A practical approach for 3D building modeling from uncalibrated video sequences

Yong Liu and H. T. Tsui
Dept. of Electronic Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
yliu@ee.cuhk.edu.hk

ChengKe Wu
Dept. of Information Engineering
Xidian University
Xi'an 710071, China

Abstract

This paper presents an approach for reconstructing a realistic 3D model of a building from its uncalibrated video sequences taken by a hand-held camera. The novelty of this approach lies in the integration of some prior scene knowledge in the auto-calibration stage of the **structure from motion** *(***SFM***) problem. The line parallelism and plane orthogonality are transformed into the constraints on the absolute quadric during camera auto-calibration. This make some critical cases solvable and the reconstruction more Euclidean. The approach is implemented and validated using simulated data and real image data. The experimental results in the end of the paper show the effectiveness of our approach.*

1 Introduction

The structure from motion (**SFM**) problem is a central problem in computer vision. Recently, much attention is focused on solving the **SFM** problem without calibrating the viewing cameras: an uncalibrated structure from motion problem (**USFM**). It is often divided into two stages to solve the USFM problem. First, a reconstruction is obtained in a projective frame by referring to the first camera as the reference camera and determining other cameras under a multiple geometry constraint, cf. [1, 2]. Second, the transformation from the projective frame to the Euclidean frame is solved by the auto-calibration approach, then the Euclidean (scaled) reconstruction can be derived cf. [3, 4, 2]. It is attractive that USFM approaches allow obtaining 3D graphical model of an object from its uncalibrated images, which means that the viewing camera can be a hand-held camera with freely zooming and focusing. The flexibility and automaticity of the approaches meet the demanding from the computer graphics community. Previously, many 3D graphical models were created manually or by using CAD-programs. Recently, in the computer graphics community, it is a tendency to apply these achievements of the computer vision community into some applications. For example, extracting realistic 3D models and using these models in special movie effects, internet selling, or creating a **VR** (Virtual Reality) world etc. As buildings are the most common objects, obtaining their realistic 3D graphical models are even more attractive.

This paper presents a practical approach for reconstructing the 3D graphical model of a building from its uncalibrated video sequences. The novelty of this approach lies in the following aspect. For camera auto-calibration, it is quite often for camera motions to close to a critical motion (straight line or circular motions, for example), resulting ambiguities in auto-calibration and 3D reconstruction . We show the fact that the auto-calibration ambiguities are caused by the ambiguities in determining the absolute quadric. For this problem, we show that the common characteristics of buildings : line parallelism and plane orthogonality, can be transformed into the constraints on the absolute quadric. The classical camera based constraints and the scene based constraints can be integrated to form a uniform constraints on the absolute quadric. In this sense, for the previous critical motion sequences, the ambiguities can be removed by using scene based constraints on the absolute quadric. Even for the non-critical motion sequences, the scene based constraints on the absolute quadric can enforce the camera based constraints and derive more robust results. Furthermore, because the scene based constraints directly come from the scenes, thus using scene based constraints to modify the auto-calibration parameters will give more Euclidean looking of the reconstructed scene for viewing.

The paper is organized as follows: Section 2 introduces the camera auto-calibration using camera based constraints. Section 3 describes the scene based constraints from parallel lines. Section 4 explains the scene based constraints

from orthogonal planes. The experimental results are shown in Section 5. Finally, the conclusions are given in Section 6.

1.1 Notations and background

In a 2D projective space, a point is represented by a homogeneous 3-vector $\mathbf{x}$, a line by 3-vector $\mathbf{l}$, a plane homography by a 3×3 matrix $\mathbf{H}$, a conic is represented by a 3×3 symmetric matrix ω , and its dual conic by ω^*. In a 3D projective space, a point is represented by a homogeneous 4-vector $\mathbf{X}$, a plane is represented by its normal direction $\mathbf{d}$, and a Quadric is represented by a 4×4 symmetric matrix $\mathbf{Q}$. "$\sim$" represents equality up to a scale factor and "$\perp$ " stands for orthogonality. A vector $\mathbf{a}$ is a column vector and $\mathbf{a}^T$ is a row vector. A 3D point $\mathbf{X}$ is projected on a 2D image plane as

$$\mathbf{x} \sim \mathbf{K}\begin{pmatrix}\mathbf{R} & -\mathbf{R}\mathbf{t}\end{pmatrix}\mathbf{X} \ , \tag{1}$$

where $\mathbf{K} = \begin{pmatrix} rf & s & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix}$ is the camera intrinsic matrix: f is the focal length, s is the skew, r is the aspect ratio and (u_0, v_0) is the coordinate of the principle point. $\mathbf{R}$ is a rotation matrix and $\mathbf{t}$ is a translation vector. Both of them are called extrinsic parameters, representing the camera orientation and position.

Definition 1.1. The absolute quadric is defined as a special dual quadric with the form as in [5]

$$\mathbf{Q}_\infty \sim \begin{pmatrix}\mathbf{I}_{3\times 3} & \mathbf{0} \\ \mathbf{0} & 0\end{pmatrix} \ . \tag{2}$$

Being as a dual quadric, its transform law likes that of points. Under a transform $\mathbf{T}$, it becomes

$$\mathbf{Q}' \sim \mathbf{T}\mathbf{Q}\mathbf{T}^T \tag{3}$$

It has three important properties

- ***Property 1***. All Euclidean transforms define a group $\mathcal{G}_E$. For any member of $\mathcal{G}_E$, noted as $\mathbf{T}_e$, the operation $\mathbf{T}_e\mathbf{Q}_\infty\mathbf{T}_e^T$ leaves $\mathbf{Q}_\infty$ invariant. i.e., $\mathbf{T}_e\mathbf{Q}_\infty\mathbf{T}_e^T \sim \mathbf{Q}_\infty$
- ***Property 2***. The absolute quadric encodes the affine structure. The plane at infinity $\pi_\infty = [0001]^T$ is its null vector, $\mathbf{Q}_\infty\pi_\infty = \mathbf{0}$.
- ***Property 3***. The absolute quadric encodes the Euclidean structure. The angle of any two finite planes π_1 , π_2 can be calculated as

$$\cos\theta = \frac{\pi_1^T\mathbf{Q}_\infty\pi_2}{\sqrt{(\pi_1^T\mathbf{Q}_\infty\pi_1)(\pi_2^T\mathbf{Q}_\infty\pi_2)}} \tag{4}$$

When the two planes are orthogonal, $\pi_1^T\mathbf{Q}_\infty\pi_2 = 0$

It is noted that the ***Property 2*** and ***Property 3*** hold not only in a Euclidean frame, but they are also true in a projective frame. Because under an arbitrary transform $\mathbf{T}$, $\mathbf{Q}_\infty \to \mathbf{T}\mathbf{Q}_\infty\mathbf{T}^T$ and a plane $\pi \to \mathbf{T}^{-T}\pi$, it can be derived the ***Property 2*** and ***Property 3*** are still satisfied.

2 Camera based constraints

The present absolute quadric based auto-calibration approaches [5, 2, 4] are mainly derived from the ***Property 1*** of the absolute quadric. In our world, the camera motions are always Euclidean transforms. They leave the absolute quadric invariant, which means the different viewing cameras will capture the absolute quadric independent of their orientations and positions. The images of the absolute quadric depend only on the camera intrinsic matrices. The absolute quadric can be related to the camera intrinsic matrices as:

$$\begin{aligned}\mathbf{P}\mathbf{Q}_\infty\mathbf{P}^T &\sim \mathbf{K}[\mathbf{R}\ \mathbf{t}]\begin{pmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & 0\end{pmatrix}[\mathbf{R}\ \mathbf{t}]^T\mathbf{K}^T \\ &\sim \mathbf{K}\mathbf{K}^T \ .\end{aligned} \tag{5}$$

In a projective space, $\mathbf{P}' \sim \mathbf{P}\mathbf{T}^{-1}$, where $\mathbf{T}$ is a transform from the Euclidean to the projective space. The absolute quadric becomes

$$\mathbf{Q}' \sim \mathbf{T}\mathbf{Q}_\infty\mathbf{T}^T \ , \tag{6}$$

thus

$$\begin{aligned}\mathbf{P}'\mathbf{Q}'\mathbf{P}'^T &\sim \mathbf{P}\mathbf{T}^{-1}\mathbf{T}\mathbf{Q}_\infty\mathbf{T}^T\mathbf{T}^{-T}\mathbf{P}^T \sim \mathbf{P}\mathbf{Q}_\infty\mathbf{P}^T \\ &\sim \mathbf{K}\mathbf{K}^T \ .\end{aligned} \tag{7}$$

For different cameras,

$$\mathbf{P}_i'^T\mathbf{Q}'\mathbf{P}_i' \sim \mathbf{K}_i\mathbf{K}_i^T \quad (i = 1, 2, \ldots, m) \ . \tag{8}$$

These are the classical auto-calibration constraints, which are independent of the choice of a projective basis. They relate the constraints on the calibration matrices to the constraints on the absolute quadric $\mathbf{Q}'$ in the projective space. As clarified in [11, 2] some prior knowledge about the viewing camera is needed to solve these equations, zero skew or known aspect ratio for example. For this reason, we classify these constraints into the camera based constraints.

3 Scene based constraints from parallel lines

From the ***Property 2***, we see that the absolute quadric encodes the affine structure. In other words, if some affine information can be extracted from the scenes, then they can be used to improve the estimation of the absolute quadric.

Consider the absolute quadric $\mathbf{Q}'$. Given a finite plane π, $\mathbf{Q}'\pi$ is the point at infinity representing its normal direction. The plane at infinity π_∞ is $\mathbf{Q}'$'s null vector

$$\mathbf{Q}'\pi_\infty = \mathbf{0} \ . \tag{9}$$

A set of parallel lines intersect at a direction point $\mathbf{X}$ at infinity which must be on the plane at infinity

$$\pi_\infty^T \mathbf{X} = 0 \ . \tag{10}$$

Consider the symmetry of $\mathbf{Q}'$. From Eqn. (9)

$$\pi_\infty^T \mathbf{Q}' = \mathbf{0} \ , \tag{11}$$

and

$$\pi_\infty^T \left(\mathbf{Q}' \ \mathbf{X}\right) = \mathbf{0} \ . \tag{12}$$

For n sets of parallel lines, the above equation becomes

$$\pi_\infty^T \left(\mathbf{Q}' \ \mathbf{X}_1 \ \mathbf{X}_2 \cdots \mathbf{X}_n\right) = \mathbf{0} \ , \tag{13}$$

where $\mathbf{X}_i$ is the i-th direction point.

lemma 4.1 The rank of matrix $(\mathbf{Q}' \ \mathbf{X}_1 \ \mathbf{X}_2 \cdots \mathbf{X}_n)$ is three
Proof π_∞ is the 4×1 norm vector of the plane at infinity, the rank of its orthogonal complement space is three, i.e., $Rank(\mathbf{Q}' \ \mathbf{X}_1 \ \mathbf{X}_2 \cdots \mathbf{X}_n) \leq 3$. The absolute quadric has a rank three, $Rank(\mathbf{Q}' \ \mathbf{X}_1 \ \mathbf{X}_2 \cdots \mathbf{X}_n) \geq 3$. Thus, $Rank(\mathbf{Q}' \ \mathbf{X}_1 \ \mathbf{X}_2 \cdots \mathbf{X}_n) = 3$
Algorithm 1.

- Obtain initial absolute quadric $\mathbf{Q}'$ from camera based constraints
- Arrange $\mathbf{Q}'$ and the direction points as matrix $\mathbf{M} = (\mathbf{Q}' \ \mathbf{X}_1 \ \mathbf{X}_2 \cdots \mathbf{X}_n)$
- Get SVD decomposition of $\mathbf{M}$, $\mathbf{M} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T$.
- If the fourth singular value in $\mathbf{S}$ is not zero set it to zero and get new matrix $\mathbf{M}' = \mathbf{U} \cdot \mathbf{S}' \cdot \mathbf{V}^T = (\hat{\mathbf{Q}} \ \hat{\mathbf{X}}_1 \ \hat{\mathbf{X}}_2 \cdots \hat{\mathbf{X}}_n)$
- Replace $\hat{\mathbf{X}}_1 \ \hat{\mathbf{X}}_2 \cdots \hat{\mathbf{X}}_n$ with original $\mathbf{X}_1 \ \mathbf{X}_2 \cdots \mathbf{X}_n$ and repeat from the third step.

Finally, the absolute quadric $\hat{\mathbf{Q}}$ which satisfies the parallel line constraints is obtained. In this algorithm, we not require the number of the direction points be three, it could be two even one. So, it is flexible for dealing with the different cases.

4 Scene based constraints from orthogonal planes

From the ***Property 3***, we see that since the absolute quadric encodes the Euclidean structure, some Euclidean characteristics of the scenes can be used as constraints to improve the estimation of the absolute quadric. Consider two orthogonal planes π_{e1} and π_{e2} in Euclidean space with their normal directions orthogonal.

$$\pi_{e1} \perp \pi_{e2} \Leftrightarrow \pi_{e1}^T \begin{pmatrix} \mathbf{I}_{3\times 3} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} \pi_{e2} = \pi_{e1}^T \mathbf{Q}_\infty \pi_{e2} = 0 \ , \tag{14}$$

i.e., the norms of two orthogonal planes are conjugated with respect to the absolute quadric. This is also true in a projective space, because $\pi_{e1} \rightarrow \mathbf{T}^{-T}\pi_{e1}$, $\pi_{e2} \rightarrow \mathbf{T}^{-T}\pi_{e2}$ and $\mathbf{Q}_\infty \rightarrow \mathbf{T}\mathbf{Q}_\infty\mathbf{T}^T$, where

$$\pi_1^T \mathbf{Q}' \pi_2 = (\mathbf{T}^{-T}\pi_{e1})^T \cdot \mathbf{T}\mathbf{Q}_\infty\mathbf{T}^T \cdot \mathbf{T}^{-T}\pi_{e2} = \pi_{e1}^T \mathbf{Q}'_\infty \pi_{e2} = 0 \ . \tag{15}$$

The Eqn. (15) gives the linear form

$$\pi_1^T \mathbf{Q}' \pi_2 = \sum_{ij} c_{ij} \mathbf{Q}'_{ij} = 0 \ , \ i,j = 1,2,3,4 \tag{16}$$

where $\mathbf{Q}'_{ij}$ is the entries of the matrix $\mathbf{Q}'$, its corresponding coefficients are

$$c_{ij} = \pi_1(i)\pi_2(j) \ , \tag{17}$$

thus every pair of orthogonal plane will give a linear constraint on the absolute quadric. Since the absolute quadric has nine degrees of freedom, except the rank three constraints, it still needs at least eight pairs of orthogonal planes to completely determine, which is not a easy satisfied requirement. It is reasonable to combine the orthogonal constraints with camera based constraints to determine the absolute quadric.

4.1 Integration of all constraints

Before the integration, a summary of all constraints is listed as follows.

- Camera based constraints

$$\mathbf{P}_i'^T \mathbf{Q}' \mathbf{P}_i' \sim \omega_{\infty i}^* \sim \mathbf{K}_i\mathbf{K}_i^T \quad (i = 1,2,\ldots,m) \ ,$$

- Rank three constraint

$$Rank(\mathbf{Q}') = 3 \ ,$$

- Parallel line constraint

$$Rank(\mathbf{Q}' \ \mathbf{X}_1 \ \mathbf{X}_2 \cdots \mathbf{X}_n) = 3 \ ,$$

- Orthogonal plane constraint

$$\pi_1^T \mathbf{Q}' \pi_2 = 0 \ .$$

Because the absolute quadric is up to a scale factor, its norm is often normalized to one. Thus in addition to the above camera based and the scene based constraints, the norm one constraint is also considered.

- Norm constraint

$$\|\mathbf{Q}'\| = 1 \ .$$

It can be seen that rank three constraint is contained in the parallel line constraints. From these constraints, a linear method is given to obtain an initial estimation of the absolute quadric. Then a non-linear iteration algorithm is proposed to find optimized solution.

5 Experiments

Simulated data Firstly, we use simulated data to evaluate our auto-calibration algorithm. A simple scene consisting of 86 points, several planes and lines is shown in the Figure 1. There are three pairs of rothogonal planes and three sets of parallel lines in the scene. A viewing camera is simulated with its focal length distributed within 1500 ± 500 pixels and its principle point is ± 50 pixels around the image center. The image dimension is about 500×600. We evaluated our method using a short image sequence and a long sequence separately. The view numbers are 4 and 10 in the two cases. In each case, 1 and 2 pixels noise are added to images to test the robustness of our algorithm.

The results are measured by auto-calibration residual errors and scene based constraints separately. The residual errors are the fitness of the camera based constraints, which are represented by $\| \frac{\mathbf{P}\mathbf{Q}'\mathbf{P}^T}{\|\mathbf{P}\mathbf{Q}'\mathbf{P}^T\|} - \frac{\mathbf{K}\mathbf{K}^T}{\|\mathbf{K}\mathbf{K}^T\|} \|$, shown in Figure 4. The other is the fitness of the scene based constraints, which is evaluated by line parallelism and plane orthogonality. The plane orthogonality is measured by the angles of each pair of reconstructed orthogonal plane. For measuring the line parallelism, we consider the fact: If the reconstructed lines are parallel, their intersection should be a point at infinity. Thus the line parallelism is measured by the distance from the scene center to their intersection, normalized by the largest dimension of the scene. The experimental results are shown in the Table 1 and Table 2, where the Method A is classical auto-calibration method without scene based constraints and the Method B is our method using scene based constraints.

The real image sequence I and II The first real image sequence consists of 15 images, some of which are shown in Figure 2. Initially, a projective reconstruction is obtained using our method. From the projective reconstruction, the camera auto-calibration and Euclidean reconstruction are computed without using scene constraints. Since the view number is large enough, and the camera motion is general enough, the camera based constraints have redundant information to give a good Euclidean results. This can be seen in the Figure 6. From the front view and the top view, we can see that the line parallelism and the plane orthogonality are satisfied. In this case, using the scene based constraints, the results have not obvious improvements.
In the second image sequence, there are totally 19 images. some of which are shown in Figure 3. Although the residual errors of auto-calibration are small, but the reconstructed structure still has obvious projective distortions. As shown in Figure 7. In order to use scene based constraints, we extract the two point sets of the orthogonal planes using the obtained homographies, and the three normal directions corresponding to three sets of parallel lines. All these results are extracted from the projective reconstruction, as shown in Figure 9. We use these scene based constraints in our algorithm and get the improved results. The reconstructed building by using scene based constraints is shown in Figure 10, it looks more Euclidean than the previous one.

6 Conclusions

In this paper, we propose a new approach for modeling buildings from their uncalibrated video sequences integrating the scene based constraints. We transform the line parallelism and plane orthogonality into the direct constraints on the absolute quadric for camera auto-calibration. These additional scene based constraints are used to improve the estimation of the absolute quadric. Due to some ill-conditions, the previous approach give some unsatisfied results. But our approach is effective in these cases.

Acknowledgement

This research is partially supported by the RGC grant CUHK 4402/98E.

References

[1] R. Hartley, "Projective reconstruction from uncalibrated views," *Applications of Invariance in Computer Vision, Lecture Notes in Computer Science*, vol. 825, pp. 237–256, 1994.

[2] M. Pollefeys, R. Koch, and L Van Gool, "Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters," *International Journal of Computer Vision*, vol. 32, Kluwer Academic Publishers, no. 1, pp. 7–25, 1999.

[3] O. Faugeras, Q.-T. Luong, and S. Maybank, "Camera self-calibration: Theory and experiments," *Computer Vision - ECCV'92, Lecture Notes in Computer Science*, vol. 588, Springer-Verlag, pp. 321–334, 1992.

[4] A. Heyden and K.Åström, "Flexible calibration: Minimal cases for auto-calibration," *Proc. International*

Conference on Computer Vision, Kerkyra, Greece, pp. 350–355, 1999.

[5] B. Triggs, "The absolute quadric," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, vol. IEEE Computer Soc. Press, pp. 609–614, 1997.

[6] M. Pollefeys and L. Van Gool, "Self-calibration from the absolute conic on the plane at infinity," *Proc. International Conference on Computer Analysis of Images and Patterns*, pp. 175–182, Kiel. Germany, 1997.

[7] A. Heyden and K. Åström, "Euclidean reconstruction from constant intrinsic parameters," *Proc. 13th International Conference on Pattern Recongnition*, vol. IEEE Computer Soc. Press, pp. 339–343, 1996.

[8] M. Pollefeys, R. Koch, and L. Van Gool, "Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters," in *International Conference on Computer Vision*, 1998.

[9] P. Sturm, "Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction," *Proc. 1997 Conference on Computer Vision and Pattern Recognition*, vol. IEEE Computer Soc. Press, pp. 1100–1105, 1997.

[10] F. Kahl, "Critical motions and ambiguious euclidean reconstruction in auto-calibration," *International Conference on Computer Vision*, vol. 1, pp. 469–475, 1999.

[11] A. Heyden and K. Åström, "Euclidean reconstruction from image sequences with varying and unknown focal length and principle point," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, vol. IEEE Computer Soc. Press, pp. 438–443, 1997.

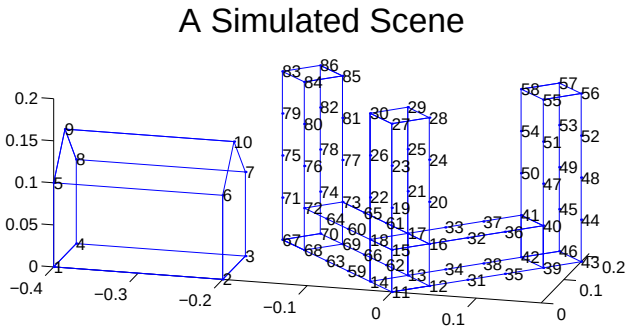


Figure 1. A simple simulated scene

Figure 2. The image sequence of a Spanish building

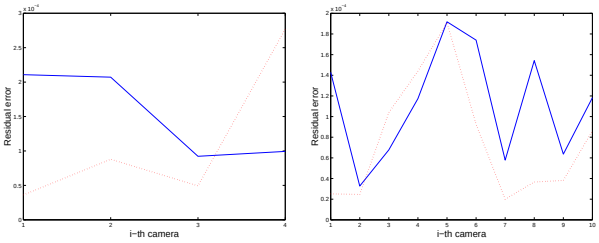
Figure 3. The image sequence of the Louvre museum

Noise = 1 pixel	View number = 4		view number = 10	
	Method A	Method B	Method A	Method B
angle (1,2)	85.8145	88.2263	91.5272	90.5277
angle (3,4)	89.6624	90.1068	90.3918	89.9589
angle (5,6)	96.9698	91.5138	92.4006	91.1760
inf point 1	8.6184	1.3550e+06	129.7338	2.9837e+05
inf point 2	7.9685	1.4829e+04	31.4970	6.5203e+04
inf point 3	7.0242	19.6086	71.7741	48.7308

Table 1. The measurements of orthogonality and parallelism using simulated data with noise = 1 pixel

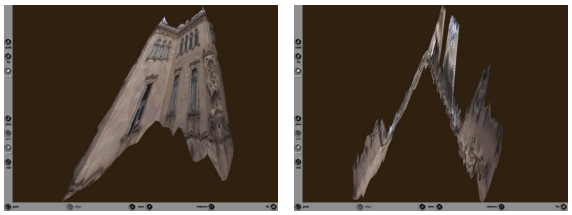
Noise = 2 pixels	View number = 4		view number = 10	
	Method A	Method B	Method A	Method B
angle (1,2)	82.6167	86.1513	94.1899	90.6617
angle (3,4)	90.7107	90.1661	97.2331	91.6818
angle (5,6)	79.1220	87.0213	81.4474	90.2430
inf point 1	5.1739	3.2744e+04	4.9076	2.6800e+04
inf point 2	5.4757	1.9183e+04	15.2913	4.2614e+03
inf point 3	4.0850	6.8521	3.7777	25.7468

Table 2. The measurements of orthogonality and parallelism using simulated data with noise = 2 pixels



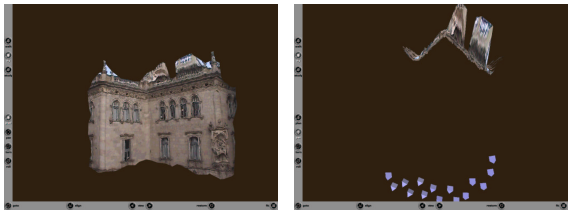
(a) view number = 4 (b) view number = 10

Figure 4. The auto-calibration residual errors of the two approaches. The solid lines are the results with scene based constraints and the dashed lines are the results without scene based constraints



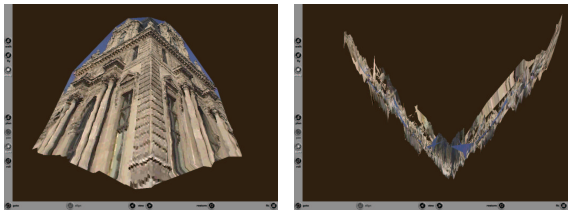
(a) The front view (b) The top view

Figure 5. The projective reconstruction of the first building



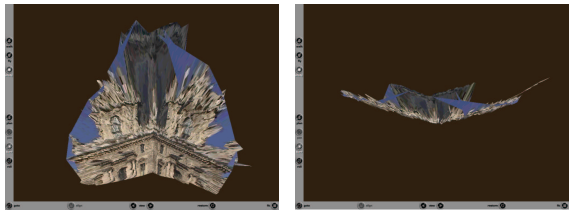
(a) The front view (b) The top view

Figure 6. The first reconstructed building



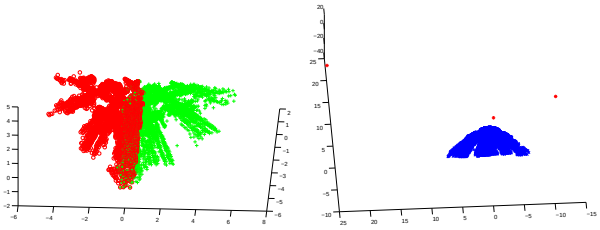
(a) The front view (b) The top view

Figure 7. The projective reconstruction of the second building



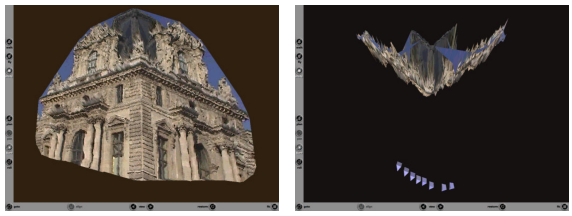
(a) The front view (b) The top view

Figure 8. The second reconstructed building without using the scene based constraints



(a) two orthogonal planes (b) three direction points

Figure 9. The extracted two orthogonal planes consisting of their coplanar points and the three direction points
(a) shows the two sets of points on the two orthogonal planes. One point set is denoted by small circles 'o' and another set by small across '+' (b) shows the projective reconstructed scene (the point cloud) and three direction points, denoted by '*'.



(a) The front view (b) The top view

Figure 10. The second reconstructed building using scene constraints