



Constrained planar motion analysis by decomposition

Long Quan^{a,*}, Yichen Wei^a, Le Lu^{b,c}, Heung-Yeung Shum^b

^a*HKUST, Department of Computer Science, Clear Water Bay, Kowloon, Hong Kong SAR, China*

^b*Microsoft Research China, Beijing 100080, China*

^c*National Lab of Pattern Recognition, Chinese Academy of Sciences, Beijing 100080, China*

Received 5 November 2003; received in revised form 24 November 2003; accepted 25 November 2003

Abstract

General SFM methods give poor results for images captured by constrained motions such as planar motion. In this paper, we propose new SFM algorithms for images captured under a common but constrained planar motion: the image plane is perpendicular to the motion plane. We show that a 2D image captured under such constrained planar motion can be decoupled into two 1D images: one 1D projective and one 1D affine. We then introduce the 1D affine camera model for completing 1D camera models. Next, we describe new subspace reconstruction methods, and apply these methods to the images captured by concentric mosaics, which undergo a special case of constrained planar motion. Finally, we demonstrate both in theory and experiments the advantage of the decomposition method over the general SFM methods by incorporating the constrained motion into the earliest stage of motion analysis.

© 2003 Elsevier B.V. All rights reserved.

Keywords: SFM; planar motion; 1D camera; Vision geometry; Image-based rendering

1. Introduction

In this paper, we investigate the relationship between planar motion and 1D cameras, and study its application for concentric mosaics (CM). A new SFM algorithm is proposed for images captured under *constrained* planar motion. A planar motion is called constrained, if the orientation of the camera is known, in particular, the image plane is perpendicular to the motion plane. We first show that the geometry under such constrained planar motion becomes greatly simplified by decomposing the 2D image into two 1D images in a simple way: one 1D projective image and one 1D affine. The 3D reconstruction is, therefore, decomposed into the reconstruction in two subspaces: a 2D metric reconstruction and a 1D affine reconstruction. To complete the 1D camera model description, we also introduce 1D affine camera and study its geometric properties. Finally, we describe subspace reconstruction methods and demonstrate both in theory and experiments the advantage of the decomposition method over general SFM methods by incorporating the constrained motion into the earliest state of motion analysis.

A preliminary short version of this paper has been published for the ICCV conference [17].

Planar motion. A planar motion consists of a translation in a plane and a rotation about an axis perpendicular to that plane. This is the typical motion a vehicle moving on the ground undergoes. The study of planar motion has found its applications in such a system, e.g. autonomous guided vehicles, which are important components for factory automation [4]. It has been shown in Refs. [5,25] that an affine reconstruction is possible, provided that the internal parameters of the camera is constant. A more complete self-calibration method with constant internal parameters for planar motion has been proposed in Refs. [1,2]. It is shown that affine calibration is recovered uniquely, and metric calibration up to a two fold ambiguity.

The orientation of the camera moving under planar motion is constant with respect to the motion plane but unknown in general. If it is known, we say that the planar motion is constrained. Without loss of generality, we assume that the image plane is perpendicular to the motion plane. If not, the image can be warped to become perpendicular to the motion plane-by-plane homography transformation. The assumption is usually true in practice, e.g. for a camera carefully mounted on a vehicle moving on

* Corresponding author.

E-mail address: quan@cs.ust.hk (L. Quan).

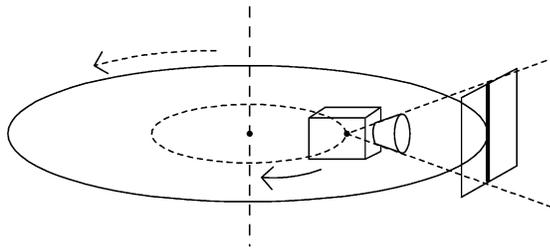


Fig. 1. The concentric mosaic acquisition set-up.

the ground. In particular, this is the case in a CM capturing device as shown in Fig. 1.

1D camera. A 1D projective camera maps a point in \mathcal{P}^2 (the projective space of 2D) onto a point in \mathcal{P}^1 with a 2×3 matrix, by analogy to a 2D projective camera that maps a point in \mathcal{P}^3 to a point in \mathcal{P}^2 . Much work has been done for both uncalibrated and calibrated 1D projective cameras [4, 16]. What is more interesting is that the usual 2D camera model could be related to this 1D camera model in various aspects. A linear algorithm is proposed in Ref. [16] for recovering 3D affine motion and shape from line correspondences for affine cameras, and those line directions are converted into 2D projective points using the 1D camera model. The relationship between planar motion and 1D cameras is revealed in Ref. [8]. It is shown that a 2D camera undergoing planar motion can be reduced to a 1D camera and a new method for self-calibrating a 2D projective camera is therefore developed.

Concentric mosaics. Recently, there has been much interest in computer vision and graphics in image-based rendering methods [13]. These methods generate new views of scenes from novel viewpoints, using a collection of images as the underlying scene representation. The major challenge is the very high dimensionality of plenoptic functions. Many simplifying assumptions that limit the underlying viewing space have been introduced: 5D plenoptic modelling [13], 4D Lightfield/Lumigraph [12, 10], 3D CM [19] and 2D panorama [15,6,21]. Among all these approaches, CM [19] is a good trade-off between the ease of acquisition and viewing space. The camera motion of images in CM is constrained to planar concentric circles, as shown in Fig. 1, and the image plane is parallel to the rotation axis, i.e. perpendicular to the plane. Therefore, the camera motion is a special case of constrained planar motion.

As the rendering using CM is performed by ray interpolation, like Lightfield, based on constant depth assumption without geometric information, it is necessary to compute the geometric structure to be capable of handling more complex scenes than constant-depth-type scenes and correcting the inherent vertical distortion problem [19]. Since, the CM capture device undergoes a constrained planar motion, this is a possible application of our new SFM algorithm. We apply our new methods to the images captured by CM and show preliminary result in the experiments.

Organisation. The paper is organized as follows. Section 2 describes the geometric analysis of images captured under constrained planar motion. Then, we describe the 1D affine camera model and 2D reconstruction from 1D images in Section 3. Experiment results are given in Section 4. Finally, the concluding remarks are given in Section 5. Throughout the paper, vectors are denoted in lower case boldface, matrices and tensors in upper case boldface. Scalars are any plain letters or lower case Greek.

2. Geometry of images under constrained planar motion

A general approach described in Refs. [1,5,24,25] consists of first computing a projective structure, then extracting fixed entities by planar motion or by assuming constant internal camera parameters. We follow a different approach [8] in which the 2D image under planar motion is reduced to the 1D trifocal line image. The 1D cameras defined this way are *virtual* and almost no physical points live in these subspaces. The virtual points on the trifocal plane could be simply created by projecting the 2D image points onto the trifocal line [8].

We take this decomposition principle further by introducing a complete decoupling of the 2D image into two complementary 1D images, one on the trifocal (motion) line and the other on the pencil of epipolar lines. The 3D space is accordingly decomposed into two orthogonal subspaces, one of 2D represented by the trifocal plane and another of 1D by the pencil of epipolar planes.

The second 1D camera is imaging the pencil of epipolar planes in space. It is in fact a sort of dual 1D imaging, as the ambient space elements are now dual elements, i.e. planes. Even more, this dual space is only of 1D, so the projection is described by a 2×2 1D homography from \mathcal{P}^1 to \mathcal{P}^1 instead of a 2×3 matrix from \mathcal{P}^2 to \mathcal{P}^1 . This is similar to 2D homography description of planar scenes by 2D cameras. Any image line intersecting the epipolar pencil produces a 1D projective image of the epipolar pencils. This decoupling of 2D image into 1D images is only possible, provided that the trifocal tensor or fundamental matrices have been estimated, this make the practical implementation of the decomposition more sensitive to the prior geometric computation.

Now, we apply this decomposition to the images captured under constrained planar motion. When the image plane is perpendicular to the motion plane and parallel to the rotation axes, the vanishing point of the rotation axes is therefore, the point at infinity of the vertical direction and the image of the motion plane, the trifocal line, is horizontal in the image space. This simplifies the decomposition of the 2D image significantly. The projection of a 2D image point onto the trifocal line is its horizontal coordinate in the image. With the vanishing point of rotation axes at vertical infinity,

the second 1D projection also gets simplified, as the vertical 1D homography now becomes an affine transformation. The vertical 1D camera is therefore, a kind of affine camera instead of projective. This motivates the definition and analysis of *1D affine camera* in Section 3.1.

Now let us introduce the space Euclidean coordinate frame such that xz plane is the ground motion plane and the y -axis is orthogonal to the ground plane. The horizontal and vertical pixel coordinates are u and v . The 2D camera matrix \mathbf{P} relates the 3D space point \mathbf{x} and 2D image point \mathbf{u} by $\lambda\mathbf{u} = \mathbf{P}\mathbf{x}$. The constrained planar motion preserves the point at infinity in y -axis direction and the ground plane $y = 0$ is imaged into the horizon line $v = 0$. The camera moving this way has the matrix of the form [25]

$$\begin{pmatrix} a & 0 & b & c \\ 0 & d & 0 & 0 \\ e & 0 & f & g \end{pmatrix} \quad (1)$$

Projecting 3D points with coordinates $(x, y, z, t)^T$ into two orthogonal subspaces of 2D with coordinates $(x, 0, z, t)^T$ and of 1D with coordinates $(0, y, 0, t)^T$. Working with points in these subspaces, the 2D camera matrix (Eq. (2)) decomposes to a 1D projective camera from \mathcal{P}^2 to \mathcal{P}^1

$$\lambda \begin{pmatrix} u \\ 1 \end{pmatrix} = \begin{pmatrix} a & b & c \\ e & f & g \end{pmatrix} \begin{pmatrix} x \\ z \\ t \end{pmatrix} \quad (2)$$

and a 1D affine camera from \mathcal{P}^1 to \mathcal{P}^1

$$\mu \begin{pmatrix} v \\ 1 \end{pmatrix} = \begin{pmatrix} d & 0 \\ 0 & g \end{pmatrix} \begin{pmatrix} y \\ t \end{pmatrix} \quad (3)$$

When these subspaces have been reconstructed, the 3D space point can be recovered by properly combining the subspace coordinates. The 1D subspace reconstruction is affine, its coordinate representation can be chosen as $\mu(0, 1, 0, 0)^T$. As the two subspaces are orthogonal, the corresponding 3D space point can be obtained by linearly combining $\lambda(x, 0, z, 1)^T + \mu(0, 1, 0, 0)^T$. The scale λ/μ can be fixed by imposing that $\lambda(x, 0, z, 1)^T + \mu(0, 1, 0, 0)^T = (\lambda x, \mu, \lambda z, \lambda)^T = (x, y, z, t)^T$. Thus, $\mu/\lambda = yt$.

Alternatively, this decoupling schema may be viewed as two 1D cameras directly from 3D spaces by

$$\begin{pmatrix} a & 0 & b & c \\ e & 0 & f & g \end{pmatrix} \text{ and } \begin{pmatrix} 0 & d & 0 & 0 \\ 0 & 0 & 0 & g \end{pmatrix}$$

but these 3D to 1D cameras being singular do not bring any new insight.

In summary, the above analysis gives a simplified SFM scheme for images captured under constrained planar motion

- decoupling the original 2D images into two 1D images by $(u, v)^T \mapsto u$ and $(u, v)^T \mapsto v$;

- reconstructing 2D points $\lambda(x, z, 1)$ from the corresponding u -coordinates in multiple views and 1D points (y, t) from the corresponding v -coordinates in multiple views;
- reconstructing 3D points by spanning the subspaces $t(x, 0, z, 1)^T + y(0, 1, 0, 0)^T$.

The 2D reconstruction methods from multiple images will be detailed in Section 3. The 1D reconstruction from a single image is straightforward, it just reads-off 1D affine coordinates $(y, 1)$ from v as $y = \gamma v$.

Another way to reconstruct 3D points is by extension of 2D Euclidean reconstruction (up to a global scale) (X, Z) of the xz -plane. With respect to the reference camera represented by $(\mathbf{I}_{3 \times 3}, \mathbf{0})$, $Y = (v - v_0)Z/f$ from

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

Notice that the 1D image decomposition method for the above-simplified SFM algorithm does not need any prior geometric computation unlike the general planar motion case. The motion constraint has been incorporated into the earliest stage of reconstruction. This is a key advantage over the general SFM methods. Another advantage is that, as the 1D trifocal line image is not physical, projecting a 2D image point to it is very simple, much more virtual points on this 1D image can be obtained. This makes the reconstruction more numerically stable.

3. 2D reconstruction from 1D images

After reducing 2D images to 1D images, the problem of 2D reconstruction from multiple 1D images will be addressed in this section. We first introduce and analyse the 1D affine camera model in Section 3.1 to complete the 1D camera model description, and then describe a reconstruction algorithm based on this concept by factorisation in Section 3.2. A complete Euclidean reconstruction algorithm from projective images based on the previous results in Refs. [4,16] is described for calibrated 1D projective cameras in Section 3.3. Since, the motion of images captured by CM is more constrained, and is actually circular motion, a further adoption of the reconstruction algorithm is described simply in Section 3.4.

3.1. Analysis of 1D affine camera

The 1D affine camera could be introduced by analogy to a 2D affine camera introduced by Mundy and Zisserman [14] as an imaging device from 2D to 1D, which preserves the affine properties characterized by the points at infinity. More importantly, it is motivated by the practical weak perspective geometry of 1D projective camera for CM images to describe a common degeneracy

of the 1D projective camera either when the viewing field is narrow or the scene is shallow compared to the average distance from the camera. This is of particular importance for CM images as many CM sequences have been captured to facilitate constant-depth rendering.

A general 1D projective camera maps a point $\mathbf{x} = (x_1, x_2, x_3)^T$ in \mathcal{P}^2 to a point $\mathbf{u} = (u_1, u_2)^T$ in \mathcal{P}^1 by a general 2×3 matrix \mathbf{M} as $\lambda \mathbf{u} = \mathbf{M}_{2 \times 3} \mathbf{x}$. If a 2D point \mathbf{x} is on the line at infinity, identified as $x_3 = 0$, to preserve the affine property, it should be mapped onto the point \mathbf{u} at infinity on the image line, identified as $u_2 = 0$. This reduces the 1D projective camera to the 1D affine camera in the following form

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ 0 & 0 & p_{23} \end{pmatrix} = \begin{pmatrix} \mathbf{m}_{1 \times 2} & t \\ \mathbf{0}_{1 \times 2} & 1 \end{pmatrix} \quad (4)$$

The 1D affine camera maps the finite points $(x, y, 1)^T$ onto finite image points $(u, 1)^T$ with

$$\lambda u = \mathbf{m}_{1 \times 2} \begin{pmatrix} x \\ y \end{pmatrix} + t$$

If we further use relative coordinates with respect to a given reference point, for instance, the centroid of the point set, $\Delta u = u - u_r$ in \mathbb{R}^1 and $(\Delta x, \Delta y)^T = (x - x_r, y - y_r)^T$ in \mathbb{R}^2 . The translation component t is cancelled and the projection for finite points in relative coordinates therefore becomes

$$\lambda \Delta u = \mathbf{m}_{1 \times 2} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \quad (5)$$

This is the basic 1D affine camera projection. We now examine the geometric constraints available for points seen in multiple views similar to the 2D camera case [7, 11, 18, 20, 23].

Assuming constant internal camera parameters, three views of the point $\Delta \mathbf{x}$ can be written as follows

$$\begin{cases} \lambda \Delta u = \mathbf{m} \Delta \mathbf{x}, \\ \lambda \Delta u' = \mathbf{m}' \Delta \mathbf{x}, \\ \lambda \Delta u'' = \mathbf{m}'' \Delta \mathbf{x} \end{cases} \quad (6)$$

or written in matrix form as

$$\begin{pmatrix} \mathbf{m} & \Delta u \\ \mathbf{m}' & \Delta u' \\ \mathbf{m}'' & \Delta u'' \end{pmatrix} \begin{pmatrix} \Delta \mathbf{x} \\ -\lambda \end{pmatrix} = \mathbf{0}$$

The vector $(\Delta \mathbf{x}, -\lambda)^T$ is not vanishing, so

$$\begin{vmatrix} \mathbf{m} & \Delta u \\ \mathbf{m}' & \Delta u' \\ \mathbf{m}'' & \Delta u'' \end{vmatrix} = \begin{vmatrix} \mathbf{a} & \mathbf{b} & \Delta u' \\ & & \Delta u'' \end{vmatrix} = 0$$

The expansion of this determinant produces a linear constraint for the three 1D affine views

$$(\mathbf{a} \times \mathbf{b})^T \begin{pmatrix} \Delta u \\ \Delta u' \\ \Delta u'' \end{pmatrix} = 0$$

or simply as

$$a \Delta u + b \Delta u' + c \Delta u'' = 0 \quad (7)$$

So the geometry of three uncalibrated 1D affine views is completely characterized by this linear constraint represented by the homogeneous 3-vector $(a, b, c)^T$ which has only 2 d.o.f. With at least three point correspondences in three views, two relative points plus the reference point, the vector $(a, b, c)^T$ could be estimated linearly.

3.2. 2D reconstruction from 1D affine camera by factorisation

The above three-view linear constraint directly encodes the relative motion parameters. For reconstruction, by analogy to the factorisation method [22] for 2D affine cameras, we could proceed in the same way by stacking p points in n images to create the measurement matrix as

$$\begin{pmatrix} \Delta u_1^1 & \cdots & \Delta u_p^1 \\ \Delta u_1^2 & & \Delta u_p^2 \\ \vdots & & \\ \Delta u_1^n & & \Delta u_p^n \end{pmatrix} = \begin{pmatrix} \mathbf{m}^1 \\ \vdots \\ \mathbf{m}^n \end{pmatrix} (\Delta \mathbf{x}_1 \quad \cdots \quad \Delta \mathbf{x}_p)$$

or in compact form as $\mathbf{U}_{n \times p} = \mathbf{M}_{n \times 2} \mathbf{S}_{2 \times p}$

The rank of the measurement matrix cannot exceed two. By taking SVD for the measurement matrix, keeping the two largest singular values and forcing all others to be zero, we obtain the best rank-two approximation of the measurement matrix $\mathbf{U} = \mathbf{M} \mathbf{S}$. However, \mathbf{M} and \mathbf{S} will in general fail to correspond to the true camera matrices and point reconstruction, since for any non-singular 2×2 affine matrix \mathbf{A} , $\mathbf{U} = \mathbf{M} \mathbf{S} = (\mathbf{M} \mathbf{A})(\mathbf{A}^{-1} \mathbf{S}) = \mathbf{M}' \mathbf{S}'$, so \mathbf{M}' and \mathbf{S}' also correspond to a possible reconstruction. In other words, the reconstruction is up to a plane affine transformation.

We need to look for an \mathbf{A} such that $\mathbf{m}^i \mathbf{A} = c \mathbf{R}_{1 \times 2}^i$, where $\mathbf{R}_{1 \times 2}^i$ is a row of a rotation matrix, the motion of i th camera and c is an independent scaling factor that each 1D affine camera may have, actually the assumed constant focal length. So that the metric constraint for an Euclidean reconstruction is

$$\mathbf{m}^i \mathbf{A} \mathbf{A}^T \mathbf{m}^{iT} = 1 \quad (8)$$

where the global scaling factor of \mathbf{A} is assumed to be fixed, $|\mathbf{A}| = 1/c$. Such an affine matrix \mathbf{A} has 3 d.o.f, and each camera matrix \mathbf{m}^i from the affine reconstruction will place one constraint on the entries of \mathbf{A} via Eq. (8). There will

generally be far more than three views used in practice, therefore, the computation of \mathbf{A} is a simple over-determined data fitting problem which, though nonlinear, can be solved efficiently and reliably.

This reconstruction might be sufficient, and if not, it can be served as an initial solution for a nonlinear optimisation. The factorisation method requires 2D points to be visible in all 1D images. For missing points, they can be handled using the linear three-view constraint Eq. (7) developed in Section 3.1.

3.3. 2D reconstruction from 1D projective camera

The affine camera model nicely approximate the geometry of projective camera when the viewing field is narrow or the scene is shallow compared to the average distance from the camera. When these assumptions do not hold, reconstruction algorithm using projective cameras is favoured instead of that using affine camera model. We describe a complete reconstruction method from three 1D projective cameras, which assembles the recent results in Refs. [3,4,16,8] for 1D cameras.

- *Computing uncalibrated 1D trifocal tensor.* The geometry of three 1D images is completely characterized by the 1D trifocal tensor \mathbf{T}_{ijk} [8]. It minimally parameterises the three uncalibrated images and can be estimated linearly with at least seven point correspondences.
- *Self-calibration of the 1D camera.* The 1D camera could be self-calibrated for constant calibration parameters via 1D trifocal tensor [8]. The knowledge of the internal parameters of a 1D camera is equivalent to that of the image points \mathbf{i} and \mathbf{j} , which are the image of the circular points in the motion plane in space, the horizontal plane, \mathcal{P}^2 [1,8]. This pair of conjugate complex points can be uniquely determined by solving the cubic equation: $T_{111}x^3 + (T_{211} + T_{112} + T_{121})x^2 + (T_{212} + T_{221} + T_{122})x + T_{222} = 0$. They are the conjugate complex roots of the cubic equation. The real part of the ratio of the projective

coordinates of the image of the circular point \mathbf{i} is the position of the principal point u_0 and the imaginary part is the focal length α .

- *Computing calibrated 1D trifocal tensor.* The internal parameters of the camera can either be self-calibrated as described in Section 3.2 or given by off-line calibration, then we come to the case of calibrated 1D camera. To handle calibrated geometry properly, the image coordinates could be first normalized by applying $\mathbf{K}^{-1}(u_i, 1)^T$ to get $(x_i, 1)^T$. To see what happens for the calibrated trifocal tensor, it suffices to notice that knowing the internal parameters of a 1D camera is equivalent to knowing two points, the pair of circular points! Substituting the circular points $(\pm i, 1)$ into the trilinear constraint gives the two following scalar constraints [4] $T_{122} + T_{212} + T_{221} - T_{111} = 0$ and $T_{112} + T_{121} + T_{211} - T_{222} = 0$.

The 1D trifocal tensor can now be linearly re-estimated by taking into account of these constraints. Substituting T_{111} and T_{222} back into the original trilinear constraint equation gives constrained trilinear constraint.

- *Recovering external projection matrix.* After the calibrated 1D trifocal tensor has been computed, the tensor components can be converted into external parameters of the cameras. The external projection matrices for three views can be written as

$$(\mathbf{R}(\theta), \mathbf{t}_{2 \times 1}), (\mathbf{R}(\theta'), \mathbf{t}'_{2 \times 1}) \text{ and } (\mathbf{R}(\theta''), \mathbf{t}''_{2 \times 1})$$

Since, the world coordinate frame can be chosen arbitrarily, it can be chosen in the way such that $\theta = 0$ and $\mathbf{t} = \mathbf{0}$. Therefore, there are totally five d.o.f for external parameters, two for θ', θ'' , and three for $\mathbf{t}', \mathbf{t}''$ (4 – the global scale). There are also five non-homogeneous tensor components for the calibrated trifocal tensor (eight entries minus one global scale and two constraints discussed earlier). The conversion from the trifocal components to external camera parameters can be solved algebraically, but up to a two-way ambiguity [4,16].

- *Reconstructing 2D point coordinates.* Each 2D point can be reconstructed by solving linear equations provided by $\lambda(u, 1)^T = \mathbf{M}_{2 \times 3}(x, y, 1)^T$.

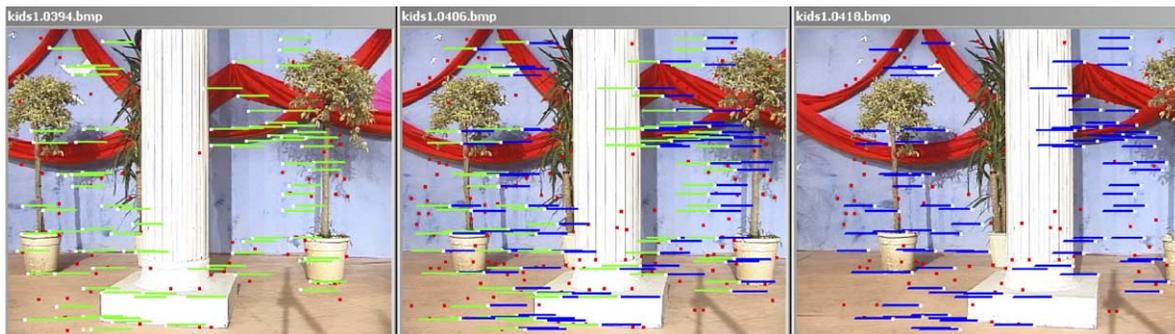


Fig. 2. Point correspondences in a triplet image from KIDS sequence.

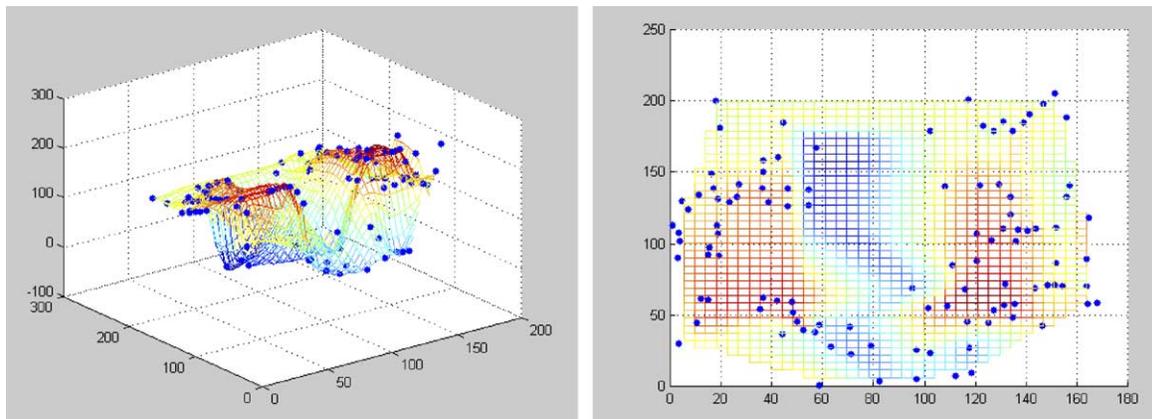


Fig. 3. 3D Affine reconstruction by 2D factorisation for KIDS sequence and projection onto (x, z) plane.

- *Nonlinear optimisation.* Finally, the reconstruction can be improved by a nonlinear optimisation method, i.e. bundle adjustment.

3.4. 2D reconstruction from calibrated 1D projective cameras under circular motion

One application of our new SFM algorithm is to apply it to CM rendering process to alleviate the problems caused by constant-depth assumption. However, the images captured by CM are not only under constrained planar motion, but have stronger motion constraint: a circular motion [9]. All the camera centers are located on a circle on the motion plane, and all the rotations between each pair of cameras are around the same axis passing through the circle center. The calibrated 2×3 projection matrices for a triplet of views can be parameterised as

$$(\mathbf{R}(\theta), \mathbf{t}), (\mathbf{R}(\theta'), \mathbf{t}), \text{ and } (\mathbf{R}(\theta''), \mathbf{t})$$

The associated trifocal tensor has also two additional constraints than the calibrated 1D trifocal tensor. One is that $T_{222} = 0$ if we choose $t_y = 0$ without loss of generality. The other has more complicated expression. This particular

parameterisation also suggests a more efficient bundle-like nonlinear optimisation.

4. Experimental results

Experiments on analysing image data by the new SFM algorithm have been carried out. In this section, we show some preliminary results based on tracking results of points of interest from triplets of original images captured by the concentric mosaic set-up in our lab. The matching points are obtained using the algorithm presented in Ref. [26].

KIDS sequence. For the KIDS triplet shown in Fig. 2, there are 159 and 107 match candidates in the first and second pairs. We obtain 89 final match triplets. The 3D affine reconstruction using standard 2D factorisation method is shown in Fig. 3. The horizontal plane is referenced by coordinates (x, z) , so the z -coordinate gives the depth and the y -coordinate the height. The 2D affine and Euclidean reconstruction using our 1D factorisation method is shown in Fig. 4.

In Fig. 5, two columns and the background wall are drawn over the reconstructed plane to illustrate

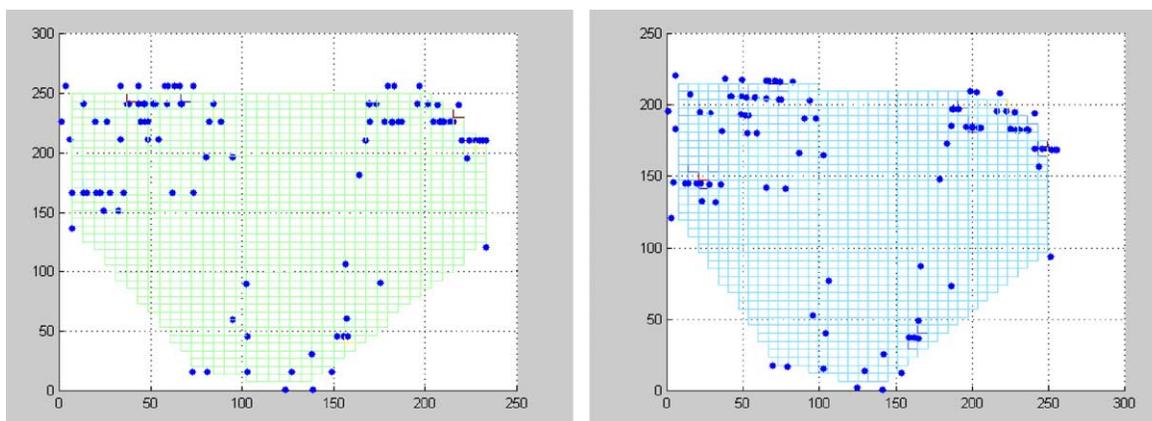


Fig. 4. 2D Affine and Euclidean reconstruction by 1D factorisation for KIDS sequence.

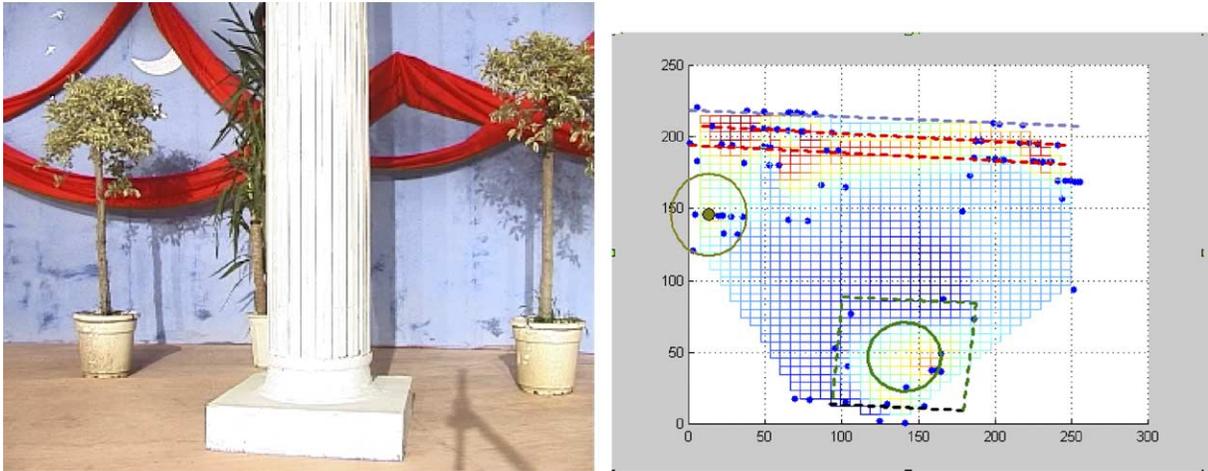


Fig. 5. One original image and the reconstructed plane by merging two triplets of KIDS sequence with manual drawing for illustration.

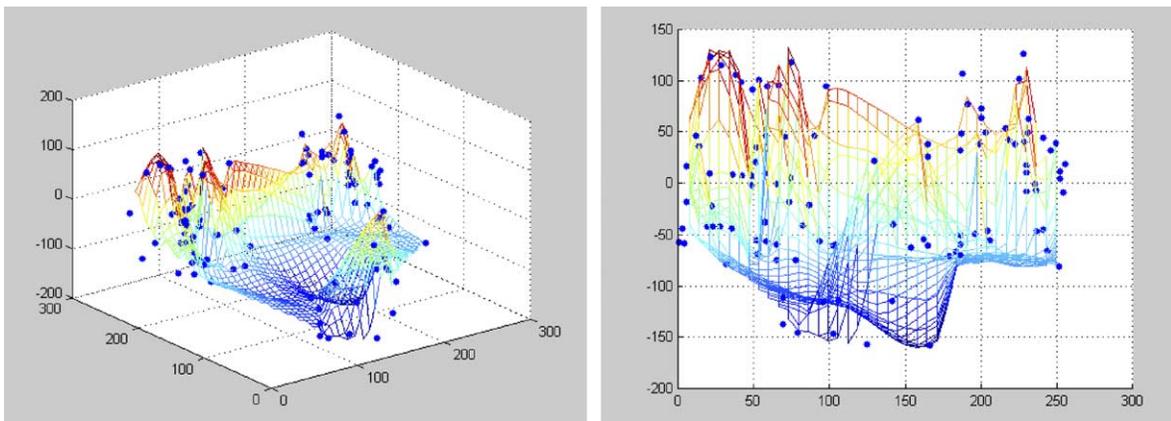


Fig. 6. The 3D Euclidean reconstruction by extending the 2D reconstruction for KIDS triplet: 3D and projection onto xy plane.

the reconstruction quality. If the camera is calibrated off-line, i.e. the aspect ratio and principal point is known, with the known depth z , the height y can be obtained by rescaling the vertical coordinate by the depth and 3D Euclidean

reconstruction is possible. Fig. 6 shows the 3D Euclidean reconstruction by extending the 2D reconstruction.

TOY sequence. For the TOY triplet shown in Fig. 7, there are 151 and 126 match candidates in the first and

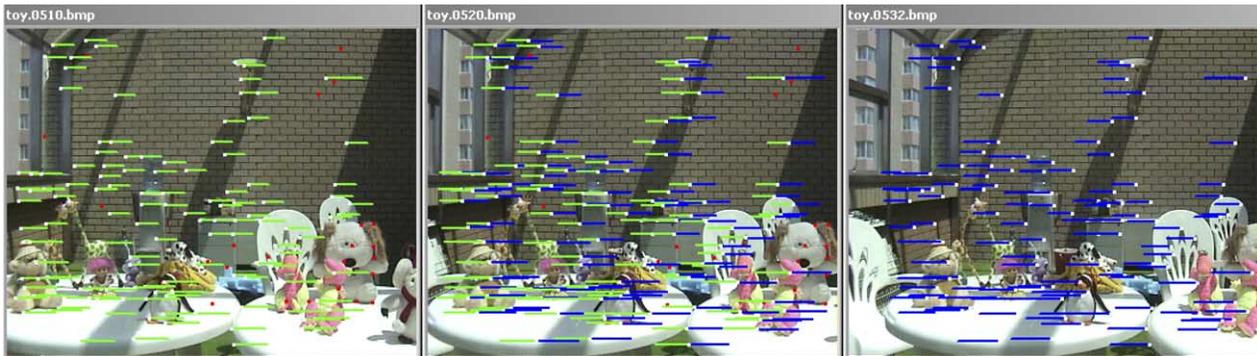


Fig. 7. Point correspondences in a triplet from TOY sequence.

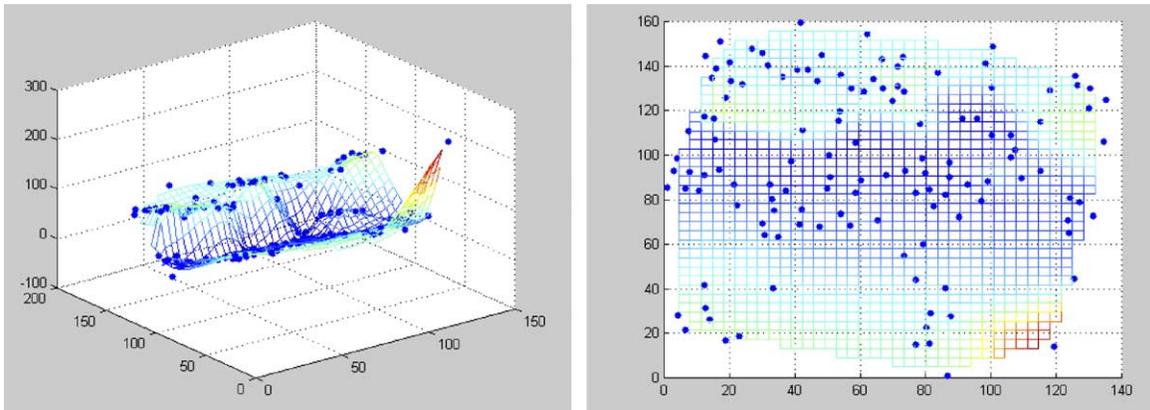


Fig. 8. 3D Reconstruction by 2D factorisation method for TOY triplet and projection onto (x, z) plane.

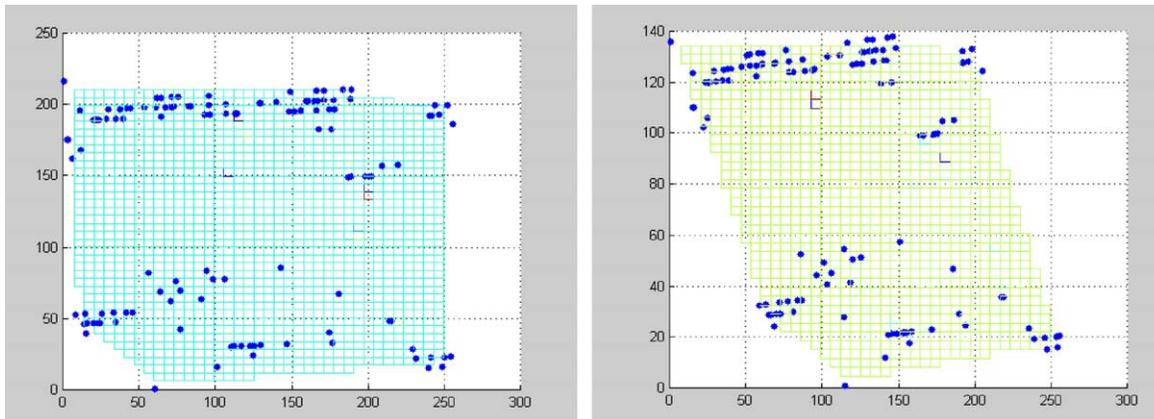


Fig. 9. 2D Affine and Euclidean reconstruction by 1D factorisation for TOY triplet.

second pairs. 76 final corresponding triplets are obtained. The 3D affine reconstruction using standard 2D factorisation method is shown in Fig. 8. The 2D affine and Euclidean reconstruction using our 1D factorisation method are shown in Fig. 9. We can notice the superior

reconstruction quality by 1D factorisation method over the 2D factorisation. The 3D Euclidean reconstruction by extending the 2D Euclidean reconstruction using off-line calibrated internal parameters are shown in Fig. 10.

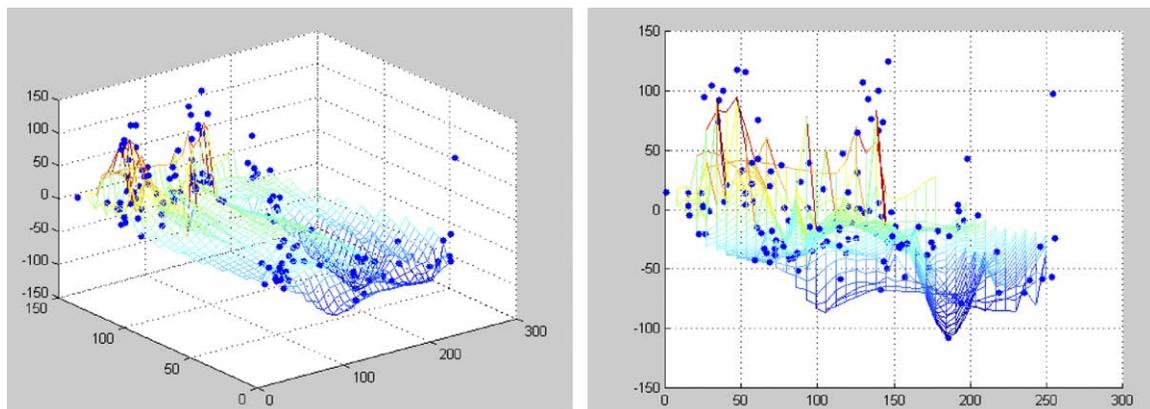


Fig. 10. The 3D Euclidean reconstruction by extending the 2D reconstruction for TOY triplet: 3D and projection onto xy plane.

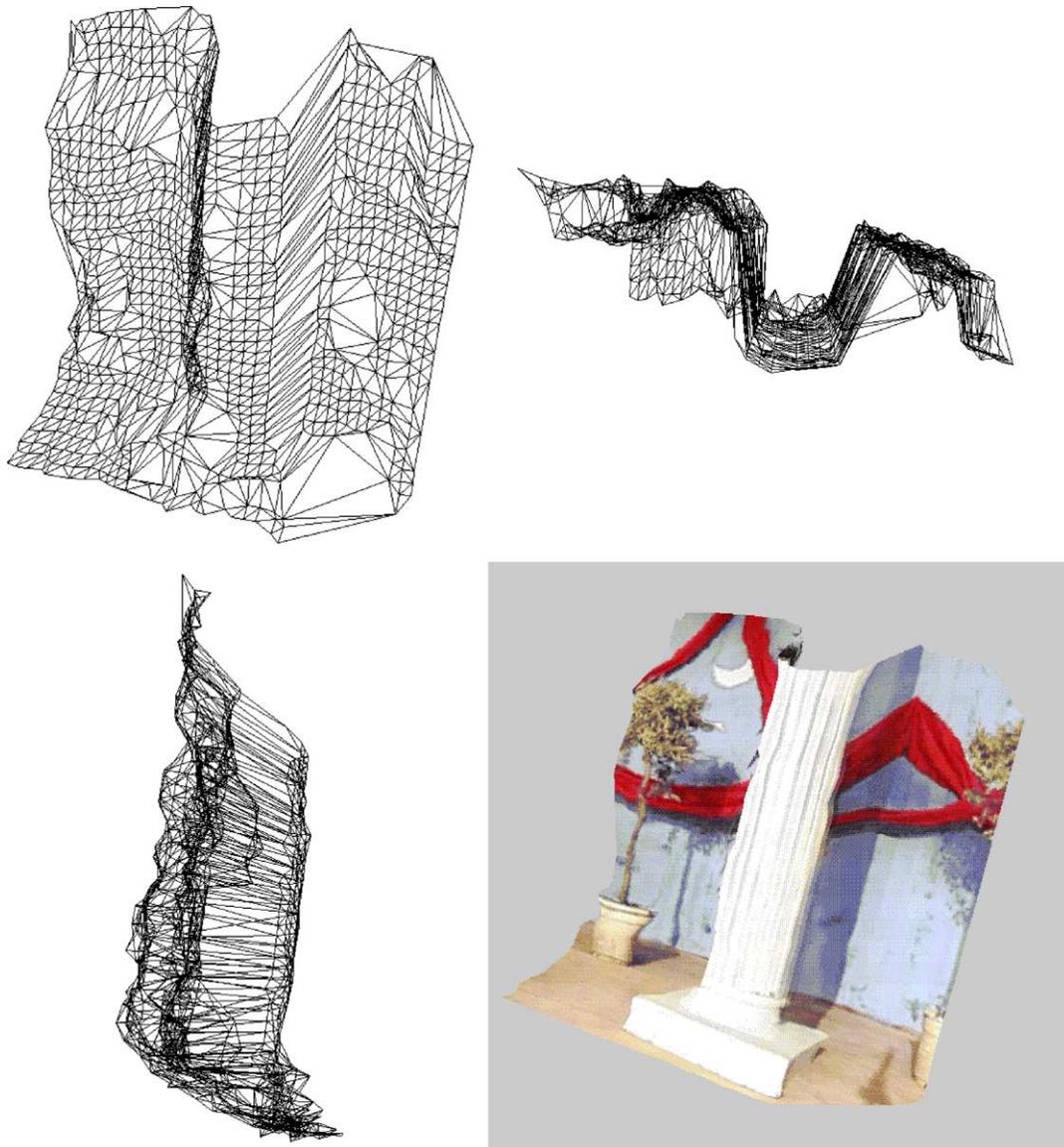


Fig. 11. 3D reconstruction of KIDS triplet in VRML. On the top are a mesh model and a top view of the mesh. On the bottom are a side view of the mesh and textured model.

The final 3D VRML model shown in Figs. 11 and 12 are reconstructed from re-sampled dense matching by the extension of Euclidean 2D reconstruction. The 3D reconstruction quality is sufficient for image-based-rendering purpose.

5. Conclusion

This paper analyses the geometry and proposes a new SFM algorithm under constrained planar motion. We have shown that the 2D image captured under constrained planar motion can be decomposed into two 1D images in an easy manner: one captured by a horizontal 1D projective camera and one captured by a vertical affine camera. The 3D

reconstruction is, therefore decomposed into the reconstruction in the subspaces, i.e. a 2D metric reconstruction combined with a 1D affine reconstruction. We have introduced the new concept of 1D affine camera and 1D factorisation method for 2D reconstruction in the horizontal subspace. The key advantage of the new algorithm is that the prior motion information has been integrated into the system. The 2D/1D image conversion does not need any geometric estimation of fundamental matrices or trifocal tensors. Another advantage of virtual 1D camera over a physical 1D camera is that it sees through the 2D scene, so it may have more virtual points independent of occlusions in different heights. Preliminary results have been demonstrated both theoretical and practical advantages of the decomposition method used in this paper over the general

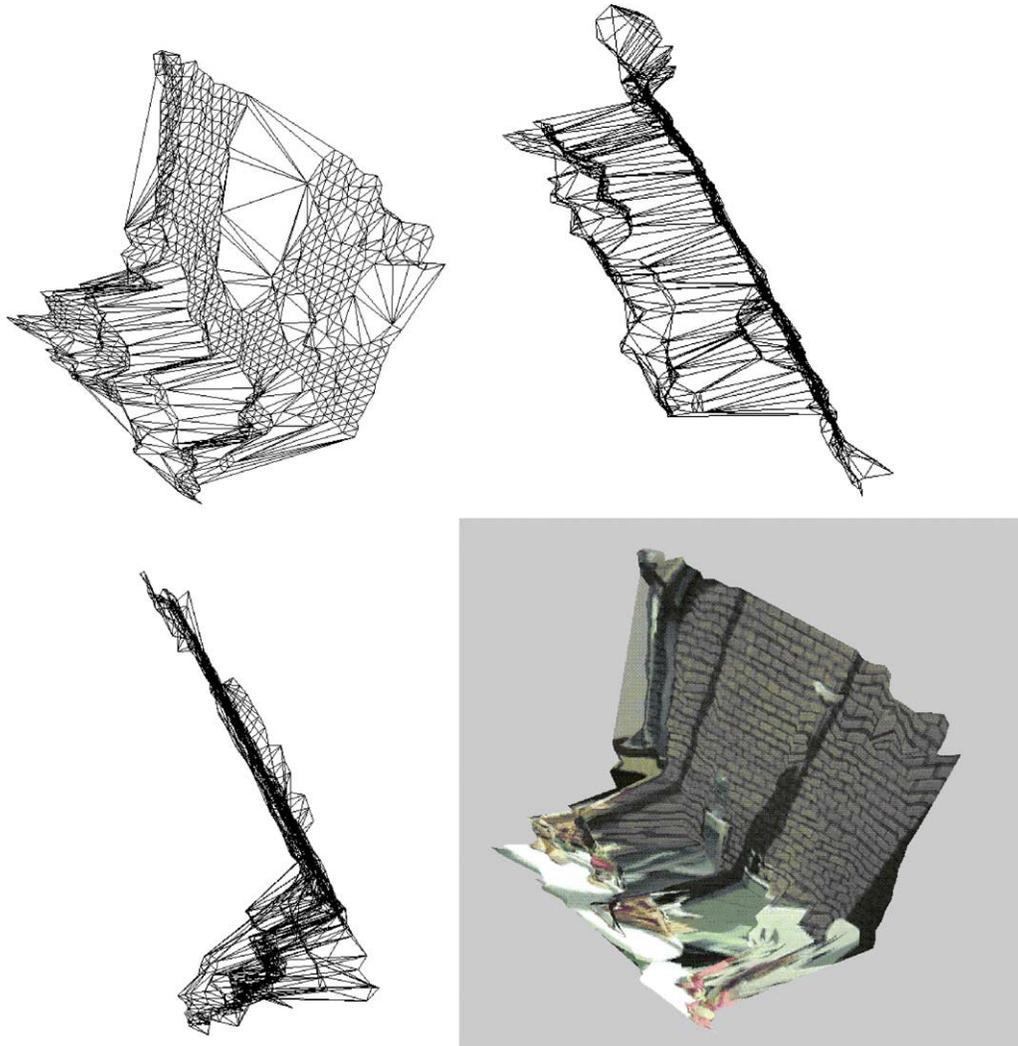


Fig. 12. 3D reconstruction of TOY triplet in VRML. On the top are a mesh model and a top view of the mesh. On the bottom are a side view of the mesh and textured model.

SFM methods which tend to be singular under the constrained motion model.

Acknowledgements

We would like to thank B. Triggs and P. Sturm for fruitful discussions. The work has also been partly supported by the Hong Kong RGC grant HKUST6188/02E.

References

- [1] M. Armstrong, A. Zisserman, R. Hartley, Self-calibration from image triplets, in: B. Buxton, R. Cipolla (Eds.), Proceedings of the Fourth European Conference on Computer Vision, Cambridge, England, Lecture Notes in Computer Science, vol. 1064, Springer, Berlin, April 1996, pp. 3–16.
- [2] M.N. Armstrong, Self-Calibration from Image Sequences, PhD Thesis, Department of Engineering Science, University of Oxford, UK, December 1996.
- [3] K. Åström, Invariance Methods for Points, Curves and Surfaces in Computational Vision, PhD Thesis, Lund University, 1996.
- [4] K. Åström, M. Oskarsson, Solutions and ambiguities of the structure and motion problem for 1d retinal vision, *Journal of Mathematical Imaging and Vision* 12 (2000) 121–135.
- [5] P.A. Beardsley, A. Zisserman, Affine calibration of mobile vehicles, in: R. Mohr, C. Wu (Eds.), Europe–China Workshop on Geometrical Modelling and Invariants for Computer Vision, Xian, China, Xidan University Press, 1995, pp. 214–221.
- [6] S.E. Chen, Quicktime VR—an image-based approach to virtual environment navigation, in: SIGGRAPH, Los Angeles, USA, 1995, pp. 29–38.
- [7] O. Faugeras, B. Mourrain, About the correspondences of points between n images, in: Workshop on Representation of Visual Scenes, Cambridge, Massachusetts, USA, 1995, pp. 37–44.
- [8] O. Faugeras, L. Quan, P. Sturm, Self-calibration of a 1d projective camera and its application to the self-calibration of a 2d projective camera, in: Proceedings of the Fifth European Conference on Computer Vision, Freiburg, Germany, June 1998, pp. 36–52.
- [9] A.W. Fitzgibbon, G. Cross, A. Zisserman, Automatic 3d model construction for turn-table sequences, in: 3D Structure from Multiple Images of Large-scale Environments SMILE'98, Springer, Berlin, 1998, pp. 154–169.

- [10] S.J. Gortler, R. Grzeszczuk, R. Szeliski, M. Cohen, The lumigraph, in: Proceedings of SIGGRAPH, New Orleans, LA, 1996, pp. 43–54.
- [11] R.I. Hartley, A. linear, A linear method for reconstruction from lines and points, in: E. Grimson (Ed.), Proceedings of the Fifth International Conference on Computer Vision, Cambridge, Massachusetts, USA, IEEE, IEEE Computer Society Press, Silver Spring, MD, June 1995, p. 887.
- [12] M. Levoy, P. Hanrahan, Light field rendering, in: Proceedings of SIGGRAPH, New Orleans, LA (1996) 31–42.
- [13] L. McMillan, G. Bishop, Plenoptic modeling: an image-based rendering system, in: SIGGRAPH, Los Angeles, USA, 1995, pp. 39–46.
- [14] J.L. Mundy, A. Zisserman, Projective geometry for machine vision, in: J.L. Mundy, A. Zisserman (Eds.), Geometric Invariance in Computer Vision, The MIT Press, Cambridge, MA, USA, 1992, pp. 463–519, (Chapter 23).
- [15] S. Peleg, J. Herman, Panoramic mosaics by manifold projection, in: Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA, 1997, pp. 338–343.
- [16] L. Quan, T. Kanade, Affine structure from line correspondences with uncalibrated affine cameras, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (8) (August 1997) 834–845.
- [17] L. Quan, L. Lu, H.Y. Shum, M. Lhuillier, Concentric mosaic(s), planar motion and 1d cameras, in: Proceedings of the Eighth International Conference on Computer Vision, Vancouver, Canada, vol. 2, 2001, pp. 193–200.
- [18] A. Shashua, Algebraic functions for recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17 (8) (August 1995) 779–789.
- [19] H.Y. Shum, L.W. He, Rendering with concentric mosaics, in: SIGGRAPH 2000, New Orleans, USA, 1999, pp. 299–306.
- [20] M. Spetsakis, J. Aloimonos, A unified theory of structure from motion, in: Proceedings of DARPA Image Understanding Workshop, 1990, pp. 271–283.
- [21] R. Szeliski, H.-Y. Shum, Creating full view panoramic image mosaics and environment maps, in: Proceedings of SIGGRAPH, Los Angeles, CA, 1997, pp. 251–258.
- [22] C. Tomasi, T. Kanade, Factoring image sequences into shape and motion, in: Proceedings of the IEEE Workshop on Visual Motion, Princeton, New Jersey, Los Alamitos, California, USA, IEEE Computer Society Press, Silver Spring, MD, October 1991, pp. 21–28.
- [23] B. Triggs, Matching constraints and the joint image, in: E. Grimson (Ed.), Proceedings of the Fifth International Conference on Computer Vision, Cambridge, Massachusetts, USA, IEEE, IEEE Computer Society Press, Silver Spring, MD, June 1995, pp. 338–343.
- [24] B. Triggs, Plane + parallax, tensors and factorization, in: Proceedings of the Sixth European Conference on Computer Vision, Dublin, Ireland, Springer, Berlin, 2000, pp. 522–538.
- [25] C. Wiles, M. Brady, Ground plane motion camera models, in: B. Buxton, R. Cipolla (Eds.), Proceedings of the 4th European Conference on Computer Vision, Cambridge, England, European, volume 1065 of Lecture Notes in Computer Science, Springer, Berlin, 1996, pp. 238–247.
- [26] Z. Zhang, R. Deriche, O. Faugeras, Q.T. Luong, A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, *Rapport de recherche* 2273, INRIA, May 1994.