Shape-based Building Detection in Visible Band Images using Shadow Information

Tran-Thanh Ngo, Vincent Mazet , Christophe Collet

***Abstract*—This paper introduces a novel methodology for automated detection of buildings from single high resolution optical images with only visible red, green, and blue bands of data. In particular, we first investigate the shadow evidence to focus on building regions. Then, a novel Markov random field (MRF) - based region growing segmentation technique is proposed. Image is oversegmented into smaller homogeneous regions which can be used to replace the rigid structure of the pixel grid. An iterative classification-merging is then applied over this set of regions. At each iteration, regions are classified using a region-level MRF model, then, according to the position of shadows, regions having the same class are merged to produce new regions whose shapes are appropriate to rectangles. The final buildings are determined using a recursive minimum bounding rectangle. 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, region growing, Markov random field, rectangularity measure, shadow, remote sensing**

I. Introduction and Motivation

AUTOMATIC detection of buildings in high resolution remotely sensed imagery is of great practical interest for a number of applications; including urban monitoring, change detection, estimation of human population, among others. Hence, developing a building detection approach that requires little or no human intervention has become one of the most challenging and widely studied topics in remote sensing literature [1]–[4].

A. Related Work

There have been a significant amount of work on building detection in the literature. Extensive reviews can be found in [1], [2]. Since this paper is devoted to the automated detection of buildings from single image, we limit the literature survey and discuss only the most relevant work that involved in the proposed framework.

Manuscript received ...; revised ...; accepted .... Date of publication ...; date of current version ....

T.-T. Ngo is with the Commissariat à l'énergie Atomique/Institut de Recherche sur la Fusion Magnétique, 13108 Saint Paul-lez-Durance, France (e-mail: tran-thanh.ngo@cea.fr).

V. Mazet and C. Collet are with ICube laboratory, University of Strasbourg, CNRS, 300 Bd Sebastien Brant - CS 10413 - 67412 Illkirch, France (e-mail: vincent.mazet@unistra.fr; c.collet@unistra.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier ....

A common feature of buildings is that they cast shadows on the ground. In the literature, shadows are used to identify buildings in two ways. On the one hand, after a building detection step, shadows are used for building hypothesis verification and height estimation. Lin and Nevatia [3] detect buildings from oblique aerial images, and hypothesized rectangular buildings were verified both with shadow and wall evidences. Sirmacek and Unsulan [4] employ color invariant features and shadows in a feature- and area-based approach. If shadows are found, the regions in opposite side of shadows are selected as candidates. A rectangle fitting method is then used to align a rectangle with the Canny edges of the image.

On the other hand, shadows can support directly the detection steps [2], [5]–[10]. In the recent work of A. Ozgun Ok [2], shadows are detected, dilated along the opposite of light direction to obtain a region of interest (ROI) for each rooftop. Iterative grabcuts are run in each ROI to label pixels inside it as rooftops or non- rooftops. Manno-Kovacs and Ok [5] improve that work by integrating urban area information to substantially revise and process the initial shadow mask. They detach dark regions from cast shadows with the aid of the solar information using a multilabel graph partitioning strategy. Manno-Kovacs and Sziranyi [6] integrate shadows with color, edge features and the illumination information to localize building candidates. In [7], Femiani *et al.* run grabcut on image and implement a self-correcting scheme that identifies falsely labeled pixels by analyzing the contours of buildings. Li *et al.* [8] segment image into homogeneous regions using the Gaussian mixture model (GMM) clustering method. Shadows and vegetation are then extracted from GMM labels. Remaining unlabeled regions are classified into probable rooftops and non-rooftops depending on shape, size, compactness and shadows. A higher order multilabel conditional random field segmentation is then performed to get final results.

In order to effectively retrieve objects from images, many methodologies partition image into smaller regions to enable region-based rather than global extraction. Image segmentation become a subsequent step in many object extraction algorithm, especially in building detection [6]–[8]. Segmentation of remotely sensed images is a difficult problem due to mixed pixels, spectral similarity, the textured appearance of land-cover types. Among many segmentation techniques, the region growing method is widely used. In [12], image is tessellated into a set of primitive regions, to build a region adjacency graph (RAG), which then undergoes vertex labeling and merging by alternating segmentation and region growing procedures. Regions are merged with the aim of minimizing the energy of MRF model [13] defined over RAG. The MRF

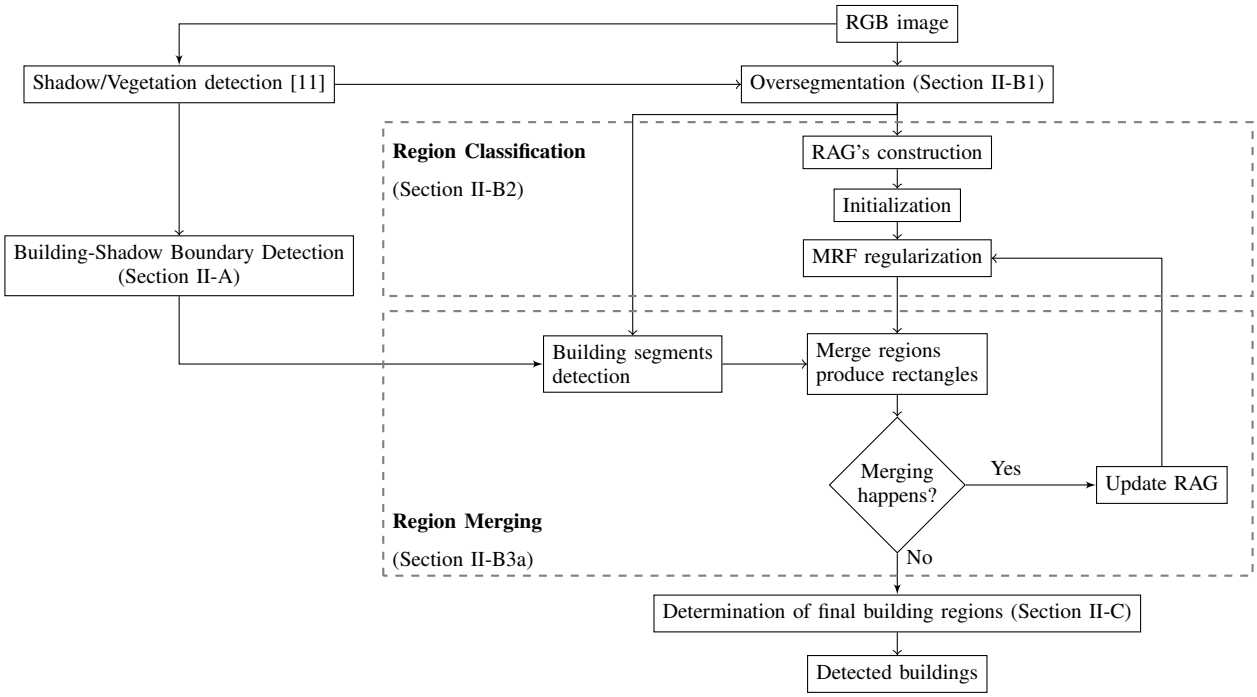


Fig. 1: Flowchart of the proposed method and related sections.

has been popular in modeling image spatial context and provides a statistically sound formulation of the segmentation problem.

Because of different materials of building rooftops, it is not easy to distinguish buildings from background using only spectral-based classification. Shape analysis can be explored for extracting rooftops. The most common shape of building is a rectangle or a combination of rectangles comprised of right-angled corners because this building structure needs less engineering effort in design, material and construction to achieve an acceptable level of seismic performance [14]. In [15], a rectangle building detection method is proposed by constructing a hierarchical framework to create various building appearance models from different elementary feature-based modules. Another approach to deal with rectangular buildings is to use rectangular boundary fitting. Several rectangularity measures are designed to evaluate how much the considered object differs from a perfect rectangle [16]–[18]. The standard method is the Minimum Bounding Rectangle (MBR) [16], [19], [20]. To our knowledge, most of these approaches have been proposed for LiDAR images [21], [22], only two studies [23], [24] have exploited the rectangularity measures to detect rectangular buildings in optical images.

B. Proposed Method and Contributions

This paper introduces a novel automatic building detection method for single RGB images. Multiple views [15], additional information such as near-infrared (NIR) [2], LiDAR [21] or any elevation data are not necessary. The method must be applicable in various areas: rural area with detached buildings and high vegetation density, suburban area with detached or semi-attached buildings and high dense urban area with high population density and attached buildings. This obliges us to follow certain assumptions about the appearance of buildings:

1) We consider that buildings have a homogeneous color. Roof homogeneity have been exploited for building detection [15], [25]. In high dense area, the spectral features are exploited to separate the attached buildings;
2) A building casts a shadow under suitable imaging conditions. Shadows must be detected beforehand. Besides, correctness and precision of the shadow detection are strongly required;
3) In this study, we focus on the buildings with right-angled corners, such as: rectangular, L-shaped, U-shaped, T-shaped buildings (Fig. 2).

(a) rectangular (b) L-shaped (c) U-shaped (d) T-shaped

Fig. 2: Considered types of building rooftops.

Starting from these assumptions, a novel method for building detection is designed (Fig. 1). Our shadow/vegetation detection method [11] is first applied to divide image into three distinct classes: *shadow*, *vegetation*, and *others*. The boundaries between shadows and their casting buildings are detected by eliminating shadows generated by vegetation objects and other non-building objects. Then, image in which shadows and vegetation are masked out is oversegmented into smaller regions. A novel MRF region growing technique is proposed, in which the radiometric and geometric information of building are

exploited. Regions are classified by a MRF-based region-level model and grouped into clusters. A merging process is performed within each cluster to merge building segments (the regions sharing the boundaries with shadows) with their neighboring regions. The goal is to produce new regions whose shapes are appropriate to rectangles. This iterative procedure continues until there is no merging happens. Finally, a recursive Minimum Bounding Rectangle (RMBR) [22] is used to determine final buildings.

The remainder of the paper is organized as follows. We introduce our building detection approach in Section II. The experiments are presented in Section III. Finally, in Section IV, we give concluding remarks and make suggestions for possible future works.

II. Methodology

A. Building-Shadow Boundary Detection

Our shadow/vegetation detection method [11] is based on Otsu's thresholding method and Dempster-Shafer (DS) fusion which aims at combining different shadow indices and vegetation indices in order to increase the information quality and to obtain a more reliable and accurate segmentation result. The DS fusion is carried out pixel by pixel and is incorporated in the Markovian context while obtaining the optimal segmentation with the energy minimization scheme associated with the MRF. Since shadows may be cast by buildings or nearby objects such as rock, vehicle, vegetation (Fig. 3(a)), it is essential to eliminate shadows that occur due to non-building objects. In reality, the detection of shadows cast by buildings is rather impossible since we have no knowledge about the height of buildings and the solar zenith angle. Therefore, as will be discussed in Subsection II-B3a, our goal is to detect the boundaries between buildings and their corresponding shadows. The illumination angle θ can be empirically estimated by counting the number of pixels horizontally and vertically from one corner of a rooftop to the corresponding corner of a building shadow. To select shadows generated by vegetation, we investigate the shadow evidence within the close neighborhoods of each vegetation object using a binary morphological dilation (Fig. 3(b)). The direction of the structuring element is determined by θ and its length l_{se} is empirically chosen. If there is more than one shadow occurring in this expansion region, we select the shadow having a border with vegetation objects (Fig. 3(c)).

Boundaries between shadows and their casting buildings are detected as follows. The opposite direction of the illumination is quantified into one of eight directions (north, northeast, east, southeast, south, southwest, west, and northwest). The contour of shadows are detected. For each pixel on the contour, if its pixel on the southeast (in our case) is shadow, it will be removed from the contour. The final building-shadow boundary mask M_{SB} are obtained by filtering out these contours whose length is below the predefined threshold d_{sh} (Fig. 3(d)).

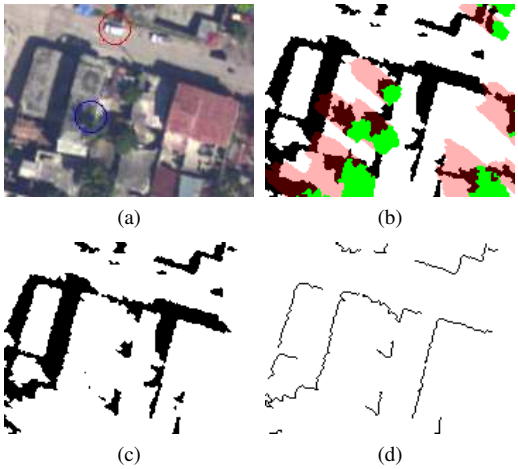


Fig. 3: Building-Shadow boundary detection. (a) RGB image. Shadow is cast by vegetation (blue circle) or other small non-building objects such as vehicle (red circle). (b) Shadow M_S (black) and vegetation M_V (green). The expansion regions (pink color) are generated by dilating vegetation objects. (c) Shadow mask after eliminating shadows generated by vegetation. (d) Building-shadow boundary mask M_{SB}.

B. Region Growing Image Segmentation

The MRF-based region growing starts with an oversegmentation (Fig. 4(b)). An iterative classification-merging is then applied. Regions are grouped into different clusters by the MRF-based region classification (Fig. 4(c)). Boundary mask M_{SB} is used to determine the building segments (the regions bordering shadows in the opposite direction of illumination angle, Fig. 4(e)) and a merging process is performed to merge building segments with their neighboring regions (Fig. 4(f)). The algorithm is detailed as follows.

1) Oversegmentation: There exists several oversegmentation algorithms, such as TurboPixels [26] and SLIC [27]. In this paper, the SLIC algorithm [27] is employed in the image in which shadow regions M_S and vegetation regions M_V are masked out. As shown in Fig. 4(b), oversegmentation generates regular-sized regions R_i ($i = \{1, \ldots, \mathrm{Q}\}$) with good boundary adherence, and fits well for region classification.

2) MRF-based Region Classification: Although MRF [13] is mostly used on the pixel graph [13], it is also proved to be a powerful model for feature-based graph (such as RAG [12], line segment graph [28]). In our approach, a RAG, $G = (S, E)$, is used, where S is the set of nodes in graph. Each node s_i corresponds to each region R_i. E is the set of edges with $(s_i, s_j) \in E$ if R_i and R_j are neighboring regions.

a) MRF's Framework: Suppose image is to be segmented into K classes. Let $\mathcal{L} = \{l_1, \ldots, l_K\}$ denote the set of class labels. Then, we want to find an assignment of all nodes s_i to $\mathcal{L}$. For each node $s_i \in S$, x_i is a realization of the label X_i of s_i. Also, let $\mathbf{X} = (X_i)_{s_i \in S}$ denote the joint random variable and the realization (configuration) $\mathbf{x} = (x_i)_{i \in S}$ of $\mathbf{X}$. $\mathbf{x}$ is estimated using $\mathbf{y} = (\mathbf{y}_i)_{s_i \in S}$ where $\mathbf{y}_i$ is the observation of all pixels in region R_i, or $\mathbf{y}_i = \{\mathbf{y}_i(s), s \in R_i\}$. 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. $\mathbf{Y}_i$). In MRF model, the optimal configuration $\hat{\mathbf{x}}$ will be a maximum

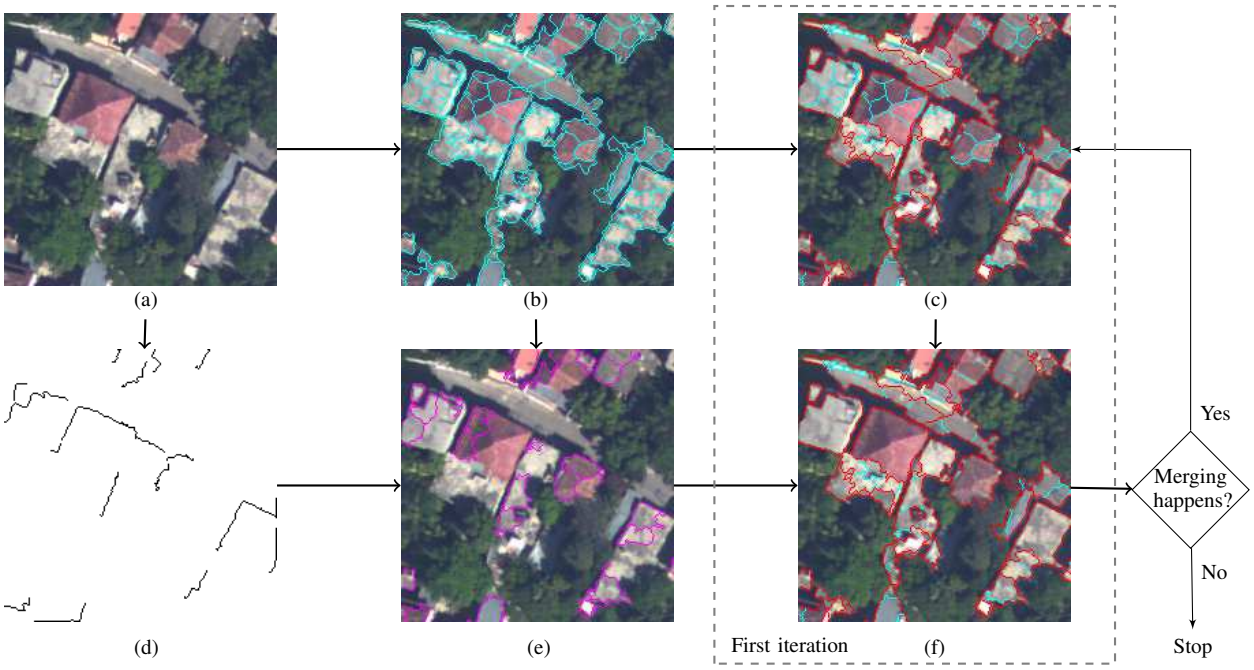


Fig. 4: An example of the first classification-merging process: (a) RGB image, (b) oversegmentation (regions are separated by cyan lines), (c) region classification (clusters are separated by red lines), (d) Building-Shadow boundary mask M_{SB}, (e) detected building segments, (f) after merging (some cyan lines are disappeared).

a *posteriori* probability (MAP) under observation $\mathbf{y}$:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} P(\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y}) \tag{1}$$

From Bayes'rule, we have:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \frac{P(\mathbf{Y} = \mathbf{y}|\mathbf{X} = \mathbf{x})P(\mathbf{X} = \mathbf{x})}{P(\mathbf{Y} = \mathbf{y})} \tag{2}$$

When the image is designed, $P(\mathbf{Y} = \mathbf{y})$ is constant. Maximizing the *a posteriori* probability leads to minimize the posterior energy function:

$$U(\mathbf{x}, \mathbf{y}) = U(\mathbf{y}|\mathbf{x}) + U(\mathbf{x}) \tag{3}$$

b) Region Classification: The first term in Eq. 3 is called as the *likelihood* term. Due to the independence assumption of the regions, the likelihood term can be written: $U(\mathbf{y}|\mathbf{x}) = \sum_{s_i \in S} U_i(\mathbf{y}_i|x_i)$. Each element $U_i(\mathbf{y}_i|x_i)$ describes the probability of region R_i with its observation $\mathbf{y}_i$ at the given region label x_i. As Gaussian distribution is a usual and effective distribution for color images, this distribution is adopted to describe the image model. So, in cases where x_i takes the class label l_k:

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

where $\boldsymbol{\mu}_k$, Σ_k are mean and standard deviation of class l_k, respectively.

The second term in Eq. 3 is the *prior* term, describing what the likely labelings $\mathbf{x}$ should be like. This knowledge can be introduced in the definition of the clique potential of the RAG. In order to reduce the computational complexity, we restrict our attention to MRF's whose clique potentials involve pairs of neighboring nodes ($\{s_i, s_j\} \in E$). The prior term is defined as follows:

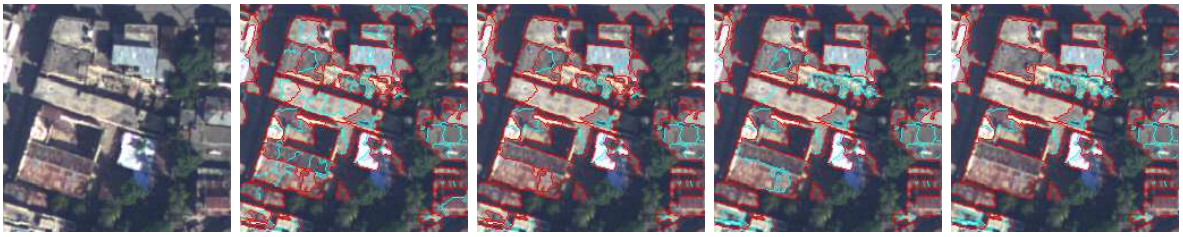
$$U(\mathbf{x}) = \sum_{i \in S} \sum_{j \in \mathcal{N}_i} n_i \times \frac{b_{ij}}{b_i} \times \frac{\beta}{|\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_j|} \times (1 - \delta(x_i - x_j))$$

where $\delta(\cdot)$ stands for the Kronecker's delta function, $\mathcal{N}_i \subset S$ is the neighbors of node s_i, n_i is the number of pixels in region R_i, b_i is the length of contour of R_i, b_{ij} is the length of common boundary of R_i and R_j, $\bar{\mathbf{y}}_i$ is the mean intensity of R_i. Two constraints, the normalized edge weight b_{ij}/b_i and the inverse 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. 3 can be found by the Iterated Conditional Mode (ICM) algorithm [13]. For the initialization, a Region-level K-Means algorithm [12] is used. The parameters of MRF model are estimated at each iteration of ICM algorithm as follows:

$$\boldsymbol{\mu}_k = \frac{\sum_{i \in \Omega_k} \sum_{s \in R_i} \mathbf{y}_i(s)}{\sum_{i \in \Omega_k} \sum_{s \in R_i} 1} \tag{4}$$

$$\Sigma_k = \frac{\sum_{i \in \Omega_k} \sum_{s \in R_i} (\mathbf{y}_i(s) - \boldsymbol{\mu}_k)(\mathbf{y}_i(s) - \boldsymbol{\mu}_k)^T}{\sum_{i \in \Omega_k} \sum_{s \in R_i} 1} \tag{5}$$

where Ω_k denotes the set of nodes whose class label is l_k. After the classification, connected regions having the same class label are grouped into cluster. An example of the classification result is shown in Fig. 4(c).



(a) Input image (b) First classification (c) First merging (d) Second segmentation (e) Second merging

Fig. 5: Iterative classification and merging: an example of the first two iterations.

3) Region Merging: This procedure is designed by using three assumptions about the target buildings. First, buildings cast shadows on the ground, so regions to be merged contain at least one region that is next to shadows in the opposite direction of the illumination angle. Second, since we focus only on buildings with right-angled corners, the merging procedure is designed to produce new regions whose shapes are appropriate to rectangles. Third, we assume that building has a homogeneous color, hence, merging is only done between regions having the same class label. Different steps of merging are designed in the following.

a) Determination of "building segment": This term refers to regions bordering building shadows in the opposite direction of the illumination (Fig. 4(e), building segments are delineated by violet lines). They are detected by measuring the boundary between each region and the building-shadow boundary M_{SB}. Since region that shares a larger border with shadows is more likely to be a building segment, only regions whose boundary with M_{SB} is larger than a predefined threshold T_S is flagged as a building segment.

Algorithm 1: Merging procedure **for each cluster**

Input:
- BdS: list of building segments that are not "visited".
- $\mathcal{L}p$: list of possible merging.
- $\mathcal{L}$: list of regions to be merged. For initialization, $\mathcal{L} \leftarrow \emptyset$.

while $\mathrm{BdS} \neq \emptyset$ **do**
- 1. $i^* \leftarrow \arg\max_i R_D[\mathcal{L}p(i)]$.
- $/*$ best of possible merging $*/$
- **if** $R_D[\mathcal{L}p(i^*)] \geq T_R$ **then**
  - 1. $\mathrm{Bd} = \mathcal{L}p(i^*) \cap \mathrm{BdS}$.
  - $/*$ building segments in $\mathcal{L}p(i^*)$ $*/$
  - 2. **if** $R_D[\mathcal{L}p(i^*)] \geq \max(R_D[\mathrm{Bd}])$ **then**
    - Add $\mathcal{L}p(i^*)$ to $\mathcal{L}$.
  - 3. Update: $\mathrm{BdS} = \mathrm{BdS} \setminus \mathrm{Bd}$.
- **else**
  - break ;

Merge regions in $\mathcal{L}$.

Output: new building segments

b) Region merging procedure: Several rectangularity measures have been proposed in the literature. The standard method is the MBR [16]. In [17], Rosin proposed three new rectangularity measures. Together with the MBR method, four methods have been tested on our building data and the discrepancy method R_D (defined in [17, section 2.4]) is concluded as the best (robust to dealing with the buildings with protrusion artifact and intrusion artifact). In this method, a rectangle is fitted to the region based on its moments. Rectangularity is measured as the normalized discrepancies between the areas of the rectangle and the region. In the next, R_D is denoted as the operator to measure the rectangularity degree of an object.

As described in Algorithm 1, since merging is done between regions having the same class label, we process each cluster independently. BdS denotes the list of building segments that are not "visited" (not merged with its neighbors). A possible merging is a group of connected regions that includes at least one building segment. $\mathcal{L}p$ denotes the list of possible merging. T_R is the predefined minimum rectangularity degree. The criteria of merging is to merge building segments with their neighboring regions while increasing the rectangularity degree of building segments. After merging, the building segments are updated for the next iteration. As shown in Fig. 4(f), some cyan lines that separate the regions are disappeared.

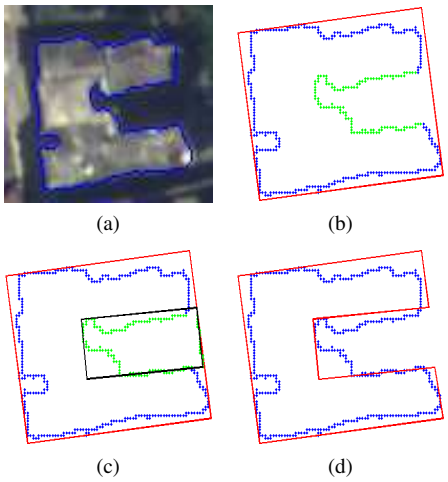
4) Iterative Classification and Merging: The main idea of the iterative classification-merging is to merge building segments with their neighboring regions while increasing their rectangularity degree R_D and avoiding merging between parts of different objects (e.g. two adjoining rectangle buildings). The building segments are expected to converge toward their full building object outlines.

After each iteration, the RAG is updated. The class label of nodes do not change since the merging is done in the limit of each cluster. The feature model class statistics ($\boldsymbol{\mu}$ and Σ) are therefore unchanged. The parameter β of prior term is kept constant. The ICM process continues with this new RAG to search for its suboptimal solution of MRF energy minimization. This iterative procedure ends when there is no merging happens. An example of the first two iterations is shown in Fig. 5.

C. Determination of Final Building Regions

The above procedure results in a segmentation map, in which the building segments have the best possible rectangular score. Because of the strict constraint of merging, a building segment whose shape is very close to a perfect rectangle can not be merged with its neighboring regions (the other parts of the same building). Moreover, other types of buildings

like L-shaped, U-shaped are partitioned into different parts. In this section, we describe how to determine the final building regions from the segmentation map.



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

Fig. 6: The step-by-step illustration of U-shaped bounding box estimation

1) Recursive Minimum Bounding Rectangle (RMBR): in a recent paper [22], the authors introduce a new framework for automatically detecting and reconstructing accurate right-angle corner building models from LiDAR data. In their approach, building boundaries are first approximated by LiDAR data and then regularized using rectangular model through a RMBR process. Given a boundary shape, we apply the RMBR algorithm in the way as follows:

1) Apply the MBR algorithm to the boundary shape and the generated MBR is denoted as $\mathrm{MBR}_{(1)}$. Track the non-overlapping boundary segments with $\mathrm{MBR}_{(1)}$ (Fig. 6(b)).
2) Apply MBR algorithm again on the non-overlapping segments and their projection onto the MBR sides (to derive $\mathrm{MBR}_{(2)}$ as illustrated in Fig. 6(c)).
3) Track the non-overlapping boundary segments with $\mathrm{MBR}_{(2)}$ (Fig. 6(d)).
4) Apply MBR algorithm again on the non-overlapping segments and their projections onto the MBR sides to derive $\mathrm{MBR}_{(3)}$.

In Fig. 6, the first-order MBR is delineated by the red rectangle. The area of this MBR outside the object is delineated by the green dot. Finally, the black rectangle is the first-order MBR of the object limited by the green dots. In our method, we limit the level of recursive MBR to 3. The final shape, denoted as RMBR, is determined as follows:

$$\mathrm{RMBR} = \mathrm{MBR}_{(1)} - \mathrm{MBR}_{(2)} + \mathrm{MBR}_{(3)} \tag{6}$$

Note that if the condition in step 3 (step 5) is not satisfied, $\mathrm{MBR}_{(2)}$ ($\mathrm{MBR}_{(3)}$) is an empty set.

2) Procedure of Determining Final Building Regions: For the sake of simplicity, we denote RMBRf as an operator to measure how much the considered object differs from its RMBR approximation:

$$\mathrm{RMBRf} = \frac{\text{Area of object}}{\text{Area of RMBR}} \tag{7}$$

Algorithm 2: Determine the final building in a cluster

Input:
- T_B: minimum RMBR-score.
- BdS: list of building segments in cluster.
- N: number of regions in cluster.

for $k \leftarrow N$ **to** $|\mathrm{BdS}|$ **do**
1. Determine $\mathcal{L}_k$: list of candidate buildings merged from k regions. $\mathcal{L}_k(j)$ is the j-th element in $\mathcal{L}_k$.
2. $j^* \leftarrow \arg\max_j \mathrm{RMBRf}[\mathcal{L}_k(j)]$.

$/*$ best of candidate building $*/$

if $\mathrm{RMBRf}[\mathcal{L}_k(j^*)] \geq T_B$ **then**
1. Merge regions in $\mathcal{L}_k(j^*)$ and get the final building region of cluster.
2. Break.

Output: final building region

(a) Segmentation result (b) Final building

Fig. 7: Determination of final buildings: 3 buildings (left) are approximated by the recursive MBR of level 2 and other 2 buildings (right) are rectangular. Their RMBR scores are respectively 0.87, 0.91, 0.84, 0.93, 0.81 (from left to right). The value of T_B is chosen as 0.8.

Similar to Algorithm 1, the determination of final buildings is done within each cluster. The main idea is to check the possibility of having a building with right-angled corners. For each cluster, we denote a candidate building as a group of connected regions that includes all building segments. Then, we verify if the candidate building has right-angled corners by measuring its RMBR score. The procedure of determining final building regions is described in Algorithm 2. An example is shown in Fig. 7.

III. Experiments

A. Image Data Set

The experiments are performed on our data set and the building detection benchmark proposed by Ok and Seneras [29]. Our data set consists of NOAA (National Oceanic and Atmospheric Administration) aerial images (24 cm) and BD ORTHO® images (50 cm), provided by the Rapid Mapping Service SERTIT (http://sertit.u-strasbg.fr/). All images convey 3 spectral bands (R, G, B) with a radiometric resolution

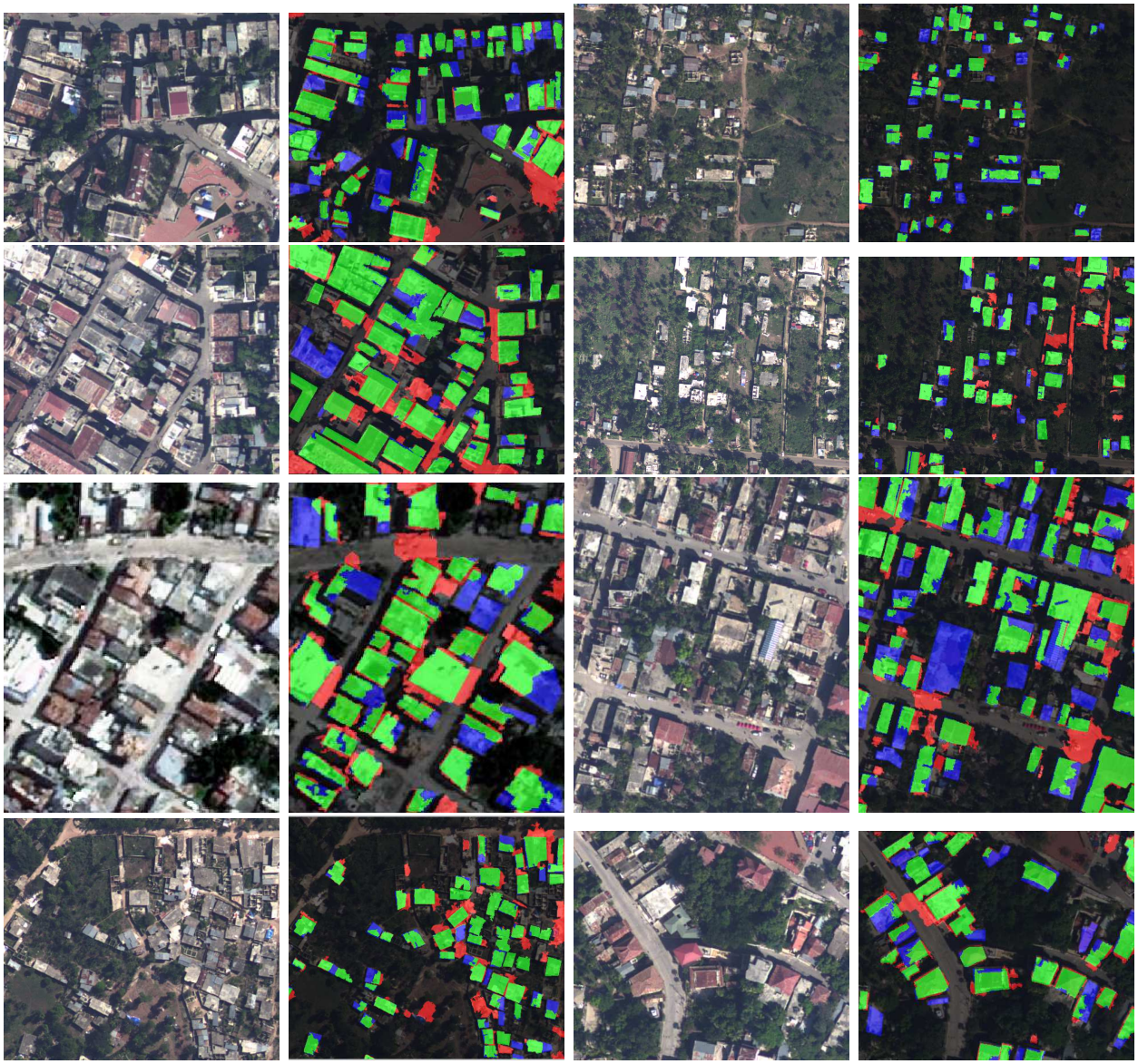


Fig. 8: The results of the proposed approach on NOAA images: (First column) Test images #1-4. (Second column) Detected buildings for test images #1-4. (Third column) Test images #5-8. (Fourth column) Detected buildings for test images #5-8.

of 8 bits per band (Fig. 8). Those images are chosen to cover different areas (high dense urban (#1-3), suburban (#7-14), rural (#4-6)), varying illumination and acquisition conditions. The benchmark of Ok and Seneras [29] consists of 14 originally orthorectified and pansharpened images selected from two different VHR satellite sensors, i.e., IKONOS-2 (1 m) and QuickBird (0.60 m). All images convey 4 bands (R, G, B, and NIR) with a radiometric resolution of 11 bits per band.

B. Assessment Strategy

The performance is assessed by comparing the results with the reference data that consist of buildings manually produced by a qualified human operator (SERTIT). Both pixel- and object-based measures are considered. For pixel-based evaluation, the common measures of precision (P), recall (R), and the F-score (F_1) [30] are used:

$$\mathrm{P} = \frac{\|\mathrm{TP}\|}{\|\mathrm{TP}\| + \|\mathrm{FP}\|}; \mathrm{R} = \frac{\|\mathrm{TP}\|}{\|\mathrm{TP}\| + \|\mathrm{FN}\|}; F_1 = \frac{2 \times \mathrm{P} \times \mathrm{R}}{\mathrm{P} + \mathrm{R}} \tag{8}$$

where TP (true positive) denotes a building pixel correctly identified, FN (false negative) indicates a building pixel identified as non-building, FP (false positive) denotes a non-building pixel identified as building. $\|.\|$ denotes the number of pixels assigned to each distinct category. The F-score F_1 captures both precision and recall into a single metric that gives each an equal importance. The object-based performance can also be evaluated in the similar way [30]. We classify a resulted building object as TP if it has at least 60% pixel overlap ratio with a building object in the reference data. Whereas, we classify a resulted object as FP if it does not coincide with any of the building objects in the reference data. In addition,

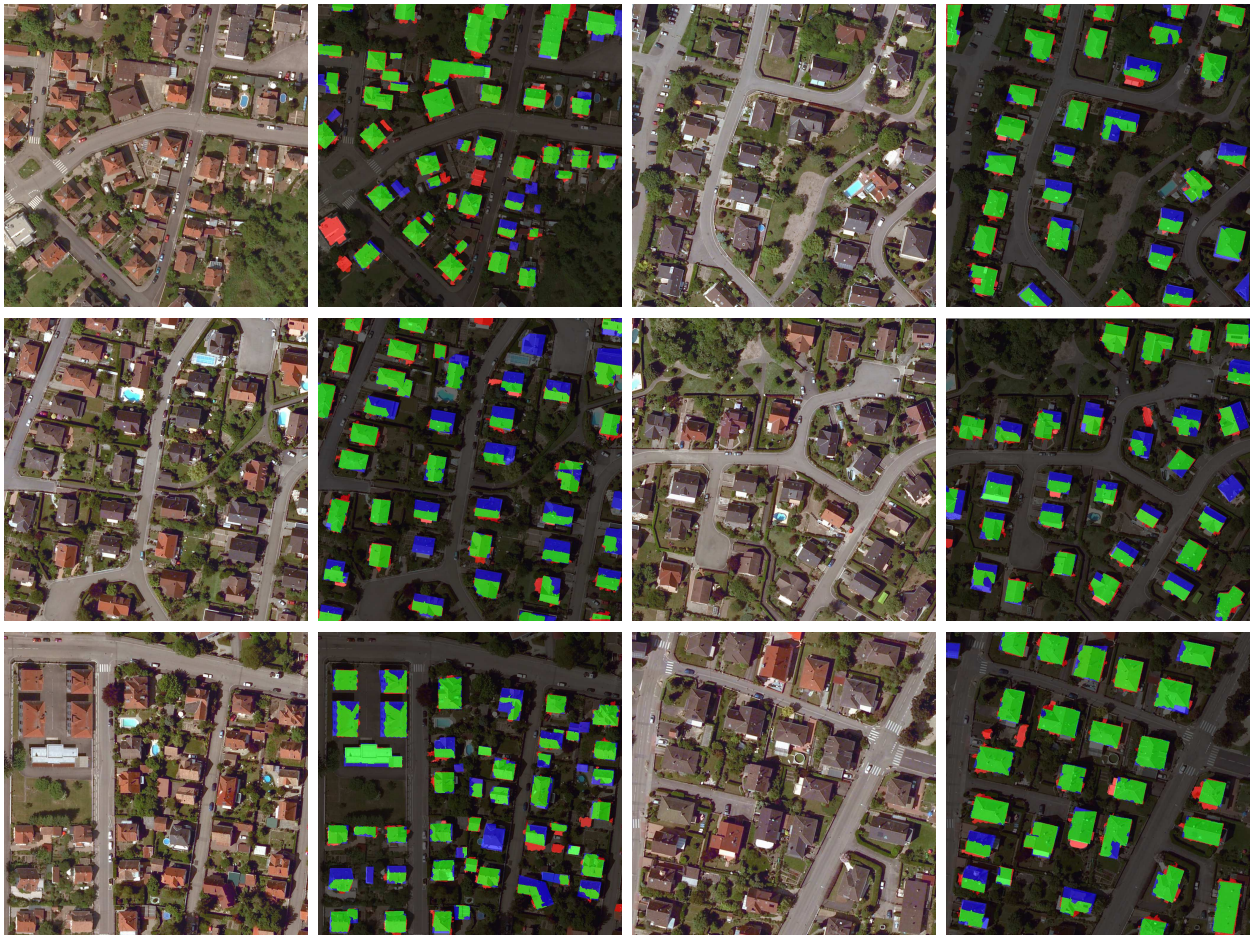


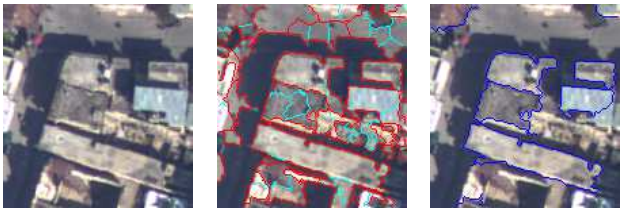
Fig. 8: (*Continued*). Test images and the results of the proposed approach on BD ORTHO® images. (First column) Test images #9-11. (Second column) Detected buildings for test images #9-11. (Third column) Test images #12-14. (Fourth column) Detected buildings for test images #12-14.

FN is assigned to a reference object when it corresponds to a resulted object with a limited amount of overlap ($< 60\%$). Thus, it is possible to compute P, R and F_1 for object-based performance measurement.

C. Results and Discussions

1) Qualitative and Quantitative Evaluation: Fig. 8 shows the building detection results by assigning the green, red and blue colors to TP, FP and FN in the pixel-based evaluation, respectively. The proposed approach revealed precision ratios ranging between 69.5% and 90.8% (Table I). Among all 14 images, 10 achieved precision levels of 75% and over; thus, the overall pixel-based precision rate is computed to be 83.75%. The pixel-based recall ratios of the proposed approach are high, ranging between 56.8% and 87.3%. Among all 14 images, 10 achieved pixel-based recall ratios of over 70%. The overall recall ratio is 74.40%, which is also considered to be relatively good. Altogether, the overall building detection F_1-score of 78.80% is good taking into account the complexities of the test images and involved imaging conditions.

As far as object-based evaluation is concerned, for all test images, our approach detected 309 of 371 buildings (total number of buildings) as long as a building is accepted to be correctly detected if we require an area overlap at least 60% of building for a correct detection of building object. Thus, $\|\mathrm{TP}\|$ and $\|\mathrm{FN}\|$ are respectively 309 and 62. For that case, our approach produced only 53 objects that do not correspond to any building object in the reference data ($\|\mathrm{FP}\|$ is equal to 53). The object-based evaluation of our approach revealed approximately 85.36% and 83.29% ratios for the precision and recall metrics, respectively. Thus, we can conclude that most of the detected buildings are nearly complete and the results



(a) Input image (b) Segmentation (c) Detected buildings

Fig. 9: Detected buildings in a dense urban area.

TABLE I: Numerical results of the method of Femiani *et al.* [7] and the proposed method on our data set

Test Image (size)	Method of Femiani *et al.* [7]			Our method(%)					
	Pixel-based Performance (%)			Pixel-based Performance (%)			Object-based Performance (%)		
	P	R	F_1	P	R	F_1	P	R	F_1
#1 (481 × 401)	59.1	66.3	62.5	75.4	67.5	71.3	87.5	72.4	79.3
#2 (480 × 400)	70.5	81.1	75.4	77.4	87.3	82.1	87.8	94.7	91.1
#3 (300 × 360)	64.4	73.1	68.5	71.3	71.6	71.4	86.3	82.6	84.4
#4 (621 × 546)	71.0	75.2	73.0	69.5	80.6	74.7	86.2	87.7	87
#5 (601 × 521)	82.3	59.7	69.2	90.8	56.8	69.9	92.8	78.8	85.3
#6 (874 × 692)	57.2	65.2	60.9	72.0	71.4	71.7	62.8	57.4	60.0
#7 (452 × 552)	60.7	62.6	61.6	71.7	71.2	71.5	89.9	79.7	91.9
#8 (480 × 400)	79.8	62.2	69.9	77.9	71.2	74.4	92.9	74.3	82.5
#9 (1024 × 1024)	82.6	75.5	78.9	80.3	85.1	82.6	91.2	79.5	84.9
#10 (1024 × 1024)	84.8	70.4	76.9	85.9	64.1	73.4	91.6	55	68.8
#11 (1024 × 1024)	88.1	73.8	80.3	90.3	70.4	79.1	91.4	66.7	77.1
#12 (1024 × 1024)	87.2	65.7	74.9	85.5	74.5	79.6	100	66.6	79.9
#13 (1024 × 1024)	80.1	75.5	77.7	86.6	67.3	75.7	91.7	73.3	81.5
#14 (865 × 865)	87.9	72.2	79.3	88.5	86.1	87.3	92	82.1	86.8
Overall (μ)	79.23	71.13	74.96	83.75	74.40	78.80	85.36	83.29	84.31
Min	57.2	59.7	62.5	69.5	56.8	69.9	62.8	55	60
Max	88.1	81.1	80.3	90.8	87.3	87.3	100	94.7	91.9

TABLE II: Numerical results of Grabcut [30], multi-level partitioning [2], Kovacs's method [5] and our method.

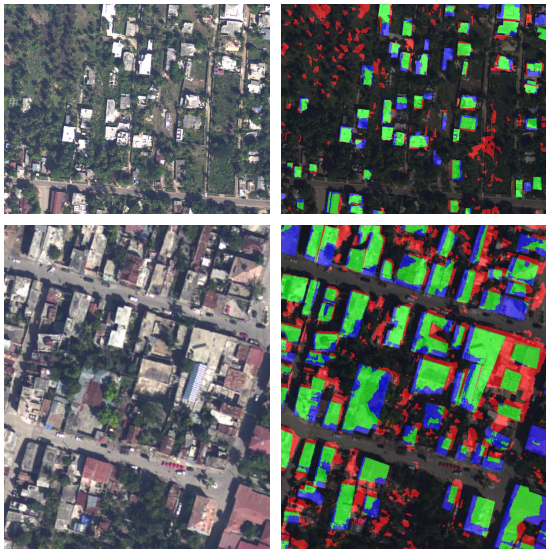
Database	Pixel-based Performance(%)											
	Ok *et al.* [30]			Ok [2]			Manno-Kovacs and Ok [5]			Proposed method		
Test Image	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
#1 (560 × 367)	59.1	58.6	58.8	36.5	56.8	44.4	81.2	75.0	78.1	66.9	74.8	70.6
#2 (554 × 483)	70.8	49.8	58.5	76.8	78.9	77.8	74.3	86.4	79.9	64.1	85.4	73.2
#3 (468 × 304)	60.4	76.3	67.4	60.1	90.2	72.1	69.2	89.0	77.9	61.9	85.9	71.9
#4 (896 × 600)	54.6	64.8	59.3	52.4	76.7	62.3	86.6	78.8	82.5	75.6	84.0	79.6
#5 (1213 × 958)	71.5	61.7	66.2	70.2	89.5	78.7	91.0	88.1	89.6	86.2	74.6	80.0
#6 (922 × 634)	46.3	80.0	58.7	23.8	74.4	36.1	87.4	68.2	76.7	83.8	73.2	78.2
#7 (928 × 639)	77.5	83.2	80.3	77.2	87.3	81.9	81.7	88.8	85.1	79.2	82.2	80.7
#8 (1009 × 695)	72.2	69.4	70.8	68.1	86.9	76.4	86.4	83.6	85.0	69.3	70.2	69.8
#9 (1615 × 1209)	47.4	62.3	53.9	40.6	74.6	52.6	89.9	90.2	90.0	81.8	88.9	85.2
#10 (1656 × 1240)	30.6	71.5	42.8	20.0	71.4	31.3	61.0	73.0	66.4	82.1	84.4	83.3
#11 (1222 × 915)	70.1	92.2	79.6	77.9	95.9	86.0	83.7	87.0	85.3	83.4	66.2	73.8
#12 (1311 × 848)	46.5	17.3	25.2	41.1	32.2	36.1	84.4	81.0	82.7	48.4	25.7	33.6
#13 (1193 × 772)	62.6	52.3	57.0	67.6	86.0	75.7	86.2	85.1	85.6	71.9	69.3	70.6
#14 (1193 × 771)	61.1	43.1	50.5	66.6	71.3	68.8	84.3	85.9	85.1	61.7	69.4	65.4
Average	57.5	61.9	59.6	53.1	78.1	63.2	83.5	84.4	83.9	74.2	70.1	72.1

are fairly acceptable in terms of an object-based point-of-view.

Let us investigate the test images in detail. First, we consider rural areas with high vegetation density and very low building density. The lowest pixel-based precision ratio (69.5%) is obtained for image #4. This poor performance is due to the reason that lots of buildings are occluded by vegetation. The number of TP pixels is therefore low. Considering image #5, regardless of its complexities, our method recovers most of the buildings. The object-based precision ratio is high (92.8%). Furthermore, the results on image #1, #2 and #3 show the efficiency of our method in detecting buildings in dense urban areas. The pixel-based F_1 are relatively good (respectively 71.3%, 82.1%, 71.4%). A zoom of the top-left corner in image #1 shows the ability of our method to separate the attached buildings (Fig. 9).

Besides, since some parts of building boundaries have similar rooftop characteristics with their surroundings, our method over-detect some building boundaries. Typical examples are visible in images #1-3, #7 and #8. This causes more FP pixels, therefore the precision ratio decreases. On the upper-left corner of image #8, two buildings are attached together but only one building generate its shadow. Only one building is detected since the shadow of the other one is missing. For BD ORTHO® images #9-14, a visual inspection of the results gives the impression that the proposed approach recovered successfully most of the buildings, without producing too many FP pixels. The pixel-based precision ratios are high (80.3% to 90.3%). As buildings are systematically located in a single-detached style, our approach achieve very high object-based precision ratios (91.2% to 100%).

2) Comparisons: We first compare our method with the method of Femiani *et al.* [7]. The results on image #6 and #7 are shown in Fig. 10. Since this method produces lots of false positives objects (small red), we only show the pixel-based evaluation results (Table I). The proposed method gives a highly competitive result with mean precision 83.75%, mean



(a) Input image (b) Method of [7]

Fig. 10: Results of building detection for test image #6 and #7 of the method of Femiani *et al.* [7].

recall 74.40% and mean F_1-score 78.80% compared with mean precision 79.23%, recall 71.13% and F_1-score 74.96% of the method of Femiani *et al.*.

Then, we compare our method with the method of Ok *et al.* [30], the method of Ok [2] and the method of Manno-Kovacs and Ok [5]. These methods were tested on a building detection benchmark proposed by Ok and Seneras [29]. As shown in Table II, the proposed method outperforms the method of Ok *et al.* [30], the method of Ok [2] but it is not better than the method of Manno-Kovacs and Ok [5]. This method gives a better result with an overall F-score of 83.9 % to be compared to 72.1 % of our method, but recall that our method uses only the visible bands R, G, B and the sun azimuth angle.

D. Sensitivity of Parameters

Table III lists the default settings of the parameters for NOAA images. A large number of tests on different parameters are performed. The effects of each parameter on the detection performance are illustrated in Fig. 13.

1) Detection of Shadows cast by Buildings: l_{se} is the length of structuring element of the binary morphological dilation. This parameter needs to be set large enough (60 pixels). d_{sh} is the minimum length of the boundary between buildings and their corresponding shadows. 10 values of d_{sh} (2.5 - 7 meters) are tested. When d_{sh} gets low values, the non-building objects having the rectangular form are considered as buildings. The FP pixels are therefore high and the precision ratio is low (Fig. 13(a)). Conversely, when d_{sh} is high, some buildings are misdetected. $\|FN\|$ is high and the recall ratio is low. Thus, d_{sh} is chosen as 5 meters, which maximizes the F_1-score.

2) Oversegmentation SLIC: Since oversegmentation SLIC [27] is considered as a preprocessing step, it is important to reduce error propagation to the final building validation. The parameter m controls the tradeoff between superpixel compactness and boundary adherence. The greater the value of m, the more spatial proximity is emphasized and the more compact the cluster is. m is set as 10, that allows to regularize regions with good boundary adherence. The sole parameter is the number of desired superpixels. This parameter is set so that the initial superpixel size is η_{sup}. As shown in Fig. 13(b), both precision and recall ratios do not significantly change for the low values of η_{sup}. However, when η_{sup} is high, $\|\text{FP}\|$ and $\|\text{FN}\|$ increases because the oversegmentation does not adhere well to building boundaries. Precision and recall ratios slightly decrease. Conversely, if η_{sup} is low, the image is segmented into more regions and we can not benefit from the interest of oversegmentation (Fig. 12). This parameter can be determined by varying its value on some test images and observing the oversegmentation results. In our algorithm, η_{sup} is set to 175, which maximizes the F_1-score and sufficiently preserve the boundaries of objects and structures without under-segmentation errors in images.

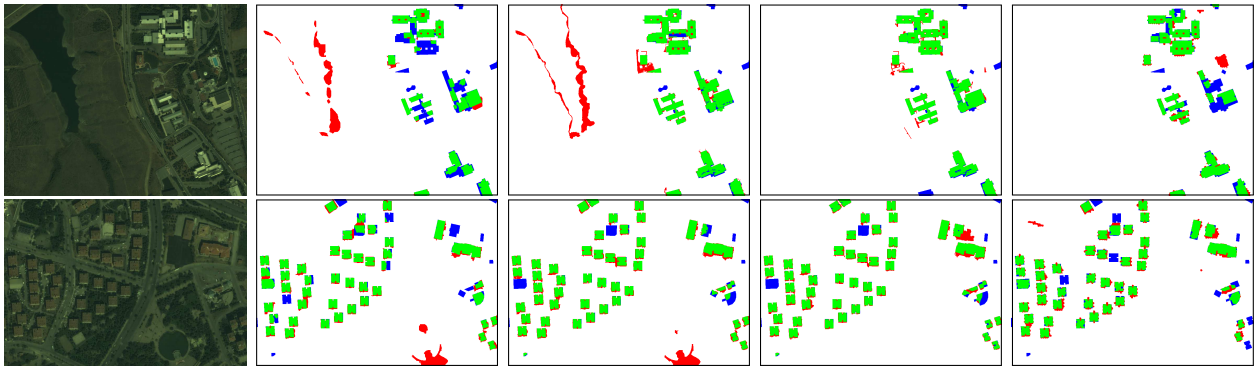
TABLE III: Parameter settings for NOAA aerial image.

Step	Parameter	Value
Building-Shadow Boundary Detection	l_{se}	60
	d_{sh}	5m
Oversegmentation SLIC [27]	η_{sup}	175
	m	10
ICM optimization - Region Classification	K	12
	β	150
	τ_{max}	500
	ϵ	0.01
Region Merging	T_S	15
	T_R	0.65
Determination of Final Building	T_B	0.8

3) Region Classification: To determine the number of classes K, the MRF model is ignored. The region growing image segmentation is therefore reduced as the initialization of region classification followed by a merging process. Fig. 13(c) indicates that small value of K can not distinguish building regions from their surroundings. The number of FP pixels is therefore high and the precision ratio is low. Conversely, when K is high, building is segmented into two or more regions. And only one region neighboring shadow is detected as building. The number of TP and FN pixels are therefore both low. We choose K as 12 for the best performance.

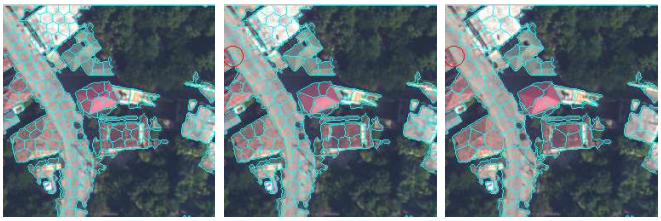
For the ICM algorithm, the maximum number of iterations τ_{max} is set to 500, which is normally not exhausted since the early stopping criterion is met with the stopping criteria ϵ is set to 0.01. The general rule for determining β is that it should be set to a large value for simple scenes and small for complex scenes. Fig. 13(d) indicates improvements as β increases for both the precision and recall ratio. When β is high, the neighboring regions of a building are classified in the same class label as building. This causes more FP pixels and the precision ratio decreases. β is chosen as 150, which maximizes the F_1-score computed for the pixel-based case.

4) Region Merging: Fig. 13(e) indicates that small values of T_S significantly reduce the precision ratio. High values of T_S do not significantly change the results because we need only one building segment to detect the entire rooftop. The



(a) Input image (b) Ok *et al.* [30] (c) Ok [2] (d) Manno-Kovacs, Ok [5] (e) Proposed method

Fig. 11: Results of some building detection methods for test patch #5 and #7 of building detection benchmark [5].



(a) η_{sup} = 125 (b) η_{sup} = 175 (c) η_{sup} = 225

Fig. 12: An example of oversegmentation SLIC with different values of η_{sup}. Low value of η_{sup} produces lots of regions and high value of η_{sup} makes the oversegmentation not adhere well to building boundaries (red circle).

optimal value of T_S is set to 15 pixels. Fig. 13(f) indicates that performance decreases in both the precision and recall ratios as the minimum rectangularity degree T_R values increases. Indeed, high values of T_R prevent the merging and as a result, $\|\mathrm{TP}\|$ decreases and $\|\mathrm{FN}\|$ increases. T_R is set to 0.65.

5) Final Determination of Buildings: As shown in Fig. 13(g), with the low value of T_B, we detect the maximum number of TP pixels, minimum number of FN pixels, but get lots of FP pixels, the precision ratio is low and the recall ration achieves the maximum. Conversely, if T_B is high, some buildings will be missed. As a result, all three measures are low. We set the optimal value of T_B to 0.8, which maximizes the F_1-score computed for the pixel-based case.

E. Performance

We implemented our algorithm in MatLab R1013a and tested on a PC (Intel Core i5 CPU 3.3 GHz with 8 GB RAM). The processing time is highly dependent on the dimension of the input images as well as the amount of details (number of building-shadow boundary objects). Our implementation took on average around 18 s per rooftop to complete, considering an average image size of 1024 $\times$ 1024 pixels with an average of 95 building-shadow boundary objects. The most time-consuming step is region classification (when the scene is more complex, the method takes more time). A migration from MatLab environment to C/C++ is expected to improve the processing time.

F. Limitations

Since shadows are used to support directly the building detection step, this approach can not detect buildings whose shadow is not visible or missing, like [2], [7], [8]. Besides, under oblique lighting, a gabled rooftop may exhibit significantly different intensities on the sloped portions of the roof, so the proposed method may lose its efficiency as it detects only one side of the rooftop . The low solar elevation angle may cause more severe shading effects on building rooftop. The dark part of building (self shadow) may be mislabeled as cast shadow and dismissed by the proposed method. Besides, the clouds and the terrains can generate shadows on the lower ground, and the proposed method systematically classifies them as buildings (e.g. image #6, Fig. 8). In urban environment, the proposed method may detect and label several non-building regions such as roads as buildings. Future work could attempt to integrate a method particularly designed for road detection. Roads can be masked out beforehand. Moreover, if the rooftop contains several components with different colors, our method fail to obtain the entire rooftops.

IV. Conclusion and Future Work

We have presented a new framework for building detection using only RGB images. We focus on buildings with right-angled corners, characterized as a collection of rectangles. The vegetation and shadow areas are first extracted. The boundaries between shadows and their corresponding buildings are then detected. Image is oversegmented into smaller homogeneous regions. An iterative region classification-merging is applied over these regions. At each iteration, regions are classified using a MRF-based image segmentation, then, according to the position of shadows, regions having the same class label are merged to produce new regions whose shapes appropriate to rectangles. The final buildings are determined using the RMBR method. We test our method on a variety of image data sets over different scenes, and the results reveal that the proposed method improves the performance of rooftop extraction, both at pixel and object levels.

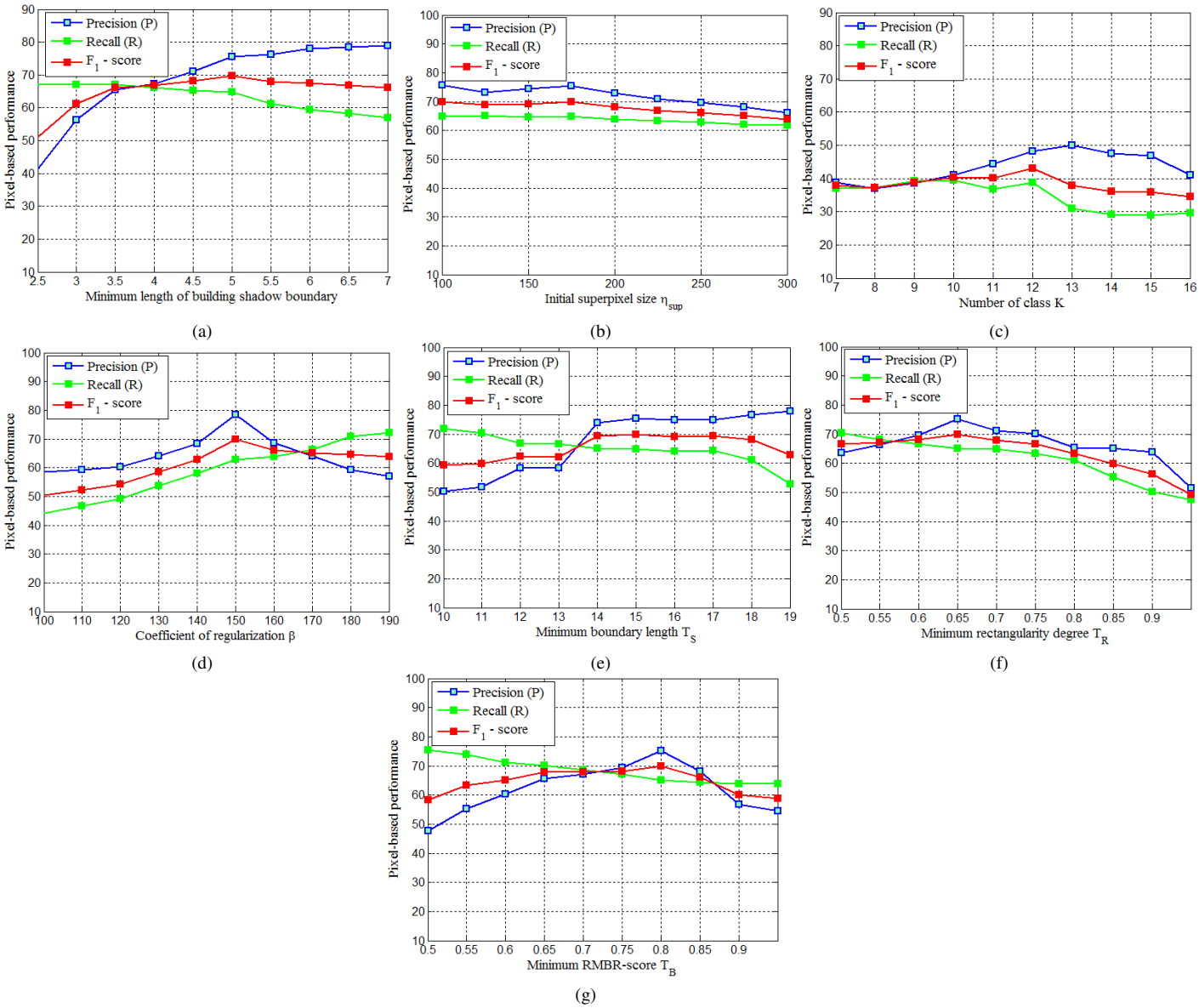


Fig. 13: Pixel-based performance for different parameter settings. The nonvarying coefficients are kept at their optimal settings.

ACKNOWLEDGEMENT

We would like to thank Rapid Mapping Service, SERTIT (http://sertit.u-strasbg.fr/) for providing data. We would also like to thank J. Femiani for processing his algorithm on our images and the reviewers for their insightful comments on the paper, as these comments led us to an improvement of the work. In addition, we would also like to thank the Institut Carnot (http://www.instituts-carnot.eu/) and the Alsace region, France for PhD funding.

REFERENCES

[1] E. Baltsavias, "Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems," *ISPRS J. Photogr. Remote Sens.*, vol. 58, no. 3, pp. 129–151, 2004.

[2] A. Ozgun Ok, "Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts," *ISPRS J. Photogr. Remote Sens.*, vol. 86, pp. 21–40, 2013.

[3] C. Lin and R. Nevatia, "Building detection and description from a single intensity image," *Comput. Vis. Image Underst.*, vol. 72, no. 2, pp. 101–121, 1998.

[4] B. Sirmacek and C. Unsalan, "Building detection from aerial images using invariant color features and shadow information," in *Proc. ISCIS*. IEEE, 2008, pp. 1–5.

[5] A. Manno-Kovacs and A. O. Ok, "Building detection from monocular VHR images by integrated urban area knowledge," *Geoscience and Remote Sensing Letters, IEEE*, vol. 12, no. 10, pp. 2140–2144, 2015.

[6] A. Manno-Kovacs and T. Sziranyi, "Orientation-selective building detection in aerial images," *ISPRS J. Photogr. Remote Sens.*, vol. 108, pp. 94–112, 2015.

[7] J. Femiani, E. Li, A. Razdan, and P. Wonka, "Shadow-based rooftop segmentation in visible band images," *IEEE J. Selected Topics Appl. Earth Observat. Remote Sens.*, vol. 8, no. 5, pp. 2063–2077, 2015.

[8] E. Li, J. Femiani, S. Xu, X. Zhang, and P. Wonka, "Robust rooftop extraction from visible band images using higher order CRF," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4483–4495, 2015.

[9] D. Chen, S. Shang, and C. Wu, "Shadow-based building detection and segmentation in high-resolution remote sensing image," *J. Multimed.*, vol. 9, no. 1, pp. 181–188, 2014.

[10] C. Senaras and F. T. Y. Vural, "A Self-Supervised Decision Fusion

Framework for Building Detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 1780–1791, 2016.

[11] T.-T. Ngo, C. Collet, and V. Mazet, "MRF and Dempster-Shafer theory for simultaneous shadow/vegetation detection on high resolution aerial color images," in *Proc. ICIP*, 2014.

[12] A. Qin and D. A. Clausi, "Multivariate image segmentation using semantic region growing with adaptive edge penalty," *IEEE Trans. Image Process.*, vol. 19, no. 8, pp. 2157–2170, 2010.

[13] J. Besag, "On the statistical analysis of dirty picture," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 48, no. 3, pp. 259–302, 1986.

[14] O. A. Lopez and E. Raven, "An overall evaluation of irregular-floor-plan-shaped buildings located in seismic areas," *Earthquake spectra*, vol. 15, no. 1, pp. 105–120, 1999.

[15] C. Benedek, X. Descombes, and J. Zerubia, "Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 33–50, 2012.

[16] G. T. Toussaint, "Solving geometric problems with the rotating calipers," in *Proc. IEEE Melecon*, vol. 83, 1983, p. A10.

[17] P. L. Rosin, "Measuring rectangularity," *Mach. Vis. Appl.*, vol. 11, no. 4, pp. 191–196, 1999.

[18] ——, "Measuring shape: ellipticity, rectangularity, and triangularity," *Mach. Vis. Appl.*, vol. 14, no. 3, pp. 172–184, 2003.

[19] D. Chaudhuri and A. Samal, "A simple method for fitting of bounding rectangle to closed regions," *Pattern Recogn.*, vol. 40, no. 7, pp. 1981–1989, 2007.

[20] D. Chaudhuri, N. Kushwaha, I. Sharif, and A. Samal, "Finding best-fitted rectangle for regions using a bisection method," *Mach. Vis. Appl.*, vol. 23, no. 6, pp. 1263–1271, 2012.

[21] L. Sahar, S. Muthukumar, and S. P. French, "Using aerial imagery and GIS in automated building footprint extraction and shape recognition for earthquake risk assessment of urban inventories," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3511–3520, 2010.

[22] E. Kwak and A. Habib, "Automatic representation and reconstruction of DBM from LiDAR data using Recursive Minimum Bounding Rectangle," *ISPRS J. Photogr. Remote Sens.*, vol. 93, pp. 171–191, 2014.

[23] T. Korting, L. Fonseca, L. Dutra, and F. Da Silva, "Image re-segmentation - A new approach applied to urban imagery," in *Proc. VISAPP*, vol. 1, 2008, pp. 467–472.

[24] T. S. Korting, L. V. Dutra, and L. M. G. Fonseca, "A resegmentation approach for detecting rectangular objects in high-resolution imagery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 4, pp. 621–625, 2011.

[25] S. Müller and D. W. Zaum, "Robust building detection in aerial images," *ISPRS Archives*, vol. 36, no. B2/W24, pp. 143–148, 2005.

[26] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "Turbopixels: Fast superpixels using geometric flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2290–2297, 2009.

[27] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, 2012.

[28] S. Krishnamachari and R. Chellappa, "Delineating buildings by grouping lines with MRFs." *IEEE Trans. Image Process.*, vol. 5, no. 1, pp. 164–168, 1995.

[29] A. Ozgun Ok and C. Senaras, "Building detection benchmark," http://biz.nevsehir.edu.tr/ozgunok/en/408/, 2015.

[30] A. Ozgun Ok, C. Senaras, and B. Yuksel, "Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1701–1717, 2013.