

# Building Detection from High-resolution Satellite Image Using Probability Model

LIU Wei      Veronique PRINET

Sino-French Laboratory in Information, Automation and Applied Mathematics

National Laboratory of Pattern Recognition

Institute of Automation, Chinese Academy of Sciences, Beijing, P.R. China

{liuw, prinet}@nlpr.ia.ac.cn

**Abstract**—This paper proposes a probability model and efficient scheme to extract buildings object from high-resolution panchromatic image in dense urban area. The overall idea of this work is to segment the image into regions, to treat all extracted regions' contours as candidates, and to make use of specific probability model to select 'true' buildings. Discriminative features that characterize buildings are proposed. Application is performed on Quickbird images over Beijing city.

**Keywords**- *High resolution satellite images; Building detection, Probability model, Contour, object recognition.*

## I. INTRODUCTION

This paper focuses on probabilistic modeling for building recognition in dense urban areas from high resolution images.

Often confined to the remote sensing (RS) community, building retrieval from high resolution images is a difficult object/pattern recognition problem. The difficulty comes at first, likewise other vision problem from the data acquisition: the image is a projection of the real 3D object, therefore introducing deformation and multi-faces view of the object; shadows appear more or less, depending on the sun position and high of the buildings; occlusions are frequent in very dense areas. RS images are also characterized by a very high level of noise: stochastic noise - i.e. thermal noise, as in all system acquisition but also deterministic noise - such as cars, white lines, i.e. all small features that disturb the recognition process. Moreover, a contrary to man-made object detection in natural scenes, the RS image is mostly composed of roads, buildings, etc, which have all similar types of primitives - straight lines and right angles; Therefore approaches relying on low level features detection would inevitably fail. Use of texture information is also biased since objects have seldom homogenous intensity distribution.

In the present paper, we propose an original approach for object - i.e. building-- extraction in dense urban areas. The core of the idea is to use probabilistic theory at its simplest level, assuming that the buildings probability distribution can be represented by a logistic function. Buildings are seen as independent objects in the image. Our main contributions are: i) to propose original features that characterize buildings; ii) to achieve recognition by computing a probability measure on each individual object.

The following of this paper is organized as follows: section II is a brief summary of traditional and recent works on recognition and object retrieval in external scenes; section III describes the proposed model and computational scheme; we give details about the features that are computed; section VI presents samples of the results and their analysis. A conclusion is finally given in section V.

## II. RELATED WORKS

Pattern/object recognition is one of the major 'unsolved problem' in image analysis and computer vision. Starting from Marr [10], recognition has been first considered as a bottom-to-top problem: in a hierarchical scheme, features are first detected, then constraints are progressively added, up to higher level recognition. One can consider this approach as an attempt to answer to the question: *what is where?* since it first considers the location, then the nature of the object. Due to high sensitivity to noise, it shows strong limitations and, in general, low robustness. At the opposite, top-to-bottom scheme defines at first what should be the object to search for, then tries to localise it in the image. The problem is therefore shifted to the question: *where is what?* (first defines what is the object, then attempt to know where it is). Indeed, the generic definition of a specific object is extremely difficult. Since the last ten years, the trend is to say: we cannot define arbitrarily the parameters of the model, but data themselves can do it. This idea has raised indeed the generation of learning-based and knowledge-based approaches. At the core of these, is probability and statistics.

Numerous works have been proposed for man-made objects extraction - building or road from aerial [11, 6, 5] or very high resolution satellite images [4]. Mayer gives in [7] a complete overview focused on buildings and proposes specific improvements. Worth to notice is the limitation of the use of one unique image: works are more and more oriented to the fusion of 2D and 3D digital surface model or Lidar data, which enables in particular to reduce occlusion and shadow problems [12, 3]. Little works have been done to study the potential of probabilistic modeling and derived approaches - maximum likelihood, markov random fields [9], discriminative random fields [1, 2], ... However intuitively, this type of approaches are the best suited to deal properly with the high level of noise - or *uncertainty* - of the RS image. One of the

difficulties is to translate the geometric shape that characterizes a building into *features* to be incorporated into the model. Widely used for natural scenes [1, 2], we study in this paper the potential of an approach derived from probability theory for building recognition in complex urban scenes of RS image.

### III. THE MODEL

A reliable probability measure relies on the appropriate choice of the functional and of the features. Our approach consists in computing this probability function over *regions* extracted from a segmentation step. Features are computed over these regions. Functional parameters are estimated from a large training data set.

#### A. The Functional

Similar to the association potential as given by Kumar [1, 2], we define our probabilistic distribution by

$$P_i = P(x_i | \mathbf{y}_i) = \frac{1}{1 + e^{\mathbf{x}_i \cdot (\boldsymbol{\omega}^T \mathbf{f}(\mathbf{y}_i))}} \quad (1)$$

Where  $P$  is used as a measure of confidence of building candidates and  $\mathbf{y}_i$  is the features vector at element  $i$ ,  $x_i$  is the label associated to the given element - with  $x_i=1$  for building and  $x_i=-1$  for non building.  $\boldsymbol{\omega}$  is the parameter vector of the model. The application function  $f$  transforms the feature vector  $\mathbf{y}$  such as:  $\mathfrak{R}^m \rightarrow \mathfrak{R}^d$  and  $\mathbf{y} \rightarrow [1, \mathbf{y}, \mathbf{y}^t \mathbf{y}, \tilde{\mathbf{y}}]$ .  $\tilde{\mathbf{y}}$  is defined as the product of each feature with another. Knowing  $m$  is the size of the feature vector, the dimension  $d$  is given by:  $1+m+m(m+1)/2$ .

#### B. Candidates Selection

We extract at first some regions of interest, that we refer as building candidates  $\mathbf{B} = \{b_i\}_{i \in S}$ , where  $S$  is the candidates space.  $\mathbf{B}$  is generated by first applying a sliding multi-window binary filter on the original image [8]. The window binary filter can be associated to a pass-band filter, where pixels having intensity level within the bandwidth are retained. This filter is sliding progressively from an arbitrarily predefined minimum value to an arbitrarily predefined maximum value, generating each time it is applied a new binary (i.e. black and white) image. In each of these images, aggregated pixels form compact regions, whose closed contours  $C_p$  are extracted (mathematical morphologic operators need to applied first to clean the data). All of these contours  $C=\{C_p\}$  localize potential candidate buildings. Smallest or largest contours are removed, overlapping contours are merged. Merging enables to generate only contours that can be considered as building candidates in  $\mathbf{B}$ . Indeed this step is critical. We merge the overlapped contours such as to maximize the probability measure  $p = P(x = 1)$ . We call this algorithm the ‘cut-and-merge’ approach.

#### C. Features Computation

In this section we focus on the way to compute the features vector  $\mathbf{y}$  from building candidates closed contours. Computation of the features vector requires to pre-compute various

information, in particular: edge points, edge line segments and shadow mask in the image. Shadow mask is obtained by finding the class of lowest intensity level of the contour region gray histogram. Edges points are extracted using Canny detector. The eight different features that are used are defined as follows:

- Distance to the Straight Lines

For each contour curve  $C_p$ , we accumulate the distance of every pixel to its nearest straight lines’ pixel. The first feature is the ratio sum of distance over the total length - i.e. total pixels number of the contour. This is motivated by the fact that building shape is polygonal and therefore constituted by joint line segments.

- Contour Region Entropy

For each contour curve  $C_p$ , we compute the entropy of the region inside it. The entropy is calculated based on the grey level histogram of this region. Roof building texture is relatively homogenous and therefore has higher entropy than non building contours.

- Contour Including Edges

Let’s say that the number of edges pixels (excluding the straight lines) included within the region delineated by the contour curve  $C_p$  is given by  $in\_ep$ . The third feature is the ratio of  $in\_ep$  over the inside area size - in term of pixel number. The motivation is that roofs are generally regions with smooth intensity level variations.

- Grey Level Average Value and Standard Deviation

Building contours regions often have lower standard deviation than non-building. If the average of the contour region is too low, the contour maybe a shadow area contour.

- Shape Features

We compute the gradient direction of each contour curve pixel and create the histogram over gradient direction [2]. Based on this histogram, the second moment is retrieved and saved as a feature. The peaks of the histogram are obtained simply by finding the local maxima of the histogram. Let  $\delta_1$  and  $\delta_2$  be the ordered gradient direction angle corresponding to the two highest peaks. Then the orientation based feature is computed as  $|\cos(2(\delta_1 - \delta_2))|$ . These two features are indicator of perpendicularity and parallelism of the straight lines that constitute the contour. We assume here that building borders may be preferentially perpendicular or parallel.

- Shadow Ratio

For the original image, it is very easy to know the orientation of the sun light. Translating the object contour  $C_i$  from  $n$  pixels (usually 3-5 pixels) according to the sun light orientation, one can create a new contour  $C_j$ . For  $C_i$  and  $C_j$ , their inside regions are noted  $\Lambda_i$  and  $\Lambda_j$  and we can make a new region  $\Lambda' = (\Lambda_i$

$\cup \Lambda j) - \Lambda i$ . The number of shadow pixels included within  $\Lambda'$  is given by  $in\_shd$ . The eighth feature is the ratio  $in\_shd$  over the area of  $\Lambda'$ . The motivation is that building contours are usually adjacent to shadow area.

#### A. Parameters Estimation

The unknown parameters  $\Theta$  are retrieved from training images. Parameters are computed by maximizing the sum of the log-likelihood of  $P_i$  calculated at each element  $i$ . Thus we have:

$$\Theta_{opt} = \arg \max_{\Theta} \sum_{n=1}^N \sum_{i \in S} \log(P_i) \quad (2)$$

where  $N$  is the total number of sub-images of the training set, and  $S$  is the space of the elements labeled as building areas  $x_i = 1$  or non-building areas  $x_i = -1$ . The optimization procedure is performed with a gradient descent algorithm.

We manually created the subset  $S$  from 4 large images -  $\dim(S)$  is close to 2650 with about 650 buildings and extracted 2000 non-building, for our experimentation. The computation time take about 3-4 minutes. Note that the computation cost comes from contour feature computation and the optimization per see only takes less than 10 seconds.

#### B. Inference

Basic probability laws enable us to write:

$$\sum_x P(x) = P(x=1) + P(x=-1) = 1 \quad (3)$$

Computing the feature vector  $\mathbf{y}$  at each building candidate and knowing  $\Theta$ , we can easily compute  $P$ . To retrieve true buildings and eliminate false candidates we should have:  $P(x=1) > P(x=-1)$ . Using equation (3), we then retain as buildings objects verifying:

$$P > 0.5. \quad (4)$$

## II. APPLICATION AND RESULTS ANALYSIS

Application and results are given on sub-images extracted from panchromatic Quickbird image, acquired in 2002. Image resolution is 0.6m/pixel.

The data have been processed and building retrieved applying equations (1) and (4). Post processing involves polygonal approximation of the recognized buildings using Douglas-Peucker line approximation method.

Table 1 summarizes simple result analysis in term of detection rate “DecR” (detection could be false or correct) and correct detection rate “CorR”. High value of detection rate and correct detection translate good results (see table 1 caption)

Note the complexity and variety of the shape and roof texture of the original images. The illustrations of the all set of contours (to right image) and the final candidate building contour after merging (image bottom left) show how much the cut-and-merge algorithm is powerful.

Bottom right figures, i.e. final retrieved buildings superimposed to the original images, show that some of the buildings can be retrieved partially, but their position or shape may not be correct. This may be due to features vector  $\mathbf{y}$  that may not be discriminative enough.

The figures of the global results however are quite significant.

TABLE I. RESULT ANALYSIS

Image	Ref	Correct	Error	Non	DecR	CorR
1	37	36	0	1	0.973	1.000
2	30	24	1	6	0.800	0.960
3	40	34	3	6	0.850	0.919
4	24	24	2	0	1.000	0.923
5	35	24	2	11	0.686	0.923
6	23	20	2	3	0.870	0.909
7	16	12	1	4	0.750	0.923
8	44	34	4	10	0.773	0.895
9	28	23	1	5	0.821	0.958
Total	277	231	16	46	0.834	0.935

**Ref:** number of buildings identified visually (ground truth); **Correct:** number of correctly detected buildings; **Error:** number of mistaken detected building; **Non:** number of undetected building; **DecR:** Correct / Ref; **CorR:** Correct / (Correct + Error)

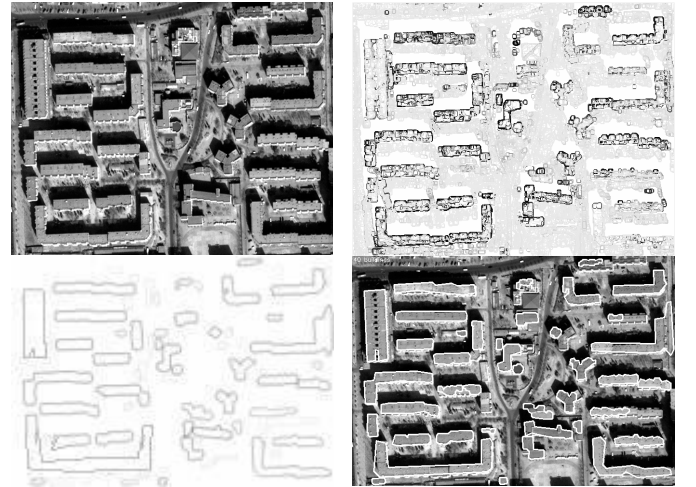


Figure 1. Illustration on Quickbird image on Beijing area (image No. 8 in the table 1), acquired in 2002. top left: original image; top right: contour extraction C from sliding multi-windows thresholding; bottom left: selected contour after merging; the gray level of each contour is proportional to its compute probability measure; bottom right: final building retrieval after polygonal approximation, overlapped to the original image

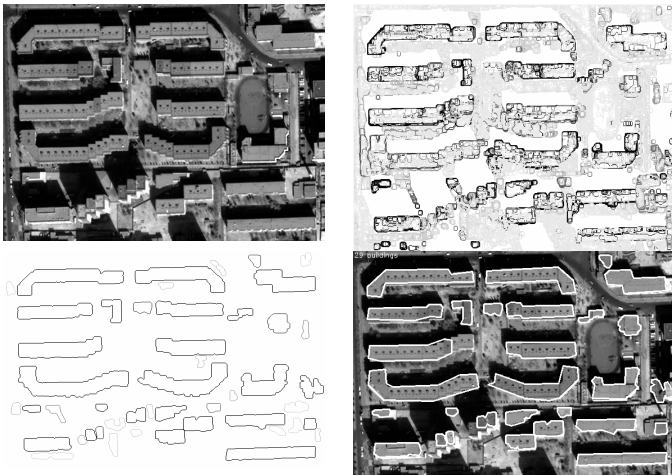


Figure 2. Illustration on Quickbird image on Beijing area (image No. 4 in the table 1), See figure 1 caption of images description

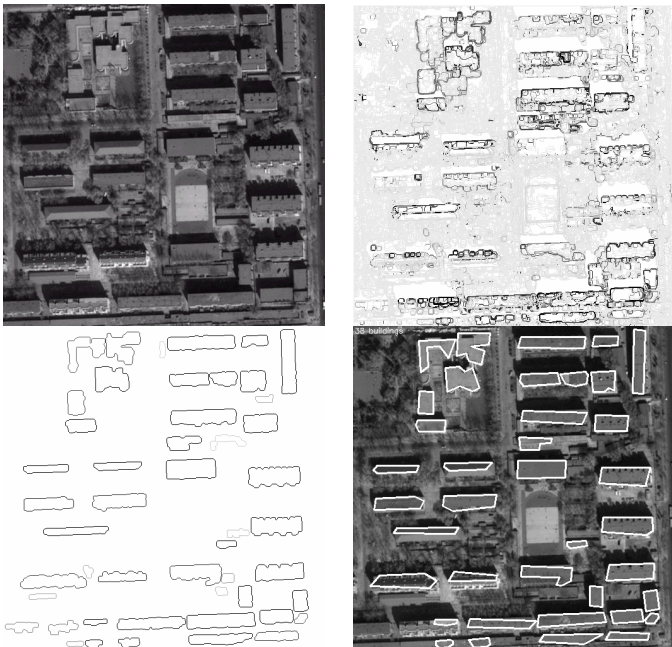


Figure 3. Illustration on Quickbird image on Beijing area (image No. 1 in the table 1), See figure 1 caption of images description

### III. CONCLUSION

We proposed in this paper an original approach for building inference based on basic probability modeling. The robustness of the approach is due in particular to the appropriate definition of probability distribution function and appropriate features

choice. However, the present approach still shows certain number of limitations. In particular, we can notice the insufficiency of the shadow model since it can not be generalized from one image to another – the orientation of the shadow with regard to the building is not an invariant. Robustness of the approach with regard to different imagery system or resolution will also be tested.

### REFERENCES

- [1] Sanjiv Kumar, Martial Hebert, "Discriminative Random Fields: A Discriminative Framework for Contextual Interaction in Classification", In Proc. IEEE Int'l Conf. on Computer Vision, 2003
- [2] Sanjiv Kumar, Martial Hebert, "Man-Made Structure Detection in Natural Images using a Causal Multiscale Random Field", In Proc. IEEE Int. Conf. on IEEE Conf. on Computer Vision and Pattern Recognition, 1:119-126, 2003
- [3] Hans-Gerd Maas, "Closed solutions for the determination of parametric building models from invariant moments of airborne laserscanner data", International Congress for Photogrammetry and Remote Sensing, vol. XXXII (B3), 1999, pp 130-137
- [4] Yanfeng Wei, Zhongming Zhao, and Jianghong Song, "Urban building extraction from high-resolution satellite panchromatic image using clustering and edge detection", Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International, Volume 3, 20-24 Sept. 2004 Page(s):2008 - 2010 vol.3
- [5] Markus Gereke, Bernd-Michael Straub & Andreas Koch, "Automatic Detection of Buildings and Trees from Aerial Imagery, Using Different Levels of Abstraction", Publications of the German Society for Photogrammetry and Remote Sensing, vol X,E. Seyfert, Ed., 2001, pp. 273-280
- [6] M. Gerke, C. Heipke and B. -M. Straub, "Building Extraction from Aerial Imagery Using a Generic Scene Model and Invariant Geometric Moments", Proc. IEEE/ISPRS Workshop on Remote Sensing and Data Fusion over Urban Areas, ISBN 0-7803-7059-7, pp.85-89.
- [7] Helmut Mayer, "Automation Object Extraction from Aerial Imagery-A Survey Focusing on Buildings", Computer vision and image understanding , col. 74, no.2, pp. 138-149, 1999
- [8] OpenCVCourses, CVPR01\_course (<http://sourceforge.net/projects/opencvlibrary>)
- [9] S Geman and D. Geman, Stochastic relaxation, Gibbs Distribution and Bayesian Restoration of Images, IEEE Transaction on Pattern Recognition and Machine Intelligence, vol.6,no.6, Nov. 1984
- [10] D. Marr , Vision, Ed.Freeman and Company, NY, 1982
- [11] Chungan Lin and Ramakant Nevatia, Building Detection and Description from a Single Intensity Image, Computer vision and image understanding, vol 72, no2, Nov.1998, pp101-121
- [12] N. Paparoditis, M. Cord, M. Jordan, and J.-P. Cocquerez, Building Detection and Reconstruction from Mid- and High-Resolution Aerial Imagery, Computer vision and image understanding, vol 72, no2, Nov.1998, pp122-142