AUTOMATIC RECTANGULAR BUILDING DETECTION FROM VHR AERIAL IMAGERY USING SHADOW AND IMAGE SEGMENTATION

Tran-Thanh Ngo, Christophe Collet, Vincent Mazet

ICube, University of Strasbourg, CNRS
300 Bd Sébastien Brant - CS 10413 - 67412 Illkirch, France
{ttngo, c.collet, vincent.mazet}@unistra.fr

ABSTRACT

This paper introduces a novel approach for the automated detection of rectangular buildings from monocular very high resolution (VHR) aerial images. The overall idea of this work is first to decompose the image into small homogeneous regions and treat all regions as candidates. According to the position of the shadows, a merging process is then performed over regions having similar spectral characteristics to produce building regions whose shapes are appropriate to rectangles. The experimental results prove that the proposed method is applicable in various areas (high dense urban, suburban, and rural) and is highly robust and reliable.

Index Terms— Building detection, image segmentation, Markov random field, very high resolution, remote sensing

1. INTRODUCTION

Automatic detection of buildings in VHR remotely sensed imagery is of great practical interest for a number of applications; including urban monitoring, change detection, estimation of population density, among others. Manual processing of images is time-consuming and expensive. Hence, developing a building detection approach that requires little or no human intervention has become one of the challenging problems in remote sensing applications.

There have been a significant amount of work on building detection from a single optical image in the literature [1–10], which are mainly based on the extraction of 2-D features, such as edge/line segments and/or corners. For the incorporation of 3-D information, shadows are valuable sources since a cast shadow is notably strong evidence of an existence of a building structure [1, 4]. For example, the authors in [4] utilize cast shadows to interpret the sides and corners of buildings. Shadows can also support directly the detection steps [5–8]. The authors [6, 7] proposed a probabilistic landscape approach to model the directional spatial relationship between buildings and their shadows. The building regions are detected by GrabCut partitioning approach over the landscape generated by shadow regions. The limitation of these approaches is that in dense urban areas, where a shadow region can be cast by a group of buildings, these adjoining buildings might be labeled as a single building.

Considering that most buildings in VHR images appear as rectangular shapes, some approaches have been proposed to deal with rectangular building detection in the literature [11–17]. Several rectangularity measures [18, 19] are designed to evaluate, on their specific way, how much an object differs from a perfect rectangle. The standard method for measuring rectangularity is to use the Minimum Bounding Rectangle (MBR) of the object. Most of these approaches have been proposed for LiDAR images [11–14], only two studies [15, 16] have exploited the rectangularity measures to detect rectangular buildings in optical images.

This paper deals with rectangular building detection in a variety of areas (high dense urban, suburban, and rural) with the assumption that buildings have homogeneous spectral features. The proposed method is based on the fact that from oblique aerial images, a 3D building structure should cast a shadow. Therefore, the methodology begins with the detection of shadows cast by building objects. In order to effectively extract building objects from image, we propose an original region-level Markov random field (MRF) image segmentation method. Buildings are extracted based on the segmentation result, their rectangularity and their location in respect to shadows. The main novelty of this paper is a new technique of utilizing shadows for identifying building regions and the combination of geometric and radiometric approaches to extract buildings.

The remaining paper is organized as follows. Section 2 deals with the detection of shadows cast by buildings. Section 3 covers our novel image segmentation method. Section 4 is devoted to the determination of final building regions. Next, results and discussion are given in Section 5 followed by Section 6 which contains our conclusions.

The authors would like to thank Rapid Mapping Service, SERTIT (http://sertit.u-strasbg.fr/) for providing data, the Institut Carnot (http://www.instituts-carnot.eu/) and the Alsace region for PhD funding.

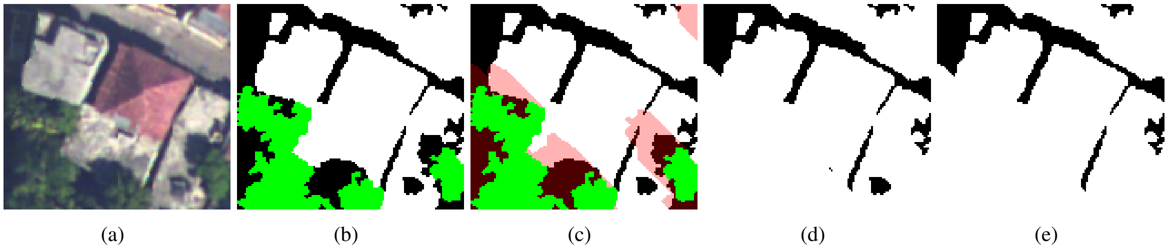


Fig. 1. Removing shadows due to non-building objects: (a) RGB aerial image (b) the detected shadow mask M_S (black) and vegetation mask M_V (green) using the method in [20] (c) the expansion regions (pink color) generated after the dilation of the vegetation objects overlaid with the shadow/vegetation map, (d) the shadow mask after eliminating shadows generated by vegetation, (e) final shadow mask M_{SB} (black) after the post-processing. In all experiments, the parameters l_{se} and d_{sh} are set to 60 pixels and 16 pixels respectively.

2. DETECTION OF BUILDING SHADOWS

Our methodology begins with the detection of shadows and vegetation. Shadows generated from vegetation or other non-building objects are then eliminated by a post-processing procedure. The remaining shadows are used to identify the possible locations of building objects.

2.1. Vegetation extraction and shadow detection

We identify vegetation and shadow regions to avoid the false alarms in building detection and to use shadows as an evidence to detect buildings. In this paper, we employ our recent shadow/vegetation detection approach [20], that allows to divide the image into three distinct classes: *shadow*, *vegetation*, and *others* with good precision (as shown in Fig. 1.(b)).

2.2. Post-processing of shadow mask

The illumination angle θ can be empirically estimated by identifying the illumination vector (e.g. a corner of a building and its estimated shadow point) and computing the angle from the north in a clockwise direction. To select shadows generated by distinct vegetation objects, we investigate, for each vegetation object of the vegetation mask M_V, the shadow evidence within the close neighborhoods of the vegetation object. To do that, binary morphological dilation is used, which allows expanding the shape of the vegetation object. The direction of the structuring element is determined by the illumination angle θ and its length l_{se} is empirically chosen. We then check for shadow evidence within this expansion region. If there is more than one shadow region occurring in the expansion region, we select the shadow region that have a border with vegetation object (as illustrated in Fig. 1). To eliminate shadows corresponding to relatively short objects, we found it necessary to compute the diameter of each shadow object and then filter out the objects whose diameter is below the predefined threshold d_{sh}. The remaining shadow mask is denoted as M_{SB} to distinguish from the original shadow mask M_S. The result of the post-processing is shown in Fig. 1.(e).

3. REGION-LEVEL MRF-BASED MULTIVARIATE IMAGE SEGMENTATION

3.1. Oversegmentation

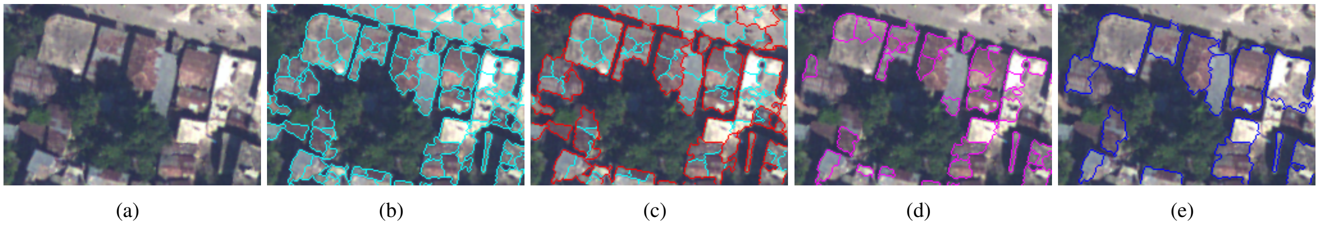
Oversegmentation is the first step performed to group spectrally similar pixels into small homogeneous regions (superpixels). In the proposed approach, oversegmentation algorithm SLIC [21] is performed in the image in which shadow regions M_S and vegetation regions M_V are masked out. The two parameters of SLIC are set as follows. The weighting factor m between color and spatial differences is set to 20, which can sufficiently preserve the boundaries of building objects and the number of superpixels is set so that the initial superpixel size η_{sup} is 200 pixels (see [21] for more details). As shown in Fig. 2.(b), oversegmentation generates regular-sized regions with good boundary adherence.

3.2. RAG and Image Segmentation Problem Statement

Starting from this set of regions (denoted by S), a region adjacency graph (RAG) is defined. Each region correspond to a node of the graph and the relationship between two regions is given by their adjacency, defining a set $\mathcal{E}$ of edges. The graph G is then $G = (S, \mathcal{E})$. For each node $i \in S$, R_i is the corresponding region of the image and x_i is a realization of the label X_i of region R_i. Also, let $\mathbf{X} = (X_i)_{i \in S}$ denote the joint random variable and the realization (configuration) $\mathbf{x} = (x_i)_{i \in S}$ of $\mathbf{X}$. Suppose image is to be segmented into K classes. Let $\mathcal{L} = \{l_1, \ldots, l_K\}$ denote the set of class labels. $\mathbf{x}$ is estimated using $\mathbf{y} = (\mathbf{y}_i)_{i \in S}$ where $\mathbf{y}_i$ is the observation of all pixels in region R_i, and therefore $\mathbf{y}_i = \{\mathbf{y}_i(s), s \in R_i\}$. For RGB images, $\mathbf{y}_i(s)$ is a 3-dimensional feature vector. $\mathbf{y}$ (resp. $\mathbf{y}_i$) is a realization of the observation field $\mathbf{Y}$ (resp. $\mathbf{Y}_i$).

3.3. Markovian regularization

Although MRF in image segmentation are mostly used on the pixel graph [22], they have also proved to be powerful models for feature-based graph (RAG [23, 24], line segment graph [2]). In MRF model, the search for $\mathbf{x}$ is defined to maximize the posterior probability $P(\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y})$, or to minimize



(a) (b) (c) (d) (e)

Fig. 2. Experimental results of different stages of the algorithm: (a) original image, (b) SLIC oversegmentation (regions are separated by cyan lines) of the image in which shadow regions M_S and vegetation regions M_V are masked out, (c) MRF-based image segmentation (clusters are separated by red lines), (d) detected building segments (boundaries are delineated by violet lines) based on shadow mask M_{SB}, (e) detected buildings (boundaries are delineated by blue lines). For all experiments, the parameters m, η_{sup}, β, $T_{bdShadow}$, R_{min} are set to 20, 200, 150, 10, 0.7 respectively.

the energy function:

$$U(\mathbf{x},\mathbf{y}) = U_1(\mathbf{y}|\mathbf{x}) + U_2(\mathbf{x}) \qquad (1)$$

Due to the independence assumption of the regions, the likelihood term can be written: $U_1(\mathbf{y}|\mathbf{x}) = \sum_{i\in S} U_i(\mathbf{y}_i|x_i)$. In this paper, Gaussian distribution is adopted to describe the image model. So, in cases where x_i takes the class label l_k:

$$U_i(\mathbf{y}_i|x_i) = \sum_{s\in R_i} \frac{1}{2}\times(\log(|\Sigma_k|)+[\mathbf{y}_i(s)-\boldsymbol{\mu}_k]^T\,\Sigma_k^{-1}\,[\mathbf{y}_i(s)-\boldsymbol{\mu}_k])$$

$\boldsymbol{\mu}_k$, Σ_k are mean and covariance of class l_k. For the prior term $U(\mathbf{x})$, we restrict our attention to MRF's whose clique potentials involve pairs of neighboring nodes ($\{i,j\} \in \mathcal{E}$). The prior term is defined as follows:

$$U_2(\mathbf{x}) = \sum_{i\in S}\sum_{j\in\mathcal{N}_i} n_i \times \frac{b_{ij}}{b_i} \times \frac{\beta}{|\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_j|} \times (1-\delta(x_i-x_j))$$

where $\delta(\cdot)$ stands for the Kronecker's delta function, $\mathcal{N}_i \subset S$ is the neighbors of the node i, n_i is the number of pixels in region R_i, b_i is the length of boundary of region R_i, b_{ij} is the length of common boundary of region R_i and region R_j. $\bar{\mathbf{y}}_i$ is the mean intensity of region R_i. Two constraints, the normalized edge weight b_{ij}/b_i and the inversed difference $|\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_j|^{-1}$ mean that if two regions share a long boundary and have similar mean intensity, they have high probability to obtain the same class label. β represents the tradeoff between fidelity to the observed image and the smoothness of the segmented image. The solution for Eq. (1) can be found by the ICM algorithm [25]. For the initialization, a Region-level K-Means algorithm [24] is used. The parameters of MRF model are estimated at each iteration of ICM algorithm as follows: $\boldsymbol{\mu}_k = \dfrac{\sum_{i\in\Omega_k}\sum_{s\in R_i}\mathbf{y}_i(s)}{\sum_{i\in\Omega_k}\sum_{s\in R_i}1}$, $\Sigma_k = \dfrac{\sum_{i\in\Omega_k}\sum_{s\in R_i}(\mathbf{y}_i(s)-\boldsymbol{\mu}_k)(\mathbf{y}_i(s)-\boldsymbol{\mu}_k)^T}{\sum_{i\in\Omega_k}\sum_{s\in R_i}1}$ where Ω_k denotes the set of nodes whose class label is l_k. The parameter β is empirically chosen. After the segmentation, connected regions that have similar spectral characteristics are grouped into cluster (a cluster is a group of neighboring regions having the same class label). An example of the segmentation result is shown in Fig. 2.(c) (clusters are separated by red lines).

4. DETERMINATION OF BUILDING REGIONS

4.1. Determination of Building Segments

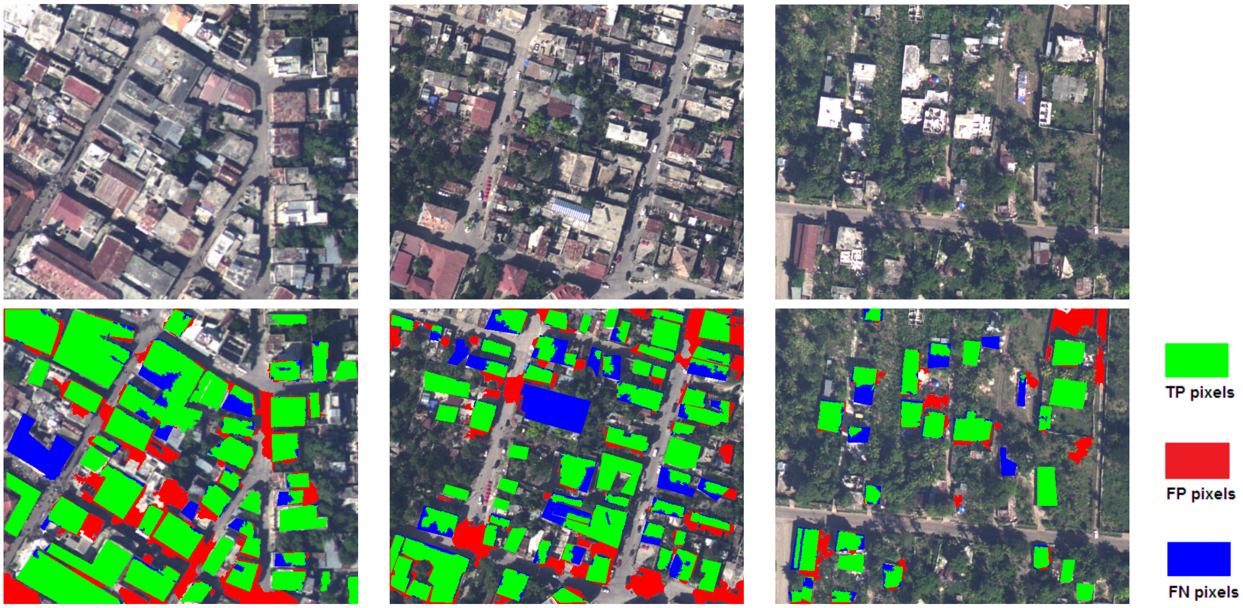
After the oversegmentation, image is decomposed into various small regions. The goal now is to determine what regions are belong to a "building". For simplicity, this type of regions is called as building segment. For each shadow object of shadow mask M_{SB} , the regions bordering the shadow object in the opposite direction of the illumination angle will be identified as building segments. Since region that shares a larger border with shadows is more likely to be a building segment, only regions whose border with shadows is larger than a predefined threshold ($T_{bdShadow}$) is flagged as a building segment (as shown in Fig. 2.(d)).

4.2. Determination of Final Buildings Regions

One approach to describe the rectangularity of an object, which relates the area of a segment and its bounding box, was proposed in [18]. In this paper, we use R_D [18, section 2.4] as the R-score (rectangularity measure). Within each cluster, we check all possible combinations of regions to produce rectangular buildings. A building is a combination of connected regions that satisfy two conditions. The first one is that it contains at least one building segment. The second one is that its R-score is superior to the preset threshold R_{min}. The final building is the biggest among all possible buildings satisfying these two conditions (as shown in Fig. 2.(e)).

5. EXPERIMENTS

The proposed method is tested on NOAA aerial images of 24 cm resolution. The reference data consisting of building regions were manually produced by a qualified human operator. The final performance of the proposed approach is assessed by comparing the results of the proposed approach with the reference data. Experimental results are shown in Fig. 3 and



(a) Dense urban area (b) Suburban area (c) Rural area

Fig. 3. Result of the proposed building detection approach in dense urban area (a), suburban area (b), rural area (c). Dense urban area is characterized by high population density and very attached buildings. Suburban area is characterized by residential strip with detached and semi-attached buildings to accommodate families. Rural area is characterized by lower population density and detached buildings. A visual inspection of the results give the strong impression that the developed approach is highly robust and that most of the buildings are successfully recovered, without producing too many FP pixels.

Test image	Pixel-Based performance (%)			Object-Based performance (%)		
	Precision	Recall	F_1 score	Precision	Recall	F_1 score
Urban area	85	94.5	89.5	85.3	88	86.6
Suburban are	92	87.4	89.6	90.3	82	85.9
Rural area	86	88.1	87.0	89.0	84	86.4

Table 1. Numerical results of the proposed building detection approach. To evaluate the performance, both pixel and object-based measures are considered. For pixel-based evaluation, comparing with the referencing data, true positive (TP), false positive (FP), false negative (FN) pixels, which represent respectively the correct detections, false alarms, missed building pixels, were counted. For object-based evaluation, like [6, 7], we label an output building object as TP if it has at least a 60% pixel overlap ratio with a building object in the reference data. We label an output object as FP if the output object of the proposed approach does not coincide with any of the building objects in the reference data, and we label an output object as FN if the output object corresponds to a reference object with a limited amount of overlap ($< 60\%$). Thus, it is possible to count TP, FP, FN for object-based evaluation. Using these counts, recall was calculated as TP/(TP + FN), precision as TP/(TP+FP), and F_1 [26] as (2 × precision × recall)/(precision + recall).

Table 1. In terms of an pixel-based point-of-view, the accuracy of our proposed method is high (the precision and recall ratios range from 85% to 95%). As far as object-based evaluation is concerned, we can conclude that most of the detected buildings are nearly complete and the results are fairly acceptable. Especially, in urban areas where a shadow region can be cast by a group of attached buildings, unlike the approaches in [6, 7], our proposed approach has the ability to separate these buildings because they are segmented into different classes. In reality, we separate these buildings based on their visual appearance.

The proposed approach cannot detect building regions whose shadow is not visible or missing. Since this method focus only on rectangular buildings, L-shaped or U-shaped buildings are partitioned into multiple rectangles. For buildings with arbitrary shapes, this method detects only the rectangular part of building that borders shadow regions.

6. CONCLUSIONS

An efficient approach is proposed for automatic rectangular building detection from monocular aerial images. Image is first decomposed into small homogeneous regions. Regions are then grouped into clusters by a region-level MRF segmentation method. Regions bordering shadows in the opposite direction of the illumination angle are flagged as building segments. A merging process is performed to merge these building segments with their neighboring regions in the same cluster to produce final building regions whose shapes are appropriate to rectangles. The experiments show that the proposed method is able to detect buildings in a variety of areas (high dense urban, suburban, and rural) with high accuracy.

7. REFERENCES

[1] Chungan Lin and Ramakant Nevatia, “Building detection and description from a single intensity image,” *Computer vision and image understanding*, vol. 72, no. 2, pp. 101–121, 1998.

[2] Santhana Krishnamachari and Rama Chellappa, “Delineating buildings by grouping lines with MRFs.,” *IEEE Transactions on image processing: a publication of the IEEE Signal Processing Society*, vol. 5, no. 1, pp. 164–168, 1995.

[3] R Bruce Irvin and David McKeown, “Methods for exploiting the relationship between buildings and their shadows in aerial imagery,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1564–1575, 1989.

[4] Andres Huertas and Ramakant Nevatia, “Detecting buildings in aerial images,” *Computer Vision, Graphics, and Image Processing*, vol. 41, no. 2, pp. 131–152, 1988.

[5] Huseyin Gokhan Akcay and Selim Aksoy, “Building detection using directional spatial constraints,” in *2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. 2010, pp. 1932–1935, IEEE.

[6] Ali Ozgun Ok, “Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 86, pp. 21–40, 2013.

[7] Ali Ozgun Ok, Caglar Senaras, and Baris Yuksel, “Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1701–1717, 2013.

[8] Mehmet Dikmen and Ugur Halici, “A learning-based resegmentation method for extraction of buildings in satellite images,” *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 12, 2014.

[9] Masoud S Nosrati and Parvaneh Saeedi, “A novel approach for polygonal rooftop detection in satellite/aerial imageries,” in *16th IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 1709–1712.

[10] Stephen Levitt and Farzin Aghdasi, “Fuzzy representation and grouping in building detection,” in *Image Processing (ICIP), 2000 7th IEEE International Conference on*, 2000, vol. 3, pp. 324–327.

[11] Liora Sahar, Subrahmanyam Muthukumar, and Steven P French, “Using aerial imagery and GIS in automated building footprint extraction and shape recognition for earthquake risk assessment of urban inventories,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 9, pp. 3511–3520, 2010.

[12] Suyoung Seo, Jeongho Lee, and Yongil Kim, “Extraction of boundaries of rooftop fenced buildings from airborne laser scanning data using rectangle models,” *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 2, 2014.

[13] Hossein Arefi and Peter Reinartz, “Building reconstruction using dsm and orthorectified images,” *Remote Sensing*, vol. 5, no. 4, pp. 1681–1703, 2013.

[14] Eunju Kwak and Ayman Habib, “Automatic representation and reconstruction of DBM from LiDAR data using recursive minimum bounding rectangle,” *ISPRS Journal of Photogrammetry and Remote Sensing*, 2013.

[15] T.S. Korting, L.M.G. Fonseca, L.V. Dutra, and F.C. Da Silva, “Image re-segmentation - a new approach applied to urban imagery,” in *VISAPP 2008 - 3rd International Conference on Computer Vision Theory and Applications, Proceedings*, 2008, vol. 1, pp. 467–472.

[16] Thales Sehn Korting, Luciano Vieira Dutra, and Leila Maria Garcia Fonseca, “A resegmentation approach for detecting rectangular objects in high-resolution imagery,” *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 621–625, 2011.

[17] Jun Wang, Xiucheng Yang, Xuebin Qin, Xin Ye, and Qiming Qin, “An efficient approach for automatic rectangular building extraction from very high resolution optical satellite imagery,” *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 3, 2015.

[18] Paul L Rosin, “Measuring rectangularity,” *Machine Vision and Applications*, vol. 11, no. 4, pp. 191–196, 1999.

[19] Paul L Rosin, “Measuring shape: ellipticity, rectangularity, and triangularity,” *Machine Vision and Applications*, vol. 14, no. 3, pp. 172–184, 2003.

[20] Tran-Thanh Ngo, Christophe Collet, and Vincent Mazet, “MRF and Dempster-Shafer theory for simultaneous shadow/vegetation detection on high resolution aerial color images,” in *21th IEEE International Conference on Image Processing (ICIP)*, 2014.

[21] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[22] S Geman and D Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–41, 1984.

[23] Florence Tupin and Michel Roux, “Markov random field on region adjacency graph for the fusion of SAR and optical data in radargrammetric applications,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 8, pp. 1920–1928, 2005.

[24] AK Qin and David A Clausi, “Multivariate image segmentation using semantic region growing with adaptive edge penalty,” *IEEE Transactions on Image Processing*, vol. 19, no. 8, pp. 2157–2170, 2010.

[25] Julian Besag, “On the statistical analysis of dirty picture,” *Journal of the Royal Statistical Society*, vol. 48, no. 3, pp. 259–302, 1986.

[26] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, and others, *Modern information retrieval*, vol. 463, ACM press New York, 1999.