

MULTIDIRECTIONAL BUILDING DETECTION IN AERIAL IMAGES WITHOUT SHAPE TEMPLATES

Andrea Manno-Kovacs^{a,b} and Tamas Sziranyi^b

^a Pazmany Peter Catholic University, Prater 50/A, 1083, Budapest, Hungary

^b MTA SZTAKI, Kende 13-17, 1111, Budapest, Hungary

andrea.kovacs@sztaki.mta.hu, sziranyi@sztaki.mta.hu

KEY WORDS: Directional Classification, Building Detection, Feature Extraction, Aerial Image Detection

ABSTRACT:

The aim of this paper is to exploit orientation information of an urban area for extracting building contours without shape templates. Unlike using shape templates, these given contours describe more variability and reveal the fine details of the building outlines, resulting in a more accurate detection process, which is beneficial for many tasks, like map updating and city planning. According to our assumption, orientation of the closely located buildings is coherent, it is related to the road network, therefore adaptation of this information can lead to more efficient building detection results.

The introduced method first extracts feature points for representing the urban area. Orientation information in the feature point neighborhoods is analyzed to define main orientations. Based on orientation information, the urban area is classified into different directional clusters. The edges of the classified building groups are then emphasized with shearlet based edge detection method, which is able to detect edges only in the main directions, resulting in an efficient connectivity map. In the last step, with the fusion of the feature points and connectivity map, building contours are detected with a non-parametric active contour method.

1 INTRODUCTION

Automatic building detection is currently a relevant topic in aerial image analysis, as it can be an efficient tool for accelerating many applications, like urban development analysis, map updating and also means a great support in crisis situations for disaster management and helps municipalities in long-term residential area planning. These continuously changing, large areas have to be monitored periodically to have up-to-date information, which means a big effort when administrated manually. Therefore, automatic processes are really welcomed to facilitate the analysis.

There is a wide range of publications in remote sensing topic for building detection, however we concentrated on the newer ones, which we also used for comparison in the experimental part. State-of-the-art methods can be divided into two main groups. The first group only localizes buildings without giving any shape information, like (Sirmaçek and Ünsalan, 2009) and (Sirmaçek and Ünsalan, 2011).

In (Sirmaçek and Ünsalan, 2009) a SIFT (Lowe, 2004) salient point based approach is introduced for urban area and building detection (denoted by SIFT-graph in the experimental part). This method uses two templates (a light and dark one) for detecting buildings. After extracting feature points representing buildings, graph based techniques are used to detect urban area. The given templates help to divide the point set into separate building subsets, then the location is defined. However, in many cases, the buildings cannot be represented by such templates, moreover sometimes it is hard to distinguish them from the background based on the given features.

To compensate the drawbacks and represent the diverse characteristics of buildings, the same authors proposed a method in (Sirmaçek and Ünsalan, 2011) to detect building positions in aerial and satellite images based on Gabor filters (marked as Gabor filters), where different local feature vectors are used to localize buildings with data and decision fusion techniques. Four different local feature vector extraction methods are proposed to be

used as observations for estimating the probability density function of building locations by handling them as joint random variables. Data and decision fusion methods define the final building locations based on the probabilistic framework.

The second group also provides shape information beside location, but usually applies shape templates (e.g. rectangles), like (Benedek et al., 2012). However, this latter case still just gives an approximation of the real building shape.

A very novel building detection approach is introduced in (Benedek et al., 2012), using a global optimization process, considering observed data, prior knowledge and interactions between the neighboring building parts (marked later as bMBD). The method uses low-level (like gradient orientation, roof color, shadow, roof homogeneity) features which are then integrated to have object-level features. After having object (building part) candidates, a configuration energy is defined based on a data term (integrating the object-level features) and a prior term, handling the interactions of neighboring objects and penalizing the overlap between them. The optimization process is then performed by a bi-layer multiple birth and death optimization.

In our previous work (Kovacs and Sziranyi, 2012) we have introduced an orientation based method for building detection in unidirectional aerial images regardless of shape, and pointed out that orientation of the buildings is an important feature when detecting outlines and this information can help to increase detection accuracy. Neighboring building segments or groups cannot be located arbitrarily, they are situated according to some bigger structure (e. g. the road network), therefore the main orientation of such area can be defined. We have also introduced Modified Harris for Edges and Corners (MHEC) point set in (Kovacs and Sziranyi, 2013) which is able to represent urban areas efficiently.

This paper presents contribution in the issue of processing multiple directional urban areas. Building groups of different orientations can be classified into clusters and orientation-sensitive shearlet edge detection (Yi et al., 2009) can be performed separately for such clusters. Finally, building contours are detected

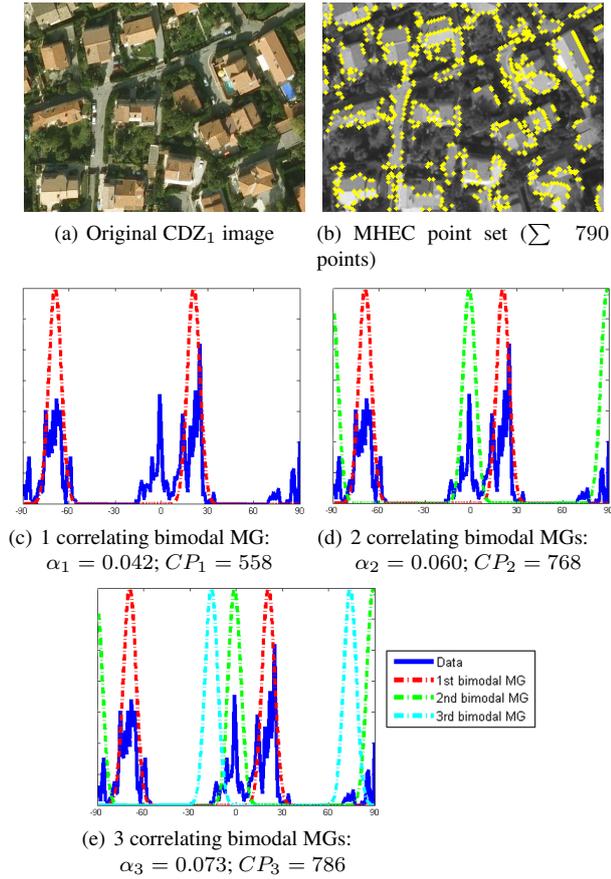


Figure 1: Correlating increasing number of bimodal Mixture of Gaussians (MGs) with the ϑ orientation density function (marked in blue). The measured α_q and CP_q parameters are represented for each step. The third component is found to be insignificant, as it covers only 18 MHEC points. Therefore the estimated number of main orientations is $\mathbf{q} = 2$.

based on the fusion of feature points and connectivity information, by applying Chan-Vese active contour method (Chan and Vese, 2001).

2 ORIENTATION BASED CLASSIFICATION

MHEC feature point set for urban area detection (Kovacs and Sziranyi, 2013) is based on the Harris corner detector (Harris and Stephens, 1988), but adopts a modified $R_{\text{mod}} = \max(\lambda_1, \lambda_2)$ characteristic function, where λ_s denote the eigenvalues of the Harris matrix. The advantage of the improved detector is that it is automatic and it is able to recognize not just corners, but edges as well. Thus, it gives an efficient tool for characterizing contour-rich regions, such as urban areas. MHEC feature points are calculated as local maxima of the R_{mod} function (see Fig.1(b)).

As the point set is showed to be efficient for representing urban areas, orientation information in the close proximity of the feature points is extracted. To confirm the assumption about connected orientation feature of closely located buildings, specific images were used in our previous work (Kovacs and Sziranyi, 2012), presenting only small urban areas and having only one main direction. In the present work, we extended the introduced, unidirectional method, to be able to handle bigger urban areas with multiple directions.

(Benedek et al., 2012) used a low level feature, called local gradient orientation density, where the surroundings of a pixel was investigated whether it has perpendicular edges or not. This method was adapted to extract the main orientation information characterizing the feature point, based on it's surroundings. Let us denote the gradient vector by ∇g_i with $\|\nabla g_i\|$ magnitude and φ_i^∇ orientation for the i^{th} point. By defining the $n \times n$ neighborhood of the point with $W_n(i)$ (where n depends on the resolution), the weighted density of φ_i^∇ is as follows:

$$\lambda_i(\varphi) = \frac{1}{N_i} \sum_{r \in W_n(i)} \frac{1}{h} \cdot \|\nabla g_r\| \cdot \kappa\left(\frac{\varphi - \varphi_r^\nabla}{h}\right), \quad (1)$$

with $N_i = \sum_{r \in W_n(i)} \|\nabla g_r\|$ and $\kappa(\cdot)$ kernel function with h bandwidth parameter.

Now, the main orientation for (i^{th}) feature point is defined as:

$$\varphi_i = \underset{\varphi \in [-90, +90]}{\operatorname{argmax}} \{\lambda_i\}. \quad (2)$$

After calculating the direction for all the K feature points, the density function ϑ of their orientation is defined:

$$\vartheta(\varphi) = \frac{1}{K} \sum_{i=1}^K H_i(\varphi), \quad (3)$$

where $H_i(\varphi)$ is a logical function:

$$H_i(\varphi) = \begin{cases} 1, & \text{if } \varphi_i = \varphi \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

In the unidirectional case, the density function ϑ is expected to have two main peaks (because of the perpendicular edges of buildings), which is measured by correlating ϑ to a bimodal density function:

$$\alpha(m) = \int \vartheta(\varphi) \eta_2(\varphi, m, d_\vartheta) d\varphi, \quad (5)$$

where $\eta_2(\cdot)$ is a two-component Mixture of Gaussian (MG), with m and $m + 90$ mean values and d_ϑ is the standard deviation for both components. The value θ of the maximal correlation can be obtained as:

$$\theta = \underset{m \in [-90, +90]}{\operatorname{argmax}} \{\alpha(m)\}. \quad (6)$$

And the corresponding orthogonal direction (the other peak):

$$\theta_{\text{ortho}} = \begin{cases} \theta - 90, & \text{if } \theta \geq 0 \\ \theta + 90, & \text{otherwise} \end{cases} \quad (7)$$

If the urban area is larger, there might be building groups with multiple orientations. However, the buildings are still oriented according to some bigger structure (like the road network) and cannot be located arbitrarily, orientation of the closely located buildings is coherent. In this case the ϑ density function of the φ_i values is expected to have more peak pairs: $2q$ peaks ($[\theta_1, \theta_{\text{ortho},1}]$, \dots , $[\theta_q, \theta_{\text{ortho},q}]$) for q main directions. As the value of q is unknown, it has to be estimated by correlating multiple bimodal Gaussian functions to the ϑ density function. The correlation is measured by $\alpha(m)$ (see Eq. 5), therefore the behavior of α values has been investigated for increasing number of $\eta_2(\cdot)$ two-component MG functions. When the number of the correlating bimodal MGs is increasing, the α value should also be increasing or remaining nearly constant (a slight decreasing is acceptable), until a correct estimation number is reached, or the correlating data involves enough points (the number of correlated points has reached a given ratio), the ratio in this case has been set to 95%.

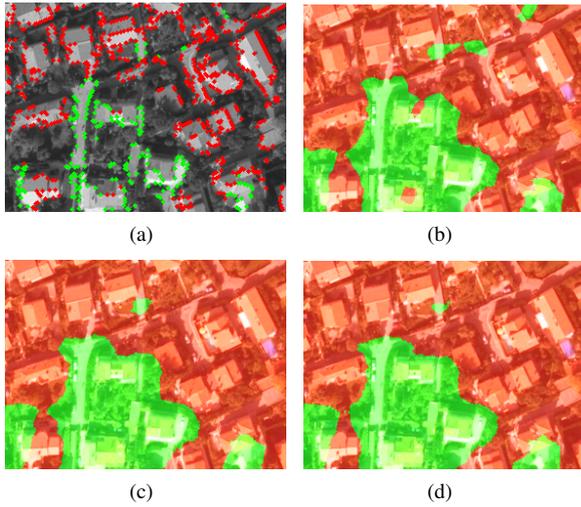


Figure 2: Orientation based classification for $q = 2$ main orientations with k -NN algorithm for image 1(a): (a) shows the classified MHEC point set, (b)–(d) is the classified image with $k = 3$, $k = 7$ and $k = 11$ parameter values. Different colors show the clusters belonging to the bimodal GMs in figure 1(d).

Based on these criteria, the value of the α_q parameter and the total number of the Correlated Points (CP_q) are investigated when correlating the data to q bimodal MGs.

Figure 1 shows the steps of defining the number of main directions (q). The calculated MHEC points for the image is in Figure 1(b), including altogether 790 points. The correlating bimodal MGs and the belonging parameters are in Fig. 1(c)–1(e). As one can see, the α_q parameter is increasing continuously and the CP_q parameter has reached the defined ratio (95%) in the second step (representing $768/790 \approx 97\%$ of the point set). The third MG (Fig. 1(e)) is just added for illustrating the behavior of the correlation step: although α_q is still increasing, the newly correlated point set is too small, containing only $CP_3 - CP_2 = 18$ points and supposed to be irrelevant. Therefore, the estimated number of main orientation is $q = 2$, with peaks $\theta_1 = 22$ ($\theta_{1,ortho} = -68$) and $\theta_2 = 0$ ($\theta_{2,ortho} = 90$).

The point set is then classified by K-means algorithm, where K is the number of main orientation peaks ($2q$) and the distance measure is the difference between the orientation values. After the classification, the 'orthogonal' clusters (2 peaks belonging to the same bimodal MG component) are merged, resulting in q clusters. The clustered point set is in Figure 2(a).

The orientation based classification is then extended to the whole image, k -NN clustering is performed to classify the image pixel-wisely. Classification has been tested with different k values (3, 7 and 11), Figure 2(b)–(d) show the results respectively, different colors marks the clusters with different orientations. The same color is picked for the correlating bimodal MG-s in Figure 1(d) and for the area belonging to the corresponding cluster in Figure 2. The tests have proved that the classification results are not sensitive to the k parameter, therefore in the further evaluation, a medium value, $k = 7$ was chosen.

The classification map defines the main orientation for each pixel of the image, therefore in the edge detection part, connectivity information in the given direction has to be extracted.

3 SHEARLET BASED CONNECTIVITY MAP EXTRACTION

Now, that the main direction is given for every pixel in the image, edges in the defined direction have to be strengthened. There are different approaches which uses directional information like Canny edge detection (Canny, 1986) using the gradient orientation; or (Perona, 1998) which is based on anisotropic diffusion, but cannot handle the situation of multiple orientations (like corners). Other single orientation methods exist, like (Mester, 2000) and (Bigun et al., 1991), but the main problem with these methods is that they calculate orientation in pixel-level and lose the scaling nature of orientation, therefore they cannot be used for edge detection. In the present case, edges constructed by joint pixels has to be enhanced, thus the applied edge detection method has to be able to handle orientation. Moreover, as searching for building contours, the algorithm must handle corner points as well. Shearlet transform (Yi et al., 2009) has been lately introduced for efficient edge detection, as unlike wavelets, shearlets are theoretically optimal in representing images with edges and, in particular, have the ability to fully capture directional and other geometrical features. Therefore, this method is able to emphasize edges only in the given directions (Fig.3(a)).

For an image u , the shearlet transform is a mapping:

$$u \rightarrow SH_\psi u(a, s, x), \quad (8)$$

providing a directional scale-space decomposition of u with $a > 0$ is the scale, s is the orientation and x is the location:

$$SH_\psi u(a, s, x) = \int u(y) \psi_{as}(x - y) dy = u * \psi_{as}(x), \quad (9)$$

where ψ_{as} are well localized waveforms at various scales and orientations. When working with a discrete transform, a discrete set of possible orientations is used, for example $s = 1, \dots, 16$. In the present case, the main orientation(s) of the image θ are calculated, therefore the aim is to strengthen the components in the given directions on different scales as only edges in the main orientations have to be detected. The first step is to define the s subband for image pixel (x_i, y_i) which includes θ_i and $\theta_{i,ortho}$:

$$\begin{aligned} \tilde{s}_{1,\dots,q} &= \left\{ s_i : (i-1) \frac{2\pi}{s} < \theta_{1,\dots,q} \leq i \frac{2\pi}{s} \right\}, \\ \tilde{s}_{1,\dots,q,ortho} &= \left\{ s_j : (j-1) \frac{2\pi}{s} < \theta_{1,\dots,q,ortho} \leq j \frac{2\pi}{s} \right\}. \end{aligned} \quad (10)$$

After this, the $SH_\psi u(a, \tilde{s}_{1,\dots,q}, x)$ and $SH_\psi u(a, \tilde{s}_{1,\dots,q,ortho}, x)$ subbands have to be strengthened at (x_i, y_i) . For this reason, the weak edges (values) have been eliminated with a hard threshold and only the strong coefficients are amplified.

Finally, the shearlet transform is applied backward (see Eq.9) to get the reconstructed image, which will have strengthened edges in the main directions. The strengthened edges can be easily detected by Otsu thresholding (Otsu, 1979). The advantage of applying shearlet method is while the pure Canny method detects the edges sometimes with discontinuities, the shearlet based edge strengthening helps to eliminate this problem and the given result represents connectivity relations efficiently.

We used the u^* component of the CIE L^*u^*v advised in (Muller and Zaum, 2005), which is also adapted in other state-of-the-art method (Benedek et al., 2012) for efficient building detection. As the u^* channel emphasizes the red roofs as well, the Otsu adaptive thresholding may also detects these pixels with high intensity

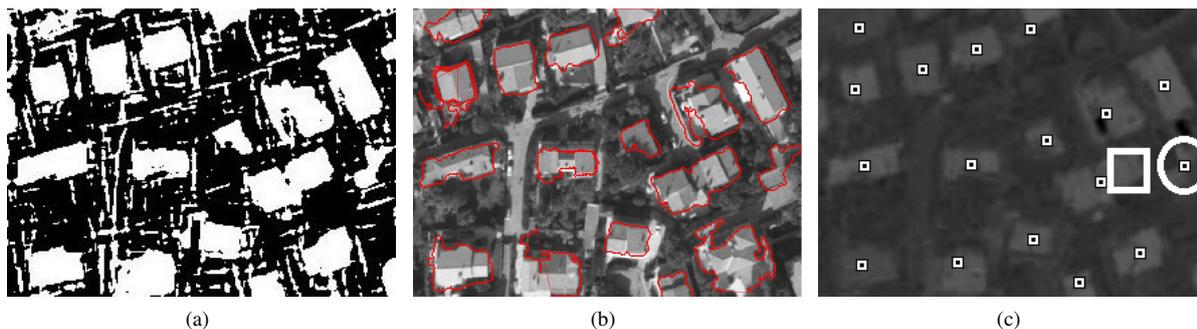


Figure 3: Steps of multidirectional building detection: (a) is the connectivity map; (b) shows the detected building contours in red; (c): marks the estimated location (center of the outlined area) of the detected buildings, the falsely detected object is marked with a white circle, missed object is marked with a white rectangle.

values in the edge strengthened map (see Figure 3(a)), therefore the extracted map is better to be called as a connectivity map. In case of buildings with altering colour (as gray or brown), only the outlining edges are detected.

4 MULTIDIRECTIONAL BUILDING DETECTION

Initial building locations can be defined by fusing the feature points as vertices (V) and the shearlet based connectivity map as the basis of the edge network (E) of a $G = (V, E)$ graph. To exploit building characteristics for the outline extraction, we have to determine point subsets belonging to the same building. Coherent point subsets are defined based on their connectivity, $v_i = (x_i, y_i)$ and $v_j = (x_j, y_j)$, the i^{th} and j^{th} vertices of the V feature point set are connected in E , if they satisfy the following conditions:

1. $S_{(x_i, y_i)} = 1$,
2. $S_{(x_j, y_j)} = 1$,
3. \exists a finite path between v_i and v_j in S .

The result after the connecting procedure is a G graph composed of many separate subgraphs, where each subgraph indicates a building candidate. However, there might be some singular points and some smaller subgraphs (points and edges connecting them) indicating noise. To discard them, only subgraphs having points over a given threshold are selected.

Main directional edge emphasis may also enhance road and vegetation contours, moreover some feature points can also be located on these edges. To filter out false detections, the directional distribution of edges ($\lambda_i(\varphi)$ in Eq. 1) is evaluated in the extracted area. False objects, like road parts or vegetation, have unidirectional or randomly oriented edges in the extracted area (see Fig. 4(b) and 4(d)), unlike buildings, which have orthogonal edges (Fig. 4(c) and 4(e)). Thus, the non-orthogonal hits are eliminated with a decision step.

Finally, contours of the subgraph-represented buildings are calculated by region-based Chan-Vese active contour method (Chan and Vese, 2001), where the initialization of the snake is given as the convex hull of the coherent point subset.

A typical detection result is shown in Figure 3(b) with the building outlines in red. In the experimental part, the method was evaluated quantitatively and compared to other state-of-the-art processes. In this case the location of the detected buildings was used, which is estimated as the centroid of the given contours (see Figure 3(c)).

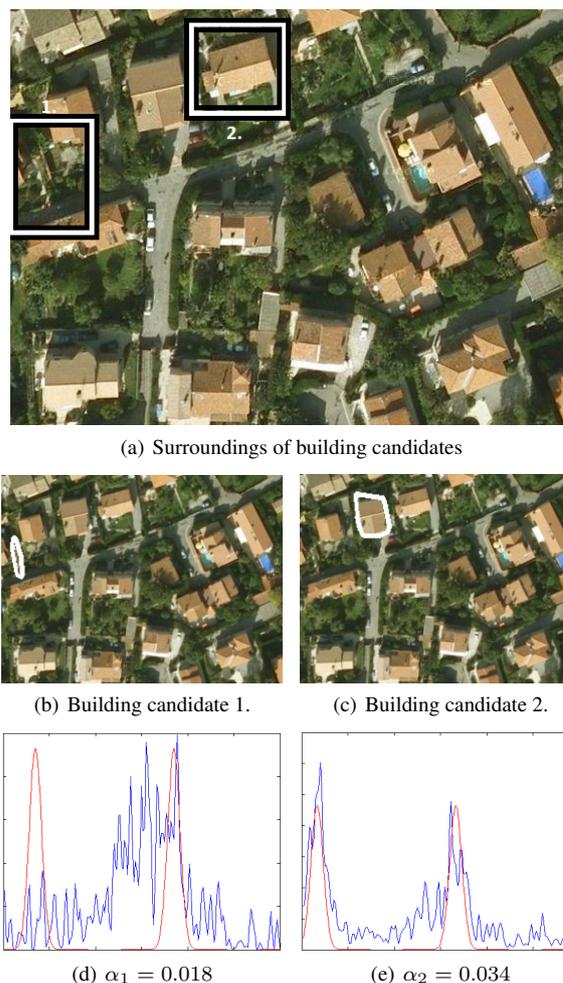


Figure 4: Elimination of false detection based on directional distribution of edges in the extracted area: 1. area is a false detection, 2. area is a building. (b)-(c): Extracted areas by the graph-based connection process. (d)-(e): The calculated $\lambda_i(\varphi)$ directional distribution and the resulting α values of the area.

5 EXPERIMENTS

The proposed method was evaluated on different databases, previously used in (Benedek et al., 2012). Smaller, multidirectional image parts (like Figure 1(a)) were collected from the databases Budapest, Côte d'Azur (CDZ) and Normandy to test the orientation estimation process. The quantitative evaluation is in Table

Database			Performance							
			SIFT-graph		Gabor features		bMBD		Prop. MultiDir	
Image name	Nr. of buildings	Nr. of directions	FD	MD	FD	MD	FD	MD	FD	MD
Budapest ₁	14	3	3	9	1	4	2	0	0	0
CDZ ₁	14	2	2	5	4	1	1	0	1	1
CDZ ₂	7	2	1	3	2	2	1	0	0	0
CDZ ₃	6	3	0	1	1	0	0	1	0	0
CDZ ₄	10	4	0	5	1	0	2	1	0	0
CDZ ₅	3	3	1	2	1	0	1	1	0	0
Normandy ₁	19	4	2	9	3	2	1	4	1	3
Normandy ₂	15	3	4	9	4	5	3	2	0	1
Total F-score			0.616		0.827		0.888		0.960	

Table 1: Quantitative results on different databases. The performance of SIFT-graph (Sirmaçek and Ünsalan, 2009), Gabor features (Sirmaçek and Ünsalan, 2011), bMBD (Benedek et al., 2012) and the proposed multidirectional (MultiDir) methods are compared. Nr. of buildings indicates the number of completely visible, whole buildings in the image. FD and MD denote the number of False and Missed Detections (false positives and false negatives). **Best** results in every row are marked in bold.

1, where the number of detected buildings were compared based on the estimated location (Fig. 3(c)). The overall performance of different techniques was measured by the F-measure:

$$P = \frac{TD}{TD + FD}, \quad R = \frac{TD}{TD + MD}, \quad F = 2 \cdot \frac{P \cdot R}{P + R}, \quad (11)$$

where TD, FD and MD denote the number of true detections (true positive), false detections (false positive) and missed detections (false negative) respectively.

Results showed that the proposed multidirectional method obtains the highest detection accuracy when evaluating the object level performance. Further tests are needed to compare the pixel level performance. By analyzing the results, we have pointed out, that the proposed method has difficulties when detecting buildings with altering colors (like gray or brown roofs). However, orientation sensitive edge strengthening is able to partly compensate this drawback. Sometimes, the closely located buildings are contracted and treated as the same object (see Figure 3). The method may also suffer from the lack of contrast difference between the building and the background and it is not able to detect the proper contours.

6 CONCLUSION

We have proposed a novel, orientation based approach for building detection in aerial images without using any shape templates. The method first calculates feature points with the Modified Harris for Edges and Corners (MHEC) detector, introduced in our earlier work. Main orientation in the close proximity of the feature points is extracted by analyzing the local gradient orientation density. Orientation density function is defined by processing the orientation information of all feature points, and the main peaks defining the prominent directions are determined by bimodal Gaussian fitting. Based on the main orientations, the urban area is classified into different directional clusters. Edges with the orientation of the classified urban area are emphasized with shearlet based edge detection method, resulting in an efficient connectivity map. The feature point set and the connectivity map is fused in the last step, to get the initial allocation of the buildings and perform an iterative contour detection with a non-parametric active contour method.

The proposed model is able to enhance the detection accuracy on object level performance, however still suffering of typical challenges (altering building colors and low contrasted outlines). In our further work, we will focus on the analysis of different color spaces, to represent altering building colors more efficiently and

enhance detection results by reducing the number of missed detections. Application of prior constraints (like edge parts running in the defined main orientations) may help in the detection of low contrasted building contours.

REFERENCES

- Benedek, C., Descombes, X. and Zerubia, J., 2012. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. *IEEE Trans. Pattern Analysis and Machine Intelligence* 34(1), pp. 33–50.
- Bigun, J., Granlund, G. H. and Wiklund, J., 1991. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Trans. Pattern Analysis and Machine Intelligence* 13(8), pp. 775–790.
- Canny, J., 1986. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence* 8(6), pp. 679–698.
- Chan, T. F. and Vese, L. A., 2001. Active contours without edges. *IEEE Trans. Image Processing* 10(2), pp. 266–277.
- Harris, C. and Stephens, M., 1988. A combined corner and edge detector. In: *Proceedings of the 4th Alvey VisionConference*, pp. 147–151.
- Kovacs, A. and Sziranyi, T., 2012. Orientation based building outline extraction in aerial images. In: *ISPRS Annals of Photogrammetry, Remote Sensing and the Spatial Information Sciences (Proc. ISPRS Congress)*, Vol. I-7, Melbourne, Australia, pp. 141–146.
- Kovacs, A. and Sziranyi, T., 2013. Improved Harris feature point set for orientation sensitive urban area detection in aerial images. *IEEE Geoscience and Remote Sensing Letters* 10(4), pp. 796–800.
- Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, pp. 91–110.
- Mester, R., 2000. Orientation estimation: Conventional techniques and a new non-differential approach. In: *Proc. 10th European Signal Processing Conference*.
- Muller, S. and Zaum, D., 2005. Robust building detection in aerial images. In: *CMRT, Vienna, Austria*, pp. 143–148.
- Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Trans. Systems, Man and Cybernetics* 9(1), pp. 62–66.

Perona, P., 1998. Orientation diffusion. *IEEE Trans. Image Processing* 7(3), pp. 457–467.

Sırmaçek, B. and Ünsalan, C., 2009. Urban-area and building detection using SIFT keypoints and graph theory. *IEEE Trans. Geoscience and Remote Sensing* 47(4), pp. 1156–1167.

Sırmaçek, B. and Ünsalan, C., 2011. A probabilistic framework to detect buildings in aerial and satellite images. *IEEE Trans. Geoscience and Remote Sensing* 49(1), pp. 211–221.

Yi, S., Labate, D., Easley, G. R. and Krim, H., 2009. A shearlet approach to edge analysis and detection. *IEEE Trans. Image Processing* 18(5), pp. 929–941.