Scene Reconstruction from Multiple Uncalibrated Views

Mei Han Takeo Kanade

January 2000
CMU-RI-TR-00-09

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

We describe a factorization-based method to reconstruct Euclidean shape and motion 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. We present three factorization-based normalization algorithms to generate the Euclidean reconstruction and the intrinsic parameters, assuming zero skews. 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. Experimental results show that out method is efficient and reliable.

Keywords: structure from motion, self-calibration, Euclidean reconstruction

1 Introduction

The problem of recovering shape and motion from an image sequence has received a lot of attention. Previous approaches include recursive methods (e.g., [11, 2]) and batch methods (e.g., [15, 13, 4]). The factorization method, first developed by Tomasi and Kanade [15] for orthographic views and extended by Poelman and Kanade [13] to weak and para perspective views, achieves its robustness and accuracy by applying the singular value decomposition (SVD) to a large number of images and feature points. Christy and Horaud [4, 5] described a method for the perspective camera model by incrementally performing reconstructions with either a weak or a para perspective camera model. 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 object, it is only possible to compute a reconstruction up to an unknown projective transformation [6]. There has been considerable progress on projective reconstruction ([6, 3, 12, 16]). Triggs proposed a projective factorization method in [17] which recovers projective depths by estimating a set of fundamental matrices to chain all the images together. Heyden [8, 9] presented methods of using multilinear subspace constraints to perform projective structure from motion.

In order to obtain a Euclidean reconstruction from the projective reconstruction some additional information about either the camera or the object is needed . Hartley recovered the Euclidean shape using a global optimization technique, assuming the intrinsic parameters are constant [7]. In [10] Heyden and Åström used a bundle adjustment algorithm to estimate the focal lengths, the principal points, the camera motion and the object shape together. Pollefeys et al. assumed that the focal length is the only varying intrinsic parameter and presented a linear algorithm [14]. Agapito et al. proposed a linear self-calibration algorithm for rotating and zooming cameras [1].

In this paper we describe a factorization-based method which recovers Euclidean shape and motion from multiple uncalibrated perspective views. Given tracked feature points, our method reconstructs the object shape, the camera motion and the intrinsic parameters (assuming zero skews). We first present an iterative algorithm to get a projective reconstruction using a bilinear factorization method. We 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. Experiments on synthetic and real images show that our method is efficient compared with the bundle adjustment method and reliable under noises.

2 Projective reconstruction

Suppose there are n perspective cameras: P_i, $i = 1 \cdots n$ and m object points $\mathbf{x}_j$, $j = 1 \cdots m$ represented by homogeneous coordinates. The image coordinates are represented by (u_{ij}, v_{ij}). The following hold

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

where λ_{ij} is a non-zero scale factor, commonly called projective depth. The equivalent matrix form is:

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

where W_s is the *scaled measurement matrix*. We propose the following projective factorization algorithm which iteratively applies factorization to the current scaled measurement matrix.

Iterative 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 rank4 factorization on W_s, generate the projective shape and motion;
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 λ_{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 (λ_{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. Experiments show that the iteration is very stable even starting with arbitrary initial depths. The convergence speed can be improved by first applying a weak perspective factorization method to get the initial values of the λ_{ij}'s. Using rough knowledge of the focal lengths and assuming that the other intrinsic parameters have generic values, we transform the tracked feature points to calibrated correspondences on which the weak perspective factorization method is applied [13]. We have,

$$\lambda_{ij} = \mathbf{k}_i \cdot \mathbf{s}_j + t_{zi} \tag{3}$$

where $\mathbf{s}_j$ denotes the jth point position, $\mathbf{k}_i$ and t_{zi} denote the motion of the ith camera: $\mathbf{k}_i$ is the third row of the rotation matrix and t_{zi} is the translation of the ith camera in z direction.

3 Euclidean reconstruction

The factorization of Equation (2) recovers the motion and shape up to a 4×4 linear projective transformation H:

$$W_s = \hat{P}\hat{X} = \hat{P}HH^{-1}\hat{X} = PX \tag{4}$$

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

Assuming 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.

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

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

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

The first linear algorithm works for the situations that the camera experiences obvious zooming in/out during the sequence. The focal lengths therefore are the main concerns of the reconstruction process. The second linear algorithm works for the situations that the camera changes the focal lengths only a little so that there is no obvious zooming effect and the principal point is very close to being constant. The aerial image sequences taken by a flying platform are examples of this case. The third algorithm, which is bilinear, works for the situations that multiple cameras are included. The focal lengths, the principal points and the aspect ratios are different. The algorithm takes advantage of the relative stable values of the principal points and the aspect ratios, therefore, it converges within a few iterations in our experiments.

3.1 Normalization algorithm outline

The projective motion matrix P_i is:

$$P_i \sim K_i \left[R_i | \mathbf{t}_i \right] \tag{5}$$

where

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

The upper triangular calibration matrix K_i encodes the intrinsic parameters of the ith camera: f_i represents the focal length, (u_{0i}, v_{0i}) is the principal point and α_i is the aspect ratio. R_i is the ith rotation matrix with $\mathbf{i}_i$, $\mathbf{j}_i$ and $\mathbf{k}_i$ denoting the rotation axes. $\mathbf{t}_i$ is the ith translation vector. Combining Equation (5) for $i = 1 \cdots n$ into one matrix equation, we get,

$$P = [M|T] \tag{6}$$

where

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

and

$$\begin{aligned} &\mathbf{m}_{xi} = \mu_i f_i \mathbf{i}_i + \mu_i u_{0i} \mathbf{k}_i \quad \mathbf{m}_{yi} = \mu_i \alpha_i f_i \mathbf{j}_i + \mu_i v_{0i} \mathbf{k}_i \quad \mathbf{m}_{zi} = \mu_i \mathbf{k}_i \\ &T_{xi} = \mu_i f_i t_{xi} + \mu_i u_{0i} t_{zi} \quad T_{yi} = \mu_i \alpha_i f_i t_{yi} + \mu_i v_{0i} t_{zi} \quad T_{zi} = \mu_i t_{zi} \end{aligned} \tag{7}$$

The shape matrix is represented by:

$$X \sim \begin{bmatrix} S \\ \mathbf{1} \end{bmatrix} \tag{8}$$

where

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

and

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

We put the origin of the world coordinate system at the center of gravity of the scaled object points to enforce

$$\sum_{j=1}^{m} \nu_j \mathbf{s}_j = 0 \tag{9}$$

We get,

$$\sum_{j=1}^{m} \lambda_{ij} u_{ij} = \sum_{j=1}^{m} (\mathbf{m}_{xi} \cdot \nu_j \mathbf{s}_j + \nu_j T_{xi}) = \mathbf{m}_{xi} \cdot \sum_{j=1}^{m} \nu_j \mathbf{s}_j + T_{xi} \sum_{j=1}^{m} \nu_j = T_{xi} \sum_{j=1}^{m} \nu_j \tag{10}$$

Similarly,

$$\sum_{j=1}^{m} \lambda_{ij} v_{ij} = T_{yi} \sum_{j=1}^{m} \nu_j \quad \sum_{j=1}^{m} \lambda_{ij} = T_{zi} \sum_{j=1}^{m} \nu_j \tag{11}$$

Define the 4×4 projective transformation H as:

$$H = [A|B] \tag{12}$$

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

$$[M|T] = \hat{P}\,[A|B] \tag{13}$$

we have,

$$T_{xi} = \hat{P}_{xi} B \quad T_{yi} = \hat{P}_{yi} B \quad T_{zi} = \hat{P}_{zi} B \tag{14}$$

From Equations (10) and (11) we know,

$$\frac{T_{xi}}{T_{zi}} = \frac{\sum_{j=1}^{m} \lambda_{ij} u_{ij}}{\sum_{j=1}^{m} \lambda_j} \quad \frac{T_{yi}}{T_{zi}} = \frac{\sum_{j=1}^{m} \lambda_{ij} v_{ij}}{\sum_{j=1}^{m} \lambda_j} \tag{15}$$

we set up $2n$ linear equations of the 4 unknown elements of the matrix B. Linear least squares solutions are then computed.

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 (7):

$$\begin{array}{rcl} |\mathbf{m}_{xi}|^2 & = & \mu_i^2 f_i^2 + \mu_i^2 u_{0i}^2 \\ |\mathbf{m}_{yi}|^2 & = & \mu_i^2 \alpha_i^2 f_i^2 + \mu_i^2 v_{0i}^2 \\ |\mathbf{m}_{zi}|^2 & = & \mu_i^2 \\ \mathbf{m}_{xi} \cdot \mathbf{m}_{yi} & = & \mu_i^2 u_{0i} v_{0i} \\ \mathbf{m}_{xi} \cdot \mathbf{m}_{zi} & = & \mu_i^2 u_{0i} \\ \mathbf{m}_{yi} \cdot \mathbf{m}_{zi} & = & \mu_i^2 v_{0i} \end{array} \tag{16}$$

Based on the three different assumptions of the intrinsic parameters (three cases), we translate the above constraints to linear constraints on MM^{T} (see Section 3.2, 3.3 and 3.4 for details). Since

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

we can compute least squares solutions for the 10 unknown elements of the symmetric 4×4 matrix $Q = AA^{\mathrm{T}}$. Then we get the matrix A from Q by rank3 matrix decomposition.

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

$$\mu_i = |\mathbf{m}_{zi}| \tag{18}$$

We then compute the principal points (if applied)

$$u_{0i} = \frac{\mathbf{m}_{xi} \cdot \mathbf{m}_{zi}}{\mu_i^2} \quad v_{0i} = \frac{\mathbf{m}_{yi} \cdot \mathbf{m}_{zi}}{\mu_i^2} \tag{19}$$

and the focal lengths as

$$f_i = \frac{\sqrt{|\mathbf{m}_{xi}|^2 - \mu_i^2 u_{0i}^2}}{\mu_i} \tag{20}$$

The aspect ratios (if applied) are

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

Therefore, the motion parameters are

$$\begin{array}{lll} \mathbf{k}_i = \dfrac{\mathbf{m}_{zi}}{\mu_i} & \mathbf{i}_i = \dfrac{\mathbf{m}_{xi} - \mu_i u_{0i}\mathbf{k}_i}{\mu_i f_i} & \mathbf{j}_i = \dfrac{\mathbf{m}_{yi} - \mu_i v_{0i}\mathbf{k}_i}{\mu_i \alpha_i f_i} \\ t_{zi} = \dfrac{T_{zi}}{\mu_i} & t_{xi} = \dfrac{T_{xi} - \mu_i u_{0i} t_{zi}}{\mu_i f_i} & t_{yi} = \dfrac{T_{yi} - \mu_i v_{0i} t_{zi}}{\mu_i \alpha_i f_i} \end{array} \tag{22}$$

The normalization process is summarized by the following algorithm.

Normalization Algorithm

1. Perform SVD on W_s and get $\hat{P}$ and $\hat{X}$;
2. Sum up each row of W_s and compute the ratios between them as in Equation (15);

3. Set up $2n$ linear equations of the 4 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 compute Q;
5. Factorize Q to get A from $Q = AA^{\mathrm{T}}$;
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 motion matrix using $P = \hat{P}H$;
8. Recover the intrinsic parameters, the rotation axes and the translation vectors according to Equations (19)–(22).

3.2 Case 1: Unknown focal lengths

Assuming that the focal lengths are the only unknown intrinsic parameters, we have

$$u_{0i} = 0 \quad v_{0i} = 0 \quad \alpha_i = 1 \tag{23}$$

We combine the constraints in Equation (16) to impose the following linear constraints on 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 $\mu_1 = 1$:

$$|\mathbf{m}_{z1}|^2 = 1 \tag{24}$$

Totally 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 (20) and (21):

$$f_i = \frac{|\mathbf{m}_{xi}| + |\mathbf{m}_{yi}|}{2\mu_i} \tag{25}$$

3.3 Case 2: Unknown focal lengths and constant principal point

Assuming that the focal lengths are unknown and the principal point is constant, that is,

$$u_{0i} = u_0 \quad v_{0i} = v_0 \quad \alpha_i = 1 \tag{26}$$

We translate the constraints in Equation (16) to the following constraints.

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

and

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

where $j = i + 1, \text{if } i \neq n; j = 1, \text{if } i = n$. 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 symmetric matrix $Q_1 = \mathbf{q}\mathbf{q}^{\mathrm{T}}$, where $\mathbf{q}$ is a 10×1 vector composed of the 10 unknown elements of the matrix Q. Therefore, we get $7n+1$ linear equations of the 55 unknown elements of the matrix $Q1$.

Once Q_1 has been computed, $\mathbf{q}$ is generated by rank1 matrix decomposition of Q_1. We then put the 10 elements of $\mathbf{q}$ into a symmetric 4×4 matrix Q which is factorized as AA^{T}.

We compute the principal point as the average of Equation (19):

$$u_0 = \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{m}_{xi} \cdot \mathbf{m}_{zi}}{\mu_i^2}$$

$$v_0 = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{m}_{yi}\cdot\mathbf{m}_{zi}}{\mu_i^2} \tag{30}$$

and the focal lengths as the average of Equations (20) and (21):

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

3.4 Case 3: Unknown focal lengths, principal points and aspect ratios

Assuming that the focal lengths, the principal points and the aspect ratios are all unknown and varying, we represent the constraints in Equation (16) 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 ratios, we impose linear constraints on the elements of the matrix Q:

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

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 motion matrix is $P = \hat{P}H$. We compute the refined principal points, the current recovered focal lengths and the refined aspect ratios according to Equations (19), (20) and (21) respectively. The current motion parameters are then computed as in Equation (22).

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

4 Experiments

In this section we apply the perspective factorization method to synthetic and real image sequences. Given tracked feature points, we first generate the projective reconstruction as described in Section 2, then choose one of the three normalization algorithms described in Section 3 to get the Euclidean reconstruction and the intrinsic parameters. The three real image sequences demonstrate the three cases respectively. Experimental results on synthetic and real data show that our method is reliable under noises and efficient compared with the nonlinear bundle adjustment method.

4.1 Synthetic data

Synthetic experiments have been conducted to study the performance of the perspective factorization method of its convergence under various motions and data noises, and of the reconstruction accuracy of the object shape, the camera motion and the camera intrinsic parameters. The experimental results show that the method converges reliably even under 5% image noises. The errors of the recovered object shape and the camera locations 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 recovered principal points and the aspect ratios are more accurate than the focal lengths. The errors of the principal points are less than 1.5 pixels and the errors of the aspect ratios are less than 0.5% of the true values.

4.2 Real data 1: Building sequence

This sequence was taken by a hand-held camera in front of a building. The camera was very far from the building at first, then moved close toward the building, and moved away again. The camera was zoomed in when it was far and zoomed out when it was close so that the images of the building have almost the same size in the sequence. The largest focal length is about 3 times of the smallest one according to the rough readings on the camera. The sequence includes 14 frames. Figure 1(a) shows three of them. 50 feature points were manually selected along the building windows and the corners. 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 projective factorization algorithm described in Section 2 and the normalization algorithm described in Section 3.2 to this example.

Figure 1(b) demonstrates the reconstructed building and the camera trajectories. The top view shows that the recovered camera moves from far to close to the building then moves 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 that the same person took the sequence while walking in front of the building. Figure 1(c) 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 of pairs of the expected parallel lines is 0.89° and the average angle between pairs of the expected perpendicular lines is 91.74°. Figure 2 plots the recovered focal lengths, which shows that the focal lengths are changing with the camera motion as we expected.

We compare our results with those of the nonlinear bundle adjustment method. The nonlinear method starts with the weak perspective factorization results as initial values assuming the generic intrinsic parameters and the rough focal lengths, then uses bundle adjustment to refine the reconstruction. Figure 2 shows the focal lengths recovered by our method and by the nonlinear method. Table 1 lists the computation cost (in terms of the number of float operations) of the two methods. The comparisons show that the focal lengths recovered by the two methods are very close while our method is more efficient.

4.3 Real data 2: Grand Canyon sequence

The second example is an aerial image sequence taken from a small airplane 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. Two frames from the sequence are shown in Figure 3(a). We assume that the focal lengths and the principal point are unknown and the principal point is fixed over the sequence. The normalization algorithm of Section 3.3 is used here. Figures 3(b) and (c) show the reconstructed camera trajectories and terrain map. Figure 4 is the plot of the recovered focal lengths by our method and by the nonlinear method. The camera focal lengths changed little when taking the sequence. The plot shows the focal lengths recovered by the two

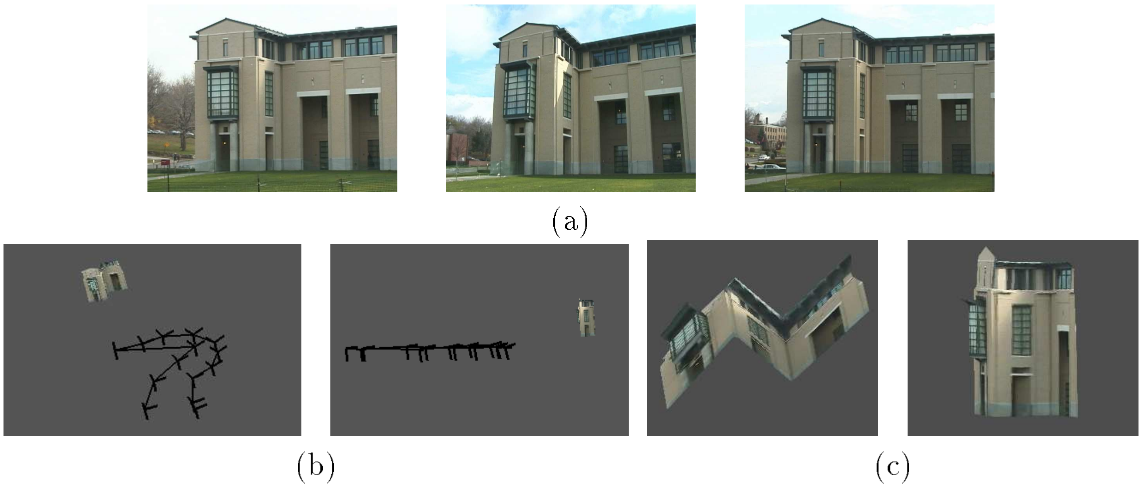


Figure 1: (a) 1st, 9th and 14th images of the building sequence. (b)Top and side view of the reconstruction, the 3-axis figures denote the recovered cameras. The top view shows that the recovered camera moves from far to close to the building, then moves away again as we expected. The side view shows that the recovered locations of the cameras are at the same height and the orientations are tilted upward. (c)Bottom and side view of the reconstructed building with texture mapping.

methods are both relatively constant. The principal point recovered by our method is $(159, 119)$ and $(157, 118)$ by the nonlinear method. Table 1 lists the comparison of computation cost of our method and the nonlinear method. It shows that the perspective factorization method is more efficient than the nonlinear method.

4.4 Real data 3: Calibration setup

In this experiment we test our method on a setup for multi-camera calibrations. The setup includes 51 cameras arranged in a dome and a bar of LEDs which is moved around under the dome. Each camera takes video sequences of the moving bar, and the images taken by one camera are combined into one image of the bar at different positions. The bar is put at known positions. Therefore, the setup generates 51 images which are used as calibration data for the cameras. Tsai's calibration algorithm [19] is used on this setup to calibrate the 51 cameras. The calibration results of Tsai's algorithm are compared with the results of our method. 232 feature points (LED positions)

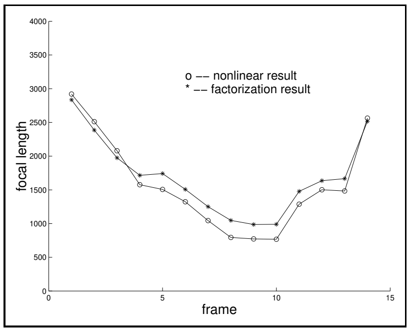
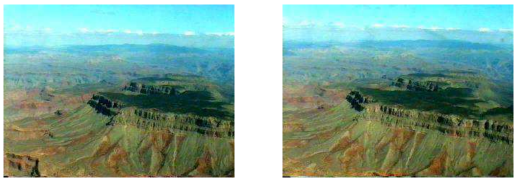
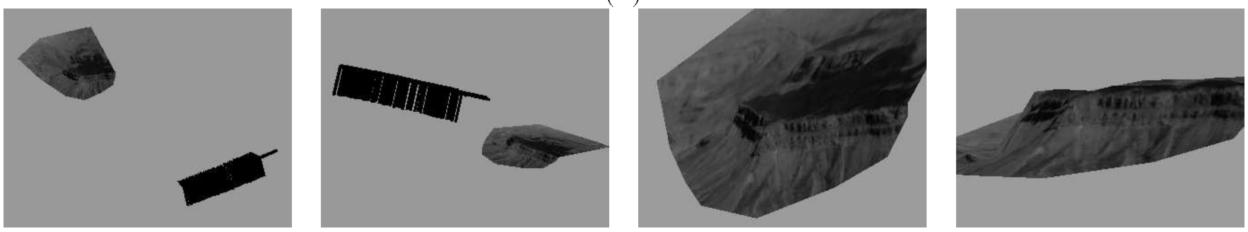


Figure 2: Focal lengths of the building sequence recovered by the perspective factorization method and by the nonlinear method. The recovered values are changing with the camera motion as expected.



(a)



(b) (c)

Figure 3: (a) 1st and 91st images of the Grand Canyon sequence. (b)Top and side view of the reconstruction, the 3-axis figures denote the recovered cameras. (c)Top and side view of the reconstructed Grand Canyon with texture mapping.

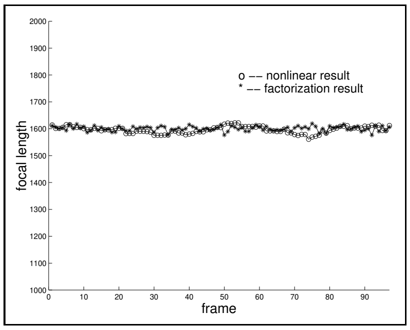


Figure 4: Focal lengths of the Grand Canyon sequence recovered by the perspective factorization method and the nonlinear method. The recovered values are both relatively constant as expected.

are included.

In this example we assume that all the intrinsic parameters (except the skews) are unknown and different from camera to camera. The normalization algorithm described in Section 3.4 is applied. We take the generic values of the principal points (the middle of the images) and the aspect ratios (1) as the initial values for the bilinear normalization algorithm. Figure 5 shows the reconstructed LED positions and the camera orientations and locations. The reconstructed LED positions are compared with their known positions. The maximal distance is 20mm which is about 0.61 percent of the object size. The recovered camera locations and orientations are compared with Tsai's calibration results. The maximal distance between the recovered camera locations by the two methods is 32mm which is about 0.98 percent of the object size, the maximal angle between the recovered camera orientations is 0.3°.

Figure 6 are the 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 our method are very close to those of Tsai's algorithm. Table 1 compares the computation cost of our method and the nonlinear bundle adjustment method. It shows that our method is more efficient.

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 one stick randomly in the room. The stick has marks

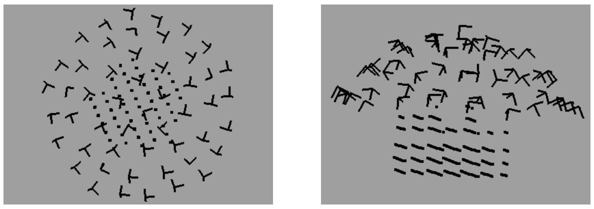
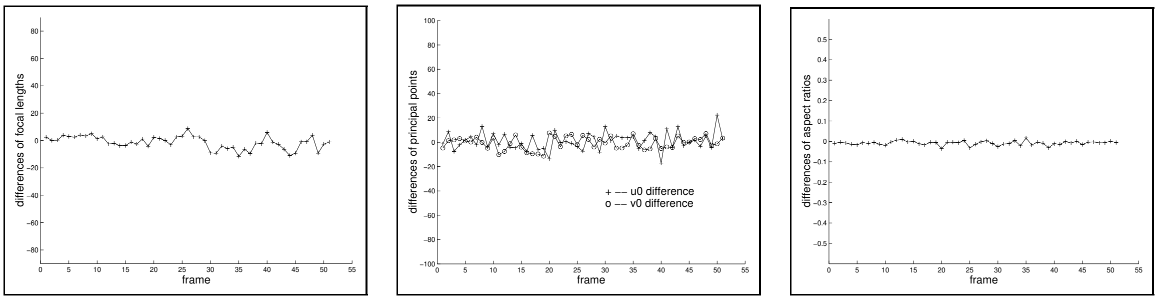


Figure 5: Top and side view of the reconstruction of the calibration setup, the points denote the recovered LED positions, the 3-axis figures are the recovered cameras.



(a) (b) (c)

Figure 6: Differences of (a)the focal lengths (b) the principal points (u_0, v_0) (c) the aspect ratios of the calibration setup data recovered by the perspective factorization method and by Tsai's calibration algorithm.

which enable the fast and easy computation of correspondences. Given these tracked feature points, the perspective factorization algorithm is applied to recover the camera extrinsic and intrinsic parameters simultaneously.

5 Discussion

Given tracked feature points from multiple uncalibrated perspective views, we recover the object shape, the extrinsic camera parameters and the intrinsic parameters simultaneously. We first perform a projective reconstruction using a bilinear factorization algorithm, then generate the Euclidean reconstruction from the projective solution and recover the focal lengths, the principal points and the aspect ratios using metric constraints. We present three factorization-based normalization algorithms dealing with different assumptions on the intrinsic parameters. The algorithm for the case that the focal

Image sequence	Comparison	Perspective factorization method		Nonlinear method
		Projective factorization	Normalization	
Building	Iterations	18	1	27
	Computation cost	2,926k	1,613k	15,745k
Grand Canyon	Iterations	38	1	43
	Computation cost	28,558	1,771k	834,991k
Calibration setup	Iterations	16	4	34
	Computation cost	285,579k	157,170k	1,765,700k

Table 1: Computation cost comparisons of the perspective factorization method and the nonlinear method. Computation cost in the table is the number of float operations per iteration which are computed according to: $22m^3 + 18mn^2$ for each projective iteration, $20n^3$ for each normalization iteration, and $27m^3 + 189m^2n + 41mn^2 + 343n^3$ for each nonlinear iteration, where n is the number of frames and m is the number of points. The perspective factorization method includes iterations for the projective reconstruction and normalization, which is 1 for the linear normalization methods. The results show that our method is more efficient.

lengths are the only unknown intrinsic parameters is linear, as well is the algorithm for the case that the focal lengths and the constant principal point are unknown. The algorithm is bilinear in the case that the focal lengths, the principal points and the aspect ratios are all unknown.

Our projective reconstruction is still a nonlinear parameter-fitting process. It is based on the bilinear relationship of projective depth and linear reconstructions, which provides a compromise between linear methods (up to affine camera models) and nonlinear optimization in terms of reconstruction quality and computation cost. Our method achieves reliability and accuracy by uniformly considering all the data in all the images like in most factorization methods.

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 [18, 14]. 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 center of gravity of the object and aligning its orientation with the first camera. Compared with the method presented by Pollefeys et al. in [14], the normalization algorithm described in Section 3.2 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 Section 3.3 and 3.4).

References

[1] L. de Agapito, R.I. Hartley, and E. Hayman. Linear self-calibration of a rotating and zooming camera. In *CVPR99*, 1999.

[2] A. Azarbayejani and A.P. Pentland. Recursive estimation of motion, structure, and focal length. *PAMI*, 17(6):562–575, June 1995.

[3] P.A. Beardsley, P.H.S. Torr, and A. Zisserman. 3d model acquisition from extended image sequences. In *ECCV96*, pages II:683–695, 1996.

[4] S. Christy and R. Horaud. Euclidean reconstruction: From paraperspective to perspective. In *ECCV96*, pages II:129–140, 1996.

[5] S. Christy and R. Horaud. Euclidean shape and motion from multiple perspective views by affine iterations. *PAMI*, 18(11):1098–1104, November 1996.

[6] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *ECCV92*, pages 563–578, 1992.

[7] R.I. Hartley. Euclidean reconstruction from uncalibrated views. In *CVPR94*, pages 908–912, 1994.

[8] A. Heyden. Projective structure and motion from image sequences using subspace methods. In *SCIA97*, 1997.

[9] A. Heyden. Reduced multilinear constraints: Theory and experiments. *IJCV*, 30(1):5–26, October 1998.

[10] A. Heyden and K. Astrom. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *CVPR97*, pages 438–443, 1997.

[11] P.F. McLauchlan, I.D. Reid, and D.W. Murray. Recursive affine structure and motion from image sequences. In *ECCV94*, volume 1, pages 217–224, 1994.

[12] R. Mohr, L. Quan, and F. Veillon. Relative 3d reconstruction using multiple uncalibrated images. *IJRR*, 14(6):619–632, December 1995.

[13] C. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. *PAMI*, 19(3):206–218, 1997.

[14] M. Pollefeys, R. Koch, and L. VanGool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *ICCV98*, pages 90–95, 1998.

[15] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *IJCV*, 9(2):137–154, 1992.

[16] B. Triggs. Matching constraints and the joint image. In *ICCV95*, pages 338–343, 1995.

[17] B. Triggs. Factorization methods for projective structure and motion. In *CVPR96*, pages 845–851, 1996.

[18] B. Triggs. Autocalibration and the absolute quadric. In *CVPR97*, pages 609–614, 1997.

[19] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *RA*, 3(4):323–344, 1987.