Concentric Mosaic(s), Planar Motion and 1D Cameras

Long QUAN[1] Le LU[3,2] Heung-Yeung SHUM[2] Maxime LHUILLIER[1]

[1]CNRS-INRIA, 655 avenue de l'Europe, 38330 Montbonnot, France
[2]Microsoft Research China, Beijing 100080, P.R. China
[3]National Lab of Pattern Recognition, Chinese Academy of Sciences, Beijing 100080, P.R. China

Abstract

General SFM methods give poor results for images captured by constrained motions such as planar motion of concentric mosaics (CM). In this paper, we propose new SFM algorithms for both images captured by CM and composite mosaic images from CM. We first introduce 1D affine camera model for completing 1D camera models. Then we show that a 2D image captured by CM can be decoupled into two 1D images: one 1D projective and one 1D affine; a composite mosaic image can by rebinned into a calibrated 1D panorama projective camera. Finally we describe subspace reconstruction methods and 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.

Key words: SFM, planar motion, 1D camera, vision geometry, image-based rendering.

1 Introduction

Image-Based Rendering and Concentric mosaics Recently there has been much interest in computer vision and graphics in image-based rendering methods [16]. They generate new views of scenes from novel viewpoints, using a collection of images as the underlying scene representation. When the sampling rate is sparse, this is very similar to classical computer vision based 3D reconstruction, either explicit [18, 21, 13] or implicit [14, 5] with texture mapping. When sampling is dense, a large amount of work [16, 15, 11] has been developed based on plenoptic functions. This models all sets of rays seen from all points, considering each image as a set of rays. The major challenge is the very high dimensionality of such plenoptic functions. Many simplifying assumptions that limit the underlying viewing space have been introduced: 5D plenoptic modeling [16], 4D Lightfield/Lumigraph [15, 11], 3D concentric mosaics [23] and 2D panorama [19, 7, 25]. Among all these approaches, concentric mosaics [23] is a good trade-off between the ease of acquisition and viewing space. The camera motion of CM is constrained to planar concentric circle, CMs are created by composing slit images taken at different locations along each circle. CMs index all imput image rays naturally in 3 parameters: radius, rotation angle and vertical elevation.

As concentric mosaics are rendered by ray interpolation, like Lightfield, based on constant depth without geometric correction, it is necessary to compute the geometric structure to be able to handle more complex than constant-depth-type scenes and correct the inherent vertical distortion [23].

Planar motion A camera moving in a plane and rotating about an axis perpendicular to that plane is called *planar motion*. This is the typical motion undergone by a vehicle moving on a plane. It has been shown in [6, 29] that an affine reconstruction is possible provided the constant internal parameters of the camera. A more complete self-calibration method with constant internal parameters from planar motion has been proposed in [1, 2].

1D camera A 1D projective camera maps a point in $\mathcal{P}^2$ (the projective space of dimension 2) 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 $\mathcal{P}^2$. Many work has been done for both uncalibrated and calibrated 1D projective cameras [4, 20]. What is more interesting is that the usual 2D camera model could be related to this 1D camera model in [20] for affine camera case and in [9] for planar motion case.

The first observation is that CM capture device undergoes a perfect planar motion. The geometry of CMs inherits that of planar motion.

For a complete decomposition of concentric mosaic images, we introduce a new 1D affine camera model.

We first show that a image captured by CM can be decoupled into two 1D cameras, one horizontal 1D projective camera and one vertical 1D affine camera. The scene depth can be recovered by the horizontal 1D projective camera while the height from the vertical 1D affine camera. Then we show that a composite mosaic image can be rebinned into a calibrated 1D panorama projective camera. The unifying thema is that the geometry of concentric mosaics reduces to the geometry of 1D camera. To complete 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 the general SFM algorithms by integrating the prior motion constraints into the estimation procedure.

The paper is organised as follows. Section 2 describes the geometric analysis of images captured by CM and Section 2 the geometric analysis of composite mosaic images. Then, we describe 1D affine camera model and 2D reconstruction from 1D images in Section 4. Experimental results are given in Section 5. Finally some concluding remarks are given in Section 6. 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 captured by CMs

The camera mounted on the CM capture system rotates around the fixed vertical axis. The original images (before mosaic composition) of concentric mosaics are therefore constrained by a planar motion. A general approach described in [29, 6, 1, 28] consists of first computing a projective structure, then extracting fixed entities by planar motion or by constant internal calibration. We follow a different approach [9] in which 2D image of a planar motion is reduced to the trifocal line image. We take this decomposition principle futher 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 dimension 2 represented by the trifocal plane and another of dimension 1 by the pencil of epipolar planes.

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 [9]. 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 dimension 1, 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 applying this decomposition to the images of the concentric mosaics, the planar motion of CM is parallel to the horizontal ground plane and the rotation axes are vertical. The vanishing point of the rotation axes is therefore the point at infinity of the vertical direction. This simplifies the projection of 2D image points onto the trifocal line: the horizontal u-coordinates of points makes up the 1D image line. 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 4.1.

Now let introduce the space Euclidean coordinate frame such that xz plane is the ground plane and the y-axis is normal to the ground plane. The horizontal and vertical pixel coordinates are u and v. The 2D camera matrix $\mathbf{P}$ relates them by $\lambda\mathbf{u} = \mathbf{Px}$. The constrained 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 [29]:

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

Projecting 3D points with coordinates $(x, y, z, t)^T$ into two orthogonal subspaces of dimension 2 with coordinates $(x, 0, z, t)^T$ and of dimension 1 with coordinates $(0, y, 0, t)^T$. Working with points in these subspaces gives

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

which is a 1D projective camera from $\mathcal{P}^2$ to $\mathcal{P}^1$; and

$$\mu \begin{pmatrix} v \\ 1 \end{pmatrix} = \begin{pmatrix} d & 0 \\ 0 & g \end{pmatrix} \begin{pmatrix} y \\ t \end{pmatrix}.$$

which is a 1D affine camera from $\mathcal{P}^2$ to $\mathcal{P}^1$.

Whence these subspaces have been reconstructed, the 3D space point can be recovered by linearly combining $\lambda(x, 0, z, 1)^T + \mu(0, 1, 0, 0)^T$ with proper scales λ/μ fixed by y/t.

Alternatively, this decoupling schema may be viewed as two 1D cameras directly from 3D spaces by $\left(\begin{smallmatrix} a & 0 & b & c \\ e & 0 & f & g \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 0 & d & 0 & 0 \\ 0 & 0 & 0 & g \end{smallmatrix}\right)$, but these 3D to 1D cameras being singular do not bring new insight.

In summary, this gives a simplified SFM algorithm for images captured by CMs.

- decoupling the original 2D images into two 1D images by $(u, v)^T \mapsto u$ and $(u, v)^T \mapsto v$;
- reconstructing 2D points (x, z, t) and 1D points (y, t);
- 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 **??**. The 1D reconstruction is straightforward from single image, 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. Back to the reference camera represented by $(\mathbf{I}_{3\times3}, \mathbf{0})$, $Y = (v - v_0)Z/f$ 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 this decomposition method for this simplified SFM algorithm does not need any prior geometric computation unlike 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 1D cameras being not physical, take much more virtual points. This makes the subspace reconstruction more numerically stable.

3 Geometry of CMs

We have just looked at the geometric structure of images captured by concentric mosaic set-up. We study the geometry of the composite mosaic images. The CM set-up undergoes not only a planar motion that we have incorporated into the image analysis, but also a continuous and uniform circular path.

Thinking of one CM now, theoretically it is a kind of multi-perspective image with a circular trajectory of the camera center, hence the cencentric mosaic is not a homographical mosaic, but a manifold one. However as CM moves uniformly and continuously along a perfect circle; at each camera position, only the central vertical line image is selected to construct the mosaic. This mosaic construction is equivalent to creating a kind of mecanical 1D panorama camera with $f + R$ as its virtual focal length as illustrated in Figure 1. Here f is supposed to be the original camera focal length and R is the circle radius. In fact, it could be understood as that the mecanical or electronical motor realizes a physical 1d camera engin which measures angles just like a laser scanner or angle meter. For each vertical line of the CM at angle θ, its direction vector or its homogeneous coordinates are

$$\theta \mapsto \mathbf{u} = \lambda \begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix},$$

and for the CM located at $\mathbf{t}$ and oriented with $\mathbf{R}(\alpha)$, the virtual 1D camera is described by

$$\lambda \mathbf{u} = \begin{pmatrix} \mathbf{R}(\alpha) & -\mathbf{R}(\alpha)\mathbf{t} \end{pmatrix}.$$

In Image-based-rendering language, a CM is *rebinned* into a 1D panorama projective camera with a larger focal length! It has exactly the same form as a lasar scanner

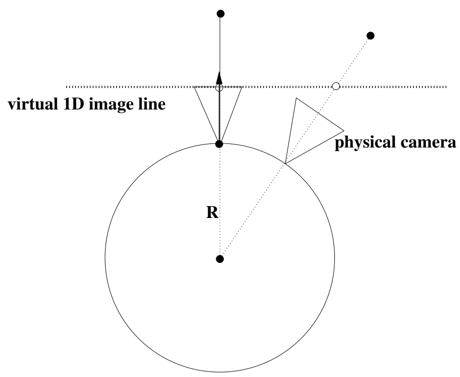

Figure 1. A concentric mosaic is rebinned into a calibrated 1D panorama projective camera.

model studied in [4]. The relative orientation of CMs at different positions, essential for large-scale scenery rendering, can then be determined by the geometry of calibrated 1D projective cameras in Section 4.3.

4 2D Reconstruction from 1D images

Have been reduced 2D images to 1D images, 2D reconstruction from multiple 1D images will be adressed in this section. We first introduce the 1D affine camera model, then describe a reconstruction algorithm based on this concept by factorization method, A complete euclidean reconstruction algorithm from projective images based on the previous results in [4, 20] is described for calibrated 1D projective cameras.

4.1 1D affine camera

The 1D affine camera could be introduced by analogy to 2D affine camera introduced by Mundy and Zisserman [17] as an imaging device from 2-dimension to 1-dimension which preserves the affine properties characterized by the points at infinity. More importantly, it is motivated by 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 shalow 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 algorithm as described in the introduction.

A 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 rank 2 matrix $\mathbf{M}$ of 2×3 as $\lambda\mathbf{u} = \mathbf{M}_{2\times3}\mathbf{x}$. If the line at infinity is identified as $x_3 = 0$, to preserve the affine spaces, it should mapped onto the point at infinity of the image line $u_2 = 0$. This gives the following projection matrix:

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

It maps the finite points $(x, y, 1)^T$ onto finite image points $(u, 1)^T$ with

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

If we further use relative cooridinates 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 translational component t is cancelled and the projection for finite points in relative coordinates are therefore:

$$\Delta u = \mathbf{m}_{1\times 2} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}.$$

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 [22, 24, 12, 27, 8].

Let three views of the point $\Delta \mathbf{x}$ be given as follows:

$$\begin{cases} \Delta u &= \Delta \mathbf{m}\mathbf{x}, \\ \Delta u' &= \Delta \mathbf{m}'\mathbf{x}, \\ \Delta u'' &= \Delta \mathbf{m}''\mathbf{x}. \end{cases} \quad (2)$$

These can be rewritten in matrix form as

$$\begin{pmatrix} \mathbf{m}^T & \Delta u \\ \mathbf{m}'^T & \Delta u' \\ \mathbf{m}''^T & \Delta u'' \end{pmatrix} \begin{pmatrix} \Delta \mathbf{x} \\ -\lambda \end{pmatrix} = \mathbf{0}.$$

The vector $(\Delta \mathbf{x}, -\lambda)^T$ cannot be zero, so

$$\begin{vmatrix} \mathbf{m}^T & \Delta u \\ \mathbf{m}'^T & \Delta u' \\ \mathbf{m}''^T & \Delta u'' \end{vmatrix} = \begin{vmatrix} \mathbf{a} & \mathbf{b} & \begin{matrix} \Delta u \\ \Delta u' \\ \Delta u'' \end{matrix} \end{vmatrix} = 0.$$

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

$$a\Delta u + b\Delta u' + c\Delta u'' = 0.$$

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

4.2 2D reconstruction by factorisation

The above three-view linear constraint directly encodes the relative motion parameters. For reconstruction, similar to factorization method [26] for 2D affine cameras, we could proceed the same way by stacking p points in n images to create the measurement matrix as

$$\begin{pmatrix} \Delta u_1 & \dots & \Delta u_p \\ \Delta u'_1 & \dots & \Delta u'_p \\ \dots & \dots & \dots \\ \Delta u_1^{(n)} & \dots & \Delta u_p^{(n)} \end{pmatrix} = \begin{pmatrix} \mathbf{m}^T \\ \dots \\ \mathbf{m}^{(n)T} \end{pmatrix} \begin{pmatrix} \Delta \mathbf{x}_1 & \dots & \Delta \mathbf{x}_p \end{pmatrix}$$

or in compact form as $\mathbf{U}_{n\times p} = \mathbf{M}_{n\times 2}\mathbf{S}_{2\times p}$. The rank of the $n \times p$ measurement matrix can not exceed 2. By keeping the two largest singular values and zeroing all others of the measurement matrix, we obtain the best rank-two camera matrices $\mathbf{m}_i$ and point reconstruction $\Delta \mathbf{x}_i$.

As for any non singular 2×2 matrix $\mathbf{A}$ representing a plane affine transformation, $\mathbf{U} = \mathbf{MS} = (\mathbf{MA})(\mathbf{A}^{-1}\mathbf{S}) = \mathbf{M}'\mathbf{S}'$. The camera matrix and shape are still affine. We need to look for a $\mathbf{A}$ such that $\mathbf{m}_i\mathbf{A} = c\mathbf{R}_{1\times 2}$ where c is an independent scaling factor that each 1D affine camera may have. So that the metric constraint for euclidean reconstruction assuming constant scaling factor is

$$\mathbf{m}_i \mathbf{A}\mathbf{A}^T \mathbf{m}_i^T = 1.$$

This reconstruction might be sufficient if not it can be served as an initial solution for a nonlinear optimisation. The factorisation method requires all points in all images. For missing points, they can be handled using the linear three-view constraint developed in the previous section.

4.3 Reconstruction from calibrated 1D panorama projective images

The 1D affine camera nicely describes the geometry for the original images within the same concentric mosaic. For recovering relative orientation and depth from concentric mosaics at different positions for large-scale environement modelling, each concentric mosaic is assimilated into a calibrated 1D panorama projective camera. We describe a complete reconstruction method from 3 panorama 1d projective cameras which assembles the recent results in [3, 4, 20, 9] for 1D cameras.

- **Computing uncalibrated 1D trifocal tensor** The geometry of three 1D images is completely characterised by the 1D trifocal tensor $\mathbf{T}_{ijk}$. It minimally parametrizes the three uncalibrated images and it can be estimated linearly with at least 7 point correspondences.
- **Solving a cubic equation for the internal parameters** The 1D camera could be self-calibrated for constant calibration parameters via 1D trifocal tensor [9]. The knowledge of the internal parameters

of a 1D camera is equivalent to that of the image points **i** and **j** of the circular points in $\mathcal{P}^2$ [1, 9]. This pair of complex conjugate points can be uniquely determined by solving the cubic equation: $T_{111}x^3 + (T_{211} + T_{112} + T_{121})x^2 + (T_{212} + T_{221} + T_{122})x + T_{222} = 0$.. The real part of the ratio of the projective coordinates of the image of the circular point **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 the previous section 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 normalised 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 knowning 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 calibrated projection matrix** For reconstruction, we need to convert the tensor compoenent into camera projection matrices. Each projection matrix can be parameterized by

$$\begin{pmatrix} c & -s & t_x \\ s & c & t_y \end{pmatrix}$$

  Actually there are 5 non-homogeneous tensor components, and 5 d.o.f. for P-matrices, θ and θ', $\mathbf{t}$ and $\mathbf{t}'$ up to a scale. This can be solved algebraically, but up to a two-way ambiguity [4, 20].

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

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

4.4 2D reconstruction from calibrated 1D projective images of a circular motion

Though the affine approximation is often sufficient for images from the same CM as described in Section 4.2, we may still apply a full 1D projective camera model. The same reconstruction algorithm as described in the previous subsection is still valid, but the images from the same CM has much weaker geometry than those from the panoramic images of different CM. However the images from the same CM has stronger motion constraint: a circular motion [10]. The calibrated 2×3 projection matrices for a triplet of views can be parameterized 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 parametrisation also suggests a more efficient bundle-like nonlinear optimisation.

5 Experimental results

Experiments on analyzing concentric mosaic data by 1D cameras 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.

KIDS sequence For the KIDS triplet shown in 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 factorization method is shown in Figure 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 Figure 4.

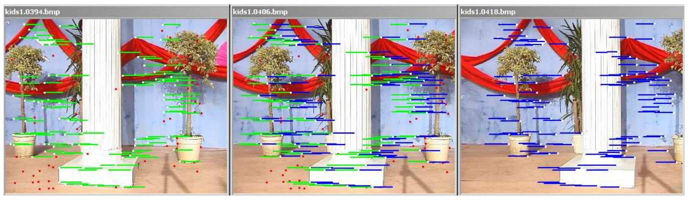

Figure 2. Point correspondences in a triplet from KIDS sequence.

In Figure 5, two columns and the backgroup wall are drawn over the reconstructed plane to illustrate the reconstruction quality.

TOY sequence For the TOY triplet illustrated in Figure 6, there are 151 and 126 match candidates in the first and second pairs. 76 final corresponding triplets are obtained. The 3D affine reconstruction using standard 2D factorisation method is shown in Figure 7. The 2D affine and Euclidean reconstruction using our 1D factorisation method are shown in Figure 8. We can notice the superior reconstruction by 1D factorisation method over the 2D factorisation.

The final 3D VRML model shown in Figures 10 and 9 are reconstructed from re-sampled dense matching by the

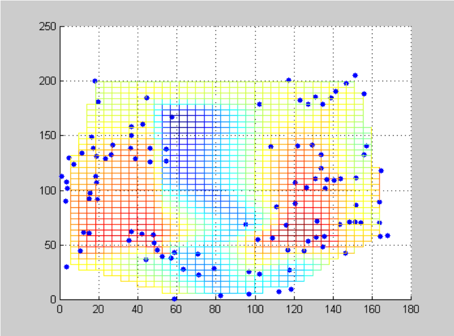

Figure 3. 3D Affine reconstruction by 2D factorisation for KIDS sequence: projection onto (x, z) plane.

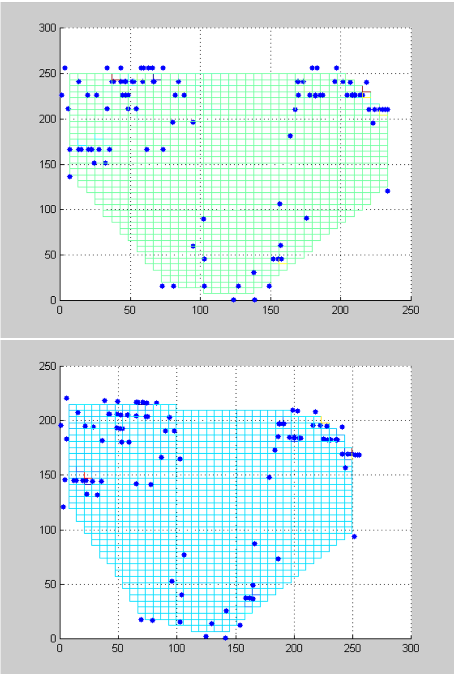

Figure 4. 2D Affine and Euclidean reconstruction by 1D factorisation for KIDS sequence.

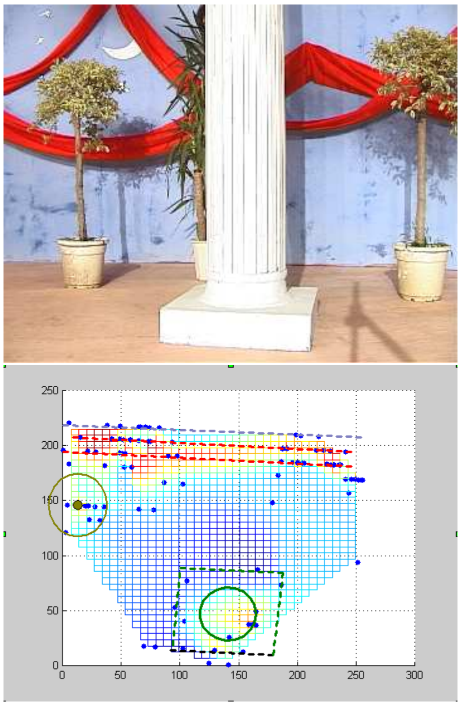

Figure 5. One original image and the reconstructed plane by merging two triplets of KIDS sequence with manual drawing for illustration.

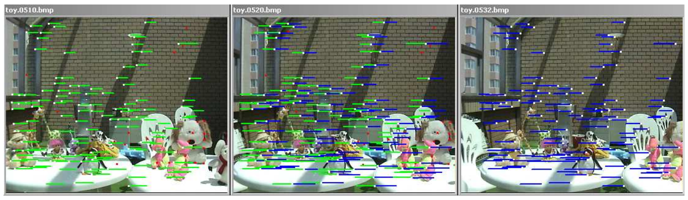

Figure 6. Point correspondences in a triplet from TOY sequence.

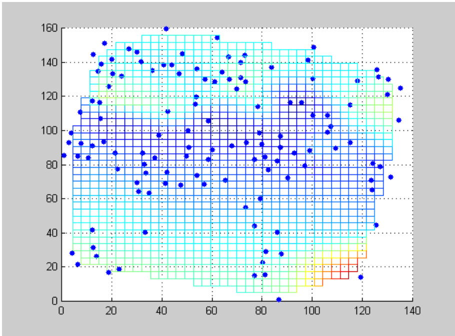

Figure 7. 3D Reconstruction by 2D factorisation method for TOY triplet.

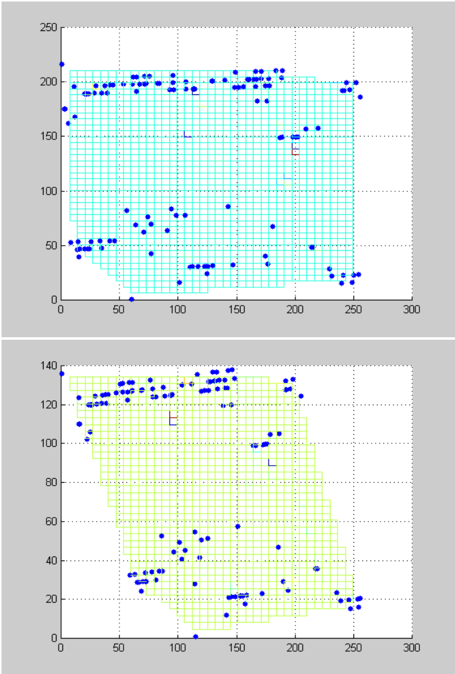

Figure 8. 2D Affine and Euclidean reconstruction by 1D factorisation for TOY triplet.

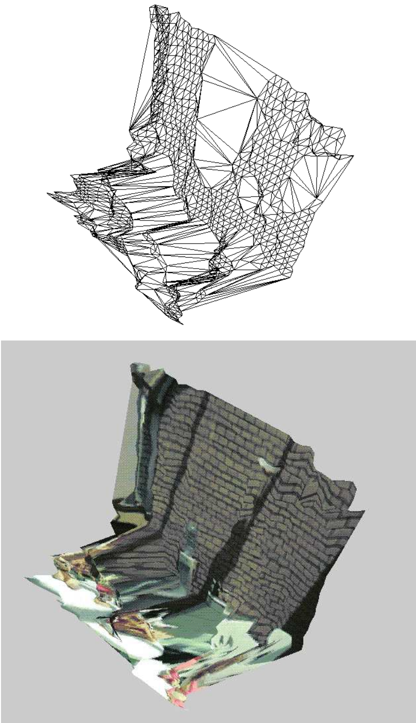

Figure 9. 3D reconstruction of TOY triplet in VRML.

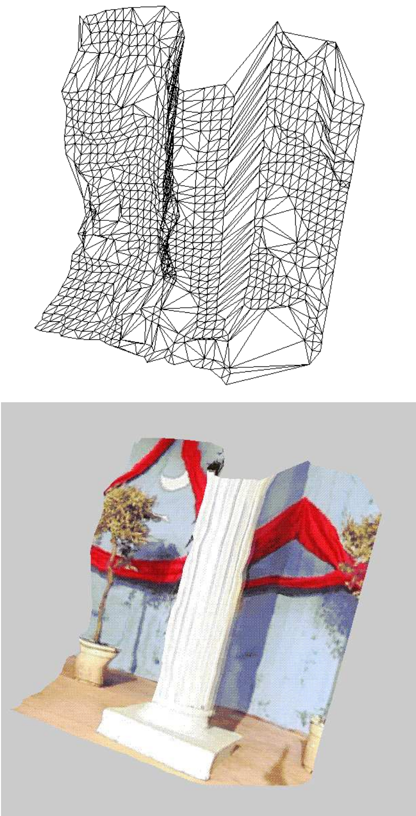

Figure 10. 3D reconstruction of KIDS triplet in VRML.

extension of Euclidean 2D reconstruction. The 3D reconstruction quality is sufficient for image-rendering purpose. Actually we are integrating the 3D reconstruction into the CM viewer for real-time applications.

6 Conclusion

This paper investigated the geometric aspect for depth recovery of the concentric mosaics—a dense sampling of 3 dimensional plenoptic function for image-based rendering. We have introduced the new concept of 1D affine camera and 1D factorisation method for 2D reconstruction. We have shown that the images captured by a concentric mosaic set-up can be completely decomposed into one horizontal 1D projective camera and one vertical affine camera. And each single concentric mosaic can be rebinned into a calibrated 1D panorama projective camera. The key advantage of CM analysis in terms of 1D camera is that the prior motion information has been integrated into the system. The 2D/1D image convertion for concentric mosaics 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 SFM methods which tend to be singular for concentric mosaics. More intensive experiments for combining with the concentric mosaic renderer are undergoing.

Aknowledgements We would like to thank Microsoft Research Beijing for support and B. Triggs and P. Sturm for fruitful discussions. The work has also been partly supported by EU IST Project EVENTS.

References

[1] M. Armstrong, A. Zisserman, and R. Hartley. Self-calibration from image triplets. ECCV'96.

[2] M.N. Armstrong. *Self-Calibration from Image Sequences*. PhD thesis, University of Oxford, 1996.

[3] K. Åström. *Invariance Methods for Points, Curves and Surfaces in Computational Vision*. PhD thesis, Lund University, 1996.

[4] K. Åström and M. Oskarsson. Solutions and ambiguities of the structure and motion problem for 1d retinal vision. JMIV, 2000.

[5] S. Avidan and A. Shashua. Novel view synthesis in tensor space. CVPR'97.

[6] P.A. Beardsley and A. Zisserman. Affine calibration of mobile vehicles. *Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision*, 1995.

[7] S.E. Chen. Quicktime VR - an image-based approach to virtual environment navigation. SIGGRAPH'95.

[8] O. Faugeras and B. Mourrain. About the correspondences of points between n images. *Workshop on Representation of Visual Scenes*, 1995.

[9] O. Faugeras, L. Quan, and P. Sturm. Self-calibration of a 1d projective camera and its application to the self-calibration of a 2d projective camera. ECCV'98.

[10] A.W. Fitzgibbon, G. Cross, and A. Zisserman. Automatic 3D Model Construction for Turn-Table Sequences. SMILE'98.

[11] S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The lumigraph. SIGGRAPH'96.

[12] R.I. Hartley. A linear method for reconstruction from lines and points. ICCV'95.

[13] E. Izquierdo and S. Kruse. Image analysis for 3d modeling, rendering and virtual view generation. CVIU, 1998.

[14] S. Laveau and O. Faugeras. 3D scene representation as a collection of images and fundamental matrices. TR-INRIA, 1994.

[15] M. Levoy and P. Hanrahan. Light field rendering. SIGGRAPH'96.

[16] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. SIGGRAPH'95.

[17] J.L. Mundy and A. Zisserman. Projective geometry for machine vision. *Geometric Invariance in Computer Vision*, 1992.

[18] P.J. Narayanan, P.W. Rander, and T. Kanade. Constructing virtual worlds using dense stereo. ECCV'98.

[19] S. Peleg and J. Herman. Panoramic mosaics by manifold projection. CVPR'97.

[20] L. Quan and T. Kanade. Affine structure from line correspondences with uncalibrated affine cameras. PAMI'97.

[21] P.W. Rander, P.J. Narayanan, and T. Kanade. Recovery of dynamic scene structure from multiple image sequence. *Int'l Conf on Multisensor Fusion and Integration for Intelligent Systems*, 1996.

[22] A. Shashua. Algebraic functions for recognition. PAMI'95.

[23] H.Y. Shum and L.W. He. Rendering with concentric mosaics. SIGGRAPH'99.

[24] M. Spetsakis and J. Aloimonos. A unified theory of structure from motion. *DARPA Image Understanding Workshop*, 1990.

[25] R. Szeliski and H.-Y. Shum. Creating full view panoramic image mosaics and environment maps. SIGGRAPH'97.

[26] C. Tomasi and T. Kanade. Factoring image sequences into shape and motion. IJCV'92.

[27] B. Triggs. Matching constraints and the joint image. ICCV'95.

[28] B. Triggs. Plane + parallax, tensors and factorization. ECCV'2000.

[29] C. Wiles and M. Brady. Ground plane motion camera models. ECCV'96.