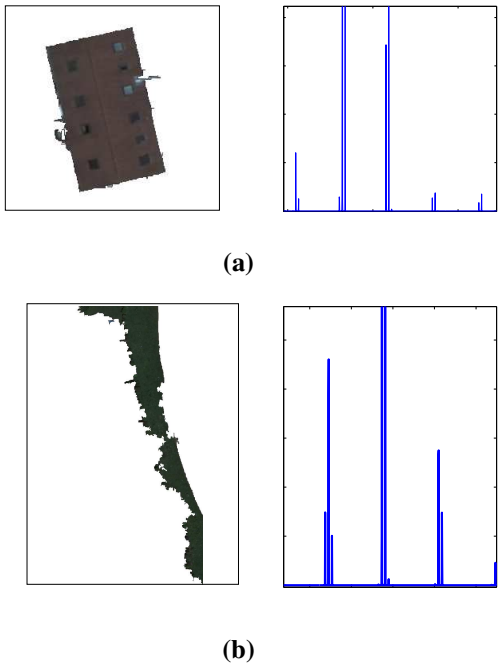


Fig. 2. **(a)** A segmented roof region and its color histogram. **(b)** A segmented background region and its color histogram.

 http://dx.doi.org/10.17562/PB-51-2

ISSN 2395-8618

2) Local Binary Patterns: Local Binary Patterns have proved to be a good texture descriptor. The original LBP operator labels the pixels of an image with decimal numbers, which are called LBPs or LBP codes that encode the local structure around each pixel [8], [9], [10]. It proceeds thus, as illustrated in Figure 3: Each pixel is compared with its eight neighbors in a neighborhood by subtracting the center pixel value; the resulting strictly negative values are encoded with 0, and the others with 1. For each given pixel, a binary number is obtained by concatenating all these binary values in a clockwise direction, which starts from the one of its top-left neighbor. The corresponding decimal value of the generated binary number is then used for labeling the given pixel. The histogram of LBP labels (the frequency of occurrence of each code) calculated over a region or an image can be used as a texture descriptor.

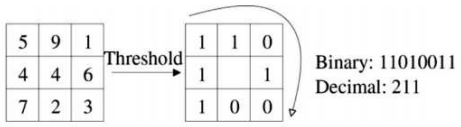
The size of the histogram is 2^P since the operator $LBP(P, R)$ produces 2^P different output values, corresponding to 2^P different binary patterns formed by P pixels in the neighborhood. Several LBP variants have been developed recently to improve performance in different applications [11], [12], [13]. These variants focus on different aspects of the original LBP operator.

For describing a segmented region, we use eight points ($P = 8$) with three radii ($R = 1$, $R = 2$, $R = 3$) each with three modes (uniform, rotation invariant, uniform and rotation invariant). Thus, there are nine LPB descriptors. The final descriptor is given by the concatenation of all. It is worth noting that despite the use of nine LBP descriptors the final one is described by $3 \times (59 + 36 + 10) = 315$ variables only.

3) Hybrid Descriptors: We propose to combine color and texture information in our region descriptor. This is done by simply concatenating the color descriptor and the LBP descriptor. Once the descriptor is computed we apply the square root on all its elements. The motivation for using the square root is that descriptor vectors consist of histograms, and applying square root prior to the distance calculations corresponds to the Hellinger distance between probabilities [14]. Moreover, some recent papers in face recognition literature has shown that the use of the square root of LBP histograms can enhance the recognition performance.

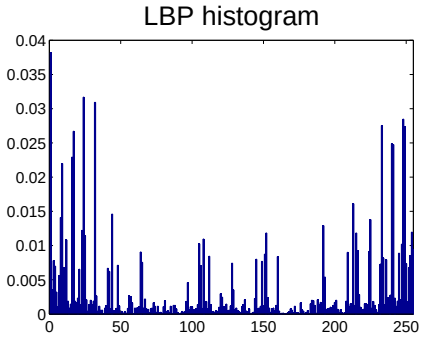
III. Performance Evaluation

In this section, we evaluate several classifiers on the detected regions. This aims at studying the performance of binary classifications on the segmented regions. We stress that the evaluation addresses the overall framework (segmentation and classification) for the problem at hand, namely detecting the building regions in an orthophoto. Firstly, we briefly describe the classifiers used. Secondly, we present the performance of the system for classifying the segmented regions. Thirdly, we present the performance of the system for building detection at pixel level using manually delineated building footprints. We consider six orthophotos depicting one hundred buildings.



(a)

(b)



(c)

Fig. 3. **(a)** Example of basic LBP operator. **(b)** Example of original image. **(c)** Example of LBP descriptor.

A. Machine Learning Approaches

Classification is the sub-field of supervised learning which is concerned with the prediction of the category of a given input. The classification model or classifier is trained using a labelled training set (i.e. a data set containing observations whose category membership is known). Each observation in the data set is a n-dimensional vector, and each element of the vector is called a *feature* (also *attribute* or *variable*). We have used four classifiers: K Nearest Neighbor (K-NN) with (K=1 and K=3), Support Vector Machines (SVM), and Classification Trees (C4.5). A brief description of all of them is included below.

a) Instance Based Learning: Instance Based Learning (IBL) belongs to the K-NN paradigm, a distance based classifier. It computes the distance of a new case to be classified to each of the observations in the database it uses as model and decides the class it will assign based on the K nearest cases. We have used the IBL algorithm described in [15], [16].

ISSN 2395-8618

b) Classification Trees: A Classification Tree is a classifier composed by nodes and branches which break the set of samples into a set of covering decision rules. In each node, a single test is made to obtain the partition. The starting node is called the root of the tree. In the final nodes or leaves, a decision about the classification of the case is made. In this work, we have used the C4.5 algorithm [17]. Note that C4.5 algorithm is also called J48.

c) Support vector machines (SVMs): SVMs are a set of related supervised learning methods used for classification and regression. In a bi-class problem, SVM views the input data as two sets of vectors (one set per class) in a n-dimensional space. The SVM will construct a separating hyperplane in that space, one which maximizes the margin between the two data sets. To calculate the margin, two parallel hyperplanes are constructed, one on each side of the separating hyperplane, which are "pushed up against" the two data sets. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the neighboring datapoints of both classes since, in general, the larger the margin the lower the generalization error of the classifier [18]. SVMs were extended to classify data sets that are not linearly separable through the use of non-linear kernels. In our work, we use non-linear SVMs with radial kernel.

d) Partial Least Square (PLS): The Partial Least Squares (PLS) classifier or regressor [19] is a statistical method that retrieves relations between groups of observed variables X and Y through the use of latent variables. It is a powerful statistical tool which can simultaneously perform dimensionality reduction and classification/regression. It estimates new predictor variables, known as components, as linear combinations of the original variables, with consideration of the observed output values. In our work, in both types of PLS, the number of latent components is fixed to 50.

B. Training

In order to get a training set which contains regions belonging to two classes (background and building) with ground-truth labels, we proceed as follows. The buildings are first manually delineated in each orthophoto. Each such ground-truth map is then intersected with the corresponding automatically over-segmented orthophoto. The label of any segmented region can be inferred by using the size of the overlap with the ground-truth building region. Any segmented region whose overlap with a building region exceeds 90% of its size will be labeled as building. Any segmented region whose overlap is below 3% of its size will have the non-building label. The regions that do not meet any of the two conditions are discarded and will not be used in as a training sample. The reason behind using these thresholds is the fact that an automatically segmented region may be shared by a building region and a background region. So it would be advantageous to use only quasi pure regions in the training set.

C. Region classification performance

It would be interesting to study the ability of descriptors and classifiers for recognizing the label of a given region. To this end, we collect a large number of labeled segmented regions, each is assumed to be a region that either belongs to the building category or to the background category. To achieve that we adopt the filtering process explained above. By adopting this filtering scheme, we collect 5656 regions with known labels. We then apply on them the 10-fold-cross validation scheme using the 1-NN, 3-NN, J48 and SVM classifiers. The obtained results are summarized in Tables I and II.

Table I depicts the number of misclassified regions and the rate of correct classification for three types of descriptors (color descriptor, LBP descriptor, and hybrid descriptor) and for four classifiers. The color descriptor is described by 805 features, the LBP descriptor by 315 features and the hybrid descriptor by 1120 features. In this evaluation, the use of hybrid descriptor has not improved the region classification over the color descriptor. However, the hybrid descriptor will improve the pixel classification as will be shown in the sequel. The main reason behind the difference in the obtained performance for regions and pixels is the fact that the SRM segmentation algorithm provides regions whose sizes (number of pixels) vary a lot. In other words, the region misclassification occurs mostly with small regions.

TABLE I
OVERALL REGION CLASSIFICATION RESULTS OBTAINED WITH 10-FOLD-CROSS VALIDATION.

		Color (805)		*LBP (315)*		*Hybrid (1120)*	
Classifier	Regions	Err.	Acc.	Err.	Acc.	Err.	Acc.
1-NN	5656	165	97.08	487	91.38	205	96.37%
3-NN	5656	144	97.45	424	92.50	205	96.37%
J48	5656	205	96.37	653	88.45	228	95.96%
SVM	5656	129	97.71	396	92.99	149	97.36%

Table II depicts the Recall, Precision, and F1 measure for building and background categories. From these two tables, we can observe that the non-linear SVM classifier has provided the best performance. We can also observe that the ability of all classifiers to discriminate background regions was better than that associated with building regions.

TABLE II
RECALL, PRECISION AND F1 FOR BACKGROUND AND BUILDING AND FOR ALL CLASSIFIERS.

	Background			*Building*		
Classifier	Recall	Precision	F1	Recall	Precision	F1
1-NN	98.1%	98.7%	98.4%	88.9%	84.5%	86.7%
3-NN	98.4%	98.8%	98.6%	89.9%	86.7%	88.3%
NB	88.9%	99.1%	93.7%	93.0%	50.0%	65.0%
J48	98.0%	98.0%	98.0%	83.1%	82.9%	83.0%
SVM	98.7%	98.8%	98.7%	89.7%	89.0%	89.3%

 http://dx.doi.org/10.17562/PB-51-2

ISSN 2395-8618

D. Segmentation and Classification Performance

In this section, we study the performance of the segmentation and classification at pixel level. Before presenting the quantitative evaluation, we first present in Figure 4 the results of building detection on the set of processed images using hybrid descriptors and SVM classifier. In each row of this figure, we show the initial orthophoto, the segmented image and the corresponding building roofs extraction where the final detected building boundaries are shown superimposed on the original orthophoto. Based on the visual evaluation of the results, we can state that the developed approach demonstrates excellent accuracy in terms of building boundary extraction, i.e., the majority of the building roofs present in the image are detected with good boundary delineation. Indeed, our method gives reliable results for complex environments having buildings with red and non-red rooftop buildings and/or buildings with the same color and texture with road areas.

In order to get a quantitative evaluation, we use the ground-truth building maps. The manually delineated buildings were used as a reference building set to assess the automated building-extraction accuracy. The extracted buildings and the manually delineated buildings are compared pixel-by-pixel. All pixels in the test image are categorized into four types.

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

From these measures it is straightforward to compute the following scores associated with the building regions in the test image: recall, precision, F1 measure, accuracy, and Matthews correlation coefficient (MCC). The MCC is in essence a correlation coefficient between the observed and predicted binary classifications; it returns a value between -1 and +1. A coefficient of +1 represents a perfect prediction, 0 no better than random prediction and -1 indicates total disagreement between prediction and observation.

Table III (upper part) illustrates the above scores when color descriptors are used. The classifier used is the non-linear SVM. Table III (lower part) illustrates the above scores when the hybrid descriptors are used. In this table, the evaluation adopted a similar protocol to the 6-fold cross validation in the sense that each orthophoto is used as a test set and the remaining orthophotos (i.e., their descriptors retained in the training set) are used as training samples.

Tables IV, V, VI, and VII illustrate the same evaluation obtained with 1-NN, 3-NN, tree, and PLS classifiers, respectively. We can observe that the use of the hybrid descriptor has improved the average performance of building detection. This is true for all the classifiers used. For example, let's consider SVM classifier and **orthohpoto5**. The rate of correct classification of its pixels is 89.6 % when only color information is used. This rate becomes 93.9 % when the hybrid descriptor is used. Since the size of **orthophoto5** is 652392 pixels this means that with the hybrid descriptor 28053 more pixels are correctly classified. We can also observe that the non-linear SVM and 3-NN classifiers adopting the hybrid descriptor give the best performances. It should be noticed that all results are obtained by using a binary classification without any post-processing.

TABLE III
RECALL, PRECISION, F1, AND MATTHEWS CORRELATION COEFFICIENT (MCC) CORRESPONDING TO A BINARY CLASSIFICATION (PIXEL LEVEL) USING BOTH COLOR AND HYBRID DESCRIPTORS (COLOR HISTOGRAMS WITH LBPS). THE RESULTS ARE OBTAINED WITH SVM.

Image	Recall (%)	Precision (%)	F1-measure (%)	Accuracy (%)	MCC
Color descriptor					
Orthophoto1	78.1	89.2	83.3	96.8	0.82
Orthophoto2	88.7	94.0	91.2	95.4	0.88
Orthophoto3	80.6	85.3	82.9	92.3	0.78
Orthophoto4	92.4	87.8	90.0	96.5	0.88
Orthophoto5	77.4	82.4	79.8	89.6	0.73
Orthophoto6	95.7	76.4	84.9	94.5	0.82
Average	85.5	85.8	85.4	94.2	0.82
Hybrid descriptor					
Orthophoto1	88.1	90.0	89.0	97.8	0.88
Orthophoto2	93.0	95.8	94.4	97.0	0.92
Orthophoto3	83.9	91.9	87.7	94.5	0.84
Orthophoto4	93.6	93.8	93.7	97.8	0.92
Orthophoto5	85.7	91.0	88.3	93.9	0.84
Orthophoto6	95.8	87.4	91.4	97.1	0.90
Average	90.0	91.6	90.8	96.4	0.88

TABLE IV
RECALL, PRECISION, F1, AND MATTHEWS CORRELATION COEFFICIENT (MCC) CORRESPONDING TO A BINARY CLASSIFICATION (PIXEL LEVEL) USING BOTH COLOR AND HYBRID DESCRIPTORS (COLOR HISTOGRAMS WITH LBPS). THE RESULTS ARE OBTAINED WITH 1-NN.

Image	Recall (%)	Precision (%)	F1-measure (%)	Accuracy (%)	MCC
Color descriptor					
Orthophoto1	74.4	86.1	79.8	96.1	0.78
Orthophoto2	89.0	93.0	91.0	95.3	0.88
Orthophoto3	95.5	84.7	89.8	94.9	0.87
Orthophoto4	92.5	90.7	91.6	97.1	0.90
Orthophoto5	76.7	84.0	80.2	89.9	0.74
Orthophoto6	89.6	86.6	88.1	96.1	0.86
Average	86.3	87.5	86.7	94.9	0.84
Hybrid descriptor					
Orthophoto1	93.9	86.8	90.2	97.9	0.89
Orthophoto2	95.4	92.0	93.7	96.5	0.91
Orthophoto3	98.0	87.2	92.3	96.2	0.90
Orthophoto4	93.9	90.8	92.3	97.3	0.91
Orthophoto5	86.5	86.5	86.5	92.8	0.82
Orthophoto6	92.3	86.0	89.1	96.3	0.87
Average	93.3	88.2	90.7	96.2	0.88

IV. CONCLUSION

In this paper, we have introduced a method which accounts for automatic and accurate multiple objects recognition

ISSN 2395-8618

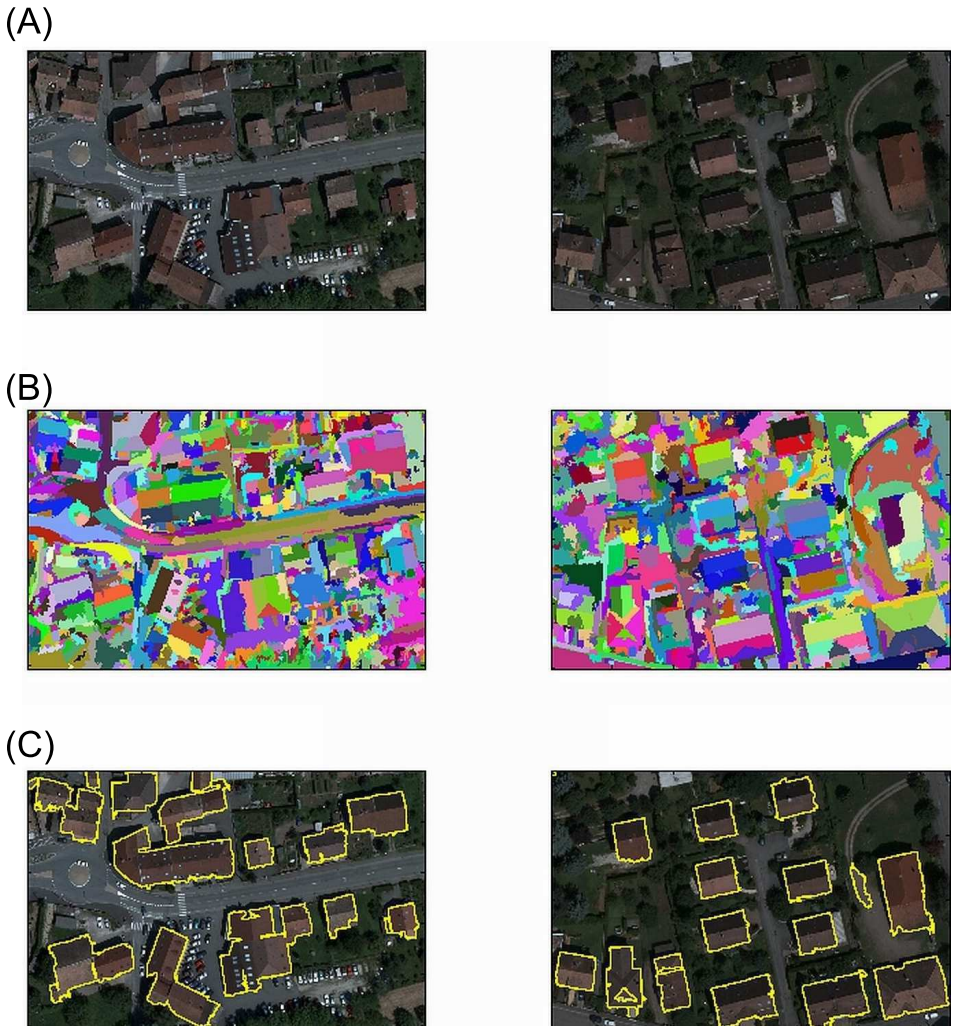


Fig. 4. **(A)** Orthophotos. **(B)** Segmented orthophotos. **(C)** Countours of detected roof regions.

TABLE V
RECALL, PRECISION, F1, AND MATTHEWS CORRELATION COEFFICIENT (MCC) CORRESPONDING TO A BINARY CLASSIFICATION (PIXEL LEVEL) USING BOTH COLOR AND HYBRID DESCRIPTORS (COLOR HISTOGRAMS WITH LBPS). THE RESULTS ARE OBTAINED WITH 3-NN.

Image	Recall (%)	Precision (%)	F1-measure (%)	Accuracy (%)	MCC
Color descriptor					
Orthophoto1	79.7	87.1	83.3	96.7	0.82
Orthophoto2	89.4	95.5	92.3	96.0	0.90
Orthophoto3	98.3	85.6	91.5	95.8	0.89
Orthophoto4	93.8	93.9	93.8	97.9	0.93
Orthophoto5	75.8	82.5	79.0	89.3	0.72
Orthophoto6	87.2	86.7	87.0	95.8	0.84
Average	87.4	88.5	87.8	95.2	0.85
Hybrid descriptor					
Orthophoto1	94.3	86.4	90.2	97.9	0.89
Orthophoto2	93.9	92.4	93.1	96.3	0.91
Orthophoto3	98.1	86.4	91.9	96.0	0.89
Orthophoto4	95.8	94.3	95.0	98.3	0.94
Orthophoto5	85.9	87.0	86.4	92.8	0.82
Orthophoto6	93.4	86.3	89.7	96.5	0.88
Average	93.6	88.8	91.1	96.3	0.89

TABLE VI
RECALL, PRECISION, F1, AND MATTHEWS CORRELATION COEFFICIENT (MCC) CORRESPONDING TO A BINARY CLASSIFICATION (PIXEL LEVEL) USING BOTH COLOR AND HYBRID DESCRIPTORS (COLOR HISTOGRAMS WITH LBPS). THE RESULTS ARE OBTAINED WITH TREE (C.45).

Image	Recall (%)	Precision (%)	F1-measure (%)	Accuracy (%)	MCC
Color descriptor					
Orthophoto1	68.7	90.8	78.2	96.0	0.77
Orthophoto2	81.6	90.5	85.8	92.8	0.81
Orthophoto3	67.0	91.7	77.4	90.9	0.73
Orthophoto4	86.7	92.0	89.3	96.4	0.87
Orthophoto5	68.6	84.2	75.6	88.2	0.68
Orthophoto6	83.2	89.5	86.2	95.7	0.84
Average	76.0	89.8	82.1	93.3	0.78
Hybrid descriptor					
Orthophoto1	88.3	88.3	88.3	97.6	0.87
Orthophoto2	86.4	94.5	90.3	95.0	0.87
Orthophoto3	75.8	88.8	81.8	92.1	0.77
Orthophoto4	86.3	91.4	88.8	96.3	0.87
Orthophoto5	72.3	86.0	78.6	89.5	0.72
Orthophoto6	83.2	90.8	86.8	95.9	0.85
Average	82.0	90.0	85.8	94.4	0.82

 http://dx.doi.org/10.17562/PB-51-2

ISSN 2395-8618

TABLE VII
RECALL, PRECISION, F1, AND MATTHEWS CORRELATION COEFFICIENT (MCC) CORRESPONDING TO A BINARY CLASSIFICATION (PIXEL LEVEL) USING BOTH COLOR AND HYBRID DESCRIPTORS (COLOR HISTOGRAMS WITH LBPS). THE RESULTS ARE OBTAINED WITH NON LINEAR PARTIAL LEAST SQUARE (PLS).

Image	Recall (%)	Precision (%)	F1-measure (%)	Accuracy (%)	MCC
Color descriptor					
Orthophoto1	71.9	90.1	80.0	96.3	0.79
Orthophoto2	86.3	95.8	90.8	95.3	0.88
Orthophoto3	78.5	94.3	85.7	93.9	0.82
Orthophoto4	91.1	92.4	91.8	97.2	0.90
Orthophoto5	72.9	90.6	80.8	90.8	0.76
Orthophoto6	83.6	89.3	86.4	95.7	0.84
Average	80.7	92.2	85.9	94.9	0.83
Hybrid descriptor					
Orthophoto1	82.9	91.1	86.8	97.4	0.85
Orthophoto2	93.3	95.7	94.5	97.1	0.93
Orthophoto3	86.4	93.1	89.6	95.3	0.87
Orthophoto4	92.8	93.7	93.3	97.7	0.92
Orthophoto5	84.4	88.7	86.5	93.0	0.82
Orthophoto6	93.7	88.0	90.8	96.9	0.89
Average	88.9	91.7	90.3	96.2	0.88

from images. Unlike methods that rely on the interactive image segmentation, our approach does not require any user interaction or any setting of initial algorithm parameters (a threshold of similarity for example). The proposed method involves a supervised scheme in which offline manual delineation and automatic segmentation are carried out to build descriptors and classifiers. At running time, after an over-segmentation of the image, one can classify the segmented regions as object parts or background image using region classification.

In order to show its performance, the proposed method was applied to extract building roofs from orthophotos. This problem is very challenging given the complexity of objects in the orthophotos. While orthophotos construction used Digital Surface Maps, our adopted building detection used image information only. Future work may investigate the use of covariance matrix descriptors as hybrid descriptors. Furthermore, we may investigate whether the use of superpixels algorithms (e.g., [20], [21]) could improve the initial segmentation. The choice of more adapted color space could be also an interesting way to improve the results.

REFERENCES

[1] O. Tournaire, M. Brédif, and M. Boldo, D.and Durupt, "An efficient stochastic approach for buildings footprint extraction from digital elevation models," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, pp. 317–327, 2010.

[2] O. Wang, S. K. Lodha, and D. P. Helmbold, "A Bayesian approach to building footprint extraction from aerial LIDAR data," in *International Symposium on 3D Data Processing, Visualization, and Transmission*, 2006.

[3] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut—interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics (SIGGRAPH'04), New York, NY, USA*, 2004.

[4] K. McGuinness and N. E. O'Connor, "A comparative evaluation of interactive segmentation algorithms," *Pattern Recognition*, vol. 43, pp. 434–444, 2010.

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

[6] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 16, pp. 641–647, 1994.

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

[8] T. Ojala, M. Pietikäinen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, 2002.

[9] V. Takala, T. Ahonen, and M. Pietikäinen, "Block-based methods for image retrieval using local binary patterns," in *Image Analysis, SCIA*, vol. LNCS, 3540, 2005.

[10] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[11] M. Bereta, P. Karczmarek, W. Pedrycz, and M. Reformat, "Local descriptors in application to the aging problem in face recognition," *Pattern Recognition*, vol. 46, pp. 2634–2646, 2013.

[12] D. Huang, C. Shan, M. Ardabilian, and Y. Wang, "Adaptive particle sampling and adaptive appearance for multiple video object tracking," *IEEE Trans. on Systems, Man, and Cybernetics-Part C: Applications and reviews*, vol. 41, no. 6, pp. 765–781, November 2011.

[13] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in *Faces in Real-Life Images Workshop in ECCV*, 2008.

[14] D. Pollard, *A user's guide to measure theoretic probability*. Cambridge University Press, 2002.

[15] D. Aha, D. Kibler, and M. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, pp. 37–66, 1991.

[16] D. Mena-Torres, J. Aguilar-Ruiz, and Y. Rodriguez, "An instance based learning model for classification in data streams with concept change," in *11th Mexican International Conference on Artificial Intelligence (MICAI)*, 2012, pp. 58–62.

[17] J. Quinlan, *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[18] D. Meyer, F. Leisch, and K. Hornik, "The support vector machine under test," *Neurocomputing*, vol. 55, pp. 169–186, 2003.

[19] R. Rosipal and N. Kramer, *Subspace, Latent Structure and Feature Selection Techniques*. Springer, 2006, ch. Overview and recent advances in partial least squares, pp. 34–51.

[20] A. Levinshtein, A. Stere, K. Kutulakos, D. Fleet, S. Dickinson, and K. Siddiqi, "Turbopixels: Fast superpixels using geometric flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2290–2297, 2009.

[21] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.