Efficient Visualization of Architectural Models from a Structure and Motion Pipeline

M. Farenzena, A. Fusiello and R. Gherardi

Dipartimento di Informatica - University of Verona Strada Le Grazie, 15 - 37134 Verona

Abstract

State of the art three dimensional reconstruction pipelines can nowadays produce models up to several million polygons without any human intervention from a set of digital images or video. Such models are able to stretch the rendering capabilities of current hardware. We propose to augment a typical structure from motion pipeline with two additional steps, automatic fitting of high-level solid primitives and relief maps extraction, thus recovering both the overall structure of a building and its fine geometry. This not only gives birth to a more tractable and semantic model of the imaged scene, but allows for efficient and compelling rendering. We substantiate our claims showing a complete example of the described system.

Categories and Subject Descriptors (according to ACM CCS): I.4.8 [Computer Vision]: Scene Analysis; I.3.5 [Computer Graphics]: Object Representations

1. Introduction

The recent advances in Structure and Motion pipelines coupled with the availability of large repositories of digital photos and aerial images have enabled the creation of some of the largest architectural models ever composed [GSC*07, Mea07].

Even if the problems arising in the visualization of large, detailed urban environments have been actively investigated, rendering such massive amounts of data is still problematic. The major source of difficulty lies in the fact that there is virtually no limit to the nature and quantity of the acquired details. State of the art systems can already build scenes with features spanning more than three scales of magnitude. On top of that, recovered meshes suffer from uneven sampling and connectivity problems.

Such magnitude and complexity is able to stretch the rendering capabilities of current rendering platforms, even when taking into account their steady power growth. To speed up the visualization process and to counter the exponential increase in size of the recovered data we propose to augment the typical structure from motion pipeline with two additional steps, high-level primitive fitting and relief map extraction.

High-level primitives such as planes and generalized cones are ideal descriptors for architectural buildings and in general human manufacts. Automatically fitting such primitives to the outputs of a reconstruction pipeline enable the characterization of structure and the extraction of high level properties (such as symmetry, or function) and unseen geometry. Relief map extraction recovers the fine geometry that is lost in the previous step, and stores it in a compact format directly usable by graphic hardware.

The final output of our system is a set of automatically recovered geometric primitives, relief maps and textures that can be used to concisely describe and to efficiently render the imaged scene. The process leverages the former dense point cloud to a sensible, editable representation ready for manipulation in a CAD software.

The approaches covered in the literature for solving the problem of urban reconstruction can be categorized in two main branches: a first one [SSS06, BL05, Kea06] is composed of the *Structure and Motion* (SaM) pipelines that are able to handle the reconstruction process making no assumptions on the imaged scene and without manual intervention. These methods usually share a common structure and produce as output, along with camera parameters, an arbitrarily dense but ultimately unorganized point cloud which fails to model surfaces.

The second category comprises the methods specifically tailored for urban environments and engineered to be mounted on survey vehicles [Mea07, CCG06]. These systems usually rely on a host of additional information, such as GPS and inertial sensors, and output dense polygonal maps using stereo triangulation.

The recovery of the semantic structure of urban elements has been tackled by fewer researchers. In this respect, the two most similar articles to the work presented here are [Dea04] and [SB03]. In [Dea04] is described a system that specializes in creating a architectural models from a limited number of images. Initially a coarse set of planes is extracted by grouping point features; the models are subsequently refined by casting the problem in a Bayesian framework where priors for architectural parts such as doors and windows are incorporated or learnt. A similar deterministic approach is developed in [SB03] where dominant planes are recovered using a orthogonal linear regression scheme: façade features, which are modeled as shaped protrusions or indentations, are then selected from a set of predefined templates. Both methods rely on a large amount of prior knowledge to operate, either implicitly or explicitly, and make strict assumption on the imaged scene.

In our approach instead, the amount of injected prior knowledge is limited to the non-critical type and number of primitives used: the recovery process rather than being top-down is entirely data-driven, and structure emerges from the data rather than being dictated by a set of potentially incorrect architectural priors. Relief maps, not present in the two aforementioned methods, serve both to preserve the information necessary for accurate rendering and to decouple the numerical errors inherently present in the stereo reconstruction process from the recovery of structure.

Our approach is based on a SaM pipeline that will be outlined in the next section.

2. Reconstruction pipeline

Given a collection of uncalibrated images of the same scene, with constant intrinsic parameters, we aim to recover camera parameters, pose estimates and a sparse 3D points cloud of the scene. Our SaM pipeline is basically an incremental greedy approach, similar to [SSS06], made up of algorithms already known in Computer Vision. The most efforts have been made in the direction of a robust and automatic approach, avoiding unnecessary parameters tuning.

Multimatching. Initially, keypoints and matches among the views must be obtained. We extract SIFT features [Low04] from each image then, for each pair of images, features are matched by thresholding the ratio of the distance from the first best match to the distance from the second best match. After that, putative matches are pruned by estimating the fundamental matrix with RANSAC followed by outlier rejection with the X84 rejection rule [HRea86] on the geometric error. After that, keypoints matching in multiple images are connected into tracks, rejecting as inconsistent those tracks in which more than one keypoint converges.

Figure 1: *Two of the pictures from the dataset used in our experiments.*

Autocalibration. The next step is autocalibration, in order to recover the intrinsic camera parameters, that we assume unknown but constant in the whole sequence. Using the fundamental matrices calculated during the matching phase, we estimate the matrix of intrinsic parameters with a global approach based on the Huang-Faugeras constraint, using [AMea04].

Initialization. After autocalibration, we recover the position of each view as well as the 3D location of the tracks. The choice of the two views for initialization is quite critical in an incremental approach. We require that the matching points must be well spread in the two images, and that the fundamental matrix must explain the data far better than other models (homography or affine), according to the Geometric Robust Information Criterion (GRIC) [Tor97]. The relative pose between the selected views is then obtained by factorizing the essential matrix, as in [Har92].

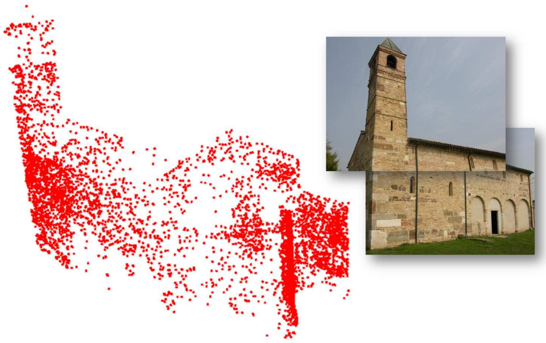

Figure 2: *The cloud of points produced by SaM. There are more than five thousands points.*

Incremental Step Loop. After initialization, we select a new camera at a time, choosing the one that contains the largest number of tracks whose 3D position has already been estimated. The new camera pose is initialized by solving an exterior orientation problem [Fio01], then the solution is refined using bundle adjustment [LA04]. Afterwards, we try to add new tracks. Candidates are those tracks that have been seen in at least one of the already estimated cameras. Outliers are rejected on the basis of the reprojection error, using the X84 rule. Finally we run bundle adjustment again with the new tracks. If bundle adjustment do not converge, then the camera is rejected.

3. High-level primitive fitting and relief map extraction

In this section, we describe the two proposed additional steps that we employ to recover semantic structure along with fine geometry.

High-level primitive fitting. Literature offers several algorithms for model estimation: we selected the approach described in [ZK06] because natively developed for multiple structures.

Given a distribution of points corrupted by outliers, the algorithm generates a set of model hypotheses by repeteadly drawing at random the minimal required number of samples for each desired structure, such as planes, cylinder or spheres. The number of models is estimated analyzing for each data point the peaks in the histogram of the hypotheses residuals. This approach enables data self-organization and requires fewer samples than solutions based on naive RANSAC algorithms. The final number of models is calculated taking the median of all the estimates: for each hypothesis the correct supporting cluster is then identified.

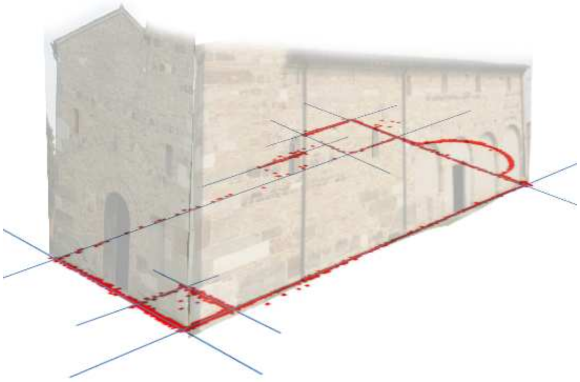

Figure 3: *Automatically recovered perimetral planes from the 3D point cloud.*

Construction of elevation maps. For the recovery of relief maps we developed a simplified version of a recent stereo algorithm based on gestalt principles [YK05]. While based on local methods, it can achieve good performance by employing large disparity neighbourhoods. The problem usually associated with large correlation windows are minimized by weighting the stereo cost function with a measure of similarity and proximity between candidate matches, thus mimicking the behaviour of stereo algorithms based on explicit segmentation. Candidate views for disparity estimation are selected by identifying those that both contain a large set of visible points from the considered surface. The views are first rectified, discarting during the process the couples with excessive distorsions. Conflicts is depth arising from different couples are resolved taking the median of the estimates. Once disparity has been obtained recovering bump, normal and displacement maps is straighforward; these data enables the simulation of fine geometry and the use of modern rendering algorithm such as [KTea01] and its more recent derivations. Since this scene description is a common standard in consumer rendering pipeline, our work can therefore potentially close the gap currently open between acquisition and visualization of urban environments.

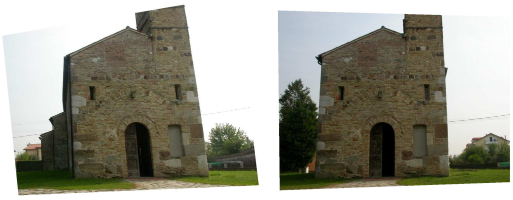

Figure 4: *Rectified images used for the recovery of the front façade.*

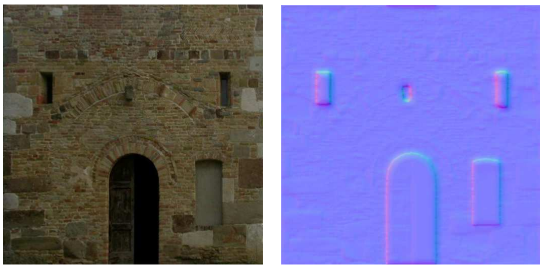

Figure 5: *Color and normal textures automatically generated for the front of the church.*

4. Experiments

Our pipeline was tested with successful results on several architectural models. Here we will present what was obtained processing a dataset of 54 images of a small medieval church. Pictures were acquired with a consumer camera at a resolution of 1024x768 pixels from the ground plane, at different times with automatic exposure (Fig. 1). Photos contain occlusions, scale changes, uneven brightness, sun flares and an outlier that we inserted purposely to verify its rejection.

The church itself has a fairly simple planimetry: the perimeter is composed of straight walls, with a bell tower and a slanted roof covered with bent tiles. A cylindrical apse protrudes from the back; several arches and slit windows open into the well-textured brick walls.

In Fig. 2 the complete point cloud generated from the SaM pipeline (described in Sec. 3) is shown; the good con-

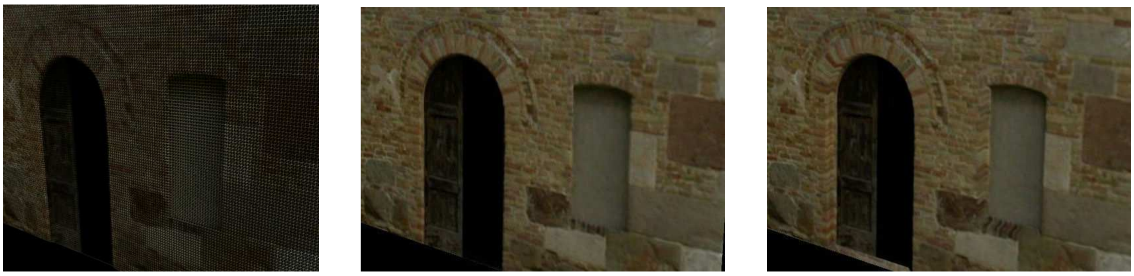

Figure 6: *From left to right, the dense mesh generated from stereo matching, a single quad textured without fine geometry and the same quad with parallax mapping.*

tinuity properties and its remarkable accuracy in modeling the perimetral walls are evident when projecting recovered features onto the ground plane (Fig. 3). The same pictures shows the planar surfaces, all correctly recovered using ten thousand different hypotheses in the primitive fitting.

Figure 4 shows two views after homographical rectification [FTV00] and Fig. 5 the color and normal maps resulting from the relief map extraction. The extracted map encodes both fine geometry and architectural features, modeling wall extrusions as well as windows and arches. A detail of the façade is analyzed in Fig. 6 where we compare side by side the dense regular mesh generated by the matching process with two renderings composed of a single polygon. The effect of parallax mapping, enabled in the last image, are particularly noticeable in correspondance of the door extrusion. With the exception of some minor artifacts, the resulting visualization is remarkably faithful to the original model.

5. Conclusions

We presented a complete SaM pipeline for large architectural scenes capable of automatically reconstructing from a set of sparse pictures a compact representation composed of high-level geometric primitives, textures and relief maps. This format, which conveys the semantic structure of the imaged environment, has obvious advantages when compared with unorganized point clouds or overly dense meshes produced by competing approaches. Based on already common, accepted standards, it has the potential of narrowing the gap between acquisition, editing and visualization of urban scenes.

References

[AMea04] A.Fusiello, M.Farenzena, et al.: Globally convergent autocalibration using interval analysis. *IEEE Trans. on PAMI 26*, 12 (December 2004), 1633–1638.

[BL05] Brown M., Lowe D. G.: Unsupervised 3D object recognition and reconstruction in unordered datasets. In *Int. Conf. 3DIM* (June 2005).

[CCG06] Cornelis N., Cornelis K., Gool L. V.: Fast compact city modeling for navigation pre-visualization. In *Proceedings of CVPR* (2006), vol. 2, pp. 1339–1344.

[Dea04] Dick A. R., et al.: Modelling and interpretation of architecture from several images. *IJCV 60*, 2 (2004), 111–134.

[Fio01] Fiore P. D.: Efficient linear solution of exterior orientation. *IEEE Trans. on PAMI 23*, 2 (2001), 140–148.

[FTV00] Fusiello A., Trucco E., Verri A.: A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications 12*, 1 (2000), 16–22.

[GSC*07] Goesele M., Snavely N., Curless B., Hoppe H., Seitz S. M.: Multi-view stereo for community photo collections. In *Proceedings of ICCV* (October 14-20 2007).

[Har92] Hartley R. I.: Estimation of relative camera position for uncalibrated cameras. In *ECCV* (1992), pp. 579–587.

[HRea86] Hampel F., Rousseeuw P., et al.: *Robust Statistics: the Approach Based on Influence Functions*. Wiley, 1986.

[Kea06] Kamberov G., et al.: 3D geometry from uncalibrated images. In *2nd Intl. Symposium on Visual Computing* (2006), Springer Lecture Notes in Computer Science.

[KTea01] Kaneko T., Takahei T., et al.: Detailed shape representation with parallax mapping. In *Proceedings of ICAT* (2001), pp. 205–208.

[LA04] Lourakis M., Argyros A.: *Design and Implementation of a Generic Sparse BA Software Package Based on the LM Algorithm*. Tech. Rep. 340, FORTH, Aug. 2004.

[Low04] Lowe D.: Distinctive image features from scale-invariant keypoints. *IJCV 60*, 2 (2004), 91–110.

[Mea07] Mordohai P., et al.: Real-time video-based reconstruction of urban environments. In *Workshop 3D-ARCH 2007* (July 12-13 2007).

[SB03] Schindler K., Bauer J.: A model-based method for building reconstruction. In *Proceedings of HLK'03* (Washington, DC, USA, 2003), IEEE Computer Society, p. 74.

[SSS06] Snavely N., Seitz S. M., Szeliski R.: Photo tourism: exploring photo collections in 3D. In *SIGGRAPH'06* (New York, NY, USA, 2006), pp. 835–846.

[Tor97] Torr P. H. S.: An assessment of information criteria for motion model selection. *Proceedings of CVPR* (1997), 47–53.

[YK05] Yoon K.-J., Kweon I.-S.: Locally adaptive support-weight approach for visual correspondence search. *IEEE Conf. CVPR 2* (20-25 June 2005), 924–931 vol. 2.

[ZK06] Zhang W., Kosecká J.: Nonparametric estimation of multiple structures with outliers. In *WDV* (2006), pp. 60–74.