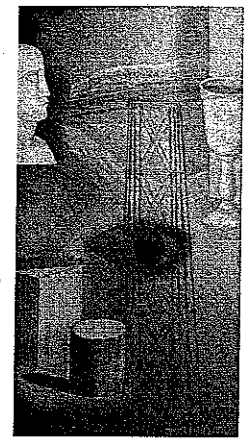
THE JOURNAL OF VISUALIZATION AND COMPUTER ANIMATION
*J Visual Comput Animat* 2002; 13: 211–223 (DOI: 10.1002/vis.290)

# A perspective factorization method for Euclidean reconstruction with uncalibrated cameras



By Mei Han*† and Takeo Kanade

*Structure from motion (SFM), which is recovering camera motion and scene structure from image sequences, has various applications, such as scene modelling, robot navigation, object recognition and virtual reality. Most of previous research on SFM requires the use of intrinsically calibrated cameras. In this paper we describe a factorization-based method to recover Euclidean structure from multiple perspective views with uncalibrated cameras. The method first performs a projective reconstruction using a bilinear factorization algorithm, and then converts the projective solution to a Euclidean one by enforcing metric constraints. The process of updating a projective solution to a full metric one is referred as normalization in most factorization-based SFM methods. We present three normalization algorithms which enforce Euclidean constraints on camera calibration parameters to recover the scene structure and the camera calibration simultaneously, assuming zero skew cameras. The first two algorithms are linear, one for dealing with the case that only the focal lengths are unknown, and another for the case that the focal lengths and the constant principal point are unknown. The third algorithm is bilinear, dealing with the case that the focal lengths, the principal points and the aspect ratios are all unknown. The results of experiments are presented.* Copyright © 2002 John Wiley & Sons, Ltd.

*Revised: 2 May 2002*

KEY WORDS: structure from motion; 3D modelling; camera calibration; Euclidean reconstruction; computer vision

## Introduction

When a camera moves around in a scene, the images taken contain information about the scene structure, the camera motion and the camera intrinsic parameters. Structure from motion (SFM), which is recovering camera motion and scene structure from monocular image sequences, has various applications, such as scene modelling, robot navigation, object recognition and virtual reality. Much work has been done on reconstruction with intrinsically calibrated cameras. However, in practice, many image sequences are taken with uncalibrated cameras and the intrinsic parameters, such as the focal lengths, are changing throughout the sequences. In this paper we present a factorization-based method for Euclidean reconstruction from multiple uncalibrated views.

The factorization method, first developed by Tomasi and Kanade[1] for orthographic views and extended by Poelman and Kanade[2] to weak and paraperspective views, achieved its robustness and accuracy by applying singular value decomposition (SVD) to a large number of images and feature points. The decomposition solutions are converted to Euclidean reconstructions by the normalization process which enforces metric constraints on the camera motion parameters. Yu *et al*.[3] presented a new approach based on a higher-order approximation of perspective projection by using Taylor expansion of depth. The accuracy of the approximation depended on the order of Taylor expansion and

*Correspondence to: M. Han, NEC USA, Inc., 10080 North Wolfe Road, SW3-350, Cupertino, CA 95014, USA.
E-mail: meihan@ccrl.sj.nec.com

†The research described in this paper was conducted while the first author was a PhD student in the Robotics Institute at CMU.

the computation increased exponentially as the order increased. Christy and Horaud[4,5] described a method for the perspective camera model by incrementally performing reconstructions with either a weak or a paraperspective camera model. Recently, some work has been done to extend the factorization methods from feature-based methods to plane-based methods. Ma and Ahuja[6] recovered a dense shape, which is composed of the recovered plane positions and normals, from region correspondences by factorization. Sturm[7] presented a factorization-based method to estimate poses of multiple planes. One major limitation with most factorization methods, however, is that they require the use of intrinsically calibrated cameras.

When nothing is known about the camera intrinsic parameters, the extrinsic parameters or the scene, it is only possible to compute a reconstruction up to an unknown projective transformation[8]. There has been considerable progress on projective reconstruction.[8–14] Some methods use only two, three or four images to obtain a projective reconstruction by a linear least squares method.[15,16] On the other hand, some projective reconstruction methods take advantage of the large amount of information from image sequences.[17–20] Triggs proposed a projective factorization method[19] which recovered projective depths by estimating a set of fundamental matrices to chain all the images together. Sturm and Triggs[18] described an iterative factorization method to recover the projective structure. Heyden *et al*.[20–22] presented methods of using multilinear subspace constraints to perform projective structure from motion. Mahamud and Hebert described an iterative projective reconstruction method[23] and proved its monotonic convergence.

In order to obtain a Euclidean reconstruction from the projective reconstruction, some additional information about either the camera or the scene is needed. Hartley recovered the Euclidean shape using a global optimization technique, assuming that the intrinsic parameters were constant.[24] Heyden also proposed a new formulation of Kruppa equations to calibrate cameras with constant parameters.[25] Heyden and Åström[26] used a bundle adjustment algorithm to estimate the focal lengths, the principal points, the camera motion and the object shape together. Triggs calibrated the cameras by recovering the absolute quadric.[27] Pollefeys *et al*. assumed that the focal length was the only varying intrinsic parameter and presented a linear algorithm.[28] Agapito *et al*. proposed a linear self-calibration algorithm for rotating and zooming cameras.[29] Ponce upgraded projective reconstructions to metric ones based on a sequence of linear steps for zero-skew uncalibrated cameras.[30]

In this paper we describe a factorization-based method which recovers Euclidean shape and motion from multiple uncalibrated perspective views. In practice, most of the camera intrinsic parameters are unknown and varying from image to image, especially when multiple cameras are involved. The only parameter that rarely changes is the skew of the camera, which is the difference between $\Pi/2$ and the angle between the rows and columns of an image. The skew is always zero. It is proved that the absence of skew is sufficient to yield a metric reconstruction.[28] Given tracked feature points, the factorization-based method reconstructs the object shape, the camera motion and the intrinsic parameters. We first apply an iterative algorithm to get a projective reconstruction, then propose three normalization algorithms to impose metric constraints on the projective reconstruction. The normalization algorithms recover the unknown intrinsic parameters and convert the projective solution to a Euclidean one simultaneously. The first algorithm deals with the case that the focal lengths are the only unknown parameters. The second one deals with the case that the focal lengths and the principal point are unknown, while the principal point is fixed. These two algorithms are linear. The third algorithm, which is bilinear, works in the case that the focal lengths, the principal points and the aspect ratios are all unknown. Experimental results are presented.

## Projective reconstruction

We decouple the uncalibrated reconstruction process into projective reconstruction and Euclidean reconstruction. In this section we describe the bilinear projective reconstruction algorithm.

Suppose there are $n$ perspective cameras: $P_i, i = 1 \ldots n$ and $m$ feature points $\mathbf{x}_j, j = 1 \ldots m$ represented by homogeneous coordinates. The image coordinates are represented by $(u_{ij}\ v_{ij})$. Using the symbol $\sim$ to denote equality up to a scale, the following hold:

$$\begin{bmatrix} u_{ij} \\ v_{ij} \\ 1 \end{bmatrix} \sim P_i \mathbf{x}_j \quad \text{or} \quad \lambda_{ij} \begin{bmatrix} u_{ij} \\ v_{ij} \\ 1 \end{bmatrix} = P_i \mathbf{x}_j \qquad (1)$$

where $\lambda_{ij}$ is a non-zero scale factor, commonly called projective depth. Each $P_i$ is a $3 \times 4$ matrix and each

feature point $\mathbf{x}_j$ is a $4 \times 1$ vector. The equivalent matrix form is

$$W_s = \begin{bmatrix} \lambda_{11}\begin{bmatrix} u_{11} \\ v_{11} \\ 1 \end{bmatrix} & \cdots & \lambda_{1m}\begin{bmatrix} u_{1m} \\ v_{1m} \\ 1 \end{bmatrix} \\ \vdots & \vdots & \vdots \\ \lambda_{n1}\begin{bmatrix} u_{n1} \\ v_{n1} \\ 1 \end{bmatrix} & \cdots & \lambda_{nm}\begin{bmatrix} u_{nm} \\ v_{nm} \\ 1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} P_1 \\ \vdots \\ P_n \end{bmatrix} [\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_m] = PX \tag{2}$$

where $W_s$ is a $3n \times m$ matrix, called *scaled measurement matrix*. It encodes the projected image information and the projective depths. Since each $P_i$ is a $3 \times 4$ matrix, $W_s$ is at most rank 4. We therefore apply the following projective factorization algorithm, which is very similar to Triggs's approach[19] and Heyden's [21]. This algorithm iteratively applies rank 4 factorization to the current scaled measurement matrix.

### Bilinear Projective Factorization Algorithm

1. Set $\lambda_{ij} = 1$, for $i = 1 \cdots n$ and $j = 1 \cdots m$
2. Compute the current scaled measurement matrix $W_s$ by Equation (2).
3. Perform rank 4 factorization on $W_s$, generate the projective motion and shape
4. Reset $\lambda_{ij} = P_i^{(3)}\mathbf{x}_j$, where $P_i^{(3)}$ denotes the third row of the projection matrix $P_i$
5. If all $\lambda_{ij}$'s are the same as the previous iteration, stop; else go to step 2

The goal of the projective reconstruction process is to estimate the values of the projective depths ($\lambda_{ij}$'s) which make Equation (2) consistent. The reconstruction results are iteratively improved by back-projecting the current projective reconstruction to refine the depth estimates.

## Euclidean Reconstruction

The factorization of Equation (2) recovers the motion and shape up to a $4 \times 4$ linear projective transformation $H$,

$$W_s = \hat{P}\hat{X} = \hat{P}HH^{-1}\hat{X} = PX \tag{3}$$

where $P = \hat{P}H$ and $X = H^{-1}\hat{X}$. $\hat{P}$ and $\hat{X}$ are referred to as the projective motion and projective shape. Any non-singular $4 \times 4$ matrix could be inserted between $\hat{P}$ and $\hat{X}$ to get another motion and shape pair.

Let us assume zero skews. We impose metric constraints to the projective motion and shape in order to simultaneously reconstruct the intrinsic parameters (i.e., the focal lengths, the principal points and the aspect ratios) and the linear transformation $H$, from which we can get the Euclidean motion and shape. We call this process *normalization*. We classify the situations into three cases:

Case 1: Only the focal lengths are unknown.

This case includes the situations that the camera undergoes zooming in/out during the sequence. The focal lengths are therefore the main concerns of the reconstruction process.

Case 2: The focal lengths and the principal point are unknown, and the principal point is fixed.

In this case we are interested in the situations in which the camera focal length changes only a little, so that there is no obvious zooming effect and the principal point is very close to being constant. Aerial image sequences taken by a flying platform are examples of this case.

Case 3: The focal lengths, the principal points and the aspect ratios are all unknown and varying.

This case covers the situations that multiple cameras are included. The focal lengths, the principal points and the aspect ratios all vary from image to image.

We present three factorization-based normalization algorithms to deal with these three cases respectively. The algorithms are linear for the first two cases and bilinear for the third case.

## Normalization Algorithm Outline

The projection matrix $P_i$ is

$$P_i \sim K_i[R_i \quad \mathbf{t}_i] \tag{4}$$

where

$$K_i = \begin{bmatrix} f_i & 0 & u_{0i} \\ 0 & \alpha_i f_i & v_{0i} \\ 0 & 0 & 1 \end{bmatrix} \quad R_i = \begin{bmatrix} \mathbf{i}_i^T \\ \mathbf{j}_i^T \\ \mathbf{k}_i^T \end{bmatrix} \quad \mathbf{t}_i = \begin{bmatrix} t_{xi} \\ t_{yi} \\ t_{zi} \end{bmatrix}$$

The upper triangular calibration matrix $K_i$ encodes the intrinsic parameters of the $i$th camera: $f_i$ represents the focal length, $(u_{0i}\ v_{0i})$ is the principal point and $\alpha_i$ is the aspect ratio. $R_i$ is the $i$th rotation matrix with $\mathbf{i}_i$, $\mathbf{j}_i$ and $\mathbf{k}_i$ denoting the rotation axes. $\mathbf{t}_i$ is the $i$th translation vector. Combining Equation (4) for $i = 1 \ldots n$ into one matrix equation, we get

$$P = [M \quad T] \tag{5}$$

where

$$M = \begin{bmatrix} \mathbf{m}_{x1} & \mathbf{m}_{y1} & \mathbf{m}_{z1} & \cdots & \mathbf{m}_{xn} & \mathbf{m}_{yn} & \mathbf{m}_{zn} \end{bmatrix}^T$$
$$T = \begin{bmatrix} T_{x1} & T_{y1} & T_{z1} & \cdots & T_{xn} & T_{yn} & T_{zn} \end{bmatrix}^T$$

and

$$\mathbf{m}_{xi} = \mu_i f_i \mathbf{i}_i + \mu_i u_{0i} \mathbf{k}_i \quad \mathbf{m}_{yi} = \mu_i \alpha_i f_i \mathbf{j}_i + \mu_i v_{0i} \mathbf{k}_i \quad \mathbf{m}_{zi} = \mu_i \mathbf{k}_i$$
$$T_{xi} = \mu_i f_i t_{xi} + \mu_i u_{0i} t_{zi} \quad T_{yi} = \mu_i \alpha_i f_i t_{yi} + \mu_i v_{0i} t_{zi} \quad T_{zi} = \mu_i t_{zi} \tag{6}$$

$\mu_i$ is the scale factor of the homogeneous representation in Equation (4). The shape matrix is represented by

$$X \sim \begin{bmatrix} S \\ \mathbf{1} \end{bmatrix} \tag{7}$$

where

$$S = [\mathbf{s}_1 \ \mathbf{s}_2 \ \cdots \ \mathbf{s}_m]$$

and

$$\mathbf{s}_j = \begin{bmatrix} x_j & y_j & z_j \end{bmatrix}^T$$
$$\mathbf{x}_j = \begin{bmatrix} \nu_j \mathbf{s}_j^T & \nu_j \end{bmatrix}^T$$

where $\nu_j$ represents the scale factor of the homogeneous representation in Equation (7).

## World Coordinate System Location

We place the origin of the world coordinate system at the centre of gravity of all the scaled feature points to enforce

$$\sum_{j=1}^{m} \nu_j \mathbf{s}_j = 0 \tag{8}$$

We get

$$\sum_{j=1}^{m} \lambda_{ij} u_{ij} = \sum_{j=1}^{m} (\mathbf{m}_{xi} \cdot \nu_j \mathbf{s}_j + \nu_j T_{xi}) = \mathbf{m}_{xi} \cdot \sum_{j=1}^{m} \nu_j \mathbf{s}_j + T_{xi} \sum_{j=1}^{m} \nu_j = T_{xi} \sum_{j=1}^{m} \nu_j \tag{9}$$

Similarly,

$$\sum_{j=1}^{m} \lambda_{ij} v_{ij} = T_{yi} \sum_{j=1}^{m} \nu_j \quad \sum_{j=1}^{m} \lambda_{ij} = T_{zi} \sum_{j=1}^{m} \nu_j \tag{10}$$

Define the $4 \times 4$ projective transformation $H$ as

$$H = [A \quad B] \tag{11}$$

where $A$ is $4 \times 3$ and $B$ is $4 \times 1$.
Since $P = \hat{P}H$,

$$[M \quad T] = \hat{P}[A \quad B] \tag{12}$$

we have

$$T_{xi} = \hat{P}_{xi} B \quad T_{yi} = \hat{P}_{yi} B \quad T_{zi} = \hat{P}_{zi} B \tag{13}$$

From Equations (9) and (10), we know

$$\frac{T_{xi}}{T_{zi}} = \frac{\sum_{j=1}^{m} \lambda_{ij} u_{ij}}{\sum_{j=1}^{m} \lambda_j} \quad \frac{T_{yi}}{T_{zi}} = \frac{\sum_{j=1}^{m} \lambda_{ij} v_{ij}}{\sum_{j=1}^{m} \lambda_j} \tag{14}$$

We set up $2n$ linear equations of the four unknown elements of the matrix $B$. Linear least squares solutions are then computed.

## Normalization

As $\mathbf{m}_{xi}$, $\mathbf{m}_{yi}$ and $\mathbf{m}_{zi}$ are the sum of the scaled rotation axes, we get the following constraints from Equation (6):

$$\begin{aligned} |\mathbf{m}_{xi}|^2 &= \mu_i^2 f_i^2 + \mu_i^2 u_{0i}^2 \\ |\mathbf{m}_{yi}|^2 &= \mu_i^2 \alpha_i^2 f_i^2 + \mu_i^2 v_{0i}^2 \\ |\mathbf{m}_{zi}|^2 &= \mu_i^2 \\ \mathbf{m}_{xi} \cdot \mathbf{m}_{yi} &= \mu_i^2 u_{0i} v_{0i} \\ \mathbf{m}_{xi} \cdot \mathbf{m}_{zi} &= \mu_i^2 u_{0i} \\ \mathbf{m}_{yi} \cdot \mathbf{m}_{zi} &= \mu_i^2 v_{0i} \end{aligned} \tag{15}$$

Based on the three different assumptions of the intrinsic parameters (three cases), we translate the above constraints to linear constraints on $MM^T$, as will be explained later. According to Equation (12),

$$M = \hat{P}A \tag{16}$$

therefore,

$$MM^T = \hat{P}AA^T\hat{P}^T \tag{17}$$

Define

$$Q = AA^T \tag{18}$$

we can translate the linear constraints on $MM^T$ to the constraints on the 10 unknown elements of the symmetric $4 \times 4$ matrix $Q$. Linear least squares solutions are computed. Then we get the matrix $A$ from $Q$ by rank 3 matrix decomposition

### Motion and Shape Recovery

Once the matrices $A$ and $B$ have been found, the projective transformation is $H = [A\,B]$. The shape is computed as $X = H^{-1}\hat{X}$ and the motion matrix as $P = \hat{P}H$. We first compute the scales $\mu_i$,

$$\mu_i = |\mathbf{m}_{zi}| \tag{19}$$

We then compute the principal points (if applied),

$$u_{0i} = \frac{\mathbf{m}_{xi} \cdot \mathbf{m}_{zi}}{\mu_i^2} \quad v_{0i} = \frac{\mathbf{m}_{yi} \cdot \mathbf{m}_{zi}}{\mu_i^2} \tag{20}$$

and the focal lengths as

$$f_i = \frac{\sqrt{|\mathbf{m}_{xi}|^2 - \mu_i^2 u_{0i}^2}}{\mu_i} \tag{21}$$

The aspect ratios (if applied) are

$$\alpha_i = \frac{\sqrt{|\mathbf{m}_{yi}|^2 - \mu_i^2 v_{0i}^2}}{\mu_i f_i} \tag{22}$$

Therefore, the motion parameters are

$$\begin{aligned} \mathbf{k}_i &= \frac{\mathbf{m}_{zi}}{\mu_i} \quad \mathbf{i}_i = \frac{\mathbf{m}_{xi} - \mu_i u_{0i}\mathbf{k}_i}{\mu_i f_i} \quad \mathbf{j}_i = \frac{\mathbf{m}_{yi} - \mu_i v_{0i}\mathbf{k}_i}{\mu_i \alpha_i f_i} \\ t_{zi} &= \frac{T_{zi}}{\mu_i} \quad t_{xi} = \frac{T_{xi} - \mu_i u_{0i} t_{zi}}{\mu_i f_i} \quad t_{yi} = \frac{T_{yi} - \mu_i v_{0i} t_{zi}}{\mu_i \alpha_i f_i} \end{aligned} \tag{23}$$

### Algorithm Outline

The normalization process is summarized by the following algorithm.

#### Normalization Algorithm

1. Perform SVD on $W_s$ and get the projective motion $\hat{P}$ and the projective shape $\hat{X}$.
2. Sum up each row of $W_s$ and compute the ratios between them as in Equation (14)
3. Set up $2n$ linear equations of the four unknown elements of the matrix $B$ based on the ratios and compute $B$
4. Set up linear equations of the 10 unknown elements of the symmetric matrix $Q$ and get $Q$
5. Perform rank 3 matrix decomposition on $Q$ to get $A$ from $Q = AA^T$
6. Put matrices $A$ and $B$ together and get the projective transformation $H = [A\,B]$.
7. Recover the shape using $X = H^{-1}\hat{X}$ and the motion matrix using $P = \hat{P}H$.
8. Recover the intrinsic parameters, the rotation axes and the translation vectors according to Equations (20)–(23).

## Case 1: Unknown Focal Lengths

Assume that the focal lengths are the only unknown intrinsic parameters. Then we have

$$u_{0i} = 0 \quad v_{0i} = 0 \quad \alpha_i = 1 \tag{24}$$

We combine the constraints in Equation (15) to impose the following linear constraints on the unknown elements of the matrix $Q$:

$$\begin{aligned} |\mathbf{m}_{xi}|^2 &= |\mathbf{m}_{yi}|^2 \\ \mathbf{m}_{xi} \cdot \mathbf{m}_{yi} &= 0 \\ \mathbf{m}_{xi} \cdot \mathbf{m}_{zi} &= 0 \\ \mathbf{m}_{yi} \cdot \mathbf{m}_{zi} &= 0 \end{aligned}$$

We can add one more equation assuming $\mu_1 = 1$:

$$|\mathbf{m}_{z1}|^2 = 1 \tag{25}$$

In total we have $4n + 1$ linear equations of the 10 unknown elements of $Q$.

The only intrinsic parameters to be recovered in this case are the focal lengths. As the aspect ratios are 1, the focal lengths are computed by the average of Equations (21) and (22):

$$f_i = \frac{|\mathbf{m}_{xi}| + |\mathbf{m}_{yi}|}{2\mu_i} \tag{26}$$

## Case 2: Unknown Focal Lengths and Constant Principal Point

In case 2, we assume that the focal lengths are unknown and the principal point is constant. Then,

$$\begin{aligned} \frac{\mathbf{m}_{xi} \cdot \mathbf{m}_{yi}}{\mathbf{m}_{xi} \cdot \mathbf{m}_{zi}} &= \frac{\mathbf{m}_{yi} \cdot \mathbf{m}_{zi}}{\mathbf{m}_{zi} \cdot \mathbf{m}_{zi}} \\ (|\mathbf{m}_{xi}|^2 - |\mathbf{m}_{yi}|^2)(\mathbf{m}_{zi} \cdot \mathbf{m}_{zi}) &= (\mathbf{m}_{xi} \cdot \mathbf{m}_{zi})^2 - (\mathbf{m}_{yi} \cdot \mathbf{m}_{zi})^2 \end{aligned} \tag{27}$$

and

$$\begin{aligned}
\frac{\mathbf{m}_{zi}\ \mathbf{m}_{zi}}{\mathbf{m}_{zj}\ \mathbf{m}_{zj}} &= \frac{|\mathbf{m}_{xi}|^2 - |\mathbf{m}_{yi}|^2}{|\mathbf{m}_{xj}|^2 - |\mathbf{m}_{yj}|^2} \\
\frac{|\mathbf{m}_{xi}|^2 - |\mathbf{m}_{yi}|^2}{|\mathbf{m}_{xj}|^2 - |\mathbf{m}_{yj}|^2} &= \frac{\mathbf{m}_{xi}\ \mathbf{m}_{yi}}{\mathbf{m}_{xj}\ \mathbf{m}_{yj}} \\
\frac{\mathbf{m}_{xi}\ \mathbf{m}_{yi}}{\mathbf{m}_{xj}\ \mathbf{m}_{yj}} &= \frac{\mathbf{m}_{xi}\ \mathbf{m}_{zi}}{\mathbf{m}_{xj}\ \mathbf{m}_{zj}} \\
\frac{\mathbf{m}_{xi}\ \mathbf{m}_{zi}}{\mathbf{m}_{xj}\ \mathbf{m}_{zj}} &= \frac{\mathbf{m}_{yi}\ \mathbf{m}_{zi}}{\mathbf{m}_{yj}\ \mathbf{m}_{zj}} \\
\frac{\mathbf{m}_{yi}\ \mathbf{m}_{zi}}{\mathbf{m}_{yj}\ \mathbf{m}_{zj}} &= \frac{\mathbf{m}_{zi}\ \mathbf{m}_{zi}}{\mathbf{m}_{zj}\ \mathbf{m}_{zj}}
\end{aligned} \tag{28}$$

where $j = i + 1$, if $i \neq n$; $j = 1$, if $i = n$. We also have the following equation assuming $\mu_1 = 1$:

$$|\mathbf{m}_{z1}|^4 = 1 \tag{29}$$

These are linear equations of the unknown elements of the symmetric matrix $Q' = \mathbf{q}\mathbf{q}^T$, where $\mathbf{q}$ is a $10 \times 1$ vector composed of the 10 unknown elements of the matrix $Q$. In total, we can have $12n + 1$ linear equations of the 55 unknown elements of the matrix $Q'$.

Once $Q'$ has been computed, $\mathbf{q}$ is generated by rank 1 matrix decomposition of $Q'$. We then put the 10 elements of $\mathbf{q}$ into a symmetric $4 \times 4$ matrix $Q$ which is factored as $AA^T$.

We compute the principal point as the average of Equation (20):

$$\begin{aligned}
u_0 &= \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{m}_{xi}\ \mathbf{m}_{zi}}{\mu_i^2} \\
v_0 &= \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{m}_{yi}\ \mathbf{m}_{zi}}{\mu_i^2}
\end{aligned} \tag{30}$$

and the focal lengths as the average of Equations (21) and (22):

$$f_i = \frac{\sqrt{|\mathbf{m}_{xi}|^2 - \mu_i^2 u_0^2} + \sqrt{|\mathbf{m}_{yi}|^2 - \mu_i^2 v_0^2}}{2\mu_i} \tag{31}$$

## Case 3: Unknown Focal Lengths, Principal Points and Aspect Ratios

Case 3 includes the situations that the focal lengths, the principal points and the aspect ratios are all unknown and varying. We then represent the constraints in Equation (15) as bilinear equations on the focal lengths and the principal points plus the aspect ratios. Starting with the rough values of the principal points and the aspect ratio of the first camera ($\alpha_1$), we impose linear constraints on the unknown elements of the matrix $Q$:

$$\begin{aligned}
\mathbf{m}_{xi}\ \mathbf{m}_{yi} &= u_{0i}v_{0i}\,\mathbf{m}_{zi}\ \mathbf{m}_{zi} \\
\mathbf{m}_{xi}\ \mathbf{m}_{zi} &= u_{0i}\,\mathbf{m}_{zi}\ \mathbf{m}_{zi} \\
\mathbf{m}_{yi}\ \mathbf{m}_{zi} &= v_{0i}\,\mathbf{m}_{zi}\ \mathbf{m}_{zi}
\end{aligned} \tag{32}$$

We add two more equations assuming $\mu_1 = 1$,

$$\begin{aligned}
\alpha_1^2(|\mathbf{m}_{x1}|^2 - u_{01}^2) &= |\mathbf{m}_{y1}|^2 - v_{01}^2 \\
|\mathbf{m}_{z1}|^2 &= 1
\end{aligned} \tag{33}$$

Once the matrix $H$ has been found, the current shape is $X = H^{-1}\hat{X}$ and the current motion matrix is $P = \hat{P}H$. We compute the refined principal points, the currently recovered focal lengths and the refined aspect ratios according to Equations (20), (21) and (22) respectively. The current motion parameters are then computed as in Equation (23).

Taking the refined principal points and the first aspect ratio, the normalization steps are performed again to generate the matrix $H$, then the focal lengths, the current shape and motion, the refined principal points and aspect ratios. The above steps are repeated until the principal points and the first aspect ratio do not change.

## Discussion

The normalization process is computationally equivalent to recovering the absolute quadric, which is computed by translating the constraints on the intrinsic camera parameters to the constraints on the absolute quadric.[27,28] Our representation is explicit in the motion parameters (rotation axes and translation vectors) and enables the geometric constraints to be naturally enforced. The representation also deals with the similarity ambiguity problem directly by putting the world coordinate system at the centre of gravity of the object and aligning its orientation with the first camera. Compared with the method presented by Pollefeys *et al*,[28] our normalization algorithm (case 1) is based on the same constraints as their method, but our representation enables natural extensions to the reconstruction of other intrinsic parameters (normalization algorithms of case 2 and case 3).

Heyden and Åström[26] described a method to estimate the focal lengths and the principal points. The method was based on a bundle adjustment algorithm. Ponce used a sequence of linear operations to upgrade projective reconstructions to metric ones.[30] His method solved a similar problem as in our case 3. We build projective and Euclidean reconstructions in the unified factorization framework in which the normalization step is naturally incorporated to take care of the metric constraints. The advantage of the factorization-based method is its reliability since it uniformly considers all the data in all the images. However, it requires all points to be visible in all views, which is a major problem for most factorization methods. Some progress has been achieved to incorporate the information of missing data into the reconstruction framework. Shum *et al* proposed an iterative method which minimized the sum of square differences between the fitted low rank matrix and the elements that are not missing in the data matrix.[31] Urban *et al* presented a linear projective reconstruction method from image sequences with missing data.[32] It requires that the images share a common reference view. Rother and Carlsson described a linear algorithm for scene reconstruction and camera recovery based on having four points on a reference plane visible in all views.[33] Jacobs[34] fit a low rank matrix to a matrix with missing elements by combining constraints on the solution derived from small submatrices of the full matrix. He also presented the application of the linear fitting method to structure from motion problem. The basic idea is to regard the missing data estimation and recovery problem as an EM process in order to find maximum likelihood estimates for unknown values. We are exploring the possibilities of incorporating the linear fitting idea into the reconstruction method presented in this paper.

It is beyond the scope of this paper to derive a complete solution of critical motions that lead to inherent ambiguities in uncalibrated Euclidean reconstructions. Basically, the critical motions correspond to the situations when the linear/bilinear estimations of $Q$ will fail, that is, the constraint matrix has rank less than the number of unknown values in $Q$.

There has been much work done about the critical motions. Sturm analysed the case where all intrinsic parameters are fixed.[35] Kahl *et al* applied subgroup approach to self-calibration constraints when some of the intrinsic parameters can vary.[36] They proved that *given the plane at infinity and known skew, aspect ratio and principal point, then a motion is critical if and only if there is only one viewing direction*. The explicit geometric descriptions of the corresponding critical motion sequences are: (i) arbitrary rotations about the optical axis and translations, (ii) arbitrary rotations about at most two centres, (iii) forward-looking motions along an ellipse and/or a corresponding hyperbola in an orthogonal plane. These are the critical motion sequences for case 1. They also analyse the case with zero skew and unit aspect ratio, which covers case 2 of the uncalibrated reconstruction method. The critical motion sequences for this case happen when *there are at most two viewing directions*. Sturm described the critical motion sequences for stereo systems with varying focal lengths.[37] Ponce gave three classes of critical motions for arbitrary zero-skew cameras:[30] (i) pure translations: the optical centre of the camera may change in an arbitrary manner but the camera's orientation is held constant, (ii) planar motions: the optical centre is held in the plane $y = 0$ and the camera is allowed to rotate about the $y$ axis, (iii) straight-line motions: the optical centre moves along a straight line but the orientation of the camera is allowed to change arbitrarily. These correspond to the critical motions for case 3.

## Experiments

In this section we demonstrate experimental results of the perspective factorization method. Given tracked feature points, we first generate the projective reconstruction, and then recover the Euclidean reconstruction and the camera intrinsic parameters using one of the three normalization algorithms. First, synthetic experiments are conducted to evaluate the quality of the reconstruction method. Then, results for real image sequences corresponding to each of the three cases are shown as well. Experimental results on synthetic and real data show that this method is reliable under noisy conditions.

## Synthetic Data

We synthesize 50 sequences of 20 frames with eight feature points representing a cube in the scene. The distance between the camera and the cube is between 4 and 15 times the cube size. The camera rotation is through 30–65 degrees around the cube. The image size is $640 \times 480$. We add 2 pixels standard noise to the feature locations. The experimental results show that the method converges reliably. The errors of the recovered

feature points positions, averaged over three cases, are less than 0.8% of the object size. The recovered focal lengths are always within $1 \pm 1.8\%$ of the true values. The errors of the principal points are less than 0.25% of the image size and the errors of the aspect ratios are less than 0.5% of the true values. The maximum distance between the recovered camera locations and the corresponding ground truth values is 2.4% of the object size and the maximum difference between the recovered camera orientations and the true values is 0.33°.

## Real Data I: Building Sequence

The building sequence was taken using a hand-held camera in front of a building. The camera was very far from the building at first, then moved toward the building, and away again. The camera was zoomed in when it was far from the building and zoomed out when it was close so that the building appeared to be almost the same size in every image of the sequence. The longest focal length was about three times the shortest one according to the rough readings on the camera. The sequence includes 14 frames, of which three are shown in Figure 1(a)–(c). Fifty feature points were manually selected along the building windows and the corners as shown in Figure 1(d). In this example we assume the focal lengths are unknown while the principal points are given (the middle of the images) and the aspect ratios are 1. We apply the perspective factorization algorithm for case 1 to this example.

Figure 2(a) shows the reconstructed building and camera trajectories. The top view shows that the recovered camera moves toward the building and then away again as expected. The recovered camera positions and orientations shown in the side view demonstrate that all the cameras have the almost same height and tilt upward a little bit, which are the expected values when the same person took the sequence while walking in front of the building. Figure 2(b) shows the reconstructed building with texture mapping. To quantify the results, we measure the orthogonality and parallelism of the lines composed of the recovered feature points. The average angle between pairs of expected parallel lines is 0.89° and the average angle between pairs of expected perpendicular lines is 91.74°. Figure 3 plots the recovered focal lengths, which shows that the focal lengths are changing with the camera motion as we expected.

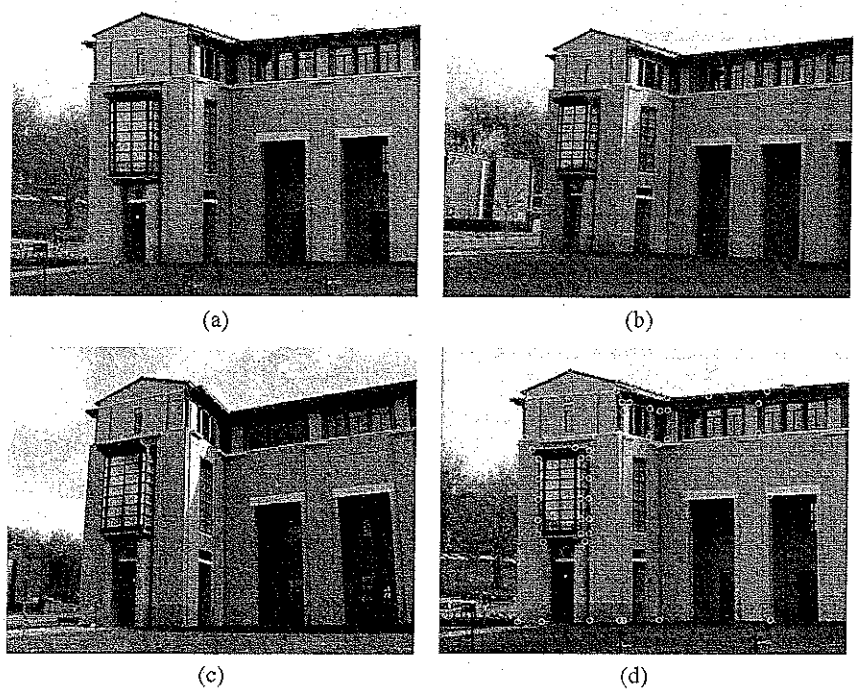



*Figure 1. Building sequence input: (a) 1st image, (b) 4th image, (c) 9th image of the building sequence. (d) 1st image of the building sequence with the feature points overlaid*

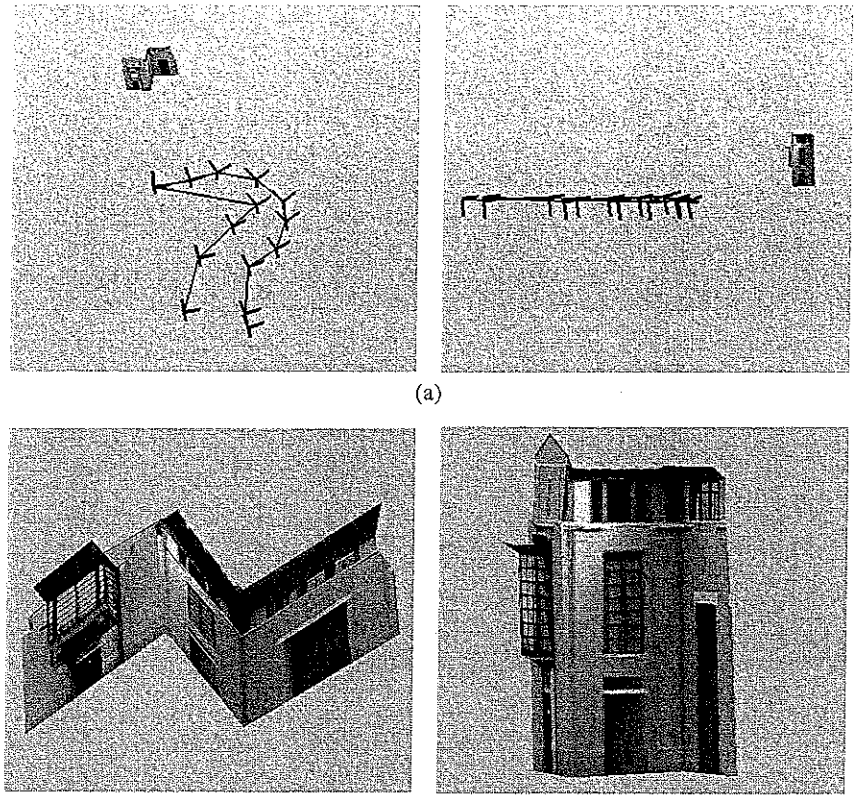



*Figure 2 Building sequence results: (a) Top and side view of the reconstruction; the 3-axis figures denote the recovered cameras. The top view shows that the recovered camera moves toward the building, then away again as expected. The side view shows that the recovered locations of the cameras are at the same height and the orientations are tilted upward. (b) Bottom and side view of the reconstructed building with texture mapping*

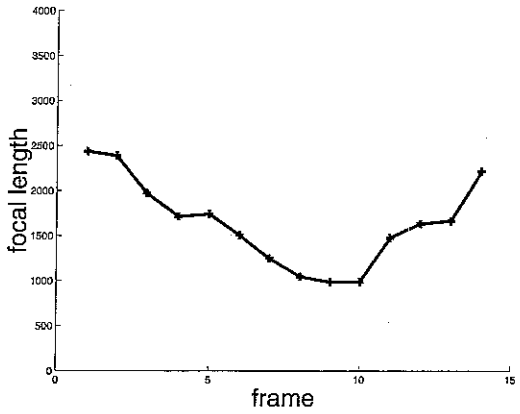



*Figure 3 Building sequence results. Focal lengths of the building sequence recovered by the perspective factorization method. The recovered values are changing with the camera motion as expected*

## Real Data 2: Grand Canyon Sequence

The second example is an aerial image sequence taken from a small aeroplane flying over the Grand Canyon. The plane changed its altitude as well as its roll, pitch and yaw angles during the sequence. The sequence consists of 97 images, and 86 feature points were tracked through the sequence. Three frames from the sequence are shown in Figure 4(a)–(c), and the tracked feature points are shown in Figure 4(d). We assume that the focal lengths and the principal point are unknown, but that the principal point is fixed over the sequence. The normalization algorithm for case 2 described above is used here. Figures 5(a) and (b) show the reconstructed camera trajectories and terrain map. The camera focal lengths changed little when taking the sequence. Figure

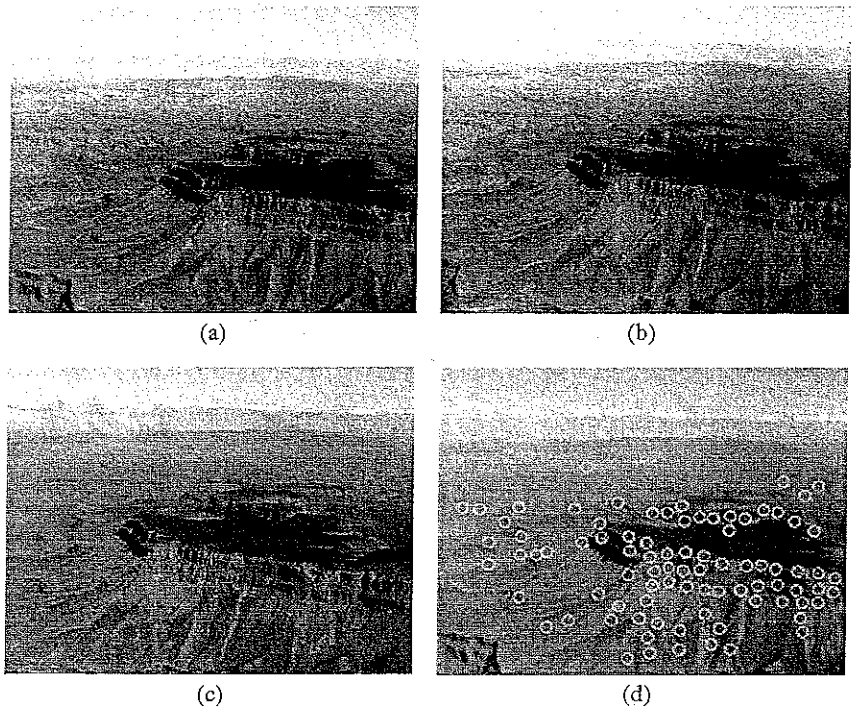



*Figure 4 Grand Canyon sequence input: (a) 1st image, (b) 46th image, (c) 91st image of the Grand Canyon sequence (d) 1st image of the Grand Canyon sequence with the feature points overlaid*

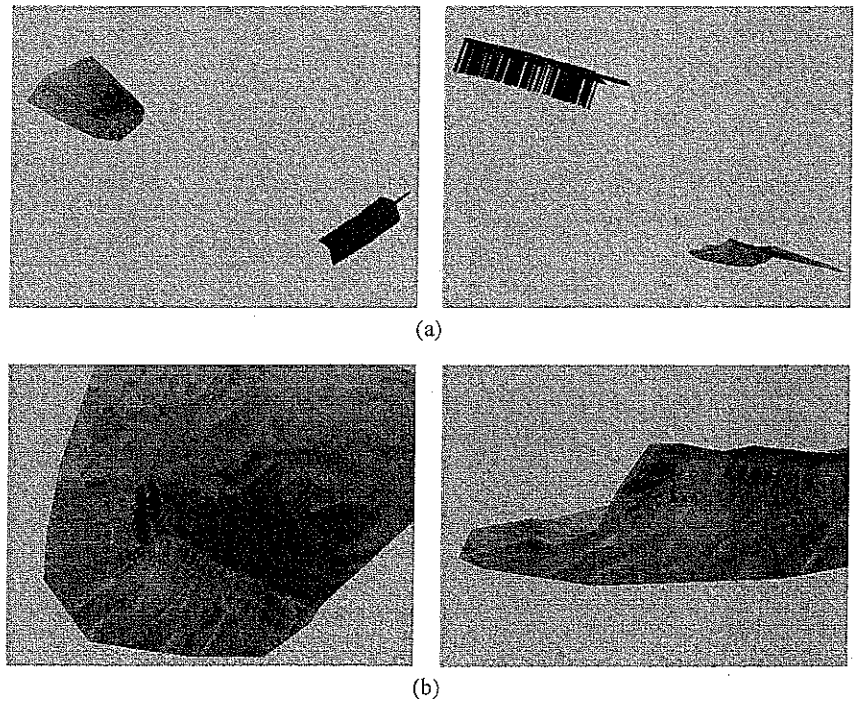



*Figure 5 Grand Canyon sequence results: (a) Top and side view of the reconstruction; the 3-axis figures denote the recovered cameras (b) Top and side view of the reconstructed Grand Canyon with texture mapping*

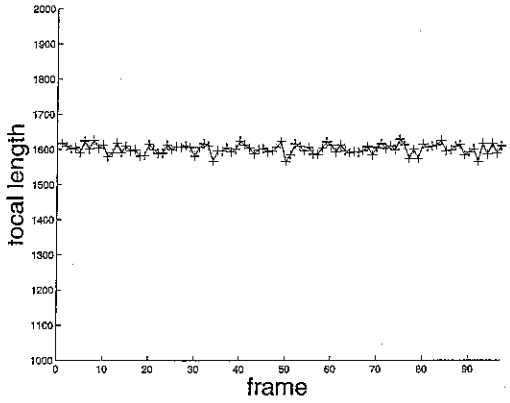



*Figure 6. Grand Canyon sequence: Focal lengths of the Grand Canyon sequence recovered by the perspective factorization method. The recovered values are relatively constant, as expected.*

6 is a plot of the recovered focal lengths, and shows that the focal lengths are relatively constant. The principal point recovered by our method is (159, 119) (with the image size of $320 \times 240$).

## Real Data 3: Calibration Set-Up

In this experiment we test our method on a set-up for multi-camera calibration. In this set-up 51 cameras are placed in a dome, and a bar of LEDs is moved around under the dome. The bar is imaged by each camera as it is moved through a series of known positions. Since the intrinsic parameters of each camera do not change as the bar is moved, the images taken by one camera are combined into one image containing multiple bars. This composite image includes 232 feature points (LED positions). Therefore, the set-up generates 51 images; each contains 232 features, which are to be used as calibration data for the cameras. Tsai's calibration algorithm[39] is used on this set-up to calibrate the 51 cameras. The calibration results of Tsai's algorithm are compared with the results of the perspective factorization method.

In this example we assume that all the intrinsic parameters (except the skews) are unknown, and may differ from camera to camera. The normalization algorithm for case 3 is applied. We initialize the aspect ratios as 1 and the principal points at the middle of the images. Figure 7 shows the reconstructed LED positions and the reconstructed camera orientations and locations. The reconstructed LED positions are compared with their known positions. The maximum distance is 20 mm, which is about 0.61% of the bar length. The recovered camera locations and orientations are compared with Tsai's calibration results. The maximum distance between the recovered camera locations by the two methods is 32 mm which is about 0.98% of the bar length, the maximum angle between the recovered camera orientations is 0.3°.

Figure 8 are plots of the differences of the focal lengths, the principal points and the aspect ratios recovered by the perspective factorization method and by Tsai's calibration algorithm. The plots show that the calibration results of these two methods are very close.

Obtaining the ground truth is difficult and time-consuming in camera calibration. This example demonstrates a good calibration method for multi-camera systems. Instead of carefully putting objects at accurate positions, a person can wave a stick with LEDs randomly in the room. The LEDs enable fast and easy computation of correspondences. Given these tracked feature points, the perspective factorization method can

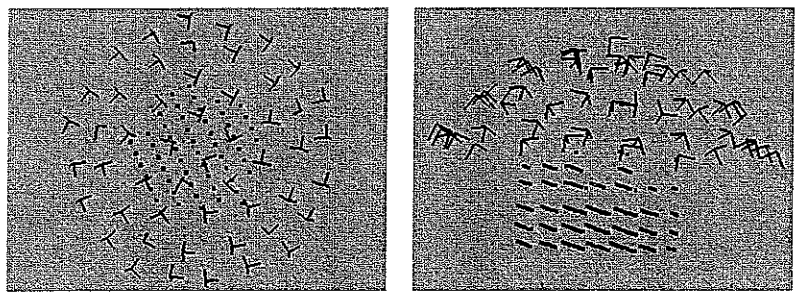
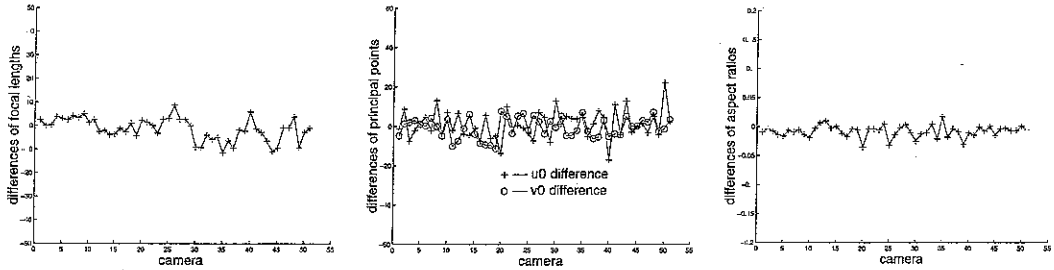


*Figure 7. Calibration set-up results: Top and side view of the reconstruction of the calibration set-up; the points denote the recovered LED positions; the 3-axis figures are the recovered cameras.*




(a) (b) (c)

*Figure 8 Calibration set-up: Differences of (a) the focal lengths, (b) the principal points ($u_0$, $v_0$) and (c) the aspect ratios of the calibration set-up data recovered by the perspective factorization method and by Tsai's calibration algorithm*

be applied to recover the camera extrinsic and intrinsic parameters simultaneously.

## Conclusion

Given image sequences taken with uncalibrated cameras, the perspective factorization method creates 3D models of the scene and recovers the extrinsic and intrinsic parameters of the cameras simultaneously. The reconstruction process consists of two steps: first, an iterative bilinear factorization method is applied to the *measurement matrix*, which is composed of the image locations of all the feature points. The output of this step is the *scaled measurement matrix*, which is the product of the projective motion and shape; second, the factorization-based *normalization* is performed on the scaled measurement matrix, which imposes metric constraints on the projective reconstruction to recover the projective transformation (matrix $H$) and generate the Euclidean shape and motion. In comparison, the normalization in orthographic and weak perspective factorization methods[1,2] is applied directly to the measurement matrix, which imposes similar metric constraints to recover the affine transformation.

The perspective factorization method is able to build 3D models from image sequences taken with one or multiple uncalibrated cameras. In this paper we show the results of applying this method to indoor object modelling, outdoor scene recovery and multi-camera calibration. The results are promising.

#### ACKNOWLEDGEMENTS

We would like to thank Simon Baker, Daniel Morris and Teck Khim Ng for fruitful discussions.

## References

1 Tomasi C, Kanade T. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision* 1992; **9**(2): 137–154.

2 Poelman C, Kanade T. A paraperspective factorization method for shape and motion recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1997; **19**(3): 206–218.

3 Yu H, Chen Q, Xu G, Yachida M. 3D shape and motion by SVD under higher-order approximation of perspective projection. In *International Conference on Pattern Recognition 96*, 1996; A80.22.

4 Christy S, Horaud R. Euclidean reconstruction: from paraperspective to perspective. In *European Conference on Computer Vision 96*, 1996; II: 129–140.

5 Christy S, Horaud R. Euclidean shape and motion from multiple perspective views by affine iterations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1996; **18**(11): 1098–1104.

6 Ma J, Ahuja N. Dense shape and motion from region correspondences by factorization. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 98*, 1998; 219–224.

7 Sturm P. Algorithms for plane-based pose estimation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 00*, 2000; I: 706–711.

8 Faugeras OD. What can be seen in three dimensions with an uncalibrated stereo rig? In *European Conference on Computer Vision 92*, 1992; 563–578.

9 Mohr R, Quan L, Veillon F. Relative 3D reconstruction using multiple uncalibrated images. *International Journal of Robotics Research* 1995; **14**(6): 619–632.

10 Triggs B. Matching constraints and the joint image. In *International Conference on Computer Vision 95*, 1995; 338–343.

11 Quan L. Invariants of 6 points and projective reconstruction from 3 uncalibrated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1995; **17**(1): 34–46.

12 Quan L. Self-calibration of an affine camera from multiple views. *International Journal of Computer Vision* 1996; **19**(1): 93–105.

13 Beardsley PA, Torr PHS, Zisserman A. 3D model acquisition from extended image sequences. In *European Conference on Computer Vision 96*, 1996; II: 683–695
14 Carlsson S, Weinshall D. Dual computation of projective shape and camera positions from multiple images. *International Journal of Computer Vision* 1998; **27**(3): 227–241
15 Hartley RI. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision* 1997; **22**(2): 125–140.
16 Hartley RI. Computation of the quadrifocal tensor. In *European Conference on Computer Vision 98*, 1998; 20–35.
17 Shashua A, Avidan S. The rank 4 constraint in multiple ($\geq 3$) view geometry. In *European Conference on Computer Vision 96*, 1996; 196–206.
18 Sturm P, Triggs B. A factorization based algorithm for multi-image projective structure and motion. In *European Conference on Computer Vision 96*, 1996; II: 709–720
19 Triggs B. Factorization methods for projective structure and motion. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 96*, 1996; 845–851
20 Heyden A. Reduced multilinear constraints: theory and experiments. *International Journal of Computer Vision* 1998; **30**(1): 5–26.
21 Heyden A. Projective structure and motion from image sequences using subspace methods. In *Scandinavian Conference on Image Analysis 97*, 1997.
22 Heyden A, Berthilsson R, Sparr G. An iterative factorization method for projective structure and motion from image sequences. *Image and Vision Computing* 1999; **17**(13): 981–991.
23 Mahamud S, Hebert M. Iterative projective reconstruction from multiple views. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 00*, 2000; II: 430–437
24 Hartley RI. Euclidean reconstruction from uncalibrated views. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 94*, 1994; 908–912.
25 Heyden A, Åström K. Euclidean reconstruction from constant intrinsic parameters. In *International Conference on Pattern Recognition 96*, 1996; A8m 4.
26 Heyden A, Åström K. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 97*, 1997; 438–443
27 Triggs B. Autocalibration and the absolute quadric. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 97*, 1997; 609–614
28 Pollefeys M, Koch R, VanGool L. Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. *International Journal of Computer Vision* 1999; **32**(1): 7–25.
29 de Agapito L, Hartley RI, Hayman E. Linear self-calibration of a rotating and zooming camera. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 99*, 1999; 15–21
30 Ponce J. On computing metric upgrades of projective reconstructions under the rectangular pixel assumption. In *SMILE 2000 Workshop on 3D Structure from Multiple Images of Large-Scale Environments*, 2000; 52–67
31 Shum HY, Ikeuchi K, Reddy R. Principal component analysis with missing data and its application to polyhedral object modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1995; **17**(9): 854–867
32 Urban M, Pajdla T, Hlavac V. Projective reconstruction from n views having one view in common. In *ICCV Workshop Vision Algorithms: Theory and Practice*, 1999; 116–130.
33 Rother C, Carlsson S. Linear multi view reconstruction and camera recovery. In *International Conference on Computer Vision 01*, 2001; I: 42–49
34 Jacobs DW. Linear fitting with missing data: applications to structure from motion and to characterizing intensity images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 97*, 1997; 206–212
35 Sturm P. Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 97*, 1997; 1100–1105
36 Kahl F, Triggs B, Astrom K. Critical motions for autocalibration when some intrinsic parameters can vary. *Journal of Mathematical Imaging and Vision* 2000; **13**(2): 131–146
37 Sturm PF. Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length. *Image and Vision Computing* 2002; **20**(5–6): 415–426
38 Tsai RY. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Transactions on Robotics and Automation* 1987; **3**(4): 323–344