

Zernike Moments and SVM for Shape Classification in Very High Resolution Satellite Images

Habib Mahi¹, Hadria Isabaten², and Chahira Serief¹

¹Earth Observation Division, Centre of Space Techniques, Algeria

²Faculty of Computing Science, Boudiaf University, Algeria

Abstract: *In this paper, a Zernike moments-based descriptor is used as a measure of shape information for the detection of buildings from Very High Spatial Resolution (VHSR) satellite images. The proposed approach comprises three steps. First, the image is segmented into homogeneous objects based on the spectral and spatial information. MeanShift segmentation method is used for this end. Second, a Zernike feature vector is computed for each segment. Finally, a Support Vector Machines (SVM)-based classification using the feature vectors as inputs is performed. Experimental results and comparison with Environment for Visualizing Images (ENVI) commercial package confirm the effectiveness of the proposed approach.*

Keywords: *Zernike moments, building extraction, mean shift, SVM, VHSR satellite images.*

Received September 23, 2011; accepted December 30, 2011; published online January 29, 2013

1. Introduction

Urban buildings are topographic features of great importance for urban planning, environmental management and monitoring. With the emergence of commercial satellites (e.g., Ikonos and Quickbird) with on-board sensors characterized by Very High Spatial Resolution (VHSR) (from 2.5 to 0.60m), identification and localization of these features become conceivable. VHSR sensors, in fact, provide images with significant amount of geometrical details that yield new kinds of information [17]. However, although the reduction of the satellite image pixel size as a result of the improvement in spatial resolution has facilitated the recognition of urban man-made objects such as buildings and roads, it poses a challenge to classical techniques for image analysis such as pixel-based classifications [17]. Indeed, as the spatial resolution increases, buildings may have very complicated appearances and may have complex structures with very different spectral signatures due to their large variability in material composition, color, shape, size and orientation. Furthermore, due to technical constraints, the spectral resolution of VHSR sensors is limited in comparison with other sensors, such as Landsat TM [14]. Therefore, different materials appear with similar spectral signatures, making spectral discrimination between objects difficult. Consequently, classification algorithms based on single-pixel analysis and using only spectral information do not give the desired result when applied to VHSR images due to large within-class spectral variations and between-class

spectral confusions that characterize urban man-made features.

If advantage is to be taken of the spatial resolution of VHSR sensors, it is therefore necessary to expand the object feature base to include geometrical characteristics in addition to spectral ones. Object-based classification approaches offer this possibility. These methods are generally based on the concept that important semantic information is not represented in single pixels alone but in meaningful image objects of different shapes and sizes and their mutual relations (i.e., context). Each object can then be modeled both with shape and topological measures which can be used and integrated with spectral and textural features to improve the classification accuracy [3]. A segmentation of the image is carried out as a first step in the analysis. After segmentation, every image segment becomes a homogeneous unit for which a set of features (spectral, textural, spatial...) can be calculated. These features can then be used for segment classification. In this context, a variety of methods with varying degrees of success has been recently developed for buildings extraction from VHSR satellite images. The most common is the method implemented in "e-cognition" software [15]. This method can be used to extract buildings in urban areas according to the author. In [23], a combined fuzzy pixel-based and object-based approach for classification of urban land cover from high-resolution multispectral image data was presented. This combined pixel/object approach has demonstrated to be able to accurately identify buildings and roads in urban areas. In [2], a new multispectral segmentation

method is proposed to automate the choice of segmentation parameters. It uses the valuable information content of existing maps. Then, the segmented image is used in a rule-based classification method that considers spectral, geometric, and contextual information.

Nevertheless, the main drawback shared by the various object-based approaches proposed in the literature is related to the set of attributes used to characterize the segment object shape. The attributes frequently used to describe the shape are area, perimeter, convex hull, and compactness. However, these spatial attributes do not exhibit sufficient discrimination between object shapes. Moreover, they are sensitive to object boundary deformation and are not invariant under geometrical transformations.

A new object-based classification method is introduced in this paper for the automated detection of buildings with complex shapes in VHSR images by exploiting shape information. The main contribution of this work is the use of Zernike moments as shape descriptor in the classification process. This choice is motivated by the fact that Zernike moments are powerful region-based shape descriptors that are invariant against linear transformations and especially against object boundary deformation. In addition, Zernike moments have been extensively used in a wide range of two- and three-dimensional pattern recognition and computer vision applications [1, 13, 18, 19] and they have been shown to be superior in terms of their robustness to image noise and distortions, expression efficiency, fast computation, and ability to provide faithful image representation [25].

In the proposed approach, image objects segments extraction is performed using the Mean Shift segmentation algorithm [10], as a first step. Then, a Zernike feature vector is constructed for each of the segments. Finally, a Support Vector Machines (SVMs)-based classification using the feature vectors as input instead of the original objects is carried out to assign a class label to each of the segments. Compared to well-known machine-learning methods (both neural and kernel methods) such as neural networks and fuzzy logic approaches, SVMs have proven to be superior in terms of efficiency and robustness [4].

The rest of paper is organized as follows: Section 2 describes the Mean Shift algorithm. Section 3 presents a detailed description of the Zernike moment-based shape invariant descriptor. Section 4 describes briefly the SVM-based classification method. Data sets used for the experiments and reports on experimental results are presented in section 5. A summary and conclusions are presented in section 6.

2. Image Segmentation

Proposed in 1975 by Fukunaga and Hostetler [12] and largely forgotten until Cheng's paper [7] rekindled

interest in it for the purpose of image analysis and more recently popularized in the computer vision literature by Comaniciu and Meer [10], for image segmentation, the Mean Shift represents a simple iterative nonparametric procedure for density mode seeking. The main idea behind the Mean Shift procedure is to treat the points in the d -dimensional feature space as an empirical probability density function, where the densest regions in the feature space correspond to the local maxima or modes of the underlying distribution. Let $\{x_i\}_{i=1..n}$ be set of n points in the d -dimensional Euclidean space R^d . The multivariate kernel density is given by:

$$\hat{f}_K(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \tag{1}$$

where h is the radius of hypersphere at point x and called bandwidth parameter. $K(x)$ is termed the kernel and is defined as:

$$K(x) = c_{k,d} k(\|x\|^2) \tag{2}$$

The function $k(x)$ is called the profile of the kernel $K(x)$, and $c_{k,d}$ is the normalization constant which makes $K(x)$ integrate to one. The modes of the density function are located at the zeros of the gradient function $\nabla \hat{f}_K(x)$. The gradient of the density estimator is given by:

$$\hat{f}_K(x) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^n (x-x_i) k\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \tag{3}$$

The gradient function can also be written as:

$$\begin{aligned} \nabla \hat{f}_G(x) &= \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^n (x-x_i) g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \\ &= \frac{2c_{k,d}}{nh^{d+2}} \left[\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \right] \left[\frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \right] \end{aligned} \tag{4}$$

where $g(z)=-k'(z)$. The first term is proportional to density estimate at x computed with kernel $G(x)=c_{g,d} g(\|x\|^2)$. The second square bracket term is called the Mean Shift vector and it is given by:

$$M_h(x) = \left[\frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \right] \tag{5}$$

We can also formulate $M_h(x)$ as the ratio of $\nabla \hat{f}_K(x)$ to $\hat{f}_G(x)$:

$$M_h(x) = \left[\frac{c_{g,d} h^2}{2c_{k,d}} \right] \left[\frac{\nabla \hat{f}_K(x)}{\hat{f}_G(x)} \right] \tag{6}$$

From this last formulation, we can see that when:

$$M_h(x) \rightarrow 0, \nabla f_K(x) \rightarrow 0.$$

For color image, the feature space is composed of the spatial domain and the range or color domain. The spatial domain is composed by pixel coordinates vectors:

$$x^s = \begin{bmatrix} x \\ y \end{bmatrix} \quad (7)$$

and the corresponding color information in the range domain generally, of three dimensions. For color images using RGB space for example, the range vector is expressed as:

$$x^r = \begin{bmatrix} r(x^s) \\ g(x^s) \\ b(x^s) \end{bmatrix} \quad (8)$$

where $r(x^s)$, $g(x^s)$, $b(x^s)$ are respectively the red, green and blue components of the pixel x^s . Considering now the Gaussian kernel:

$$G(x) = (2\pi)^{-d/2} \exp\left(\frac{-\|x\|^2}{2}\right) \quad (9)$$

with its profile:

$$g(x) = \exp\left(\frac{-x}{2}\right) \quad (10)$$

and knowing that the two previous domains are independent, the new kernel in the join domain with the bandwidths h^s and h^r respectively, is defined as the product of two different radially symmetric kernels (superscript s will refer to the spatial domain, and r to the range domain):

$$g = g^s g^r \quad P \quad g' = [g^s]' [g^r]' \quad (11)$$

The Mean Shift vector becomes $M_{h^s, h^r}(x)$ and is expressed as:

$$M_{h^s, h^r}(x) = \frac{\sum_{i=1}^n x_i \exp\left(-\frac{1}{2} \left\| \frac{x^s - x_i^s}{h^s} \right\|^2\right) \exp\left(-\frac{1}{2} \left\| \frac{x^r - x_i^r}{h^r} \right\|^2\right)}{\exp\left(-\frac{1}{2} \left\| \frac{x^s - x_i^s}{h^s} \right\|^2\right) \exp\left(-\frac{1}{2} \left\| \frac{x^r - x_i^r}{h^r} \right\|^2\right)} - x \quad (12)$$

The Mean Shift segmentation procedure follows the following steps:

1. Convert image from RGB space to L^*a^*b space.
2. Initialize h^s , h^r , ε and s .
3. Build k hyperspheres with radius h^s and centers $\{x_1, x_2, \dots, x_k\}$ within the converted image. The centers must verify: $\{d(x_i, x_j) = 2h^s; i, j = 1..k; i \neq j\}$ with d is the Euclidean distance.
4. For each hypersphere at center x_i

Repeat

Compute the Mean Shift vector $M_{h^s, h^r}(x_i^t)$.

Translate density estimate $x_i^{t+1} = x_i^t + M_{h^s, h^r}(x_i^t)$.

Until $(x_i^{t+1} - x_i^t) < \varepsilon$.

5. The final positions of the centers constitute the set of modes.
6. Group adjacent modes if the difference is less than s .
7. Replace each pixel with its mode.

3. Invariant Zernike Moments-based Shape Descriptors

On VHRS satellite images, similar objects may have different sizes and orientations. Consequently, an effective shape description constitutes a key part in the classification process. For this purpose, a favorable shape descriptor should have high discrimination ability so that it can group similar shapes together and separate dissimilar shapes into different groups. It is also expected that a reliable shape descriptor should be invariant to common geometrical transformations and robust to shape boundary deformations. During the past decades, numerous shape descriptors have been proposed in the context of content-based image retrieval and pattern recognition applications. According to [5, 20, 27], these can be broadly categorized into two groups, namely, contour- and region-based descriptors. Contour-based shape descriptors such as Fourier descriptors [22], ignore potentially important information in the shape interior. Consequently, they are sensitive to variations of object boundaries and cannot deal with disjoint shapes where contour information may not be available. Compared with contour-based ones, region-based methods are more suitable for general applications [27]. Indeed, region-based shape descriptors exploit information from both boundaries and interior regions of the shape. Among them, the best-known ones are moments-based descriptors which have been very popular since they were first introduced in the 60's [16]. These include geometric, legendre [25] and Zernike moments [24]. Comparative studies [5, 27], [28] show that Zernike moments descriptor outperforms the other moments descriptors in terms of invariance, computation complexity, compact representation and robustness to noise and distortions.

3.1. Zernike Moments

The two-dimensional complex Zernike moments of a digital image with current pixel $P(x,y)$ are defined as:

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y P(x,y) V_{nm}^*(x,y), \quad x^2 + y^2 \leq 1 \quad (13)$$

n and m are generally called order and repetition, respectively. The order n is a nonnegative integer, and

the repetition m is an integer subject to the conditions: $n-|m|$ even, $0 \leq |m| \leq n$. $V_{nm}(x,y)$ are the Zernike polynomials defined in polar coordinates as:

$$V_{nm}(x,y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) \exp(jm\theta) \quad (14)$$

where ρ and θ represent polar coordinates over the unit disk and $R_{nm}(\rho)$ is the real-valued radial polynomial of ρ given as follows:

$$R_{nm}(x,y) = \sum_{s=0}^{\frac{n-|m|}{2}} \frac{(-1)^s (n-s)! \rho^{n-2s}}{s! \left(\frac{n+|m|-s}{2}\right)! \left(\frac{n-|m|-s}{2}\right)!} \quad (15)$$

To calculate the Zernike moments, the image (or region of interest) is first mapped to the unit disc using polar coordinates, where the centre of the image is the origin of the unit disc. Those pixels falling outside the unit disc are not used in the calculation. The coordinates are then described by ρ which is the length of the vector from the origin to the coordinate point and θ which is the angle from the x axis to the vector ρ , by convention measured from the positive x axis in a counter clockwise direction. Since their moment functions are defined using a polar coordinate representation of the image space, Zernike moments are by nature rotational invariants where their magnitude values are unaffected and remain the same for original and rotated image. However, this coordinate representation does not easily yield translation and scale invariance [8]. One of the indirect approaches to achieve scale and translation invariance is through expressing Zernike moments using centralized and normalized regular moments.

3.2. Zernike Moments-based Shape Descriptor Construction

Real-valued radial polynomials defined by equation 15 could be rewritten in a simple way as follows [9]:

$$R_{nm}(\rho) = \sum_{k=m; n-k=\text{even}}^n B_{nmk} \rho^k \quad (16)$$

The coefficients B_{nmk} are defined as:

$$B_{nmk} = \frac{(-1)^{\frac{n-k}{2}} \left[\frac{n+k}{2}\right]!}{\left[\frac{n-k}{2}\right]! \left[\frac{m+k}{2}\right]! \left[\frac{k-m}{2}\right]!} \quad (17)$$

Zernike moments can then be expressed using central and normalized regular moments as [21]:

$$Z_{nm} = \frac{n+1}{\pi} \sum_{k=m}^{k_l} B_{nmk} \mu_{k-2j-m+j1, 2j+m-j1} \quad (18)$$

where μ_{pq} are the central and normalized moments of order $p+q$ and k_l is given by:

$$k_l = n \sum_{j=0}^q \sum_{l=0}^m (-1)^{j+l} \frac{j!}{[j-q]! q!} \frac{j!}{[j-l-m]! m!} \quad (19)$$

Since Zernike moments are defined over a unit disk,

this disk must be specified before moments can be calculated. In our implementation, the radius of a circle is determined to enclose the shape completely from the centroid of the binarized shape in the segmented image to the outer most pixel of the shape. The unit disk is then centered on the shape centroid. This makes the obtained moments scale and translation invariant. Rotation invariance is achieved by only using magnitudes of the moments. The magnitudes are then normalized into $[0, 1]$ by dividing them by the mass of the shape.

A feature vector \vec{v} is constructed with the Zernike moments for each segment as $\vec{v} = (v_1, v_2, \dots, v_L) = (|Z_{20}|, |Z_{22}|, |Z_{31}|, \dots, |Z_{n_{max}, n_{max}}|)$, where L is the length of the feature vector and n_{max} is the optimal order of Zernike moments. In this study, the shape descriptor is required to have high discrimination ability and to be invariant to common geometrical transformations and especially robust to shape boundary deformation. Therefore, two experiments were performed to select the values of Zernike order n and repetition m that achieve a compromise between these requirements. Set CE-1 part B of MPEG-7 region-based shape database has been used in the experiments. It consists of 1406 binary shape. In the first experiment, the discrimination ability of the Zernike shape descriptor is assessed to determine the optimal value of order n . The binary image “Device8_20” shown on Figure 1-a has been used as original image. Figure 2 shows some sample shapes from a set of 105 binary shapes taken as test images.

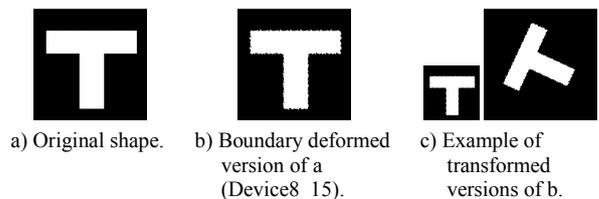


Figure 1. Examples of boundary deformation and geometrical transformations of the shape.



Figure 2. Examples from MPEG-7 CE1 B data set.

The absolute root-mean-square-error (Armse) is then used as similarity measure between the original image and the test images. Hence, we measure the distance between two Zernike feature vectors using Armse function given by:

$$Armse(\%) = \sqrt{\frac{\sum_{i=1}^L [ZV_i^T - ZV_i^0]^2}{L}} \times 100\% \quad (20)$$

where ZV^0 is the Zernike feature vector of the original image and ZV^T is the Zernike feature vector of the test image.

The Armse values computed between the original image and each of the 105 binary images are plotted in Figure 3 for different values of order n up to four. A big value of Armse means a higher discriminative power. The graphs shown on Figure 3 suggest that the higher discriminative power is achieved for orders $n=8, 9, 10$ because they give the bigger values of Armse for all the test images.

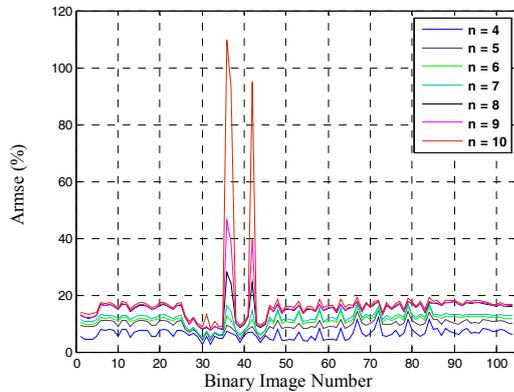


Figure 3. Plot of the armse values computed between the original image and each of the 105 binary images for different values of order n up to four.

To determine which of these values of order n achieves a compromise between the discrimination ability of the Zernike shape descriptor and the invariance to rotation, scale changes and shape boundary deformation, the second experiment is performed using two binary shapes from the database. “Device8_20” has also been used as original shape (reference object) for this experiment. Figure 1-b shows the binary shape “Device8_15” representing the boundary deformed version of the original shape “Device8_20”. Several transformed versions of the deformed boundary shape are then created according to a combination of rotation through different angles from 5° to 155° at the interval of 5° and scale changes by different factors from 0.2 to 1.8 at the interval of 0.3. In this way, 36 images are obtained. Figure 1-c gives an example of transformed versions. Figure 4 is a plot of the Armse computed between the original binary image and each of the 36 transformed images for different values of order n up to four. For this experiment, a small value of the Armse means good performance in terms of invariance against geometrical transformation and boundary deformation. It can be seen from the graphs shown on Figure 4 that the value $n=9$ gives the small Armse and consequently achieves the compromise between discrimination ability and invariance against common geometrical transformations and boundary deformation.

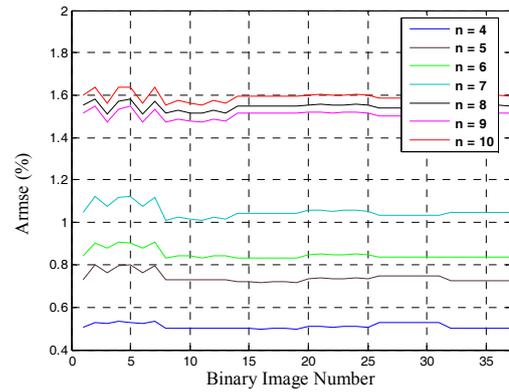


Figure 4. Plot of the armse computed between the original binary image and each of the 36 transformed images for different values of order n up to four.

4. Zernike Feature Vectors Classification

After calculating the Zernike feature vectors, a class label can be assigned to each of the segments in the image by performing a classification method using these feature vectors as inputs instead of the image segments. In this way, only shape information measured by Zernike moments descriptor is used in the classification process. Developed by Vapnik [26], SVMs have been proposed as an efficient method for pattern classification and nonlinear regression. Their appeal lies in their strong connection to the underlying statistical learning theory where an SVM is an approximate implementation of the method of Structural Risk Minimization (SRM) [26]. This principle states that a better solution in terms of generalization capabilities can be found by minimizing an upper bound of the generalization error.

SVMs have many attractive features. For instance, the solution of quadratic programming problem is globally optimized. In addition, SVMs can handle large input spaces, can effectively avoid overfitting by controlling the margin, and can automatically identify a small subset made up of informative points, namely Support Vectors (SVs) [4]. Consequently, they have been used in a wide range of applications and their popularity within the remote sensing community is constantly on the increase [11], due to their properties and intrinsic effectiveness. For greater detail on SVMs, we refer the reader to [11, 26].

5. Experimental Results and Discussion

The first experiment was conducted using a synthetic gray-scale image with sixteen distinct regions (Figure 5-a). The labeled regions image is shown on Figure 5-b. In order to classify this image using only shape information, the first step consists to extract homogeneous segments by applying the Mean Shift segmentation algorithm. The result is illustrated on Figure 5-c.

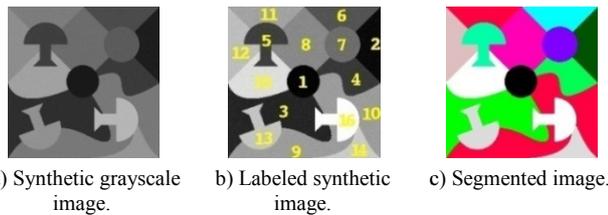


Figure 5. Synthetic gray-scale image and its Mean Shift segmentation used in the first experiment.

After the image was segmented, a Zernike feature vector is computed for each of the segments in the image. Figures 6-a, 6-b and 6-c illustrate the second, fourth and eighth component of each feature vector of all segments in the image respectively. Then, these vectors are used as inputs to the SVM classifier with an RBF kernel. Seven samples are selected according to the segment shape for training the SVM (see Figure 7-a). The classification result is shown on Figure 7-b. It can be seen that all the segments having the same shape are correctly classified with the proposed method.

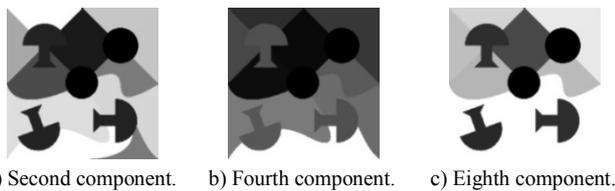


Figure 6. Illustration of the second, fourth and eighth component of each feature vector of all segments in the synthetic image.

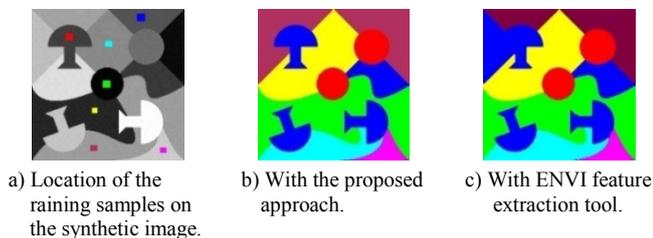


Figure 7. Classification results obtained by the proposed approach and ENVI feature extraction tool with training set.



Figure 8. Sub-scene of a Quickbird image.

The second experiment was conducted using a sub-scene of Quickbird image datasets of Algiers (northern Algeria). This one was acquired on August 23 2003 and consists of three multispectral images with a spatial resolution of 0.61m. Figure 8 shows the sub-scene of size 550×450 pixels.

The first step in the proposed method consists of extracting homogeneous segments by applying the Mean Shift segmentation algorithm. Thus the

performance of the segmentation method has a great effect on the shape descriptor and the classification result. Bearing this in mind, we carried out several trials on the first sub-scene to determine the optimal values of the Mean Shift parameters which give the best segmented image. For Mean Shift, we have four parameters: the spatial bandwidth h^s , the range bandwidth h^r , the merging threshold s , and the stopping threshold ε . The last parameter was kept fixed to 0.0005 in all tests. Several combinations of the three parameters h^s , h^r and s were done as follows: Tests from 1 to 5: $h^r=3$, $s=5$, h^s gradually increased from 3 to 7 at the interval of 1. Tests from 6 to 12: $h^s=7$ (best segmentation result in the previous test set), $h^r=3$, s gradually increased from 4 to 10 at the interval of 1. Tests from 13 to 17: $h^s=7$, $s=8$ (best segmentation result in the second test set), h^r gradually increased from 4 to 10 at the interval of 1. In this way, seventeen segmentation results were obtained. Quantitative assessments of these results were done to select the optimal values of the parameters h^s , h^r and s . The quantitative assessment makes possible to quantify the quality of segmentation result. For this end, we used Borsotti criterion [6], Q , which is given by:

$$Q = \frac{\sqrt{R}}{10000 \times N \times M} \sum_{i=1}^R \left[\frac{e_i^2}{1 + \log A_i} + \left(\frac{\psi(A_i)}{A_i} \right)^2 \right] \quad (21)$$

where $N \times M$ is the size of the segmented image S , R is the number of segments of S , A_i is the area of the i^{th} segment, e_i is the color error of the segment i and $\psi(A_i)$ the number of segments with the same area. A small value of Q means a good segmentation quality. The value of Borsotti criterion computed for each of the segmentation results corresponding to the combinations of the three parameters shows that the smaller value is equal to 5.0955.



Figure 9. The best segmentation result according to the Borsotti criterion.

This value corresponds to the parameters values $h^s=7$, $h^r=3$ and $s=9$. Figure 9 shows the best segmentation result corresponding to these values.

After the image was segmented, a Zernike feature vector of thirty one components is computed for each segment in the image. Then, these vectors are used as inputs to the SVM classifier using RBF kernel with a parameter γ equal to 0.091 and a cost parameter C equal to 100. The classification is performed according to eight classes. The training samples set

consists of pixels selected from buildings, roads, grass areas and bare soil. Figure 10 shows the location of the selected samples on subsets of the original image. The classification map of the sub-scene is illustrated in Figure 11.



Figure 10. Location of training samples set on subsets of the original image.

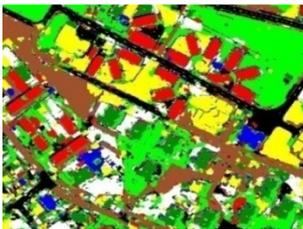


Figure 11. Classification map of the quickbird sub-scene.



Figure 12. Extracted buildings classes from the first sub-scene with the original image as background.

Figure 12 shows the extracted buildings classes with the original image as background: buildings #1 (red color), buildings #2 (green color) and buildings #3 (blue color). It can be seen that overall most of building segments were correctly identified with a good visual quality. Indeed, the Building segments are homogeneous and look like the reality. By visual inspection, the detection accuracy is more than 83% in the case of buildings #1 and is more than 84% in the case of buildings #2. Here the accuracy is defined as the ratio between number of buildings classified correctly and total buildings in the sub-scene. However, the analysis of the results shows that the most important source of errors lies with confusion between buildings and grass areas (red and blue dashed regions). These confusions are mainly due to the fact that grass areas and buildings have similar shapes. This problem can be avoided by computing the Normalized Difference Vegetation Index (NDVI). Indeed, the regions whose average NDVI values are higher than an NDVI threshold are denoted as vegetation areas. In this way, vegetation areas can be eliminated. It can also be seen

from the results that the class building #4 was misclassified. This drawback is mainly related to the segmentation result. In fact, the object in the original image is formed by heterogeneous spectral regions and it is very difficult to find the best set of parameters values to obtain one segment for this object without loss of other best segments.

• Comparison with ENVI Feature Extraction Tool

To assess the performance of the proposed method, a comparative evaluation against the ENVI feature extraction tool is performed. The same training samples were used. For the case of the synthetic image, SVM classifier with RBF kernel was adopted. As shown in Figure 7-c, the ENVI feature extraction tool failed to classify the segment with label 12. For the quickbird sub-scene, the classification was performed by applying rule-based method with two spatial attributes namely: rectangular fit and area in order to identify only the buildings with rectangular form. The spatial attribute rectangular fit indicates how well the shape is described by a rectangle. The classification result is shown in Figure 13.



Figure 13. Extracted buildings classes from the quickbird sub-scene obtained with ENVI feature extraction tool.

A classification accuracy of 79% was obtained which is less than the accuracy obtained by our approach. This can be explained by the fact that this method uses simple spatial attributes as scalar descriptors such as convexity, elongation and area. These attributes are not discriminative because in most cases they give similar values for different shapes.

6. Conclusions

In this paper, a new approach for automated buildings extraction from VHRS images was proposed. First, image objects extraction is performed using the Mean Shift segmentation algorithm. Then, a Zernike moments-based descriptor is calculated for each object or segment. Finally, a SVMs-based classification using the feature vectors as input instead of the original objects is carried out to assign a class label to each of the segments. The main outcome of this work is the use of Zernike moment's descriptor as a measure of shape information. Consequently, by performing the classification on the obtained Zernike feature vectors instead of the original objects, a good discrimination between object shapes can be achieved.

The preliminary experimental results clearly demonstrated the potentials and the efficiency of the proposed approach as well as some application limitations, which encourages new reflections both for method enhancement and for future research.

Future research will involve using the adaptive Mean Shift algorithm in the segmentation stage and extending the proposed approach to include the textural and the spectral information in the classification process in order to increase the building identification accuracy.

References

- [1] Abdel-Qader H., Ramli A., and Al-Haddad S., "Fingerprint Recognition using Zernike Moments," *International Arab Journal of Information Technology*, vol. 4, no. 4, pp. 372-376, 2007.
- [2] Bouziani M., Goita K., and He D., "Rule-based Classification of a Very High Resolution Image in an Urban Environment using Multispectral Segmentation Guided by Cartographic Data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 48, no. 8, pp. 3198-3211, 2010.
- [3] Bruzzone L. and Carlin L., "A Multilevel Context-Based System for Classification of Very High Spatial Resolution Images," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2587-2600, 2006.
- [4] Camps-Valls G., Gomez-Chova L., Calpe-Maravia J., Martin-Guerrero J., Soria-Olivas E., Alonso-Chorda L., and Moreno J., "Robust Support Vector Method for Hyperspectral Data Classification and Knowledge Discovery," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 42, no. 7, pp. 1530-1542, 2004.
- [5] Celebi M. and Aslandogan Y., "A Comparative Study of Three Moment-based Shape Descriptors," in *Proceedings of IEEE International Conference on Information Technology: Coding and Computing*, Las Vegas, vol. 1, pp. 788-793, 2005.
- [6] Chabrier S., Emile B., Laurent H., Rosenberger C., and March P., "Unsupervised Evaluation of Image Segmentation: Application to Multispectral Images," in *Proceedings of the 17th International Conference Pattern Recognition*, vol. 1, pp. 576-579, 2004.
- [7] Cheng Y., "Mean Shift, Mode Seeking, and Clustering," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790-799, 1995.
- [8] Chong C., Raveendran P., and Mukundan R., "Translation Invariants of Zernike Moments," *Pattern Recognition*, vol. 36, no. 8, pp. 1765-1773, 2003.
- [9] Chong C., Raveendran P., and Mukundan R., "A Comparative Analysis of Algorithms for Fast Computation of Zernike Moments," *Pattern Recognition*, vol. 36, no. 3, pp. 731-742, 2003.
- [10] Comaniciu D. and Meer P., "Mean Shift: A Robust Approach Towards Feature Space Analysis," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603-619, 2002.
- [11] Cristianini N. and Shaew-Taylor J., *An Introduction to Support Vector Machines and other Kernel Based Learning Methods*, Cambridge University Press, UK, 2000.
- [12] Fukunaga K. and Hostetler L., "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition," *IEEE Transaction on Information Theory*, vol. 21, no. 1, pp. 32-40, 1975.
- [13] Gope C., Kehtarnavaz N., and Hillman G., "Zernike Moment Invariants based Photo-Identification using Fisher Discriminant Model," in *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Francisco, vol. 1, pp. 1455-1458, 2004.
- [14] Herold M., Gardner M., and Roberts D., "Spectral Resolution Requirement for Mapping Urban Areas," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 1907-1919, 2003.
- [15] Hofmann P., "Detecting Urban Features from IKONOS Data using an Object-Oriented Approach," in *Proceedings of the Remote Sensing and Photogrammetry Society*, pp. 79-91, 2003.
- [16] Hu M., "Visual Pattern Recognition by Moment Invariants," *IRE Transaction on Information Theory*, vol. 8, no. 2, pp. 179-187, 1962.
- [17] Inglada J. and Michel J., "Qualitative Spatial Reasoning for High-Resolution Remote Sensing Image Analysis," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 47, no. 2, pp. 599-612, 2009.
- [18] Kim H. and Lee H., "Invariant Image Watermark using Zernike Moments," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 766-775, 2003.
- [19] Li S., Lee M., and Pun C., "Complex Zernike Moments Features for Shape-Based Image Retrieval," *IEEE Transaction on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 39, no. 1, pp. 227-237, 2009.
- [20] Loncaric S., "A Survey of Shape Analysis Techniques," *Pattern Recognition*, vol. 31, no. 8, pp. 983-1001, 1998.
- [21] Nixon M. and Aguado A., *Feature Extraction and Image Processing*, Academic Press, UK, 2008.
- [22] Persoon E. and Fu K., "Shape Discrimination

using Fourier Descriptors,” *IEEE Transaction on Systems, Man, and Cybernetics*, vol. 7, no. 3, pp. 170-179, 1977.

- [23] Shackelford A. and Davis C., “A Combined Fuzzy Pixel-Based and Object-Based Approach for Classification of High Resolution Multispectral Data Over Urban Areas,” *IEEE Transaction on Geoscience and Remote Sensing*, vol. 41, no. 10, pp. 2354-2363, 2003.
- [24] Teague M., “Image Analysis via the General Theory of Moments,” *Journal of the Optical Society of America*, vol. 70, no. 8, pp. 920-930, 1980.
- [25] Teh C. and Chin R., “On Image Analysis by the Methods of Moments,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 496-513, 1988.
- [26] Vapnik V., *Statistical Learning Theory*, Wiley, New York, 1998.
- [27] Zhang D. and Lu G., “Review of Shape Representation and Description Techniques,” *Pattern Recognition*, vol. 37, no. 1, pp. 1-19, 2004.
- [28] Zhang D. and Lu G., “Content-Based Shape Retrieval using Different Shape Descriptors: A Comparative Study,” in *Proceedings of IEEE International Conference on Multimedia and Expo*, Japan, pp. 317-320, 2001.



Chahira Serief received her Engineering, Ms, and PhD degrees in Electrical Engineering from the University of Constantine in Algeria, in 1995, 2000, and 2009 respectively. Since 2002 she has been working as a permanent researcher at the Earth Observation Division of the Centre of Space Techniques in Algeria. Her principal research interests are in the fields of satellite and medical image processing, image feature extraction; image registration, image segmentation and contour detection.



Habib Mahi received his Engineering degree in computer science from the University of Mohamed Boudiaf in Algeria, in 1994 and MS degree in Techniques Spatiales et Applications from the Centre of Space Techniques in Algeria, in 2004. During 2004, he is a permanent researcher at the Earth Observation Division of CTS. His research interests include: image processing, image feature extraction, classification algorithms development for remotely sensed data.



Hadria Isabaten received her Engineering degree in Electrical Engineering from the University of Mohamed Boudiaf in 1981, the “Docteur Ingénieur” degree in industrial computer science and automation from the University of Lille 1 in France, in 1987 and the PhD degree in computer science from the University of Mohamed Boudiaf in 2005. Currently she is a professor in the Faculty of Computer Science in the University of Mohamed Boudiaf. Her research interests cover image processing and the application of pattern recognition to remotely sensed data.