Combining meshes and geometric primitives for accurate and semantic modeling

Florent Lafarge[1]
http://imagine.enpc.fr/
Renaud Keriven[1]
http://imagine.enpc.fr/
Mathieu Brédif[2]
http://ign.fr/

[1] IMAGINE, ENPC/CSTB, LabIGM
Université Paris est
Paris, France

[2] Matis laboratory, IGN
Saint-Mandé, France

1 Introduction

3D-models of urban scenes are very useful for many applications such as urban planning, virtual reality, disaster recovery or computer games. The reconstruction of such scenes is a well known computer vision problem which has been addressed by various approaches providing integral building representations such as [4, 6, 12, 22], but remains an open issue [16, 29].

1.1 Building modeling

With the new perspectives offered for the aid to navigation by general public softwares such as *Street View* (Google) or *GeoSynth* (Microsoft), 3D building modeling is a topic of growing interest. Many works have been recently proposed. Two main families of approaches may be distinguished.

3D-primitive modeling - The first family represents buildings as 3D-object layouts [6, 11, 13, 17, 18, 26, 28]. These works efficiently detect and insert various urban objects such as windows or doors in 3D building models. However, these limited parametric descriptions fail to model fine details.

Mesh representation - The reconstruction of buildings with high order details, such as ornament, statues and other irregular shapes, is mainly addressed by mesh generation techniques using Laser scanning [3, 8] or multi-view stereo processes [9, 10, 25]. Multi-view stereo techniques have significantly progressed during recent years as underlined in the comparative studies [20, 23]. Figure 1 highlights the quality of a mesh, especially for describing high order details. However, buildings are man made objects containing many regular components such as planar or cylindrical shapes. Such a mesh representation gives a large amount of useless information concerning these regular elements which could be more relevantly described by parametric objects (*e.g.* wall facets by planes or columns by cylinders).

© 2009. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

The two families have complementary advantages : semantic knowledge and model compaction for the former, detail modeling and non-restricted use for the latter. A natural idea, but still unexplored, would consist in merging both the families in order to propose a **hybrid modeling**. Regular elements would be representing by 3D primitives whereas irregular structures would be described by meshes. In this paper, we propose a process for substituting regular mesh patches by 3D-objects. This is of interest for several reasons: *(i)* the introduction of semantic knowledge in the mesh; *(ii)* the simplification of the modeling while preserving details; and *(iii)* the corrections of some errors generated by the multi-view stereo processes.

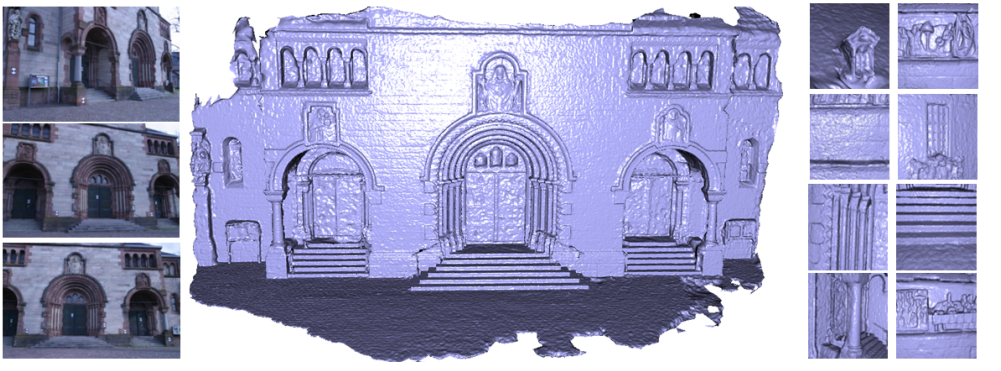

Figure 1: Modeling by mesh representation output by a multi-view stereo process (data from [23], results from [25]).

1.2 Strategy

Extracting 3D-primitives from meshes without a preliminary segmentation is a difficult problem [7]. It has been addressed by [5] for simplifying a mesh into a 3D-plane layout, and then extended by [27] for modeling with quadrics. However, such an approach cannot be efficiently adapted for image-based modeled meshes which contain noise, facet density variations, multi-scale components and errors/approximations resulting of multi-view stereo processes as shown on Figure 1. We adopt a more robust two step strategy consisting in *(i)* segmenting the mesh, and *(ii)* fitting 3D-primitives on the obtained partition where it is relevant. Section 2 presents the segmentation process based on a curvature analysis of the mesh. A multi-label energy taking topological smoothness constraints into account is formulated. The optimal labeling is estimated by α-expansion. The 3D-primitive extraction from the obtained partition is then described in Section 3. An error parameter controls the fitting quality and decides whether a mesh cluster has to be substituted by a plane, sphere, cylinder, cone or torus. In addition, a refinement process corrects the eventual errors generated during the segmentation step. Experimental results on real building meshes and also on synthetic data are given in Section 4. Basic conclusions are outlined in Section 5.

2 Mesh segmentation

Let us consider a three dimensional boundary mesh M defined as a tuple $\{V,E,F\}$ of vertices V, edges E and triangular faces F. We aim to segment the vertices of the mesh M into subsets corresponding to regions of interest.

2.1 Geometric attributes based on curvature analysis

Many kinds of local geometric attributes have been proposed in the literature for segmenting synthetic meshes such as multi-scale blowing bubbles, 3D feature descriptors or skeleton knowledge. The comparative studies proposed in [1, 20] present the most efficient techniques for extracting information from synthetic meshes. Most of these techniques cannot be adapted to meshes generated by multi-view stereo processes due to the problems mentioned in Section 1. Local differential geometry estimates are known to be robust for analyzing the mesh topology. The principal curvatures k_{min} and k_{max} and their associated direction vectors $\mathbf{w_{min}}$ and $\mathbf{w_{max}}$ measure how the surface bends by different amounts in different directions (see Figure 2). In order to distinguish the various types of shapes, this curvature information is used to label the mesh according to four labels of interest: *planar* ($k_{max} = k_{min} = 0$), *developable convex* ($k_{min} = 0 < k_{max}$), *developable concave* ($k_{min} < k_{max} = 0$) and *non developable* surfaces ($k_{min}k_{max} \neq 0$).

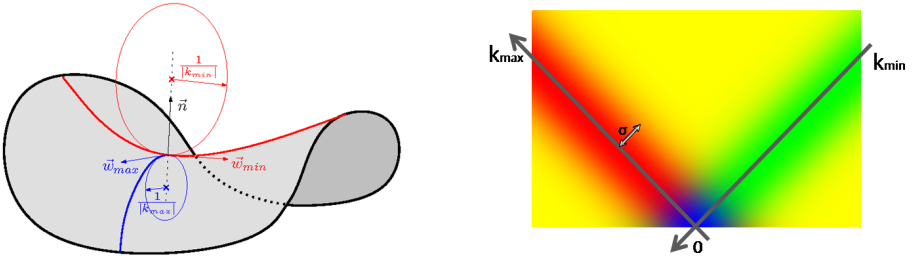

Figure 2: Principal curvatures - *left:* representation of (k_{min}, k_{max}), *right:* map of the label dominance in function of k_{min} and k_{max} (blue sector indicate that the highest probability is obtained for the '*planar*' label, red for '*developable convex*', green for'*developable concave*' and yellow for'*non developable*').

Let us consider $\mathscr{L} = \{1,2,3,4\}$, the label set corresponding to the classes mentioned above respectively. Let $l = (l_1, ..., l_N)$ be a label configuration in $\mathscr{L}^N$, associated with the N vertices of the mesh M. By denoting $G_\sigma(k) = \exp(-k^2/2\sigma^2)$ the non normalized centered Gaussian distributions with a standard deviation σ, we can express the probability of each label at the vertex i as a combination of the curvature distributions:

$$Pr(l_i|k_{min}^{(i)}, k_{max}^{(i)}) = \begin{cases} G_\sigma(k_{min}^{(i)})G_\sigma(k_{max}^{(i)}) & \text{if } l_i = 1 \\ G_\sigma(k_{min}^{(i)})(1 - G_\sigma(k_{max}^{(i)})) & \text{if } l_i = 2 \\ (1 - G_\sigma(k_{min}^{(i)}))G_\sigma(k_{max}^{(i)}) & \text{if } l_i = 3 \\ (1 - G_\sigma(k_{min}^{(i)}))(1 - G_\sigma(k_{max}^{(i)})) & \text{if } l_i = 4 \end{cases} \quad (1)$$

Figure 2 presents the behavior of this probability in function of the couple (k_{min}, k_{max}). The label configuration maximizing $\prod_{i \in V} Pr(l_i|k_{min}^{(i)}, k_{max}^{(i)})$, denoted by $\widehat{l_P}$, is simple to compute and provides an interesting estimator in the case of synthetic meshes as we can see with the Fandisk and Cup models presented on Figure 3. However, the results obtained from non synthetic meshes are clearly more limited. Additional information has to be taken into account to improve results.

2.2 A multi-label energy model

The energy of the configuration l is formulated using both a consistency term and topological smoothness constraints, balanced by the parameter $\beta > 0$:

$$U(l) = \sum_{i \in V} D_i(l_i) + \beta \sum_{\{i,j\} \in E} V_{ij}(l_i, l_j) \tag{2}$$

Consistency The consistency $D_i(l_i)$ which measures the coherence of the label l_i at the vertex i is computed using the probability $Pr(l_i | k_{min}^{(i)}, k_{max}^{(i)})$ (see Eq.1) such as:

$$D_i(l_i) = 1 - Pr(l_i | k_{min}^{(i)}, k_{max}^{(i)}) \tag{3}$$

The sensitivity of the consistency term is controlled by the standard deviation σ of the principal curvature distributions (See Figure 2). Taking a low σ value makes the consistency term more selective with *planar* and *developable* labels and favors the *non developable* one. On the contrary, a high value has to be chosen for dealing with noise corrupted meshes.

Topological smoothness constraints The term V_{ij} represents a pairwise interaction potential between adjacent vertices i and j. It expresses prior knowledge about the optimal labeling.

$$V_{ij}(l_i, l_j) = \begin{cases} 1 & \text{if } l_i \neq l_j \\ \min(1, a||\mathbf{W_i} - \mathbf{W_j}||_2) & \text{otherwise} \end{cases} \tag{4}$$

where a is a scale factor fixed proportionately to the mean edge length $\widehat{e}$ of M, and $\mathbf{W_i}$ and $\mathbf{W_j}$ are 6×1 vectors combining the principal direction vectors and their curvatures:

$$\mathbf{W} = \begin{pmatrix} k_{min}.\mathbf{w_{min}} \\ k_{max}.\mathbf{w_{max}} \end{pmatrix} \tag{5}$$

This term introduces spatial smoothness constraints which take into account the mesh topology. Two principles define the behavior of V_{ij}:

- **Smoothness on regular surfaces -** In order to favor the label homogeneity in a neighborhood, adjacent vertices are penalized if their labels are different. This principle acts like the Potts model (See Figure 3-2^{nd} and 4^{th} rows).
- **Edge preservation -** The boundaries are preserved by taking into account the principal direction vector variations of adjacent vertices with similar labels. The mesh is then partitioned according to changes of local differential geometry. For example, it allows the separation of two connected planes with different normals (See Figure 3-*Corner* model).

2.3 Optimization by α-expansions

Finding the label configuration that minimizes the energy U requires advanced optimization techniques since U is a non convex function defined in a multi-label space. We use the α-expansion algorithm [4] based on the Graph-cuts theory. One can easily check that our energy fits the requirements for this method. This algorithm allows us to quickly reach an approximate solution close to the global one. To accelerate the convergence, $\widehat{l_P}$ is chosen as

initialization. Note that faster algorithms such as Logcut [14] could be used. However, the time savings would be minor since we have a small number of labels.

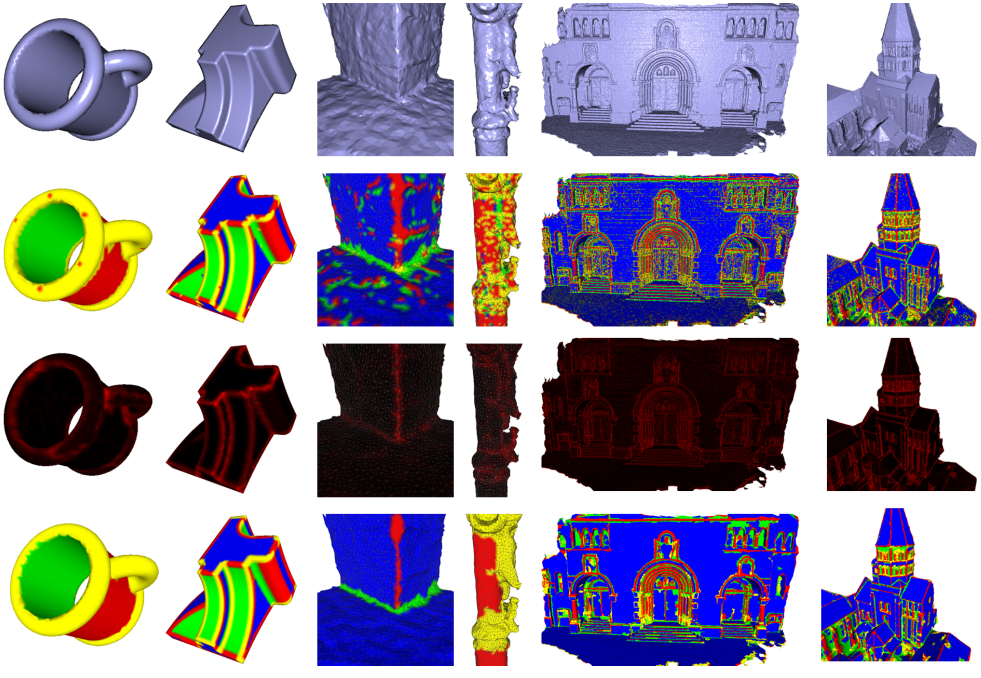

Figure 3: Mesh segmentation - *from top to down:* original mesh, $\widehat{l_P}$ estimator (blue='*planar*', red='*developable convex*', green='*developable concave*' and yellow='*non developable*'), edge term of the regularizing part of the energy ($||\mathbf{W_i} - \mathbf{W_j}||_2$) (red=high values, black=low values), our labeling result after energy minimization.

Figure 3 shows results of the segmentation stage. The proposed multi-label energy significantly improves the results compared to the $\widehat{l_P}$ estimator for the non synthetic meshes. The various parts are correctly identified: walls, roofs or stairs are associated with the *planar* label - columns, corners or vaultings with *developable convex* or *developable concave* labels - and ornaments or statues with the *non developable* one. The edges are accurately localized due to the detection of principal direction vector variations (See 3-3^{rd}row). It allows us to extract these components easily by a region growing process (see Figure 4). The next stage consists in fitting 3D-primitives to the obtained partition.

3 Geometric shape extraction

In the sequel, we call *cluster* a connected region of same label extracted by the previous process. Each cluster of the segmented mesh is then compared to a set of 3D-primitives composed of planes, spheres, cylinders, cones and tori. They represent the most common regular shapes which can be found on buildings.

Extraction strategy- In order to avoid an exhaustive comparison between a cluster and all

the types of 3D-primitives, the labeling information obtained in the previous stage is used to drive the shape extraction. A cluster labeled as a *planar* component is then compared to a plane, *developable convex* and *developable concave* clusters to cylinders and cones, and *non developable* clusters to spheres and tori. An error parameter ξ controls the fitting quality. If the quadratic error between the optimal primitive and the cluster is lower than ξ, the cluster is substituted by the detected primitive. Otherwise, the rejected cluster is compared to the other types of 3D-primitives. This second fitting test prevents wrong labelings generated by scale ambiguities. For example, the large vaultings on Figure 3 are mistakenly labeled as '*planar*' clusters due to the low values of their principal curvatures. This additional test correctly fits these vaultings to cylinders (See Figure 5). Finally, if the cluster is still rejected during this second test, it keeps its triangular mesh representation.

Object fitting - Severals works such as [19, 24] have been proposed to detect shapes in point clouds containing outliers. Contrary to point clouds, meshes have generally less outliers and exhibit useful topological information. Outlier rejection based techniques such as the RANSAC algorithm are not required for our problem due to our preliminary segmentation. Plane fitting can be easily performed using a Principal Component Analysis (PCA). However, fitting spheres, cylinders, cones or tori has no closed-form solution when the dataset only represents an unknown portion of the whole shape. Thus, it requires an iterative non-linear minimization, typically using a Levenberg-Marquardt optimization. We base our fitting on [15], that proposes a parametrization and a first order Euclidean distance approximation to spheres, cylinders, cones and tori, that behaves well as curvatures vanish. This allows numerically stable fittings of more complex shapes on a dataset close to a simpler shape (sphere, cone, cylinder or torus fitting of an almost planar patch, cone fitting of an almost cylindrical patch, torus fitting of a spherical or a conical patch...).

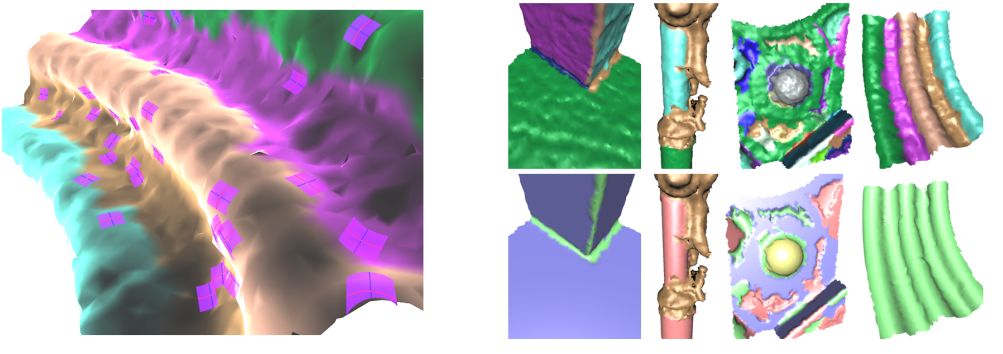

Figure 4: 3D-primitive extraction - *left:* multi-initialization using local differential geometry estimates (highlighted here as small purple patches), *right:* examples of fitted primitives with segmented meshes *(top)* and hybrid representations *(down)*.

A multi-initialization strategy using local differential geometry estimates - Relying on a non-linear optimization, the quality of the fitting process depends on its initialization. [15] estimates an initialization from a global criterion. Here, for more robustness, we propose to let multiple initializations based on various local estimates compete, and keep the overall best fit (see Figure 4-left). Differential geometry estimates have already been computed for each vertex to drive the segmentation. Considering a small set of seed vertices covering the

whole patch, we initialize a non-linear optimization for each seed vertex position using its differential geometry estimates. The parameterizations in [15] use an arbitrary point on the shape to parametrize the whole shape using its local differential geometry (normal vector, principal curvatures and directions). Spheres and cylinders are completely parameterized using the local estimates of a seed vertex. Cones, which are generalized cylinders with a center at infinity, are initialized using the locally estimated cylinder. Turning to tori, they contain an inner and an outer circle, where the normals are orthogonal to the axis of revolution. Supposing a seed vertex is on such a circle then yields two possible torus initializations which are optimized independently.

4 Experiments

Our approach is tested on real meshes generated by the multi-view stereo technique proposed in [25]. Figure 4 shows the potential of the method on some details whereas Figure 5 presents results on various larger scenes. There is, to our knowledge, no other method proposing hybrid representations. However, we evaluate our results qualitatively and quantitatively with a visual evaluation, a compression rate study and an accuracy improvement experiment.

Visual evaluation - The obtained hybrid representations are promising and provide interesting simplified modelings of the original meshes while preserving details. The overall rough components of buildings are reconstructed by 3D-primitive layouts with an accuracy controlled by ξ. Such object layouts are very useful since they allow the introduction of semantic information in the modeling. Structural components such as walls, roofs, windows or dormer windows can be easily identified from the obtained primitives by a subsequent basic analysis[1] as one can see on Figure 5-4^{th} row. The results reveal the reconstruction of interesting fine details such as thin pipes located at the vaultings on Figure 5-2^{nd} row or small statue heads on Figure 5-last row.

Compression rate- The compression rate, defined as the ratio between the original mesh and the hybrid representation, is function of the error parameter ξ. Table 1 shows that it indeed also depends on the scene: a scene containing many regular components (e.g. Church) has a better factor than one composed of many irregular shapes (e.g. Fountain-P11). The experiments presented Figure 5 are conducted with $\xi = \widehat{e}$ and give both a good compression ratio and visually acceptable results. Figure 6 compares our method with a state-of-the-art decimation method (the one in CGAL library[2]). For a given compression rate, our representation gives a better description than the decimated mesh which is uniformly degraded with no semantic awareness. Indeed, taking the geometric regularity of the scene into account is relevant for buildings: on this detail, planes and cylinders are clearly identified.

Accuracy improvement- In order to quantify in what extent our representation is better than the decimated one, we evaluate the error with respect to the (range scanned) ground truth. In fact, our method even allows the corrections of some errors already contained in the non-decimated model! For instance, the noisy balls and waveform ornament are regularized by half a sphere and a set of toroidal patches respectively on Figure 4. The error occupancy

[1] For example, one can detect walls by just selecting tall surface planar primitives which are vertical.
[2] http://www.cgal.org/

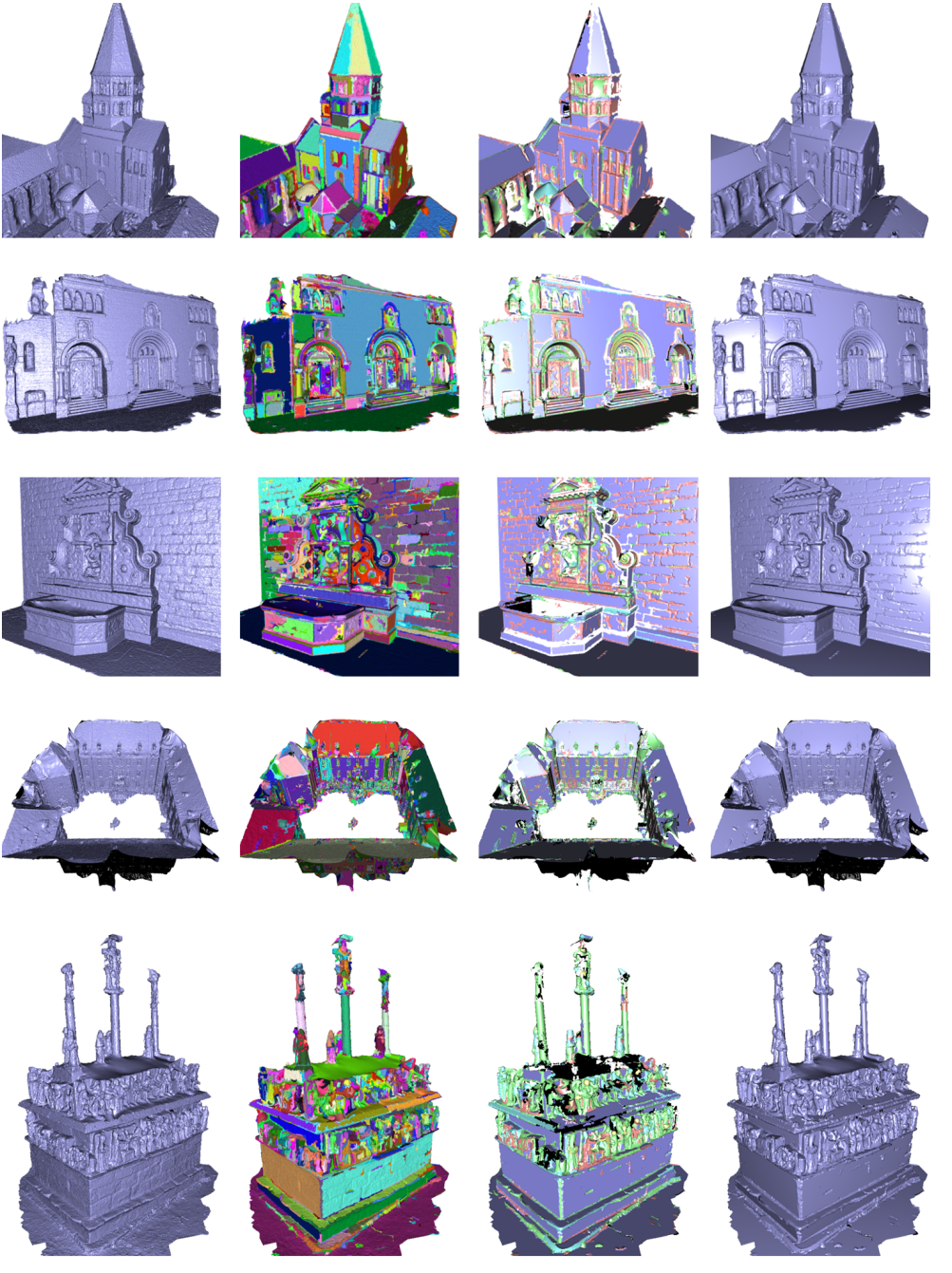

Figure 5: Large scenes- *from left to right:* original mesh, segmented mesh, 3D-primitives *(purple=plane, pink=cylinder, blue=cone, yellow=sphere, green=torus)*, hybrid representation. *From top to bottom:* Church, Herz-Jesu-P25, Fountain-P11, Castle-P30, Calvary.

Table 1: Compression rates in function of the error ξ.

	$\xi = 0.1\widehat{e}$	$\xi = 0.5\widehat{e}$	$\xi = \widehat{e}$	$\xi = 5\widehat{e}$	$\xi = 10\widehat{e}$
Church	1.27	3.55	4.59	10.43	10.43
Herz-Jesu-P25	1.12	3.57	5.93	10.78	11.34
Fountain-P11	1.09	2.33	3.6	6.34	11.51
Castle-P30	1.1	2.19	3.96	8.87	11.21

histogram, measured with respect to the standard deviation Σ of the ground truth accuracy (see [23]), quantifies this improvement. The number of low-error vertices if higher for the hybrid representation than for the original stereo mesh, mainly transferring from the 2Σ bin to the Σ one (Figure 6 bottom). This is indeed a first step toward a more extensive evaluation, since this improvement seems to concern subset of the mesh the has to be identified and closely analyzed. Yet, the benchmark website of [23] only outputs global statistics and does not easily allow this investigation.

Limitations- Our approach cannot extract piecewise 3D-primitives merged in a single cluster. For example, the Ω shape ornaments above the doors of the Herz-Jesu mesh (see Figure 5-2^{nd} row) are not reconstructed because they are composed of cylinders and tori. The compression rate could be improved by proposing a process for fitting several objects per cluster. Moreover, the process is not well adapted to smoothness variations over the mesh. To solve this problem, β could be locally adjusted according to some estimated local quality.

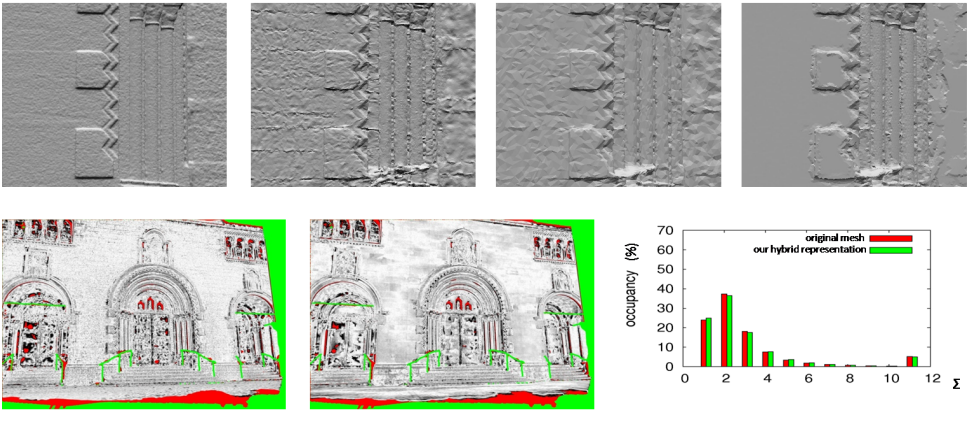

Figure 6: *From left to right, top :* details of ground truth [23], original mesh [25], state-of-the-art decimated mesh (compression rate=3.6), and our representation (same compression rate). *Bottom row :* error of the original mesh with respect to ground truth (white=low, black=high), error of our representation, histogram of the errors [23].

5 Conclusion

We propose an hybrid representation of noisy 3D models such as buildings obtained by multi-view stereo. This representation merges meshes and 3D-primitives. It provides high compression rates while keeping details, introduces semantic knowledge despites noise cor-

ruption, and even improves accuracy of the original reconstruction. Both the proposed multi-label energy formulation for mesh segmentation and the contributions for 3D-primitive fitting could be used for others meshing applications. In the future, we will study the simultaneous generation of meshes and 3D-primitives during the multi-view stereo process. This would allow us to take interactions between meshes and primitives into account, but would require more complex models and advanced 3D-primitive samplers.

Acknowledgments

The authors are grateful to the EADS foundation for partial financial support.

References

[1] M. Attene, S. Katz, M. Mortara, G. Patane, M. Spagnuolo, and A. Tal. Mesh segmentation - a comparative study. In *SMI*, Washington, United States, 2006.

[2] C. Baillard and A. Zisserman. Automatic reconstruction of piecewise planar models from multiple views. In *CVPR*, Los Alamitos, United States, 1999.

[3] A. Banno, T. Masuda, T. Oishi, and K. Ikeuchi. Flying laser range sensor for large-scale site-modeling and its applications in bayon digital archival project. *IJCV*, 78 (2-3):207–222, 2008.

[4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222 – 1239, 2001.

[5] D. Cohen-Steiner, P. Alliez, and M. Desbrun. Variational shape approximation. *ACM Transactions on Graphics*, 23(3):905–914, 2004.

[6] A.R. Dick, P.H.S. Torr, and R. Cipolla. Modelling and interpretation of architecture from several images. *IJCV*, 60(2):111–134, 2004.

[7] C. Dorai and A.K. Jain. COSMOS - a representation scheme for 3D free-form objects. *PAMI*, 19(10):1115–1130, 1997.

[8] C. Fruh and A. Zakhor. An automated method for large-scale, ground-based city model acquisition. *IJCV*, 60(1):5–24, 2004.

[9] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. In *CVPR*, Minneapolis, United States, 2007.

[10] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S.M. Seitz. Multi-view stereo for community photo collections. In *ICCV*, Rio de Janeiro, Brazil, 2007.

[11] F. Han and S.C. Zhu. Bottom-up/top-down image parsing by attribute graph grammar. In *ICCV*, Beijing, China, 2005.

[12] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny. Building reconstruction from a single DEM. In *CVPR*, Anchorage, United States, 2008.

[13] S.C. Lee and R. Nevatia. Extraction and integration of window in a 3d building model from ground view images. In *CVPR*, Washington, United States, 2004.

[14] V. Lempitsky, C. Rother, and A. Blake. Logcut - efficient graph cut optimization for markov random fields. In *ICCV*, Rio de Janeiro, Brazil, 2007.

[15] D. Marshall, G. Lukacs, and R. Martin. Robust segmentation of primitives from range data in the presence of geometric degeneracy. *PAMI*, 23(3):304–314, 2001.

[16] H. Mayer. Object extraction in photogrammetric computer vision. *JPRS*, 63(2), 2008.

[17] P. Muller, G. Zeng, P. Wonka, and L. Van Gool. Image-based procedural modeling of facades. *ACM Transactions on Graphics*, 26(3):614–623, 2007.

[18] M. Pauly, N. J. Mitra, J. Wallner, H. Pottmann, and L. Guibas. Discovering structural regularity in 3D geometry. *ACM Transactions on Graphics*, 27(3), 2008.

[19] R. Schnabel, R. Wahl, and R. Klein. Efficient ransac for point-cloud shape detection. *Computer Graphics Forum*, 26(2):214–226, 2007.

[20] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR*, New York, United States, 2006.

[21] A. Shamir. A survey on mesh segmentation techniques. *Computer Graphics Forum*, 27(6):1539–1556, 2008.

[22] S. N. Sinha, D. Steedly, R. Szeliski, M. Agrawala, and M. Pollefeys. Interactive 3D architectural modeling from unordered photo collections. *ACM Transactions on Graphics*, 27(5), 2008.

[23] C. Strecha, W. Von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *CVPR*, Anchorage, United States, 2008.

[24] V. Verma, R. Kumar, and S. Hsu. 3D building detection and modeling from aerial LIDAR data. In *CVPR*, New York, United States, 2006.

[25] H. Vu, R. Keriven, P. Labatut, and J.-P Pons. Towards high-resolution large-scale multi-view. In *CVPR*, Miami, United States, 2009.

[26] J. Xiao, T. Fang, P. Tan, P. Zhao, E. Ofek, and L. Quan. Image-based façade modeling. *ACM Transactions on Graphics*, 27(5), 2008.

[27] D.M. Yan, Y. Liu, and W.P. Wang. Quadric surface extraction by variational shape approximation. In *GMP*, Pittsburgh, United States, 2006.

[28] 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 *ECCV*, Marseille, France, 2008.

[29] Z. Zhu and T. Kanade. Special issue on modeling and representations of large-scale 3d scenes. *International Journal of Computer Vision*, 78(2-3), 2008.