

A JOINT APPROACH OF BUILDING LOCALIZATION AND OUTLINE EXTRACTION

Andrea Kovács^{1,2}

¹Department of Information Technology
Péter Pázmány Catholic University
H-1083 Práter utca 50/A
email: kovacs1@digitus.itk.ppke.hu

Csaba Benedek and Tamás Szirányi²

²Distributed Events Analysis Research Group
Computer and Automation Research Institute
H-1111, Kende utca 13-17, Budapest, Hungary
email: {bcsaba,sziranyi}@sztaki.hu

ABSTRACT

In this paper we introduce a novel method for building localization and 2D outline extraction in remotely sensed images. A robust Marked Point Process (MPP) model attempts to detect and separate the individual building segments and gives a rough rectangular estimation about the geometry of each entity. The refinement of the detection is achieved by an active contour model, which is initialized by the convex hull of the Harris feature points extracted around the MPP step's output rectangles. The method is tested in real aerial images provided by the Hungarian Institute of Geodesy, Cartography and Remote Sensing.

KEY WORDS

Marked point process, remote sensing, building detection, active contour, Harris detector

1 Introduction

Building extraction is a key task in several aerial surveillance, exploration and reconnaissance applications. In the present paper we deal with a single optical image input (an aerial photo or a satellite image), where building identification is a hard monocular object recognition problem. We can divide the previous methods into three different groups based on the levels of solving the task. First, some techniques address purely detection and counting the houses [12], by extracting the estimated center points of the buildings without any geometrical information. Although these approaches are efficient in computational complexity and less sensitive for parameter tuning, they can hardly be adopted for building classification or change detection which require more detailed shape, size and spectral characterization of the entities. A second group of the existing approaches uses simplified geometric approximations of the buildings or building parts, such as rectangles [11] or hexagons [10]. These models are suitable for a more comprehensive interpretation of the scenes, however fine details and the irregular patterns can not at all or only inaccurately be retrieved. As the resolution of the available remotely sensed images increases rapidly, such rough modeling of the building shapes yields a notable amount of lost information for the higher level interpretation modules. These effects increase the significance of a third group of

approaches, which aim to describe the accurate building outlines by general parametric curves [6]. Variational techniques are efficient tools to address contour approximation, meanwhile they can embed prior information for the building outlines. However, contour initialization is challenging, especially, if the scene can not be described by a few template building shapes as in [6]. To achieve an efficient outline estimation, corner detectors [5] can be applied to extract contour points of the objects. Based on these contour points, active contour algorithms [13, 2, 8, 9] are able to detect the outline of the buildings.

2 Proposed approach

The proposed approach consists in two steps. *First*, we attempt to localize the buildings in the scene and roughly estimate the shape of each rooftop. This task is fulfilled by a Marked Point Process (MPP) model [1], which fits rectangles to the observed building candidates. *Second*, using the coarse location and shape information we identify feature points inside and around the rectangular rooftop masks, and based on the selected key points we initialize an active contour [7] to obtain the building outlines. The main contribution of this paper lies in the linking process between the two steps: we propose a novel key point selection technique for the initial contour, which jointly exploits low level features from the original image and structural descriptors from the MPP output mask. Finally the efficiency of the cascade procedure is tested on real aerial images and discussion is given about the results.

3 Preliminary Building Mask Estimation

The goal of this model part is to assign to each building segment of the scene a unique rectangle which estimates the mask of the observed rooftop.

Let us denote by S the 2-D pixel lattice of the input aerial image, and by $s \in S$ a single pixel. A rectangular building segment candidate u is described by four parameters: $s_c = [c_x, c_y] \in S$ center pixel, $e_L, e_t \in [L_{\min}, L_{\max}]$ side lengths and $\theta \in [-90^\circ, +90^\circ]$ orientation [see Fig. 1(a)].

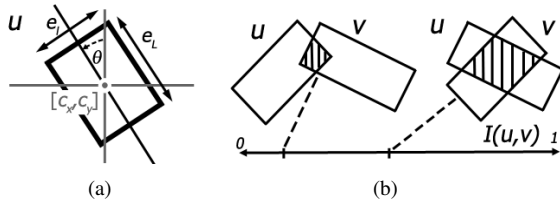


Figure 1. Demonstration of the (a) object rectangle parameters and (b) calculation of the interaction potentials

3.1 Marked Point Process Model

We model an aerial image as composition of an arbitrary number of u rectangular Building Segments (BS) whose positions and geometric parameters are realizations of a MPP [4]. Let us denote by Ω the space of all configurations of a finite number of BS objects. We will refer to a given object configuration in Ω by ω , where $\omega = \emptyset$ or $\omega = \{u_1, \dots, u_n\}$ for an arbitrary positive integer n . We also define a neighborhood relation in the BS object space: $u \sim v$ if their rectangles intersect.

Next, we introduce a $\Phi(\omega)$ energy function on the Ω configuration space, so that we aim to fulfill that the optimal configuration exhibits the minimal energy. The energy function takes into account the interactions between the geometric objects (the prior energy $\Phi_p(\omega)$), and the way they fit to the image data (the data energy $\Phi_d(\omega)$):

$$\Phi(\omega) = \Phi_p(\omega) + \Phi_d(\omega)$$

The optimal BS population $\hat{\omega}$ is obtained as the Maximum Likelihood (ML) configuration estimate:

$$\hat{\omega} = \underset{\omega \in \Omega}{\operatorname{argmin}} \Phi(\omega) \quad (1)$$

3.1.1 Prior energy

We begin with the $\Phi_p(\omega)$ energy part, which prescribes *prior geometric* constraints in the model. Since we aim to extract individual building entities, we must penalize overlapping between different BS rectangles. Therefore, we define $I(u, v)$ *interaction* potentials, which penalize object pairs according to the $I(u, v)$ normalized intersection area [Fig. 1(b)]:

$$I(u, v) = \frac{\operatorname{Area}\{u \cap v\}}{\operatorname{Area}\{u\} + \operatorname{Area}\{v\}}$$

where $u \cap v$ means the intersection of objects u and v .

Finally, the complete prior energy term of the population is calculated as:

$$\Phi_p(\omega) = \sum_{\substack{u, v \in \omega \\ u \sim v}} I(u, v)$$

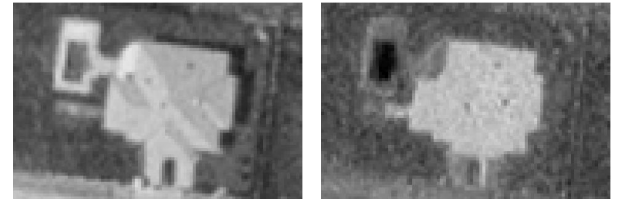


Figure 2. Different color components: (a) is the R component of the RGB color space, (b) is the u^* component of the $L^*u^*v^*$ color space

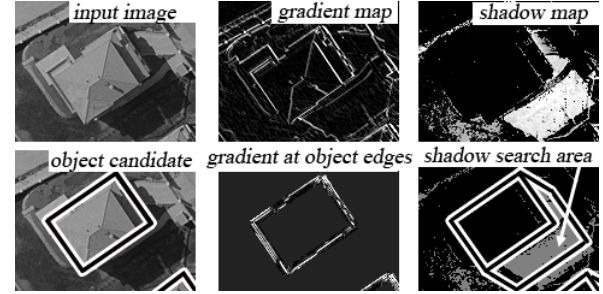


Figure 3. Edge and shadow features

3.1.2 Data energy

The *data energy*, $A_{\mathcal{D}}(u)$, associated with object u , characterizes a proposed building segment $u = \{c_x, c_y, e_L, e_L, \theta\}$ depending on the local image data, but independently of other objects of the population. The data term of the whole configuration is obtained as:

$$\Phi_d(\omega) = \sum_{u \in \omega} A_{\mathcal{D}}(u)$$

A BS u with $A_{\mathcal{D}}(u) < 0$ is called *attractive object*. Since according to Sec. 3.1.1 all prior terms are non-negative, the optimal population (1) should consist of attractive objects exclusively: if $A_{\mathcal{D}}(u) > 0$, removing u from the configuration results in a lower $\Phi(\omega)$ global energy.

To fit the above framework to the building detection task, we need to define the $A_{\mathcal{D}}(u)$ data term appropriately. This term must assigns to each proposed object candidate an energy value (i.e. negative fitness value), which evaluates the hypothesis that u is a building in the image. The proposed energy function, detailed in [1], can integrate different feature-information such as roof color, roof edge and shadow. On one hand, *red* roofs [10] can be detected in color images using the u^* and v^* color components of the corresponding pixel values in CIE $L^*u^*v^*$ color space representation (See Fig. 2). Thereafter with notation R_u for the mask and \tilde{R}_u for the dilated mask of u 's rectangle, the color term prefers objects which contain in majority roof colored pixels inside R_u and background pixels in $\tilde{R}_u \setminus R_u$.



Figure 4. Rectangular footprint results, obtained by the MPP based detector module

For non-red roofs we can rely on the gradient and shadow maps exploiting that under the roof edges strong intensity changes should be observed in the images, while in sunny weather dark shadow blobs are present next to the buildings in the shadow direction [11] (See Fig. 3).

3.2 Optimization

For the estimation of the optimal object configuration (1) we have applied the recent Multiple Birth and Death (MBD) algorithm [4]. The steps are as follows:

Initialization: start with an empty population $\omega = \emptyset$.

Main program: initialize the inverse temperature parameter $\beta = \beta_0$ and the discretization step $\delta = \delta_0$ and a constant basic birth frequency b_0 . Alternate the following three steps, till convergence is reached

1) *Birth step:* Visit all pixels of the S lattice one after the other. At each pixel $s \in S$ call *Object Generation* procedure with probability $\delta \cdot b_0$: create a new BS object u with center s ; set u 's side length and orientation parameters randomly following prior size distributions and add u to the current configuration ω .

2) *Death step:* Consider the current configuration ω and create a list of the $u \in \omega$ objects sorted from the highest to the lowest $A_D(u)$ values. For each object u taken in this order, compute the cost of deleting u from ω w.r.t. the global configuration energy:

$$\Delta\Phi_\omega^u = \Phi(\omega/\{u\}) - \Phi(\omega) = -A_D(u) - \sum_{\substack{v \in \omega \\ v \sim u}} I(u, v)$$

Next, we derive the $p_d(u)$ death rate as follows:

$$p_d(u) = f(\Delta\Phi_\omega^u) = \frac{\delta \cdot \exp(-\beta \cdot \Delta\Phi_\omega^u)}{1 + \delta \cdot \exp(-\beta \cdot \Delta\Phi_\omega^u)} \quad (2)$$

Finally, we remove u from ω with a probability $p_d(u)$.

4) *Convergence test:* if the process has not converged yet, increase the inverse temperature β and decrease the discretization step δ with a geometric scheme, and go back

to the *Birth step*. The convergence is obtained when all the BS objects added during the birth step, and only these ones, have been killed during the death step.

3.3 Discussion of the MPP detector results

The above detailed MPP detector results in a set of oriented rectangles which estimate the building footprints. As seen in Fig. 4, most building segments are correctly detected, however some roofs are only partially covered by the extracted masks, and connected roof parts may break up into many pieces (see the building in the bottom left corner). Moreover, since the roofs in the scene have various and irregular shapes, the proposed rectangular outline approximation proves to be often coarse. For this reason, we continue with a second step in the next section, which is responsible for the extraction of more faithful contours.

4 Refinement of the MPP Detection

In the first step, a feature (or saliency) point extraction technique is applied, which is based on the Harris corner detector [5] and helps to find saliency points in the surroundings of the estimated building locations.

Secondly, the generated saliency points and the edge information of the image are combined in a graph based representation to get the initial contour estimation of the buildings.

Finally, the Gradient Vector Flow (GVF) [13] active contour method detects the accurate object boundaries.

4.1 Feature point extraction

The Harris corner detector was introduced in [5]. The algorithm based on the principle that at corner points intensity values change largely in multiple directions. By considering a local window in the image and determining the average changes of image intensity result from shifting the window by a small amount in various directions, all the shifts will result in large change in case of a corner point. Thus corner can be detected by finding when the minimum change produced by any of shifts is large.

The method first computes the Harris matrix (M) for each pixel in the image:

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix}, \quad (3)$$

where $A = \dot{x}^2 \otimes w$, $B = \dot{y}^2 \otimes w$, $C = \dot{x}\dot{y} \otimes w$. \dot{x} and \dot{y} denote the approximation of the first order derivatives, w is a Gaussian window [5].

Then, instead of computing the eigenvalues of M , an R corner response is defined:

$$R = \text{Det}(M) - k * \text{Tr}^2(M), \quad (4)$$

where Det and Tr denote the determinant and trace and k is a coefficient (usually around 0.04).

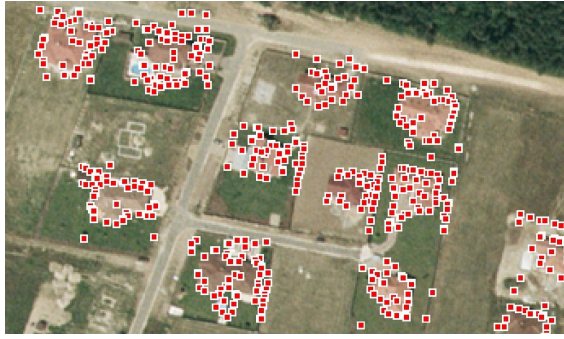


Figure 5. Extracted feature points around the object locations estimated by the MPP based detector

This R characteristic function is used to detect corners. R is large and positive in corner regions, and negative in edge regions. By searching for local maxima of R , the Harris key points can be found.

In our case, when searching for object contours, we need to extract both corner and edge points. Hence, the traditional R characteristic function need to be modified. According to [7] the following modification of R can emphasize corner and edge points equally:

$$R_{\log\max} = \max(0, \log[\max(\lambda_1, \lambda_2)]), \quad (5)$$

where λ_1 and λ_2 are the eigenvalues of the M Harris matrix (Eq. 3).

The eigenvalues define separate cases: both of them are large in corner regions, only one of them is large in edge regions and both of them are small in flat regions [5]. When emphasizing corners and edges, they both have one large component, thus $\max(\lambda_1, \lambda_2)$ function separates the flat and non-flat regions accurately. To produce a steady feature map, the dynamics of the characteristic function should be compressed into a balanced distribution by keeping the necessary strength of the main attractors. The natural logarithmic (\log) mapping function satisfies this condition: it has a balanced output for both corner and edge saliency. The target set of the $R_{\log\max}$ is the positive domain, thus the outer max function is responsible for replacing negative values of small λ s (flat points) with zeros. Saliency points are chosen as the local maxima of $R_{\log\max}$.

When searching for object outlines, only saliency points around the estimated building locations are extracted. The size of the interest area is determined by the resolution of the image. Fig. 5 shows the result of feature point extraction.

4.2 Graph based contour localization

Now a set of saliency points is given, denoting possible objects, which serves as the basis for building contour de-

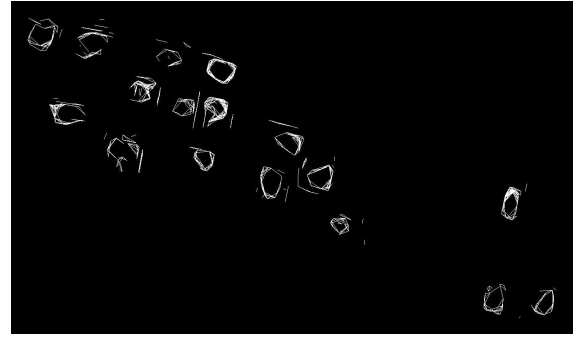


Figure 6. Subgraphs given after matching procedure

tection. We redefine the problem in terms of graph theory [12].

A graph G is represented as $G = (V, E)$, where V is the vertex set, E is the edge network. In our case, V is already defined by the set of feature points. Therefore, E needs to be formed. Information about how to link the vertices can be gained from edge maps. These maps can help us to only match vertices belonging to the same building.

If objects have sharp edges, we need such image modulations, which emphasize these edges as strong as it is possible. By referring back to Fig. 2(a) and 2(b), we can see that u^* component of $L^*u^*v^*$ color space can intensify *red* color component and R component of RGB is suitable for emphasizing *gray* color.

By generating the R and u^* components of the image, Canny edge detection [3] with large threshold ($Thr = 0.4$) is executed on them. C_r and C_u marks the result of Canny detection.

The process of matching is as follows. Given two vertices: $v_i = (x_i, y_i)$ and $v_j = (x_j, y_j)$. We match them if they satisfy the following conditions:

- (1.) $d(v_i, v_j) = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} < \epsilon_5$,
- (2.) $C_{...}(x_i, y_i) = \text{true}$,
- (3.) $C_{...}(x_j, y_j) = \text{true}$,
- (4.) \exists a finite path between v_i and v_j in $C_{...}$.

$C_{...}$ indicates either C_r or C_u . ϵ_5 is a tolerance value, which depends on the resolution and average size of the objects. We apply $\epsilon_5 = 20$.

These conditions guarantee that only vertices connected either in C_r or C_u edge maps are matched and closely located buildings are separated correctly. After this step we obtain a graph composed of many separate subgraphs, which can be seen in Fig. 6. The unmatched points indicate noise and are discarded. Only subgraphs surrounding the given mask centers are retained for contour detection step.

4.3 Object contour detection

To determine the exact contour of the buildings, we applied the classical Gradient Vector Flow (GVF) active contour



Figure 7. Results of the joint building localization and outline extraction approach. In the first column the original images can be seen, the second column shows the detected buildings.

method [13]. The evolution of the active contour is controlled by different energies. Internal energy is responsible for elasticity and rigidity of the contour, while external energy represents the constraints of the image. GVF snake defined a new external force as a diffusion of the gradient vectors of a gray-level or binary edge map derived from the image and proved to be able to detect concave boundaries as well.

To achieve the best performance, the initial contour of the GVF has to be defined closely to the object boundaries. For this reason, we calculated the convex hull of the generated subgraphs, which could be suitable as starting outline. Two cases were defined: if the convex hull surrounds a location estimated by the MPP detector, then it was used as initial contour; else the outline of the rectangle mask was used as initial contour.

After the GVF iterative process the detected contours

are checked. If the detection was successful, the obtained boundary should contain the center of the given rectangle mask. If this condition is not satisfied, then the detection has to be reconsidered. Depending on the initial contour, two cases were separated again: if the convex hull was used as initial contour, the GVF method is repeated with the rectangle mask used as initial contour and the resulted contour is checked again with the same procedure. If the outline of the rectangle mask was used as initial contour, then the unsatisfied condition shows that the active contour method could not refine the result of the Marked Point Process. In this case the given rectangle mask will be the detection result.

5 Experiments and Conclusion

Results of the joint building localization and outline extraction approach can be seen in Fig. 7 for different remotely sensed images.

The advantage of the MPP detector (see Sec. 3.1) is the accurate localization of buildings, but the rectangular contour estimation should be refined. Therefore a complex object extraction method has been introduced, which retrieves saliency points of the buildings, connects the coherent points by a graph based localization step, then detects the object contours with an active contour segmentation method. By comparing Fig. 4 and Fig. 7(b) the improvement of the joint approach can be observed. The various outline of buildings were detected more accurately and broken building segments were merged into one contour (building in the bottom left corner).

Acknowledgement

This work was supported by the Hungarian Scientific Research Fund under grant number 80352. The work of the second author was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. The authors would like to thank for the registered image pairs to Hungarian Institute of Geodesy, Cartography and Remote Sensing (FÖMI).

References

- [1] C. Benedek, X. Descombes, and J. Zerubia. Building detection in a single remotely sensed image with a point process of rectangles. In *Proc. ICPR*, August 2010.
- [2] Xavier Bresson, Selim Esedoglu, Pierre Vandergheynst, Jean-Philippe Thiran, and Stanley Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, 2007.
- [3] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.
- [4] X. Descombes, R. Minlos, and E. Zhizhina. Object extraction using a stochastic birth-and-death dynamics in continuum. *J. Mathematical Imaging and Vision*, 33:347–359, 2009.
- [5] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. of the 4th Alvey Vision Conf.*, pages 147–151, 1988.
- [6] K. Karantzas and N. Paragios. Recognition-driven two-dimensional competing priors toward automatic and accurate building detection. *IEEE Trans. GRS*, 47(1):133–144, 2009.
- [7] Andrea Kovacs and Tamas Sziranyi. High definition feature map for GVF snake by using Harris function. In *Advanced Concepts for Intelligent Vision Systems ACIVS 2010*, Sydney, Australia, 2010.
- [8] B. Li and T. Acton. Active contour external force using vector field convolution for image segmentation. *IEEE Trans. on Image Processing*, 16(8):2096–2106, 2007.
- [9] Akshaya K. Mishra and Alexander Wong. KPAC: A kernel-based parametric active contour method for fast image segmentation. *IEEE Signal Processing Letters*, 17(3):312–315, 2010.
- [10] S. Muller and D.W. Zaum. Robust building detection in aerial images. In *CMRT*, pages 143–148, Vienna, Austria, 2005.
- [11] B. Sirmacek and C. Unsalan. Building detection from aerial imagery using invariant color features and shadow information. In *ISCIS*, Istanbul, Turkey, 2008. [CD-ROM].
- [12] B. Sirmacek and C. Unsalan. Urban-area and building detection using SIFT keypoints and graph theory. *IEEE Trans. GRS*, 47(4):1156–1167, April 2009.
- [13] Chenyang Xu and Jerry L. Prince. Gradient vector flow: A new external force for snakes. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 66–71, 1997.