COMPETING 3D PRIORS FOR OBJECT EXTRACTION IN REMOTE SENSING DATA

Konstantinos Karantzalos and Nikos Paragios

Ecole Centrale de Paris
Grande Voie des Vignes, 92295
Chatenay-Malabry, France
{konstantinos.karantzalos, nikos.paragios}@ecp.fr
http://users.ntua.gr/karank/Demos.html

Commission III

KEY WORDS: Computer Vision, Pattern Recognition, Variational Methods, Model-Based, Evaluation, Voxel-Based

ABSTRACT:

A recognition-driven variational framework was developed for automatic three dimensional object extraction from remote sensing data. The essence of the approach is to allow multiple 3D priors to compete towards recovering terrain objects' position and 3D geometry. We are not relying, only, on the results of an unconstrained evolving surface but we are forcing our output segments to inherit their 3D shape from our prior models. Thus, instead of evolving an arbitrary surface we evolve the selected geometric shapes. The developed algorithm was tested for the task of 3D building extraction and the performed pixel- and voxel-based quantitative evaluation demonstrate the potentials of the proposed approach.

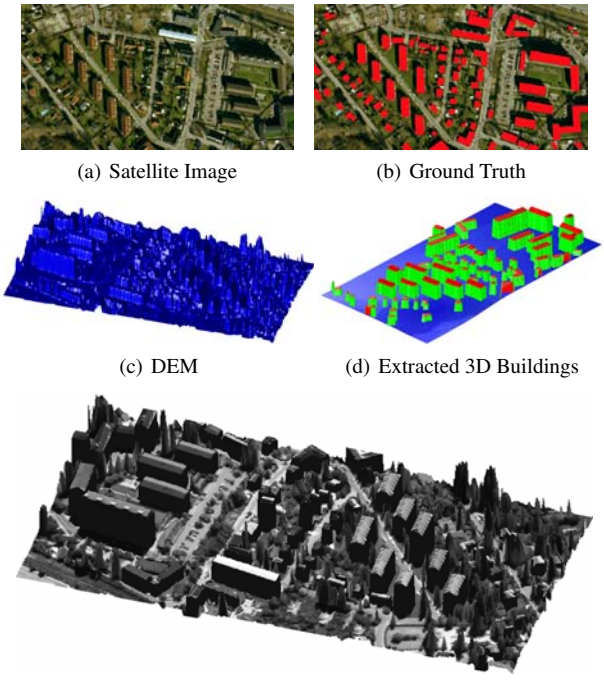
1 INTRODUCTION

Although, current remote sensing sensors can provide an updated and detailed source of information related to terrain analysis, the lack of automated operational procedures regarding their processing impedes their full exploitation. By using standard techniques based, mainly, on spectral properties, only the lower resolution earth observation data can be effectively classified. Recent automated approaches are not, yet, functional and mature enough for supporting massive processing on multiple scenes of high- and very high resolution data.

On the other hand, modeling urban and peri-urban environments with engineering precision, enables people and organizations involved in the planning, design, construction and operations lifecycle, in making collective decisions in the areas of urban planning, economic development, emergency planning, and security. In particular, the emergence of applications like games, navigation, e-commerce, spatial planning and monitoring of urban development has made the creation and manipulation of 3D city models quite valuable, especially at large scale.

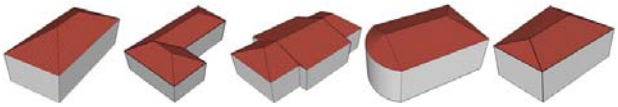
In this perspective, optimizing the automatic information extraction of terrain features/objects from new generation satellite data is of major importance. For more than a decade now, research efforts are based on the use of a single image, stereopairs, multiple images, digital elevation models (DEMs) or a combination of them. One can find in the literature several model-free or model-based algorithms towards 2D and 3D object extraction and reconstruction [(Hu et al., 2003),(Baltsavias, 2004),(Suveg and Vosselman, 2004),(Paparoditis et al., 2006),(Drauschke et al., 2006),(Rottensteiner et al., 2007),(Sohn and Dowman, 2007),(Verma et al., 2006),(Lafarge et al., 2007),(Karantzalos and Paragios, 2009) and the references therein]. Despite this intensive research, we are, still, far from the goal of the initially envisioned fully automatic and accurate reconstruction systems (Brenner, 2005),(Zhu and Kanade (Eds.), July, 2008),(Mayer, 2008). Processing remote sensing data, still, poses several challenges.

In this paper, we extend our recent 2D prior-based formulations (Karantzalos and Paragios, 2009) aiming at tackling the problem of automatically and accurately extracting 3D terrain objects from optical and height data. Multiple 3D competing priors are considered transforming reconstruction to a labeling and an estimation problem. In such a context, we fuse images and DEMs towards recovering a 3D prior model. We are experimenting with buildings but, similarly, any other terrain object can be modeled. Our formulation allows data with the higher spatial resolution to constrain properly the footprint detection in order to achieve the optimal spatial accuracy (Figure 1). Therefore, we are proposing a variational functional that encodes a fruitful synergy between observations and multiple 3D grammar-based models. Our models refer to a grammar, which consists of typologies of 3D shape priors (Figure 2). In such a context, firstly one has to select the

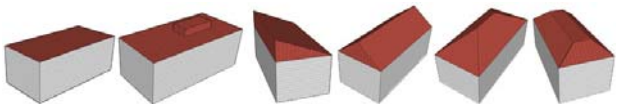


(a) Satellite Image (b) Ground Truth (c) DEM (d) Extracted 3D Buildings (e) Reconstructed Scene

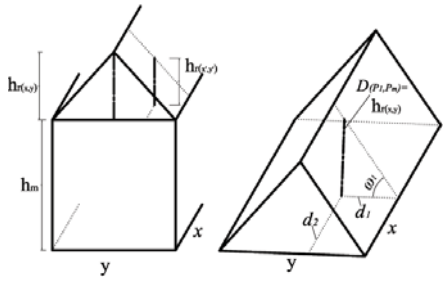
Figure 1: 3D Building Extraction through Competing 3D Priors



(a) Prior Building Models ($\tilde{\Phi}_{\mathbf{i},\mathbf{j}}$): $\mathbf{i}$ determines the shape of the footprint and $\mathbf{j}$ the roof type



(b) The family $\tilde{\Phi}_{1,\mathbf{j}}$ which has a rectangular footprint ($\mathbf{i} = 1$).



(c) Building's main height h_m and roofs height $h_r(x, y)$

Figure 2: Hierarchical Grammar-Based 3D Prior Models. The case of Building Modeling: Building's footprint is determined implicitly from the E_{2D}. h_m and $h_r(x, y)$ are recovered for every point (E_{3D}) and thus all the different type of roofs $\mathbf{j}$ are modeled.

most appropriate model and then determine the optimal set of parameters aiming to recover scene's geometry (Figure 1). The proposed objective function consists of two segmentation terms that guide the selection of the most appropriate typology and a third DEM-driven term which is being conditioned on the typology. Such a prior-based recognition process can segment both rural and urban regions (similarly to (Matei et al., 2008)) but is able, as well, to overcome detection errors caused by the misleading low-level information (like shadows or occlusions), which is a common scenario in remote sensing data.

Our goal was to develop a single generic framework (with no step-by-step procedures) that is able to efficiently account for multiple 3D building extraction, no matter if their number or shape is *a priori* familiar or not. In addition, since usually for most sites multiple aerial images are missing, our goal was to provide a solution even with the minimum available data, like a single panchromatic image and an elevation map (produced either with classical photogrammetric multi-view stereo techniques either from LIDAR or INSAR sensors), contrary to approaches that were designed to process multiple aerial images or multispectral information and cadastral maps (like in (Suveg and Vosselman, 2004),(Rottensteiner et al., 2007),(Sohn and Dowman, 2007)), data which much ease scene's classification. Doing multiview stereo, using simple geometric representations like 3D lines and planes or merging data from ground sensors was not our interest here. Moreover, contrary to (Zebedin et al., 2008), the proposed, here, variational framework does not require as an input dense height data, dense image matching processes and *a priori* given 3D line segments or a rough segmentation.

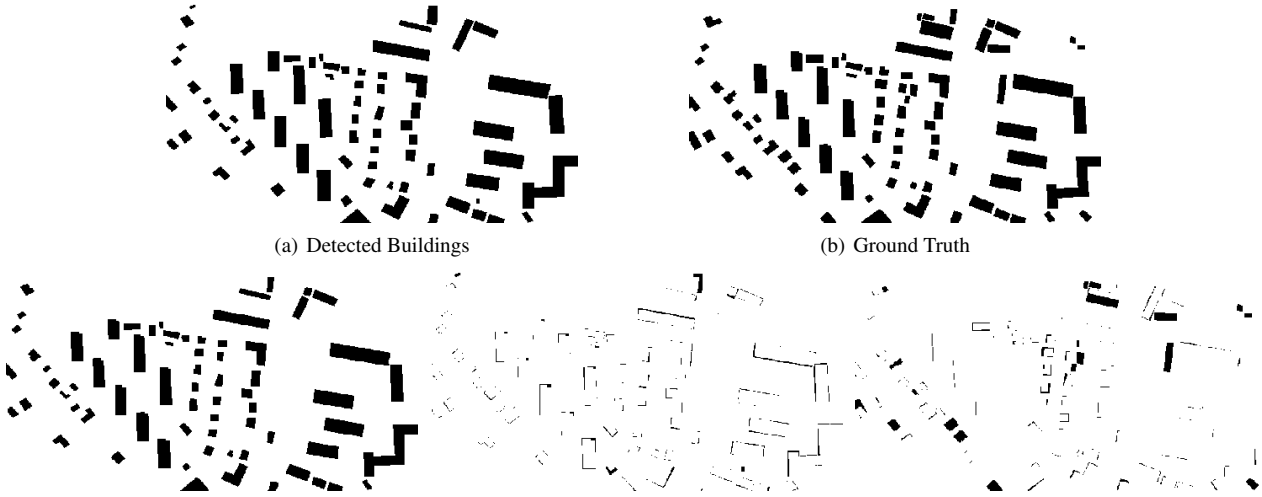
2 MODELING TERRAIN OBJECTS WITH 3D PRIORS

Numerous 3D model-based approaches have been proposed in literature. Statistical approaches (Paragios et al., 2005), aim to describe variations between the different prior models by measuring the distribution of the parameter space. These models are capable to model building with rather repeating structure and of limited complexity. In order to overcome this limitation, methods using generic, parametric, polyhedral and structural models have been considered (Jaynes et al., 2003),(Kim and Nevatia, 2004),(Suveg and Vosselman, 2004),(Dick et al., 2004),(Wilczkowiak et al., 2005),(Forlani et al., 2006),(Lafarge et al., 2007). The main strength of these models is their expressional power in terms of complex architectures. On the other hand, inference between the models and observations is rather challenging due to the important dimension of the search space. Consequently, these models can only be considered in a small number. More recently, procedural modeling of architectures was introduced and vision-based reconstruction in (Muller et al., 2007) using mostly facade views. Such a method recovers 3D using an L-system grammar (Muller et al., 2006) that is a powerful and elegant tool for content creation. Despite the promising potentials of such an approach, one can claim that the inferential step that involves the derivation of models parameters is still a challenging problem, especially when the grammar is related with the building detection procedure.

Hierarchical representations are a natural selection to address complexity while at the same time recover representations of acceptable resolution. Focusing on buildings, our models involve two components, the type of footprint and the type of roof (Figure 2). Firstly, we structure our prior models space $\tilde{\Phi}$ by ascribing the same pointer $\mathbf{i}$ to all models that belong to the family with the same footprint. Thus, all buildings that can be modeled with a rectangular footprint are having the same index value $\mathbf{i}$. Then, for every family (i.e. every $\mathbf{i}$) the different types of building tops (roofs) are modeled by the pointer $\mathbf{j}$ (Figure 2b) Under this hierarchy $\tilde{\Phi}_{\mathbf{i},\mathbf{j}}$, the priors database can model from simple to very complex building types and can be easily enriched with more complex structures. Such a formulation is desirously generic but forms a huge search space. Therefore, appropriate attention is to be paid when structuring the search step.

Given the set of footprint priors, we assume that the observed building is a homographic transformation of the footprint. Given, the variation of the expressiveness of the grammar, and the degrees of freedom of the transformation, we can now focus on the 3D aspect of the model. In such a context, only building's main height h_m and building's roof height $h_r(x, y)$ at every point need to be recovered. The proposed typology for such a task is shown in Figure 2. It refers to the rectangular case but all the other families can respectively be defined. More complex footprints, with usually more than one roof types, are decomposed to simpler parts which can, therefore, similarly recovered. Given an image $\mathcal{I}(x, y)$ at domain (bounded) $\Omega \in R^2$ and an elevation map $\mathcal{H}(x, y)$ -which can be seen both as an image or as a triangulated point cloud- let us denote by h_m the main building's height and by P_m the horizontal building's plane at that height. We proceed by modeling all building roofs (flat, shed, gable, etc.) as a combination of four inclined planes. We denote by P_1, P_2, P_3 and P_4 these four roof planes and by $\omega_1, \omega_2, \omega_3$ and ω_4, respectively, the four angles between the horizontal plane h_m and each inclined plane (Figure 2). Every point in the roof rests strictly on one of these inclined planes and its distance with the horizontal plane is the minimum compared with the ones formed by the other three planes.figure

With such a grammar-based description the five unknown parameters to be recovered are: the main height h_m (which has a constant value for every building) and the four angles ω. In this way all -but two- types of buildings tops/roofs can be modeled. For example, if all angles are different we have a totally dissymmetric roof (Figure 2b - $\tilde{\Phi}_{\mathbf{1},\mathbf{5}}$), if two opposite angle are zero we have a



(a) Detected Buildings (b) Ground Truth

(c) Horizontal True Positives (HTP) (d) Horizontal False Positives (HFP) (e) Horizontal False Negatives (HFN)

Figure 3: Horizontal Qualitative Evaluation: The recognition-driven process efficiently detects, in an unsupervised manner, scene buildings and recovers their 3D geometry.

gable-type one ($\tilde{\Phi}_{\mathbf{1},\mathbf{4}}$) and if all are zero we have a flat one ($\tilde{\Phi}_{\mathbf{1},\mathbf{1}}$). The platform and the gambrel roof types can not be modeled but can be easily derived in cases where the fit energy metric is assumed on local minima. The platform one ($\tilde{\Phi}_{\mathbf{1},\mathbf{2}}$), for instance, is the case where all angles have been recovered with small values and a search around their intersection point will estimate the dimensions of the rectangular-shape box above main roof plane P_m. With the aforementioned formulations, instead of searching for the best among $\mathbf{i}\mathbf{x}\mathbf{j}$ (e.g. $5\text{x}6 = 30$) models, their hierarchical grammar and the appropriate defined energy terms (detailed in the following section) are able to cut down effectively the solutions space.

3 MULTIPLE 3D PRIORS IN COMPETITION EXTRACTING MULTIPLE OBJECTS

Let us consider an image ($\mathcal{I}$) and the corresponding digital elevation map ($\mathcal{H}$). In such a context, one has to separate the desired for extraction objects from the background (natural scene) and, then, determine their geometry. The first segmentation task is addressed through the deformation of a initial surface $\phi : \Omega \rightarrow \mathcal{R}^+$ that aims at separating the natural components of the scene from the man-made parts. Assuming that one can establish correspondences between the pixels of the image and the ones of the DEM, the segmentation can be solved in both spaces through the use of regional statistics. In the visible image we would expect that buildings are different from the natural components of the scene. In the DEM, one would expect that man-made structures will exhibit elevation differences from their surroundings. Following the formulations of (Karantzalos and Paragios, 2009), these two assumptions can be used to define the following segmentation function

$$\begin{aligned} E_{seg}(\phi) \; = \; & \int |\nabla \phi(\mathbf{x})| \, d\mathbf{x} \\ + \; & \int_{\Omega} H_{\epsilon}(\phi) \, r_{obj} \left(\mathcal{I}(\mathbf{x})\right) \; + [1 - H_{\epsilon}(\phi)] \, r_{bg} \left(\mathcal{I}(\mathbf{x})\right) \; d\mathbf{x} \\ + \rho \int_{\Omega} & H_{\epsilon}(\phi) \, r_{obj} \left(\mathcal{H}(\mathbf{x})\right) \; + [1 - H_{\epsilon}(\phi)] \, r_{bg} \left(\mathcal{H}(\mathbf{x})\right) \; d\mathbf{x} \end{aligned} \tag{1}$$

where H is the Heaviside, r_{obj} and r_{bg} are *object* and *background* positive monotonically decreasing data-driven functions driven from the grouping criteria. The simplest possible approach would involve the Mumford-Shah approach that aims at separating the means between the two classes. Above equation can be straightforwardly extended in order to deal with other optical or radar data like for example in cases where multi- or hyper-spectral remote sensing data are available.

Furthermore, instead of relying only on the results of an unconstrained evolving surface, we are forcing our output segments to inherit their 2D shape from our prior models. Thus, instead of evolving an arbitrary surface we evolve selected geometric shapes and the 2D prior-based segmentation energy term takes the following form:

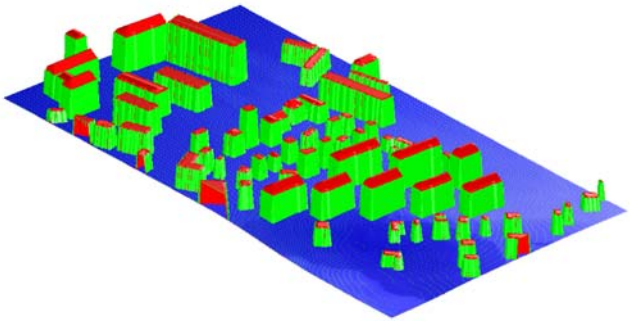
$$\begin{aligned} & E_{2D}(\phi, \mathcal{T}_i, \mathbf{L}) = \\ & \sum_{i=1}^{m-1} \int \left(\frac{H_{\epsilon}(\phi(\mathbf{x})) - H_{\epsilon}(\tilde{\phi}_i \left(\mathcal{T}_i(\mathbf{x})\right))}{\sigma_i} \right)^2 x_i(\mathbf{L}(\mathbf{x})) d\mathbf{x} \; + \\ & \int \lambda^2 x_m(\mathbf{L}(\mathbf{x})) d\mathbf{x} + \rho \sum_{i=1}^{m} \int |\nabla L(\mathbf{x})| d\mathbf{x} \end{aligned} \tag{2}$$

with the two parameters $\lambda, \rho > 0$ and the k-dimensional labeling formulation able for the dynamic labeling of up to $m = 2^k$ regions.

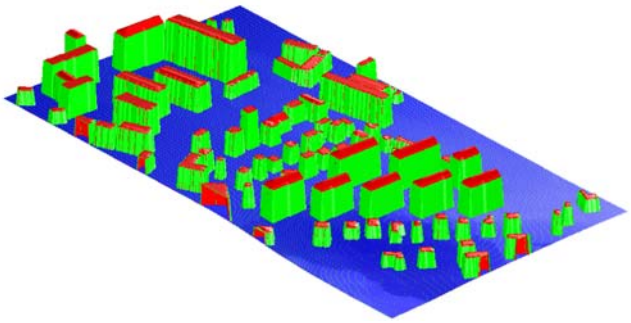
In this way, during optimization the number of selected regions $m = 2^k$ depends on the number of the possible building segments according to ϕ and thus the k-dimensional labeling function $\mathbf{L}$ obtains incrementally multiple instances. It should be, also, mentioned that the initial pose of the priors are not known. Such a formulation $E_{seg} + E_{2D}$ allows data with the higher spatial resolution to constrain properly the footprint detection in order to achieve the optimal spatial accuracy. Furthermore, it solves segmentation simultaneously in both spaces (image and DEM) and addresses fusion in a natural manner.figure

3.1 Grammar-based Object Reconstruction

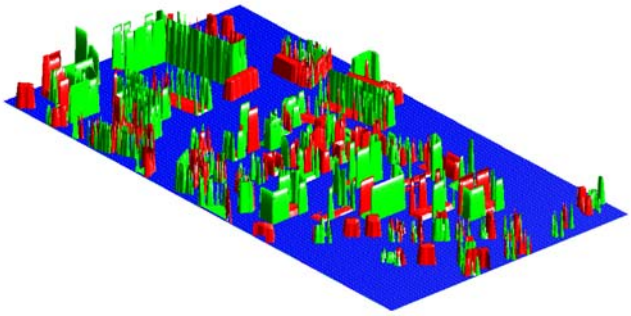
In order to determine the 3D geometry of the buildings, one has to estimate the height of the structure with respect to the ground and the orientation angles of the roof components i.e. five unknown parameters: the building's main height h_m which is has a constant value for every building and the four angles ω of the



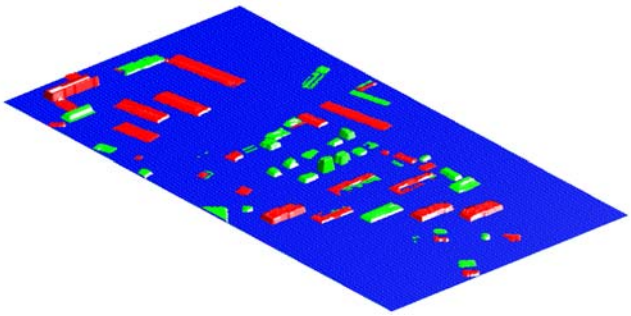
(a) 3D View of the Extracted Buildings



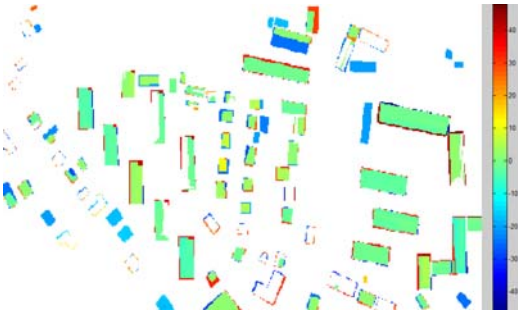
(b) 3D View of the Ground Truth



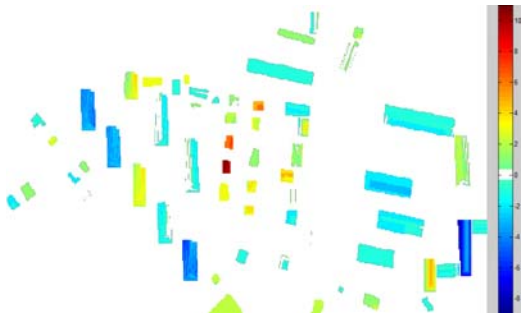
(c) Vertical/Hypsometric Difference (absolute values)



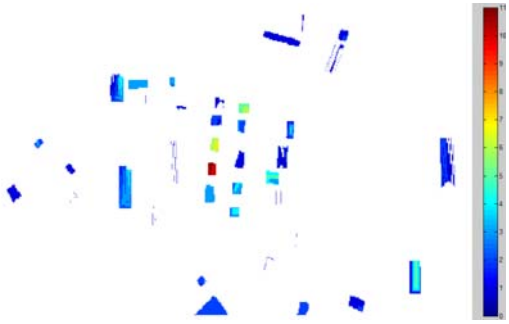
(d) Vertical Difference among the HTP (absolute values)



(e) Vertical Difference



(f) Vertical Difference among the HTP



(g) Vertical False Positives among the HTP

(h) Vertical False Negatives among the HTP

Figure 4: Vertical/Hypsometric Difference between the Extracted Buildings and the Ground Truth

roof's inclined planes ($\Theta_i = (h_m, \omega_1, \omega_2, \omega_3, \omega_4)$). These four angles (Figure 2) along with the implicitly derived dimensions of every building's footprint (from E_{2D}) can define the roof's height at every point (pixel) $h_r(x, y)$:

$$\begin{aligned} h_r(x,y) = \\ \min\left[\mathcal{D}(P_1,P_m); \mathcal{D}(P_2,P_m); \mathcal{D}(P_3,P_m); \mathcal{D}(P_4,P_m)\right] \\ = \min\left[d_1 \tan\omega_1;\ d_2 \tan\omega_2;\ d_3 \tan\omega_3;\ d_4 \tan\omega_4\right] \end{aligned} \quad (3)$$

where $\mathcal{D}$: is the perpendicular distance between the horizontal plane P_m and roof's inclined plane $P_{1:4}$. The distance for e.g. between P_1 and P_m in Figure 2 is the actual roof's height at that point (x, y) and can be calculated as the product of the tangent of plane's P_1 angle and the horizontal distance d_1 lying on plane P_m. $\mathcal{D}(P_1, P_m)$ is, also, the minimum distance in that specific point comparing with the ones that are formed with the other three inclined planes.

Utilizing the 3D information from $\mathcal{H}$ -either from point clouds or from a height map- the corresponding energy E_{3D} that recovers

our five unknowns for a certain building i has been formulated as follows:

$$E_{3D}(\Theta_i) = \sum_{i=1}^{m} \int_{\Omega_i} \left(h_{m_i} + h_{r_i}(\mathbf{x}) - \mathcal{H}(\mathbf{x})\right)^2 \, d\mathbf{x} \quad (4)$$

Each prior that has been selected for a specific region is forced to acquire such a geometry so as at every point its total height matches the one from the available DEM. It's a heavily constrained formulation and thus robust. The introduced, here, recognition driven framework now takes the following form in respect to ϕ, $\mathcal{T}_i$, $\mathbf{L}$ and Θ_i:

$$E_{total} \;=\; E_{seg}(\phi) + \mu E_{2D}(\phi, \mathcal{T}_i, \mathbf{L}) + \mu E_{3D}(\Theta_i) \quad (5)$$

The energy term E_{seg} addresses fusion in a natural way and solves segmentation ϕ in both $\mathcal{I}(\mathbf{x})$ and $\mathcal{H}(\mathbf{x})$ spaces. The term E_{2D} estimates which family of priors, i.e which 2D footprint $\mathbf{i}$, under any projective transformation T_i best fit at each segment ($\mathbf{L}$). Finally, the energy E_{3D} recovers the 3D geometry Θ_i of every prior by estimating building's h_m and h_r heights.

4 QUALITATIVE AND QUANTITATIVE ASSESSMENT OF THE PRODUCED 3D MODELS

The quality assessment of 3D data [(Meidow and Schuster, 2005),(Sargent et al., 2007) and their references therein] involves the assessment of both the geometry and topology of the model. During our experiments the quantitative evaluation was performed based on the 3D ground truth data which were derived from a manual digitization procedure. The standard quantitative measures of Completeness (detection rate), Correctness (under-detection rate) and Quality (a normalization between the previous two) were employed. To this end, the quantitative assessment is divided into two parts: Firstly, for the evaluation of the extracted 2D boundaries i.e. the horizontal localization of the building footprints (Figure 3) and secondly, for the evaluation of the hypsometric differences i.e. the vertical differences between the extracted 3D building and the ground truth (Figure 4).

In order to assess the horizontal accuracy of the extracted building footprints the measures of Horizontal True Positives (HTP), Horizontal False Positives (HFP) and Horizontal False Negatives (HFN), were calculated.

$$\text{2D Completeness} = \frac{\text{area of correctly detected segments}}{\text{area of the ground truth}} = \frac{HTP}{HTP + HFN}$$

$$\text{2D Correctness} = \frac{\text{area of correctly detected segments}}{\text{area of all detected segments}} = \frac{HTP}{HTP + HFP}$$

$$\text{2D Quality} = \frac{HTP}{HTP + HFP + HFN}$$

Moreover, for the evaluation of the hypsometric differences between the extracted buildings and the ground truth the measures of Vertical True Positives (VTP), Vertical False Positives (VFP) and Vertical False Negatives (VFN) were, also, calculated. The VTP are the voxels among, the corresponding Horizontal True Positive pixels, that have the same altitude with the ground truth. Note that Horizontal True Positives may correspond *(i)* to voxels with the same altitude as in the ground truth (VTP) and *(ii)* to voxels with a lower or higher altitude than the ground truth (VFN and VFP, respectively). Thus, the Vertical False Positives are the voxels with an hypsometric difference with the ground truth, containing all the corresponding voxels from the HFP and the corresponding ones from the HTP (those with a higher altitude than the ground truth). Respectively, the Vertical False Negatives are the voxels with an hypsometric difference with the ground truth, containing all the corresponding voxels from the HFN and the corresponding ones from the HTP (those with a lower altitude than the ground truth). To this end, the 3D quantitative assessment was based on the measures of the 3D Completeness (detection rate), 3D Correctness (under-detection rate) and 3D Quality (a normalization between the previous two), which were calculated in the following way:

2D Quantitative Measures		
Completeness	Correctness	Quality
0.84	0.90	0.76

3D Quantitative Measures		
Completeness	Correctness	Quality
0.86	0.86	0.77

Table 1: Pixel- and Voxel-Based Quality Assessment

$$\text{3D Completeness} = \frac{VTP}{VTP + VFN}$$

$$\text{3D Correctness} = \frac{VTP}{VTP + VFP}$$

$$\text{3D Quality} = \frac{VTP}{VTP + VFP + VFN}$$

The developed algorithm has been applied to a number of scenes where remote sensing data was available. The algorithm managed in all cases to accurately recover their footprint and overcome low-level misleading information due to shadows, occlusions, etc. In addition, despite the conflicting height similarity between the desired buildings, the surrounding trees and the other objects the developed algorithm managed to robustly recover their 3D geometry as the appropriate priors were chosen (Figure 1). This complex landscape contains a big variety of texture patterns, more than 80 buildings of different types (detached single family houses, industrial buildings, etc) and multiple other objects of various classes. Two aerial images (with a ground resolution of appx. 0.5m) and a the coarser digital surface model (of appx. 1.0m ground resolution) were available. The robustness and functionality of the proposed method is illustrated, also, on Figures 3 and 4, where one can, clearly, observe the Horizontal and the Vertical True Positives, respectively. The proposed generic variational framework managed to accurately extract the 3D geometry of scene's buildings, searching among various footprint shapes and various roof types. The performed quantitative evaluation reported an overall horizontal detection correctness of 90% and an overall horizontal detection completeness of 84% (Table 1).

In Figure 4c, the hypsometric/vertical difference between the extracted buildings and the ground truth is shown. With a red color are the VFN voxels and with a green color the VFP ones. Similarly, at Figure 4c where the -corresponding among the HTP pixels- VFN and VFP voxels are shown. The performed quantitative evaluation reported an overall 3D completeness and correctness of appx. 86% (Table 1).

5 CONCLUSIONS AND FUTURE WORK

We have developed a generalized variational framework which addresses large-scale reconstruction through information fusion and competing grammar-based 3D priors. We have argued that our inferential approach significantly extends previous 3D extraction and reconstruction efforts by accounting for shadows, occlusions and other unfavorable conditions and by effectively narrowing the space of solutions due to our novel grammar representation and energy formulation. The successful recognition-driven results along with the reliable estimation of buildings 3D geometry suggest that the proposed method constitutes a highly promising tool for various object extraction and reconstruction tasks.

Our a framework can be easily extended to process spectral information, by formulating respectively the region descriptors and to account for other types of buildings or other terrain features. For real-time applications, the labeling function straightforwardly allows a parallel computing formulation by concurrently recovering the transformations for every region. In order to address the sub-optimality of the obtained solution, the use of the compressed sensing framework by collecting a comparably small number of measurements rather than all pixel values is currently under investigation. Last, but not least introducing hierarchical procedural grammars can reduce the complexity of the prior model and provide access to more efficient means of optimization.

ACKNOWLEDGEMENTS

This work has been partially supported from the Conseil General de Hauts-de-Seine and the Region Ile-de-France under the TERRA NUMERICA grant of the Pole de competitivite CapDigital.

REFERENCES

Baltsavias, E., 2004. 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 58, pp. 129–151.

Brenner, C., 2005. Building reconstruction from images and laser scanning. International Journal of Applied Earth Observation and Geoinformation 6, pp. 187–198.

Dick, A. R., Torr, P. H. S. and Cipolla, R., 2004. Modelling and interpretation of architecture from several images. International Journal of Computer Vision 60(2), pp. 111–134.

Drauschke, M., Schuster, H.-F. and Förstner, W., 2006. Detectibility of buildings in aerial images over scale space. In: ISPRS Symposium of Photogrammetric Computer Vision, Vol. XXXVInumber Part 3, pp. 7–12.

Forlani, G., Nardinocchi, C., Scaioni, M. and Zingaretti, P., 2006. Complete classification of raw LIDAR data and 3D reconstruction of buildings. Pattern Anal. Appl. 8(4), pp. 357–374.

Hu, J., You, S. and Neumann, U., 2003. Approaches to large-scale urban modeling. IEEE Computer Graphics and Applications 23(6), pp. 62–69.

Jaynes, C., Riseman, E. and Hanson, A., 2003. Recognition and reconstruction of buildings from multiple aerial images. Computer Vision and Image Understanding 90(1), pp. 68–98.

Karantzalos, K. and Paragios, N., 2009. Recognition-Driven 2D Competing Priors Towards Automatic and Accurate Building Detection. IEEE Transactions on Geoscience and Remote Sensing 47(1), pp. 133–144.

Kim, Z. and Nevatia, R., 2004. Automatic description of complex buildings from multiple images. Computer Vision and Image Understanding 96(1), pp. 60–95.

Lafarge, F., Descombes, X., Zerubia, J. and Pierrot-Deseilligny, M., 2007. 3D city modeling based on hidden markov model. In: Proc. IEEE International Conference on Image Processing (ICIP), Vol. II, pp. 521–524.

Matei, B.C.and Sawhney, H., Samarasekera, S., Kim, J. and Kumar, R., 2008. Building segmentation for densely built urban regions using aerial lidar data. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8.

Mayer, H., 2008. Object extraction in photogrammetric computer vision. ISPRS Journal of Photogrammetry and Remote Sensing 63(2), pp. 213–222.

Meidow, J. and Schuster, H., 2005. Voxel-based quality evaluation of photogrammetric building acquisitions. In: ISPRS International archives of photogrammetry, remote sensing and spatial information sciences (Stilla U, Rottensteiner F, Hinz S (Eds)), Vol. XXXVI, Part 3/W24.

Muller, P., Wonka, P., Haegler, S., Ulmer, A. and Gool, L., 2006. Procedural modeling of buildings. Proceedings of ACM SIGGRAPH / ACM Transactions on Graphics 25(3), pp. 614–623.

Muller, P., Zeng, G., Wonka, P. and Gool, L., 2007. Image-based procedural modeling of facades. Proceedings of ACM SIGGRAPH 2007 / ACM Transactions on Graphics.

Paparoditis, N., Souchon, J.-P., Martinoty, G. and Pierrot-Deseilligny, M., 2006. High-end aerial digital cameras and their impact on the automation and quality of the production workflow. ISPRS Journal of Photogrammetry and Remote Sensing 60(6), pp. 400–412.

Paragios, N., Chen, Y. and Faugeras, O., 2005. Handbook of Mathematical Models of Computer Vision. Springer.

Rottensteiner, F., Trinder, J., Clode, S. and Kubik, K., 2007. Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis. ISPRS Journal of Photogrammetry and Remote Sensing 62(2), pp. 135–149.

Sargent, I., Harding, J. and Freeman, M., 2007. Data Quality in 3D: Gauging Quality Measures From Users' Requirements. In: International Symposium on Spatial Quality, Endchede, Netherlands.

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

Suveg, I. and Vosselman, G., 2004. Reconstruction of 3D building models from aerial images and maps. ISPRS Journal of Photogrammetry and Remote Sensing 58, pp. 202–224.

Verma, V., Kumar, R. and Hsu, S., 2006. 3D building detection and modeling from aerial lidar data. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2213–2220.

Wilczkowiak, M., Sturm, P. and Boyer, E., 2005. Using geometric constraints through parallelepipeds for calibration and 3D modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(2), pp. 194–207.

Zebedin, L., Bauer, J., Karner, K. and Bischof, H., 2008. Fusion of feature-and area- based information for urban buildings modeling from aerial imagery. In: European Conference on Computer Vision, Vol. 5305, Lecture Notes in Computer Science, pp. 873–886.

Zhu, Z. and Kanade (Eds.), T., July, 2008. Special Issue: Modeling and Representations of Large-Scale 3D scenes. International Journal of Computer Vision.