UNSUPERVISED AUTOMATIC BUILDING EXTRACTION USING ACTIVE CONTOUR MODEL ON UNREGISTERED OPTICAL IMAGERY AND AIRBORNE LIDAR DATA

T. H. Nguyen[1,2,*], S. Daniel[2], D. Guériot[1], C. Sintès[1], J.-M. Le Caillec[1]

[1] IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France
-(thanh.nguyen, didier.gueriot, christophe.sintes, jm.lecaillec)@imt-atlantique.fr
[2] Université Laval, Québec City, QC G1V 0A6, Canada - sylvie.daniel@scg.ulaval.ca

ICWG II/III: Pattern Analysis in Remote Sensing

KEY WORDS: Building extraction, Optical imagery, Airborne LiDAR, Active contour model, Snake model, Polygonization.

ABSTRACT:

Automatic extraction of buildings in urban scenes has become a subject of growing interest in the domain of photogrammetry and remote sensing, particularly with the emergence of LiDAR systems since mid-1990s. However, in reality, this task is still very challenging due to the complexity of building size and shape, as well as its surrounding environment. Active contour model, colloquially called snake model, which has been extensively used in many applications in computer vision and image processing, has also been applied to extract buildings from aerial/satellite imagery. Motivated by the limitations of existing snake models dedicated to the building extraction, this paper presents an unsupervised and automatic snake model to extract buildings using optical imagery and an unregistered airborne LiDAR dataset, without manual initial points or training data. The proposed method is shown to be capable of extracting buildings with varying color from complex environments, and yielding high overall accuracy.

1. CONTEXT

Automatic extraction of buildings from aerial/satellite imagery in urban and residential scenes has become a subject of growing interest in the domain of photogrammetry and remote sensing. Indeed, a large number of building detection and extraction techniques have been reported over the last few decades, particularly with the emergence of LiDAR (Light Detection and Ranging) systems. For instance, (Rottensteiner, Briese, 2003) proposed a building detection method exploiting primarily LiDAR data while removing vegetation using imagery data. (Sohn, Dowman, 2007) focused on exploiting the synergy of IKONOS multispectral imagery combined with a hierarchical segmentation of a LiDAR digital elevation model (DEM) to extract buildings. Another method of building detection was proposed by (Awrangjeb et al., 2010) based on building masks obtained from LiDAR and multispectral imagery. These methods using both complementary sources, namely LiDAR data and optical imagery, achieve better building extraction results. Other building detection and extraction methods utilize only LiDAR data such as (Khoshelham et al., 2013, Zhang et al., 2013). They involve segmentation algorithms and classification using attributes such as building size, shape, height and PCA (Principal Component Analysis) features. However, these approaches usually face problem of misclassification of vegetation as buildings (Zhang, Lin, 2017).

Originally introduced by (Kass et al., 1988), snake model has been studied for building extraction from urban area using aerial and satellite imagery. While (Guo, Yasuoka, 2002) used snake model with balloon force to extract buildings, (Peng et al., 2005) focused in increasing stability of snake convergence. (Kabolizade et al., 2010) proposed to improve the model with imagery data coupled with a Digital Surface Model (DSM) generated from LiDAR data. This improvement involves using

two energy terms based on the variances of height and gray-level between snake points. Hence, the height variance energy term requires height information for every pixel of the image, in other words, the DSM must be of same size and resolution as the imagery data. This can be problematic since LiDAR dataset does not often have a spatial resolution as high as aerial imagery, and height data interpolation may be unreliable. (Fazan, Dal Poz, 2013) proposed another approach based on exhaustive searches of rectilinear building corners from the image, optimized by dynamic programing. However, this method depends heavily on initial points to have decent results. (Ahmadi et al., 2010) proposed an area-based geometrical snake model to detect building boundaries without height information or initial points. Nevertheless, this method requires *a priori* sampled gray-levels of buildings and grounds (i.e. training data) to attract the snakes toward desired buildings. It also may not work well with building roofs with varying gray levels.

To summarize, the main common issues of snake models in building boundary extraction are:

- Sensitivity to noise and image details;
- Dependence on initial points or training data;
- Weak convergence to building corners;
- Snake's convergence sensitivity to its number of points and weighting parameters.

Motivated by the limitations of existing snake models listed above, this paper presents an unsupervised snake model to extract buildings using optical imagery and an unregistered airborne LiDAR dataset, without manual initial points or training data. Instead, we propose to initialize our proposed snake model with projected LiDAR building boundary points. The movement of snake model is then governed by an additional energy term related to its similarity to the LiDAR building boundary. Our method is also able to extract buildings with varying color, from complex environments.
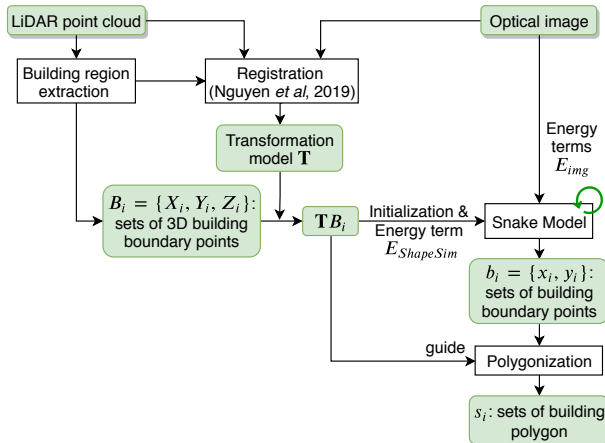
---

*Corresponding author

Figure 1. Flowchart of the proposed method.

Figure 2. Building region extraction from LiDAR point cloud.

The remainder of this paper is organized as follows. Section 2 is devoted to the description of the proposed method. Then, experimental results are presented in Section 3. Finally, Section 4 provides conclusions and perspectives of this work.

2. METHOD

The novelty of our approach resides in an unsupervised method that carries out effectively building extraction from urban scene. This method is based on a snake model initialized and enhanced by integrating with LiDAR data, followed by an improved building polygonization. Presented by Fig. 1, the proposed method is composed of three parts: *(i)* determination of initial points involving the registration of datasets, *(ii)* the snake model, and *(iii)* the improved building boundary polygonization.

2.1 Determination of initial points

2.1.1 Registration This research study involves an airborne LiDAR data set which has been acquired independently of the optical imagery, i.e. two different surveys, done at different time, with different platforms. Such a context yields spatial discrepancies between data sets that provide the snake model with wrong initial points. Therefore, a registration has been carried out beforehand (Nguyen et al., 2019). Indeed, such a registration is substantially important to a building extraction procedure, since it involves many problems exemplified by the work of (Gilani et al., 2016).

The registration consists in extracting building regions, on one hand, from the LiDAR point cloud, and on the other hand, from the optical image using a mean-shift segmentation. Extracted building segments are then matched using the Graph Transformation Matching algorithm proposed by (Aguilar et al., 2009). The resulting pairwise segments establish a set of correspondences between the two data sets. They are used to estimate a transformation model using the Gold Standard algorithm (Hartley, Zisserman, 2003, p. 187). This transformation model is used to register the image and LiDAR data sets with an accuracy good enough to get relevant snake initial points.

2.1.2 Building region extraction from LiDAR point cloud Presented by Fig. 2, the extraction of building regions from LiDAR point cloud is carried out through a series of steps. First, non-ground points are separated from ground points using an elevation thresholding. The threshold is set as follows,
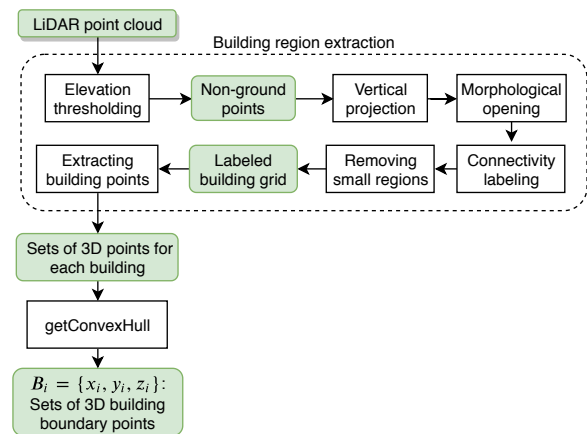
$T_e = \text{mean}(z_G) + \max\{2.5, \text{std}(z_G)\}$, where z_G denotes the elevation of ground points. All non-ground points are then vertically projected onto the plan $z = 0$. A raster grid representing these projected points is created. The resolution of the grid is set according to the LiDAR point cloud density in order to avoid null-valued pixels. A binary grid of same resolution is also generated. Its cell value is set to 1 or 0 according to the presence or absence of projected non-ground points in the cell ('1': presence, '0': absence). A morphological opening operator is then applied to remove small artifacts on the binary grid. Remaining grid cells with value set to 1 are grouped into labeled segments based on their connectivity. Next, small segments (e.g. smaller than 10 m^2) are removed. The resulting grid consists of a number of relatively large labeled segments that relate to buildings. These segments are then used to select the building points in the LiDAR point cloud. A convex hull is calculated on each set of 3D building points, yielding a set of boundary points, denoted by B_i. These points are then projected onto the optical image using the transformation model $\mathbf{T}$. They will be used as the initial points of the snake model for each building.

2.2 Snake models for building boundary extraction

2.2.1 Traditional snake models An active contour, or a snake is a dynamic curve $\mathbf{x}(s) = (x(s), y(s))$, where $s \in [0, 1]$ is the normalized arc length, defined within an image domain that is deformable under the influence of internal and external forces (Xu, Prince, 1997). Mathematically, the behaviors of the snake are governed by an energy function, which is defined as follows,

$$E_{\text{snake}} = \int_0^1 (E_{\text{int}}(\mathbf{x}(s)) + E_{\text{ext}}(\mathbf{x}(s)))ds$$
$$E_{\text{int}}(\mathbf{x}(s)) = \frac{1}{2}\left(\alpha|\mathbf{x}'(s)|^2 + \beta|\mathbf{x}''(s)|^2\right) \quad (1)$$
$$E_{\text{ext}}(\mathbf{x}(s)) = E_{\text{img}}(\mathbf{x}(s)) + E_{\text{con}}(\mathbf{x}(s))$$

where E_{int} and E_{ext}, respectively, represent the internal and external energy terms. The internal energy relates to the tension (the amount of stretch) and the rigidity (the amount of curvature) of the snake, respectively controlled by weighting parameters α and β. $\mathbf{x}'(s)$ and $\mathbf{x}''(s)$ denote the first and second derivatives of $\mathbf{x}(s)$ with respect to s. The external energy E_{ext} is composed of the forces due to the image itself E_{img}, and other constraint forces introduced by the users E_{con}, e.g. inflation force introduced by balloon model (Cohen, 1991).

The external energy of an image I related to its salient features i.e. lines, edges and terminations (i.e. line end-points, corners)

can be generally formulated as follows,

$$E_{\text{img}} = w_{line}E_{line} + w_{edge}E_{edge} + w_{term}E_{term} \qquad (2)$$

where $w_{line}, w_{edge}, w_{term}$ are the weights of the respective salient features. Mathematical formulation of these energy terms is described in (Kass et al., 1988).

A snake that minimizes E_{snake} must satisfy the Euler equation

$$\alpha\mathbf{x}''(s) - \beta\mathbf{x}''''(s) - \nabla E_{\text{ext}} = 0 \qquad (3)$$

In order to solve (3), $\mathbf{x}$ is regarded as a function of time t as well as of s. Then, the partial derivative of $\mathbf{x}$ with respect to t is then set equal to the left hand side of (3), as follows,

$$\mathbf{x}_t(s,t) = \alpha\mathbf{x}''(s) - \beta\mathbf{x}''''(s) - \nabla E_{\text{ext}} \qquad (4)$$

When $\mathbf{x}(s,t)$ stabilizes, the partial derivative term $\mathbf{x}_t(s,t)$ vanishes and we obtain a solution for (3). This dynamic equation can also be viewed as a gradient descent algorithm designed to solve (1). A numerical solution to (4) can be found by discretizing the equation and solving the discrete system iteratively (Kass et al., 1988).

2.2.2 Gradient vector flow (GVF) is proposed by (Xu, Prince, 1997) to allow more flexible initialization of snake and encourage convergence to boundary concavities, as well as improving the model's robustness versus image noise. GVF field is defined as the vector field $\mathbf{v}(x,y) = (u(x,y), v(x,y))$ that minimizes the energy functional

$$E_{\text{GVF}} = \int\int \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2|\mathbf{v} - \nabla f|^2 dxdy \qquad (5)$$

with μ is a controllable smoothing term, and f represents external forces from Eq. 3, i.e. $f(x,y) = -E_{\text{ext}}$. Using the *calculus of variations* (Courant, Hilbert, 2008), the GVF can be found by solving:

$$\mu\nabla^2 u - (u - f_x)(f_x^2 + f_y^2) = 0$$
$$\mu\nabla^2 v - (v - f_y)(f_x^2 + f_y^2) = 0 \qquad (6)$$

where ∇^2 is the Laplacian operator. The Euler equations (6) can be solved by regarding u and v as functions of time and solving

$$u_t(x,y,t) = \mu\nabla^2 u(x,y,t) -$$
$$[u(x,y,t) - f_x(x,y)] \cdot [f_x(x,y)^2 + f_y(x,y)^2]$$
$$v_t(x,y,t) = \mu\nabla^2 v(x,y,t) -$$
$$[v(x,y,t) - f_y(x,y)] \cdot [f_x(x,y)^2 + f_y(x,y)^2] \qquad (7)$$

Once computed, $\mathbf{v}(x,y)$ will replace the potential force $-\nabla E_{\text{ext}}$ in the dynamic equation (3), yielding

$$\mathbf{x}_t(s,t) = \alpha\mathbf{x}''(s) - \beta\mathbf{x}''''(s) + \mathbf{v} \qquad (8)$$

This equation is solved similarly as the traditional snake model, i.e. by discretization and iterative solution. The parametric curve solving the above dynamic equation is thus called a GVF snake.

2.2.3 Proposed snake model In this paper, we propose adding a new energy term as a constrained force (i.e. E_{con}), which is

Figure 3. 3D building boundary points projected onto the optical image, i.e. $\mathbf{T}B_i$, which are used to initialize the snake model and for the energy term E_{ShapeSim}.

calculated based on the similarity between the shape formed by snake points and the projected LiDAR building boundary, as follows,

$$E_{\text{ShapeSim}}(\mathbf{x}) = 1 - \exp\left(-\frac{d_H^2(\mathbf{x}, \mathbf{T}B_i)}{\delta}\right) \qquad (9)$$

where d_H denotes the Hausdorff distance, δ is the scale factor set accordingly to the size of the building, and $\mathbf{x}$ is the current snake. B_i represents the 3D building boundary extracted from LiDAR data, and $\mathbf{T}B_i$ is its projection onto the image using the transformation model $\mathbf{T}$ provided by the registration. The role of this energy term is to encourage the snake to maintain the same shape as the building boundary extracted from the LiDAR data while being attracted by the image salient features and under the influence of GVF fields.

In our approach, the projected LiDAR building boundaries, denoted by $\mathbf{T}B_i$, have an important role. They provide the snake algorithm with relevant initial points, and they are used in the energy term E_{ShapeSim} and the polygonization. Fig. 3 shows three examples of 3D building boundary points projected onto the optical image.

Up to this point, the proposed snake model has involved a number of parameters, such as α, β (from E_{int}), $w_{line}, w_{edge}, w_{term}$ and σ (from E_{img}), μ (from GVF) and lastly δ (from E_{ShapeSim}). To the best of our knowledge, the know-how to set these parameters for the snake to extract buildings from an imagery dataset still remains unresolved. This has driven many existing works to perform a trial-and-error approach to determine them (Peng et al., 2005, Kabolizade et al., 2010, Ahmadi et al., 2010). Similar approach is applied for our snake model, empirically setting $\alpha = \beta = 0.01, w_{line} = 0.04, w_{edge} = 2, w_{term} = 0.01$, with standard deviation $\sigma = 10$ used for image smoothing in E_{img}, $\mu = 0.2$ and lastly $\delta = 50$.

2.3 Polygonization of building boundary

The polygonization step is then applied to each set of building boundary points extracted with the snake model, i.e. b_i. It involves formulating these boundary points into regular building polygons. Such polygon usually consists of several parallel and perpendicular straight line segments. Many researches have addressed this topic for various purposes, such as building boundary (Dutter, 2007), cartography (Samsonov, Yakimova, 2017), and classification maps (Maggiori et al., 2017).

In this paper, we modified the method proposed by (Dutter, 2007) which polygonizes a building into three levels of shape: rectangular; Z-, T- or L-shape; U-shape. The original method starts with a Minimum Bounding Rectangle (MBR) (Freeman,
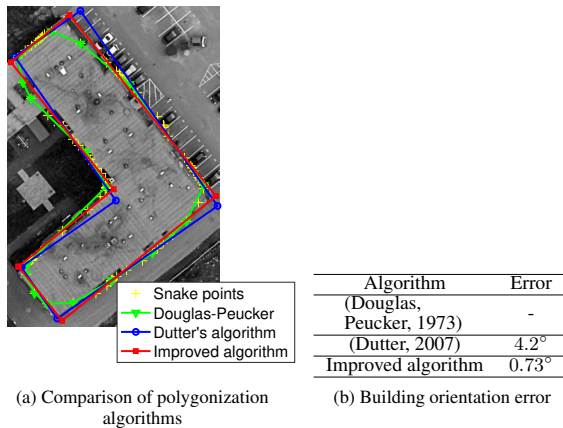
+ Snake points	
▼ Douglas-Peucker	
● Dutter's algorithm	
■ Improved algorithm	

Algorithm	Error
(Douglas, Peucker, 1973)	-
(Dutter, 2007)	4.2°
Improved algorithm	0.73°

(a) Comparison of polygonization algorithms (b) Building orientation error

Figure 4. Polygonization of the proposed snake model on an L-shaped building.

Shapira, 1975) containing all snake points, and the edges of building polygon are set to be parallel to the MBR edges. This setting may bias the polygonization result and eventually lead to errors in building orientation. In our proposed approach, instead of detecting the MBR containing snake points, we detect and use the MBR of the projected LiDAR building boundary points which yields more reliable building orientation.

A comparison of building boundary polygonization methods has been carried out. It involves our approach, Dutter's original polygonization (Dutter, 2007) and Douglas-Peucker line simplification algorithm (Douglas, Peucker, 1973). Results are provided by Fig. 4, which underline the significant reduction of building orientation error yielded by our approach. Since Douglas-Peucker algorithm only reduces the number of points of a boundary, it does not yield a building orientation.

3. RESULTS

This section is devoted to the assessments of the snake model and the overall building extraction performance.

3.1 Evaluation metrics

In order to assess performance of a building extraction method with respect to a ground truth, the following metrics are used:

1. Intersection over Union (IoU) (Jaccard, 1901):

$$\text{IoU} = \frac{\mathcal{A}(E \cap R)}{\mathcal{A}(E \cup R)} \times 100\% \quad (10)$$

where $\mathcal{A}(\cdot)$ represents the area measurement. IoU measures the ratio between the area of the intersection of an extracted building boundary E and the ground-truth building boundary R, over the area of their union.

Two related criteria, namely *Completeness* (C_p) and *Correctness* (C_r), are also used to facilitate the comparison of our approach with other works, e.g. (Fazan, Dal Poz, 2013, Gilani et al., 2016). Full descriptions of these metrics can be found in the paper of (Rutzinger et al., 2009).

- Completeness, measured by the *Recall* criterion:

$$C_p = \frac{TP}{TP + FN} = \frac{\mathcal{A}(E \cap R)}{\mathcal{A}(R)} \times 100\% \quad (11)$$

- Correctness, measured by the *Precision* criterion:

$$C_r = \frac{TP}{TP + FP} = \frac{\mathcal{A}(E \cap R)}{\mathcal{A}(E)} \times 100\% \quad (12)$$

where TP, FP and FN respectively denote the number of true positive (i.e. building pixels correctly identified as building pixels), false positive (i.e. non-building pixels identified as building pixels) and false negative pixels (i.e. building pixels not identified).

All three metrics IoU, C_p and C_r reach their best value at 100% and worst at 0%.

While the IoU metric reflects the overall accuracy of a building extraction method according to a ground truth, C_p represents the fraction of relevant identified building pixels over total number of actual building pixels, and C_r represents the fraction of relevant identified building pixels among all identified pixels.

2. Euclidean distance between centroids (EDC):

$$\text{EDC} = d(C_E, C_R) \quad (13)$$

where $d(\cdot)$ is the Euclidean distance, whereas C_E and C_R, respectively, stand for the centroid coordinates of the extracted building and the ground truth building.

3. Dominant angle rotation error (DARE):

$$\text{DARE} = |\theta_E - \theta_R| \quad (14)$$

where θ_E and θ_R represent the dominant angle of the extracted building and the ground truth building. The dominant angle of a polygonized building is determined according to its longest line segment.

3.2 Data sets

3.2.1 Quebec City The first data sets used for these assessments were acquired in Quebec City (QC, Canada) and are described in Table 1. Boundary of twenty buildings on this test area are manually determined and used as ground truth.

Data type	Aerial optical imagery	Airborne LiDAR
Spectral resolution	R, G, B, IR	1064 nm
Spatial resolution/ Point density	15 cm	2 points/m^2
Acquisition time (season)	June 2016 (summer)	Oct-Nov 2011 (winter)
Geometry/Properties	Orthorectified Georeferenced	Multi-return (4) Classified
Relative misalignment	1.41 m (before registration) 0.49 m (after registration)	

Table 1. Description of Quebec City data set (LiDAR data ©Ville de Québec, aerial imagery data ©Communauté Métropolitaine de Québec).

3.2.2 Vaihingen The proposed method is also tested using the ISPRS benchmark data set on Vaihingen, Germany (Cramer, 2010). This additional test aims to demonstrate its effectiveness on complex environments, and to compare it with other existing methods.

We use the true orthophoto mosaic given as a RGB image of 9-cm resolution generated from aerial images that are taken between July 24 and August 06, 2008; whereas the airborne
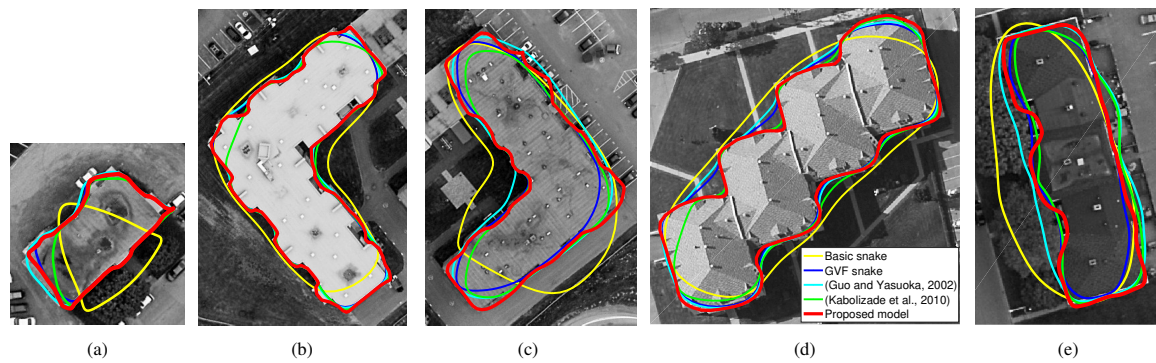
Figure 5. Results of snake models on simple and complex buildings. From left to right, (a) a rectangular building with varying color, then (b) an L-shaped building, (c) an L-shaped building with similar gray level to the background, (d) a gable roof building with its shadow, and (e) a gable roof building with similar gray level to the background.

LiDAR dataset of an average point density of 4 points/m^2 is acquired on August 21, 2008. Since the misalignment between the orthophoto and the airborne LiDAR data was already small, a registration between the datasets was not carried out.

This benchmark data set provides three test areas consisting of buildings with diversified characteristics, as well as their ground truth boundaries. Area 1 is situated in the center of the city and characterized by dense construction consisting of historic buildings with complex shapes and some trees. Area 2 is composed of high-rise residential buildings surrounded by trees. Finally, Area 3 is purely residential with detached houses and many surrounding trees.

3.3 Evaluation of snake models

Fig. 5 presents the results of different snake models applied to extract buildings from the Quebec City data set: basic snake without GVF and E_{img}, GVF snake, snake model of (Guo, Yasuoka, 2002), snake model of (Kabolizade et al., 2010), and our proposed method. In this comparison, all snake models are initialized by the LiDAR building boundaries projected on the optical image (i.e. $\mathbf{T}B_i$). As we focus on an unsupervised approach for snake models without a priori information of building gray-level, method of (Ahmadi et al., 2010) is not considered. Visual comparison shows that our proposed snake model always converges better toward the building true boundaries than other snake models on all presented buildings.

Table 2 summarizes the IoU metrics computed based on snake models with respect to ground truth building boundaries. All the snake models except the basic snake model show a high average IoU value, i.e. more than 88 %, which demonstrate the gain of using the projected LiDAR building boundaries as initial points for snake models.

However, we notice unstable IoU metric values from GVF snake and snake model of (Kabolizade et al., 2010), underlined by the IoU minimum values. Furthermore, IoU value of snake model of (Kabolizade et al., 2010) worsens if the building has similar color to its background (cf. Fig. 5c and 5e). On the other hand, snake model of (Guo, Yasuoka, 2002) yields good IoU but it is usually unable to converge toward building corners and boundary concavities (cf. Fig. 5c-5e).

3.4 Overall performance assessment

This sub-section is dedicated to the evaluation of the overall performance of our proposed method on both data sets. In addi-

Snake models	IoU		
	Mean	Min	Max
Basic snake	61.45 %	36.42 %	72.61 %
GVF snake	88.55 %	58.10 %	97.57 %
(Guo and Yasuoka, 2002)	89.04 %	79.85 %	96.83 %
(Kabolizade et al., 2010)	88.32 %	57.68 %	97.52 %
(Proposed model)	90.36 %	74.23 %	97.74 %

Table 2. Performance of snake models initialized by $\mathbf{T}B_i$ on Quebec City data set, based on IoU metric measured between snake models and ground truth building boundaries.

tion, using the Quebec City data set, we compare it with the approach using only LiDAR data, in order to demonstrate the benefit of using jointly imagery data with LiDAR data. Then, using the ISPRS Vaihingen data set, we compare the result provided by our method with six existing non-snake building extraction methods.

3.4.1 Performance results using Quebec City data set

Building extraction result yielded by our proposed method on urban area of Quebec City can be visually assessed through Fig. 6. Snake initial points and snake result (before polygonization) are also depicted. Many buildings with varying color, or color similar to the background or gable roofs are well delineated. The metric values averaged on the area are summarized by Table 3.

Metric	LiDAR only	Proposed method
IoU	85.51 %	91.12%
C_p	87.83 %	97.07 %
C_r	97.05 %	92.88 %
EDC	1.48 m	0.89 m
DARE	0.62°	0.81°

Table 3. Average pixel-based building extraction accuracy metrics on Quebec City data set.

Compared with metrics of LiDAR-only method, the proposed method yields more accurate IoU and EDC metrics. However, it can be remarked that the value of *Correctness* C_r of the LiDAR-only method is higher than the proposed method, i.e. 97.05% versus 92.88%. The reason is because in many cases, LiDAR boundary points can be found inside the true building boundaries, e.g. Fig. 3. These cases yield a C_r value equal to 100% (since $E \cap R = E$ when $E \subseteq R$), which increase the average value of C_r. Also, the DARE value of the proposed method is relatively small which confirms its reliability and accuracy in determining building orientation.

3.4.2 Performance results using ISPRS Vaihingen data set

The proposed method is applied on all three test areas, respect-
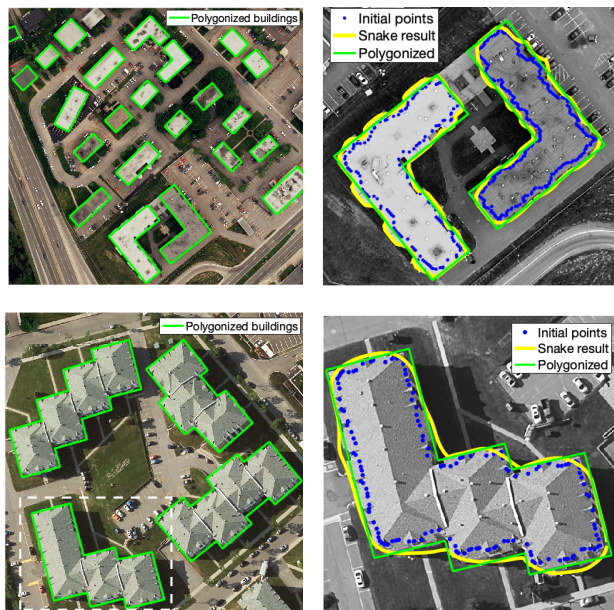
Figure 6. Building extraction result on urban area (Quebec City).

ively presented by Fig. 7a, 7c and 7e. Extraction results are then evaluated according to the ground truth building boundaries, provided within the benchmark data set. Fig. 7b, 7d and 7f illustrate this evaluation denoting true positive, false positive and false negative pixels. Their significations were mentioned in sub-section 3.1.

First, the results are visually assessed. Although a high percentage of building pixels are well extracted (represented by yellow pixels), several issues can be identified in all three test areas. Firstly, several buildings are not entirely extracted but only partially, causing false negative pixels. This issue is due to complex building roofs and connected buildings. Also, several small buildings are also not extracted. Secondly, several vegetation regions that are close to buildings are also extracted as buildings, causing false positive pixels.

Area	IoU	C_p	C_r
1	82.27 %	89.43 %	91.13 %
2	87.35 %	92.13 %	94.39 %
3	80.07 %	87.21 %	90.72 %
Avg	83.23 %	89.59 %	92.08 %

Table 4. Pixel-based building extraction accuracy metrics on ISPRS Vaihingen benchmark data set.

Pixel-based assessment metrics on each area are provided by Table 4. These metrics are then compared with other methods that are selectively proposed by (Gilani et al., 2016), and synthesized in Table 5. The first two methods are categorized *supervised*, as they involve a supervised classification methodology based on training data, according to the taxonomy by (Rottensteiner et al., 2014). On the other hand, the others proposed *unsupervised* and *data-driven* approaches which consist in extracting buildings without predefined constraint on the building appearance (Gilani et al., 2016).

Fig. 8 depicts graphically the performance of each method with relation to IoU, Correctness and Completeness metrics. Despite the discussed issues on this data set, our proposed method, being automatic and unsupervised, shows very competitive quantitative results. Indeed, it achieves relevant accuracy results on all three test areas. For instance, on Area 1 and 3, our proposed

(a) Area 1   (b) Result on Area 1

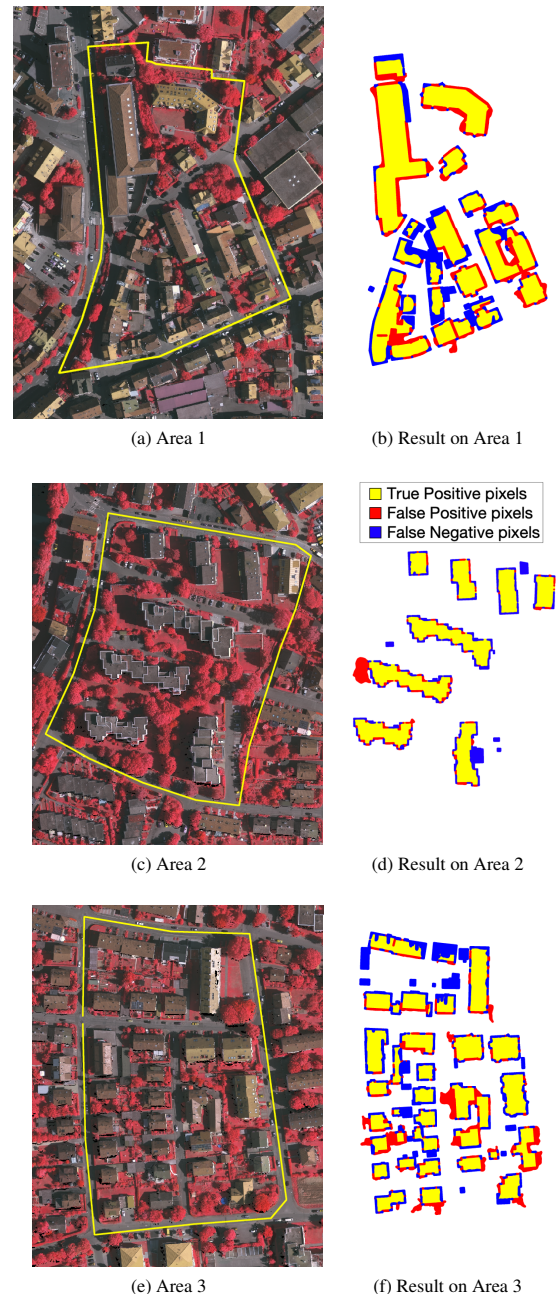(c) Area 2   (d) Result on Area 2

(e) Area 3   (f) Result on Area 3

Figure 7. Evaluation of building extraction using our proposed method with respect to the ground truth building boundaries on test areas of the ISPRS Vaihingen benchmark data set.

method yields better IoU than most of the other methods, except the supervised method KNTU, or on Area 2 where ours exceeds all the others.

However the tests conducted on the ISPRS data set underlined a limitation in our approach. Indeed, there are several null-valued regions in the orthophoto as shown in Fig. 9. These regions perturb the snake model convergence as they involve high gradient values in the image-based external energy term that attract the snake model.

4. CONCLUSIONS AND PERSPECTIVES

In this paper, we proposed and evaluated an unsupervised method to extract buildings from urban scenes, using snake model on

Method	Data types	Processing strategy
KNTU (Zarea, Moham-madzadeh, 2015)	LiDAR + image	supervised
Whuz (Zhan et al., 2012)		
IIST (Gilani et al., 2016)		unsupervised and data-driven
Fed_2 (Gilani et al., 2016)		
Mon2 (Awrangjeb et al., 2014)	LiDAR	
Yang (Yang et al., 2013)		

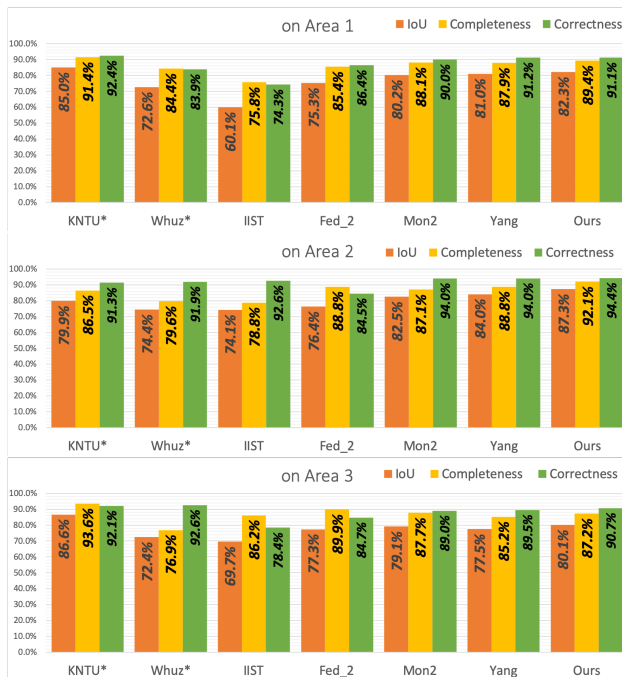Table 5. Methods that have been compared with the proposed method, using ISPRS Vaihingen data set.

Figure 8. Comparison of building extraction accuracy metrics on ISPRS Vaihingen benchmark data set among various methods. The '∗' sign indicates that a method is supervised.

aerial optical imagery with airborne LiDAR data. Without needing manual initial points nor training data, the proposed method is highly automatic and achieves a high accuracy of building extraction, indicated by different metrics on many test areas (cf. Tables 3 and 4). Indeed, compared to other methods dedicated to building extraction, our work provide an alternative solution that is unsupervised and yields a better accuracy than most of them, except the supervised method proposed by (Zarea, Mohammadzadeh, 2015).

These results also demonstrate significantly the advantage of the conjoint use of optical imagery with LiDAR data, even with a LiDAR dataset collected a long time before, e.g. five years. Indeed, by the virtue of LiDAR-based initialization and guidance, our proposed method succeeds to extract buildings with gable roofs and/or having varying color, or even when the roof has very similar color to its background. It also overcomes the problem of disturbed shadow regions near buildings which is usually problematic to a building extraction method, e.g. (Fazan, Dal Poz, 2013). Compared with methods that use only LiDAR data, the proposed approach provides more accurate building extraction results. Moreover, the fact that this approach uses an optical image with a LiDAR dataset that is acquired several years in advance also means a cheaper and timelier solution, instead of requiring updated LiDAR data.
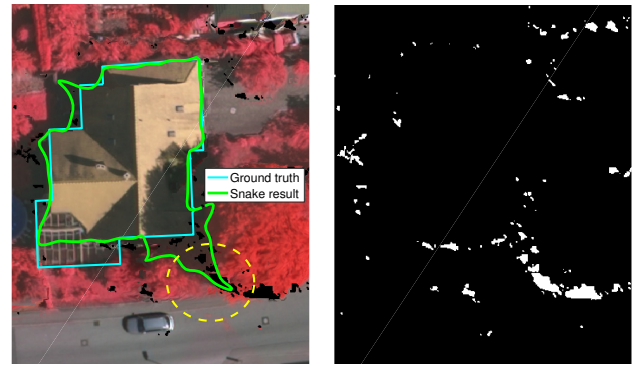
(a) Building ground truth (cyan) and extracted boundary (green).

(b) Binary mask denoting null-valued pixels (white: null, black: non-null).

Figure 9. Illustration of a wrongly extracted building due to null-valued image pixels.

However, like other snake model-based building extraction methods, our proposed method is also still dependent to snake parametrization in order to work well on complex areas. Future work will investigate on an automated contextualization of scene (either residential, industrial, or mixed) allowing to automatically parametrize the snake model. Nevertheless, more efforts will also be put on increasing effectiveness of the snake model on complex environments, such as the areas from the ISPRS benchmark data set. As a matter of fact, the results obtained on Vaihingen data set, even though being relevant compared to other methods, can still be improved in order to reach performances similar to those of Quebec City data set, i.e. average IoU values respectively of 83.23% versus 91.12%. There are two reasons that explain this difference. The process of building boundary extraction from the LiDAR point cloud of Vaihingen data set is less effective, due to the misclassification problem of vegetation and buildings. LiDAR points from Quebec City data set have a class value which facilitate the discrimination of building regions from trees. The second reason is the problem of null-valued regions in the orthophoto as mentioned above. Future works will also concentrate on these two issues.

5. ACKNOWLEDGMENT

REFERENCES

Aguilar, W., Frauel, Y., Escolano, F., Martinez-Perez, M. E., Espinosa-Romero, A., Lozano, M. A., 2009. A robust graph transformation matching for non-rigid registration. *Image and Vision Computing*, 27(7), 897–910.

Ahmadi, S., Zoej, M. V., Ebadi, H., Moghaddam, H. A., Mohammadzadeh, A., 2010. Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours. *International Journal of Applied Earth Observation and Geoinformation*, 12(3), 150–157.

Awrangjeb, M., Lu, G., Fraser, C., 2014. Automatic building extraction from LiDAR data covering complex urban scenes. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(3), 25.

Awrangjeb, M., Ravanbakhsh, M., Fraser, C. S., 2010. Automatic detection of residential buildings using LIDAR data and multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(5), 457–467.

Cohen, L. D., 1991. On active contour models and balloons. *CVGIP: Image understanding*, 53(2), 211–218.

Courant, R., Hilbert, D., 2008. *Methods of Mathematical Physics: Partial Differential Equations*. John Wiley & Sons.

Cramer, M., 2010. The DGPF-test on digital airborne camera evaluation–overview and test design. *Photogrammetrie-Fernerkundung-Geoinformation*, 2010(2), 73–82.

Douglas, D. H., Peucker, T. K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2), 112–122.

Dutter, M., 2007. Generalization of building footprints derived from high resolution remote sensing data.

Fazan, A. J., Dal Poz, A. P., 2013. Rectilinear building roof contour extraction based on snakes and dynamic programming. *International Journal of Applied Earth Observation and Geoinformation*, 25, 1–10.

Freeman, H., Shapira, R., 1975. Determining the minimum-area encasing rectangle for an arbitrary closed curve. *Communications of the ACM*, 18(7), 409–413.

Gilani, S., Awrangjeb, M., Lu, G., 2016. An automatic building extraction and regularisation technique using lidar point cloud data and orthoimage. *Remote Sensing*, 8(3), 258.

Guo, T., Yasuoka, Y., 2002. Snake-based approach for building extraction from high-resolution satellite images and height data in urban areas. *Proceedings of the 23rd Asian Conference on Remote Sensing*, 25–29.

Hartley, R., Zisserman, A., 2003. *Multiple view geometry in computer vision*. Cambridge university press.

Jaccard, P., 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat*, 37, 547–579.

Kabolizade, M., Ebadi, H., Ahmadi, S., 2010. An improved snake model for automatic extraction of buildings from urban aerial images and LiDAR data. *Computers, Environment and Urban Systems*, 34(5), 435–441.

Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active contour models. *International journal of computer vision*, 1(4), 321–331.

Khoshelham, K., Elberink, S. O., Xu, S., 2013. Segment-based classification of damaged building roofs in aerial laser scanning data. *IEEE geoscience and remote sensing letters*, 10(5), 1258–1262.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Polygonization of remote sensing classification maps by mesh approximation. *2017 IEEE International Conference on Image Processing (ICIP)*, IEEE, 560–564.

Nguyen, T. H., Daniel, S., Gueriot, D., Sintes, C., Le Caillec, J.-M., 2019. Robust Building-Based Registration of Airborne Lidar Data and Optical Imagery on Urban Scenes. *IGARSS 2019*, IEEE GRSS, Yokohama, Japan.

Peng, J., Zhang, D., Liu, Y., 2005. An improved snake model for building detection from urban aerial images. *Pattern Recognition Letters*, 26(5), 587–595.

Rottensteiner, F., Briese, C., 2003. Automatic generation of building models from LIDAR data and the integration of aerial images. *Proceedings of the ISPRS Working Group III/3 Workshop '3-D Reconstruction from Airborne Laserscanner and InSAR Data'*, 34(PART 3), 174-180.

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J. D., Breitkopf, U., Jung, J., 2014. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93, 256–271.

Rutzinger, M., Rottensteiner, F., Pfeifer, N., 2009. A comparison of evaluation techniques for building extraction from airborne laser scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2(1), 11–20.

Samsonov, T. E., Yakimova, O. P., 2017. Shape-adaptive geometric simplification of heterogeneous line datasets. *International Journal of Geographical Information Science*, 31(8), 1485–1520.

Sohn, G., Dowman, I., 2007. Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(1), 43–63.

Xu, C., Prince, J. L., 1997. Gradient vector flow: A new external force for snakes. *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, IEEE, 66–71.

Yang, B., Xu, W., Dong, Z., 2013. Automated extraction of building outlines from airborne laser scanning point clouds. *IEEE Geoscience and Remote Sensing Letters*, 10(6), 1399–1403.

Zarea, A., Mohammadzadeh, A., 2015. A novel building and tree detection method from LiDAR data and aerial images. *IEEE Journal of selected topics in applied Earth Observations and Remote Sensing*, 9(5), 1864–1875.

Zhan, Q., Liang, Y., Wei, C., Xiao, Y., 2012. Ground object recognition using combined high resolution airborne images and DSM. *Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39, B3.

Zhang, J., Lin, X., 2017. Advances in fusion of optical imagery and LiDAR point cloud applied to photogrammetry and remote sensing. *International Journal of Image and Data Fusion*, 8(1), 1–31.

Zhang, J., Lin, X., Ning, X., 2013. SVM-based classification of segmented airborne LiDAR point clouds in urban areas. *Remote Sensing*, 5(8), 3749–3775.