CENTER FOR
MACHINE PERCEPTION

CZECH TECHNICAL
UNIVERSITY

REPRINT

Automatic Factorization-Based Reconstruction from Perspective Images with Occlusions and Outliers

Daniel Martinec and Tomáš Pajdla

{martid1, pajdla}@cmp.felk.cvut.cz

Available at
ftp://cmp.felk.cvut.cz/pub/cmp/articles/martinec/Martinec-CVWW2003.pdf

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Prague 6, Czech Republic
fax +420 2 2435 7385, phone +420 2 2435 7637, www: http://cmp.felk.cvut.cz

Automatic Factorization-Based Reconstruction from Perspective Images with Occlusions and Outliers

Daniel Martinec and Tomáš Pajdla

Center for Machine Perception
Department of Cybernetics
Czech Technical University in Prague
Karlovo nám. 13, 121 35 Praha, Czech Republic
{martid1, pajdla}@cmp.felk.cvut.cz

Abstract *This paper presents a method for automatic 3D-reconstruction from a set of plain images. The main contribution of this paper is a robust integration of the partial correspondences from image pairs provided by an existing correspondence estimator into a reconstruction consistent with all images of the scene. Projective shape and motion is estimated by factorization of a matrix containing the images of all scene points. Outlier detection is based on* RANSAC *paradigm and the trifocal tensors. Compared to previous methods, this method can handle perspective views, occlusions, and outliers in image correspondences jointly. The main novelty of this paper is the fusion of a correspondence estimator [8] for wide base-line stereo and the method for outlier detection [6]. It appears that the method is able to detect outliers which cannot be detected using the epipolar geometry only and therefore it is suitable for integration with WBS from image pairs. The new method is demonstrated by experiments with laboratory and outdoor image sets and some results on metric reconstruction are shown.*

1 Introduction

RANSAC on correspondences from image pairs usually exploits the epipolar geometry constraint which says that an image point must lie on the corresponding epipolar line. However, there still may be wrong correspondences left among those satisfying epipolar geometry. Such wrong correspondences, called *outliers*, cannot be detected using the epipolar geometry. Moreover, it can happen in some degenerate cases, e.g. when the overlapping area of the two images is almost planar, that the estimated epipolar geometry is wrong at all and there is no way to detect it using the epipolar geometry. Because of these two reasons, geometric constraints from more than two views have to be used. Our motivation in this paper was to use multiview constraints for detecting outliers in such cases when the two-view geometry fails at all, see Fig. 1.

Tomasi & Kanade [10] developed a factorization method of the measurement matrix for scene reconstruction with an orthographic camera. Their method as well as Jacobs' method [5] can handle occlusions. Sturm and Triggs [9] extended this method from affine to perspective projections

(a) Outliers satisfying epipolar geometry in an image pair

(b) Correctly matched correspondences using the multiview geometry

Figure 1: Corresponding points satisfying the epipolar geometry may be outliers (a). These can be detected using the multiview geometry (b)

but without occlusions. Martinec & Pajdla [7] solved reconstruction with both perspective projections and occlusions. Heyden [4] presented a reconstruction method from affine images with outliers but occlusions are not handled. Recently he extended the method into the perspective case [3]. We presented a method [6] for outlier detection so that reconstruction from perspective images is solved when occlusions and outliers are present jointly. The method is independent of image ordering and treats all data uniformly. No six-tuple of points seen in all images is needed.

The contribution of this paper is a robust integration of the partial correspondences from image pairs provided by correspondence estimator [8] into a more consistent reconstruction from all images of the scene. An automatic method for 3D-reconstruction from plain images is achieved as the combination of the method [6] for outlier detection by factorization applied on the automatic correspondence estimator [8] and followed by projective reconstruction [7].

After problem formulation, the way of obtaining correspondences will be explained in Section 3. Outlier detection and projective reconstruction will be briefly sketched in Sections 4 and 5, respectively. Experiments and conclusion come in Sections 6 and 7.

2 Problem Formulation

Suppose a set of n 3D points is observed by m perspective cameras. Not all points are visible in all views. There may be *outliers*, i.e. mismatches in correspondences. The goal is to reject outliers and to recover 3D structure (point locations) and motion (camera locations) from the remaining image measurements that are called *inliers*.

Let $\mathbf{X}_p$ be the unknown homogeneous coordinate vectors of 3D points, $\mathtt{P}^i$ the unknown 3×4 projection matrices, and $\mathbf{x}_p^i$ the measured homogeneous coordinate vectors of image points, where $i = 1, .., m$ labels images and $p = 1, .., n$ labels points. Due to occlusions, $\mathbf{x}_p^i$ are unknown for some i and p.

The basic image projection equation says that $\mathbf{x}_p^i$ are the projections of $\mathbf{X}_p$ up to unknown scale factors λ_p^i, which will be called (*projective*) *depths*:

$$\lambda_p^i \mathbf{x}_p^i = \mathtt{P}^i \mathbf{X}_p$$

The complete set of image projections can be gathered into a matrix equation:

$$\begin{bmatrix} \lambda_1^1\mathbf{x}_1^1 & \lambda_2^1\mathbf{x}_2^1 & \dots & \lambda_n^1\mathbf{x}_n^1 \\ \times & \lambda_2^2\mathbf{x}_2^2 & \dots & \times \\ \vdots & & \ddots & \vdots \\ \lambda_1^m\mathbf{x}_1^m & \times & \dots & \lambda_n^m\mathbf{x}_n^m \end{bmatrix} = \underbrace{\begin{bmatrix} \mathtt{P}^1 \\ \vdots \\ \mathtt{P}^m \end{bmatrix}}_{\mathtt{P}} \underbrace{[\mathbf{X}_1 \dots \mathbf{X}_n]}_{\mathtt{X}}$$

where marks $\times$ stand for unknown elements which could not be measured due to occlusions, $\mathtt{X}$ and $\mathtt{P}$ stand for structure and motion, respectively. The $3m \times n$ matrix $[\mathbf{x}_p^i]_{i=1..m,p=1..n}$ will be called the *measurement matrix*, shortly MM. MM may have (and in most cases does have) some missing elements and outliers.

3 Obtaining Correspondences

Correspondences across all images were established from matches between pairs of images obtained using the method by Matas et al. [8] exploiting a special kind of distinguished regions, so called extremal regions. Matches obtained using this method satisfy epipolar constraint but are not guaranteed to be the true correspondences. Therefore, an outlier detection technique may still be needed to reject remaining outliers that can only be found using more than two images.

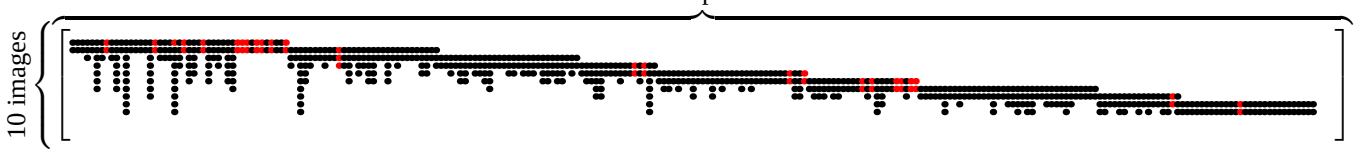
The transitivity assumption was adopted in the process of building the measurement matrix. Whenever there are pair-wise correspondences between images 1–2 and 2–3, it can be expected that there is a pair-wise correspondence between images 1–3. The assumption holds for true correspondences, however, this is not always the case using method [8]. If some pair-wise matches are conflicting, i.e. the found match between images 1–3 differs from the match expected because of matches between images 1–2 and 2–3, it is not clear which pair-wise match is true and which is wrong. The task could be stated so that a maximal set of not conflicting pair-wise correspondences could be searched for. Such a problem leads to a greedy algorithm which is highly non-linear. We did not try to solve it in an optimal manner. Thanks to the possibility of using the outlier detection method [6], conflicts among pair-wise correspondences could be simply ignored and false matches revealed in the subsequent outlier detection stage. Image pairs were read in a random order. The matches between image pairs were placed into the measurement matrix so that the overlapping matches were joined and the conflicting ones were ignored.

4 Outlier Detection

Method [6] was used for outlier detection. The method assumes that the amount of inliers is significantly larger than the amount of outliers. This is true thanks to matcher [8] that discards all matches that do not satisfy consistent epipolar geometries between image pairs. The main idea is that minimal configurations of points in triples of images are sufficient to validate inliers reliably. The RANSAC paradigm is used. Trifocal tensors are computed from randomly selected minimal 6-tuples of points in triples of images. After the tensor estimation, the number of points consistent with the tensor is counted. If there are sufficiently enough consistent points, those not used to estimate the trifocal tensor receive one positive vote. The voting is repeated until points in the measurement matrix are sufficiently sampled. The points that obtain zero or a very small number of votes are rejected as outliers. Inliers are used by the method described in [7] to obtain a projective reconstruction. The set of inliers can be further enlarged by an iterative process.

Method [6] can inspect correspondences among at least three images because the trifocal tensors are used. However, there may be some right correspondences between two images only. Each image point marked by method [6] as outlier is finally verified by sampling pair-wise correspondences in the corresponding column of MM. Each such sample is checked by reconstructing the 3D point using the known cameras. If the reprojection errors in both images are small,

Scene *Valbonne sequence (Oxford)*	14 images [768×512]
Correspondence number / missing data	297 / 80.74 %
λ_p^i estimation	*sequence* (82.52 % of λ_p^i known)
Outliers	**50** (6.24 % out of 801 image points)
Reconstructed / not-reconstructed cameras	10 / 4
Reconstructed / partially rec. / not-rec. corresp.	233 / 24 / 64 out of 297
Mean / maximal reprojection error	**0.36** / 1.82 [pxl] (from inliers only)



Scene *Valbonne all pairs (Oxford)*	14 images [768×512]
Correspondence number / missing data	376 / 73.27 %
λ_p^i estimation	*sequence* (51.53 % of λ_p^i known)
Outliers	**396** (28.14 % out of 1407 image points)
Reconstructed / not-reconstructed cameras	14 / 0
Reconstructed / partially rec. / not-rec. corresp.	271 / 32 / 105 out of 376
Mean / maximal reprojection error	**0.45** / 3.66 [pxl] (from inliers only)

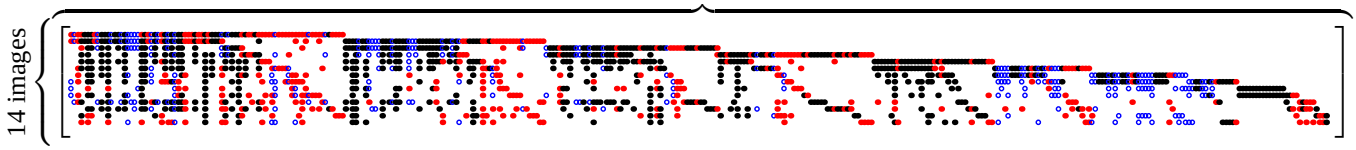


Figure 2: *The Valbonne scene reconstructed using sequence only (top) and using 'all image pairs' (bottom).* Note significantly larger amount of the reconstructed data for all image pairs: number of cameras as well as points is higher

the correspondence (and the image point) is validated.

5 Projective Reconstruction

Method [7] was used for projective reconstruction because it can deal with occlusions in various configurations like in the data from wide base-line multiple view stereo. It is a method for recovery of projective shape and motion from multiple images by factorization of a matrix containing the images of all scene points (MM). Compared to previous methods, this method can handle perspective views and occlusions jointly. The projective depths of image points are estimated by the method of Sturm & Triggs [9] using epipolar geometry. Occlusions are solved by the extension of the method by Jacobs [5] for filling of missing data. This extension can exploit the geometry of the perspective camera so that both points with known and unknown projective depths are used. Many ways of combining the two methods exist, and the one with the best results was presented in [7].

6 Experiments

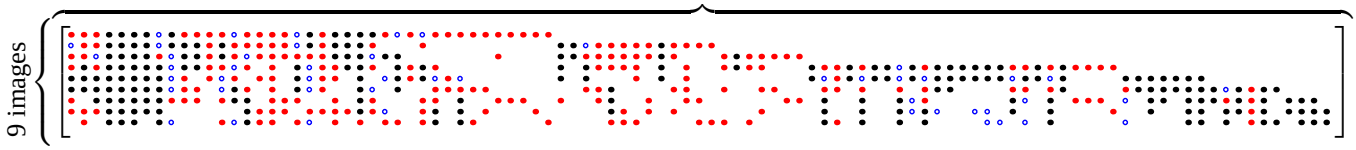
In all experiments, pair-wise matches were found using method [8]. The accuracy for outlier detection using method [6] was set to 5 pixels. For each experiment, one image, an error table, and the structure of MM are provided. The table includes the scene name, number of images and their sizes, number of found correspondences, amount of the missing data, the chosen strategy for depth estimation (see [7]), amount of detected outliers, reconstructed and not-reconstructed cameras, reconstructed, partially reconstructed, and not-reconstructed correspondences, and the reprojection errors of the reconstruction without outliers. In structure of MM, "●" stands for outliers, "●" for image points with depths estimated using the fundamental matrices while "○" stands for image points with depths estimated in the subsequent process of filling MM, and " " stand for the missing data.

The Valbonne scene was reconstructed in two different ways, see Fig. 2. First, $m-1$ image pairs were taken in a sequence and second, all $\binom{m}{2}$ image pairs were used. As could be expected, the latter produced significantly larger amount of the reconstructed data: number of cameras as well as points is higher. Note that the sequence could be reconstructed using some sequential technique, e.g. [1], on the other hand, the data from all image pairs need to be treated by some robust method with missing data. Our factorization method [6] was well suited for the task. Moreover, in the Valbonne data in sequence, the number of correspondences among the four last images was too low to form a trifocal constraint; that is why these images could not be reconstructed by any sequential algorithm using the trifocal constraint (like [1]) at all.

Results of reconstructions of the Movi-house, the Shelf and the House scene are summarized in Fig. 3. Maximal errors in the Movi-house and the House scene are higher than 5 pixels set in the outlier detection accuracy. It is because of

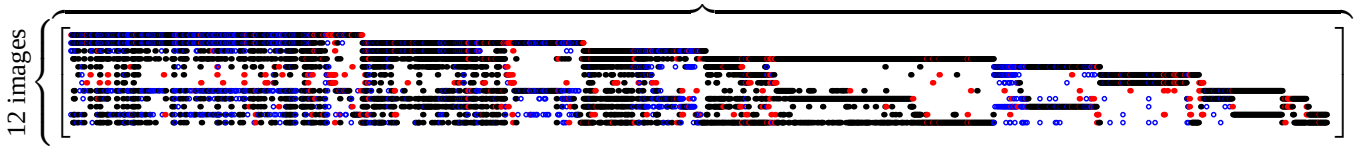
Scene *Movi-house (CMP)*	14 images [512×512]
Correspondence number / missing data	101 / 49.28 %
λ_p^i estimation	*sequence* (44.47 % of λ_p^i known)
Outliers	**207** (44.90 % out of 461 image points)
Reconstructed / not-reconstructed cameras	9 / 0
Reconstructed / partially rec. / not-rec. corresp.	67 / 33 / 34 out of 101
Mean / maximal reprojection error	**0.75** / 5.27 [pxl] (from inliers only)

101 correspondences



Scene *Shelf (CMP)*	14 images [1200×1600]
Correspondence number / missing data	1953 / 72.64 %
λ_p^i estimation	*central image No. 4* (75.00 % of λ_p^i known)
Outliers	**414** (6.46 % out of 6411 image points)
Reconstructed / not-reconstructed cameras	12 / 0
Reconstructed / partially rec. / not-rec. corresp.	1839 / 72 / 114 out of 1953
Mean / maximal reprojection error	**0.51** / 4.90 [pxl] (from inliers only)

1953 correspondences



Scene *House (CMP)*	14 images [640×800]
Correspondence number / missing data	1073 / 73.64 %
λ_p^i estimation	*central image No. 1* (50.26 % of λ_p^i known)
Outliers	**1412** (38.40 % out of 3677 image points)
Reconstructed / not-reconstructed cameras	13 / 0
Reconstructed / partially rec. / not-rec. corresp.	787 / 175 / 286 out of 1073
Mean / maximal reprojection error	**0.51** / 8.31 [pxl] (from inliers only)

1073 correspondences

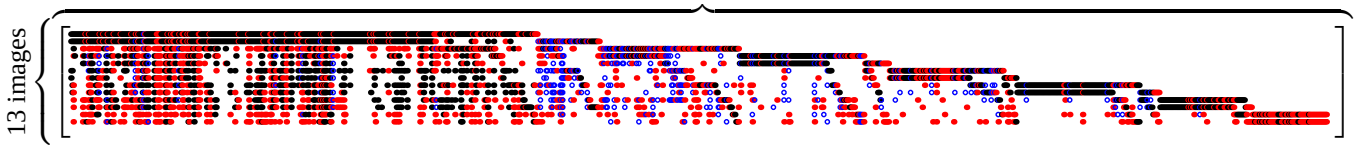


Figure 3: *Scenes reconstructed using all image pairs*

the following reason. The outlier detection accuracy is used for validation of the sampled three- and two-tuples of image points in columns of MM. There is no additional checking on the join of the validated k-tuples of image points. That is why the 3D-point reconstructed using more images can have larger reprojection error in some of them (although such 3D-point can be expected to be reconstructed more precisely). For more details, see [6].

6.1 Metric Reconstruction

To demonstrate correctness of the reconstructions, simple approximation of a metric reconstruction was used. The linear method for auto-calibration using the absolute dual quadric [2] was applied. Fig. 4 shows the metric reconstruction of the Valbonne scene. Only this scene is shown because the linear algorithm for auto-calibration produced some reasonable output for this scene only. It is recommended in [2] to use the non-linear bundle adjustment on the output of the linear method.

7 Conclusion

A fully automatic method for 3D-reconstruction from plain images from perspective cameras was developed. It is the combination of the method [6] for outlier detection applied on the automatic correspondence estimator [8] and followed by projective reconstruction [7]. Tests on laboratory and outdoor scenes showed its applicability in wide base-line multiple view stereo while its applicability for sequences was presented in [6].

Acknowledgement

Jiří Matas and Ondřej Chum from the Czech Technical University in Prague provided the image pair matches, Ondřej Chum provided the routine for the trifocal tensor estimation, Andrew Zisserman from the University of Oxford kindly provided the Valbonne data, and Tomáš Werner from the University of Oxford provided the routine for the bundle adjustment.

This research was supported by the Grant Agency of the Czech Republic under projects GACR 102/01/0971, by The Bilateral Czech-Austrian project Aktion 34p24, project CTU 0209313, by Net CEEPUS SK-042, by The FP5 EU under Project BeNoGo IST-2001-39184, by MSMT Kontakt 22-2003-04, and by MSM 212300013.

References

[1] A. W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *Proc. ECCV*, volume I, pages 311–326. Springer-Verlag, June 1998.

[2] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.

[3] A. Heyden. Personal communication, 2002.

[4] D. Q. Huynh and A. Heyden. Outlier detection in video sequences under affine projection. In *Proc. IEEE Conf. on CVPR*, pages 695–701, December 2001.

[5] D. Jacobs. Linear fitting with missing data: Applications to structure from motion and to characterizing intensity images. In *CVPR*, pages 206–212, 1997.

[6] D. Martinec and T. Pajdla. Outlier detection for factorization-based reconstruction from perspective images with occlusions. In *Proceedings of the Photogrammetric Computer Vision*, volume B, pages 161–164, 2002.

[7] D. Martinec and T. Pajdla. Structure from many perspective images with occlusions. In *Proceedings of the European Conference on Computer Vision*, volume II, pages 355–369, Berlin, Germany, May 2002. Springer-Verlag.

[8] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proceedings of the British Machine Vision Conference*, volume 1, pages 384–393, London, UK, September 2002. BMVA.

[9] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *ECCV96(II)*, pages 709–720, 1996.

[10] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. In *IJCV(9), No. 2*, pages 137–154, November 1992.

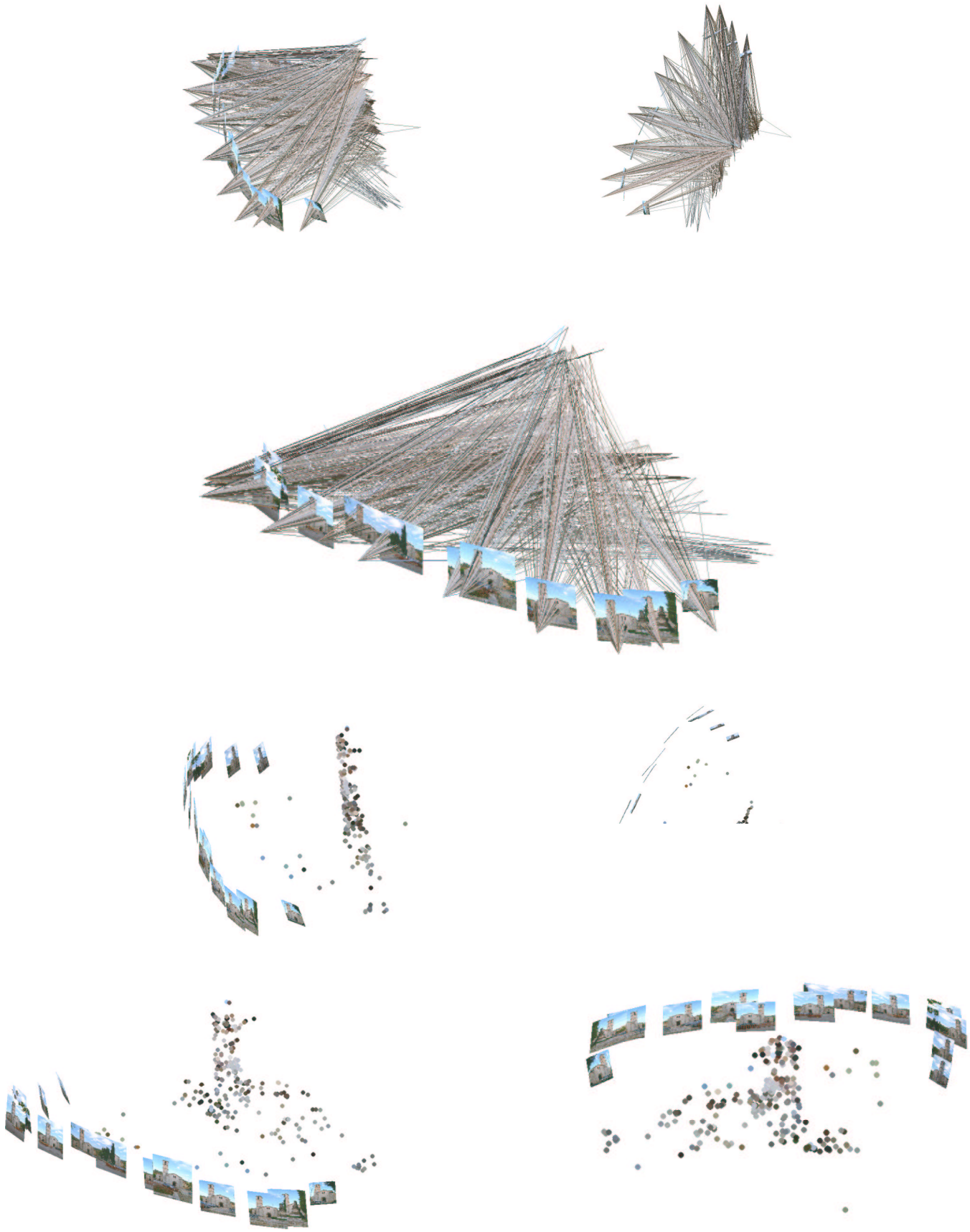


Figure 4: Metric reconstruction of the Valbonne scene