Extracting polygonal building footprints from digital surface models: A fully-automatic global optimization framework

Mathieu Brédif, Olivier Tournaire, Bruno Vallet, Nicolas Champion

ISPRS Journal of Photogrammetry and Remote Sensing

Extracting polygonal building footprints from digital surface models: A fully-automatic global optimization framework

Mathieu Brédif [*], Olivier Tournaire [1], Bruno Vallet, Nicolas Champion

Université Paris Est, IGN, MATIS 73, avenue de Paris, 94165 Saint-Mandé, France

ARTICLE INFO

ABSTRACT

This paper presents a fully automatic framework to extract building footprints from a Digital Surface Model (DSM). The proposed approach may be decomposed in two steps, each of them relying on a global optimization solver. The first step aims to extract rectangular building footprints directly from the DSM using a Marked Point Process (MPP) of rectangles. We introduce an energy that prevents overlapping rectangles and aligns rectangle edges with DSM discontinuities. This energy is then embedded in a RJMCMC sampler coupled with a simulated annealing to find its global optimum. Then, the second step of our framework refines these extracted rectangles into polygonal building footprints. We first create an arrangement of line segments supporting the rectangle edges. The dual graph of this arrangement is then considered in a maximum flow optimization scheme to remove edges in the arrangement which do not correspond to building edges in the DSM. Finally, 3D results illustrate a fully automatic process to build a 3D city model from a DSM only.

© 2012 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.

1. Introduction

1.1. Context and related work

Building extraction and reconstruction from remotely sensed images has been a very motivating topic for both research and industry in the last two decades. Indeed, 3D building models are now embedded in many scientific or large public applications. We can mention, among many others, wave propagation models for telecommunication operators, virtual tourism or realistic urban environments simulation for video games.

However, to construct a volumetric representation of buildings, most approaches of the literature rely on the existence of two-dimensional building outlines (Brunn and Weidner, 1997; Lafarge et al., 2008). If cadastral maps provide these polygonal footprints, they are however not always available. In addition, they may be outdated, incomplete or more generally suffer from discrepancies with the remotely-sensed dataset (Pfeifer et al., 2007). Besides, as both the surfaces to be mapped and the resolution of the remotely-sensed datasets tend to increase, it becomes unrealistic to rely

on human operators to delineate building footprint manually. In this respect, automating the extraction of these polygonal building outlines is highly desirable.

Many methods tackling this issue have been presented in the literature. Some methods operate a segmentation and classification to extract building regions from input data (LiDAR, DSM and/or aerial images). This is the case of e.g. the two state-of-the-art methods (Rottensteiner et al., 2005; Rutzinger et al., 2006) presented and compared in (Pfeifer et al., 2007). If these two methods are shown to be efficient, they only output noisy building outlines that need to be vectorized and simplified as a post-process. Arefi et al. (2007) proposed two simplifications methods. One is based on a minimum bounding rectangle that is very efficient but assumes rectilinear buildings. The other is based on RANSAC-extracted line segments combined using a rule-based algorithm. This simplification may also be achieved using Hough transform and least squares adjustment (Guercke and Sester, 2011) or Douglas–Peucker line simplification, graph and RANSAC-based generalization followed by a least squares adjustment to enforce parallelism and orthogonality (Neidhart and Sester, 2008).

Among the methods that deliver simplified 2D building outlines directly , we can find (Frédéricque et al., 2008) that populates a database by first detecting building footprints with a geometric approach. Regions of interest are then extracted from a DSM and their skeletons are computed to obtain the main directions. Eventually, they are used to generate a set of rectangle hypotheses which is then filtered using an iterative algorithm to reconstruct building

* Corresponding author.
E-mail addresses: mathieu.bredif@ign.fr (M. Brédif), olivier.tournaire@cstb.fr (O. Tournaire), bruno.vallet@ign.fr (B. Vallet), nicolas.champion@ign.fr (N. Champion).
[1] Olivier Tournaire is now a member of the CSTB / MODEVE team (84, Avenue Jean-Jaurès, Marne-la-Valée). He is also a member of the Université Paris Est, Center for Visual Computing.

blocks. (Ortner et al., 2007) also extracts rectangular building footprints from a DSM. This approach is based on a stochastic optimization framework and consists in modeling an energy which defines how a rectangle fits a building in a DSM, and how rectangles are arranged. This method is all the more interesting as it has been applied successfully to photogrammetric DSM, similar to those used in this paper (Lafarge et al., 2008). We give more details about this method in Section 2, which proposes a simplification and optimization of this method as a first step to extract polygonal footprints (see Fig. 1).

1.2. Overview of our framework

None of the previous approaches is able to extract polygonal building footprints with a fully automatic framework and in a global optimization scheme. To do so, we have chosen to work on DSM which are rasterized height maps (Fig. 2). They can be directly obtained from the 3D point cloud provided by LiDAR sensors, or with surface reconstruction techniques using multiple images (Hirschmuller, 2008; Pierrot-Deseilligny and Paparoditis, 2006).

As illustrated in Fig. 2, our strategy relies on two distinct steps. The first one is presented in Section 2 and aims at extracting rectangular footprints in the DSM. This is performed using a marked point process of rectangles. We try to find the optimum configuration of rectangles in the DSM, *i.e.* the one which minimizes an energy fitting rectangle edges on DSM height discontinuities and penalizing overlapping objects. In Section 3, which is our main contribution, we propose a new procedure to turn the previously extracted rectangles into polygonal footprints, based on a max-flow-optimized energy defined on the graph which is dual to an arrangement of line segments supporting the extracted rectangle edges. Section 4 presents some results and show how the extracted polygonal footprints may be used to automatically reconstruct 3D buildings models.

2. Extracting rectangular building footprints

The first step of our process consists in extracting rectangular building footprints from the DSM image. This section summarizes the global strategy proposed to reach this goal by Tournaire et al. (2010). Its implementation is publicly available in the open-source *librjmcmc* library (Brédif and Tournaire, 2012).

2.1. Marked point process of rectangles

A solution of the first global optimization step is a simple collection of rectangles. This allows to introduce the prior that buildings tend to be rectilinear and thus decomposable into non-overlapping unions of rectangles. Each rectangle is described by its center position c_i within the DSM image bounds $\mathcal{C} = [0, w] \times [0, h]$ and by additional geometric *marks* (Fig. 3): the aspect ratio $r_i = \frac{l_i}{L_i}$ and a vector $\overrightarrow{v_i}$ coding the semi-major axis, where L_i and l_i are respectively the

rectangle dimensions along and across $\overrightarrow{v_i}$ (i.e. $\|\overrightarrow{v_i}\| = \frac{l_i}{2}$). Using this original parameterization optimizes greatly many geometric computations (Tournaire et al., 2010). Finally, object parameters are sampled within the compact sets of the image bounds $\mathcal{C}$ and the rectangle marks $\mathcal{M}$:

$$(c_i, \overrightarrow{v_i}, r_i) \in \underbrace{[0, w] \times [0, h]}_{\mathcal{C}} \times \underbrace{\left[\frac{-w}{2}, \frac{w}{2}\right] \times \left[\frac{-h}{2}, \frac{h}{2}\right] \times [0, 1]}_{\mathcal{M}} \qquad (1)$$

A key issue is that the number itself of rectangles required to describe a scene is unknown which leads to using a Marked Point Process (MPP) framework, which extends our search space over all collections of rectangles to a probability space. This framework is able to deal with spaces of varying and very high dimension and also allows the introduction of an energy (Section 2.2) in the probability space density, enabling its minimization (Section 2.3). Mathematical details can be found in (van Lieshout, 2000). MPP have recently shown their versatility to extract objects in images in various remote sensing and computer vision applications. They have for instance been used to extract road marking in high resolution aerial images (Tournaire and Paparoditis, 2009), linear structures such as roads or rivers in satellite images (Lacoste et al., 2005) or to reconstruct 3D building models (Lafarge et al., 2010). This proposed extraction of building footprints as a realization of a MPP of rectangles is a complete reformulation and optimization (Tournaire et al., 2010) of the earlier similar work of (Ortner et al., 2007).

2.2. Energy model

A collection X of rectangles $x_i \in \mathcal{C} \times \mathcal{M}$ is valued using the following energy:

$$\mathcal{U}(X) = \sum_{x_i \in X} \mathcal{U}_{\text{data}}(x_i) + \beta \sum_{\substack{x_i, x_j \in X \\ i \neq j}} \mathcal{U}_{\text{prior}}(x_i, x_j) \qquad (2)$$

where β tunes the importance of the data energy $\mathcal{U}_{\text{data}}$ versus the prior energy $\mathcal{U}_{\text{prior}}$ (see Tournaire et al. (2010) for a study of β's influence; $\beta = 10$ generally gives good results). Fig. 4 illustrates these energy terms on various configurations of objects.

2.2.1. Data term
The data attachment term aims at measuring the consistency of the extracted objects with respect to the image DSM and defines what an attractive object is. As we aim at extracting building footprints from DSM, i.e. rasterized height data, we are naturally looking for altimetric discontinuities. The best rectangle edge candidates should thus be well positioned and oriented on a high vertical discontinuity. This is measured using an estimated façade area covered by each edge $e_i^j = \left[p_i^j, p_i^{j+1}\right]$ of a rectangle x_i:
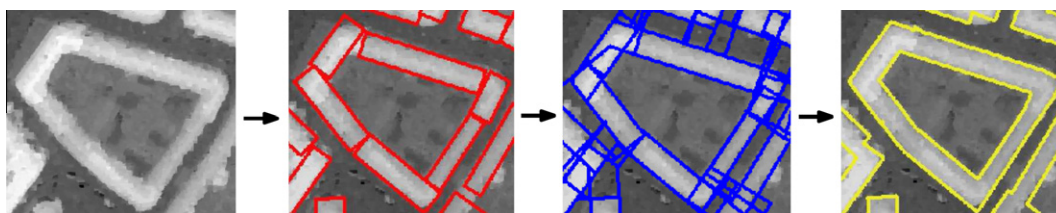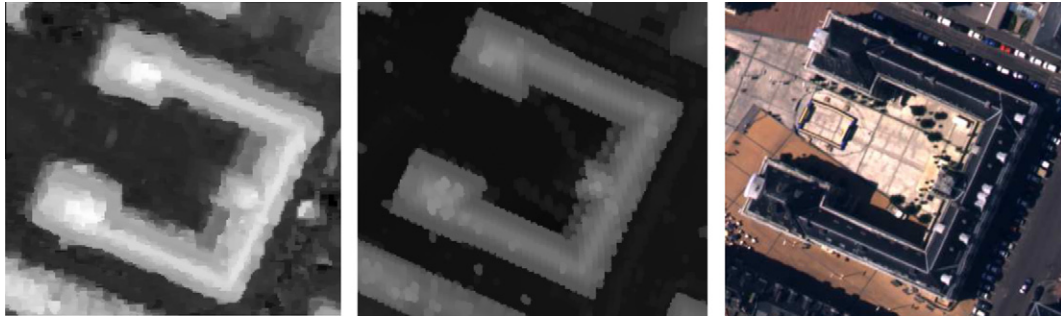

Fig. 1. Overview of our framework. From left to right: a multiview based *photogrammetric DSM*; The rectangles detected using a *marked point process*; *Arrangement of line-segments* supporting the rectangle edges; *Polygonal footprints* obtained by a max-flow optimization on the dual graph of the arrangement.

Fig. 2. DSM obtained from multiple images (*left*) or from a 3D LiDAR point cloud (*middle*) on a urban area (*right*) (the brighter the pixel, the higher the height). Notice that a LiDAR based DSM is generally less noisy than a photogrammetric DSM and has sharper edges.
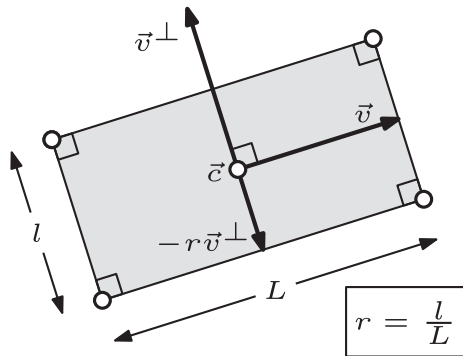

Fig. 3. The $(c, \vec{v}, r)$ parameterization of rectangular objects.

$$\mathcal{A}_{(c_i, \overrightarrow{v_i}, r_i) \in \underbrace{[0, w] \times [0, h]}_{C} \times \underbrace{\left[\frac{-w}{2}, \frac{w}{2}\right] \times \left[\frac{-h}{2}, \frac{h}{2}\right] \times [0, 1]}_{\mathcal{M}}} \left(e_i^j\right)$$

$$= \max\left(0, \overrightarrow{n_i^j} \cdot \int_0^1 \overrightarrow{\nabla z}\left(p_i^j + t \, \overrightarrow{e_i^j}\right)dt\right) \tag{3}$$

where $\overrightarrow{e_i^j} = \left(p_i^{j+1} - p_i^j\right)$ is the edge vector, $\overrightarrow{n_i^j} = \overrightarrow{e_i^j}^\perp$ its inner-pointing normal and $\overrightarrow{\nabla z}(p)$ is the DSM gradient magnitude at pixel p. The estimated façade area

$$\mathcal{A}_{(c_i, \overrightarrow{v_i}, r_i) \in \underbrace{[0, w] \times [0, h]}_{C} \times \underbrace{\left[\frac{-w}{2}, \frac{w}{2}\right] \times \left[\frac{-h}{2}, \frac{h}{2}\right] \times [0, 1]}_{\mathcal{M}}} \left(e_i^j\right) \text{ is}$$

thus the dot product of the edge normal $\overrightarrow{n_i^j}$ and the average gradient across the edge $\int \overrightarrow{\nabla z}$.

$$\mathcal{A}_{(c_i, \overrightarrow{v_i}, r_i) \in \underbrace{[0, w] \times [0, h]}_{C} \times \underbrace{\left[\frac{-w}{2}, \frac{w}{2}\right] \times \left[\frac{-h}{2}, \frac{h}{2}\right] \times [0, 1]}_{\mathcal{M}}} \text{ is}$$

clamped to nonnegative values to disable any penalization that would for instance prevent the extraction of a low-rise building next to a higher one as the height discontinuity would contribute negatively to the detection of the low-rise building on their common edge. Then, the data energy of a rectangle x_i combines the estimated façade areas of its four edges e_i^j:
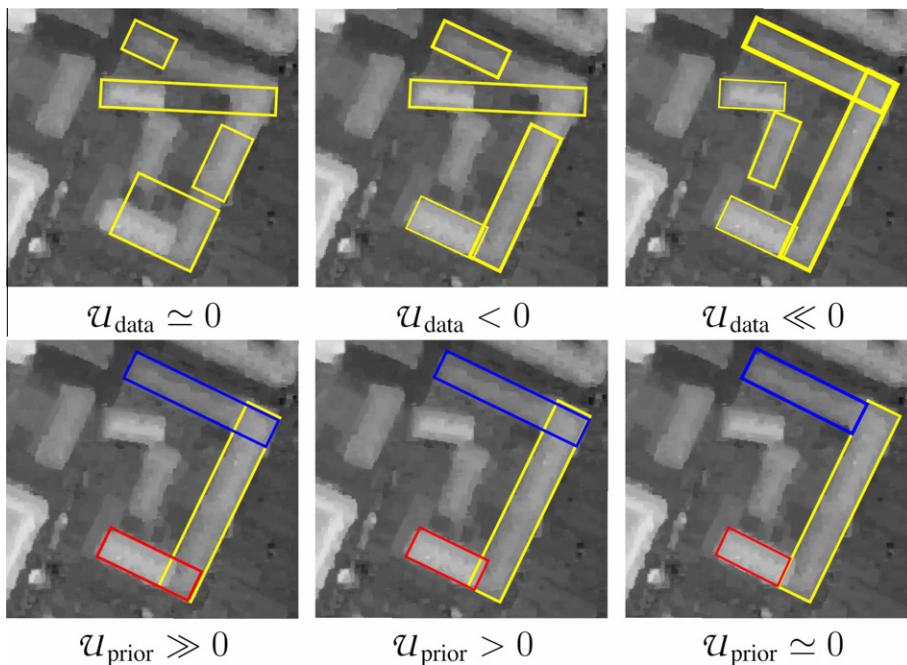
Fig. 4. Configurations with decreasing energy terms (from left to right), as the objects are increasingly better aligned with height discontinuities ($\mathcal{U}_{\text{data}}$), or less overlapping ($\mathcal{U}_{\text{prior}}$).
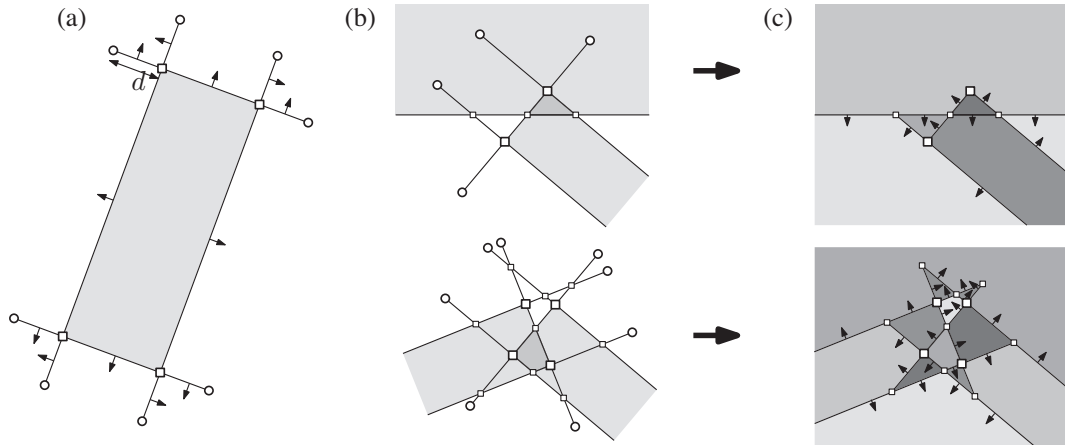
Fig. 5. (a) Rectangle with extended oriented line segments. (b) Arrangement close-ups on interacting rectangle pairs. (c) Cleaned-up arrangement faces with edge orientations (black arrows).

Fig. 6. (a) MPP-extracted rectangles. (b) Extended line-segment arrangement. (c) Cleaned-up arrangement. (d) Graph-cut-extracted polygonal footprints.

$$\mathcal{U}_{\text{data}}(x_i) = \mathcal{A}_0 - \sum_{j=0}^{3} \mathcal{A}$$

$$(c_i, \overrightarrow{v_i}, r_i) \in \underbrace{[0, w] \times [0, h]}_{\mathcal{C}} \times \underbrace{\left[\frac{-w}{2}, \frac{w}{2}\right] \times \left[\frac{-h}{2}, \frac{h}{2}\right] \times [0, 1]}_{\mathcal{M}} \left(e_i^i\right) \quad (4)$$
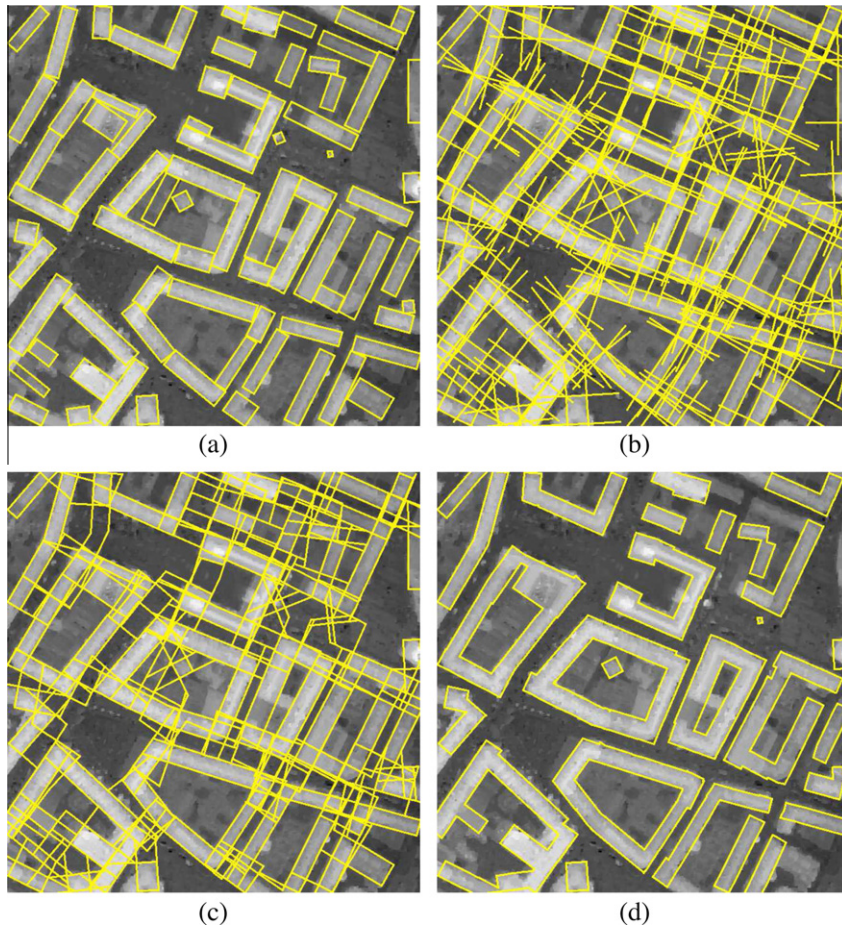
where the constant $\mathcal{A}_0$ can be considered as the minimal façade surface required for a building to be detected, and thus acts as a generalization parameter. Since this parameter has a physical meaning and can be expressed in squared meters, its tuning does not depend on the DSM resolution. In our experiments, we used $\mathcal{A}_0 = 100$ m².

2.2.2. Prior term

The proposed prior term is repulsive as it only penalizes overlapping objects and multiple detection, using their intersection area.
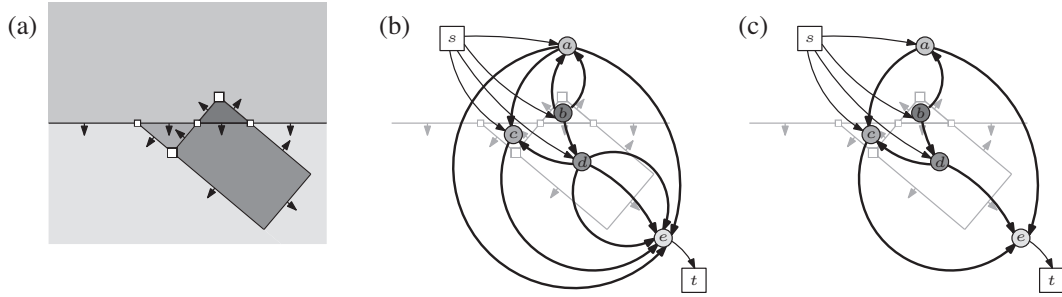
Fig. 7. (a) Input arrangement. (b) Corresponding graph with source node *s*, sink node *t* and regular nodes *a*, *b*, *c*, *d*, *e*. Each oriented arrangement edge *g* yields an oriented graph edge between regular nodes (bold), weighted by $\mathcal{U}_{edge}(g, 1, 0)$. (c) Multiple edges linking identical nodes are merged by summing up their weights. They may correspond to contiguous arrangement edges (e.g. 3 edges from *d* to *e*) or discontinuous arrangement edges (e.g. 2 edges from *a* to *e*). The ∞-weighted edges (not depicted here for readability) are reverse edges of the bold edges.

$$\mathcal{U}_{prior}(x_i, x_j) = \mathcal{A}_{intersection}(x_i, x_j) \tag{5}$$

This simple energy proved to be sufficient as the attractivity needed to combine close rectangles into building blocks is provided by the data term, which tries to fit rectangle edges to every meaningful height discontinuity present in the DSM.

2.3. RJMCMC optimization

Let us now consider a marked point process X_t of rectangles defined through a (unnormalized) probability density f with respect to the law $\pi_v(\cdot)$ of a reference Poisson process:

$$f(X_t) = e^{-\mathcal{U}(X_t)} \tag{6}$$

Extracting building footprints consists in finding the configuration of objects $\widetilde{X}_t$ minimizing the energy $\mathcal{U}$, thus maximizing the posterior probability: $\widetilde{X}_t = argmax\, f(\cdot)$. This is where the RJMCMC sampler and the simulated annealing take place.

Classical MCMC methods such as Metropolis–Hastings cannot handle dimension jumps, i.e., changes in dimension between samples. That is why we used a RJMCMC sampler, which consists in simulating a Markov Chain on the configuration space: a transition from one step of the chain to another is guided through a set of proposition kernels (Green, 1995). Each kernel randomly proposes a new configuration based on the current configuration, by modifying an already existing object, or by adding/removing an object (thus changing the dimension space).

The RJMCMC sampler is coupled with a simulated annealing (Salamon et al., 2002) in order to find the optimum of the density $f(\cdot)$. Instead of $f(\cdot)$, we use in the optimization process $f(\cdot)^{\frac{1}{T_t}}$, where T_t is a sequence of decreasing temperatures which tends to zero as t tends to infinity. Theoretically, convergence is guaranteed whatever the initial configuration X_0 is if the decrease of the temperature follows a logarithmic scheme. In practice it is impossible to use such a scheme since it is too slow. Thus, it is generally replaced with a geometric decrease which gives a good solution close to the optimal one.

3. Extracting polygonal building footprints

Once a meaningful configuration of rectangles has been extracted, a subsequent refinement step is required to retrieve nonoverlapping polygonal building footprints comparable to cadastral maps. First a normalized Digital Surface Model (nDSM) is derived from the DSM using the extracted rectangles (Section 3.1). We then formulate the polygonal building footprint extraction as a global optimization problem. Sections 3.2 and 3.3 define its search space as a selection of faces of an arrangement (Halperin, 2004) of line segments. Then Section 3.4 introduces the scoring

function as a refinement of the energy used during the MPP extraction step. The best polygonal footprints are finally extracted using a global optimization based on maximum flow (Section 3.5).

3.1. Rectangle-guided DTM and nDSM derivation

The polygonal footprint extraction step requires building heights above the ground (Pfeifer et al., 2007), that are encoded by the normalized Digital Surface Model (nDSM). To produce this nDSM, we first derive an off-ground mask from the rectangles detected during the MPP step. This mask is then used to guide the derivation of a Digital Terrain Model (encoding the height of bare earth) from the DSM, with the surface-based approach proposed in (Champion et al., 2009). In the end, the nDSM is then obtained by subtracting the DTM from the DSM.

3.2. Extending rectangular line segments

The rectangles extracted by the marked point process faithfully retrieve the supporting lines of the façades but their extent is less accurately estimated, due to the rectangular and nonoverlapping priors. Therefore, it seems natural to extend the line segments bounding the MPP-extracted rectangular footprints by some distance *d* to construct the set of line segment hypotheses supporting the polygonal building footprints (5a and 6b). *d* tunes the confidence on the localization of the line segments extracted by the MPP step (*d* = 0: complete confidence, $d \to \infty$: endpoints are discarded so that only the supporting lines of the extracted line segments are considered). As *d* can be expressed in meters, its value can easily be set independently of the nDSM resolution ($d \in [5; 10]$ was used in our experiments). Note that the parallelism and orthogonality priors introduced by the rectangular extraction step are preserved, as line segment orientations are not modified.

3.3. Selecting arrangement faces

The arrangement of a set of line segments is a well-known computational geometry data structure (Halperin, 2004) with readily available infinite precision implementations.[2] It refers to the subdivision of the plane into vertices (the endpoints of the line segments and their intersections), edges (a sub-segment of an extended line-segment between two arrangement vertices) and faces (polygons, possibly with holes and/or unbounded, delimited by arrangement edges). The sets of vertices, edges and faces of an arrangement are respectively denoted *V*, *E* and *F*. The faces in *F* define, by construction, a partition of the plane (Fig. 5b and 6b): they only meet at their boundaries and their union is $\mathbb{R}^2$.

---

[2] Computational Geometry Algorithms Library, http://www.cgal.org.

Whereas previous methods (Lafarge et al., 2010) to derive polygonal footprints from rectangle layouts relied on a complex case analysis that was only able to consider interactions between a single pair of rectangles at a time, the proposed approach is based on the arrangement of the extended rectangular line segments to be both robust and generic. First, all the geometric robustness concerns are segregated into the computation of the extended line segment arrangement. Second, its genericity is implied by the arrangement framework, which does not limit the considered geometric cases or even the number of simultaneously interacting rectangles.

To simplify subsequent processing, the arrangement of the extended rectangular line segments is first cleaned up (Fig. 5c and 6c) by discarding edges that do not separate distinct facets (such as the antennae of Fig. 5b), as well as the resulting isolated vertices. We then propose to extract the polygonal building footprints by selecting a subset S of the face set F. The union of the selected faces indeed represents a set of polygonal shapes (Fig. 6d). For instance, selecting all the faces inside at least one extracted rectangle results in the polygonal shape of the union of the extracted rectangles.

Finally, line segments are oriented, as they describe a height discontinuity oriented outward the extracted rectangle (Fig. 5a and 5c). This orientation information is kept as an attribute for each edge of the arrangement, so that an edge may only be used to model a façade with the detected orientation. This orientation defines for each arrangement edge e, oriented from face f to face g (denoted $f \xrightarrow{e} g$) its source f and target g.

3.4. Energy model

A selection S of arrangement faces is valued according to the following energy, where the λ parameter balances the effects of a unary energy $\mathcal{U}_1^S(f)$ summed over all faces and a binary energy $\mathcal{U}_2^S(f, g)$ summed over all face pairs:

$$\mathcal{U}(S) = \sum_{f \in F} \mathcal{U}_1^S(f) + \lambda \sum_{f,g \in F,\, f \neq g} \mathcal{U}_2^S(f, g) \tag{7}$$

3.4.1. Unary energy

The union of all the MPP-extracted rectangles provides an initial classification of the nDSM pixels into building and non-building areas. From this preliminary segmentation, histograms of nDSM values of the building $h_1(z)$ and non-building $h_0(z)$ classes may be computed. These two histograms may then be combined into the probability $\mathbb{p}_1(z) = 1 - \mathbb{p}_0(z) = \frac{h_1(z)}{h_0(z) + h_1(z)}$ whether a given height z belongs to a building or not. This probability is finally turned into a Gibbs energy $\mathcal{U}_1^S$, assuming the independence of the height values z_p of the nDSM pixels p within an arrangement face f:

$$\mathcal{U}_1^S(f) = \mathcal{U}_{\text{face}}(f, \chi_S(f))$$
$$\text{with } \mathcal{U}_{\text{face}}(f, i) = \sum_{(x,y) \in f} -\log(\mathbb{p}_i(z_p))$$

where $\chi_S(f) = 1$ if $f \in S$ and 0 otherwise. This unary energy learns the building and non-building heights from the preliminary segmentation given by the extracted rectangular building approximations.

3.4.2. Binary energy

$\mathcal{U}_2^S(f, g)$ has two goals: ensuring the orientation constraint (Section 3.3), and placing the interface between selected *building* faces and deselected *non-building* faces at height discontinuities. $\mathcal{U}_2^S(f, g)$ accumulates the contributions of the possibly multiple edges of the arrangement from f to g:

$$\mathcal{U}_2^S(f, g) = \sum_{e \in E\, f \to g} \mathcal{U}_{\text{edge}}(e, \chi_S(f), \chi_S(g)) \tag{8}$$

$$\begin{cases} \mathcal{U}_{\text{edge}}(e, 0, 0) = \mathcal{U}_{\text{edge}}(e, 1, 1) = 0 \\ \mathcal{U}_{\text{edge}}(e, 0, 1) = +\infty \\ \qquad \mathcal{U}_{\text{edge}}(e, 1, 0) = \max(0, H|e| - \mathcal{A} \\ (c_i, \overrightarrow{v_i}, r_i) \in \underbrace{[0, w] \times [0, h]}_{\mathcal{C}} \times \underbrace{\left[\frac{-w}{2}, \frac{w}{2}\right] \times \left[\frac{-h}{2}, \frac{h}{2}\right] \times [0, 1]}_{\mathcal{M}}(e)) \end{cases} \tag{9}$$

where H is a height constant, $|e|$ the length of e and $\mathcal{A}_{(c_i, \overrightarrow{v_i}, r_i) \in \underbrace{[0, w] \times [0, h]}_{\mathcal{C}} \times \underbrace{\left[\frac{-w}{2}, \frac{w}{2}\right] \times \left[\frac{-h}{2}, \frac{h}{2}\right] \times [0, 1]}_{\mathcal{M}}}(e)$ is the estimated façade area (Eq. (3)). Thus, $\mathcal{U}_{\text{edge}}$ forbids wrongly oriented façades using the $+\infty$ value, and penalizes edges with poor nDSM support if their mean height discontinuity is less than H (we usually set H to 3 m which is suitable in the most common cases).

3.5. Graph-cut optimization

The proposed energy is graph-representable (Kolmogorov and Zabih, 2004), it may thus be globally minimized using a maximum flow/graph cut optimization. The flow graph is dual to the line segment arrangement and is constructed as follows (Fig. 7):

- Each face of the extended line-segment arrangement is associated with a node, denoted regular by opposition to the auxiliary source and sink nodes.
- A pair of opposite oriented edges is added between the nodes representing adjacent facets of the arrangement.
- An oriented edge is added at each face node from the source or to the sink according to the sign of the unary energy of the face (Kolmogorov and Zabih, 2004).

The result of a graph-cut problem is a partition of its nodes into two sets, each containing an auxiliary node. Such a partition provides S by selecting all the faces of the arrangement represented by regular nodes of the graph partitioned in the same set as the source node. The polygonal footprint is then finally the union of the faces in S.

To represent the proposed energy $\mathcal{U}^S$ into this dual graph, each edge must be weighted by the cost of simultaneously selecting its source node in S and deselecting its target node as a non-building region in $F \backslash S$. Edge pairs of the graph between regular nodes correspond to an arrangement edge e. They are weighted by $\mathcal{U}_{\text{edge}}(e, 0, 1) = +\infty$, modeling the orientation constraint (Section 3.3), embedded into the graph using an arbitrarily high constant weight K, and $\mathcal{U}_{\text{edge}}(e, 1, 0)$, ensuring façades are placed at height discontinuities. An edge between an auxiliary node (source or sink) and a regular node f encodes a term of the unary energy $\mathcal{U}_1^S(f)$. Such an edge is weighted by $\Delta \mathcal{U}_1(f)$:

$$\Delta \mathcal{U}_1(f) = |\mathcal{U}_{\text{face}}(f, 1) - \mathcal{U}_{\text{face}}(f, 0)| = \left| \sum_{(x,y) \in f} \log \frac{h_1(z_p)}{h_0(z_p)} \right| \tag{10}$$

Fig. 8 illustrates the unary energy $\mathcal{U}_1^S$ on the faces of the partition of Fig. 6c. Fig. 8a shows $\mathcal{U}_1^S$ directly while Fig. 8b is normalized by the face area, which illustrates that faces are pixelwise well classified even on oversegmented facets (top-right region).

Note that it would have been simpler and more intuitive to define $\mathcal{U}_{\text{edge}}(e, 1, 0) = -\mathcal{A}_{(c_i, \overrightarrow{v_i}, r_i) \in \underbrace{[0, w] \times [0, h]}_{\mathcal{C}} \times \underbrace{\left[\frac{-w}{2}, \frac{w}{2}\right] \times \left[\frac{-h}{2}, \frac{h}{2}\right] \times [0, 1]}_{\mathcal{M}}}(e)$, preventing the need for a parameter H, but the resulting energy would not have been graph-representable and thus not minimizable using graph-cuts (Kolmogorov and Zabih, 2004). The binary
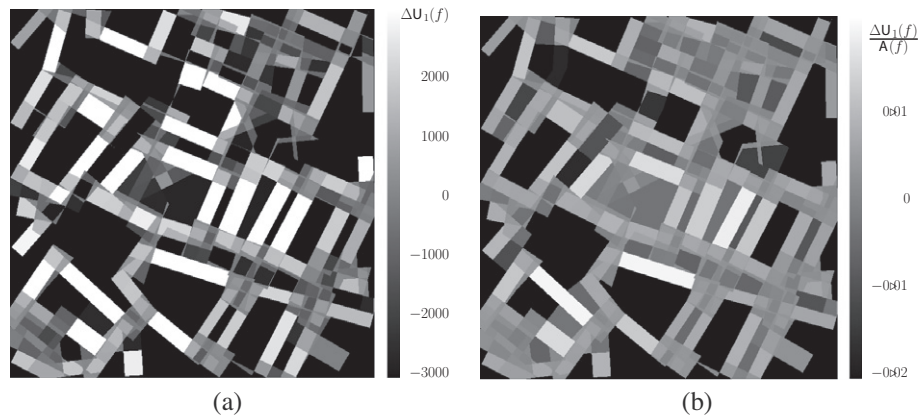
Fig. 8. Unary energy terms: each arrangement face f is uniformly filled with (a) its $\Delta\mathcal{U}_1(f)$ unary term and (b) the same term normalized by the face area $\frac{\Delta\mathcal{U}_1(f)}{\mathcal{A}(f)}$.

term $\mathcal{U}_2$ that fits polygonal boundaries to the oriented gradients alone is not sufficient by itself. Discarding the pixelwise unary term $\mathcal{U}_1$ that biases the polygonal segmentation towards the preliminary rectangle-based segmentation would however yield a trivial segmentation (all building or all non-building). This also grounds from the required submodularity of the graphcut-minimized energy (Kolmogorov and Zabih, 2004).

4. Results

4.1. Qualitative analysis

Our framework was applied on both photogrammetric (Photo) and Lidar DSMs with Ground Sampling Distances (GSDs) of 10 cm and 50 cm, with the following number of objects and computation times for the two main steps:

Input DSM			MPP		Graph cut	
Type (fig)	GSD (cm)	Image size	♯ Rect.	Time (s)	♯ Nodes	Time (s)
Photo (6)	50	650 × 650	76	177	1045	5
Photo (9)	10	800 × 1400	31	104	147	5
Lidar (10)	50	650 × 650	109	179	2801	7
Photo (11)	50	3634 × 2502	426	480	12954	352

There is still room for optimization as the 352 s of graph-cut processing in Fig. 11 are not due to the graph-cut itself (10 s), but to the elevation histograms computation (128 s) and to the edge and node weight evaluations (214 s).
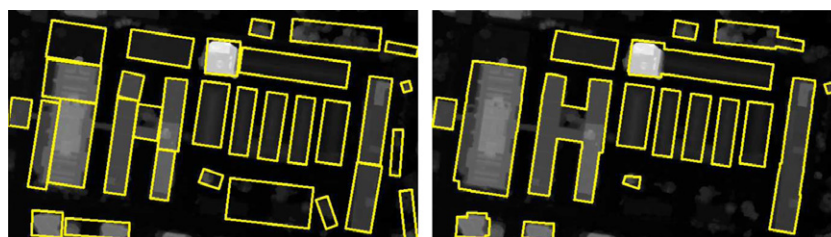
On photogrammetric DSMs, the MPP-extracted rectangles delineate most of the buildings (Figs. 6 and 11). Undetected buildings have low height and are mainly located in inner courtyards.

This underdetection is due to our parameterization since $\mathcal{A}_0$ was tuned high to prevent false positives and get an uncluttered result. The second step to reconstruct polygonal footprints gives good results: all building blocks composed of several rectangles are now given by only one polygon (possibly with holes), and complex building shapes are clearly recovered (see large buildings in the middle left of Fig. 11). The main problem is that low height buildings are not properly segmented during the graph-cut step: they are not easily distinguishable from ground regions in the elevation histograms. Fig. 9 presents the MPP rectangle extraction and polygonal building footprints graph-cut optimization on a 10 cm GSD photogrammetric DSM. At this high resolution, all building parts are correctly detected, but some rectangles are located on trees. These false detections may easily be avoided by introducing a vegetation mask. The interesting point is that the graph cut optimization to obtain polygonal footprints is well adapted to this resolution without tuning the parameters again. Moreover, some false detection on trees are also filtered as their supporting arrangement edges exhibit too weak a height discontinuity.

On the Lidar DSM (Fig. 10), the extraction results are equivalent to the ones obtained with a photogrammetric DSM on the same area (Fig. 6): similar discontinuities have been detected and thus polygonal footprints are extracted similarly. Some buildings are however better delineated, due to sharper discontinuities in the Lidar DSM (*e.g.* the 9-shaped building in the center of the image).

4.2. Quantitative analysis

A quantitative analysis has been performed to assess the quality of the proposed approach. Reference polygonal footprints have been manually extracted and compared to the output of our automatic process (Fig. 11c). It follows a pixel-based comparison of the two resulting building/non-building binary images. Each pixel is then classified as either True Positive (TP) for successfully extracted building areas, False Positive (FP) for overdetected

Fig. 9. Results on a photogrammetric DSM, 10 cm GSD: extracted rectangles (left) and extracted polygons (right).
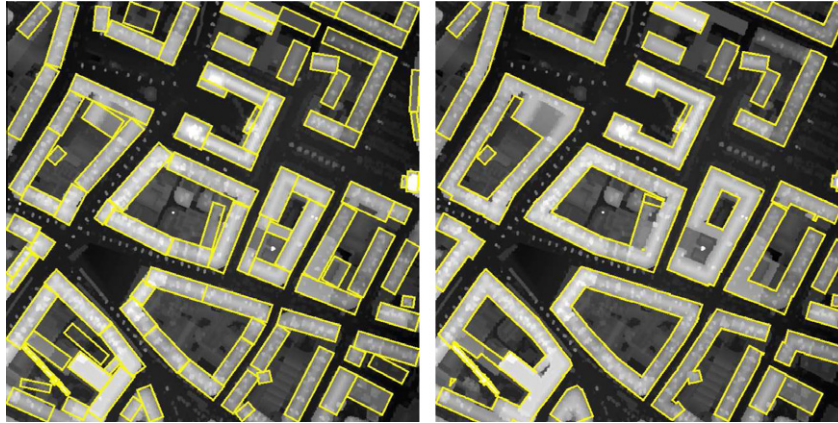
Fig. 10. 50 cm GSD Lidar DSM on the same area as Fig. 6: extracted rectangles (left) and extracted polygons (right).
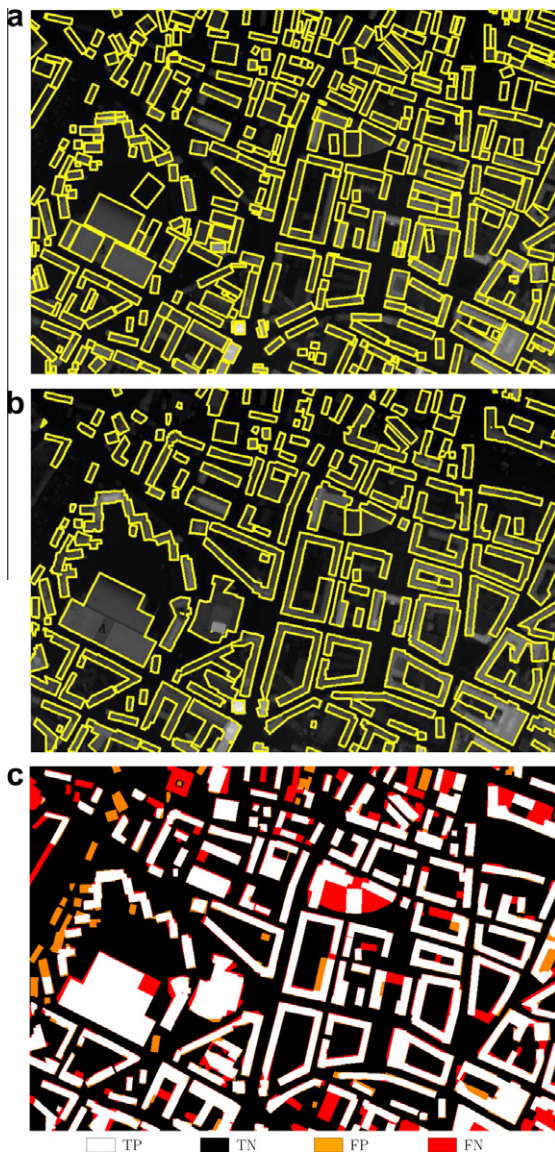

Fig. 11. Results on a 50 cm GSD Photogrammetric DSM (containing Fig. 6) proving the scalability of the proposed approach. From top to bottom: (a) MPP rectangle extraction, (b) graph-cut optimization and (c) pixel-based evaluation.
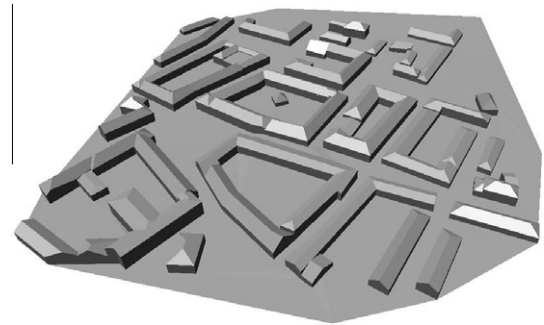

Fig. 12. Automatic 3D reconstruction of the area of Fig. 6.

buildings (false alarms), False Negative (FN) for underdetected buildings or True Negative (TN) for successfully extracted non-building areas. This allows us to introduce the two quality measures of *completeness* and *correctness*:

$$\text{Completeness} = \frac{\text{TP}}{\text{TP} + \text{FN}} = 80.3\%$$

$$\text{Correctness} = \frac{\text{TP}}{\text{TP} + \text{FP}} = 90.5\%$$

Although these completeness/correctness values are dataset dependent, they are in line and compare favorably with the values of 85%/73% and 85%/63% obtained in Pfeifer et al. (2007) for two state-of-the-art methods Rottensteiner et al. (2005), Rutzinger et al. (2006).

To further analyze the causes of the errors, FN and FP pixels have been manually subclassified for the following discussion. Most of FN errors (54.9%) are due to completely undetected buildings. This is mainly due to the low sensitivity of our method to the detection of low-rise buildings surrounded by higher buildings. Two other noticeable underdetections are due to the presence of a large curved mall that breaks our rectilinearity assumption and to the misreconstruction of a large glassroof in the photogrammetric DSM. Turning to FP errors, 40.9% are trees detected as buildings. Thus correctness could then easily be improved above 94% by using a vegetation mask. Finally, the geometric accuracy of correctly detected buildings accounts for 45.1% of FN errors and 46.1% of FP errors. Dividing the number of the corresponding FN and FP pixels by the total façade length of the reference yields an estimate of the localization performance of our method of 1.85

m. This complies with the expected precision of the DSM and of the reference dataset ($\sim$1 m).

5. Conclusion and future works

This paper proposes a succession of two global optimization techniques to successfully extract polygonal building outlines from a raw DSM only. These polygonal footprints may be used to fully automatically reconstruct 3D building models (Durupt and Taillandier, 2006; Vallet et al., 2009) (Fig. 12). Thus, we demonstrate a scalable and fully-automatic process that converts robustly raw DSMs into 3D city models at extremely low costs.

The two optimization procedures proposed in this paper are fully automatic and only involve a few easily tunable parameters ($\beta, \mathcal{A}_0, d, \lambda, H$) that feature intuitive physical interpretations. In our experiments, their value was set using prior knowledge on the scene. The MPP-based optimization relying on a RJMCMC and simulated annealing framework (Tournaire et al., 2010) has been extended with a novel graph-cut-based global optimization scheme. This second step relaxes the rectilinear constraints by casting its minimization problem as a classification problem over a partition of the DSM induced by the extracted rectangles. The main advantages of this formulation lies in its simplicity, its robustness and its geometric genericity.

For future work, artefacts due to the two successive steps of global optimization (such as missing low-rise buildings) should be reduced by performing a single step of global optimization. For instance, a RJMCMC-based optimization performed directly over the set of piecewise linear partitions of the input DSM is a promising idea. Unifying the energy terms between the two successive MPP and graphcut-based steps is subject to future work but it is constrained by the submodularity requirement of the graphcut formulation and the desired independence of the energy terms to the extended line segment oversegmentation. Automatic parameter estimation is also a very important task. The work presented in (Ben Hadj et al., 2010) can be of great interest. Last but not least, the proposed energies have been tuned for simplicity and operational efficiency. Future work will explore alternative energies taylored to more specific problems, such as energies that introduce alignment priors. Finally the result may exhibit topological clutter (*e.g.* saw-tooth patterns) that may require simplification or generalization. A final regularization step could be added to balance the topological complexity of the extracted outlines with their geometric accuracy (*i.e.* the data alignment of facades).

References

Arefi, H., Engels, J., Hahn, M., Mayer, H., 2007. Approximation of building boundaries. In: Coors, V., Rumor, M., Fendel, E., Zlatanova, S. (Eds.), Proc. of the 26th Urban Data Management Symposium, Stuttgart, 10–12 October. UDMS, pp. 25–33.

Ben Hadj, S., Chatelain, F., Descombes, X., Zerubia, J., 2010. Parameter estimation for a marked point process within a framework of multidimensional shape extraction from remote sensing images. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (Part 3A), 1–6.

Brédif, M., Tournaire, O., 2012. librjmmc: an open-source generic C++ library for stochastic optimization. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (3B), 259–264, <http://librjmcmc.ign.fr> (accessed 28.12.12).

Brunn, A., Weidner, U., 1997. Extracting buildings from digital surface models. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 32 (3-4 W2), 27–34.

Champion, N., Boldo, D., Pierrot-Deseilligny, M., Stamon, G., 2009. Automatic estimation of fine terrain models from multiple high resolution images. In: IEEE International Conference on Image Processing (ICIP), Cairo, 7–10 November. pp. 577–580.

Durupt, M., Taillandier, F., 2006. Automatic building reconstruction from a digital elevation model and cadastral data: an operational approach. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (3) (On CD–ROM).

Frédéricque, B., Daniel, S., Bédard, Y., Paparoditis, N., 2008. Populating a building multi-representation data base with photogrammetric tools: recent progress. ISPRS Journal of Photogrammetry and Remote Sensing 64 (4), 441–460.

Green, P.J., 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. Biometrika 82 (4), 711–732.

Guercke, R., Sester, M., 2011. Building footprint simplification based on Hough transform and least squares adjustment. In: 25th International Cartographic Conference (ICC'11) Workshop on Generalization and Multiple Representation, 3–8 July, 2011. Paris (On CD–ROM).

Halperin, D., 2004. Arrangements. In: Goodman, J.E., O'Rourke, J. (Eds.), Handbook of Discrete and Computational Geometry. CRC Press LLC, Boca Raton, FL, pp. 529–562.

Hirschmuller, H., 2008. Accurate and efficient stereo processing by semi-global matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (2), 328–341.

Kolmogorov, V., Zabih, R., 2004. What energy functions can be minimized via graph cuts? IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (2), 147–159.

Lacoste, C., Descombes, X., Zerubia, J., 2005. Point processes for unsupervised line network extraction in remote sensing. IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (10), 1568–1579.

Lafarge, F., Descombes, X., Zerubia, J., Pierrot-Deseilligny, M., 2008. Building reconstruction from a single DEM. In: Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24–26, June 2008. Anchorage, AK, pp. 1–8.

Lafarge, F., Descombes, X., Zerubia, J., Pierrot-Deseilligny, M., 2010. Structural approach for building reconstruction from a single DSM. IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (1), 135–147.

Neidhart, H., Sester, M., 2008. Extraction of building ground plans from lidar data. The international archives of the photogrammetry. Remote Sensing and Spatial Information Sciences 37 (Part B2), 405–410.

Ortner, M., Descombes, X., Zerubia, J., 2007. Building outline extraction from digital elevation models using marked point processes. International Journal of Computer Vision 72 (2), 107–132.

Pfeifer, N., Rutzinger, M., Rottensteiner, F., Muecke, W., Hollaus, M., 2007. Extraction of building footprints from airborne laser scanning: comparison and validation techniques. In: Urban Remote Sensing Joint Event. IEEE/ISPRS, Paris, pp. 1–7.

Pierrot-Deseilligny, M., Paparoditis, N., 2006. A multiresolution and optimization-based image matching approach: an application to surface reconstruction from SPOT5-HRS stereo imagery. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 1/W41), On CD–ROM.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2005. Using the Dempster-Shafer method for the fusion of LiDAR data and multi-spectral images for building detection. Information Fusion 6 (4), 283–300.

Rutzinger, M., Höfle, B., Geist, T., Stötter, J., 2006. Object-based building detection based on airborne laser scanning data wit-hin GRASS GIS environment. In: Proc. 25th Urban Data Management Symposium, Aalborg, 15–17 May. UDMS, pp. 37–48.

Salamon, P., Sibani, P., Frost, R., 2002. Facts, Conjectures and Improvements for Simulated Annealing. Society for Industrial and Applied Mathematics, Philadelphia, USA.

Tournaire, O., Brédif, M., Boldo, D., Durupt, M., 2010. An efficient stochastic approach for building footprint extraction from digital elevation models. ISPRS Journal of Photogrammetry and Remote Sensing 65 (4), 317–327.

Tournaire, O., Paparoditis, N., 2009. A geometric stochastic approach based on marked point processes for road marks detection from high resolution aerial images. ISPRS Journal of Photogrammetry and Remote Sensing 64 (6), 621–631.

Vallet, B., Pierrot-Deseilligny, M., Boldo, D., 2009. Building footprint database improvement for 3D reconstruction: a direction aware split and merge approach. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 3A), 139–144.

van Lieshout, M.N.M., 2000. Markov Point Processes and Their Applications. Imperial College Press, London.