3D CITY MODELING BASED ON HIDDEN MARKOV MODEL

Florent Lafarge[1,2] *, *Xavier Descombes*[1], *Josiane Zerubia*[1]

[1]Ariana Reseach Group - INRIA
2004, routes des Lucioles, BP93
06902 Sophia Antipolis, Cedex France
E-mail=Name.Lastname@inria.fr

Marc Pierrot-Deseilligny[2]

[2]French Mapping Agency (IGN)
2/4 avenue Pasteur
94165 Saint-Mandé, Cedex France
E-mail=Name.Lastname@ign.fr

ABSTRACT

In this paper, we present an automatic method for the 3D building reconstruction from satellite images. The proposed approach consists in reconstructing buildings by assembling simple urban structures extracted from a grammar of 3D parametric models, as a "LEGO" game. First, the building footprints are extracted through sequences of quadrilaterals: it allows to define the problem as a causal process. Then, the 3D reconstruction stage is realized through a Hidden Markov Model and the optimal sequences of 3D parametric objects are found using the Viterbi algorithm.

***Index Terms*—** 3D modeling, building reconstruction, Hidden Markov Model, Digital Elevation Model

1. INTRODUCTION

Scene modeling and representations of 3D urban areas are critical in many applications such as urban planning, radiowave reachability tests for wireless communications or disaster recovery. Many automatic methods based on varied approaches such as parametric models [1], perceptual organization [2], ground based models [3] or polyhedral approach [4, 5], have been proposed. These methods provides convincing results using aerial images or/and laser scanning.
With the recent progress in the spatial domain, this problem can nowadays be tackled by the sub-metric satellite images. Such data are very interesting, especially for the developing countries where the aerial and terrestrial data acquisition is often difficult and the cadastral maps do not exist. However, satellite images have "relatively low" resolution and SNR to deal with 3D reconstruction problems (0.7 meter resolution for the satellite data used in this paper vs 0.08 meter resolution for aerial images used in [4, 5] for example). Consequently, contrary to the aerial methods, it is necessary to introduce important prior knowledge concerning urban structures in the satellite methods in order to face the difference of data quality.
In a previous work [6], a 3D building reconstruction method adapted to the satellite context has been proposed. This method suffers from several drawbacks: the generation of many artefacts, a lack of genericity, the tuning of many parameters and the computing time. In this paper, we propose a new method correcting all these defaults.
The inputs are Digital Elevation Models (DEMs), generated from satellite images, which are well adapted to global geometric descriptions. The proposed method is based on a structural approach: it consists in reconstructing buildings by assembling simple urban structures extracted from a grammar of 3D parametric models (as a "LEGO" game). First, the building footprints are extracted through sequences of quadrilaterals: it allows to define the problem as a causal process. Then, the 3D reconstruction stage is realized through a Hidden Markov Model (HMM) and the optimal sequences of 3D parametric objects are found using the Viterbi algorithm. Finally, results are shown on complex buildings.

*The author would like to thank the IGN and the French Space Agency (CNES) for partial financial support during his PhD. The authors thank the CNES for providing PLEIADES simulations.

2. BUILDING EXTRACTION

In a previous work [6], the building footprints were modeled by rectangle layouts (see Figure 1-(b)) estimated by marked point processes [7]. Such footprints generate many artefacts in the 3D reconstruction stage (see Figure 6-(d)).
We propose to regularize these rectangular footprints by improving the connections between the neighboring rectangles and detecting the roof height discontinuities inside the footprints (see Figure 1-(c)). More details concerning this regularization process are available in [8]. Each building is represented by sequences of connected quadrilaterals (i.e. quadrilaterals with common edges - quadrilaterals can even be triangles) which allow to avoid artefacts. This regularization especially introduces a causal process and so low computing time. In most of the cases, a building is represented by an unique sequence of connected quadrilaterals (which can be closed as we can see in Figure 1-top). However, some buildings owning complex roof junctions are modeled by several sequences (see Figure 1-bottom).
The obtained footprints allow to define the sequences of 3D-model supports. The 3D reconstruction stage is detailed in the following.

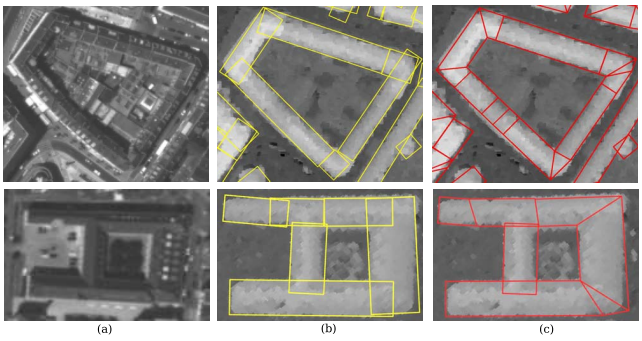


Fig. 1. **(a):** Satellite images of two buildings - **(b):** Rectangular footprints obtained by a marked point process - **(c):** Final footprints (after the regularization process).

3. 3D RECONSTRUCTION

3.1. Grammar of 3D-models

The contents of the grammar is a key point. If the grammar is too limited (such as in [6] where only flat and gable roof models are available), the method loses genericity. The proposed grammar, denoted by $\mathcal{M}$ and presented in Figure 2, allows to reconstruct a large range of buildings. It is composed of the most common roof types including monoplane roofs ($\mathcal{M}_{1x}$), multi-plane roofs ($\mathcal{M}_{2x}$) and curved roofs ($\mathcal{M}_{3x}$). Each roof type is modeled by a set of parameters knowing the quadrilateral footprint. H_t, H_g, ϕ and η are parameters which respectively represent the rooftop height, the roof gutter height, the orientation of the roof w.r.t. the quadrilateral base, and the choice between hipped and straight ends at the roof extremities. All the models of this grammar are detailed in [8].

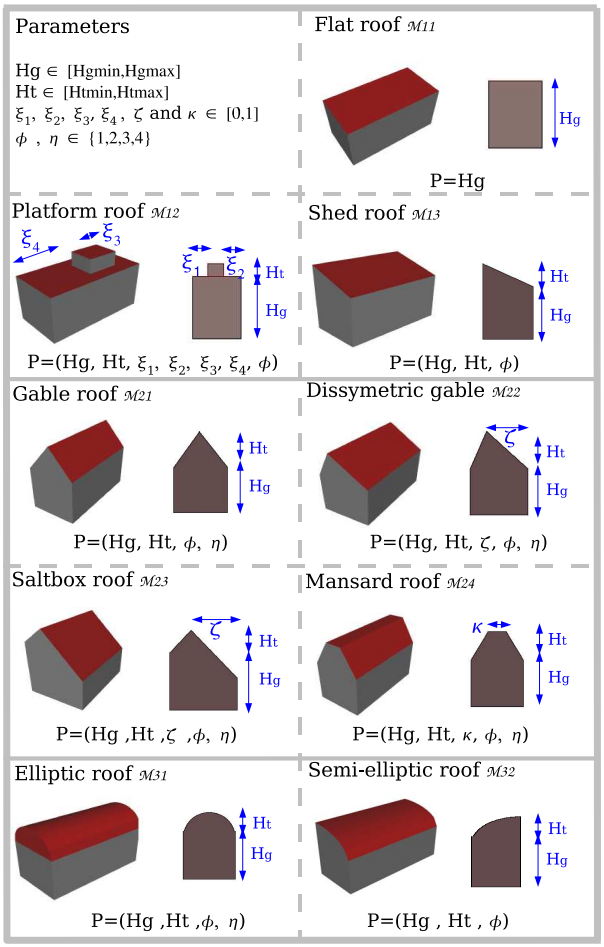


Fig. 2. The grammar of 3D-models (3D and profile views) - P, the set of parameters of each model.

3.2. Hidden Markov Model formulation

The HMM formulation requires some notations, summarized below:

- S, a set of sites and $I = \{\Lambda_s / s \in S\}$, a part of a DEM where Λ_s represents the intensity of the site s.
- $\mathcal{C}$, the sequence of quadrilaterals representing the building footprints associated with I. N represents the number of quadrilaterals of the sequence (see Figure 1-(c)).
- S_t, the subset of S whose sites are inside the t^{th} quadrilateral of $\mathcal{C}$.
- $y = (y_t)_{t=1..N}$, the data set where $y_t = \{\Lambda_s \in I / s \in S_t\}$
- x, a sequence of 3D-parametric objects knowing the quadrilateral footprints $\mathcal{C}$. $x = (x_t)_{t=1..N} = (m_t, \theta_t)_{t=1..N}$ where each object x_t is defined by both a roof type $\mathcal{M}_{m_t} \in \mathcal{M}$ and a set of parameters θ_t associated with $\mathcal{M}_{m_t}$. In the following, $x_t = (m_t, \theta_t)$ and $\mathcal{M}_{m_t}$ will be respectively called an object and a model.
- $\mathcal{S}_{x_t}$, the function from S_t to $\mathbb{R}$ which associates the roof altitude of the object x_t to each site of S_t.

Let us consider $(X_t, Y_t)_{t=1}^N$, a homogeneous hidden Markov chain where $X = (X_t)_t$ represents the hidden states (i.e the sequences x of 3D parametric objects) and $Y = (Y_t)_t$ corresponds to the sequence of observations (i.e. the data set y). Figure 3 represents the dependence graph of this chain. A requirement is to be able to build

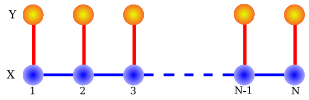


Fig. 3. Dependence graph of the hidden Markov chain $(X_t, Y_t)_{t=1}^N$

both the transition probability $P(X_{t+1} = x_{t+1} | X_t = x_t)$ and the local likelihood $P(Y_t = y_t / X_t = x_t)$ [1]. In the following, these two terms are detailed.

3.2.1. *Likelihood*

The local likelihood represents the probability of observing the data y_t knowing the object x_t. It can be expressed as :

$$P(y_t / x_t) = \frac{1}{Z_t} \exp -\Gamma^{\alpha}_{(t)}(\mathcal{S}_{x_t}, y_t) \quad (1)$$

where Z_t is a normalizing constant and $\Gamma^{\alpha}_{(t)}(.,.)$ is the distance from $\mathbb{R}^{card(S_t)} \times \mathbb{R}^{card(S_t)}$ to $\mathbb{R}$ defined by :

$$\forall a, b \in \mathbb{R}^{card(S_t)}, \Gamma^{\alpha}_{(t)}(a, b) = \left(\sum_{s \in S_t} |a_s - b_s|^{\alpha} \right)^{\frac{1}{\alpha}} \quad (2)$$

To sum-up, the local likelihood corresponds to the Z-error of the L_α norm between the DEM and the object. In practice, $\alpha = \frac{3}{2}$ is a good compromise between robustness and sensitivity to the DEM errors.

3.2.2. *Transition probability*

The transition probability, which corresponds to the introduction of prior knowledge concerning the assembling of urban structures, is a key point in the structural approach and allows to propose realistic building reconstructions. It is defined through interactions between neighboring objects. In a previous work [6], too many interactions have been set up. Their number must be minimal in order to keep

[1] To simplify notation, we will denote the discrete probability $P(X = x)$ as $P(x)$

robustness and avoid problems of parameter settings. We propose a simple and efficient transition probability which is defined through a unique interaction.
All the objects cannot be assembled together (for example, it is not possible to merge a curved roof with a mansard roof). So, it is necessary to define an assembling law. Two objects $x_t = (m_t, \theta_t)$ and $x_{t+1} = (m_{t+1}, \theta_{t+1})$ are said "**joinable**" (noted $x_t \sim_a x_{t+1}$) if they verify the three following points:

- $m_t = m_{t+1}$ or $\{m_t, m_{t+1}\} = \{11, 12\}$
- roof orientations ϕ_t and ϕ_{t+1} are the same
- the common edge of the t^{th} and $(t+1)^{th}$ quadrilateral footprints of $\mathcal{C}$ is not a roof height discontinuity.

The first point verifies that the two models belong to the same roof family. The second and third points test whether the rooftops of the two objects can be connected.
The transition probability consists in favoring the "joinable" objects. More precisely, in order to avoid the artefacts, the common parameters of two "joinable" objects are attracted to have similar values. To do so, the transition probability $P(x_{t+1}|x_t)$ is expressed through a Gibbs energy U (i.e. $P(x_{t+1}|x_t) = \frac{1}{Z} \exp -U(x_{t+1}, x_t)$) defined as follows:

$$U(x_{t+1}, x_t) = \mathbb{1}_{\{x_{t+1} \sim_a x_t\}} \beta g(x_{t+1}, x_t) \quad (3)$$

where $\beta \in \mathbb{R}^+$ is the parameter which weights the importance of the prior with respect to the likelihood. The function g, living in $[-1, 0]$, measures the distance between the common parameters of two "joinable" objects.

$$g(x_i, x_j) = \frac{D(x_i, x_j)}{D_{max}} - 1 = \frac{\sum_k \omega_k |\widetilde{\theta}_{i,(k)} - \widetilde{\theta}_{j,(k)}|}{D_{max}} - 1 \quad (4)$$

$\widetilde{\theta}_{i,(k)}$ and $\widetilde{\theta}_{j,(k)}$ correspond to the k^{th} element of the set of the common parameters of the objects x_i and x_j respectively. $D_{max} = \max_{x_i, x_j} D(x_i, x_j)$ represents 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 knowing the XY and Z resolutions and the sequence of quadrilaterals $\mathcal{C}$.
Figure 4 shows the principle of this interaction. If the two models belong to different roof types (for example a mansard roof model and an elliptic roof model on the top right) or if the two objects do not have compatible roof orientations (see bottom right), they will not be "joinable" and so, 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.

3.3. Optimization

The Viterbi algorithm [9] is especially adapted to find the optimal sequence of objects x of this HMM. The algorithm speed depends on the size of the state space (i.e the discretization step of the state space). We propose an optimization process composed of two stages. First, the Viterbi algorithm is applied on the "global" state space (i.e. with a large discretization step of metric accuracy). The best sequence allows to select the relevant models and parameter values and so to define a new state space which is reduced and focused on the objects of interest . Then, a second Viterbi algorithm is used on this detailed and reduced state space (i.e. with a small discretization step of submetric accuracy) in order to have an accurate solution. This process does not certify to obtain the optimal sequence of the global state space with the small discretization step. However, it gives in practice a convincing solution close to the optimal one by reducing computing time by a factor 10 compared to [6]. Concerning the algorithm initialization, the $P(x_1)$ are equiprobable which means no model of $\mathcal{M}$ is favored. For global scenes, all the sequences are independently optimized using this process.

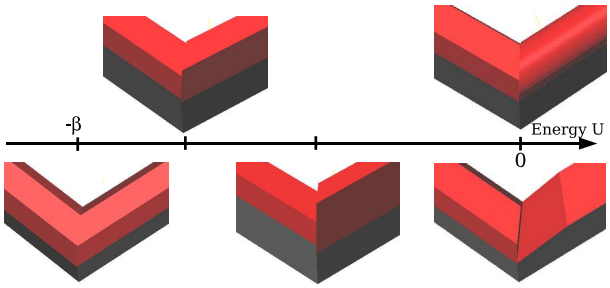


Fig. 4. Principle of the prior energy - examples of various interaction cases.

4. RESULTS

The results have been obtained from satellite images (PLEIADES simulations) on a dense downtown owning an important variety of roof types. DEMs have been generated from 3-view images using a multi-resolution implementation of a Cox and Roy optimal flow matching image algorithm [10]. The ground truths are raster images provided by the French Mapping Agency (IGN).
Figure 5 presents varied examples of reconstructed buildings (owning various roof types, roof height discontinuities, closed structures or/and complex roof junctions) associated with ground truths, satellite images and building footprints. These results are convincing in general. The five first examples provide good descriptions. Even if some details are omitted, the global shapes of buildings are respected compared to the ground truths and the generalization level is acceptable for satellite data in an automatic context. Few artefacts are generated on the 4^{th} and 5^{th} examples which means the causal process is adapted to buildings owning complex roof junctions. The last example shows the limits of this approach. Some footprints (especially the curved footprints) cannot be modeled by quadrilateral layouts. Figure 6 shows a result on a piece of downtown which is convincing compared to the modeling obtained by [6] on the same urban scene. The Z-RMSE on common building footprints has been reduced by 30% (i.e. Z-RMSE = 2.1 meters). The esthetic aspect of the modeling has also been improved. More results and evaluation details can be found in [8].

5. CONCLUSION

The proposed method provides convincing results: the global shape of buildings is respected and the generalization level is acceptable for satellite data in an automatic context. The causal process defined through a HMM allows to have low computing times and avoid artefacts. In future works, it would be interesting to estimate the parameter β of the HMM using EM algorithm. Moreover, we should evaluate the potential of this method on other kinds of cities such as typical north American urban areas.

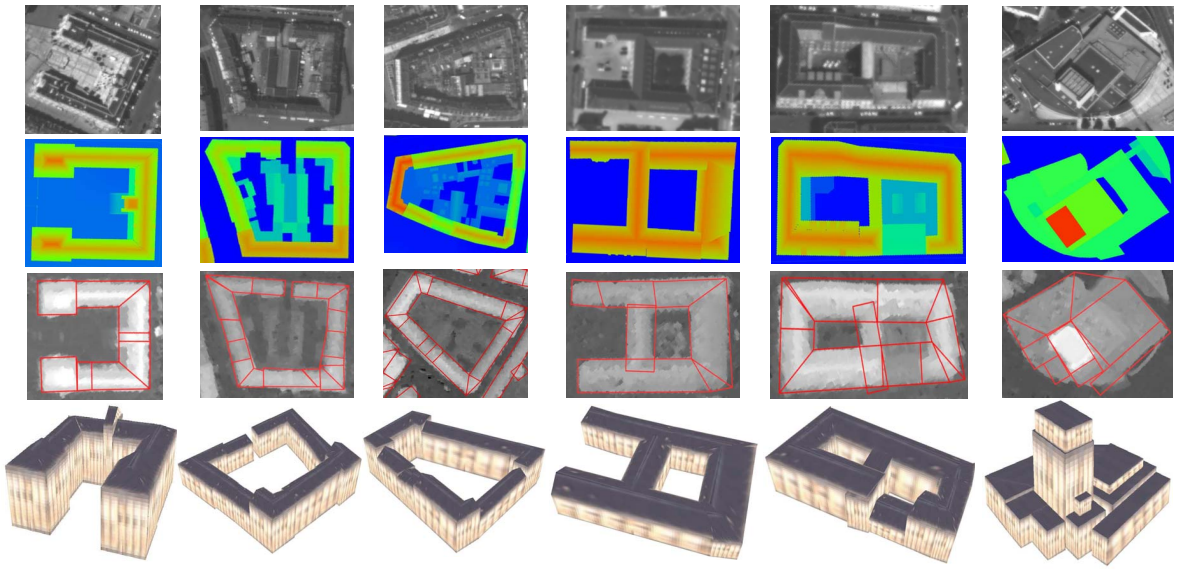
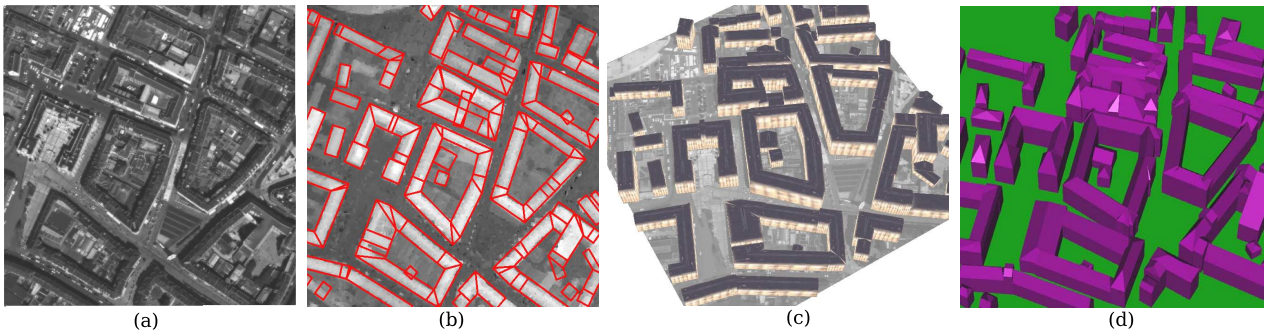


Fig. 5. Examples of reconstructed buildings (4^{th} row) associated with satellite images (1^{st} row), ground truths (2^{nd} row) and quadrilateral footprints (3^{rd} row).



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

Fig. 6. Satellite image of a piece of downtown **(a)** - quadrilateral footprints **(b)** - 3D results **(c)** - 3D results obtained by [6] **(d)**.

6. REFERENCES

[1] U. Weidner and W. Forstner, "Fowards Automatic Building Reconstruction from HR Digital Elevation Models," *Journal of Photogrammetry and Remote Sensing*, vol. 50(4), 1995.

[2] R. Nevatia and K. Price, "Automatic and interactive modeling of buildings in urban environments from aerial images," in *IEEE Proc. ICIP*, New York, 2002.

[3] C. Früh and A. Zakhor, "Constructing 3D city models by merging aerial and ground views.," *IEEE Trans. Computer Graphics and Applications*, vol. 23, no. 6, pp. 52–61, 2003.

[4] C. Baillard and A. Zisserman, "A plane-sweep strategy for the 3D reconstruction of buildings from multiple images," in *19th ISPRS Congress*, Amsterdam, The Netherlands, 2000.

[5] S. Scholze, T. Moons, and L.J. Van Gool, "A probabilistic approach to building roof reconstruction using semantic labelling," in *Proceedings of the 24th DAGM Symposium on Pattern Recognition*, London, UK, 2002, pp. 257–264.

[6] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny, "An automatic building reconstruction method : A structural approach using HR satellite images," in *IEEE Proc. ICIP*, Atlanta, USA, Oct 2006.

[7] M. Ortner, X. Descombes, and J. Zerubia, "Building outline extraction from DEMs using marked point processes," *International Journal of Computer Vision*, 2006, online version.

[8] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny, "A structural approach for 3D building reconstruction," Research Report 6048, INRIA, France, 2006.

[9] A.J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. on Information Theory*, vol. 13, pp. 260–269, 1967.

[10] S. Roy and I.J. Cox, "A maximum-flow formulation of the n-camera stereo correspondence problem," in *IEEE Proc. ICCV*, Bombay, 1998.