VARIATIONAL MODEL-BASED 3D BUILDING EXTRACTION FROM REMOTE SENSING DATA

Konstantinos Karantzalos and Nikos Paragios

Laboratoire de Mathematiques Appliquees aux Systemes (MAS)
Ecole Centrale de Paris
Chatenay-Malabry, France
{konstantinos.karantzalos, nikos.paragios}@ecp.fr

ABSTRACT

In this paper, we introduce a variational framework towards automatic 3D building reconstruction from optical and Lidar data. Multiple 3D competing building priors are considered under a recognition-driven way. These models, under a certain hierarchical representation, describe the space of solutions and under a fruitful synergy with an inferential procedure recover the observed scene's geometry. Our formulation allows the cue with the higher spatial resolution to constrain properly the boundaries detection procedure ensuring, in this way, optimal results in terms of accuracy. Such an integrated approach is defined in a variational context, solves segmentation in both spaces, addresses fusion in a natural manner and allows multiple competing priors to determine the pose and 3D geometry from the observed data. Very promising experimental results demonstrate the potentials of our approach.

Index Terms— Pattern Recognition, Variational Methods, Object Detection, Segmentation, Competing Priors

1. INTRODUCTION

Modeling urban and peri-urban environments with engineering precision, enable 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 has made the creation and manipulation of 3D city models quite valuable, especially at large scale. 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 3D building extraction and reconstruction ([1, 2, 3, 4, 5, 6] 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 [7, 8, 9]. Processing remote sensing data, still, poses several challenges.

In this paper, we aim to address these challenges by introducing a novel variational framework towards large-scale building reconstruction through information fusion and grammar-based building priors. 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 model. Our formulation allows data with the higher spatial resolution to constrain properly the footprint detection. Therefore, we are proposing a variational functional that encodes a fruitful synergy between observations and multiple 3D grammar-based building models. Our models refer to a grammar, which consists of typologies of 3D shape priors. In such a context, firstly one has to select the most appropriate model and then determine the optimal set of parameters aiming to recover scene's geometry. The proposed objective function consists of two segmentation terms that guide the selection of the most appropriate typology and a DEM-driven term which is being conditioned on the typology. Looking forward to large scale reconstruction and since usually for most sites very high resolution data are missing, our aim was not to produce high quality 3D maps from video sequences or numerous high resolution stereopairs but rather to fuse the fewest available data (e.g. a single satellite image and a coarser DTM) with prior models towards large scale reconstruction. Doing multiview stereo or using simple geometric representations like 3D lines or planes was not our interest here.

2. HIERARCHICAL GRAMMAR-BASED BUILDING PRIORS

Hierarchical representations are a natural selection to address complexity while at the same time recover representations of acceptable resolution. Our models involve two components, the type of footprint and the type of roof (Fig.(1)). 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

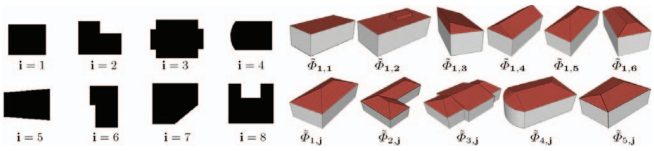

Fig. 1. Prior Building Models ($\tilde{\Phi}_{\mathbf{i,j}}$): $\mathbf{i}$ determines the shape of building's footprint and $\mathbf{j}$ its roof type. Left: 8 binary templates from the different type of building footprints. Right (top): The family $\tilde{\Phi}_{1,\mathbf{j}}$ of buildings which have a rectangular footprint ($\mathbf{i} = 1$). Right (bottom): The family $\tilde{\Phi}_{\mathbf{i=1:5,j}}$ of prior models.

every family (i.e. every $\mathbf{i}$) the different types of building tops (roofs) are modeled by the pointer $\mathbf{j}$ (Fig.(1)) Under this hierarchy $\tilde{\Phi}_{\mathbf{i,j}}$, 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 Fig.(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 (Fig.(3)). 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.

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 (Fig.(1) - $\tilde{\Phi}_{\mathbf{1,5}}$), if two opposite angle are zero we have a gable-type one (Fig.(1) - $\tilde{\Phi}_{\mathbf{1,4}}$) and if all are zero we have a flat one ($\tilde{\Phi}_{\mathbf{1,1}}$). The platform and the gambrel roof types can not be modeled but can be easily derived. The platform one ($\tilde{\Phi}_{\mathbf{1,2}}$), for instance, is the case where

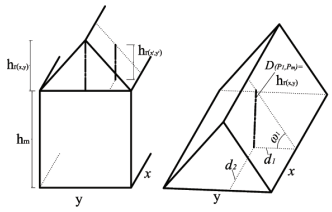

Fig. 2. Hierarchical grammar-based 3D description for the building models. Building's footprint is determined implicitly from the E_{2D}. Building's main height h_m and roofs height $h_r(x, y)$ at every point are recovered (E_{3D}) and thus all $\mathbf{j}$ different roof types are modeled or easily derived.

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 $\mathbf{i}$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 BUILDING PRIORS IN COMPETITION

Let us consider a pair of images: one that corresponds to the visible domain ($\mathcal{I}$) and the corresponding digital elevation map ($\mathcal{H}$). In such a context, one has first to separate buildings from background (natural scene), extract the corresponding footprint types and determine their geometry. Let $\phi : \Omega \rightarrow \mathcal{R}^+$ be a level set representation defined at the dense image resolution level. Then, 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. On top of that, in the DEM one would expect that man-made structures will exhibit elevation differences from the natural part of the scene. These two assumptions can be used to define the following segmentation function

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

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. In order, to cope with the lack of visual support from such a purely data-driven term, one can consider the use of prior knowledge. This can be achieved, through the integration of global shape prior constrains into the segmentation process. These constraints can encode both 2D as well as 3D measurements. The 2D constraint (footprint) can be determined from the image and the

DEM while the 3D one from the DEM. Let us now consider an abuse of notation and introduce an additional prior components in the process $E_{prior} = E_{2D} + E_{3D}$.

Let us first consider the footprint prior. Following the formulations of [10], we employ a k-dimensional labeling function, which is able for the dynamic labeling of up to $m = 2^k$ regions. Thus, the following cost functional can account for a recognition-driven segmentation, based on multiple competing shape priors:

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

with the two parameters $\lambda, \rho > 0$.

3.1. Grammar-based Building 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 roof's inclined planes ($\Theta_i = (h_m, \omega_1, \omega_2, \omega_3, \omega_4)$). These four angles (Fig.(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) $h_r(x, y)$:

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

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 Fig.(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:

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

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 reconstruction framework now takes the following form in respect to ϕ, $\mathcal{T}_i$, $\mathbf{L}$ and Θ_i:

$$E_{total} = E_{seg}(\phi) + \mu E_{2D}(\phi, \mathcal{T}_i, \mathbf{L}) + \mu E_{3D}(\Theta_i) \quad (4)$$

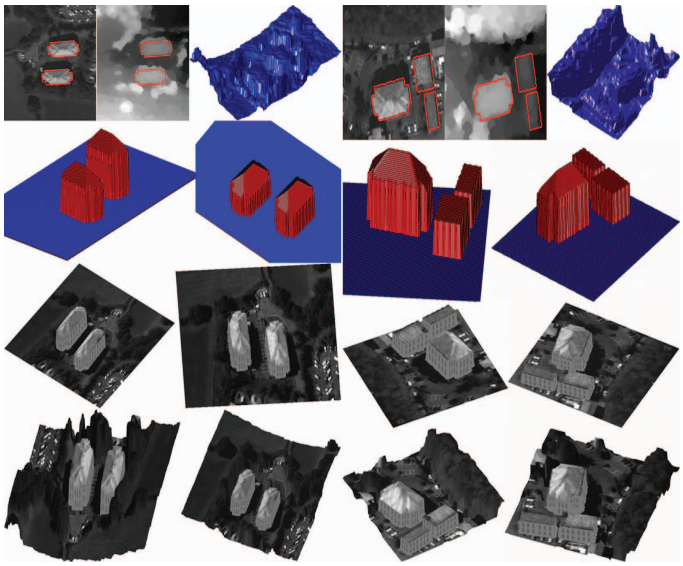

Fig. 3. First row: Detected building footprints superimposed on data and a 3D visualization of the DEM. Second and third row: 3D views of the reconstructed buildings with and without texture. Fourth row: 3D views of scene's reconstruction.

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 **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. EXPERIMENTAL RESULTS

The developed algorithm has been applied to a number of scenes where remote sensing data was available. In Fig.3 results for the detection and reconstruction of a small number of buildings are presented. The algorithm managed in all cases to accurately recover their boundaries 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

Fig. 4. Left: Detected building boundaries superimposed data. Middle: 3D visualization of scene's DEM. Right: Reconstructed scene.

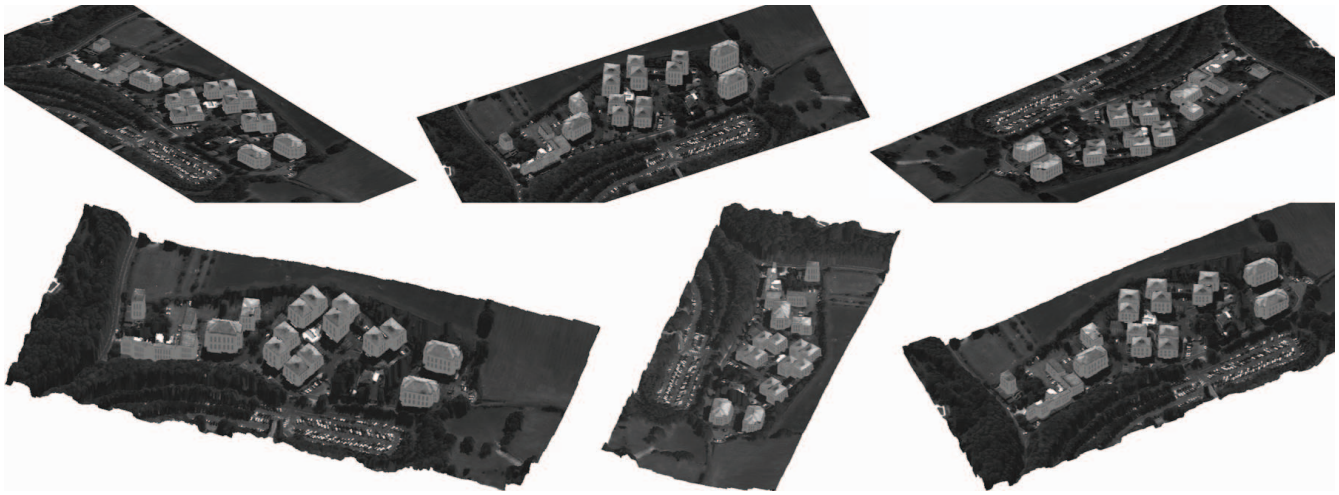

Fig. 5. Large-scale building reconstruction. Different views of the reconstructed buildings (first row) and views of the entire scene's reconstruction.

recover their 3D geometry as the appropriate priors were chosen. In both cases of Fig.3, the performed quantitative evaluation indicated that the algorithm's completeness, correctness and overall quality -standard quantitative measures for man-made object extraction- were above 96%.

In Fig.(4) and Fig.(5) results are shown for a quite complex scenario. The considered areas, consist of complex landscape, multiple objects of various classes, shadows, occlusions, different texture patterns and an important terrain variability. For both test site just a single panchromatic aerial image with appx. 0.7m spatial resolution was available and the corresponding DEM in a lower resolution (of appx. 2.5m). The detected building footprints superimposed on data are shown in (Fig.4) and different views of their recovered 3D geometry are shown in (Fig.5). All buildings, except one, were extracted and reconstructed. All of them have been recognized with a different identity (have been labeled and numbered uniquely) apart from the three-building segment at the top right corner of the scene. It was poorly detected but, also, appears as one segment in the ground truth data.

5. CONCLUSIONS

A novel recognition-driven variational framework, has been introduced, towards multiple 3D building extraction and reconstruction. It is an inferential approach that fuses optical images and digital elevation maps, is defined in a variational context, solves segmentation in both spaces and allows multiple competing priors to determine their pose and 3D geometry from the observed data. By describing our numerous building models with a certain hierarchy and grammar and formulating, respectively, our energy terms we narrow, effectively, the search space during optimization. Apart from new building models, other classes of terrain features can be added or removed from the database, controlling respectively the type of objects that can be addressed by the system. Last but not least, 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.

6. REFERENCES

[1] J. Hu, S. You, and U. Neumann, "Approaches to large-scale urban modeling," *IEEE Computer Graphics and Applications*, vol. 23, no. 6, pp. 62–69, 2003.

[2] A. Zakhor and C. Frueh, "Automatic 3d modeling of cities with multimodal air and ground sensors.," in *Multimodal Surveillance, Sensors, Algorithms and Systems.*, Z. Zhu and T.S. Huang, Eds., vol. Chapter 15, pp. 339–362. Artech House, 2007.

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

[4] V. Verma, R. Kumar, and S. Hsu, "3D building detection and modeling from aerial lidar data," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 2213–2220.

[5] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny, "3D city modeling based on hidden markov model," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2007, vol. II, pp. 521–524.

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

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

[8] Z. Zhu and T. Kanade (Eds.), "Special Issue: Modeling and Representations of Large-Scale 3D scenes," *International Journal of Computer Vision*, vol. 78, no. 2-3, July, 2008.

[9] H. Mayer, "Object extraction in photogrammetric computer vision," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 63, no. 2, pp. 213–222, 2008.

[10] K. Karantzalos and N. Paragios, "Recognition-driven 2D competing priors towards automatic and accurate building detection," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 133–144, 2009.