A Probabilistic Approach to Detect Urban Regions from Remotely Sensed Images Based on Combination of Local Features

Beril Sırmaçek
German Aerospace Center (DLR)
Remote Sensing Technology Institute
Weßling, 82234, Germany
E-mail: Beril.Sirmacek@dlr.de

Cem Ünsalan
Computer Vision Research Laboratory
Yeditepe University
Istanbul, 34755, Turkey
E-mail: unsalan@yeditepe.edu.tr

***Abstract*—Detecting urban regions from very high resolution aerial and satellite images provides very useful results for urban planning, and land use analysis. Since manual detection is very time consuming and prone to errors, automated systems to detection of urban regions from very high resolution aerial and satellite images are needed. Unfortunately, diverse characteristics of urban regions, and uncontrolled appearance of remote sensing images (illumination, viewing angle, etc.) increase difficulty to develop automated systems. In order to overcome these difficulties, herein we propose a novel urban region detection method using local features and a probabilistic framework. First, we introduce four different local feature extraction methods. Extracted local feature vectors serve as observations of the probability density function to be estimated. Using a variable kernel density estimation method, we estimate the corresponding probability function. Using modes of the estimated density, as well as other probabilistic properties, we detect urban region boundaries in the image. We also introduce data and decision fusion methods to fuse information coming from different feature extraction methods. Extensive tests on very high resolution grayscale aerial and panchromatic Ikonos satellite images indicate practical usefulness of proposed method to detect urban regions automatically in a robust and fast manner.**

I. Introduction

Remotely sensed satellite and aerial images provide very valuable information. However, their covering areas and resolution make manual analysis very difficult and prone to errors. Furthermore, especially urban areas are dynamic environments. Hence, they should be monitored and analyzed periodically. Due to these problems, developing algorithms to detect particular objects in remotely sensed images is a very important research field. Especially, automatic detection of urban region boundaries can provide useful information to municipalities, mapping agencies, military, government agencies, or unmanned aerial vehicle developers. Unfortunately, automatic object detection algorithms cope with some difficulties on remotely sensed images. First, buildings generally have diverse characteristics with different texture, color, and shape. In addition to that, in their appearance on the image, the illumination, view angle, scaling, occlusion effects are uncontrolled. Therefore, classical object detection algorithms cannot provide an acceptable detection performance. Therefore, more advanced methods are required for robust detection of the objects in remotely sensed images.

In related literature, due to the importance of the problem, many researchers concentrated on developing automated systems to detect urban regions. Karathanassi *et al.* [7] used building density information to classify residential regions. They benefit from texture information and segmentation to extract the 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 region boundaries in satellite images. Their method is based on using a neural network which is trained using training urban area regions to classify input images. Ünsalan and Boyer [16], [18] 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 [17] associated structural features with graph theoretical measures in order to grade the satellite images and extract the residential regions from them. Fonte *et al.* [5] considered corner detectors to obtain the type of the structure in a satellite image. They concluded that corner detectors might 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 and 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 [20] extracted urban regions in grayscale satellite images using a multiple-classifier approach. These last three studies also need training data for urban area classification. In a related study, Sırmaçek and Ünsalan [16], 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. In a following study, we proposed a method to detect urban region boundaries using Gabor features alone [14]. In this study, we extend our previous work by a new approach for robust detection of urban region boundaries based on four sets of local invariant features, their fusion in different levels, and a probabilistic framework. To this end, first we introduce four different local feature extraction methods as: Harris corners, Gradient Magnitude Support Region (GMSR) based features, Gabor features, and FAST features. The extracted local features serve as observations of the probability density function (pdf) to be estimated. We formed pdf by using the nonparametric variable-kernel density estimation method. Using the estimated density, we detect urban region boundaries in the given image. To increase robustness, and to merge information coming from different local features, we present two different fusion methods in data and decision levels. In our experiments, we use a dataset obtained from different cities by two different sensors to test our method. Specifically the data set is formed of very high resolution panchromatic Ikonos satellite images and grayscale aerial images. These test images have also different spatial resolutions and diverse characters. This data set is the same as in our previous study. Therefore, we have a chance to compare both our existing and novel method on the same basis.

II. Local Feature Extraction

We detect the urban area in a test image using local feature points. As first step, before extracting local features, is to apply smoothing to the input test image by using median filter [15]. This step eliminates small noise terms in the image. Then, we apply Gabor filtering in different directions. The maxima in these filter responses lead to local feature points. Next, we will explore these steps in detail.

A. Gabor Features

Gabor filters are extensively used in texture segmentation and object recognition [8]. They exhibit desirable characteristics as spatial locality and orientation selectivity [19]. Mathematically, the 2D Gabor filter can be defined as the product of a Gaussian and a complex exponential function as follows,

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

Here, $u = x\cos\varphi + y\sin\varphi$ and $v = -x\sin\varphi + y\cos\varphi$. 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 with respect to the image resolution at hand.

We can detect the 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 NxM), we benefit from the real part of the Gabor filter response as

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

where $*$ stands for the 2D convolution operation. Here, $G_\varphi(x,y)$ is the maximum for image regions having similar characteristics with the filter.

In order to extract Gabor features, 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 [9]. Therefore, we eliminate the weak candidate local feature points in future calculations.

To represent each candidate local feature point further, we assign an orientation and weight to them. We obtain the weight for each local feature vector similar to the methods in previous sections. Here, we obtain our binary image $B_m(x,y)$ by thresholding $G_\varphi(x,y)$ with α for weight calculations. However, we assign the orientation different than the two previous methods as follows. We check for the orientations in the eight-neighborhood of (x_j,y_j) and pick the orientation, θ_j as the one having highes magnitude. We apply this procedure to obtain a robust orientation information. We apply this procedure in all φ directions and obtain Gabor filtering based local feature vectors as $k_f = (x_j,y_j,\theta_j,w_j)$ for $j = 1,...,K_f$, where K_f is the total number of extracted Gabor features.

B. Harris Corner Features

Fonte *et al.* [5] considered Harris and Susan corner detectors to obtain the type of structure in a satellite image. They concluded that, corner detectors are not sufficient alone to give distinctive information on the type of structure in an image. Schmid *et al.* [11] on the other hand evaluated and compared different corner detectors for general image processing applications. They concluded that the best results are provided by the Harris corner detector [6]. Therefore, in our study we also benefit from information coming from Harris corner detector.

Harris corners are extracted in three steps: gradient calculation, matrix formation, and eigenvalue computation. First, we should calculate smoothed gradients in x and y directions to detect corners in a given grayscale image $I(x,y)$. We define smoothed gradient filters for x and y directions as;

$$g_x(x,y) = \frac{-x}{2\pi\tau_g^4} exp(-\frac{x^2+y^2}{2\tau_g^2}) \tag{3}$$

$$g_y(x,y) = \frac{-y}{2\pi\tau_g^4} exp(-\frac{x^2+y^2}{2\tau_g^2}) \tag{4}$$

where τ_g is the smoothing parameter that we select as unity due to the scale of Ikonos and aerial images at hand. Although our method is fairly robust to this parameter, it should be adjusted by the resolution of the image to be analyzed in future studies.

We calculate the smoothed gradients for the image $I(x,y)$ as

$$I_x = g_x(x,y) * I(x,y) \tag{5}$$

$$I_y = g_y(x,y) * I(x,y) \tag{6}$$

where $*$ stands for the two dimensional convolution operation.

Harris corner detector depends on calculating a matrix (related to autocorrelation function) as

$$A(x,y) = \begin{pmatrix} a_{xx} & a_{xy} \\ a_{xy} & a_{yy} \end{pmatrix} \tag{7}$$

where

$$a_{xx} = \sum_{x_i \in W} \sum_{y_i \in W} I_x^2(x_i, y_i) \tag{8}$$

$$a_{xy} = \sum_{x_i \in W} \sum_{y_i \in W} I_x(x_i, y_i) I_y(x_i, y_i) \tag{9}$$

$$a_{yy} = \sum_{x_i \in W} \sum_{y_i \in W} I_y^2(x_i, y_i) \tag{10}$$

Here, a_{xx}, a_{xy}, a_{yy} are gradient magnitudes averaged over a window W. We pick this averaging window width as seven pixels in this study. Further analysis of choosing this parameter can be found in our previous study [13]. Eigenvalues of the matrix A provide information about the edge in a given location. If both eigenvalues of the matrix at a given location is large, then there is a corner there. Harris and Stephens suggested that, exact eigenvalue computation can be avoided by calculating the response function

$$R(A) = |A| - \kappa trace^2(A) \tag{11}$$

where κ is a tunable parameter with values from 0.04 to 0.15 were reported as appropriate in the literature. We picked $\kappa = 0.06$ in this study. We also tested the effect of using this parameter in remote sensing image in our previous study [13]. Corner points are detected by checking the local maxima of $R(A)$.

As we obtain corner points with their spatial coordinates, we define our local features using them. Besides spatial coordinates, we also add orientation and weight information to features. First, we calculate the gradient orientation $O(x,y)$, and magnitude $M(x,y)$, for each image as,

$$O(x,y) = arctan(\frac{I_y(x,y)}{I_x(x,y)}) \tag{12}$$

$$M(x,y) = \sqrt{I_x^2(x,y) + I_y^2(x,y)} \tag{13}$$

For the corner point coordinate (x_j, y_j), the corresponding orientation is $\theta_j = O(x_j, y_j)$. To assign a weight for the local feature vector, we threshold $M(x,y)$ using Otsu's method in an adaptive manner [9]. As a result, we obtain $B_m(x,y)$ as a binary image. In this image, pixels having values one corresponding to strong responses. We obtain connected pixels to (x_j, y_j), we assign their sum as the weight w_j. Therefore, if a candidate local feature vector has more connected pixels, it has more weight. Finally, we have our Harris corner based local features as $k_h(j) = (x_j, y_j, \theta_j, w_j)$ for $j = 1, ..., K_h$ where K_h is the total number of detected Harris features.

C. GMSR Features

Next we pick our previous study, to extract gradient magnitude based support regions (GMSR) [16]. There, we used these features to extract structural and conditional statistical features to classify land use. Herein, we extract support regions using smoothed gradient values, namely I_x and I_y given in 5 and 6 respectively. To extract support regions, we threshold $M(x,y)$ with 10% of the maximum gradient magnitude in the considered image $I(x,y)$. The rationale here is as follows. We take the maximum gradient magnitude as a benchmark. After experiments, we observed that even 10% of this value still gives information about the structure in the image. Therefore, we have an adaptive threshold value. Similar to the Harris corner detection method, we obtain $B(x,y)$ as a binary image after thresholding. In this image, pixels having a value of one correspond to support regions.

We define local feature vectors based on the extracted support regions. Therefore, we pick each support region pixel as a local feature vector coordinate. Assume that we have a local feature vector (x_j, y_j). By definition, $B(x_j, y_j) = 1$. We define the orientation and magnitude of the local feature vector having spatial coordinate (x_j, y_j) with the same method that we used in the previous section. As a result, we obtain our local features as $k_g(j) = (x_j, y_j, \theta_j, w_j)$ for $j = 1, ..., K_g$ where K_g is the total number of extracted GMSR features.

D. FAST Features

Rosten *et al.* [10] introduced the FAST method to detect corners in images in a fast and reliable manner. The method depends on wedge-model-style corner detection and machine learning techniques. This method can briefly be explained as follows. For each corner candidate pixel, its 16 neighbors are checked. If there exist nine contiguous pixels passing a set of tests, the candidate pixel is labeled as a corner. These tests are done using machine learning techniques to speed up the operation.

In our study, we finally pick FAST to extract local features from given test image. Assume that, we have a local feature (FAST corner) at (x_j, y_j) coordinate. We define the orientation and magnitude of the local feature, using the same method that we used in the Harris-corner-based local feature extraction. As result, we obtain local features as $k_s(j) = (x_j, y_j, \theta_j, w_j)$ for $j = 1, ..., K_s$ where K_s is the total number of extracted FAST-based features.

III. Urban Region Detection

Each extracted local feature indicates a characteristic part of a building to be detected in the image. However, only one feature is not sufficient enough to detect urbanized region. In fact, more local features increases detection probability of interest region. To solve problem, we formulate our urban region detection method with a probabilistic framework. To do so, we represent locations which are holding possible building characteristics as discrete joint random variables. We then estimate their pdf by taking local features as observations. Here, we benefit from a variable-kernel density estimation method. Next, we introduce our probabilistic urban region detection framework using it. Finally, we introduce data and decision fusion methods based on our probabilistic framework to detect buildings.

A. Kernel based Density Estimation

Silverman [12]defines the kernel density estimator for a discrete and bivariate pdf as follows. First, the bivariate kernel function, $N(x,y)$ should satisfy the conditions

$$\sum_x \sum_y N(x,y) = 1 \tag{14}$$

and

$$N(x,y) \geq 0 \quad \forall (x,y) \tag{15}$$

The pdf estimator with kernel $N(x,y)$ is defined by

$$p(x,y) = \frac{1}{nh} \sum_{i=1}^{n} N\left(\frac{x-x_i}{h}, \frac{y-y_i}{h}\right) \tag{16}$$

where h is the window width (also called the smoothing parameter) and (x_i, y_i) for $i = 1, ..., n$ are observations from the pdf to be estimated.

If observations can not be represented reliably by a fixed kernel function, then a variable kernel function can be used. This is achieved by adaptation of the amount of smoothing to the local density of the data (observation). Hence, the scale parameter is allowed to vary from one observation point to another. Besides, the estimate is constructed similarly to the classical kernel estimate. The pdf estimate given in Eqn. 16 then becomes

$$p_v(x,y) = \frac{1}{nh} \sum_{i=1}^{n} \frac{1}{\sigma_i} N\left(\frac{x-x_i}{h\sigma_i}, \frac{y-y_i}{h\sigma_i}\right) \tag{17}$$

where σ_i is the variable scale parameter for $i = 1, ..., n$.

B. Detecting Urban Regions Using Variable Kernel based Density Estimation

As we mentioned previously, we use local features (k_h, k_g, k_f, k_s) as observations to estimate the pdf. Without loss of generality, we explain pdf estimation on a generic local feature vector $k = (x_i, y_i, \theta_i, w_i)$ for $i = 1, ..., K_i$. These features provide information about urban region having buildings characteristics. In order to estimate pdf, each local feature will have its effect on (x_i, y_i) coordinate. Using $N(x,y)$ in Eqn. 17 as a Gaussian symmetric pdf, which is used in most density estimation applications, we form the estimated pdf as

$$p(x,y) = \frac{1}{R} \sum_{i=1}^{K_i} \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left(-\frac{(x-\hat{x}_i)^2 + (y-\hat{y}_i)^2}{2\sigma_i}\right) \tag{18}$$

where $\sigma_i = w_i$ and R is the normalizing constant. We will use $p(x,y)$ pdf function to detect urban region boundaries. To do so, we apply automatic thresholding on obtain pdf ???

C. Data Fusion and Decision Fusion for Urban Region Detection

The four local feature extraction methods extract different information from the same image. In the previous section, we introduced how to separately use this information to detect urban region boundaries. However, their fusion can also improve our detection performance. Fortunately, our probabilistic urban region detection approach allows fusion of information. Therefore, in this section we introduce two feature fusion methods using probabilistic framework.

Our first method is based on data fusion. This method is straightforward, such that we use all the local features extracted with different methods as one unique group. In other saying, $k_F = \{k_h, k_g, k_f, k_s\}$. We estimate the pdf using Eqn. 18 with the local feature set k_F. We detect urban region boundaries from the estimated pdf with the same method in the previous section.

Our second method is based on decision fusion. Here, we mix the estimated pdfs by different methods and obtain a final pdf. While mixing the estimated pdfs, we assign a weight to each of them directly proportional to their maximum mode value. As we mentioned in the previous section, in detecting building locations from the estimated pdf we label the mode with the maximum value as a building. By normalizing four different pdfs this way, we can mix them and obtain the final pdf estimate as

$$p_D(x,y) = \frac{1}{R} \sum_{l=\{h,g,f,s\}} \frac{p_l(x,y)}{\max_{(x,y)} p_l(x,y)} \tag{19}$$

where $p_h(x,y)$, $p_g(x,y)$, $p_f(x,y)$, and $p_s(x,y)$ are the estimated pdfs from k_h, k_g, k_f, and k_s. R is again the normalizing constant. We call this method as decision fusion, since we apply the fusion operation close to the urban region detection step. Again, we use the urban region detection method in the previous section on $p_D(x,y)$ to detect urban region boundaries.

IV. Experiments

We test our probabilistic urban region detection methods on panchromatic Ikonos and grayscale aerial images of Istanbul and Adana cities of Turkey.

V. Conclusion

Herein, we introduced a novel urban region detection method based on a probabilistic framework. To do so, we defined urban region pixels to be detected as joint random variables. We formed a pdf which shows probabilities of belonging to an urban region for each image pixel. In estimating the pdf, we used local features which are extracted from image using four different methods. However, our probabilistic building detection framework is not limited to these four methods. It can be applied to other local feature extraction methods as well. Then, we detected urban region boundaries using the estimated pdf. We further improved our urban region detection method by introducing two feature fusion methods; data fusion and decision fusion. Obtained performances of the proposed approach on panchromatic Ikonos satellite and grayscale aerial images show robustness of algorithm even on images obtained from very different kind of sensors. We can conclude that our probabilistic urban region detection method can be used in real-life applications in a very fast and reliable manner.

REFERENCES

[1] J. 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 (9), pp. 1940–1949, Sep. 2003.

[2] S. Bhagavathy and B. Manjunath, "Modeling and detection of geospatial objects using texture motifs," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44 (12), pp. 3706–3715, Dec. 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 (9), pp. 2587–2600, Sep. 2006.

[4] M. Fauvel, J. Chanussot, and J. Benediktsson, "Decision fusion for the classification of urban remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44 (10), pp. 2828–2838, Oct. 2006.

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

[6] C. Harris and M. Stephens.

[7] 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 (9), pp. 1807–1823, Jun. 2000.

[8] V. Kyrki, J. Kamarainen, and H. Kalviainen, "Simple gabor feature space for invariant object recognition," *Pattern Recognition Letters*, vol. 25 (3), pp. 311–318, Feb. 2004.

[9] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-9 (1), pp. 62–66, Jan. 1979.

[10] E. Rosten, R. Porter, and T. Drummond.

[11] C. Schmid, R. Mohr, and C. Bauckhage.

[12] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, 1st ed. Chapman Hall, 1986.

[13] B. Sırmaçek and C. Ünsalan.

[14] ——, "Urban area detection using local feature points and spatial voting," *IEEE Geoscience and Remote Sensing Letters*, vol. 7 (1), pp. 146–150, Jan. 2010.

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

[16] C. Ünsalan, "Gradient magnitude based support regions in structural land use classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 3 (4), pp. 546–550, Oct. 2006.

[17] C. Ünsalan and K. Boyer, "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 (4), pp. 575–589, Apr. 2005.

[18] ——, "Classifying land development in high resolution satellite imagery using hybrid structural multispectral features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42 (12), pp. 2840–2850, Dec. 2004.

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

[20] 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 (12), pp. 3978–3988, Dec. 2007.