Structural approach for building reconstruction from a single DSM

Florent Lafarge, Xavier Descombes, Josiane Zerubia, Marc Pierrot-Deseilligny

Structural Approach for Building Reconstruction from a Single DSM

Florent Lafarge, Xavier Descombes, Josiane Zerubia, and Marc Pierrot-Deseilligny

Abstract—We present a new approach for building reconstruction from a single Digital Surface Model (DSM). It treats buildings as an assemblage of simple urban structures extracted from a library of 3D parametric blocks (like a LEGO® set). First, the 2D-supports of the urban structures are extracted either interactively or automatically. Then, 3D-blocks are placed on the 2D-supports using a Gibbs model which controls both the block assemblage and the fitting to data. A Bayesian decision finds the optimal configuration of 3D-blocks using a Markov Chain Monte Carlo sampler associated with original proposition kernels. This method has been validated on multiple data set in a wide resolution interval such as 0.7 m satellite and 0.1 m aerial DSMs, and provides 3D representations on complex buildings and dense urban areas with various levels of detail.

Index Terms—3D reconstruction, urban area, Digital Surface Model, Stochastic models, Monte Carlo simulations.

❖

1 INTRODUCTION

Three dimensional models of urban areas are very useful for many applications such as urban planning, radiowave reachability tests for wireless communications, disaster recovery or computer games. A standard technique for creating 3D city models in an automated or semi-automated way consists in applying stereo vision techniques on aerial or satellite imagery.

Problem statement - Many methods have been proposed. It is difficult to compare these methods efficiently since they have been developed in different contexts (kinds of data, types of reconstructed buildings, level of user interactivity, *etc*) and use different evaluation criteria [1][2].
There are two main families of approaches in 3D building reconstruction. Generic representations are theoretically able to reconstruct any shape of building through connected planar facets, but they demand high resolution data and often require very high computing times. Scholze *et al.* extract 3D-lines and group them into faces which allow the building reconstruction through a semantic interpretation [3]. Rooftop hypotheses are generated from 3D-lines and junction information by Kim *et al.*[4]. Baillard *et al.* present a method based on planar facet hypothesis which can be generated from single 3D-lines [5]. Taillandier *et al.*[6] combine several kinds of primitives such as 3D-lines, planes and facade hypothesis. These methods provide accurate descriptions of common buildings. However, they are not adapted to model large urban scenes and complex buildings owning many facets since they have combinatorial problems. Parametric representations are known to be robust with respect to data quality and adapted to large scenes [7], but these reconstructions are limited - most parametric representations consider a symmetric two-plane roof reconstruction. Collins *et al.* propose a complete interactive system restricted to the representation of flat roof buildings [8]. Brédif *et al.* present a method including the parametric roof superstructure reconstruction using a Minimum Description Length energy minimization [9]. Several systems based on the Constructive Solid Geometry (CSG) representation have also been proposed [10][11]. They combine single primitives by Boolean operations in order to obtain complete building models. One of the most efficient methods uses a hierarchical model which combines three different levels of detail [12]. Laser scans are also popular inputs for acquiring 3D city models thanks to the measurement accuracy [13][14]. Früch *et al.*[15] use laser scans to model buildings with a detailed reconstruction of the facades. Other works consist in constructing artificial urban scenes. For example, Müller *et al.* develop a procedural model based on a shape grammar [16].
These methods provide convincing 3D-models using aerial images, ground views or laser scans. Most of them have been developed using a specific kind of data and cannot easily be adapted if the image characteristics change (resolution, Signal to Noise Ratio, *etc*). Here, we propose a new method adapted to varying data resolution by using a single data - a Digital Surface Model (DSM) - as illustrated on Figure 1. DSMs describe the altimetry of an urban scene through a regular grid of points - a height is associated with each point of the grid.

- F. Lafarge, X. Descombes and J. Zerubia are with the Ariana Research Group, INRIA Sophia Antipolis, 2004 route des Lucioles 06902 Sophia Antipolis, France. E-mail: Florent.Lafarge@inria.fr, Xavier.Descombes@inria.fr, Josiane.Zerubia@inria.fr
- F. Lafarge and M. Pierrot-Deseilligny are with the Matis Laboratory, French Mapping Agency, 2/4 avenue Pasteur, 94165 Saint-Mandé, France. E-mail: Marc.Pierrot-Deseilihny@ign.fr

They are well adapted to global geometric descriptions of urban scenes. DSMs are mostly generated by using stereo vision techniques from multiview images. Then, this kind of data is often very noisy as we can see on Figure 1. Using DSMs as inputs of a 3D reconstruction process is very interesting since we do not need to match the multiview images during the process: it strongly reduces the combinatorial problems.
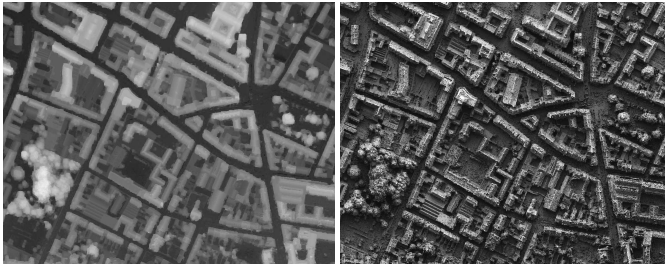

Fig. 1. Digital Surface Model (DSM) of an urban area - raster version *(left)*, and shaded version *(right)*.

Global strategy - We propose an approach based on a structural concept, which is halfway between generic and parametric representations. It consists in reconstructing buildings by assembling simple urban structures extracted from a library of 3D parametric blocks, as a LEGO®kit (see Figure 2). The block assemblage is controlled by stochastic Gibbs models. This concept, which uses a CSG representation, has been addressed in a previous work [17]. However, it was limited and suffered from many drawbacks: generation of many artefacts, lack of 3D-modeling generality, tuning of many parameters, long computation times and restriction to satellite data.
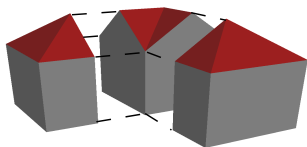

Fig. 2. Principle of the structural approach.

This approach is particularly interesting since it combines the advantages of generic and parametric representations:

1- the robustness of parametric approaches is preserved since the library objects are defined by parameter sets,

2- an efficient library permits modeling a large range of buildings. It is even possible to reconstruct buildings such as curved roof structures that some generic models cannot construct,

3- assembly of urban structures is particularly adapted to various data resolutions and allows us to obtain both basic roof shapes from 0.7 m resolution satellite images or details including superstructures (chimneys, dormer-windows,

etc) from 0.1 m resolution aerial images. Figure 3 shows the difference, in terms of level of detail, between 0.25 meter resolution aerial data and 0.7 meter resolution satellite data.
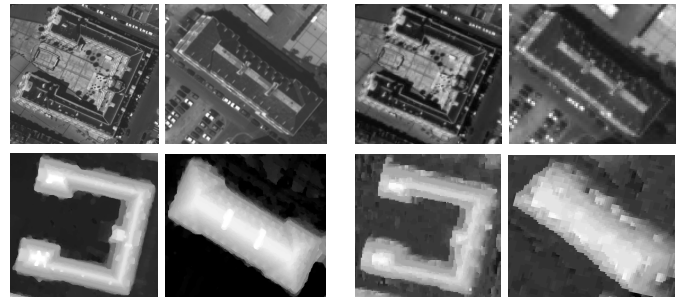

Fig. 3. Images of building *(top)* and associated DSM *(bottom)* obtained from 0.25 meter resolution aerial views *(left examples)* and 0.7 meter resolution satellite views *(right examples)*.

This approach is based on important prior knowledge concerning urban structures and their assembly. It is necessary to correctly define the interactions between blocks to have a convincing modeling without artefacts. A stochastic framework is especially well adapted to introduce such knowledge.

This paper extends the work we presented in [18] by detailing both the model and the optimization technique, as well as presenting new results and comments on various aerial/satellite scenes. First, the proposed system extracts the 2D-supports of the urban structures either automatically or interactively using previous works. Then, 3D-blocks are positioned on the 2D-supports using a Gibbs model described in Section 3. A Bayesian decision finds the optimal configuration of 3D-blocks using a Monte Carlo sampler. Experimental results on complex buildings and dense urban areas are shown using data of various resolutions in Section 4.

2 BUILDING EXTRACTION

The first step extracts the 2D-supports of the urban structures from a DSM. Each 2D-support is represented by a quadrilateral (or triangle) and is associated with a specific part of a building. The 2D-supports of a building correspond to sets of connected quadrilaterals (*i.e.* non-overlapping quadrilaterals with common edges). We propose two different ways for the user to extract the 2D-supports: an automatic one and an interactive one.

- **Automatic extraction -** The automatic extraction of building footprints from aerial and satellite data is a difficult problem which has been addressed by various techniques such as pixel based classification [19], contour vectorization [20], or active contours [21]. The used method, which has been developed in previous works [22][23], has an important advantage compared to other approaches: it is independent of the initialization state. The

method is performed in two steps. First, a rough approximation of the 2D-supports is realized by using an object approach based on marked point processes [22]. This method generates a set of rectangles representing the building footprints. Faced with the complexity and the diversity of the building forms, such an approach is well adapted since it provides a modeling by simple geometric objects (i.e rectangles) and allows the introduction of a prior knowledge concerning the object layout and doesn't need any initialization. An energy is associated to each configuration. The global minimum of this energy is then found by applying a simulated annealing [24]. Then, the rectangle layout is regularized into 2D-supports adapted to a structural concept [23]. The rectangle layout is transformed into sets of connected quadrilaterals by fusing neighboring rectangles. Moreover, each quadrilateral is partitioned by detecting the roof height discontinuities from the DSM: each resulting element represents a specific part of an urban structure. The 2D-supports obtained by this automatic method are convincing (see Figure 4). However, they are not as accurate as those obtained by an interactive extraction controlled by an operator. Some errors are generated, especially on complex buildings as we can see on Figure 4. That is why it is important to propose an alternative way to the user which allows to correct interactively these errors.

- **Interactive extraction -** An operator controls interactive extraction: four clicks add a 2D-support (these clicks are the four points of the quadrilateral). This method allows accurate extraction as we can see on Figure 4, but takes high operator time. It is especially interesting to choose the interactive extraction on some complex buildings or urban areas for which the automatic one gives bad results.
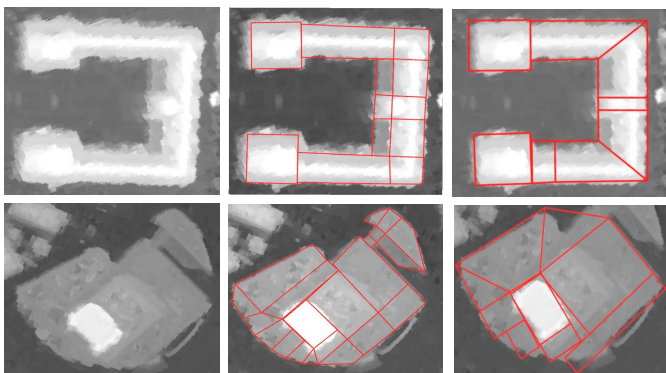

Fig. 4. Extraction of 2D-supports: DSMs *(left)*, interactive *(center)* and automatic extraction *(right)*.

3 3D RECONSTRUCTION

Once the 2D-supports have been extracted, the buildings are automatically reconstructed through a density

formulation. The first step consists in specifying the 3D objects.

3.1 Library of 3D-blocks

The content of the library is a key point: if it is too limited (*e.g.* Lafarge *et al.*[17] had only flat and gable roof forms), the method loses generality. The proposed library, denoted by $\mathcal{M}$ and presented in Figure 5, allows the reconstruction of a large range of buildings through an association of blocks. Each block possesses both a roof form and a variant:

- The proposed **roof forms** (denoted by $\mathcal{F}$ and illustrated in Figure 5-*top*) include monoplane ($\mathcal{F}_{1x}$), multi-plane ($\mathcal{F}_{2x}$) and curved roofs ($\mathcal{F}_{3x}$). Each roof form has a specific set of parameters F (the number of parameters varies between 1 and 6).
- The **variants** (denoted by $\mathcal{V}$ and shown in Figure 5-*bottom* for a gable roof type) are specific to a roof form. They correspond to types of structure ends (hipped or straight ends) or structure junctions ("-","L","T" or "+" junctions). The variants also specify the orientation of the roof with respect to the quadrilateral 2D-support (see Figure 6). The set of the variants, denoted by V, owns up to two parameters.

To sum up, each block of the library $\mathcal{M}$ is defined by a tuple $(\mathcal{F}, \mathcal{V})$ and an associated parameter set $\theta = (F, V)$. Some blocks can topologically be degenerated in some situations (*e.g.* a semi-elliptic roof on a triangular support): these cases are not allowed in the process in practice. More details concerning the parameters of the models are available in Tables 1 and 2.
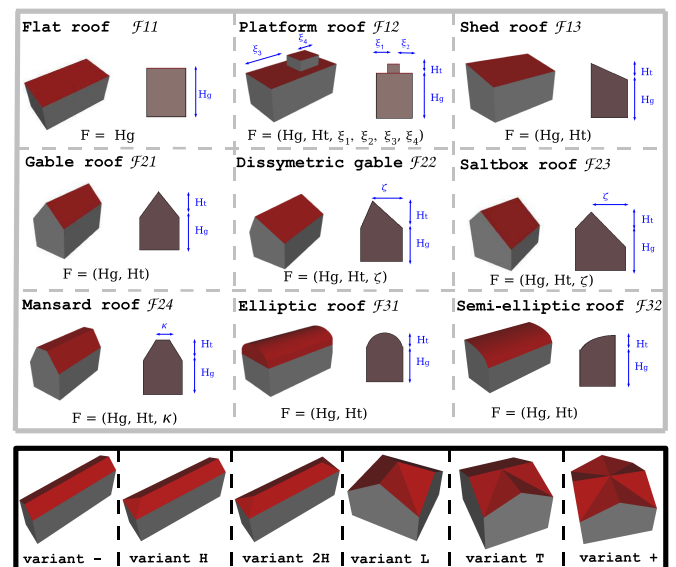

Fig. 5. Library of 3D-blocks - the roof forms (3D and profile views) *(top)* and the variants of a gable roof *(bottom)*.

TABLE 1
Parameter definition.

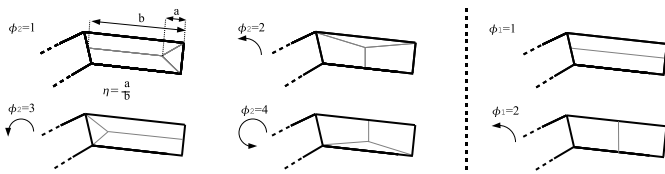| Parameter	Meaning	Domain
H_g	Gutter height	$[H_{gmin}, H_{gmax}]$
H_t	Rooftop height	$[H_{tmin}, H_{tmax}]$
$\xi_1, \xi_2, \xi_3, \xi_4$	Platform positioning	$[0,1]^4$
ζ	Rooftop positioning	$[0,1]$
κ	Pitch break positioning	$[0,1]$
ϕ_1	Symmetric roof orientation	$\{1,2\}$
ϕ_2	Dissymmetric roof orientation	$\{1,2,3,4\}$
η	Hipped end positioning	$[0,1]$

Fig. 6. Parameters ϕ and η: variants having one axis of symmetry *(left)*, and variants having two axes of symmetry *(right)*.

TABLE 2
Table of parameters for each model.

Roof forms	Form parameters F	Variants	
		types $\mathcal{V}$	parameters V
$\mathcal{F}_{11}$	H_g	$\mathcal{V}_-$	
$\mathcal{F}_{12}$	$H_g, H_t, \xi_1, \xi_2, \xi_3, \xi_4$	$\mathcal{V}_-$	
$\mathcal{F}_{13}$	H_g, H_t	$\mathcal{V}_-$	ϕ_2
		$\mathcal{V}_L$	ϕ_2
$\mathcal{F}_{21}$	H_g, H_t	$\mathcal{V}_-$	ϕ_1
		$\mathcal{V}_V$	ϕ_2, η
		$\mathcal{V}_{2V}$	ϕ_1, η
		$\mathcal{V}_L$	ϕ_2
		$\mathcal{V}_T$	ϕ_2
		$\mathcal{V}_+$	
$\mathcal{F}_{22}$	H_g, H_t, ζ	$\mathcal{V}_-$	ϕ_1
		$\mathcal{V}_V$	ϕ_2, η
		$\mathcal{V}_{2V}$	ϕ_1, η
		$\mathcal{V}_L$	ϕ_2
$\mathcal{F}_{23}$	H_g, H_t, ζ	$\mathcal{V}_-$	ϕ_2
		$\mathcal{V}_L$	ϕ_2
$\mathcal{F}_{24}$	H_g, H_t, κ	$\mathcal{V}_-$	ϕ_1
		$\mathcal{V}_L$	ϕ_2
		$\mathcal{V}_T$	ϕ_2
		$\mathcal{V}_+$	
$\mathcal{F}_{31}$	H_g, H_t	$\mathcal{V}_-$	ϕ_1
		$\mathcal{V}_L$	ϕ_2
		$\mathcal{V}_T$	ϕ_2
		$\mathcal{V}_+$	
$\mathcal{F}_{32}$	H_g, H_t	$\mathcal{V}_-$	ϕ_2
		$\mathcal{V}_L$	ϕ_2

3.2 Bayesian formulation

Let us introduce the notation for the Bayesian formulation:
• S, a set of sites and $\Lambda = \{\Lambda(s) : s \in S\}$, a given DSM where $\Lambda(s)$ represents the elevation of the site, s.
• $\mathcal{C}$, the quadrilateral configuration representing the building 2D-supports associated with Λ. N is the number of quadrilaterals (see Figure 4).
• $S_i = \{s \in int(i) : \Lambda(s) > H_e\}$, the subset of S whose sites are inside the quadrilateral, $i \in \mathcal{C}$, and have

elevation higher than half a floor height ^{1} H_e above ground (in practice, $H_e = 1.5$ meter).
• $\mathcal{D} = (\mathcal{D}_i)_{i \in \mathcal{C}}$, the set of data where $\mathcal{D}_i = \{\Lambda(s) : s \in S_i\}$.
• x, an element of the configuration space, $\mathcal{T}$, which corresponds to a configuration of 3D-parametric blocks knowing the 2D-supports $\mathcal{C}$. $x = (x_i)_{i \in \mathcal{C}} = (m_i, \theta_i)_{i \in \mathcal{C}}$ where each block, x_i, is specified by both a model, m_i of the library $\mathcal{M}$ and an associated set of parameters, θ_i. In the following, $x_i = (m_i, \theta_i)$ and m_i will be referred to as an object (or block) and a model, respectively.
• d_m, the number of continuous parameters describing the model m.
• S_{x_i}, the function from S_i to $\mathbb{R}$ which associates the roof altitude of the object, x_i, to each site of S_i.

Let us consider $\mathbb{1}_{\{.\}}$, the characteristic function. The measure associated with the set of 3D-blocks is given by:

$$\mu(u) = \sum_{k \in \mathcal{M}} \mathbb{1}_{\{m=k\}} \nu_k(\theta) \qquad (1)$$

where $u = (m, \theta)$ is an object of the library, and $\nu_k(.)$ is the measure associated with the model k. Most models own both continuous and discrete parameters. In this case, $\nu_k(\theta)$ is the product of two measures $\nu_k^{(c)}(\theta^{(c)}) \times \nu_k^{(d)}(\theta^{(d)})$ where $\nu_k^{(c)}(.)$ corresponds to the Lebesgue measure on $\mathbb{R}^{d_k}$ associated with the continuous parameters $\theta^{(c)}$, and $\nu_k^{(d)}(.)$ corresponds to the counting measure on $\mathbb{N}$ associated with the discrete parameters $\theta^{(d)}$. $\nu_k(.)$ corresponds to the Lebesgue measure on $\mathbb{R}^{d_k}$ if the model k does not own discrete parameter.
We then consider the measurable space $(\mathcal{T}, \mathcal{B}(\mathcal{T}), \mu^N(.))$ associated with the set of object configurations $\mathcal{T}$, where $\mathcal{B}(.)$ represent the Borel set. We denote by X, the random variable distributed in $\mathcal{T}$, which follows an unnormalized density, h against μ^N. h is the posterior density of a configuration, x, of objects, given $\mathcal{D}$. In a Bayesian framework, this density can be obtained from:

$$h(x) = h(x/\mathcal{D}) \propto h_p(x) \mathcal{L}(\mathcal{D}/x) \qquad (2)$$

The next step consists in building both a prior density, $h_p(x)$, and a likelihood, $\mathcal{L}(\mathcal{D}/x)$.

3.2.1 Likelihood

The likelihood represents the probability of observing the data, $\mathcal{D}$, knowing the configuration, x. By considering the hypothesis of conditional independence, it can be expressed by the local likelihood of objects, $\mathcal{L}(\mathcal{D}_i/x_i)$:

$$\mathcal{L}(\mathcal{D}/x) = \prod_{i \in \mathcal{C}} \mathcal{L}(\mathcal{D}_i/x_i) \qquad (3)$$

The usual solution to define the local likelihood consists in introducing a distance measuring the difference between the object and the DSM, such as the Tukey or

---

1. During the building extraction stage, some $2D$ supports are not correctly positioned. It is necessary to remove from the supports the pixels corresponding to the ground in order to preserve robustness.

Cauchy distances [25]. Then, the local likelihood can be expressed as follows:

$$\mathcal{L}(\mathcal{D}/x_i) = \frac{1}{Z(x_i)} \exp -\Gamma_{(i)}(\mathcal{S}_{x_i}, \mathcal{D}_i) \qquad (4)$$

where $Z(x_i)$ is the normalization constant of the local likelihood, and $\Gamma_{(i)}(.,.)$ corresponds to a distance from $\mathbb{R}^{card(S_i)} \times \mathbb{R}^{card(S_i)}$ to $\mathbb{R}$.

The used distance consists in measuring the Z-error of the L_α norm between the DSM and the object:

$$\Gamma_{(i)}(\mathcal{S}_{x_i}, \mathcal{D}_i) = \left(\sum_{s \in S_i} |\mathcal{S}_{x_i}(s) - \Lambda(s)|^\alpha \right)^{\frac{1}{\alpha}} \qquad (5)$$

This distance is based on the comparison between pixels of the DSM and the roof altitude of object x_i denoted by $\mathcal{S}_{x_i}$. Even if this distance seems to be simple, it efficiently models the data term of our problem. This choice is motivated by two main reasons:

- Outliers are not taken into account in the computation of the likelihood (see the definition of the data $\mathcal{D}$ - Section 3.2). Then, the use of complex distances specially adapted to noisy data such as Tukey distance can be avoided.
- This distance allows to make the normalization constant, $Z(x_i)$, independent of the object[2] x_i. It implies that the computation of the normalization constant is not necessary: we deal with unnormalized local likelihood.

In practice, we choose $\alpha = \frac{3}{2}$. The L_2 norm is too sensible to high variations: it is not adapted to our problem since DSMs are strongly noisy. The L_1 norm is more robust to noise. However, we have experimentally noticed that the roof profiles in the DSM are generally more concave than in reality: the L_1 norm does not ideally react to this phenomenon by under-estimating the roof slope. The $L_{1.5}$ norm is finally selected with respect to realized experiments on DSM fitting detailed in [26].

3.2.2 Prior

The prior introduces interactions between neighboring objects: it is a key point in our structural concept. It allows us to both assemble objects in order to propose a realistic building and compensate for the lack of information contained in the DSM. A neighborhood relationship on $\mathcal{C}$ must be set up to define the interactions: two distinct quadrilaterals, i and $j \in \mathcal{C}$, are said to be neighbors if they have a common edge. The neighborhood relationship is denoted by $\bowtie$ ($i \bowtie j$ represents the set of neighboring pairs in $\mathcal{C}$). In a previous work, too many interactions were set up [17]. The number must be minimal to preserve robustness and avoid problems in parameter setting. We propose a simple and efficient prior which is defined through a single interaction.

To do so, we define an assembly law which tests whether two objects can be assembled together. Two neighboring objects $x_i = (m_i, \theta_i)$ and $x_j = (m_j, \theta_j)$ are said "**joinable**" (denoted by $x_i \sim_a x_j$), if:

1 - $\mathcal{F}_{f_i} = \mathcal{F}_{f_j}$ and
2 - roof top orientations are compatible and
3 - the common edge of the quadrilateral 2D-supports, i and j, is not a roof height discontinuity[3].

The first condition checks that the two blocks have the same roof form. The second and third conditions test whether the roof tops of the two objects can be connected.

The prior favors "joinable" objects, *i.e.* homogeneous structures. However, heterogeneous structures can also be reconstructed. Moreover, in order to avoid artefacts, the parameters of two "joinable" objects are encouraged to have similar values. To do so, the unnormalized density, h_p, is expressed through a Gibbs energy, U_p (*i.e.* $h_p(x) = \exp -U_p(x)$), defined by:

$$\forall x \in \mathcal{T}, U_p(x) = \beta \sum_{i \bowtie j} \mathbb{1}_{\{x_i \sim_a x_j\}} g(x_i, x_j) \qquad (6)$$

where $\mathbb{1}_{\{.\}}$ is the characteristic function. The parameter, $\beta \in \mathbb{R}^+$, weights the importance of the prior density relatively to the likelihood. The function, g, taking values in $[-1, 0]$, measures the distance between the parameters of two "joinable" objects:

$$g(x_i, x_j) = \frac{D(x_i, x_j)}{D_{max}} - 1 = \frac{\sum_k \omega_k |\widetilde{\theta}_{i,(k)} - \widetilde{\theta}_{j,(k)}|}{D_{max}} - 1 \quad (7)$$

$\widetilde{\theta}_{i,(k)}$ and $\widetilde{\theta}_{j,(k)}$ are the kth element of the parameter sets F_i and F_j of the objects x_i and x_j, respectively. $D_{max} = \max_{x_i, x_j} D(x_i, x_j)$ is the maximum value of the distance. ω_k are weights which are introduced in this distance in order to normalize the parameter values according to the metric system. These weights are computed from the X, Y and Z resolutions and the configuration of quadrilaterals $\mathcal{C}$.

Figure 7 illustrates this interaction. If the two blocks belong to different roof types (for example a mansard roof and a semi-elliptic roof on the top right) or if the two objects do not have compatible roof orientations (bottom right), they will not be "joinable" and the energy will be null. On the contrary, if the two objects are "joinable", the energy will be negative : these configurations are favored. The nearer the parameters of the two objects, the lower the energy. The left configuration is the best one with respect to the prior. To sum up, the proposed density $h(.)$ is expressed as follows:

$$h(x) = \exp - \left(\sum_{i \in \mathcal{C}} \Gamma_{(i)}(\mathcal{S}_{x_i}, \mathcal{D}_i) + \beta \sum_{i \bowtie j} \mathbb{1}_{\{x_i \sim_a x_j\}} g(x_i, x_j) \right)$$

$$(8)$$

---

2. The proof consists in making the substitution of variables $u(s) = \mathcal{S}_{x_i}(s) - \Lambda(s)$ when we integrate on the data space in the expression of $Z(x_i)$: the resulting expression is then independent of x_i by assuming that the data space is not bounded.

3. The detection of the roof height discontinuities from the DSM is detailed in previous works [23].
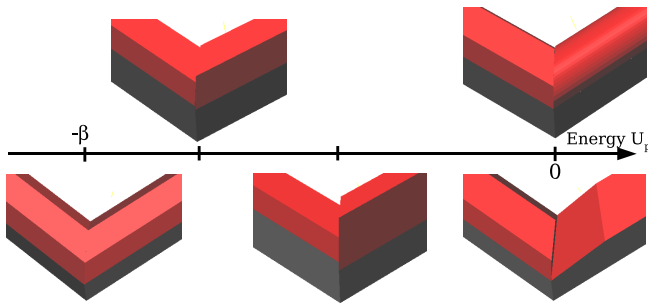
Fig. 7. Prior energy - local configurations of various energies.

This density only depends on the parameter β which is computed using the Maximum Likelihood estimator under regularization constraints detailed in [26]. It provides robustness to the process and avoids the parameter setting problems. In the following, the Gibbs energy associated with h will be denoted by U ($U = -\ln h$).

3.3 Optimization

We now find the object configuration maximizing the posterior density, $h(.)$, *i.e.* the Maximum A Posteriori (MAP) estimator, x_{MAP}. This is a non convex optimization problem in a high and variable dimension space, $\mathcal{T}$, since the blocks in the library, $\mathcal{M}$, are defined by a different number of parameters.

3.3.1 RJMCMC sampler

The Reversible Jump Markov Chain Monte Carlo (RJM-CMC) algorithm [27] is well adapted to our problem. This technique is an extension of the formalism introduced by Hastings [28], allowing to deal with variable dimension state spaces. Several papers have shown the efficiency of the RJMCMC sampler for the problem of multiple parametric object recognition. For example, Dick *et al.*[29] use such a sampler to reconstruct architectural buildings from terrestrial images where the parametric object set includes structures such as columns, buttresses, entablatures or drainpipes. Brenner *et al.*[30] extract facades through a grammar based approach driven by a RJMCMC sampler.

The RJMCMC sampler simulates a discrete Markov Chain $(X_t)_{t\in\mathbb{N}}$ on the configuration space $\mathcal{T}$, which converges towards an invariant measure (specified by the posterior density, $h(.)$). The transitions of this chain correspond to local perturbations[4] of the current configuration. The chain is built in order to be ergodic: it allows, under certain conditions of probabilistic relaxation, to ensure convergence towards the target measure for any initial configuration. Each iteration of the sampler is composed of two steps. The first one consists in proposing a new state by perturbing the current one. The

4. Local perturbations mean that only one object of the configuration is generally concerned by a perturbation.

second step decides whether the perturbation is accepted to define the new state. We denote by:

- π, the target measure defined on $\mathcal{T}$ and specified by the posterior density h.
- $Q_k(.,.)$, the proposition kernels defined on $\mathcal{T} \times \mathcal{B}(\mathcal{T})$. They allow to propose different types of perturbations specified in Section 3.3.2.
- $R_k(x, y)$, the acceptation ratio of a proposition from $x \to y$ given by:

$$R_k(x,y) = \frac{\pi(dy)Q_k(y,dx)}{\pi(dx)Q_k(x,dy)} \qquad (9)$$

The acceptance probability of a perturbation from $x \to y$ is then expressed by $\min(1, R_k(x, y))$.

In summary, the RJMCMC sampler is: at iteration t, if $X_t = x$:

1 - Choose the kernel $Q_k(x, .)$ with probability q_k.
2 - According to Q_k, propose a new state y.
3 - Take $x^{(t+1)} = y$ with probability $\min(1, R_k(x, y))$, and take $x^{(t+1)} = x$ otherwise.

3.3.2 Proposition kernels

The kernel specification plays a crucial role in the efficiency of the sampler. Appropriate kernels allow to accelerate the convergence of the process by proposing object configurations of interest.

Let us consider two models $\mathcal{M}_m$ and $\mathcal{M}_n$ and a perturbation from an object $x_i = (m, \theta_i)$ to an object $\widehat{x}_i = (n, \widehat{\theta}_i)$ such that the current object configuration $x = (x_p)_{p\in\mathcal{C}}$ is perturbed into the configuration $y = (x_p)_{p\in\mathcal{C}-\{i\}} \cup \widehat{x}_i$. The idea of Green [27] is to create a bijection between the parameter spaces of the models $\mathcal{M}_m$ and $\mathcal{M}_n$. θ_i is completed by auxiliary variables u_{mn} simulated under a law $\varphi_{mn}(.)$ to provide (θ_i, u_{mn}), and $\widehat{\theta}_i$ by $v_{nm} \sim \varphi_{nm}(.)$ into $(\widehat{\theta}_i, v_{mn})$ such that the mapping Ψ_{mn} between (θ_i, u_{mn}) and $(\widehat{\theta}_i, v_{mn})$ is a bijection :

$$(\widehat{\theta}_i, v_{mn}) = \Psi_{mn}(\theta_i, u_{mn}) \qquad (10)$$

The ratio of the kernels in the acceptation ratio is then expressed by:

$$\frac{Q_k(y,dx)}{Q_k(x,dy)} = \frac{J_{nm}^{(k)}\varphi_{nm}^{(k)}(v_{nm})}{J_{mn}^{(k)}\varphi_{mn}^{(k)}(u_{mn})} \left| \frac{\partial \Psi_{mn}(\theta_i, u_{mn})}{\partial(\theta_i, u_{mn})} \right| \qquad (11)$$

where J_{mn} corresponds to the probability of choosing a jump from $\mathcal{M}_m$ to $\mathcal{M}_n$. Examples of the computation of the bijection Ψ_{mn} are detailed in Appendix. We propose three kernels, *i.e.* three sets of distributions (J_{mn}, φ_{mn}). More details concerning the computation of these kernels are given in [26].

- **Kernel Q_1: uniform jumps** This is the classic kernel proposing a new state according to uniform distributions. It is enough to ensure that the Markov chain can visit any configuration of the state space. However, using only this kernel requires long computation time [17]. Thus we propose two efficient additional kernels Q_2 and Q_3.

- **Kernel Q_2: data-driven jumps** This kernel cleverly explores the state space using a data-driven process which is efficient for similar problems [31]. To do so, the state, x, is proposed knowing the data, *i.e.* according to a probability, $p(x|\mathcal{D})$. More precisely, it firstly estimates the gutter roof height $\widehat{H_g}$ and the roof top height $\widehat{H_t}$ of the object concerned by the jump by using a median filtering on DSM areas of interest. Secondly, it chooses its height values according to the Gaussian distributions, $\mathcal{N}(\widehat{H_g}, \sigma)$ and $\mathcal{N}(\widehat{H_t}, \sigma)$, respectively (in practice, $\sigma = 1$ m). Moreover, this kernel takes into account knowledge about roof form occurrences in the urban scene by using a counting process. The probabilities J_{mn} are then focused on models of interest.

- **Kernel Q_3: regularization jumps** In our application, the visual aspect of the result is very important: we need a kernel which proposes well-regularized objects, *i.e.* objects which are perfectly connected with their neighbors. The new object, x_i, must be proposed knowing its neighboring objects $\{x_j/j \bowtie i\}$, *i.e.* according to $p(x_i|\{x_j/j \bowtie i\})$. The model is uniformly selected according to the models of the neighboring objects. The parameter values are chosen according to Gaussian mixtures depending on the parameter values of the neighboring objects. Although this kernel is very useful to regularized objects, it can block the current configuration in a local optimum. That is why it must be mainly used at the end of the process, *i.e.* when the current configuration is close to the optimal one.

3.3.3 Simulated annealing

Simulated annealing [24] is used to ensure convergence: the density, $h(.)$, is substituted by $h(.)^{\frac{1}{T_t}}$, where T_t is a sequence of temperatures which tends to zero as t tends to infinity. Simulated annealing theoretically ensures convergence to the global optimum for any initial configuration, x_0, using a logarithmic temperature decrease. In practice, we use a geometric decrease which is faster and gives an approximate solution close to the optimal one. Such a decrease, detailed in [32], is defined as follows:

$$T_t = T_o.\alpha^t \tag{12}$$

where α and T_0 are respectively the decrease coefficient and the initial temperature. The decrease coefficient α can vary and be adapted according to the variation of the energy [33][34][35]. However, the time savings are usually relatively minor in practice. That is why we prefer using a constant decrease coefficient depending on the number of objects (in practice, $\alpha = 0.9999^{\frac{1}{N}}$).

T_0 is estimated through the variation of the energy U on random configurations. More precisely, T_0 is chosen as twice the standard deviation of U at infinite temperature [36] :

$$T_0 = 2.\sigma(U_{T=\infty}) = 2.\sqrt{\langle U_{T=\infty}^2 \rangle - \langle U_{T=\infty} \rangle^2} \tag{13}$$

where $\langle U \rangle$ is the means of the energy of the samples (several thousands of samples are necessary to obtain a good estimation - it is negligible w.r.t. the number of iterations of the optimization process).

The process has two stages. At the beginning, *i.e.* when the temperature is high (see Figure 8-*(d)* showing two simple examples of simulations), the process explores the density modes and favors configurations which have a high density. In this exploration stage, the data-driven kernel Q_2 is mainly used ($q_1 = q_3 = \frac{1}{8}q_2 = 0.1$). At low temperature[5], the configuration is close to the optimal solution and does not evolve very much : it involves a detailed adjustment of the 3D-block parameters. In this second stage, the regularization kernel Q_3 is mainly used ($q_1 = q_2 = \frac{1}{8}q_3 = 0.1$). Figure 10 shows the efficiency of the kernel triplet with respect to a single uniform kernel. The obtained configuration is better in terms of energy (U=971 *vs* U=986 for the single uniform kernel) by using a temperature decrease which is three times faster.
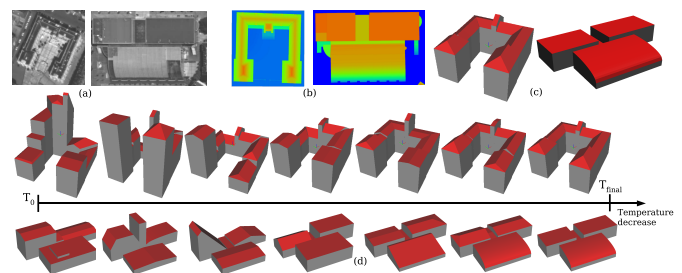

Fig. 8. Two simple examples of the optimization process - evolution of the configuration as the temperature decreases *(d)*, associated with the satellite images *(a)*, the ground truth *(b)* and the final results *(c)*.
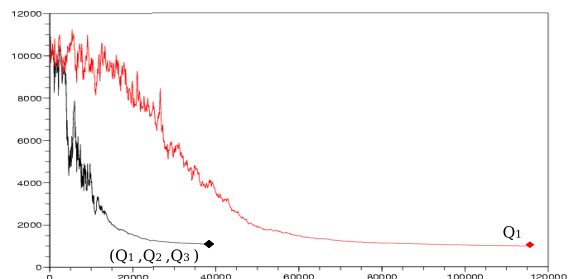

Fig. 9. Energy decrease $x = t$, $y = U$ with a single uniform kernel *(red)* and with the combination of the three kernels *(black)*.

4 EXPERIMENTS

The results were obtained from satellite DSMs (0.7 m resolution) and aerial DSMs (0.1 and 0.25 m resolution).

5. In practice, the second stage is detected when the accepted proposition rate computed on 1000 iterations becomes lower than 0.05.
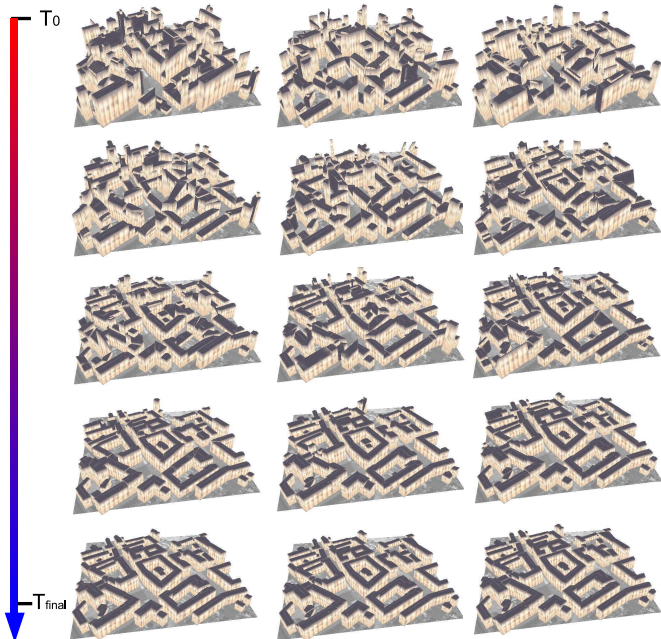
Fig. 10. Optimization - evolution of the configuration as the temperature decreases *(left to right, and top to down)*.

The results show the reconstruction of complex buildings and dense urban areas whose level of detail depends on both the choice of the extraction process (automatic or interactive) and the kind of data. There is no comparison with other approaches for the reasons underlined in the introduction, except where the context is similar [17]. DSMs have been generated from 3-view images by an algorithm detailed in [37], and based on a multi-resolution implementation of a Cox and Roy optimal flow image matching algorithm [38]. The 3D ground truth are raster images. Generic textures were applied to the objects of Figures 11 and 12 for visualization.

4.1 With automatic extraction

The following results have been obtained by the automatic process from single DSMs, *i.e.* without operator control or cadastral information. Figures 11 and 12 present various examples of reconstruction (showing different roof types, roof height discontinuities, closed structures or complex roof junctions). These results are convincing. The 3D-blocks are correctly assembled and few artefacts are generated which means that the process adapts to buildings with complex roof junctions.

Results shown in Figure 11 were obtained from satellite DSMs with automatic 2D extraction [22][23]. Even if some details are omitted, the shapes of buildings compare well to the ground truth and the generalization level is satisfactory with respect to the context. The roof height discontinuities were accurately located.

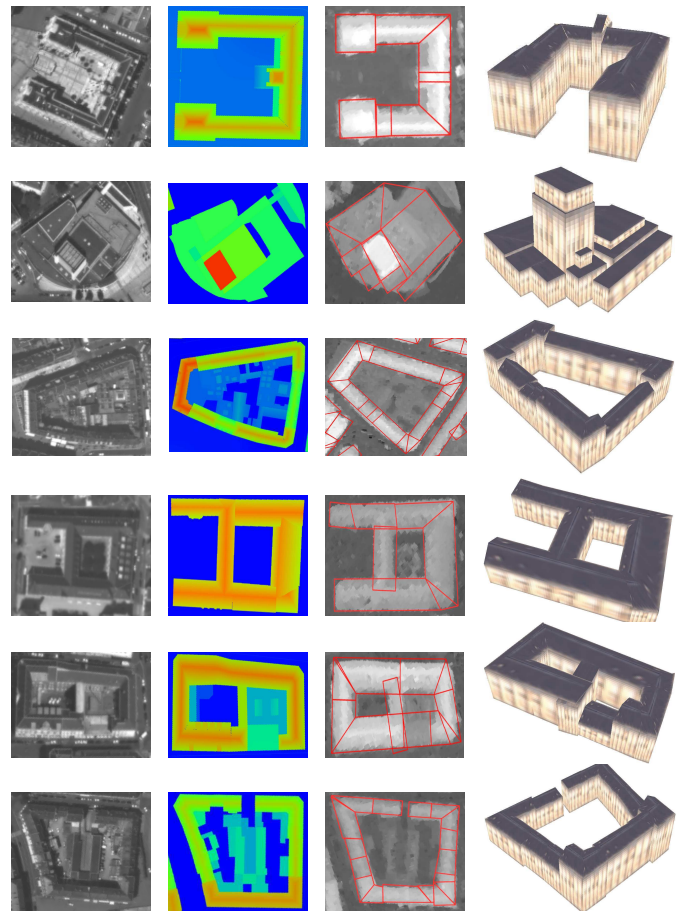Globally speaking, the different roof types were correctly identified as we can see on Figure 11. When the roof

Fig. 11. Reconstructed buildings with automatic 2D extraction from satellite data *(4th column)*, satellite images *(1st column)*, ground truth *(2nd column)* and DSMs with 2D-supports *(3rd column)*.
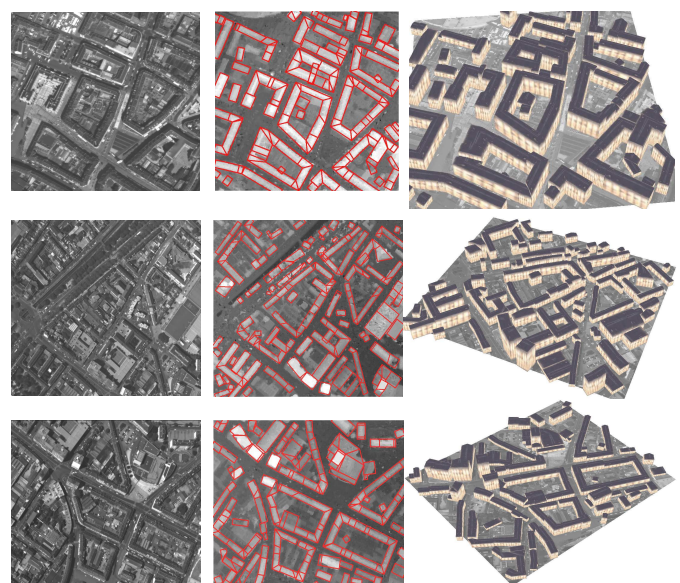

Fig. 12. Urban areas reconstructed with automatic 2D extraction from satellite data *(3rd column)*, satellite images *(1st column)*, and DSMs with 2D-supports *(2nd column)*.
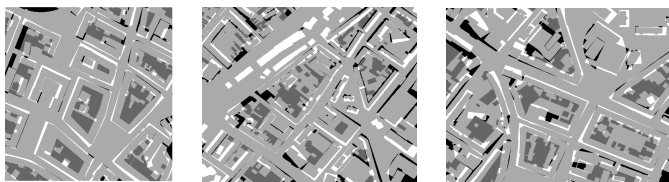
Fig. 13. Ground errors of results presented on Figure 12 (dark gray: missed corresponding to low flat buildings of inner courtyards, black: missed, white: over-detected).
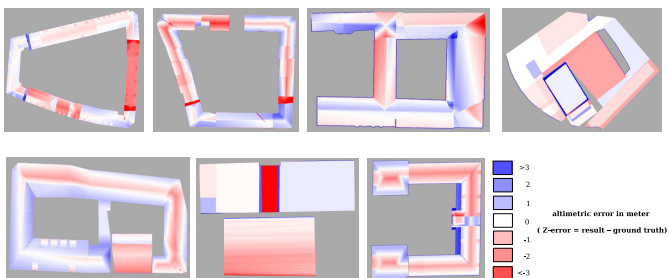

Fig. 14. Altimetric errors of buildings shown in Figure 11.

model selection is not well discriminated by the distance Γ introduced in the likelihood formulation (see Section 3.2.1), the prior provides helpful information concerning the roof models of the neighboring blocks: this point represents one of the main advantages of the Bayesian decision for such a structural concept. Nevertheless, some errors are generated between the similar roof forms, especially with the mansard roof form. When the parameter κ of the mansard roof is close to 0 (respectively close to 1), the discrimination with a gable roof (respectively with a flat roof) become delicate. Other confusions can appear between gable roofs with low slope (*i.e.* H_t close to 0) and flat roofs as we can see on Figure 15 (1st and 2nd examples). Such confusions are not important in terms of altimetric accuracy, but can corrupt the visual rendering of the scene.

The ground errors for the automatic 2D extraction process were satisfactory. The over-detection rate (in terms of surface) was 9.7%. This rate can be improved by adding a vegetation mask to prevent the detection of trees. The rate of missed detection was quite high (15%). However, it was mainly due to low flat buildings in inner courtyards (one floor height structures) that the 2D extraction process cannot detect since these buildings have low DSM discontinuities. Without taking into account these low flat buildings, this rate falls to 4.5% (see Figure 13). The altimetric Root Mean Square Error (RMSE) against ground truth in this context (satellite data / fully-automatic process) is 2.3 m. This is better than the 3.2 m error obtained earlier [17] in the same context, but still remains high. As we can see on Figure 14, it is mainly due to both a non optimal positioning of 2D-supports in the automatic extraction process (which engenders important local altimetric errors at some lo-

cations) and inaccuracies in the DSMs (which mainly correspond to matching problems of non Lambertian surfaces such as glass roofs). The second example of Figure 11 underlines the limits of the automatic 2D extraction process: some footprints (especially curved footprints) cannot be modeled accurately by sets of quadrilaterals. The proposed kernels achieved acceptable computation times. Less than one minute is necessary to obtain the buildings of Figure 11 using a 3Ghz processor (*vs* 5 minutes [17] with the same processor).

4.2 With interactive extraction

Although the results presented in Section 4.1 are convincing, the automatic extraction of the 2D-supports generates some errors. The use of the interactive extraction is then an interesting alternative to reduce these errors. Figure 15 presents examples of buildings reconstructed by the interactive 2D extraction process. In the two first examples, the results are clearly better since the 2D-supports are accurately located. The reconstructed buildings are more detailed and the altimetric RMSE is 1.1 m from the satellite data. Table 3 summerizes the different altimetric RMSEs with respect to both the type of data (0.7 m resolution satellite DSMs or 0.25 m resolution aerial DSMs) and the level of automation (with automatic or interactive extraction processes).
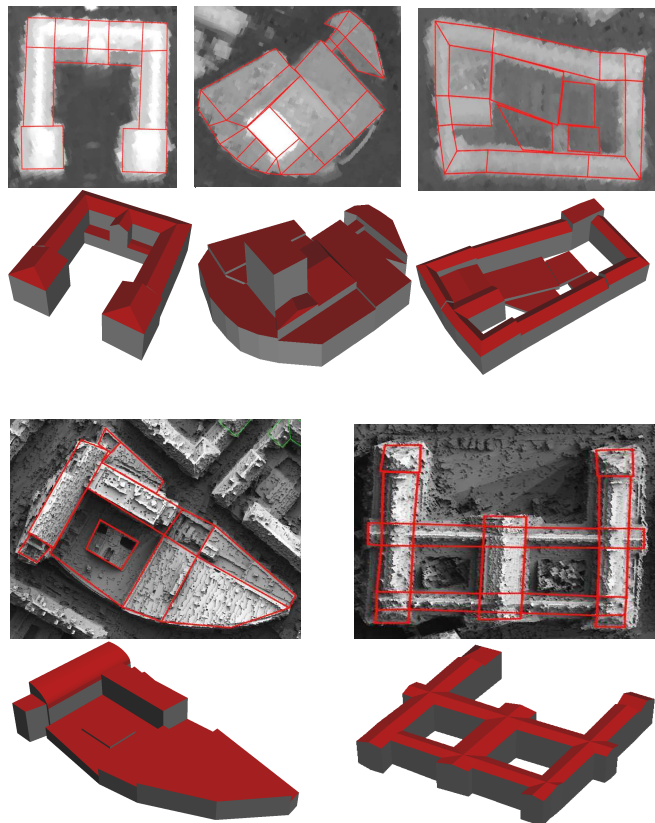

Fig. 15. Reconstructed buildings with interactive 2D extraction from satellite data (top examples) and 0.25 m resolution aerial data (bottom examples).

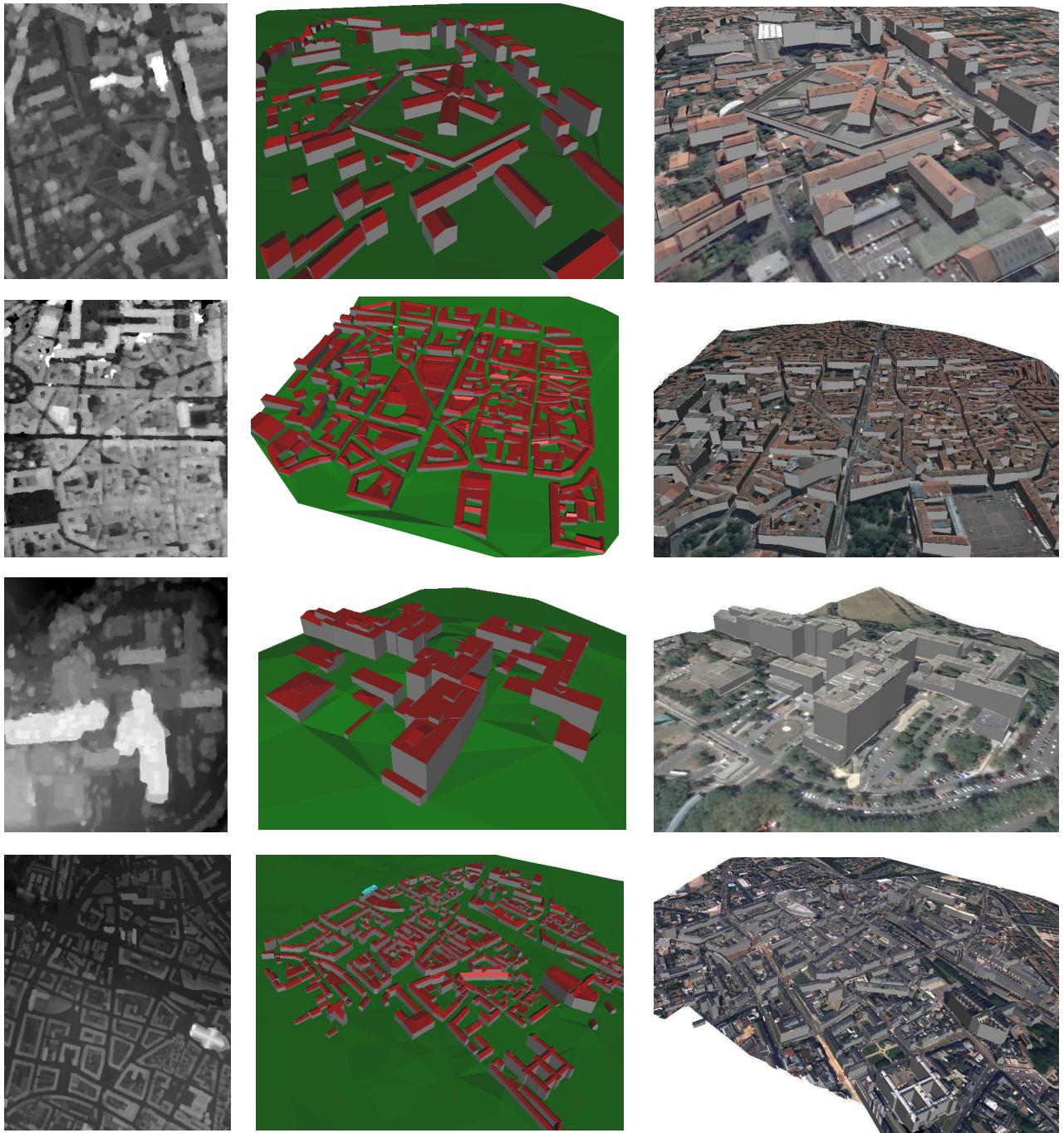Fig. 16. Urban area reconstruction without and with textures *(3rd and 2nd columns respectively)* from DSMs *(1st column)*. The three first examples correspond to 0.7 meter resolution satellite DSMs while the last one is a 0.25 meter resolution aerial DSM.
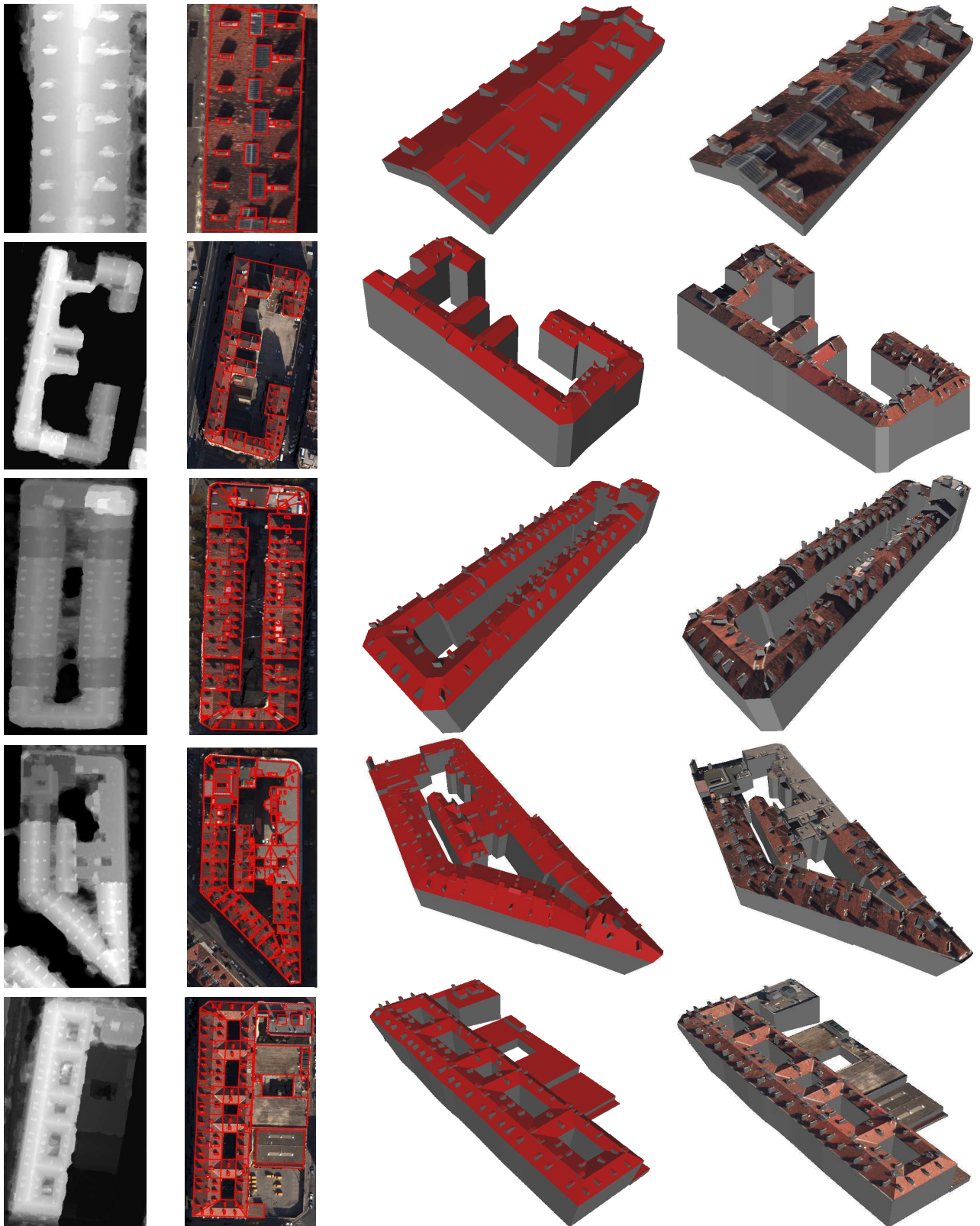
Fig. 17. Building superstructure reconstruction without and with textures *(3rd and 4th columns respectively)* from 0.1 m resolution aerial DSMs *(1st column)* and interactive 2D-supports *(2nd column)*.

TABLE 3
Altimetric evaluation (RMSE) w.r.t. type of data / level of automation.

	with automatic extraction	with interactive extraction
from satellite DSMs (0.7m)	2.3 m	1.1 m
from aerial DSMs (0.25m)	1.3 m	0.8 m

Figure 16 shows results on four typical European down town areas. The computation time for the fourth example was 35 minutes ($0.6 km^2$ - about 800 objects) which is very good. However, the operator time is quite high (about 3 hours to create the 800 2D-supports).

The proposed method also allows modeling roof details such as chimneys or dormer-windows. In fact, the library of 3D-blocks is general enough to reconstruct such roof superstructures. Figure 17 presents accurate results of building superstructure reconstruction from a 0.1 m resolution aerial DSM. The buildings and their superstructures are correctly represented. Most chimneys, dormer-windows and glass roofs are modeled by flat roofs, gable roofs and shed roofs respectively. However, frequent errors often appear on the chimney reconstruction. The system has difficulties to recognize the small structures, and often selects an incorrect roof form. A solution could be to introduce prior knowledge concerning the object size. The altimetric RMSE for these images is 0.5 m which is a very good result.

5 CONCLUSION

This new method is an interesting alternative to generic and parametric approaches. This structural concept is both generic (primitives are combined using Gibbs models) and parametric (primitives are extracted from a library of parametric objects). It presents several important characteristics. First, it provides very good results from a single DSM. Moreover, this approach works efficiently on various data resolutions: a global description of the buildings from 0.7 m resolution satellite data or a detailed building reconstruction including roof superstructures from 0.1 m aerial data. The user can also choose the level of automation of the process since the 2D-supports can be extracted either interactively or automatically. Finally, it is an adaptive method since other 3D-block types can be added to the library depending on the context.

In future works, it would be interesting to improve the optimization step to achieve both higher precision and shorter computing time. Belief propagation techniques for graphical models could be used or Jump-Diffusion processes [39] which are efficient for similar optimization problems [40]. Moreover, we should evaluate the potential of this method on other kinds of cities such as typical North American urban areas, and test the system on other DSMs generated by methods such as [41].

APPENDIX

The library $\mathcal{M}$ is composed of 26 models. Among these models, there are 12 different parameter sets. Then, $12^2 - 12 = 132$ bijections and associated completion parameters must be computed. The model parameters have been defined in order to simplify the computation (most parameters taking values in $[0, 1]$). In this appendix, we present two different examples of bijection computation. Let us consider two models $\mathcal{M}_m$ and $\mathcal{M}_n$ and a perturbation from an object $x = (m, \theta)$ to an object $\widehat{x} = (n, \widehat{\theta})$. In order to create a bijection Ψ_{mn} between the parameter spaces of the models $\mathcal{M}_m$ and $\mathcal{M}_n$, θ is completed into (θ, u_{mn}) by simulating $u_{mn} \sim \varphi_{mn}(.)$, and $\widehat{\theta}$ into $(\widehat{\theta}, v_{mn})$ by $v_{nm} \sim \varphi_{nm}(.)$ such that $(\widehat{\theta}, v_{mn}) = \Psi_{mn}(\theta, u_{mn})$.

First example: Let us consider a jump from the model $(\mathcal{F}_{11}, \mathcal{V}_-)$ (denoted by $\mathcal{M}_1$) to the model $(\mathcal{F}_{21}, \mathcal{V}_-)$ (denoted by $\mathcal{M}_2$). We move from $\theta = H_g$ to $\widehat{\theta} = (H_g, H_t, \phi_1)$. The parameter H_g exists in both models: we need to complete the model $\mathcal{M}_1$ by $u_{12} = (H_t, \phi_1)$. We then obtain $\widehat{\theta} = \Psi_{12}(\theta, u_{12})$ where Ψ_{12} corresponds to the identity function. The jacobian is equal to 1.

Second example: Let us consider a jump from the platform model $(\mathcal{F}_{12}, \mathcal{V}_-)$ (denoted by $\mathcal{M}_1$) to the dissymetric model with one hipped end $(\mathcal{F}_{22}, \mathcal{V}_V)$ (denoted by $\mathcal{M}_2$). We have $\theta = (H_g, H_t, \xi_1, \xi_2, \xi_3, \xi_4)$ and $\widehat{\theta} = (H_g, H_t, \zeta, \phi_2, \eta)$. The common parameters are H_g and H_t. Moreover, it exists, for all kernels, a direct relation between the parameters ξ_1, ξ_2 from $\mathcal{M}_1$ and the parameters ζ, η from $\mathcal{M}_2$. But, it is not possible to find a bijection between ξ_3 et ϕ_2. Then, the completion parameters are $u_{12} = \phi_2$ and $v_{21} = (\xi_3, \xi_4)$. We obtain $(\widehat{\theta}, v_{21}) = \Psi_{12}(\theta, u_{12})$ where Ψ_{12} corresponds to the identity function. In the case of the uniform kernel for example, we have $\varphi_{21}^{(1)}(v_{21}) = \mathcal{U}_{[0,1]^2}(v_{21}) = 1$ and $\varphi_{12}^{(1)}(u_{12}) = \frac{1}{4}$.

REFERENCES

[1] H. Mayer, "Automatic object extraction from aerial imagery -a survey focusing on buildings", *Computer Vision and Image Understanding*, vol.74, no.2, 1999.

[2] E.P. Baltsavias, "Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.58, no.3-4, 2004.

[3] S. Scholze, T. Moons and L. Van Gool, "A Probabilistic Approach to Building Roof Reconstruction Using Semantic Labelling", *Proc. of the 24th DAGM Symposium on Pattern Recognition*, London, UK, 2002.

[4] Z. Kim and R. Nevatia, "Automatic description of complex buildings from multiple images", *Computer Vision and Image Understanding*, vol.96, no.1, 2004.

[5] C. Baillard, C. Schmid, A. Zisserman and A. Fitzgibbon, "Automatic line matching and 3D reconstruction of buildings from multiple views", *ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery*, 1999.

[6] F. Taillandier and R. Deriche, "Automatic Builings Reconstruction from Aerial Images : a Generic Bayesian Framework", *Proc. of ISPRS Congress*, Istanbul, Turkey, 2004.

[7] U. Weidner and W. Förstner, "Towards Automatic Building Reconstruction from High Resolution Digital Elevation Models", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.50, no.4, 1995.

[8] R. Collins, C. Jaynes, Y. Cheng, X. Wang, F. Stolle, E. Riseman and A.R. Hanson, "The Ascender System: Automated Site Modeling from Multiple Aerial Images", *Computer Vision and Image Understanding*, vol.72, no.2, 1998.

[9] M. Brédif, D. Boldo, M. Pierrot-Deseilligny and H. Maître, "3D building reconstruction with parametric roof superstructures", *Proc. of IEEE International Conference on Image Processing*, San Antonio, US, 2007.

[10] A. Fischer, T. Kolbe, F. Lang, A. Cremers, W. Förstner, L. Plümer and V. Steinhage, "Extracting Buildings from Aerial Images using Hierarchical Aggregation in 2D and 3D", *Computer Vision and Image Understanding*, vol.72, no.2, 1998.

[11] E. Gülch and H. Müller, "New applications of semi-automatic building extraction", *Proc. of the Third International Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Ascona, Switzerland, 2001.

[12] S.C. Lee and R. Nevatia, "Interactive 3D building modeling using a hierarchical representation", *IEEE Workshop on higher-level knowledge in 3D modeling and motion*, Nice, France, 2003.

[13] H.G. Maas and G. Vosselman, "Two algorithms for extracting building models from raw laser altimetry data", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.54, no.2-3, 1999.

[14] C. Brenner, "Building reconstruction from images and laser scanning", *International Journal of Applied Earth Observation and Geoinformation*, vol.5, 2005.

[15] C. Früh, S. Jain and A. Zakhor, "Data Processing Algorithms for Generating Textured 3D Building Facade Meshes from Laser Scans and Camera Images", *International Journal of Computer Vision*, vol.61, no.2, 2005.

[16] P. Müller, P. Wonka, S. Haegler, A. Ulmer and L. Van Gool, "Procedural modeling of buildings", *Proc. of ACM SIGGRAPH*, Boston, US, 2006.

[17] F. Lafarge, X. Descombes, J. Zerubia and M. Pierrot-Deseilligny, "An Automatic Building Reconstruction Method : A Structural Approach Using High Resolution Images", *Proc. of IEEE International Conference on Image Processing*, Atlanta, US, 2006.

[18] F. Lafarge, X. Descombes, J. Zerubia and M. Pierrot-Deseilligny, "Building reconstruction from a single DEM", *Proc. of IEEE Computer Vision and Pattern Recognition*, Anchorage, US, 2008.

[19] N. Haala and C. Brenner, "Extraction of buildings and trees in urban environments", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.54, no.2-3, 1999.

[20] C. Vestri and F. Devernay, "Using Robust Methods for Automatic Extraction of Buildings", *Proc. of IEEE Computer Vision and Pattern Recognition*, Hawaii, US, 2001.

[21] T. Bailloeul, V. Prinet, B. Serra and P. Marthon, "Spatio-temporal Prior Shape Constraint for Level Set Segmentation", *Proc. of Energy Minimization Methods in Computer Vision and Pattern Recognition*, St. Augustine, US, 2005.

[22] M. Ortner, X. Descombes and J. Zerubia, "Building outline extraction from Digital Elevation Models using marked point processes", *International Journal of Computer Vision*, vol.72, no.2, 2007.

[23] F. Lafarge, X. Descombes, J. Zerubia and M. Pierrot-Deseilligny, "Automatic Building Extraction from DEMs using an Object Approach and Application to the 3D-city Modeling", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.63, no.3, 2008.

[24] M. Metropolis, A.W. Rosenbluth, A.H. Teller and E. Teller, "Equation of state calculations by fast computing machines", *Journal of Chemical Physics*, vol.21, 1953.

[25] G. Xu and Z. Zhang, "Epipolar Geometry in stereo, motion and object recognition", *Kluwer Academic Publishers*, 1996.

[26] F. Lafarge, "Stochastic approaches for 3D building reconstruction", *Ph.D. dissertation, Ecole Nationale Supérieure des Mines de Paris*, Paris, France, 2007.

[27] P.J. Green, "Reversible Jump Markov Chains Monte Carlo computation and Bayesian model determination", *Biometrika*, vol.82, 1995.

[28] W.K. Hastings, "Monte Carlo sampling using Markov chains and their applications", *Biometrika*, vol.57, no.1, 1970.

[29] A.R. Dick, P.H.S. Torr and R. Cipolla, "Modelling and Interpretation of Architecture from Several Images", *International Journal of Computer Vision*, vol.60, no.2, 2004.

[30] C. Brenner and N. Ripperda, "Extraction of facades using RJMCMC and constraint equations", *Proc. of ISPRS Commission III Symposium on Photogrammetric and Computer Vision*, Bonn, Germany, 2006.

[31] Z.W. Tu and S.C. Zhu, "Image Segmentation by Data-Driven Markov Chain Monte Carlo", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.24, no.5, 2002.

[32] P.J.M. Van Laarhoven and E.H.L. Aarts, "Simulated Annealing : Theory and Applications", Boston, US, 1987.

[33] G. Perrin, X. Descombes and J. Zerubia, "Adaptive Simulated Annealing for Energy Minimization Problem in a Marked Point Process Application", *Proc. of Energy Minimization Methods in Computer Vision and Pattern Recognition*, St Augustine, US, 2005.

[34] J.M. Varanelli, "On the acceleration of the simulated annealing", *Ph.D. dissertation, University of Virginia*, Charlottesville, US, 1996.

[35] H. Haario and E. Saksman, "Simulated annealing process in general state space", *Advances in Applied Probability*, vol.23, 1991.

[36] S.R. White, "Concepts of scale in Simulated Annealing", *Proc. of IEEE International Conference on Computer Design*, 1984.

[37] M. Pierrot-Deseilligny and N. Paparoditis, "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*, vol.36 (part 1/w41), 2006.

[38] S. Roy and I.J. Cox, "A maximum-flow formulation of the N-camera stereo correspondence problem", *Proc. of IEEE International Conference on Computer Vision*, Bombay, India, 1998.

[39] U. Grenander and M.I. Miller, "Representations of Knowledge in Complex Systems", *Journal of the Royal Statistical Society*, vol.56, no.4, 1994.

[40] F. Han, Z.W. Tu and S.C. Zhu, "Range Image Segmentation by an Effective Jump-Diffusion Method", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.26, no.9, 2004.

[41] H. Hirschmüller, "Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information", *Proc. of IEEE Computer Vision and Pattern Recognition*, San Diego, US, 2005.

PLACE PHOTO HERE

Florent Lafarge received the BSc degree in aeronautics from the French School of Aeronautics in 2004 and the MSc degree in applied mathematics from the University of Toulouse, France, in the same year. From 2004 to 2007, he prepared the PhD degree at the INRIA and at the French Mapping Agency. He received the PhD degree in applied mathematics from the Ecole Nationale Supérieure des Mines de Paris, in 2007. From 2007 to 2008, he was a research visitor at the University of Auckland, New Zealand. He is now working as a postdoctoral researcher at the CERTIS laboratory of the Ecole Nationale des Ponts et Chaussées, Paris, France. His research interests are probabilistic modeling in image processing, stereo vision techniques and 3D reconstruction processes.

Xavier Descombes received the bachelors degree in telecommunications from the Ecole Nationale Superieure des Telecommunications de Paris (ENST) in 1989, the master of science degree in mathematics from the University of Paris VI in 1990, the PhD degree in signal and image processing from the ENST in 1993, and the "habilitation" from the University of Nice-Sophia Antipolis in 2004. He has been a postdoctoral researcher at ENST in 1994, at the Katolieke Universitat Leuven in 1995, and at the Institut National de Recherche en Informatique et en Automatique (INRIA) in 1996 and a visiting scientist in the Max Planck Institute of Leipzig in 1997. He is currently at INRIA as a research director. His research interests include Markov Random Fields, stochastic geometry, and stochastic modeling in image processing.

Josiane Zerubia has been a permanent research scientist at INRIA since 1989, and director of research since July 1995. She was head of the PASTIS remote sensing laboratory (INRIA Sophia-Antipolis) from mid-1995 to 1997. Since January 1998, she has been head of the Ariana research group (INRIA/CNRS/University of Nice), which also works on remote sensing. She has been adjunct professor at SUPAERO (ISAE) in Toulouse since 1999. Before that, she was with the Signal and Image Processing Institute of the University of Southern California (USC) in Los-Angeles as a postdoc. She also worked as a researcher for the LASSY (University of Nice/CNRS) from 1984 to 1988 and in the Research Laboratory of Hewlett Packard in France and in Palo-Alto (CA) from 1982 to 1984. She received the MSc degree from the Department of Electrical Engineering at ENSIEG, Grenoble, France in 1981, and the Doctor of Engineering degree, her PhD, and her "Habilitation", in 1986, 1988, and 1994 respectively, all from the University of Nice Sophia-Antipolis, France. She is a Fellow of the IEEE. She is a member of the IEEE IMDSP and IEEE BISP Technical Committees (SP Society). She was associate editor of IEEE Trans. on IP from 1998 to 2002; area editor of IEEE Trans. on IP from 2003 to 2006; guest co-editor of a special issue of IEEE Trans. on PAMI in 2003; and member-at-large of the Board of Governors of the IEEE SP Society from 2002 to 2004. She has also been a member of the editorial board of the French Society for Photogrammetry and Remote Sensing (SFPT) since 1998, of the International Journal of Computer Vision since 2004, and of the Foundation and Trends in Signal Processing since 2007. She has been associate editor of the on-line resource: Earthzine (IEEE CEO and GEOSS). Her current research interests are in image processing using probabilistic models and variational methods. She also works on parameter estimation and optimization techniques.

Marc Pierrot-Deseilligny has been working for twenty years in image analysis applied to remote sensing, photogrammetry and geographic information system. He obtained the master of science from Ecole Polytechnique in 1985, the PhD from the university of Paris V in 1994 and the "habilitation" in 1999. Since 2001 he has been working at French Mapping Agency (IGN). His current research interest are focussing on photogrametry applied to cultural heritage.