AUTOMATIC STRUCTURE RECOVERY AND VISUALIZATION

Michela Farenzena, Andrea Fusiello, Riccardo Gherardi and Roberto Toldo

Department of Computer Science, University of Verona
Strada Le Grazie 15, 37134 Verona (ITALY)
andrea.fusiello@univr.it

KEY WORDS: Reconstruction, Structure and Motion, Model Extraction, Visualization

ABSTRACT:

We describe an automated pipeline for the reconstruction and rendering of three dimensional objects, with particular emphasys for urban environments. Our system can robustly recover 3D points and cameras from uncalibrated views, without manual assistance. The reconstructed structure is augmented by fitting geometrical primitives such as planes and cylinders to the sparse point cloud obtained. Such information is the key to obtain a higher level understanding of the scene; we use this knowledge to efficiently render the recovered environment, capturing its global appearance while preserving scalability. Several examples display our system in action.

1 INTRODUCTION

In this paper we describe the fully automated approach to object reconstruction from images that we developed over the last five years. Our integrated approach is composed by a geometric back-end and a graphical front-end.

The former is constituted by a Structure and Motion pipeline specifically tailored for robustness, that is able to automatically reconstruct 3D points and cameras from uncalibrated views. The resulting unorganized point cloud is subsequently augmented by fitting its elements with geometrical primitives such as planes and cylinders, gaining a higher level understanding of the scene. In this regard, we developed a specific approach that enables data self-organization and copes naturally with multiple structures.

The latter is a visualizer that enables both researchers and end-users to analyze and navigate the results of the reconstruction process. Users can jump from photo to photo in 3D, thus perceiving the relative position of the images, or they can navigate in the virtual space, roaming freely around the point cloud and image pyramids, similarly to the Virtual Tourism project (Snavely et al., 2006).

Since our model is richer than a sparse unorganized point cloud, thanks to the geometric primitives fitted to 3D data, we can offer a third, novel visualization modality. The original pictures are projected on the recovered surfaces, thus producing a model that captures both the appearance and structure of the scene. This enriches the picture context and increases the user scene awareness.

The final system brings together previous art and novel solutions in an unsupervised framework which needs relatively few assumptions or restrictions.

The rest of this paper is organized as follows: in section two we will review the relevant state of the art; our approach to the problem of uncalibrated reconstruction will be described in the following three section, dealing respectively with the structure and motion pipeline, the high-level model fitting and the visualization stage. Section 6 will present several experimental results; conclusions are drawn in section 7.

2 PREVIOUS ART

Literature covers several approaches for solving the problem of architectural/urban reconstruction: these can be categorized in two main branches: a first one (Snavely et al., 2006, Vergauwen and Gool, 2006, Brown and Lowe, 2005, Kamberov et al., 2006) is composed of 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 ((Goesele et al., 2007) being the notable exception).

The second category comprises the methods specifically tailored for urban environmentsand engineered to be mounted on survey vehicles (P. Mordohai et al., 2007, Cornelis et al., 2006). These systems usually rely on a host of additional information, such as GPS and inertial sensors, and output dense polygonal maps using stereo triangulation.

Both approaches produce large amounts of data, making it difficult to store, render, analyze or disseminate the results. The most scalable approach was shown in (Cornelis et al., 2006), developed for compact visualization on consumer navigation products. Road ground and building façades were forced to lie on textured, mutually-orthogonal, gravity-aligned, geo-located planes.

The recovery of the semantic structure of urban elements, in order to produce simpler and more tractable models, has been tackled by fewer researchers. In this respect, the two most similar articles to the work presented here are (Dick et al., 2004) and (Schindler and Bauer, 2003). In (Dick et al., 2004) is described a system that specializes in creating 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 (Schindler and Bauer, 2003) 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 pre-determined architectural priors.

3 STRUCTURE AND MOTION PIPELINE

Given a collection of uncalibrated images of the same scene, with constant intrinsic parameters, the SaM pipeline outputs camera parameters, pose estimates and a sparse 3D points cloud of the scene. Our SaM pipeline is made up of state-of-the-art algorithms and follows an incremental greedy approach, similar to (Snavely et al., 2006) and (Pollefeys et al., 2002). The most efforts have been made in the direction of a robust and automatic approach, avoiding unnecessary parameters tuning and user intervention. More details are reported in (Farenzena et al., 2008b). A sample output is shown in Fig. 1.

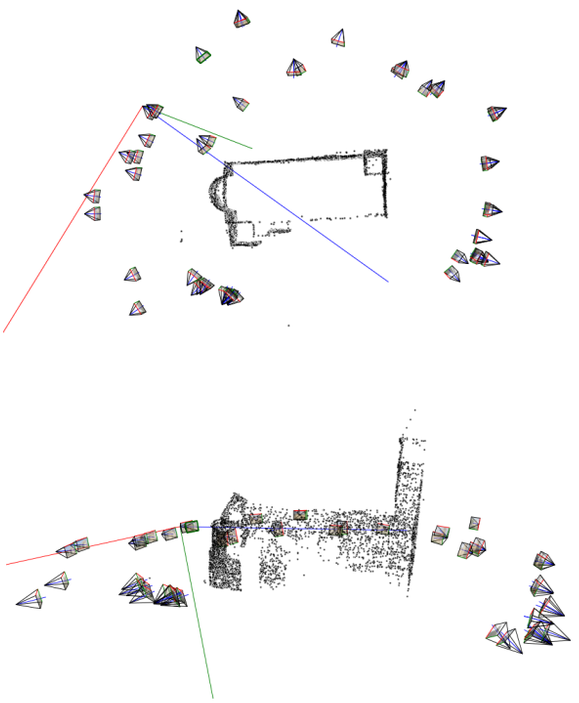

Figure 1: Reconstruction of the "Pozzoveggiani" dataset.

3.1 Multimatching

Initially, keypoints are extracted and matched over different images. This is accomplished using SIFT (Lowe, 2004) for detection and description of local point features. Matching follows a nearest neighbor approach (Lowe, 2004), with rejection of those keypoints for which the ratio of the nearest neighbor distance to the second nearest neighbor distance is greater than 2.0.

Homographies and fundamental matrices between pairs of images are then computed using RANSAC (Fischler and Bolles, 1981). At this point we have a set of matches that are considered inliers for a certain model. However, in order to increase the robustness of the method further, we apply the x84 (Hampel et al., 1986) outlier rejection rule. The best-fit model (homography or fundamental matrix) is selected according to the Geometric Robust Information Criterion (GRIC) (Torr, 1997). The matches from this model go through a validation gate as in (Brown and Lowe, 2003). The idea is to compare the probabilities that this set of inliers/outliers was generated by a correct image match or by a false image match.

After that, keypoints matching in multiple images (at least three) are connected into *tracks*, rejecting as inconsistent those tracks in which more than one keypoint converges (Snavely et al., 2006).

The fundamental matrices so far obtained are fed into a globally convergent autocalibration algorithm (Fusiello et al., 2004) that recovers the intrinsic parameters K of the cameras by minimizing the following cost function using Interval Analysis:

$$\chi(K) = \sum_{i,j} w_{ij} \frac{2\,\mathrm{tr}(E_{ij}E_{ij}^{\mathsf{T}})^2 - \mathrm{tr}^2(E_{ij}E_{ij}^{\mathsf{T}})}{\mathrm{tr}^2(E_{ij}E_{ij}^{\mathsf{T}})} \qquad (1)$$

where F_{ij} is the fundamental matrix between views i and j, and $E_{ij} = K^{\mathsf{T}} F_{ij} K$.

Once the intrinsic parameters are known, the position of each view as well as the 3D location of the tracks is recovered using an incremental approach that starts from two views. The extrinsic parameters of two given views is obtained by factorizing the essential matrix, as in (Hartley, 1992). Then 3D points are reconstructed by intersection and pruned using x84 on the reprojection error. Bundle adjustment (BA) (Lourakis and Argyros, 2004) is run eventually to improve the reconstruction.

The choice of these two seed views turns out to be critical (Thormählen et al., 2004). It should be a compromise between distance of the views and the number of keypoints in common. 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 (namely, homography), according to the GRIC, as in (Pollefeys et al., 2002). Therefore, we compute the following distance measure between views:

$$\mathcal{S}_{i,j} = \frac{CH_i}{A_i} + \frac{CH_j}{A_j} \qquad (2)$$

where CH_i (CH_j) is the area of the convex hull of the keypoint in image I_i (I_j), A_i (A_j) is the total area of image I_i (I_j). Then, among the top 20% closest views we choose the one with the lowest $\mathrm{gric}(F_{i,j})/\mathrm{gric}(H_{i,j})$, where $\mathrm{gric}(F_{i,j})$ and $\mathrm{gric}(H_{i,j})$ are the GRIC scores obtained by the fundamental matrix and the homography matrix respectively.

After initialization, a new view at a time is added until there are no remaining views. The next view to be considered is the one that contains the largest number of tracks whose 3D position has already been estimated. This gives the maximum number of 3D-2D correspondences, that are exploited to solve an exterior orientation problem via a linear algorithm (Fiore, 2001). The algorithm is used inside a RANSAC iteration, in order to cope with outliers. The extrinsic parameters are then refined with BA.

Afterwards, the 3D structure is updated by adding new tracks, if possible. Candidates are those tracks that have been seen in at least one of the cameras in the current reconstruction. 3D points are reconstructed by intersection, and successively pruned using x84 on the reprojection error. As a further caution, 3D points for which the intersection is ill-conditioned are discarded, using a threshold on the condition number of the linear system.

Finally, we run BA again, including the new 3D points.

4 GEOMETRIC PRIMITIVE EXTRACTION

A widespread problem in Computer Vision is fitting a model to noisy data: The RANSAC algorithm (Fischler and Bolles, 1981) is the common practice for that task. It works reliably when data contains measurements from a single structure corrupted by gross outliers. However it has been proved to be not suited to deal with multiple structures (Zuliani et al., 2005). Mode finding in parameter space (Xu et al., 1990, Comaniciu and Meer, 2002), on the other hand, copes naturally with multiple structures, but cannot deal with high percentage of gross outliers, especially as the number of models grows and the distribution of inliers per model is uneven (Zhang and Kosecká, 2006). We developed a

new method for primitives extraction that overcome these problems (Toldo and Fusiello, 2008).

The method starts with random sampling: M Model hypothesis are generated by drawing minimal sets of data points necessary to estimate the model, called minimal sample sets (MSS). They are constructed in a way that neighbouring points are selected with higher probability, as suggested in (Kanazawa and Kawakami, 2004, Zuliani et al., 2005). Namely, if a point $\mathbf{x}_i$ has already been selected, then $\mathbf{x}_j$ has the following probability of being drawn:

$$P(\mathbf{x}_j|\mathbf{x}_i) = \begin{cases} \frac{1}{Z} \exp -\frac{||\mathbf{x}_j - \mathbf{x}_i||^2}{\sigma^2} & \text{if } \mathbf{x}_j \neq \mathbf{x}_i, \\ 0 & \text{if } \mathbf{x}_j = \mathbf{x}_i. \end{cases} \quad (3)$$

where Z is a normalization constant and σ is chosen heuristically.

The consensus set (CS) of each model, i.e. the points with a distance less than a threshold ε from the model, is then computed. We can virtually build a $N \times M$ matrix, where each column is the characteristic function of the CS of a model hypothesis. Each row of this matrix indicates which models a points has given consensus to, i.e., which models it prefers. We call this the *preference set* (PS) of a point. According to (Duin et al., 2004), this is a *conceptual representation* of a point.

Points belonging to the same structure will have similar conceptual representations, in other words, they will cluster in the *conceptual space* $\{0,1\}^M$. This is a consequence of the fact that models generated with random sampling cluster in the hypothesis space around the true models.

Therefore, models are extracted by agglomerative clustering of data points in the conceptual space. The general agglomerative clustering algorithm proceeds in a bottom-up manner: Starting from all singletons, each sweep of the algorithm merges the two clusters with the smallest distance. The way the distance between clusters is computed produces different flavours of the algorithm, namely the simple linkage, complete linkage and average linkage (Duda and Hart, 1973).

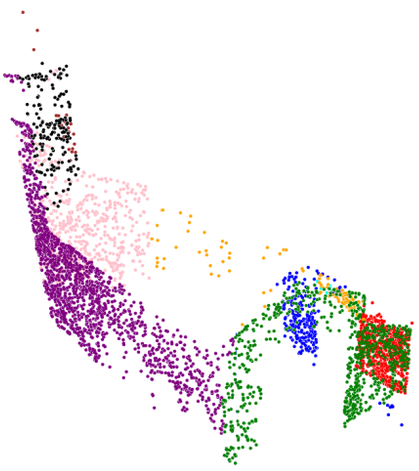

Figure 2: Planes recovered by J-Linkage. The colour of the point encodes the plane it belongs to.

We propose a variation that fits very well to our problem, called *J-linkage* (see Algorithm 1). First the preference set of a cluster is computed as the *intersection* of the preference sets of its points. Then the distance between two elements (point or cluster) is computed as the *Jaccard distance* between the respective preference sets: Given two sets A and B, the Jaccard distance is

$$d_J(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}.$$

The Jaccard distance measures the degree of overlap of the two sets and ranges from 0 (identical sets) to 1 (disjoint sets).

The cut-off value is set to 1, which means that the algorithm will only link together elements whose preference sets overlap. Please note that that the cut-off distance is not data dependent, but defines a qualitative behaviour of the J-linkage algorithm. Indeed, as a result, clusters of points have the following properties:

- for each cluster there exist at least one models that is in the PS of all the points (i.e., a model that fits all the points of the cluster)
- one model cannot be in the PS of *all* the points of two distinct clusters (otherwise they would have been linked).

Algorithm 1 J-LINKAGE

Input: the set of data points, each point represented by its preference set (PS)
Output: clusters of points belonging to the same model

1. Put each point in its own cluster.
2. Define the cluster's PS as the *intersection* of the PSs of its points.
3. Among all current clusters, pick the two clusters with the smallest Jaccard distance between the respective PSs.
4. Replace these two clusters with the union of the two original ones.
5. Repeat from step 3 while the smallest Jaccard distance is lower than 1.

Each cluster of points defines (at least) one model. If more models fit all the points of a cluster they must be very similar. The final model for each cluster of points is estimated by least squares fitting.

Outliers emerge as small clusters. In our pipeline we set rejection threshold equals to the MSS, since we don't have to deal with a large amount of gross outliers. If different kind of geometric primitives are present, a model selection step is employed (Farenzena et al., 2008b).

Fig. 2 shows the result of fitting planes to the 3D reconstruction of Pozzoveggiani.

5 STRUCTURE VISUALIZATION

In this section we show how all the data gathered in the preceding steps of the pipeline can be used to visualize and analyze the imaged environment in a efficient and compelling way.

The two main goals of the visualization phase can be described in terms of scalability and fidelity of the obtained rendering. Usually a compromise has to be reached when balancing the computational efficiency and the accuracy of the reproduction: we claim however that the knowledge of the high-level structure of the scene can reconcile elegantly these two competing goals.

Before describing our solution, it is useful to briefly review the diverse approaches to this problem: at the lowest end of the spectrum, structure can be visualized just as a collection of (possibly coloured) points in space, as shown in Fig. 1; if surfaces have been extracted, another alternative consist in connecting the

aforementioned cloud in a dense triangulated mesh. Of these solutions, neither is arbitrarily scalable, and only the second is capable of recapturing faithfully the original shape of the environment.

A different philosophy is employed in the Photosynth software (http://photosynth.net): only a single picture is ever shown at full resolution from a vantage point; other pictures that can be related to the reference one by a homography are used to provide the context for the user to understand and navigate the collection. An example of this approach produced within our system is shown in Fig 3.

Figure 3: A planar stitch of pictures in 3D.

Such a representation has a number of interesting properties: it supports spatial navigation and provides excellent visual fidelity, since the reference picture is always seen from the position from which it was shoot, and augmented with a relevant context. On the other hand however, it is structurally unable to capture image relashionships that can't be modeled with a collinearity.

These problems can be overcomed by using the high level models recovered during our reconstruction pipeline as a proxy for the scene geometry, and rendering the photo collection against them. To be scalable and effective however, this approach must be coupled with a way to select from a arbitrary position in 3D the subset of the available views which maximizes the visual fidelity while containing the computational workload.

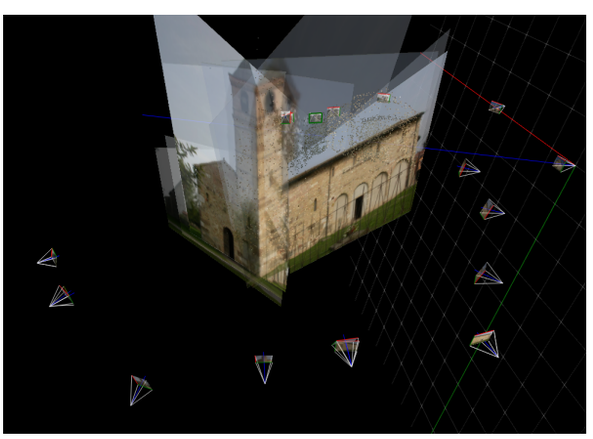

Figure 4: Unmasked rendering on the recovered primitives.

5.1 View-model affinity

We first consider the problem of selecting, given a reference view, a number of additional views that will provide the context for the reference one. This can be done in several ways: one possible solution is to simply select the nearest neighbours induced by a distance function on the camera parameters, like the following one:

$$d(G_{ref}, G) = \min(\log(||G_{ref}^{-1} G||), \log(||G G_{ref}^{-1}||))$$

where G are the extrinsic parameters of the considered camera. This metric usually gives good result when the scale and the intrinsic parameters of the cameras are roughly the same.

In the general case however, selecting views that contain a large number of common 3D features has shown itself a much more stable heuristic, capable of automatically coping with scale changes and camera tilt. Such characteristics are important for selecting a range of images with sufficient variability. The same criteria can be used also to evaluate the affinity between a collection of high level primitives and a view.

When realizing that a arbitrary position and direction in space specified by a virtual camera is akin to a regular view, it becomes possible to select both the models and cameras that have affinity with a arbitrary point in space.

With these data, each selected view can then be rendered using projective texture mapping on the proxy geometry that the high level primitives constitute. If needed, the fine details lost in the primitive extraction can be encoded in displacement or relief maps, as suggested in (Farenzena et al., 2008a).

5.2 Mask creation

The process described in the previous section however is not sufficient to guarantee an artifact-free rendering, as Fig. 4 clearly shows. These effect can be avoded masking the projection over each recovered primitive.

Figure 5: Points on two different planes and their recovered masks.

The problem can be solved creating the mask for each primitive back-projecting its points onto the image plane, and extracting a 2D neighbourhood of the obtained points. We found that just thresholding a low pass filtered binary image containing the point projections gave reasonable results.

Figure 5 shows the masks obtained from two planar surfaces: as it can be seen the recovered mask follow quite closely the underlying structure. This approach works well when the three dimensional features are evenly distributed: in that case, we obtain surfaces without connectivity problems.

The potential issues with color bleeding on the boundaries between primitives could be further corrected by constraining mask borders to align with the models intersections. In our experience however, the perceived effect of bleeding was unnoticeable.

6 EXPERIMENTAL RESULTS

Our pipeline was tested on several datasets of images. All pictures were taken in uncontrolled settings and environment, and no additional information was used.

Pozzoveggiani. Pozzoveggiani is a small church near Padua (IT) that had been used before in (Guarnieri et al., 2004) to test photogrammetric reconstruction. It has a simple planimetry: the perimeter is composed of straight walls, with a bell tower and a slanted roof covered with bent tiles (see Fig. 7). A cylindrical apse protrudes from the back; several arches and slit windows open into the well-textured brick walls.

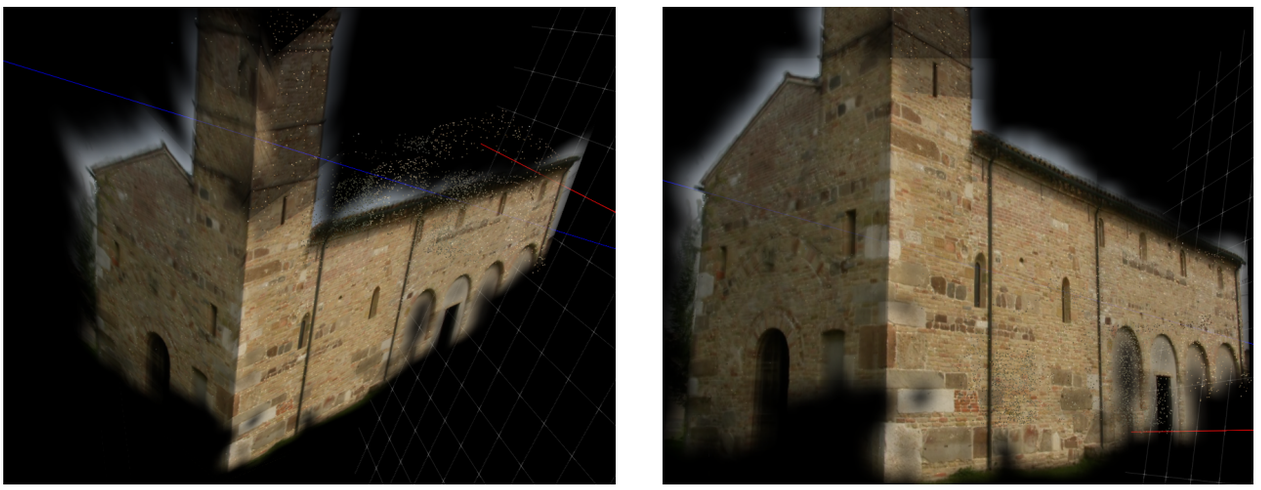

Figure 6: Two pictures of Pozzoveggiani from our interactive visualizer.

The picture set is composed of 54 images acquired from the ground plane with a consumer camera at a resolution of 1024x768 pixels, at different times and with automatic exposure. This is the dataset that was chosen to illustrate the various step of the algorithm through this paper: as was shown, our pipeline succeeds in recovering and modeling all the perimetral walls. The good properties of the reconstruction can also be assessed by measuring the average angle between orthogonal planes, which is 90.44 degrees.

Figure 7: Two views of the Pozzoveggiani church.

Two frames from our interactive are shown in Fig. 6; as it can be seen, it correctly models the two surfaces visible from the current view, while discarting the background. The missing parts from the side textures are due to a uneven distribution of the 3D features on the walls; when seen in movement, the model faithfully captures the expected appearance of the scene, guiding the user in the exploration.

Valbonne. The church of Valbonne is another small church located in France, and extensively used in the computer vision literature. This experiment comprises fifteen photos: the dataset is recorded at a resolution of 768x512 pixels, in varying condition of illumination and occlusion. Again – as shown in Fig. 8 – our system successfully recovers all dominant planes and cameras, with the front façade assimilating the contributes of the two protrusions at its sides.

7 CONCLUSIONS

We have described a completely automated structure and motion pipeline for the reconstruction and rendering of architectural and urban models. Initially tailored for robustness, our method is able to make use of the peculiar characteristics of these environments by augmenting the initial reconstruction with high level geometrical primitives which not only provide a better understanding of the scene, but have the possibility of efficiently supporting further processing on the data. This is demonstrated by employing them to faithfully render the acquired environment in a scalable way.

REFERENCES

Brown, M. and Lowe, D., 2003. Recognising panoramas. In: Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, pp. 1218–1225 vol.2.

Brown, M. and Lowe, D. G., 2005. Unsupervised 3D object recognition and reconstruction in unordered datasets. In: Proceedings of the International Conference on 3D Digital Imaging and Modeling, Ottawa, Ontario, Canada.

Comaniciu, D. and Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5), pp. 603–619.

Cornelis, N., Cornelis, K. and Gool, L. V., 2006. Fast compact city modeling for navigation pre-visualization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2, New York, NY, USA, pp. 1339–1344.

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

Duda, R. O. and Hart, P. E., 1973. Pattern Classification and Scene Analysis. John Wiley and Sons, pp. 98–105.

Duin, R., Pekalska, E., Paclik, P. and Tax, D., 2004. The dissimilarity representation, a basis for domain based pattern recognition? In: Pattern representation and the future of pattern recognition, ICPR 2004 Workshop Proceedings, Cambridge, UK, pp. 43–56.

Farenzena, M., Fusiello, A. and Gherardi, R., 2008a. Efficient visualization of architectural models from a structure and motion pipeline. In: Proceedings of Eurographics - short papers, Crete, Greece.

Farenzena, M., Fusiello, A., Gherardi, R. and Toldo, R., 2008b. Towards unsupervised reconstruction of architectural models. In: Proceedings of Vision, Modeling, and Visualization 2008, Konstanz, Germany, pp. 41–50.

Fiore, P. D., 2001. Efficient linear solution of exterior orientation. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(2), pp. 140–148.

Fischler, M. A. and Bolles, R. C., 1981. Random Sample Consensus: a paradigm model fitting with applications to image analysis and automated cartography. Communications of the ACM 24(6), pp. 381–395.

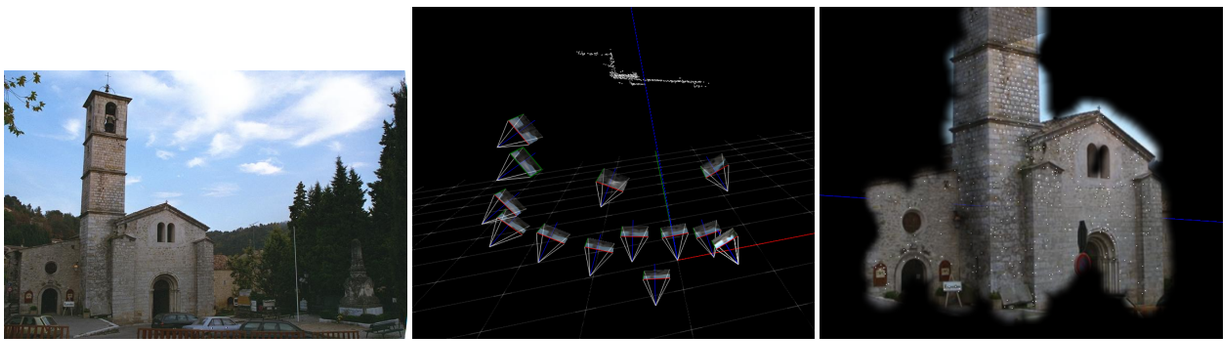

Figure 8: Results for the Valbonne dataset.

Fusiello, A., Benedetti, A., Farenzena, M. and Busti, A., 2004. Globally convergent autocalibration using interval analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(12), pp. 1633–1638.

Goesele, M., Snavely, N., Curless, B., Hoppe, H. and Seitz, S. M., 2007. Multi-view stereo for community photo collections. In: Proceedings of the International Conference on Computer Vision, Rio de Janeiro, Brazil.

Guarnieri, A., Vettore, A. and Remondino, F., 2004. Photogrammetry and ground-based laser scanning: Assessment of metric accuracy of the 3D model of pozzoveggiani church. In: FIG Working Week. TS on "Positioning and Measurement Technologies and Practices II - Laser Scanning and Photogrammetry", Athens, Greece.

Hampel, F., Rousseeuw, P., Ronchetti, E. and Stahel, W., 1986. Robust Statistics: the Approach Based on Influence Functions. Wiley Series in probability and mathematical statistics, John Wiley & Sons.

Hartley, R. I., 1992. Estimation of relative camera position for uncalibrated cameras. In: Proceedings of the European Conference on Computer Vision, Santa Margherita L., pp. 579–587.

Kamberov, G., Kamberova, G., Chum, O., Obdrzalek, S., Martinec, D., Kostkova, J., Pajdla, T., Matas, J. and Sara, R., 2006. 3D geometry from uncalibrated images. In: Proceedings of the 2nd International Symposium on Visual Computing, Springer Lecture Notes in Computer Science, Lake Tahoe, Nevada, USA.

Kanazawa, Y. and Kawakami, H., 2004. Detection of planar regions with uncalibrated stereo using distributions of feature points. In: Proceedings of the British Machine Vision Conference, pp. 247 – 256.

Lourakis, M. and Argyros, A., 2004. The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm. Technical Report 340, Institute of Computer Science - FORTH, Heraklion, Crete, Greece.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), pp. 91–110.

P. Mordohai et al., 2007. Real-time video-based reconstruction of urban environments. In: 3D-ARCH 2007: 3D Virtual Reconstruction and Visualization of Complex Architectures, Zurich, Switzerland.

Pollefeys, M., Gool, L., Vergauwen, M., Cornelis, K., Verbiest, F. and Tops, J., 2002. Video-to-3d. In: Proceedings of Photogrammetric Computer Vision 2002, Int. Archive of Photogrammetry and Remote Sensing., Graz, Austria, p. 252 258.

Schindler, K. and Bauer, J., 2003. A model-based method for building reconstruction. In: Proceedings of the First IEEE International Workshop on Higher-Level Knowledge in 3D Modeling And Motion Analysis, Washington, DC, USA, p. 74.

Snavely, N., Seitz, S. M. and Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. In: SIGGRAPH Conference Proceedings, New York, NY, USA, pp. 835–846.

Thormählen, T., Broszio, H. and Weissenfeld, A., 2004. Keyframe selection for camera motion and structure estimation from multiple views. In: Proceedings of the European Conference on Computer Vision, Lecture Notes in Computer Science, Vol. 3021, pp. 523–535.

Toldo, R. and Fusiello, A., 2008. Robust multiple structures estimation with j-linkage. In: Proceedings of the European Conference of Computer Vision, Vol. 1, Marseille, France, pp. 537–547.

Torr, P. H. S., 1997. An assessment of information criteria for motion model selection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp. 47–53.

Vergauwen, M. and Gool, L. V., 2006. Web-based 3D reconstruction service. Machine Vision and Applications 17(6), pp. 411–426.

Xu, L., Oja, E. and Kultanen, P., 1990. A new curve detection method: randomized Hough transform (RHT). Pattern Recognition Letters 11(5), pp. 331–338.

Zhang, W. and Kosecká, J., 2006. Nonparametric estimation of multiple structures with outliers. In: Workshop on Dynamic Vision, European Conference on Computer Vision 2006, Lecture Notes in Computer Science, Vol. 4358, pp. 60–74.

Zuliani, M., Kenney, C. S. and Manjunath, B. S., 2005. The multiRANSAC algorithm and its application to detect planar homographies. In: Proceedings of the IEEE International Conference on Image Processing, Genova, IT.