Urban Area Detection Using Local Feature Points and Spatial Voting

Beril Sırmaçek, *Student Member, IEEE,* Cem Ünsalan, *Member, IEEE,*

***Abstract*—Automatically detecting and monitoring urban regions is an important problem in remote sensing. Very high resolution aerial and satellite images provide valuable information to solve this problem. However, they are not sufficient alone for two main reasons. First, a human expert should analyze these very large images. There may be some errors in operation. Second, the urban area is dynamic. Therefore, detection should be done periodically and this is time consuming. To handle these shortcomings, an automated system is needed to detect the urban area from aerial and satellite images. In this study, we propose such a method based on local feature point extraction using Gabor filters. We use these local feature points to vote for candidate urban areas. Then, we detect the urban area using an optimal decision making approach on the vote distribution. We test our method on a diverse panchromatic aerial and Ikonos satellite image set. Our test results indicate the possible use of our method in practical applications.**

***Index Terms*—Aerial images, satellite images, local feature points, spatial voting, urban area detection.**

I. Introduction

Monitoring urbanization may help government agencies and urban region planners in updating land maps and forming long term plans accordingly. Very high resolution aerial and satellite images provide valuable information for this purpose. Unfortunately, these images cover very large areas. Therefore, their manual inspection is very hard and prone to errors. Also, urban areas are dynamic environments. Hence, they should be monitored periodically. Because of these difficulties, manually monitoring urbanization using very high resolution aerial and satellite images is not feasible. Therefore, in this study we propose an automated system for the first step of urban monitoring, detecting urban areas.

In the literature, many researchers considered the urban area detection problem using automated techniques. Karathanassi *et al.* [6] used building density information to classify residential regions. They benefit from texture information and segmentation to extract residential areas. Unfortunately, they had several parameters to be adjusted manually. Benediktsson *et al.* [1] used mathematical morphological operations to extract structural information to detect the urban area in satellite images. They benefit from neural networks for classification purposes. Therefore, they need training data to detect urban areas. Ünsalan and Boyer [13], [14], [12] used structural features to classify urban regions in panchromatic satellite images. Since they use statistical classifiers, they also need training data to detect the urban area in the image. In a following study, Ünsalan and Boyer [15] associated structural features with graph theoretical measures in order to grade satellite images and extract residential regions from them. In a related study, Sırmaçek and Ünsalan [10] used scale invariant feature transform (SIFT) and graph theory to detect urban areas and buildings in grayscale Ikonos images. They used template building images for this purpose. Although graph theoretical methods are suitable for urban area detection, they need considerable computation power and operation time. Fonte *et al.* [5] considered corner detectors to obtain the type of structure in a satellite image. They concluded that, corner detectors may give distinctive information on the type of structure in an image. Bhagavathy and Manjunath [2] used texture motifs for modeling and detecting regions (such as golf parks, harbors) in satellite images. They focused on repetitive patterns in the image. Bruzzone and Carlin [3] proposed a context-based system to classify very high resolution satellite images. They used support vector machines fed with a novel feature extractor. Fauvel *et al.* [4] fused different classifiers to extract and classify urban regions in panchromatic satellite images. Zhong and Wang [17] extracted urban regions in grayscale satellite images using a multiple classifier approach. These last three studies also need a training data for urban area classification.

In this study, we propose a novel method to detect urban areas in very high resolution panchromatic aerial and satellite images. Different from most of the studies in the literature, our method does not need any training data for urban area detection. Instead, we first use Gabor filters to extract spatial building characteristics (such as edges and corners) in different orientations. Then, we process filter outputs to obtain local feature points in the image. We form a voting matrix using them. Finally, we label the urban area in a given image by an optimal adaptive decision making approach. Our method is able to detect urban areas as long as buildings are dense. Besides, we do not have any other constraints. To provide an experimental justification to our method, we tested it on diverse aerial and Ikonos satellite images. We obtained encouraging results.

II. Local Feature Point Extraction

We detect the urban area in a given test image using local feature points. The first step to extract them from an image is smoothing it by median filtering [11]. This step eliminates small noise terms in the image. Then, we apply Gabor filtering in different directions. The maxima in these filter responses

The authors are with the Department of Electrical and Electronics Engineering, Computer Vision Research Laboratory, Yeditepe University, Istanbul, Turkey (e-mail: unsalan@yeditepe.edu.tr)

This work is supported by İstanbul Büyükşehir Belediyesi

lead to local feature points. Next, we explore these steps in detail.

A. Gabor Filtering

Gabor filters are extensively used in texture segmentation and object recognition [7]. They exhibit desirable characteristics as spatial locality and orientation selectivity [16]. Mathematically, the two dimensional Gabor filter can be defined as the product of a Gaussian and a complex exponential function as:

$$F_{\varphi}(x,y) = \frac{1}{2\pi\sigma_g^2} \exp\left(-\frac{u^2+v^2}{2\sigma_g^2}\right) \exp\left(j2\pi fu\right) \quad (1)$$

where $u = x\cos\varphi + y\sin\varphi$ and $v = -xsin\varphi + y\cos\varphi$. f is the frequency of the complex exponential signal, φ is the direction of the Gabor filter, and σ_g is the scale parameter. These parameters should be adjusted wrt. the image resolution at hand. In this study, we explore the effect of these parameters in Section IV-A.

We can detect edge-oriented urban characteristics (such as building edges) in a test image using Gabor filtering. Therefore, for a test image $I(x,y)$ (with size $N \times M$), we benefit from the real part of the Gabor filter response as:

$$G_{\varphi}(x,y) = \mathfrak{Re}\{I(x,y) * F_{\varphi}(x,y)\} \quad (2)$$

where $*$ stands for the two dimensional convolution operation. $G_{\varphi}(x,y)$ is maximum for image regions having similar characteristics with the filter. In the next section, we use this information to extract local feature points.

B. Local Feature Points

To extract local feature points, we first search for the local maxima in $G_{\varphi}(x,y)$ for $x = 1,...,N$ and $y = 1,...,M$. If any pixel (x_o, y_o) in $G_{\varphi}(x,y)$ has the largest value among its neighbors, $G_{\varphi}(x_o,y_o) > G_{\varphi}(x_n,y_n)\ \forall\ (x_n,y_n) \in \{(x_o-1,y_o-1),(x_o,y_o-1),...,(x_o+1,y_o+1)\}$; we call it as a local maximum. It is a candidate for being a local feature point. Next, we check the amplitude of the filter response $G_{\varphi}(x_o,y_o)$. We call our local maximum (x_o,y_o) as a candidate local feature point if and only if $G_{\varphi}(x_o,y_o) > \alpha$. To handle different images, we obtain α using Otsu's method on $G_{\varphi}(x,y)$ in an adaptive manner for each image separately [8]. Therefore, we eliminate weak candidate local feature points in future calculations.

To represent each candidate local feature point further, we assign a weight, w_o, to it as follows. We first threshold $G_{\varphi}(x,y)$ with α and obtain a binary image $B_{\varphi}(x,y)$. In this image, pixels having value one correspond to strong responses. We obtain connected pixels to (x_o,y_o) in $B_{\varphi}(x_o,y_o)$. By definition, two pixels are connected (in a binary image) to each other if there is a path (of pixels with value one) connecting them [11]. As we obtain all the connected pixels to (x_o,y_o), we assign their sum as the weight w_o. Therefore, if a candidate local feature point has more connected pixels, it has more weight. We expect the candidate local feature points to represent urban characteristics such as building clusters. Unfortunately, all candidate local feature points may not represent reliable information on the urban area. Therefore, we discard candidate local feature points having weight, w_o, less than 20 pixels. Although we applied the same weight threshold value for both aerial and satellite images (having different characteristics) in this study, the reader may need to adjust it wrt. the test image at hand. Finally, we obtain the local feature points for the given φ direction.

We apply this procedure in all φ directions and obtain a total of K local feature points as (x_k,y_k) with their weights w_k for $k = 1,...,K$. We expect these local feature points to be located on building edges in the image. We provide such an example on the sample $Adana_6$ satellite test image in Fig. 1. As can be seen, most local feature points are located on the building edges in this image.

Fig. 1. The $Adana_6$ test image, local feature points extracted

In the literature, there are also more complex feature point extraction methods [9]. However, for urban area detection we do not need perfect local point extraction from the image. Since we use their ensemble, missing a few local feature points does not affect the performance of our urban area detection method. We explore urban area detection using local feature points next.

III. Urban Area Detection

As we obtain local feature points, the next step is detecting the urban area using them. Therefore, we form a voting matrix based on spatial voting first. Then, we apply an optimum decision making method to detect the urban area from the voting matrix. We explain these steps in detail next.

A. Voting Matrix Formation

To detect an urban area, we should have many local feature points in it. These should also be closely located in spatial domain. Therefore, we define a voting matrix based on extracted local feature points as follows. We assume that, around each local feature point there is a high possibility of an urban characteristic (such as a building). Therefore, each local feature point has the highest vote at its spatial coordinate (x_k,y_k) and its votes decrease wrt. spatial distance. Based on this definition, we form the voting matrix for $x = 1,...,N$ and $y = 1,...,M$ as:

$$V(x,y) = \sum_{k=1}^{K} \frac{1}{2\pi\sigma_k^2} \exp\left(-\frac{(x-x_k)^2+(y-y_k)^2}{2\sigma_k^2}\right) \quad (3)$$

where σ_k is the parameter for voting proximity for each local feature point. For our test images, we pick $\sigma_k = 5 \times w_k$ to add some tolerance for voting. If σ_k has higher values, each local feature point will have a wider spatial effect in the voting matrix. Therefore, the false alarms in urban area detection will increase. On the other hand, if σ_k has lower values, each local feature point will have a narrower spatial effect. Hence, the correct urban area detection results will decrease.

For the $Adana_6$ test image, we provide the voting matrix in Fig. 2. In this figure, vote values are color coded (red corresponds to the highest and blue to the lowest vote value). As can be seen, votes are cumulated around buildings.

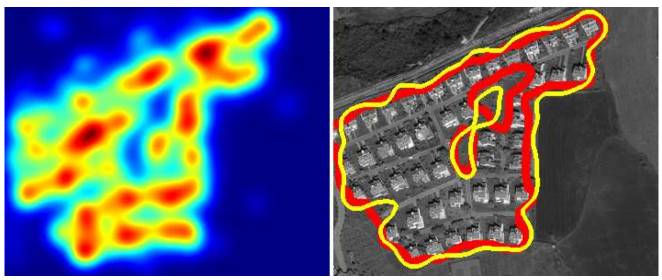

Fig. 2. The $Adana_6$ test image, the voting matrix obtained and the urban area detected (with ground truth data)

B. Optimum Decision Making

Locations with high votes in $V(x, y)$ are possible urban area pixels and locations with low votes are possible non-urban area pixels in the image. Therefore, we expect to have a bimodal pixel distribution on $V(x, y)$ (one peak corresponds to the urban area and the other to non-urban area pixel votes). We use this information and Otsu's method to detect the urban area [8]. Based on Bayes decision criteria, Otsu's method finds the optimal threshold level (assuming Gaussian probability density functions) between urban and non-urban pixel votes. In this study, the threshold value is adaptively obtained for each test image separately. Since this operation is adaptive, we do not need any manual (or predefined) parameter adjustments.

For the $Adana_6$ satellite test image, we provide the voting matrix pixel distribution in Fig. 3. As can be seen in this distribution, there are two peaks (one corresponding to urban and the other to the non-urban area votes). In the same figure, we also provide the optimum decision boundary obtained by Otsu's method by a red dashed vertical line. As can be seen, the decision boundary is correctly located. Based on this threshold value, we provide the detected urban area (as a yellow curve) from the $Adana_6$ image in Fig. 2. In the same figure, we also provide our ground truth data for the $Adana_6$ test image as a thick red curve. As can be seen, our detection result closely fits to the ground truth data.

Besides thresholding with Otsu's method, we also consider the overall votes in the test image. If their accumulation is not sufficient, we assume they can not represent an urban area. To perform this test on an image basis, we assume that if the sum of votes does not exceed 5% of the image size, there is no urban area there. After extensive testing, we obtained this

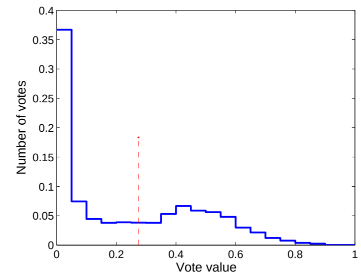

Fig. 3. The voting matrix pixel distribution and the optimum threshold value for the $Adana_6$ test image.

value for our test images. However, it is not strict and can be changed in a relaxed manner.

IV. Experiments

To test the performance of our urban area detection method, we use 21 aerial and 31 panchromatic Ikonos satellite images. Our aerial images have 0.3 m spatial resolution and Ikonos satellite images have 1 m spatial resolution. These aerial and satellite images naturally have different acquisition sensor characteristics. Therefore, experimental test results (provided in this section) can be taken as independent of the acquisition sensor characteristics and spatial resolution. Our test images are also specifically selected to represent wide and diverse urban area characteristics. On Ikonos images, we first provide the experimental justification for selected parameter values. In the following two sections, we provide the overall urban area detection performance of our method on satellite and aerial images separately. We quantify these tests by reporting the urban area detection percentage, P_d, (the ratio of correctly detected urban area pixels to manually labeled urban area pixels) and the overall false alarm percentage, P_f, (the ratio of false detected pixels to manually labeled urban area pixels). In forming the ground truth (manually labeled urban area), we label a region as urban if it contains building clusters, gardens around them and nearby street segments joining them as we did in our previous study [10]. We also provide the computation times for each operation in our urban area detection method on a sample test image. Finally, we compare our method with our previously published urban area detection method [10]. There, we used SIFT and graph theoretical tools for urban area detection.

A. Tests on Parameter Values

Although we do not have many parameters in our urban area detection method, we provide their effect on the final detection results for satellite images in this section. We first comment on Gabor filter parameters. We observed that, in panchromatic Ikonos images building edges can be represented as a ramp edge with three to four pixels width. Therefore, in this study we picked $\sigma_g = 1.5$, and $f = 0.65$ after extensive testing. We also used the same parameter values for aerial test images without any problem. These two parameters should be adjusted wrt the building characteristics in the test image. To cover differently oriented building edges, we tested different φ

values. We provide different settings in Table. I. In these tests, we use 31 Ikonos images to eliminate image dependency.

TABLE I
THE EFFECT OF GABOR FILTERING DIRECTIONS ON URBAN AREA DETECTION.

Number of φ	$\mathbf{P_d}(\%)$	$\mathbf{P_f}(\%)$
4	88.39	11.55
6	88.63	8.21
8	88.47	6.30
10	**89.33**	**5.91**
12	88.79	5.91
14	90.79	5.99

As can be seen in Table. I, the number of Gabor filtering directions have effect on decreasing P_f. Based on this test, we can conclude that choosing ten different directions for Gabor filtering ($\varphi = \{0, \pi/10, 2\pi/10, ..., 9\pi/10\}$ radians) is suitable for urban area detection. To note here, choosing six, eight or twelve directions also provide similar P_d and P_f values. Therefore, we can conclude that our method does not specifically depend on the total Gabor filtering direction.

We then test the effect of median filtering on the final detection result. We provide tests with different median filter sizes in Table. II. Again, in these tests we use 31 Ikonos images to eliminate image dependency.

TABLE II
THE EFFECT OF MEDIAN FILTERING ON URBAN AREA DETECTION.

Median filter size	$\mathbf{P_d}(\%)$	$\mathbf{P_f}(\%)$
No filtering	95.04	31.76
3×3	94.13	16.75
$\mathbf{5 \times 5}$	**89.33**	**5.91**
7×7	84.29	5.15

As can be seen in Table. II, without median filtering P_f is fairly high. Therefore, we need a median filtering operation. As can be seen, a 5×5 median filtering gives reasonable P_d and P_f values.

For aerial images, we also use the same settings. However, to make the resolution of the satellite and aerial images similar, we downsampled aerial images by 0.5. Next, we discuss the overall urban area detection performance of our method on aerial and satellite images.

B. The Overall Performance on Satellite Images

The overall performance of our method on 31 satellite images having 775714 urban area pixels is $P_d = 89.33\%$ and $P_f = 5.91\%$. Therefore, our method was able to detect 692926 urban area pixels correctly with 45854 false alarm pixels. This is a fairly good urban area detection result on such a diverse satellite image set. We also provide a sample test image for urban area detection results for satellite images in Fig. 4. As can be seen in this figure, all urban regions with different building characteristics are correctly detected.

C. The Overall Performance on Aerial Images

Next, we discuss the overall performance of our method on 21 aerial images. In this test, we have 1688936 urban area pixels. We obtained $P_d = 85.93\%$ with $P_f = 10.95\%$ for urban area detection. Therefore, 1451225 urban area pixels are correctly detected, with 184851 false alarm pixels. Similar to the satellite images test, this result on such a data set is fairly good. We also provide a sample test image for urban area detection results for aerial images in Fig. 5. As can be seen in this figure, all urban regions with different building characteristics are correctly detected.

Fig. 4. Urban area detection result on a sample satellite image

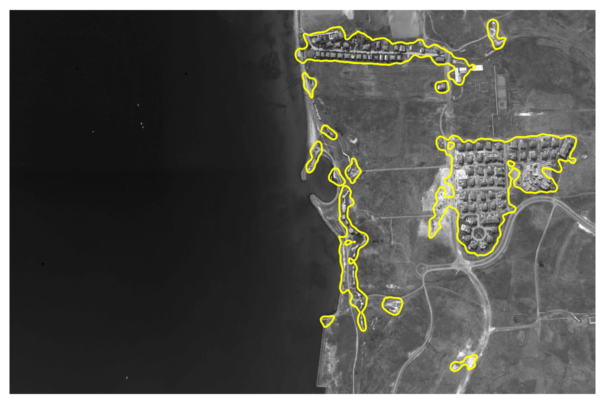

Fig. 5. Urban area detection result on a sample aerial image

D. Computation Time

The other important property of our method is its computation time. To provide an idea about the computation time needed for our method, we next consider the CPU timings for each step on the $Adana_8$ satellite test image (having size of 235×265 pixels). We picked this image as a benchmark in our previous study [10]. Therefore, we pick the same image in this study to compare both methods. In reporting computation times, we used a PC with Intel Core2Due processor with 2.13 GHz. clock speed and has 4 GB of RAM. We used Matlab as our coding platform. After testing, we obtained that for the $Adana_8$ image Gabor filtering needs 0.61 sec.; local feature point extraction needs 0.94 sec. and spatial voting needs 2.30 sec. The total time needed for urban area detection is 3.85 sec. This is a fairly good timing to obtain the urban area.

E. Comparison with SIFT Based Method

Finally, we compare our new urban area detection method with our previous study based on SIFT and graph theory [10]. In this section, we use the same data set in our previous study (without the final two Adana satellite images and all aerial images we used in this study). The overall urban area detection performance there was $P_d = 89.62\%$ with $P_f = 8.03\%$. On the same data set, with our new method we obtain $P_d = 89.09\%$ with $P_f = 5.88\%$. The detection and false alarm rates are almost the same. Interested reader can visually compare detection results on the $Adana_6$ satellite test image both with the present method and SIFT based method by checking [10].

To compare both methods in terms of CPU timings, we pick the $Adana_8$ satellite test image as a benchmark. The time needed for the SIFT based method was 81.97 sec. With our new method, we only need 3.85 sec. Therefore, our new method is fairly fast and almost has the same performance with our SIFT based urban area detection method. Besides, we do not need any template building images (as in the SIFT based method) to detect urban areas in this new method. To note here, in our SIFT based method, we also detect buildings after detecting the urban area.

V. Conclusions

This study focuses on urban area detection using very high resolution aerial and satellite images. It may be taken as the first step in monitoring urbanization. Our method depends on local feature point extraction using Gabor filtering. We use these local feature points in forming a spatial voting matrix. Then by using an optimum decision making approach, we are able to detect the urban area in a given aerial or satellite image. After extensive testings, we obtained very encouraging results with our method. Comparing with an existing algorithm, we can also conclude that our new urban area detection method is fairly fast and reliable. We can further apply probabilistic relaxation to improve our results. However, we will need extra computations for performing it. The next step in this study will be detecting the urban area type (such as dense, homogeneous, and well-structured) by analyzing the voting matrix characteristics.

References

[1] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 1940–1949, 2003.

[2] S. Bhagavathy and B. S. Manjunath, "Modeling and detection of geospatial objects using texture motifs," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 12, pp. 3706–3715, 2006.

[3] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2587–2600, 2006.

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

[5] L. M. Fonte, S. Gautama, W. Philips, and W. Goeman, "Evaluating corner detectors for the extraction of man made structures in urban areas," in *IEEE International Geoscience and Remote Sensing Symposium*, 2005, pp. 237–240.

[6] V. Karathanassi, C. Iossifidis, and D. Rokos, "A texture-based classification method for classifying built areas according to their density," *International Journal of Remote Sensing*, vol. 21, no. 9, pp. 1807–1823, 2000.

[7] V. Kyrki, J. K. Kamarainen, and H. Kalviainen, "Simple Gabor feature space for invariant object recognition," *Pattern Recognition Letters*, vol. 25, no. 3, pp. 311–318, 2004.

[8] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on System, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[9] C. Schmid, R. Mohrand, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151–172, 2000.

[10] B. Sırmaçek and C. Ünsalan, "Urban area and building detection using SIFT keypoints and graph theory," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 4, pp. 1156–1167, 2009.

[11] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*, 3rd ed. CL Engineering, 2007.

[12] C. Ünsalan, "Gradient magnitude based support regions in structural land use classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, pp. 546–550, 2006.

[13] C. Ünsalan and K. L. Boyer, "Classifying land development in high resolution panchromatic satellite images using straight line statistics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 4, pp. 907–919, 2004.

[14] ——, "Classifying land development in high resolution satellite imagery using hybrid structural - multispectral features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 12, pp. 2840–2850, 2004.

[15] ——, "A theoretical and experimental investigation of graph theoretical measures for land development in satellite imagery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 575–589, 2005.

[16] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, 1st ed. Prentice Hall, 1995.

[17] P. Zhong and R. Wang, "A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3978–3988, 2007.