International Journal of Shape Modeling

METRIC RECONSTRUCTION OF BUILDINGS FROM UNCALIBRATED IMAGE SEQUENCES *

ZEZHI CHEN

ISN National Key Lab., Xidian University,710071, Xi'an , P.R.China
Department of Computer Science, University of York, YO10 5DD, UK
zezhi.chen@cs.york.ac.uk
http://<www.cs.york.ac.uk/∼chen>

CHENGKE WU

ISN National Key Lab., Xidian University,710071, Xi'an , P.R.China
ckwu@ns2.xidian.edu.cn

YONG LIU

ISN National Key Lab., Xidian University,710071, Xi'an , P.R.China
yliu@cvmt.dk

NICK PEARS

Department of Computer Science, University of York, YO10 5DD, UK
nep@cs.york.ac.uk

Received (Day Month Year)
Revised (Day Month Year)
Accepted (Day Month Year)

Communicated by (xxxxxxxxxx)

This paper presents a new approach for reconstructing realistic 3D models of buildings from uncalibrated image sequences taken by a hand-held camera. Firstly, correspondences between image pairs are established by using various computer vision tools, and then the fundamental matrix is estimated to high accuracy. Meanwhile, homography constraints are exploited to find more correspondences, to avoid degenerate cases and to obtain more accurate results. Secondly, rectified image pairs are resampled by using epipolar geometry constraints, where epipolar lines coincide with image scan-lines and disparities between the images are in the x-direction only. This allows subsequent stereoscopic analysis algorithms to easily take advantage of the epipolar constraint and reduce the search space to one dimension, namely along the horizontal row of the rectified images. Furthermore, dense stereo matching of the original image pairs is simple and low computational cost. Finally, the 3D model can be built through self-calibration, matching and Delaunay triangulation. The self-calibration method uses prior knowledge of

*This project is supported by the National Natural Science Foundation of China (69972039; 60002007), France-China Advanced Research Program (PRA SI00-04) and Hong Kong RGC grant.

orthogonal planes (lines) and parallel lines to act as constraints on the absolute quadric. A large number of experimental results show that this method improves the speed and accuracy of reconstructed 3D models and the 3D models obtained are more realistic.

Keywords: Metric reconstruction; Epipolar constraint; Self-calibration; Rectification; Dense matching.

1. Introduction

Modeling of 3D objects from image sequence is one of the challenging problems in computer vision and has been a research topic for many years. In the last few years, interest in 3D models has dramatically increased:[1−16] The approach by Tomasi and Kanade[12] is known as the factorization method; The projective factorization method of Triggs[14] uses epipolar constraints to initialize the depth scale factors but essentially it is an iterative algorithm; Heyden et al.[15] upgrades the affine approximations to projective results by iterative optimization. The other kind of method is the camera-centered approach:[1] The first view is used as the reference camera to determine the projection matrices of other cameras in a projective frame under multiple geometric constraints. The world-centered approach is suitable for long image sequences and the camera-centered approach is suitable for the case of fewer images. The photogrammetry approach mostly focuses on accuracy problems, and the derived techniques produce 3D models of high accuracy[16]. However, they generally require heavy human interaction. More and more applications are using computer-generated models, such as synthesis, simulation, virtual and augmented reality, computer graphics etc. Although more tools are at hand to ease the generation of models, it is still a time-consuming and expensive process. In computer vision, researchers have produced a number of automatic techniques for computing structure from multiple views. From these techniques, the 3D models are produced much more easily, but they are less accurate. Traditional solutions include the use of stereo rigs, laser range scanners and other 3D digitizing devices, and these devices are often very expensive, require careful handling and complex calibration procedures and are designed for a restricted depth range only. However, in reconstruction without any knowledge about the scene, a scale ambiguity is always present, because it is impossible to distinguish between a large object far away and a small object close to the camera. This means that it is only possible to reconstruct the object up to a similarity transformation, that is Euclidean transformation plus a uniform change of scale. Such a reconstruction is termed a metric reconstruction in the computer vision literature, although it is sometimes loosely referred to as Euclidean. If scale needs to be known, this can be obtained from one ground truth measurement and the correct scale of the 3D model can be set to give a strictly Euclidean 3D model.

In this paper, we address the problem of the recovery of a realistic textured model from an image sequence, without any prior knowledge of either the parameters of the cameras or their motion. The reconstruction method proposed here avoids most of the problems mentioned above. First, the features are classified into

coplanar and non-planar sets. Each set has a different weight factor in calculating the epipolar geometry. Then, from the epipolar geometry of every pair of consecutive images, the structure-and-motion problem can be solved up to a projective transformation. Second, observing the prior knowledge that there are many orthogonal planes (lines) and parallel lines in a building, we transform these characteristics into several Euclidean invariant constraints for self-calibration. Under these new constraints, the (near) orthogonality and parallelism of planes and lines is guaranteed, so that the reconstructed buildings will be more realistic or "Euclidean".

The paper is organized as follows. In section 2, feature matching and projective reconstruction are introduced. In section 3, we show how to determine the camera projective matrix and camera intrinsic parameters. Section 4 shows the details of rectification and dense matching. Metric 3D reconstruction is presented in section 5 and experimental results are given in section 6. Section 7 concludes the paper.

2. Feature Matching and Projective Reconstruction

In order to reconstruct a 3D model, the first problem is the correspondence problem. Given a feature in an image, what is the corresponding feature (i.e. the projection of the same 3D feature) in the other image? This is an ill-posed problem and is often very hard to solve. When some assumptions are satisfied, it is possible to automatically match points or other features between images. One of the most useful assumptions is that the images are not too different from each other. In this case, the coordinates of the features and the intensity distribution around those feature are similar in both images. This allows us to restrict the search space and to match features through intensity cross-correlation. The extraction of interest points should be, as much as possible, independent of camera pose and illumination changes and the neighborhood of these points should contain as much information as possible to allow matching. An interest point detector (KLT) is used to select a certain number of matched points[17] and these points should be well located and indicate salient features that stay visible in consecutive images.

In order to determine the fundamental matrix, at least eight pairs of correspondences are needed for a linear solution and seven pairs are needed a non-linear solution.[18] Here, the LMedS algorithm of Zhang et al. in Ref. 19 and RANSAC (RANdom SAmpling Consensus) algorithm of Torr in Ref. 20 can be used.

Several sequences are used to assess the quality of the algorithm proposed in this paper. The first sequence used consists of seven images of Xi'an Bell Tower (a famous place of interest in China and a typical ancient architecture). A pair of images and the associated epipolar lines, computed using RANSAC, are shown in Fig.1. Once such an initial epipolar geometry, represented by the RANSAC epipolar lines, has been recovered, one can start looking for more matches to refine the fundamental matrix. In this case, the search region is restricted to a few pixels around the epipolar lines. Furthermore, one can initialise the projective reconstruction.

The first two images of a sequence are used to determine a reference frame.

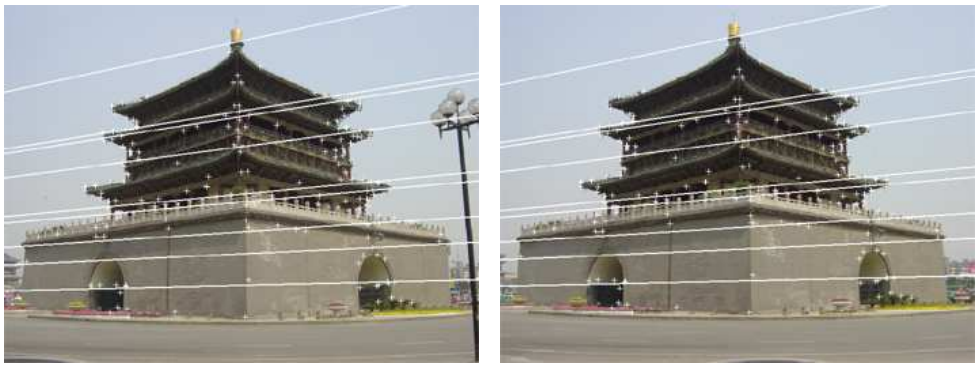


Fig. 1. A pair of images from the Xi'an Bell Tower sequence, showing 200 interest points (extracted using KLT) and some of the epipolar lines (computed using RANSAC).

The world frame is aligned with the first camera (position). The second camera is chosen so that epipolar geometry corresponds to the recovered $\mathbf{F}_{12}$

$$\mathbf{P}_1 = [\mathbf{I}_{3\times3}|\mathbf{0}_3]\,. \tag{1}$$

$$\mathbf{P}_2 = \left[[\mathbf{e}_{12}]_\times \mathbf{F}_{12}|\mathbf{e}_{12}\right]. \tag{2}$$

where $[\mathbf{e}_{12}]_\times$ indicates the vector product with $\mathbf{e}_{12}$, $\mathbf{e}_{12}$ is the epipole on the second image plane. Then, the projective reconstruction $\mathbf{X}$ can be obtained from

$$\begin{bmatrix} \mathbf{x} & \mathbf{0} & \mathbf{P}_1 \\ 0 & \mathbf{x}' & \mathbf{P}_2 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \mathbf{X} \end{bmatrix}. \tag{3}$$

where λ_1 and λ_2 are scale factors. Once the first two camera projection matrices are determined, they can be used as the reference frames for determining other projection matrices. For a pair of correspondences in the first two images $\mathbf{x}$ and $\mathbf{x}'$, their correspondence in the third image can be predicted as described in the above section. Every new corresponding point $\mathbf{x}''$ in the third image will give two constraints on the third projection matrix. At least six points in general position are needed to solve the third projection matrix. Since the reference frames are the first two images, there may be accumulated errors in the projective matrices of later frames in the image sequence. Thus, an optimal algorithm is used to refine the solutions in a least-square (LS) sense for reprojection errors for all frames. This is the well-known bundle adjustment method.[21]

3. Self-calibration

The projection of a scene onto an image can be modeled by the following equation:

$$\lambda\mathbf{m} = \mathbf{PM}. \tag{4}$$

where $\mathbf{m} = [x, y, 1]^T$ is an image point and $\mathbf{M} = [X, Y, Z, 1]^T$ is a scene point. $\mathbf{P}$ is the 3×4 camera projection matrix and λ is a scale factor. The camera projection matrix is factorized as follows:

$$\mathbf{P} = \mathbf{K}\,[\mathbf{R}| - \mathbf{R}\mathbf{t}] \quad with \;\; \mathbf{K} = \begin{bmatrix} f_x & s & u \\ & f_y & v \\ & & 1 \end{bmatrix}. \tag{5}$$

Here $(\mathbf{R}, \mathbf{t})$ denotes a rigid transformation, while the upper triangular calibration matrix $\mathbf{K}$ includes the intrinsic parameters of the camera (i.e. f_x and f_y represent the focal length divided by the pixel width and height respectively, (u, v) represents the principal point and s is a factor which is zero in the absence of skew). Note that the factorization of the camera projection matrices in Eq.(5) yields the physical parameters of the camera, which are necessary to generate a metric reconstruction. For the actual computations the absolute conic ω is used. The plane at infinity(π_∞) and the absolute conic are all invariant under Euclidean transformations and they include the affine and metric properties of buildings, which means that, when the position of π_∞ is known in a projective framework, affine invariants can be measured. Since the absolute conic is invariant under Euclidean transformations, its image only depends on the intrinsic camera parameters and not on the extrinsic camera parameters.[1,22]

The way in which the method presented here differs from other methods in the literature is as follows. Firstly, we put constraints on the absolute quadric instead of the images of the absolute conic. As a result, not only orthogonal/parallel lines but also orthogonal/parallel planes can be used as constraints directly. Secondly, we deal with several images instead of a pair of images, so the quadric forms are more robust and concise.

In 3D space, there is a special entity known as the absolute quadric $\mathbf{Q}_\infty$, which is a dual quadric.

$$\mathbf{Q}_\infty \sim \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}. \tag{6}$$

It is the tangent planes to the absolute conic which satisfy the equation of a dual quadric. A tangent plane π of the dual quadric satisfies

$$\pi^T \mathbf{Q}_\infty \pi = 0. \tag{7}$$

Consider a Euclidean transformation $\mathbf{T}_e$ applied to the absolute quadric $\mathbf{Q}_\infty$ as follows:

$$\mathbf{Q}'_\infty \sim \mathbf{T}_e \mathbf{Q}_\infty \mathbf{T}_e^T \sim \begin{bmatrix} \mathbf{R}^T & -\mathbf{R}^T\mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}^T & -\mathbf{R}^T\mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}^T = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \sim \mathbf{Q}_\infty. \tag{8}$$

The absolute quadric contains information on both the absolute conic and the plane at infinity. As introduced by Ref. 27, it is easier to be used in self-calibration than the absolute conic. Since the absolute quadric is invariant under Euclidean transformations, its images on different image planes do not depend on the motions

of the cameras and depend only on the intrinsic characteristics of the cameras. Thus, some constraints can be derived directly on the intrinsic parameters of the cameras.[22] From Eqs.(5) and (6) the absolute quadric is related to its image ω_∞^* in Euclidean space as

$$\mathbf{P}\mathbf{Q}_\infty\mathbf{P}^T \sim \mathbf{K}\left[\mathbf{R}^T - \mathbf{R}^T\mathbf{t}\right]\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & 0\end{bmatrix}\left[\mathbf{R}^T - \mathbf{R}^T\mathbf{t}\right]^T\mathbf{K}^T = \mathbf{K}\mathbf{K}^T \sim \omega_\infty^*. \tag{9}$$

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

$$\mathbf{Q}' \sim \mathbf{T}\mathbf{Q}_\infty\mathbf{T}^T. \tag{10}$$

and

$$\mathbf{P}'\mathbf{Q}'\mathbf{P}'^T \sim \mathbf{P}\mathbf{T}^{-1}\cdot\mathbf{T}\mathbf{Q}_\infty\mathbf{T}^T\cdot\mathbf{T}^{-T}\mathbf{P}^T = \mathbf{P}\mathbf{Q}_\infty\mathbf{P}^T \sim \mathbf{T}\mathbf{T}^T \sim \omega_\infty^*$$

Thus

$$\mathbf{P}_i'\mathbf{Q}'\mathbf{P}_i'^T \sim \omega_{\infty i}^* \sim \mathbf{K}_i\mathbf{K}_i^T \qquad i = (1, 2, \cdots, m). \tag{11}$$

These are absolute quadric projection constraints, which are independent of the choice of projective basis. They can translate constraints on the calibration matrices to constraints on the absolute quadric $\mathbf{Q}'$ in the projective space. What camera self-calibration has to do is to extract the information of the calibration matrices. This is equivalent to obtaining the transformed absolute quadric $\mathbf{Q}'$ or the transformation $\mathbf{T}$.

Before deriving the scene based constraints, it is required to clarify how the concepts such as orthogonality of planes, orthogonality of lines, parallel planes and parallel lines can be related to 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.

$$\mathbf{Q}'\pi_\infty = 0. \tag{12}$$

A point $\mathbf{X}$ at infinity must be on the plane at infinity

$$\pi_\infty^T\mathbf{X} = 0. \tag{13}$$

Consider the symmetry of $\mathbf{Q}'$, and from Eq.(12)

$$\pi_\infty^T\mathbf{Q}' = 0. \tag{14}$$

It can be seen that the point at infinity $\mathbf{X}$ must be in the space formed by the columns of $\mathbf{Q}'$. Thus

$$Rank\,(\mathbf{Q}' \;\; \mathbf{X}) = Rank\,(\mathbf{Q}') = 3. \tag{15}$$

For a set of parallel lines, they intersect at a point at infinity. Actually one can find several sets of parallel lines for a building, and every point at infinity can give one such constraint.

Consider two orthogonal planes π_1 and π_2 with their normal directions orthogonal under Euclidean space.

$$\pi_1 \perp \pi_2 \Longrightarrow \pi_1^T \cdot \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \cdot \pi_2 = \pi_1^T \mathbf{Q} \pi_2 = 0. \tag{16}$$

under a projective transform $\mathbf{T}$, $\pi_1 \rightarrow \mathbf{T}^{-T}\pi_1 = \pi_{p1}$, $\pi_2 \rightarrow \mathbf{T}^{-T}\pi_2 = \pi_{p2}$.

$$(\pi_{p1})^T \cdot \mathbf{Q}' \cdot \pi_{p2} = (\mathbf{T}^{-T}\pi_1)^T \cdot \mathbf{Q}' \cdot \mathbf{T}^{-T}\pi_2 = (\mathbf{T}^{-T}\pi_1)^T \cdot \mathbf{T}\mathbf{Q}\mathbf{T}^T \cdot \mathbf{T}^{-T}\pi_2 = \pi_1^T \mathbf{Q} \pi_2 = 0. \tag{17}$$

Suppose the homogeneous representation of these two planes π_{p1}, π_{p2} are $\mathbf{X}_1 = (x_{11}, x_{12}, x_{13}, x_{14})^T$ and $\mathbf{X}_2 = (x_{21}, x_{22}, x_{23}, x_{24})^T$, respectively.

$$\mathbf{X}_1^T \mathbf{Q}' \mathbf{X}_2 = 0. \tag{18}$$

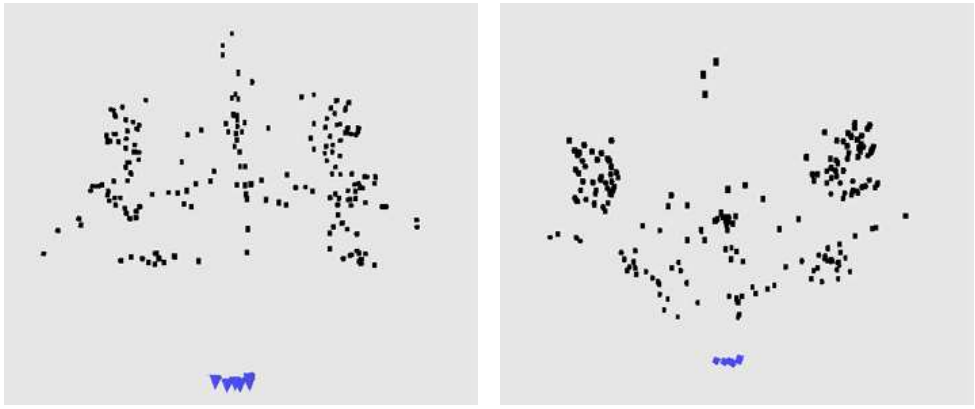
This is a linear equation set

$$\mathbf{X}_1^T \mathbf{Q}' \mathbf{X}_2 = \sum_{i,j}^{4} C_{i,j} x_{1i} x_{2j} = 0. \tag{19}$$

with $C_{i,j}$ representing the i-th row and j-th column of $\mathbf{Q}'$. In this way, some prior information about a building can be used as the constraints on the absolute quadric. Combining these constraints with the absolute quadric-based constraints, the absolute quadric can be solved firstly by assuming the principle point of the camera is fixed at the image center and only allowing the focal length to be varied. This simplified camera model will lead to a set of linear equations in the absolute quadric as shown in Ref. 22. The difference is that in Ref. 22 the linear equations of the absolute quadric do not contain scene based constraints. With the additional scene based constraints for the absolute quadric, a SVD (Singular Value Decomposition) method is used to get a solution of the set of linear equations. This can be used as the initial solution for further non-linear optimization with the actual camera model. In this model, both the focal length and the principle point can be varied and the Levenberg-Marquardt algorithm is used for the non-linear optimization. The estimated absolute quadric $\mathbf{Q}'$ (in projective space) under the combined constraints will verify not only the rigid motions of cameras between different viewpoints, but also the prior information of the building. From the absolute quadric $\mathbf{Q}'$, one can compute the transformation $\mathbf{T}$ from Eq.(10). Furthermore the extrinsic parameters $\mathbf{R}_i$ and $\mathbf{t}_i$ can be solved as shown in Eq.(5). The front view and top view of the metric 3D model of sparse matched points and the positions of cameras are shown in Fig.2 and, here, triangles represent the positions of the cameras.

4. Rectification and Dense Matching

Since we have computed the calibration between successive image pairs, we can exploit the epipolar constraint that restricts the correspondence search to a 1-D search range. It is possible to resample the image pairs to a standard geometry, so that the epipolar lines coincide with the image scan-lines. The correspondence



Front view Top view

Fig. 2. Metric 3D model of sparse matched points and the positions of the camera.

search then reduces to matching of image points along each image scan-line. The key idea of our new rectification method is based on resampling the image using epipolar geometry constraints. There are two stages in the method. (i) The first step consists of determining the common regions for both images, which means that the regions of the two images include the same 3D scene information. (ii) The second step is the pixel-by-pixel resampling step. Here the rectified image is resampled starting from one of the extreme epipolar lines, which are the epipolar lines that pass through the four corners of the image. If the epipole is in the image, an arbitrary epipolar line can be chosen as starting point. In these cases, boundary effects can be avoided by adding an overlap of the size of the stereo algorithm's matching window. Note that if the original image size is $m \times n$, an upper bound for the resampled image size is $2(m+n) \times \sqrt{m^2+n^2}$. The two steps of the rectification method and then the dense stereo matching algorithm are described in detail in the following paragraphs.

Determining the common regions. Before determining the epipolar lines common to both views, the extreme epipolar lines should be determined for each image. Suppose the image size is $m \times n$, then the corners, **A**, **B**, **C** and **D**, starting from the top left and in an anticlockwise direction, have homogenous coordinates $(0,0,1)^T$, $(0,n-1,1)^T$, $(m-1,n-1,1)^T$ and $(m-1,0,1)^T$ respectively. The epipolar lines l_3, l_4, l'_1 and l'_2 are given as follows:

$$l'_1 \sim \mathbf{FD} \quad l'_2 \sim \mathbf{FB} \quad l_3 \sim \mathbf{F}^T\mathbf{D} \quad l_4 \sim \mathbf{F}^T\mathbf{B} \tag{20}$$

where $\sim$ means equality up to a non-zero scale factor. If the line l'_1 intersects the infinite length ideal line that includes the right border of image II and the intersection is outside of image II, as shown in (Fig.3), then the extreme upper epipolar line of image I is l_3. Otherwise, the extreme upper epipolar line of the image I is l_1. The extreme lower epipolar line of the image I can be obtained through

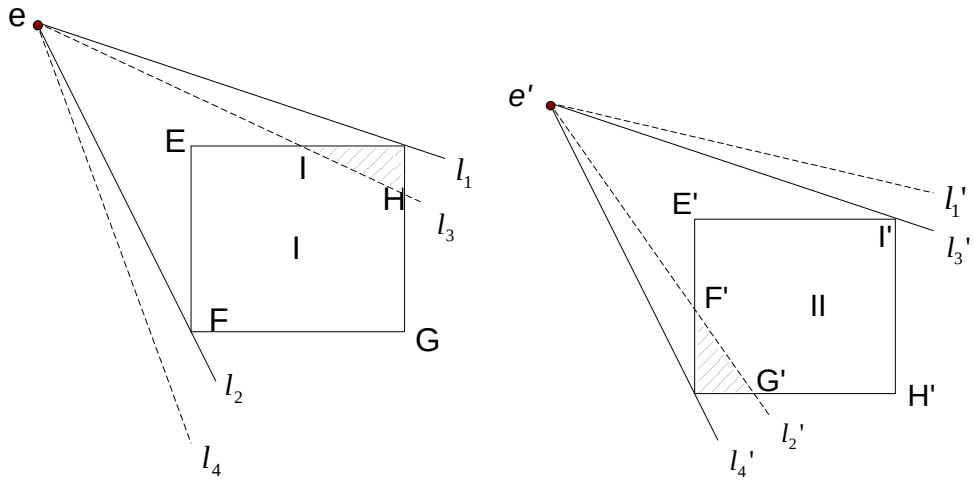


Fig. 3. Determination of the common regions

the same procedure. The common regions $EFGHI$ in image I and $E'F'G'H'I'$ in image II are thus extracted as shown in (Fig.3).

Resampling the rectified image. To avoid losing image information, the area of every pixel should be at least preserved when transforming to the rectified image. In order to draw the epipolar line and avoid information loss, every pixel along an epipolar line can be extracted by using Bresenham's algorithm.[23] Drawing new epipolar lines is accomplished by calculating intermediate positions along the linear path between two endpoints. An output device is then directed to fill in these positions between the endpoints. In this case, the obtained epipolar line should be transferred back to the first image. The minimum of both computed displacements is the one that is used. At the same time, the coordinates of every pixel along an epipolar line are saved in a list for later reference (in order to transfer back to the original images). Note that the resulting performance is superior to that of rectification using homographies.

Every pixel in the rectified images corresponds to a unique pixel in the original images, so information about a specific point in the original image can be obtained by looking it up on a pixel-by-pixel basis in the saved list. The resampled images and a few of the epipolar lines are shown side-by-side in Fig.4. Clearly, the epipolar lines coincide with the image scan-lines.

Dense stereo matching. For dense correspondence matching, a disparity estimator based on the dynamic programming scheme of Cox et al.[24] is employed and it incorporates other constraints, such as preserving the order of neighboring pixels, bi-directional uniqueness of the match, and detection of occlusions. It operates on

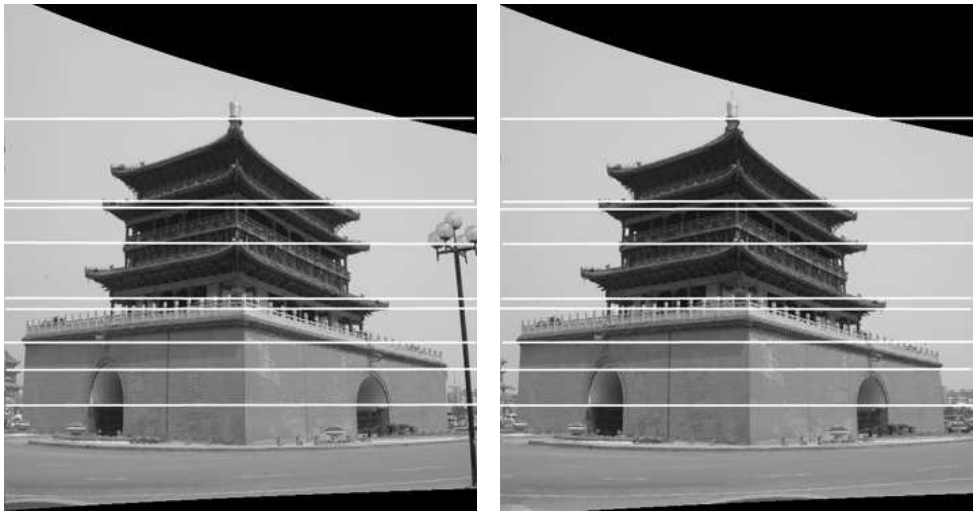


Fig. 4. The resampled images and some of the epipolar lines.

rectified image pairs (I_L, I_R), where the epipolar lines coincide with image scanlines. The point matching algorithm searches at each pixel in image I_L for maximum normalized cross correlation in I_R by shifting a small measurement window along the corresponding scan line. The selected search step size (usually 1 pixel) determines the search resolution. Matching ambiguities are resolved by exploiting the ordering constraint in the dynamic programming approach. Fig.5 shows the dense matching depth map.

5. Metric reconstruction

Once the camera projection matrices have been fully determined and dense matching is accomplished, the metric 3D model can be reconstructed through Delaunay triangulation. Fig.6 shows the result of Delaunay triangulation. Here the two dimensional nodes consist of two types, the first of which is produced from a regular grid pattern of the original images. These nodes are distributed evenly and their 3D reconstructions represent the large surfaces of the building. The second kind of node comes from the features of the disparity maps. These features are the responses of the intensity changes of the disparity maps. For a region with large changes of surface depth, more nodes are used to describe it. On the other hand, fewer nodes are used for a smooth surface. Thus, this is an adaptive 2D mesh strategy according to depth discontinuity. The reprojective residual errors of 3D reconstruction of surface features are shown in Fig.7. The average reprojective residual error is 0.74759 pixels, and the standard deviation is 0.32165 pixels.

Having completed point and line reconstruction, we now describe how to convert the sparse 3D features into a form suitable for graphical rendering. To produce

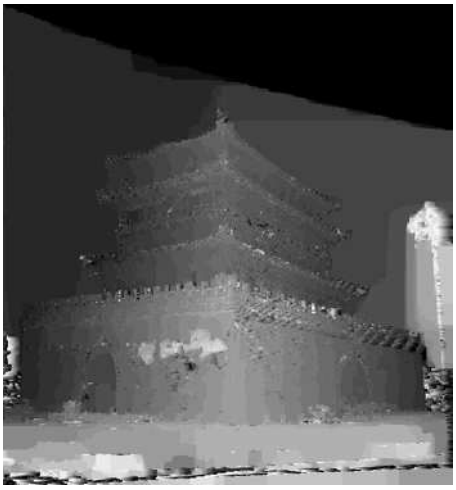


Fig. 5. Dense matching depth map (light indicates near and dark indicates far away)

Fig. 6. Delaunay triangulation

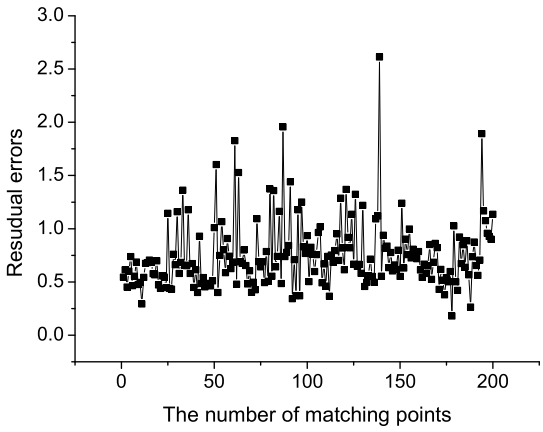


Fig. 7. Reprojective residual errors of 3D reconstruction around surface features

Delauney triangulated structure for the polyhedral examples in this system, planes are textured by selecting (automatically) the image from the sequence that is most fronto-parallel to that plane and texture mapping from the appropriate polygonal image region. As the texture mapping from the image to the plane is via an affine transformation, it is necessary to first warp the image to remove any projective distortion, which is performed automatically. For this experiment 13 lines are ex-

tracted by hand from the surfaces of the building. They are shown in Fig.8 as white lines. Meanwhile, the results of the 3D metric reconstruction of these planes/lines are given. The lines inside each object plane should be parallel to each other (angle between them should be 0 degrees), while lines of different direction should be perpendicular to each other (angle between them should be 90 degrees). Measurements on the object surface are indeed close to the expected values (see Table 1). To judge the visual quality of the final texture-mapping model, different perspective views of the model are computed and displayed in Fig.9.

Table 1. Metric measurements on the Xi'an Bell Tower reconstruction (average).

Parallelism	Orthogonality
0.507 degrees	91.088 degrees

6. Experiments with Other Two Sequences

6.1. *Sequence 1*

This sequence contains images of the library building in the Chinese University of Hong Kong. As the camera-object distance changed, the zoom was changed to keep the image size of the object constant.The first, fifth, seventh and tenth images in the sequence are shown in Fig.10 and Fig.11 shows the final texture-mapped model.

6.2. *Sequence 2*

This sequence shows a gloriette, which is a park building for recreational use. Thirteen images with VGA resolution (640×480) were capture by a hand-held camera. Eight images in the sequence are shown in Fig.12 and, in this experiment, sixty eight matched points were used. The angle between the two vertical walls of the reconstructed gloriette is 89.2354 degrees, and the average angle of the parallel lines of the reconstructed gloriette is 1.7083 degrees. Delaunay triangular surface patches and three perspective views of the reconstructed model are shown in Fig.13.

7. Conclusion

In this paper, we have described a virtual reality modeling system based on computer vision techniques. The system requires at least five images of the scene to be modeled and proceeds to estimate the perspective projection matrices corresponding to all the images by matching image features automatically. The resulting matrices do not in general allow recovery of a metric model of the scene since no metric information has been used so far. In order to do this, the system can do three things: Firstly, self-calibration to estimate the intrinsic parameters of each camera

(active zoom), and camera projection matrices. We transform the characteristics of buildings such as orthogonal/parallel planes and orthogonal/parallel lines into several constraints. These additional scene based constraints are used to improve the estimation of the absolute quadrics. This information allows the system to specialize its representation of the environment from projective to affine and finally to metric. Secondly, image rectification results in a dramatic increase of the computational efficiency and accuracy of dense stereo matching. Finally, the 3D metric model is built through self-calibration, matching and Delaunay triangulation. In producing a 2D mesh for 3D metric reconstruction, not only the 2D feature points are taken as mesh nodes, but the 3D situation is also considered by extracting mesh nodes from the depth map. In this way, a depth adaptive mesh can be obtained to give a more detailed description of the reconstructed 3D surfaces of a building. One of the advantages of this system is that it does not require any prior knowledge about the cameras, and can be usefully and easily applied to shape modeling applications.

Acknowledgements

The authors are very grateful to the anonymous reviewers, whose detailed critical comments and valuable suggestions have enormously improved the clarity of the presentation.

References

1. M. Pollefeys, R. Koch, M. Vergauwen and L. V. Gool, Metric 3D surface reconstruction from uncalibrated image sequences, in *Proc. SMILE Workshop (Post-ECCV'98), LNCS1506*, University of Freiburg, Germany (Jun. 1998),Springer-Verlag, p. 138–153.
2. O, Faugeras *Three-Dimensional Computer Vision: A Geometric Viewpoint* (Cambridge, Massachusetts, London, England: The MIT Press, 1993).
3. F. Devernay and O. Faugeras, From projective to Euclidean reconstruction, in *Proc. of Conference on Computer Vision and Pattern Recognition*, (IEEE Computer Society Press, San Francisco, California, USA, 1996), pp. 264–269.
4. R. Hartley, Euclidean reconstruction from uncalibrated views, in *Applications of invariance in computer vision*, eds. Joseph Mundy and Andrew Zisserman, Volume 825 of Lecture Notes in Computer Science, (Berlin, Germany, 1993),Springer-Verlag, pp. 237–256.
5. J. Costeira and T. Kanade, A multi-body factorization method for motion analysis, in *Proceedings of the 5th International Conference on Computer Vision*, (IEEE Computer Society Press, Boston. MA. June 1995), pp. 1071–1076.
6. Long Quan, Inherent two-way ambiguity in 2D projective reconstruction from three uncalibrated images, in ICCV'99, (Kerkyra, Greece, Sept. 1999), pp. 344–349.
7. Zhengyou Zhang, P. Anandan, Heung-Yeung Shum, What can be determined from a full and a weak perspective image? in ICCV'99, (Kerkyra, Greece, Sept. 1999), pp: 680–687.
8. Long Quan and Takeo Kanade, Affine structure from line correspondences with uncalibrated affine cameras, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **19(8)**(1997) 834–845.

9. Long Quan, Invariants of six points and projective reconstruction from three uncalibrated images, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **17(1)**(1995) 34–46.
10. C. P. Jerian and R. Jain, Structure from motion: a critical analysis of methods, *IEEE Transactions on system. Man and Cybernetics*, **21(3)**(1991) 572–587.
11. Conrad J. Poelman and Takeo Kanade, A Paraperspective Factorization Method for Shape and Motion Recovery, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19(3)**(1997) 206–218.
12. C. Tomasi and T. Kanade, Shape and motion from image streams under orthography: A factorization approach, *International Journal of Computer Vision*, **9(2)**(1992) 137–154.
13. P. Beardsley, P. Torr and A. Zisserman, 3D model acquisition from extended image sequences, in *Proceedings of European Conference on Computer Vision-ECCV'96*, (Lecture Notes in Computer Science, Vol. 1065, Springer-Verlag, 1996), pp.683–695.
14. B. Triggs, Factorization methods for projective structure and motion, in *Proc. Conference on Computer Vision and Pattern Recognition*, (San Francisco, California, USA, 1996), pp. 845–851.
15. A. Heyden and R. berthilsson, An iterative factorization method for projective structure and motion from image sequences, *Image and Vision Computing*, **17(3)**(1999) 981–991.
16. K. B. Atkinson, *Close Range Photogrammetry and Machine Vision* (Whittles Publishing, 1996).
17. Jianbo Shi and Carlo Tomasi, Good features to track, in *IEEE Conference on Computer and Pattern Recognition*, (Seattle, USA, 1994), pp. 593–600.
18. R. Hartley, In defence of the 8-point algorithm, in *Proceedings of the 5th International Conference on Computer Vision (ICCV)*, (IEEE Computer Society Press, Cambridge, MA, USA, 1995), pp. 1064–1070.
19. Zheng-You Zhang, Determining the epipolar geometry and its uncertainty: a review, *International Journal of Computer Vision*, **27(2)**(1998) 161–195.
20. P. H. S. Torr and A. Zisserman, Robust parameterization and computation of the trifocal tensor, *Image and Vision Computer*, **15**(1997) 591–605.
21. R. Hartley et al, *Multiple View Geometry in Computer Vision*, (First Edition. Cambridge: Cambridge University Press, 2000).
22. M. Pollefeys, Self-calibration and metric 3D reconstruction from uncalibrated image sequences, Ph. D Thesis, Department Elektrotechniek Afdeling ESAT, Katholieke Universiteit Leuven (1999).
23. Donald Hearn and M. Pauline Baker, *Computer Graphics* (Prentice-Hall-International Inc. Press, Second edn, 1998).
24. I. Cox, S. Hingorani and S. Rao, A maximum likelihood stereo algorithm, *Computer Vision and Image Understanding*, **63(3)**(1996) 542–567.
25. Zezhi Chen, Chengke Wu, Peiyi Shen, Yong Liu, Long Quan, A robust algorithm to estimate the fundamental matrix, *Pattern Recognition Letters*, **21**(2000) 851–861.
26. H.C.Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, *Nature*, **293**(1981) 133–135.
27. Bill Triggs, Autocalibration and the absolute quadric, in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, Puerto Rico, USA, IEEE Computer Society Press, (June 1997) pp. 609–614.

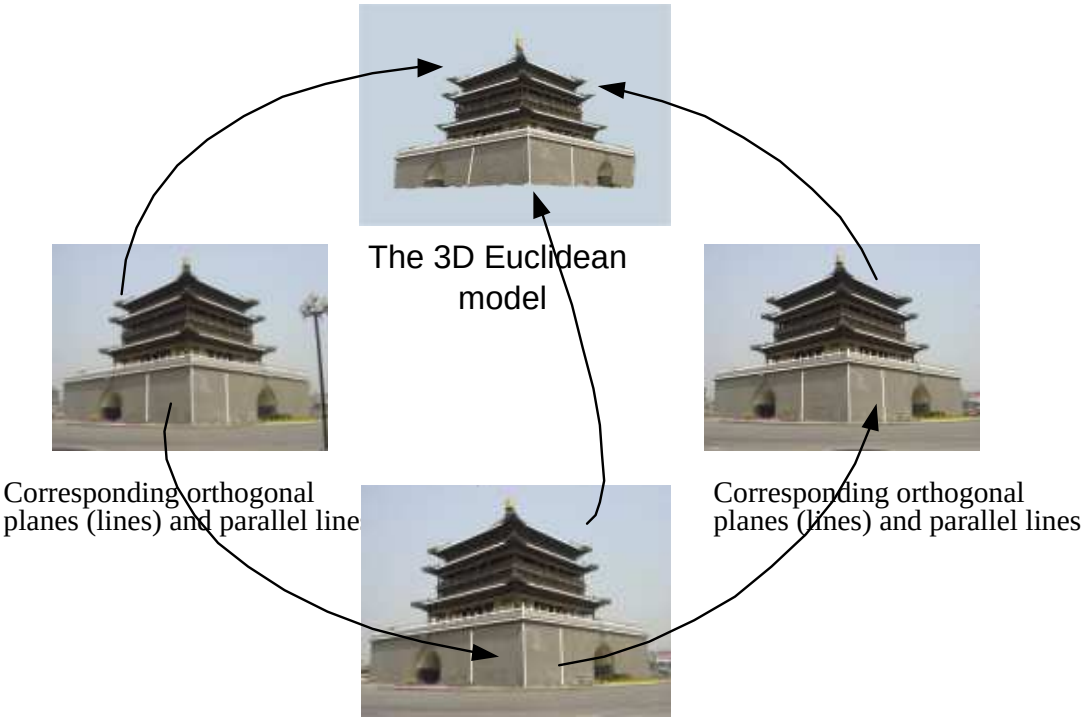


Fig. 8. The three lower images show the orthogonal planes (lines) and parallel lines in the building. The upper image shows the results of 3D metric reconstruction around these planes/lines.

Fig. 9. Perspective views of the Xi'an Bell Tower reconstruction.

Fig. 10. Partial image sequence of the library building at the Chinese University of Hong Kong.

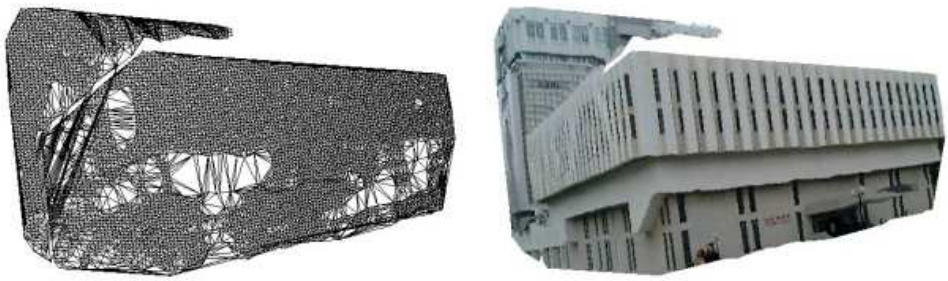


Fig. 11. 3D surface model of the library building, obtained automatically from the uncalibrated image sequence (Delaunay triangular surface patches left, texture right).

Fig. 12. The image sequence of a gloriette.

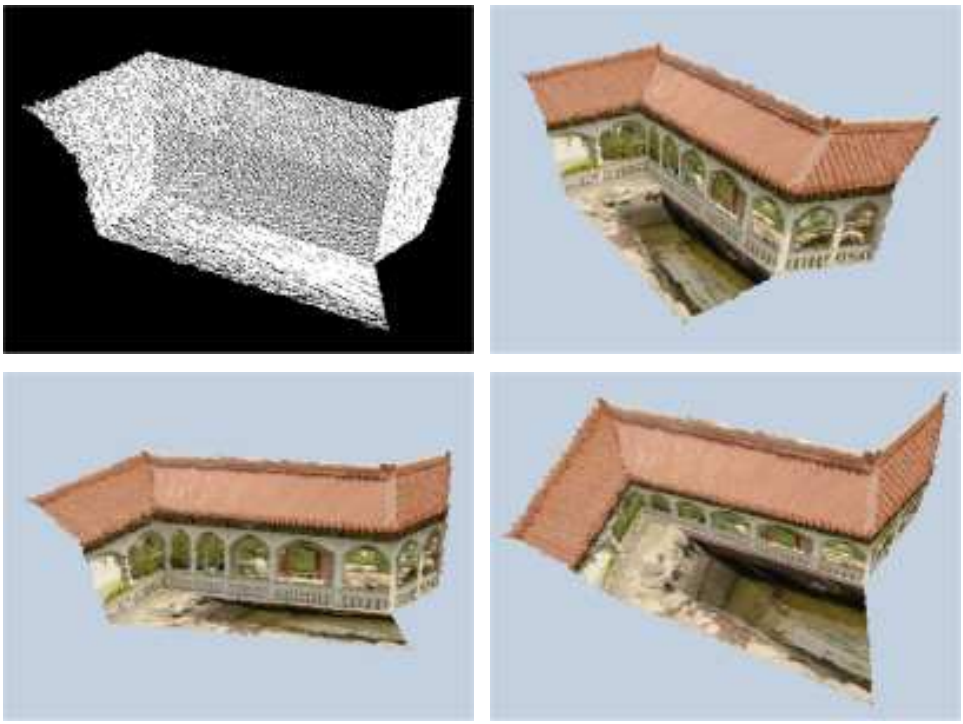


Fig. 13. Delaunay triangular surface patches and three perspective views of reconstruction.