Insertion of 3D-Primitives in Mesh-Based Representations: Towards Compact Models Preserving the Details

Florent Lafarge, Renaud Keriven, Mathieu Brédif

HAL Id: inria-00503145

https://inria.hal.science/inria-00503145

Submitted on 28 Jan 2013

Insertion of 3D-primitives in mesh-based representations: Towards compact models preserving the details

Florent Lafarge, Renaud Keriven and Mathieu Brédif

Abstract—We propose an original hybrid modeling process of urban scenes that represents three-dimensional (3D) models as a combination of mesh-based surfaces and geometric 3D-primitives. Meshes describe details such as ornaments and statues, whereas 3D-primitives code for regular shapes such as walls and columns. Starting from an 3D-surface obtained by multi-view stereo techniques, these primitives are inserted into the surface after being detected. This strategy allows the introduction of semantic knowledge, the simplification of the modeling, and even correction of errors generated by the acquisition process. We design a hierarchical approach exploring different scales of an observed scene. Each level consists first in segmenting the surface using a multi-label energy model optimized by α-expansion and then in fitting 3D-primitives such as planes, cylinders or tori on the obtained partition where relevant. Experiments on real meshes, depth maps and synthetic surfaces show good potential for the proposed approach.

Index Terms—3D reconstruction, shape extraction, urban scenes, object based representation, graph-cut.

I. INTRODUCTION

3D-models of urban scenes are useful for many applications such as urban planning, virtual reality, disaster recovery or computer games. In particular, with new perspectives offered to navigation aids by general public softwares such as *Street View* (Google) or *GeoSynth* (Microsoft), 3D building modeling is a topic of growing interest. The reconstruction of such scenes is a well known computer vision problem which has been addressed by various approaches but remains an open issue [1], [2], [3].

A. *3D-modeling of urban scenes*

Urban objects within the same scene significantly differ in term of complexity, diversity, and density. Many works have been proposed for reconstructing urban scenes. The most common inputs are generally multiview stereo images and laser scans. Three main types of 3D-representations can be distinguished in the literature for modeling these scenes.

3D-primitive arrangements constitute the most common type of 3D-representation in building reconstruction. The scenes are represented as layouts of simple geometric objects such as planes, lines or cubes which describe the urban objects well and are interesting in terms of storage capacity. For example,

F. Lafarge and R. Keriven are with the IMAGINE group, LIGM, Université Paris Est, Marne-la-Vallée, FRANCE – e-mail:lafargef@imagine.enpc.fr and keriven@imagine.enpc.fr.

M. Brédif is with the MATIS laboratory, French Mapping Agency, Saint-Mandé, FRANCE – e-mail:Mathieu.bredif@ign.fr.

buildings are fully reconstructed by 3D-planes in [4], [5]. More specific works on roof reconstruction from aerial/satellite data [6], [7], [8], [9], [10], facade modeling from terrestrial data [11], [12] or building interior reconstruction [13] underline the efficiency of the 3D-primitive based approaches. They are also used to introduce semantic information in 3D building representations by detecting and inserting various urban objects such as windows, doors or roof supertructures [14], [15], [16], [17]. However, these parametric descriptions remain a simplistic representation and fail to model fine details and irregular shapes.

Depth maps allow the description of a scene as an image. Contrary to the other representations, they offer a view dependent 2.5D representation as each pixel of the map is associated with a single depth value as we can see on Figure 1. Such pixel based representations generally remain noisy even if they provide interesting details on the observed scenes. They are particularly well adapted to describe urban areas from aerial/satellite data (also called Digital Surface Models) as underlined in [18]. Several comparative studies [19], [20], [21] outline the potential of the depth maps by stereo processing.

Meshes provide a description of urban structures with a large amount of details, such as ornament, statues and other irregular shapes. The mesh generation techniques are mainly performed using laser scanning [22], [23] or multi-view stereo processes [24], [25], [26]. Multi-view stereo techniques have significantly progressed during recent years. The comparative studies [27], [28] highlight the quality of such representations, especially for describing details. However, man made objects contain many regular components such as planar or cylindrical shapes. The obtained meshes give a large amount of useless information concerning these regular elements which could be more relevantly described by parametric objects.
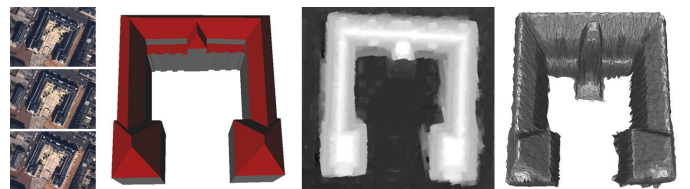

Fig. 1. Various 3D representations of an urban object - *from left to right:* multiview images ©IGN, 3D-primitive modeling, depth map and mesh with triangular facets.

In addition, other types of 3D-representations have been proposed in more specific contexts. For example, level set methods [29], [30], [31] have interesting properties to deform

surfaces. They are efficient for organ modeling, face tracking and more generally variational shape representation but are not well adapted to urban objects.

To sum-up, two main families are used in the literature: the implicit and explicit representations. The former allows object based representations by describing scenes such as 3D-primitive structured arrangements (generally, by using a Constructive Solid Geometry (CSG) principle). The latter, composed of mesh and depth map representations, is unstructured grids which do not provide semantic knowledge on the observed scene but are generally more accurate. Some works have been proposed to convert an explicit representation into an implicit one. Such methods help to extract structural knowledge in a scene. Monte Carlo sampling based models have been proposed in [32], [8] for reconstructing urban scenes as geometric primitive layouts from range data/reflectance maps and disparity maps respectively. It has been also addressed in [33] by simplifying a mesh into a 3D-plane layout, and then extended by [34], [35] for modeling with quadrics. Other works such as [36] allow the detection of regular structures from mesh-based models.

B. Combining implicit and explicit representations

The two families mentioned above have complementary advantages : semantic knowledge and model compaction for the former, detailed modeling and non-restricted use for the latter. A natural idea, but still lightly explored, would consist in merging both the families in order to propose a **hybrid model**. Regular elements would be represented by 3D-primitives whereas irregular structures would be described by mesh-based surfaces. This idea has been partially addressed by several works in specific contexts. The insertion of planar constraints in urban aerial Digital Surface Models (DSM) was proposed by [37]. However, this work is restricted to a roof simplification and cannot be extended to non-planar shapes and general depth maps easily. Both explicit and implicit representations are jointly used in [38]. The two types of information improve the conventional results in human upper-body modeling by taking advantage of their own specificity. However, the process is performed in parallel and does not combine the representations in a hybrid model.

We aim to propose a more general process which can substitute the regular structures from explicit representations by geometric 3D-objects while preserving the irregular details. This is of significant interest for several reasons:

- *Semantic insertion* - The introduction of semantic knowledge in explicit representations provides a better understanding of the urban scenes. For example, 3D-primitives as planes or cylinders give helpful information concerning the locations of walls or columns respectively.
- *Mesh simplification* - Substituting the regular components to 3D-objects strongly simplifies the size of the explicit representations with appealing compression rates.
- *Acquisition process correction* - More still, it allows the correction of some errors generated by the multi-view stereo processes. In particular, some structures which

have been partially reconstructed due to occlusions can be completed thank to 3D-primitives.

In addition, the process has to be general to deal with various types of urban scenes (*i.e.* large urban areas, building roofs and facades, indoor scenes, and generally speaking, the man made creations) and different kinds of inputs (all the explicit representations including mesh-based models and depth maps).

C. Strategy overview

In order to unify the framework, we consider the data as three dimensional boundary meshes with triangular faces. Such a modeling is general enough to represent the different kinds of aforementioned inputs.

Extracting 3D-primitives from explicit representations without a preliminary segmentation is a difficult problem [39] which has been addressed by various works such as [33], [35]. However, such approaches cannot be efficiently adapted for non synthetic explicit representations which contain noise, component density variations, and errors/approximations generated by the multi-view stereo processes.

We adopt a more robust strategy consisting in fitting 3D-primitives from a segmented surface. These two steps are embedded into a hierarchical process which allows the exploration of various object scales in the scene (See Figure 2). For every level of the hierarchy, we obtain an hybrid representation which contains the 3D-primitives extracted in the previous levels.
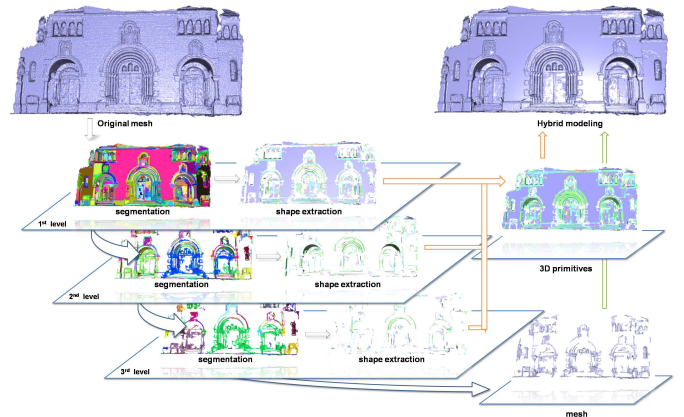

Fig. 2. Overview of our hierarchical approach. Each level of the hierarchy consists in first segmenting the remaining mesh-based surface of the previous level and second extracting 3D-primitives. The hybrid model of a given level is obtained by combining the remaining mesh-based surface of this level and the accumulated 3D-primitives of all the previous levels. Irregular elements such as statues are described by mesh-based surfaces whereas regular structures such as columns or walls are modeled by 3D-primitives.

This paper extends the work we presented in [40] by detailing the model, presenting new applications, results and comments, as well as proposing a new general hierarchical process which significantly improves the performances. The paper is organized as follows. Section II presents the segmentation process based on a curvature analysis of the surface. A multi-label energy which takes topological smoothness constraints into account is formulated. The optimal label configuration which is estimated by α-expansion, provides the segmented surface by region growing. The 3D-primitive extraction from

the obtained partition is then described in Section III. An error parameter controls the fitting quality and decides whether an element of the partition has to be substituted by a plane, sphere, cylinder, cone or torus. Section IV presents the general hierarchical process which takes into account the object scale problem and noise corruption in the urban scenes. Experimental results on real and synthetic data are given in Section V. Basic conclusions are outlined in Section VI.

II. SURFACE SEGMENTATION

Let us consider a three dimensional surface S which can be defined as a boundary mesh $\{V, E, F\}$ of vertices V, edges E and triangular faces F. We aim to segment the surface S into patches corresponding to regions of interest.

A. Geometric attributes based on curvature analysis

Many kinds of local geometric attributes have been proposed in the literature for segmenting synthetic mesh-based surfaces such as multi-scale blowing bubbles, 3D feature descriptors and skeleton knowledge. The comparative studies proposed in [41], [42] present the most efficient techniques for addressing this problem. Most of these techniques cannot be adapted to surfaces generated by multi-view stereo processes due to the meshing irregularities/errors and the significant noise corruption. Such local geometric attributes have to be embedded in a global process.

1) Principal curvatures: Local differential geometry estimates constitute efficient descriptors for analyzing the surface 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 4). In order to distinguish the various types of shapes, this curvature information is used to label each point of the surface 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 ($k_{min}k_{max} \neq 0$).
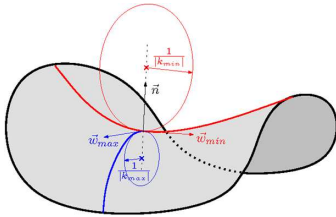

Fig. 3. Representation of the principal curvatures (k_{min}, k_{max}) and their associated direction vectors $\mathbf{w_{min}}$ and $\mathbf{w_{max}}$.

2) Local estimator of the geometry: Let us consider $\mathcal{L} = \{1, 2, 3, 4\}$, the label set corresponding to the classes mentioned above respectively. Let $l = (l_1, ..., l_N)$ be a label configuration in $\mathcal{L}^N$, associated with the N vertices of the surface S. 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:

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

Figure 4 presents the behavior of this probability in function of the couple (k_{min}, k_{max}). The label configuration maximizing the joint probability on the surface S constitutes an interesting estimator, denoted by $\widehat{l_P}$ and given by :

$$\widehat{l_P} = \arg\max_{l \in \mathcal{L}^N} \prod_{i \in V} Pr(l_i|k_{min}^{(i)}, k_{max}^{(i)}) \tag{2}$$

This estimator is simple to compute and provides a reliable labeling in the case of synthetic meshes as we can see with the first two models presented on Figure 5. However, the results obtained from non synthetic surfaces are clearly more limited for the reasons detailed above. Additional information must be taken into account to improve these results. An energy minimization method is proposed in the following to address this problem.
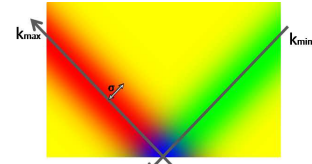

Fig. 4. Map of the label dominance in function of k_{min} and k_{max} (blue sector indicates that the highest probability is obtained for the '*planar*' label, red for '*developable convex*', green for '*developable concave*' and yellow for '*non developable*').

B. 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 $\beta > 0$:

$$U(l) = \sum_{i \in V} D_i(l_i) + \beta \sum_{\{i,j\} \in E} V_{ij}(l_i, l_j) \tag{3}$$

The function U represents the Gibbs energy of the posterior probability distribution of a Markov Random Field (MRF) where the site set corresponds to V (*i.e.* the vertex set) and the neighborhood relationship between sites is given by the edge set E.

a) 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:

$$D_i(l_i) = 1 - Pr(l_i|k_{min}^{(i)}, k_{max}^{(i)}) \tag{4}$$

We have experimentally noticed that such a linear formulation provides more relevant results compared to other functions such as a logarithmic term. It allows us to not penalize too much the non local optimal label in the energy. The sensitivity of the consistency term is controlled by the standard deviation σ of the principal curvature distributions (See Figure 4). Taking a low σ value makes the consistency term more selective with *planar* and *developable* labels. On the contrary, a high value must be chosen for penalizing the occurrence of the *non developable* label.

b) Topological smoothness constraints: The term V_{ij} represents a pairwise interaction potential between adjacent vertices i and j. Prior knowledge about the optimal labeling are expressed by using a model derived from the generalized Potts formulation.

$$V_{ij}(l_i, l_j) = \begin{cases} 1 & \text{if } l_i \neq l_j \\ \min(1, a||\mathbf{W_i} - \mathbf{W_j}||_\mathbf{2}) & \text{otherwise} \end{cases} \quad (5)$$

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

$$\mathbf{W} = \begin{pmatrix} k_{min}.\mathbf{w_{min}} \\ k_{max}.\mathbf{w_{max}} \end{pmatrix} \quad (6)$$

Note that the vectors $\mathbf{w_{min}}$ and $\mathbf{w_{max}}$ are oriented according to a criterion of component positiveness such that the difference between two vectors $\mathbf{W_i}$ and $\mathbf{W_j}$ (see Eq. 5) is computed without sign ambiguity. The potential V_{ij} introduces spatial smoothness constraints which take into account the surface topology. Two principles define the behavior of this term:

- **Smoothness on regular patches -** In order to favor the label homogeneity in a neighborhood, adjacent vertices with different labels are penalized. This principle corresponds to a classical regularization procedure (See Figure 5-2nd and 4th rows).
- **Edge preservation -** The boundaries are preserved by taking into account the principal direction vector variations of adjacent vertices with similar labels. The surface is then partitioned according to changes in local differential geometry. For example, it allows the separation of two connected planes with different normals (See Figure 5-*Corner* model).
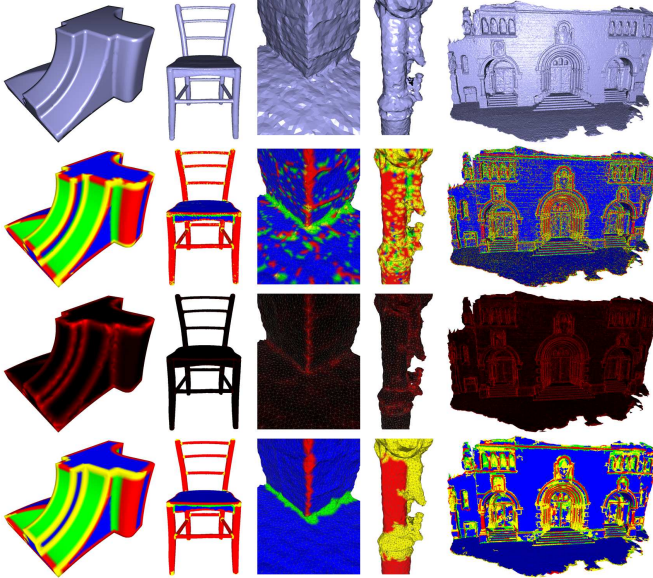

Fig. 5. Surface labeling - *from top to down:* original mesh-based surface, $\widehat{l}_P$ estimator (blue='*planar*', red='*developable convex*', green='*developable concave*' and yellow='*non developable*'), edge term of the regularizing part of the energy ($||\mathbf{W_i} - \mathbf{W_j}||_\mathbf{2}$) (red=high values, black=low values), our labeling result after energy minimization.

C. *Optimization by α-expansion*

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. The α-expansion algorithm [43] based on the Graph-cuts theory is

used. 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 optimum. To accelerate the convergence, $\widehat{l_P}$ is chosen as the initialization. Note that faster algorithms such as Logcut [44] could be used. However, the time savings would be minor since the number of labels we use remains reasonable.

D. *Comments*

Figure 5 shows results of the labeling process. 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 and vaultings with *developable convex* or *developable concave* labels - and ornaments and statues with the *non developable* one. The edges are accurately localized due to the detection of principal direction vector variations (See 5-3^{rd}row). It allows us to extract these components easily with a region growing process as we can see on Figure 6. Nevertheless, this labeling process does not take into account the different object scales of the scenes since σ is a constant value. Thus, the partitioning is not optimal for complex urban surfaces. This problem is addressed in Section IV by proposing a hierarchical approach. In the sequel, we define as *cluster* a connected region of same label extracted by region growing.
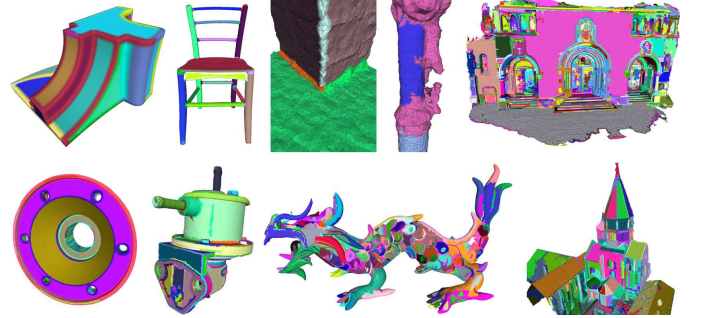

Fig. 6. Segmentation results on real and synthetic mesh-based surfaces. Each color represents a cluster.

III. GEOMETRIC SHAPE EXTRACTION

Each cluster of the segmented surface 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 in urban landscapes and more generally in scenes containing man made objects.

A. *3D-primitive fitting*

Several works such as [45], [46], [47] have been proposed to detect shapes in point clouds containing outliers. Mesh-based surfaces have generally less outliers compared to point clouds 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 performed easily using a Principal Component Analysis (PCA). However, fitting spheres, cylinders, cones and 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 algorithm on the work of [48] which proposes a first order approximation $\widetilde{d}$ of the true Euclidean distance d from points to spheres, cylinders, cones and tori, that is both easier to compute and better behaved with respect to singularities. The error ϵ between the 3D-primitive based surface denoted by s and the vertex set $(v_k)_{k \in [1,N]}$ of the original mesh-based surface is then given by Eq. 7.

$$\epsilon(s) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} \widetilde{d}(s, v_k)^2} \qquad (7)$$

We also use the parametrizations of the various primitives introduced in [48], that behave well as their curvatures vanish. As shown on Figure 7, 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...). However, this method fails in a significant number of cases due to the non-linear optimization procedure which depends on the initialization quality. Thus, we propose some improvements which are detailed in the following.
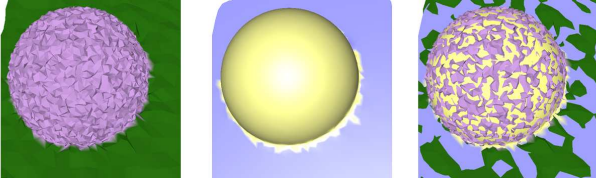

Fig. 7. 3D-primitive fitting on a planar/spherical surface corrupted by noise - *from left to right:* segmented surface, extracted primitives, and overlaid results.

B. Improving the local optimization by a multi-initialization strategy

Relying on a local optimization, the quality of the fitting process strongly depends on its initialization. One could use a global optimization technique such as stochastic relaxation based methods [49], but the computing time would be extremely high and the global configuration would not necessarily be reached. The most relevant solution consists in improving the initialization of the primitive parameters.

[48] proposes to estimate a single initialization using a single 3D point and its local differential estimates. For a more robust appreciation, we propose to let multiple such initializations based on various local estimates compete and keep the overall best fit, *i.e.* the primitive minimizing the error ϵ. 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 [48] 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. Such a process requires higher computation times but gives superior fitting rates than the original procedure [48].
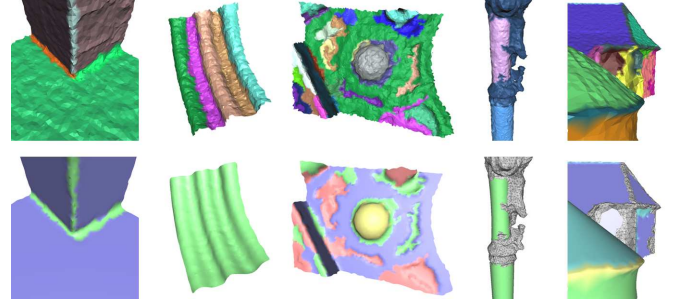

Fig. 8. Shape extraction on some examples *(purple=plane, pink=cylinder, blue=cone, yellow=sphere, green=torus)* with the associated segmented surfaces.

C. Using labeling information for accelerating the extraction procedure

The labeling information obtained in the segmentation stage is used to drive the shape extraction. This knowledge is particularly useful since it avoids an exhaustive comparison between a cluster and all the types of 3D-primitives.

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 fitting quality threshold ξ controls the choice between 3D-primitive and mesh-based surface. If the quadratic error $\epsilon(s)$ of the optimal found primitive s is lower than ξ, the cluster is substituted by the detected primitive whose border is computed by projecting the cluster border on the primitive surface. Otherwise, the rejected cluster is compared to the other types of 3D-primitives. This second fitting test prevents incorrect labellings generated by scale ambiguities. For example, the large vaultings on Figure 5 are mistakenly labeled as *planar* clusters due to the low values of their principal curvatures. The additional test correctly fits these vaultings to cylinders (See Figure 12). Finally, if the cluster is still rejected during this second test, it keeps its original mesh-based representation.

By Performing this shape extraction procedure on all the clusters resulting from the segmentation step, we can obtain a first **hybrid** model which combines 3D-primitives and mesh-based surfaces as illustrated in Figure 8-two right examples.

IV. HIERARCHICAL APPROACH

A. Object scale and noise corruption

The successive performing of the segmentation and shape extraction processes (see Sections II and III respectively) has been realized in [40]. Such an approach gives good results but the number of found 3D-primitives remains limited. In many locations the segmentation is not optimal. Some clusters

contain multiple shapes and fail to be modeled by one single 3D-primitive. As we can see in Figure 9, a high radius cylindrical shape continuously connected to a plane structure is segmented as one unique cluster if the value of σ is too high, as well as a spherical surface connected to a corrupted-by-noise structure if σ is too low. To sum-up, performing an approach with a constant σ value does not give an optimal partitioning of the surface in complex urban scenes since the various scales of the objects and the noise generated by acquisition process are not taken into account in the segmentation.
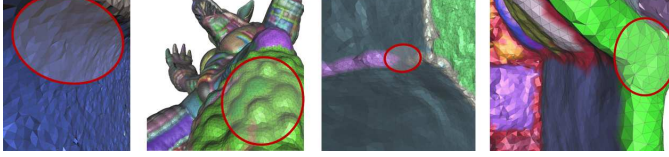

Fig. 9.   Examples of non optimal partitionings which generate the primitive extraction failure.

A local estimation of σ could be a solution in order to get a variable σ value on the surface. However, the extraction of object scale information from urban scenes is complex without semantic knowledge and makes the estimation especially difficult to perform. We adopt a hierarchical process which consists in exploring various object scales of the scene; each exploration being realized using a bidirectional segmentation with two σ values. This process is detailed in the following.

B. Bidirectional segmentation

The bidirectional segmentation allows a better partitioning of the corrupted-by-noise patches and so a better extraction of the 3D-primitives. It consists in successively partitioning twice the surface by using the segmentation process proposed in Section II.
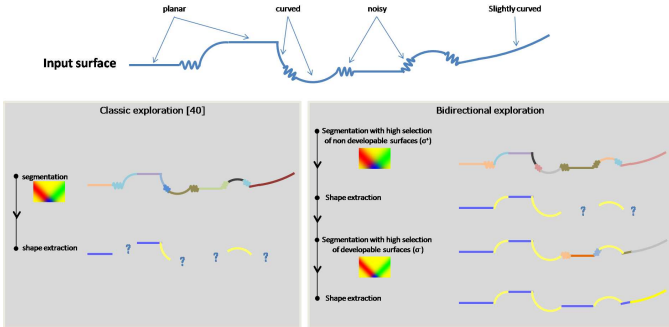

Fig. 10.   Bidirectional segmentation from a surface corrupted by noise and comparison with the procedure used in [40]. Blue and yellow segments indicate that shape extraction procedure achieves the fitting of developable and non developable patches respectively.

The first segmentation is performed using a high σ value (called σ^+) in order to be both selective with the non developable patches and tolerant with those that are noisy and developable (See Figure 10). 3D-primitives are extracted from the obtained partition by the shape extraction proposed in Section III. Then, a second segmentation is performed from the remaining mesh-based surfaces. This time, the segmentation is lead with a low σ value (called σ^-) to be both selective with the developable patches and tolerant with the noisy non

developable patches. Finally, we proceed to a second shape extraction from the obtained partition. In the following, we call $\mathcal{S}_{2D}(S, \sigma^+, \sigma^-, \beta)$, the bidirectional segmentation of the overall surface S by using the parameter set $(\sigma^+, \sigma^-, \beta)$.

C. Hierarchical exploration of the geometry

The bidirectional segmentation is embedded into a hierarchical approach which allows the exploration of the various scales of the scenes. At each iteration, we try to extract geometric shapes from the previously rejected patches by exploring a new object scale. To do so, a bidirectional segmentation is performed using new σ^+ and σ^- values. In addition, the smoothness constraints of the segmentation process are relaxed by decreasing the value of β in order to refine the surface partitioning.
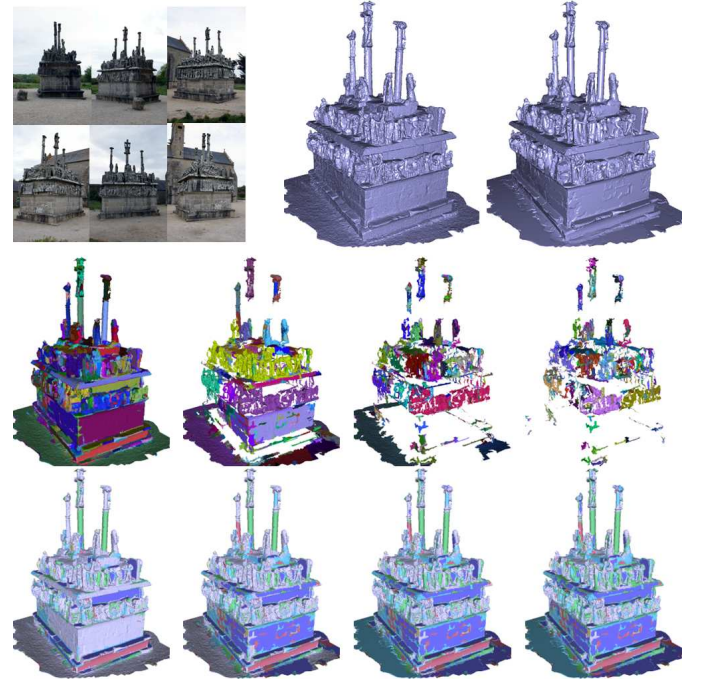

Fig. 11.   Illustration of the hierarchical process on the Calvary model - (first row) multiview images, mesh-based surface by [26], final hybrid representation, (second and third rows) segmented surface and hybrid representation at different levels.

Let $H_0 = (S_0, P_0)$ be the initial hybrid model with $S_0 = M$, the original mesh-based surface, and $P_0 = \emptyset$. Let σ_0^+, σ_0^- and β_0 be the initial parameters of the bidirectional segmentation with $\sigma_0^+ = \sigma_0^- = \sigma_0$. Then the hierarchical exploration can be formulated as follow:

At the hierarchical level $i > 0$,

1: Updating the parameter set such as $\sigma_i^+ = \sigma_{i-1}^+ \gamma_1$, $\sigma_i^- = \sigma_{i-1}^- \gamma_1^{-1}$ and $\beta_i = \beta_{i-1} \gamma_2^{-1}$

2: Performing the bidirectional segmentation $\mathcal{S}_{2D}(S_{i-1}, \sigma_i^+, \sigma_i^-, \beta_i)$ in order to obtain the couple (S_i, P_i)

3: Collecting the corresponding hybrid representation $H_i = (S_i, \bigcup_{0 \le k \le i} P_k)$

γ_1 and γ_2 are the hierarchical parameters which define how different are the exploring scales. In our experiments, we fix $\gamma_1 = \gamma_2 = \frac{3}{2}$. Figure 11 describes the hierarchical process on the Calvary model and presents hybrid representations at different levels. The first levels allow the extraction of the biggest regular components such as columns, planar background behind statues and wall sides whereas the last levels explore smaller more irregular patches in order to extract spherical shapes such as statue heads or conical elements as the statue dresses. This hierarchical process efficiently tackles the object scale and noise corruption problems presented on Figure 9. For instance, the high radius cylindrical shape continuously connected to a plane structure in the foreground is segmented as one unique cluster in the first hierarchy level and so cannot be properly fitted. Such a problem is solved in the second level where the high radius cylindrical shape and the plane structure are successfully separated.

V. EXPERIMENTS

Our approach is tested on both real surfaces generated by the multi-view stereo techniques and synthetic models. To our knowledge, there is no other method proposing hybrid representations. However, we evaluate our results qualitatively and quantitatively with a visual appreciation, a compression rate study and an accuracy improvement experiment. We also underline the interest of our method by proposing subsequent processes for semantic identification. These experiments have been lead with the error parameter ξ fixed to $0.5\widehat{e}$ and with four hierarchy levels in order to allows the extraction of both major structures and smaller components. This value constitutes an interesting compromise between good compression rates and visually convincing results. Note that the presented algorithm was implemented in C++ and relies on the CGAL library (*www.cgal.org*) for the computation of the geometric attributes.

A. Real surfaces from multiview stereo

Figure 12 shows the potential of the method on mesh-based surfaces whereas Figure 13 presents results on depth maps corresponding to aerial urban scenes (*i.e.* Digital Surface Models). The obtained hybrid representations are promising and provide simplified modelings of the original surfaces 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 useful because they allow the introduction of semantic information in the modeling. Structural components such as walls, roofs, windows and dormer windows can be identified easily from the obtained primitives by a subsequent basic analysis as we will see on Section V-D. The results reveal the reconstruction of interesting details such as thin pipes located at the vaultings on *Herz-Jesu-P25* or small statue heads on *Calvary*.
As we can see on Table I, the hybrid models are mainly composed of 3D-primitives. The unfitted patches are negligible for scenes containing many regular components such as the *Church* or *Herz-Jesu-P25* models. To the contrary, the scenes with few man made structures such as the *Aiguille du Midi* model remain more limited in terms of extracted primitives.

Nevertheless, the process gives promising results on natural elements such as pieces of rock or snow. Note that the number of 3D-primitives could be reduced by merging the neighboring primitives which are similar. However such a subsequent process is of limited interest with respect to our aims.
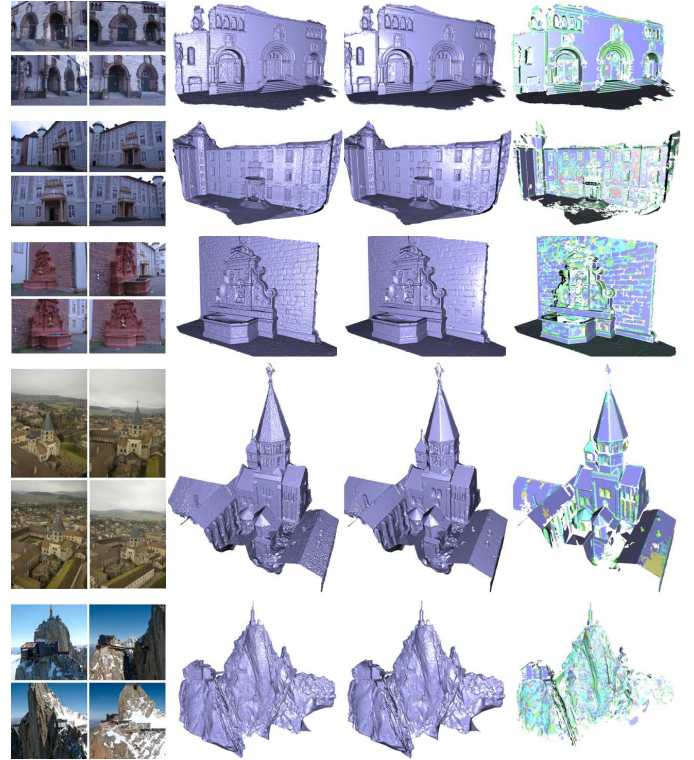

Fig. 12.   Results on large scenes- *from left to right:* some multiview images, mesh-based surfaces from [26], our hybrid models and the associated 3D-primitives *(purple=plane, pink=cylinder, blue=cone, yellow=sphere, green=torus). From top to bottom:* Herz-Jesu-P25, Castle-P10, Fountain-P11, Church, Aiguille du Midi.

TABLE I
ADDITIONAL INFORMATION ON THE TESTS PRESENTED ON FIGURE 12. FOR EACH TEST, WE SPECIFY THE NUMBERS OF VERTEX AND FACET IN BOTH THE INITIAL AND HYBRID MODELS AS WELL AS THE NUMBER OF EXTRACTED PRIMITIVES.

	initial model (vertex/facets)	number of primitives	remaining mesh (vertex/facets)
Herz-Jesu-P25	$0.72M / 1.41M$	346	$0.15M / 0.26M$
Castle-P10	$0.99M / 1.99M$	649	$0.27M / 0.46M$
Fountain-P11	$0.59M / 1.17M$	509	$0.12M / 0.22M$
Church	$0.16M / 0.33M$	173	$31K / 60K$
Aiguille Midi	$2.27M / 4.56M$	2157	$0.81M / 0.22M$
Calvary	$0.79M / 1.56M$	612	$1.45M / 0.38M$

Figure 13 presents some results on building reconstruction from Digital Surface Models. These depth maps are generally of low quality due to the restricted number of multiview aerial images which are available (in our case, no more than three), and the low value of the base to height ratio which generates a poor altimetric accuracy (in our case, the altimetric accuracy is five times more than the planimetric one). Thus, such surfaces are especially noisy and irregular. The proposed results are promising and allow us to significantly regularize and correct the initial Digital Surface Models which have been obtained by an algorithm based on [50]. Our hybrid

models are mainly composed of planes which fit well to the various roof sides of buildings. The models remain limited in terms of superstructure reconstruction (*e.g.* chimneys, dormer-windows, glass roofs) because these elements are too small and irregular to be properly extracted. A more complex shape library including plane-parallel primitives such as in [8] or [15] is required in this case. Even if our representation is less compact and accurate than the specialized building reconstruction approach proposed by [8], our method is more general and can successfully represent curved 2D-footprints as we can see on the third example. The structural approach [8], which is restricted to rectilinear 2D-footprints, cannot provide such details.
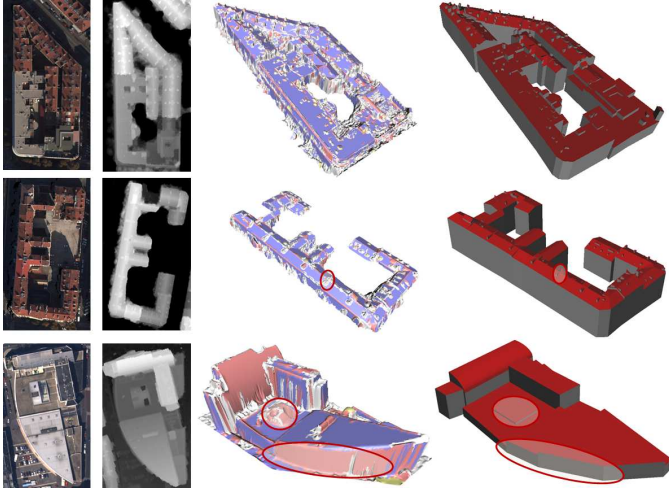

Fig. 13. Aerial building reconstruction results- *from left to right:* aerial stereo image, associated Digital Surface Model ©IGN, our hybrid model *(purple=plane, pink=cylinder, blue=cone, yellow=sphere, green=torus, white=mesh)*, parametric model by [8].

B. Synthetic surfaces

Tests from synthetic surfaces have also been realized. The interest of such experiments is more limited than previous tests. However, it can be useful for the simplification and the denoising of mesh-based surfaces. In addition, performing tests on simulated surfaces allows us to evaluate the correctness of our method and to validate it.

Figure 14 presents some hybrid models from synthetic meshes. The results are globally convincing as our hybrid models are mainly composed of 3D-primitives which correspond well to the different object scales of the initial surface. The various primitive types are correctly detected on the *Thai Statue* and no type is favored with respect to the others. We can note that the planar shape occurrence is lower in such non urban scenes. The remaining surface rate of the *Asian Dragon* is higher than the mechanical object example because it varies according to the level of the surface regularity.

Tests on noise corruption impact have also been realized (see Figure 15). If a low level of noise has almost no consequences on the result, our method loses robustness when the noise becomes very significant and impedes the visual description of the scene. Some details are lost such as the grid pattern on the Stanford Armadillo tigh which is then modeled by a toroidal shape. In addition, the noise makes the extraction of
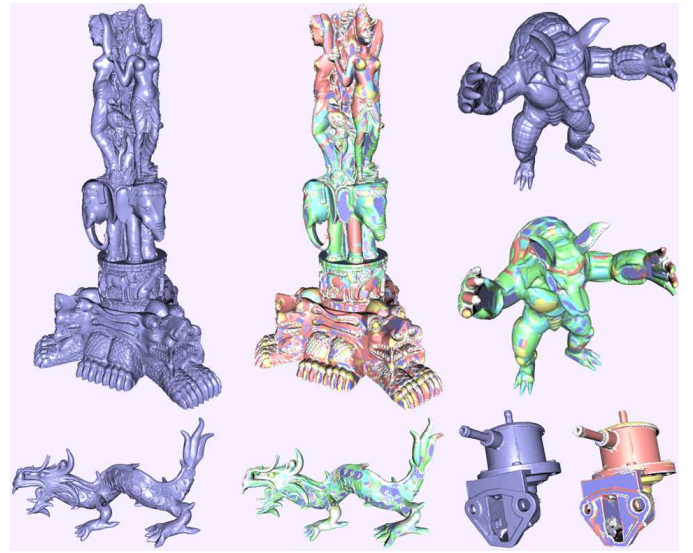

Fig. 14. Synthetic mesh simplification- original synthetic meshes and our associated hybrid models *(purple=plane, pink=cylinder, blue=cone, yellow=sphere, green=torus, white=mesh). From top to bottom, and from left to right:* Stanford Thai Statue, Stanford Armadillo, Stanford Asian Dragon and a mechanical object.

the correct shape difficult without a Ransac based algorithm. For example, the inside conical part of the Carter model is mistakenly identified as a spherical object. Even if the result remains correct, our method is not optimal when the meshes are too strongly corrupted.
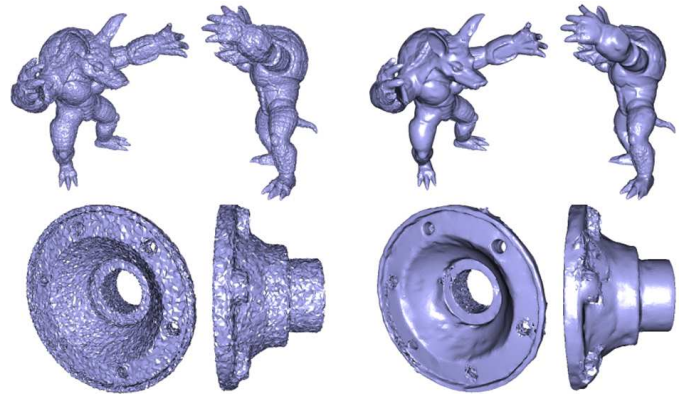

Fig. 15. Noise corruption impact - synthetic meshes corrupted by noise (*top:* Stanford Armadillo, *bottom*: Carter) and the associated hybrid models.

C. Performances

Compression rate- The compression rate, defined as the ratio between the file sizes of the original mesh-based surface and the hybrid representation, is function of the error parameter ξ. Table II 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 compression rates are significantly better than those obtained in [40]: the hierarchical approach allows a finer exploration of the surface and thus a better extraction of 3D-primitives. For example, the compression rate of the Herz-Jesu-P25 model is 4.48 with an error parameter

$\xi = 0.5$ whereas the rate obtained in [40] is only 3.57. Figure 16 compares our method with a state-of-the-art decimation method (the one in CGAL library. 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.

TABLE II
COMPRESSION RATES ON SOME MODELS IN FUNCTION OF THE PARAMETER ξ.

	$\xi = 0.1\widehat{e}$	$\xi = 0.5\widehat{e}$	$\xi = \widehat{e}$	$\xi = 5\widehat{e}$
Church	1.31	4.57	6.4	32.11
Herz-Jesu-P25	1.19	4.48	14.63	48.93
Fountain-P11	1.14	4.09	10.32	23.6
Stanford Armadillo	1.18	3.69	7.4	18.45

Accuracy improvement- In order to quantify to what extent our representation improves the original surface, we evaluate the error with respect to the (range scanned) ground truth. As we can see on Figure 16, some details are lost in the hybrid representation due to the simplification by 3D-primitives (*e.g.* walls are transformed to planes and lose their small asperities). However, this drawback remains negligible with respect to the multiview stereo error corrections which are realized in terms of surface denoising and structure completion. For instance, the noisy balls and waveform ornaments are regularized by half a sphere and a set of toroidal patches respectively in Figure 8. The partial structures can also be completed as we can see on Figure 16 where a partial column is integrally reconstructed. The error occupancy histogram, measured with respect to the standard deviation Σ of the ground truth accuracy (see [28]), quantifies this improvement. On the Herz-Jesu-P25 histogram, we note for example that close to 3% of the global error occupancy is transfered from the 3 and 4Σ bins to the 1 and 2Σ ones. This is indeed a first step toward a more extensive evaluation, as this improvement seems to concern subsets of the surface which have to be identified in detail and closely analyzed. Yet the benchmark website of [28] only outputs global statistics and does not easily allow this investigation.

Computation time and memory consumption- The proposed method achieves reasonable computation time. For example, the estimation of the various geometric attributes and the surface segmentation method are performed in less than one minute on the Castle-P10 model containing one million vertices. The primitive fitting procedure requires longer computation time but remains acceptable: around four minutes are necessary to test the primitive relevancy for all the clusters of the Castle-P10 model. Note that the computation time decreases at each level of the hierarchical procedure thanks to the primitive accumulation process. The memory consumption is quite low: the algorithm just needs to store the original mesh structure with several additional attributes on each vertex and facet.

Limitations- Our process is not well adapted to point density variations over the mesh-based surface. This problem can appear on the real surfaces from multiview stereo where the visibility from images is very weak. It mainly concerns the
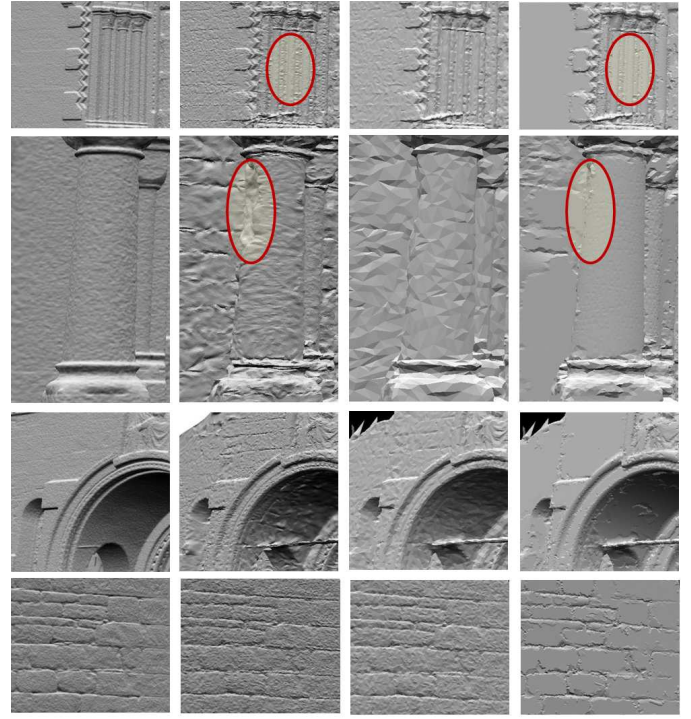

Fig. 16. Comparisons on some crops- *From left to right, top :* details of ground truth [28], original mesh-based surface [26], decimated surface, and our representation (same compression rate as the decimated surface).

borders and the hidden parts of the surfaces. For instance, this drawback is illustrated on Figure 12 where some roof parts of the castle-P10 model which have a very low point density do not fit 3D-primitives. To solve this problem the scale factor a, used in the multi-label Markov Random Field formulation (see Section II-B), should be locally estimated. a is currently computed over the surface in a global approach so that our multi-label energy check the requirements of the α-expansion optimization technique. A local estimation of the scale factor a would imply the use of alternative techniques providing higher computing time. For example, one could imagine estimating the scale factor a in a local spherical neighborhood along the surface and using Monte Carlo Markov Chain based stochastic method [51] to solve the energy minimization.
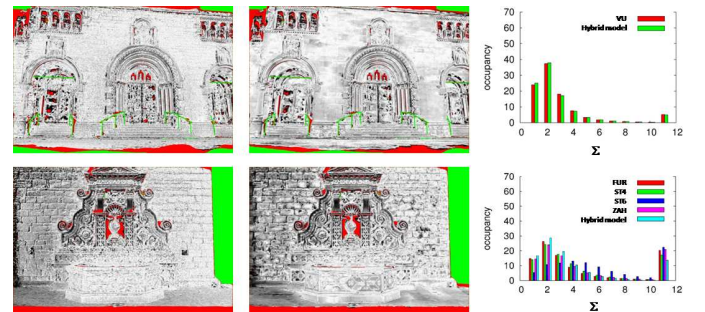

Fig. 17. Accuracy evaluation- *From left to right:* error of the original mesh-based surface with respect to ground truth (white=low, black=high), error of our representation, histogram of the errors [28].

The method is also less competitive when the surface is strongly corrupted by noise. Several shape selection errors are

shown on Figure 15, even from a simple mesh-based surface. Outlier rejection algorithms have to be performed to improve 3D-primitive extraction on such highly noisy surfaces.

D. Other applications by subsequent processes

In order to underline the interest of our method, we propose an overview of two simple subsequent processes which allow the extraction of more semantic information from our hybrid representation.

Fully implicit representation- We can convert a hybrid representation into a parametric one, *i.e.* a model which is strictly composed of 3D-primitives according to a CSG principle. In other words, we transform our model into a fully implicit representation. First, a filtering procedure is realized in order to remove useless components of the hybrid model. The mesh-based patches and the non significant 3D-primitives are deleted. Then, a primitive arrangement is performed on the remaining objects. Figure 18-*top* presents the results of a preliminary version restricted to planar structure arrangements [52]. Although this process can be used in a large range of urban scenes such as building roof/facade reconstruction, the plane arrangement has to be extended in the future to the other 3D-primitives (*i.e.* cylinder, sphere, cone and torus).

Semantic identification- We can also identify urban components from our hybrid model. Depending on the wanted elements, some procedural analysis can be easily proposed for identifying 3D-primitives and mesh-based patches according to various criteria such as the type, the size, the area, the orientation or the repetitiveness. Figure 18-*bottom* shows a result of facade classification where we have defined simple heuristics to distinguish four types of components (*i.e.* walls, roofs, windows and other). We could also propose a learning process by defining appropriate kernels which measures the similarity between 3D-primitives and takes into account neighborhood geometrical considerations and which could improve the existing methods [53] [54].
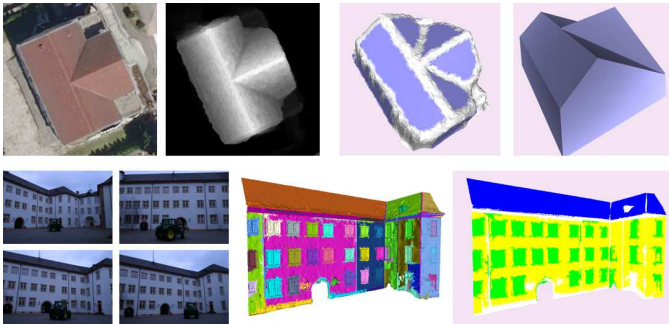

Fig. 18. Examples of subsequent applications from hybrid models- *top:* Convertion of an hybrid representation *(3rd row)* into a parametric model *(4th row)* by using a plane arrangement algorithm [52], associated aerial image *(1st row)* and Digital Surface Model *(2nd row). bottom:* facade component classification *(3rd row)* by procedural analysis (blue=roof, yellow=wall, green=window and white=others), associated multiview images *(1st row)* and segmented mesh-based model *(2nd row).*

VI. CONCLUSION

We propose an original hybrid representation of urban scenes which combines 3D-primitives and mesh-based sur-

faces. Our approach possesses several interesting characteristics. First, it provides high compression rates while keeping details, introduces semantic knowledge despite noise corruption, and even improves accuracy of the original surface. Convincing results have been obtained on various applications including semantic insertion, surface simplification or acquisition process correction. Both the proposed multi-label Markov Random Field formulation for surface segmentation and the contributions for 3D-primitive fitting can be used separately for other meshing applications. Moreover, our approach can be used in various contexts where types of urban scenes (*i.e.* large urban areas, building roofs and facades, indoor scenes) and kinds of inputs (*i.e.* multiview based meshes, depth maps, synthetic surfaces) differ. Finally, the user can control both accuracy and level of details via ξ and the number of hierarchy levels respectively.

In the future, the preliminary algorithms presented in Section V-D would have to be developed. Also of interest would be to study the simultaneous generation of mesh patches and 3D-primitives during the multi-view stereo process. This would allow us to take interactions between mesh-based surfaces and primitives into account, but would require advanced 3D-primitive samplers which strongly increase computing time.

REFERENCES

[1] Z. Zhu and T. Kanade, "Special issue on modeling and representations of large-scale 3D scenes," *International Journal of Computer Vision*, vol. 78, no. 2-3, 2008.

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

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

[4] S. N. Sinha, D. Steedly, R. Szeliski, M. Agrawala, and M. Pollefeys, "Interactive 3D architectural modeling from unordered photo collections," *ACM Transactions on Graphics*, vol. 27, no. 5, 2008.

[5] A. Dick, P. Torr, and R. Cipolla, "Modelling and interpretation of architecture from several images," *International Journal of Computer Vision*, vol. 60(2), pp. 111–134, 2004.

[6] Q. Zhou and U. Neumann, "A streaming framework for seamless building reconstruction from large-scale aerial lidar data," in *Proc. of IEEE Computer Vision and Pattern Recognition*, Miami, United States, 2009.

[7] C. Poullis and S. You, "Automatic reconstruction of cities from remote sensor data," in *Proc. of IEEE Computer Vision and Pattern Recognition*, Miami, United States, 2009.

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

[9] L. Zebedin, J. Bauer, K. Karner, and H. Bischof, "Fusion of feature- and area-based information for urban buildings modeling from aerial imagery," in *Proc. of European Conference on Computer Vision*, Marseille, France, 2008.

[10] C. Baillard and A. Zisserman, "Automatic reconstruction of piecewise planar models from multiple views," in *Proc. of IEEE Computer Vision and Pattern Recognition*, Los Alamitos, United States, 1999.

[11] J. Xiao, T. Fang, P. Tan, P. Zhao, E. Ofek, and L. Quan, "Image-based faade modeling," *ACM Transactions on Graphics*, vol. 27, no. 5, 2008.

[12] C. Brenner and N. Ripperda, "Extraction of facades using RJMCMC and constraint equations," in *Proc. of ISPRS Photogrammetric and Computer Vision*, Bonn, Germany, 2006.

[13] Y. Furukawa, B. Curless, S. Seitz, and R. Szeliski, "Reconstructing building interiors from images," in *Proc. of IEEE International Conference on Computer Vision*, Kyoto, Japan, 2009.

[14] P. Muller, G. Zeng, P. Wonka, and L. Van Gool, "Image-based procedural modeling of facades," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 614–623, 2007.

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

[16] F. Han and S. Zhu, "Bottom-up/top-down image parsing by attribute graph grammar," in *Proc. of IEEE International Conference on Computer Vision*, Beijing, China, 2005.

[17] S. Lee and R. Nevatia, "Extraction and integration of window in a 3d building model from ground view images," in *Proc. of IEEE Computer Vision and Pattern Recognition*, Washington, United States, 2004.

[18] H. Hirschmuller, "Stereo processing by semi-global matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328 – 341, 2008.

[19] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. of IEEE Computer Vision and Pattern Recognition*, Minneapolis, United States, 2007.

[20] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in Computational Stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993–1008, 2003.

[21] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1-2-3, pp. 7–42, 2002.

[22] C. Fruh and A. Zakhor, "An automated method for large-scale, ground-based city model acquisition," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 5–24, 2004.

[23] 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," *International Journal of Computer Vision*, vol. 78, no. 2-3, pp. 207–222, 2008.

[24] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," in *Proc. of IEEE Computer Vision and Pattern Recognition*, Minneapolis, United States, 2007.

[25] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. Seitz, "Multi-view stereo for community photo collections," in *Proc. of IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, 2007.

[26] H. Vu, R. Keriven, P. Labatut, and J.-P. Pons, "Towards high-resolution large-scale multi-view," in *Proc. of IEEE Computer Vision and Pattern Recognition*, Miami, United States, 2009.

[27] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proc. of IEEE Computer Vision and Pattern Recognition*, New York, United States, 2006.

[28] 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 *Proc. of IEEE Computer Vision and Pattern Recognition*, Anchorage, United States, 2008.

[29] C. Shen, J. F. O'Brien, and J. Shewchuk, "Interpolating and approximating implicit surfaces from polygon soup," *ACM Transactions on Graphics*, vol. 23, no. 3, 2004.

[30] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2003.

[31] R. Keriven and O. Faugeras, "Variational principles, surface evolution, PDEs, level set methods and the stereo problem," *IEEE Transactions on Image Processing*, vol. 7, no. 3, 1998.

[32] F. Han, Z. W. Tu, and S. Zhu, "Range image segmentation by an effective jump-diffusion method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1138 – 1153, 2004.

[33] D. Cohen-Steiner, P. Alliez, and M. Desbrun, "Variational shape approximation," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 905–914, 2004.

[34] J. Wu and L. Kobbelt, "Structure recovery via hybrid variational surface approximation," *Computer Graphics Forum*, vol. 24, no. 3, pp. 277–284, 2005.

[35] D. Yan, Y. Liu, and W. Wang, "Quadric surface extraction by variational shape approximation," in *Proc. of Geometric modeling and processing*, Pittsburgh, United States, 2006.

[36] M. Pauly, N. J. Mitra, J. Wallner, H. Pottmann, and L. Guibas, "Discovering structural regularity in 3D geometry," *ACM Transactions on Graphics*, vol. 27, no. 3, 2008.

[37] N. Chehata, M. Pierrot-Deseilligny, and G. Stamon, "Hybrid DEM generation constrained by 3d-primitives : A global optimization algorithm using graph cuts," in *Proc. of IEEE International Conference on Image Processing*, Genoa, Italy, 2005.

[38] S. Ilic and P. Fua, "Implicit meshes for surface reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 328 – 333, 2006.

[39] C. Dorai and A. Jain, "COSMOS - a representation scheme for 3D free-form objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 10, pp. 1115–1130, 1997.

[40] F. Lafarge, R. Keriven, and M. Brédif, "Combining meshes and geometric primitives for accurate and semantic modeling," in *Proc. of British Machine Vision Conference (BMVC)*, London, United Kingdom, 2009.

[41] A. Shamir, "A survey on mesh segmentation techniques," *Computer Graphics Forum*, vol. 27, no. 6, pp. 1539–1556, 2008.

[42] M. Attene, S. Katz, M. Mortara, G. Patane, M. Spagnuolo, and A. Tal, "Mesh segmentation - a comparative study," in *Proc. of IEEE International Conference on Shape Modeling and Applications*, Washington, United States, 2006.

[43] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222 – 1239, 2001.

[44] V. Lempitsky, C. Rother, and A. Blake, "Logcut - efficient graph cut optimization for markov random fields," in *Proc. of IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, 2007.

[45] V. Verma, R. Kumar, and S. Hsu, "3D building detection and modeling from aerial LIDAR data," in *Proc. of IEEE Computer Vision and Pattern Recognition*, New York, United States, 2006.

[46] R. Schnabel, R. Wahl, and R. Klein, "Efficient ransac for point-cloud shape detection," *Computer Graphics Forum*, vol. 26, no. 2, pp. 214–226, 2007.

[47] P. Labatut, J.-P. Pons, and R. Keriven, "Hierarchical shape-based surface reconstruction for dense multi-view stereo," in *Proc. of IEEE International Workshop on 3-D Digital Imaging and Modeling*, Kyoto, Japan, 2009.

[48] D. Marshall, G. Lukacs, and R. Martin, "Robust segmentation of primitives from range data in the presence of geometric degeneracy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 304–314, 2001.

[49] M. Metropolis, A. Rosenbluth, A. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, pp. 1087–1092, 1953.

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

[51] W. Hastings, "Monte Carlo sampling using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[52] F. Taillandier and R. Deriche, "3D reconstruction of linear primitives from multiple images for urban area modelisation," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 34, no. 3A, 2002.

[53] A.-L. Chauve, J.-P. Pons, J. Audibert, and R. Keriven, "Transductive segmentation of textured meshes," in *Proc. of Asian Conference on Computer Vision*, Xi'an, China, 2009.

[54] A. Delaunoy, K. Fundana, E. Prados, and A. Heyden, "Convex multi-region segmentation on manifolds," in *Proc. of IEEE International Conference on Computer Vision*, Kyoto, Japan, 2009.