Euclidean Reconstruction from Image Sequences with Varying and Unknown Focal Length and Principal Point*

Anders Heyden, Kalle Åström
Dept of Mathematics, Lund University
Box 118, S-221 00 Lund, Sweden

E-mail: {heyden,kalle}@maths.lth.se

Abstract

In this paper the special case of reconstruction from image sequences taken by cameras with skew equal to 0 *and aspect ratio equal to* 1 *has been treated. These type of cameras, here called cameras with Euclidean image planes, represent rigid projections where neither the principal point nor the focal length is known. It will be shown that it is possible to reconstruct an unknown object from images taken by a camera with Euclidean image plane up to similarity transformations, i.e., Euclidean transformations plus changes in the global scale.*

An algorithm, using bundle adjustment techniques, has been implemented. The performance of the algorithm is shown on simulated data.

1. Introduction

During the last years there has been an intensive research on the possibility to obtain reconstructions up to an unknown similarity transformation (often called *Euclidean reconstruction*), without using fully calibrated cameras. It is a well-known fact that it is only possible to make reconstruction up to an unknown projective transformations (often called *projective reconstruction*) when nothing about the intrinsic parameters, extrinsic parameters or the object is known. Thus it is necessary to have some additional information about either the intrinsic parameters, the extrinsic parameters or the object in order to obtain the desired Euclidean reconstruction.

One common situation is when the intrinsic parameters are constant during the whole (or a part) of the image sequence. This approach leads to the well-known Kruppa equations. These equations are highly nonlinear and difficult to solve numerically. Several attempts to solve this problem have been made, see [6, 2, 5]. In [3] the same problem is solved by a global optimisation technique, where a lot of smaller optimisation problems have to be solved in order to get a starting point for the last optimisation.

Another constraint, called the *modulus constraint* have been used in [8], to obtain Euclidean reconstruction from constant intrinsic parameters. This formalism has been extended to the case when the focal length is varying between the different imaging instants, see [7]. The practical implications of this result is questionable since when the focal length varies, by zooming, the principal point varies also.

The results presented in this paper is motivated by this fact, that when a CCD-camera is used in order to capture an image sequence and the zoom is used, as in active vision, both the focal length and the principal point varies. However, it is often the case that the aspect ratio is equal to 1 and the skew is equal to 0. This particular case of camera will be called a *camera with Euclidean image plane* and represents a rigid perspective transformation from 3D Euclidean space to a 2D Euclidean space, where neither the principal point nor the focal distance is known.

An interesting application of this model is reconstruction from X-ray images used in medical investigations. The X-ray images are rigid and can not deform affinely. This means that the skew is equal to 0 and that the aspect ratio is equal to 1, that is the camera has Euclidean image plane. It is furthermore not possible to assume that the principal point is located approximately in the center of the image, since the centre of projection is determined by the position of the X-ray source and the orientation of the photographic plate, which can be freely moved around.

In this paper it is shown theoretically that Euclidean reconstruction is possible even when the focal length and principal point are unknown and varying. The proof is based on the assumption of generic camera motion and known skew and aspect ratio. However, if the camera motion is not

*This work has been done within the ESPRIT Reactive LTR project 21914, CUMULI and the Swedish Research Council for Engineering Sciences (TFR), project 95-64-222

sufficiently general, e.g. pure translation or circular motion, then this is not possible. The theoretical result is verified by experiments on simulated data both for general and restricted camera motion.

2. The camera model

The image formation system (the camera) is modeled by the equation

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & fs & x_0 \\ 0 & \gamma f & y_0 \\ 0 & 0 & 1 \end{bmatrix} [R \,|\, -Rt] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad \Leftrightarrow \qquad (1)$$
$$\lambda \mathbf{x} = K[R \,|\, -Rt]\mathbf{X} = P\mathbf{X} \ .$$

Here $\mathbf{X} = [X\,Y\,Z\,1]^T$ denotes object coordinates in extended form and $\mathbf{x} = [x\,y\,1]^T$ denotes extended image coordinates. The scale factor λ, called the **depth**, accounts for perspective effects and (R,t) represent a rigid transformation of the object, i.e. R denotes a 3×3 rotation matrix and t a 3×1 translation vector. Finally, the parameters in K represent intrinsic properties of the image formation system: f represents **focal length**, γ represents the **aspect ratio**, s represents the **skew**, i.e. nonrectangular light sensitive arrays can be modelled, and (x_0, y_0) is called the **principal point** and is interpreted as the orthogonal projection of the focal point onto the image plane. The parameters in R and t are called **extrinsic parameters** and the parameters in K are called the **intrinsic parameters**.

In this paper we will deal with cameras where $s = 0$ and $\gamma = 1$. Then (1) can be written

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{bmatrix} [R \,|\, -Rt] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \ . \qquad (2)$$

Definition 2.1. A camera that can be modeled by (2) is called a **camera with Euclidean image plane**. An internal calibration matrix K of type

$$K = \begin{bmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (3)$$

is called a **Euclidean calibration matrix**. ■

Observe that it is not necessary that $\gamma = 1$ and $s = 0$ in order to use the subsequent results. It is sufficient that they are known, since then they can be compensated for by a suitable change of coordinates.

The following result, shown in [4], will be needed later.

Lemma 2.1. *A camera matrix*

$$P = \left[\begin{array}{c|c} u^T & \\ v^T & t \\ w^T & \end{array} \right] \ , \qquad (4)$$

normalised such that $w.w = 1$, represents a camera with Euclidean image plane, if and only if

$$\begin{aligned} (u \times w).(v \times w) &= 0 \ , \\ (u \times w).(u \times w) &= (v \times w).(v \times w) \ . \end{aligned} \qquad (5)$$

where a.b denotes the scalar product of a and b.

Observe that the condition $w.w = 1$ can easily be fulfilled by multiplying the camera matrix by a suitable constant, since a camera matrix is only defined up to scale.

Now we have the necessary tools to prove that it is possible to obtain a Euclidean reconstruction, when sufficiently many point correspondences are given in a sufficient number of images.

3. Euclidean reconstruction is possible

For a moment, we do not take into account the special form of the camera matrices, (2), for cameras with Euclidean image planes, and instead work with totally uncalibrated cameras, as in (1). Then it is possible to make reconstruction up to an unknown projective transformation. This means that it is possible to calculate camera matrices P_i, $i = 1, \ldots, m$ that fulfils

$$\lambda_i \mathbf{x}_i = P_i \mathbf{X}, \quad i = 1, \ldots, m \ , \qquad (6)$$

where $\mathbf{x}_i$ denotes extended image coordinates in image i and λ_i denotes the corresponding depth in image i. It can easily be seen from (6) that given one such sequence of camera matrices, P_i, $i = 1, \ldots, m$, and a reconstruction, $\mathbf{X}$, also $P_i H$, $i = 1, \ldots, m$ and $H^{-1}\mathbf{X}$ is a possible choice of camera matrices and reconstruction, where H denotes a nonsingular 4×4 matrix. Multiplication of $\mathbf{X}$ by such a matrix corresponds to projective transformations of the object. In our case H can not be chosen arbitrarily since every camera matrix has to obey the conditions in Lemma 2.1.

The next step is to show that given a sequence of camera matrices that solves the projective reconstruction problem and represents cameras with Euclidean image planes, i.e. fulfils the conditions in Lemma 2.1, then the only possible transformations H that preserve these conditions are the ones representing similarity transformations. In order to show this some notations will be introduced.

Denote by $\mathcal{M}_P$ the manifold of all 3×4 projection matrices, i.e., the set of all 3×4 matrices defined up to scale. Denote by $\mathcal{M}_E$ the manifold of all camera matrices that represents cameras with Euclidean image planes, i.e., all 3×4

matrices that can be written as in (2), and thus obeying the conditions in Lemma 2.1. Denote the group of all projective transformations, represented by 4×4 matrices, by $\mathcal{G}_P$. The subclass of transformations that preserves the properties in Lemma 2.1 is denoted by $\mathcal{G}_E$, i.e.

$$\mathcal{G}_E = \{ H \in \mathcal{G}_P \,|\, (P \in \mathcal{M}_E) \Rightarrow PH \in \mathcal{M}_E \} \ . \tag{7}$$

This group represents the ambiguities in reconstruction when using cameras with Euclidean image planes. This group tells us what kind of reconstruction we can get under the assumption of Euclidean image planes and it is our goal to determine this group. Finally, the group of all similarity transformations will be denoted by $\mathcal{G}_S$ and will be represented by

$$\mathcal{G}_S = \{ H = \begin{bmatrix} \lambda R & t \\ 0 & 1 \end{bmatrix} \,|\, RR^T = I, 0 \neq \lambda \in \mathbb{R} \} \ . \tag{8}$$

The group of similarity transformations is contained in $\mathcal{G}_E$ since

$$K[\,R\,|\,-Rt\,] \begin{bmatrix} R' & t' \\ 0 & 1 \end{bmatrix} = K[\,R'R\,|\,Rt' - Rt\,] \in \mathcal{M}_E \ ,$$

for all Euclidean calibration matrices K and all orthogonal R and R'. Thus

$$\mathcal{G}_S \subseteq \mathcal{G}_E \subseteq \mathcal{G}_P \ .$$

Theorem 3.1. *Let $\mathcal{G}_E$ denote the class of transformations in 3D-space that preserves the conditions in Lemma* 2.1 *and $\mathcal{G}_S$ the group of similarity transformations in 3D-space. Then*

$$\mathcal{G}_E = \mathcal{G}_S \ .$$

Proof. From the discussion above we have $\mathcal{G}_S \subseteq \mathcal{G}_E$.

Observe that the constraints on the camera matrices in Lemma 2.1 only involve the first 3×3 submatrix. Use the notation

$$H = \begin{bmatrix} A & b \\ c & d \end{bmatrix} \ ,$$

where A is a 3×3 matrix. Assume that P represents a camera with Euclidean image planes, H a projective transformation and

$$\begin{aligned} PH =& K[\,R\,|\,t\,] \begin{bmatrix} A & b \\ c & d \end{bmatrix} = \\ =& [\,K(RA + tc)\,|\,K(Rb + td)\,] \in \mathcal{M}_E \ . \end{aligned} \tag{9}$$

Then $K(RA+tc)$ can be factorised $K(RA+tc) = K'R'$ where K is a Euclidean calibration matrix and R' denotes an orthogonal matrix. Since (9) is valid for any P that represents a camera matrix, i.e., for any K, R and t, we first study the case $t = 0$.

Assume that A has the property that for every Euclidean calibration matrix K and orthogonal R, it is possible to factorise KRA according to $KRA = K'R'$, for some Euclidean calibration matrix K' and orthogonal R'. Then also UAV has this property for every pair of orthogonal matrices U and V, since

$$KRUAV = KR''AV = K'R'''V = K'R' \ ,$$

where R'' and R''' denotes orthogonal matrices. Now, using the singular value decomposition we can write

$$D_1 = U_1AV_1 = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \ ,$$

and by a simple permutation of the rows and columns in U_1 and V_1 respectively, we can also write

$$D_2 = U_2AV_2 = \begin{bmatrix} b & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & a \end{bmatrix} \ .$$

Replacing A by D_1 and choosing $R = I$ in (9), Lemma 2.1 gives $a = b$ and replacing A by D_2 gives $b = c$. Thus all singular values of A are equal, which means that A is a multiple of an orthogonal matrix.

Consider now the case, where $t \neq 0$, and the condition that for every Euclidean calibration matrix K, every orthogonal R and every t, $K(RA + tc)$ can be factorised as $K(RA + tc) = K'R'$ for some Euclidean calibration matrix K' and orthogonal R'. If $RA + tc$ can be factorised in this way then so can $(RA + tc)V$ for every orthogonal matrix V. Choose V such that $cV = [\,s\,0\,0\,]$, then choose $R = (AV)^{-1}$ and $t = [\,1\,0\,0\,]^T$. These choices gives

$$(RA + tc)V = RAV + tcV =$$
$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} s & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1+s & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \ ,$$

and according to Lemma 2.1, $s = 0$, which in turn implies $c = [\,0\,0\,0\,]$.

Summing up, H is of the form

$$H = \begin{bmatrix} \lambda R & b \\ 0 & d \end{bmatrix} \ ,$$

where λ is a scalar and R an orthogonal matrix. Dividing by d gives $H \in \mathcal{G}_S$. Thus $\mathcal{G}_E \subseteq \mathcal{G}_S$ from which the theorem follows. ∎

We remark that this theorem is valid only under the assumption that the camera motion is sufficiently general. This fact is used implicit in the formulation of the theorem and in the proof, by requiring that $P = K[\,R\,|\,-Rt\,]$ can be chosen arbitrarily.

4. Description of algorithm

A bundle adjustment algorithm was developed for estimating all unknown parameters. This will briefly be described. Let m denote the number of images and n the number of points. Denote by $\mathfrak{m}$ the bundle of all unknown parameters, $\mathfrak{m} = \{P_1, \ldots, P_m, \mathbf{X}_1, \ldots, \mathbf{X}_n\}$. Each such element belongs to a non-linear manifold, $\mathcal{M}$.

Introduce a *local parametrisation* $\mathfrak{m}(\Delta\mathbf{x})$, around $\mathfrak{m}_0 \in \mathcal{M}$ according to

$$\mathcal{M} \times R^N \ni (\mathfrak{m}_0, \Delta\mathbf{x}) \mapsto \mathfrak{m}(\mathfrak{m}_0, \Delta\mathbf{x}) \in \mathcal{M} \quad ,$$

where $N = 9m + 3n$. (9 parameters in each camera matrix describing a camera with Euclidean image plane and 3 parameters for the coordinates of each reconstructed point.) Let $\Delta\mathbf{x} = [\Delta a_1, \ldots, \Delta a_m, \Delta b_1, \ldots, \Delta b_n]^T$, so that Δa_i parametrise changes in camera matrix P_i and Δb_j parametrise changes in reconstructed point X_j. Each camera matrix is written

$$P_i = K_i[R_i \,|\, t_i] \quad .$$

Changes in K_i are parametrised

$$K_i(\mathfrak{m}_0, \Delta\mathbf{x}) = \begin{bmatrix} f + \Delta a_i(1) & 0 & x_0 + \Delta a_i(2) \\ 0 & f + \Delta a_i(1) & y_0 + \Delta a_i(3) \\ 0 & 0 & 1 \end{bmatrix} \quad ,$$

changes in R_i

$$R_i(\mathfrak{m}_0, \Delta\mathbf{x}) = \exp(\begin{bmatrix} 0 & \Delta a_i(6) & -\Delta a_i(5) \\ -\Delta a_i(6) & 0 & \Delta a_i(4) \\ \Delta a_i(5) & -\Delta a_i(4) & 0 \end{bmatrix})R_i \quad ,$$

changes in t_i

$$t_i(\mathfrak{m}_0, \Delta\mathbf{x}) = \begin{bmatrix} t_x + \Delta a_i(7) \\ t_y + \Delta a_i(8) \\ t_z + \Delta a_i(9) \end{bmatrix} \quad .$$

and changes in each object point, $\mathbf{X}_j$,

$$\mathbf{X}_j(\mathfrak{m}_0, \Delta\mathbf{x}) = \begin{bmatrix} X_j + \Delta b_j(1) \\ Y_j + \Delta b_j(2) \\ Z_j + \Delta b_j(3) \\ 1 \end{bmatrix} \quad .$$

Introduce a *residual vector* $\mathbf{Y}$, formed by putting all reprojected errors in a column vector.

These residuals depend on our measured image positions $\mathbf{x}_{ij}$ on our estimated parameters $\mathfrak{m}$.

The residual vector $\mathbf{Y}(\Delta\mathbf{x})$ is a non-linear function of the local parametrisation vector $\Delta\mathbf{x}$. The sum of squared residuals $f = \mathbf{Y}^T\mathbf{Y}$ was minimised with respect to the unknown parameters $\Delta\mathbf{x}$, using the Gauss-Newton method. Linearisation of $\mathbf{Y}(\Delta\mathbf{x})$ gives

$$\mathbf{Y}(\Delta\mathbf{x}) \approx \mathbf{Y}(0) + \frac{\partial\mathbf{Y}}{\partial\Delta\mathbf{x}}(0)\,\Delta\mathbf{x} \quad .$$

We want to find $\Delta\mathbf{x}$ so that $\mathbf{Y}(\Delta\mathbf{x}) = 0$, which gives

$$\Delta\mathbf{x} = -\left(\frac{\partial\mathbf{Y}}{\partial\Delta\mathbf{x}}(0)\right)^{\dagger}\mathbf{Y}(0) \quad ,$$

where † denotes the pseudo-inverse. In practice it is useful to use the Levenberg-Marquardt method. Let

$$A = \frac{\partial\mathbf{Y}}{\partial\Delta\mathbf{x}}(0), \qquad b = \mathbf{Y}(0) \quad .$$

Instead of taking

$$\Delta\mathbf{x} = -(A^TA)^{-1}A^Tb \quad , \tag{10}$$

which might be numerically sensitive if (A^TA) has small singular values one uses the update

$$\Delta\mathbf{x} = -(A^TA + \varepsilon I)^{-1}A^Tb \quad ,$$

where ε is a small positive number.

5. Experiments

The method was tested on simulated data. Two different simulations were performed in order to show the performance and robustness of the bundle adjustment algorithm. Two further experiments were carried out using a restricted camera motion. In the first case the camera is stationary and the object is purely translating and in the second the camera is stationary and the object is rotating around its center.

First simulation

First an experiment was performed with 10 points in 15 images. The points were taken as random points with coordinates between -300 and $+300$ units. The camera positions were chosen at random approximately 1000 units away. All point- and camera-parameters were estimated with the method described above. Each iteration of the minimisation involves a matrix inversion of the following type

$$(A^TA)\Delta\mathbf{x} = -A^Tb \quad .$$

The singular values and of A^TA give valuable information of the stability of the estimated parameters. A plot of the logarithm of the singular values are shown in Figure 1. Notice that the last 7 singular values are significantly smaller than the others. These correspond to changes in translation, orientation and global scale, i.e. to the unknown similarity transformation. The matrix A^TA can also be used to estimate the covariance matrix of the estimated parameters as

$$C[\Delta\mathbf{x}] \approx \frac{A^TA}{2}^{-1}\sigma^2[e] \quad , \tag{11}$$

where e is the stochastic variable representing errors in image coordinates. Under the assumptions that the errors are of equal distribution and small so that the linearisation holds with high accuracy, a non-biased estimate of the variance $\sigma^2[e]$ is given by

$$\hat{\sigma}^2[e] = \frac{f}{2mn - (9m + 3n - 7)} \ ,$$

where $2mn$ is the number of measurements and $(9m+3n-7)$ is the number of estimated parameters (see [9]); 9 camera parameters in each camera matrix describing a camera with Euclidean image planes and 3 parameters for the coordinates for each reconstructed point minus 7 parameters for the unknown similarity transformation that is impossible to recover, giving effectively $3n-7$ parameters in the reconstruction.

Thus small singular values of $A^T A$ correspond to large uncertainties in the estimated parameters.

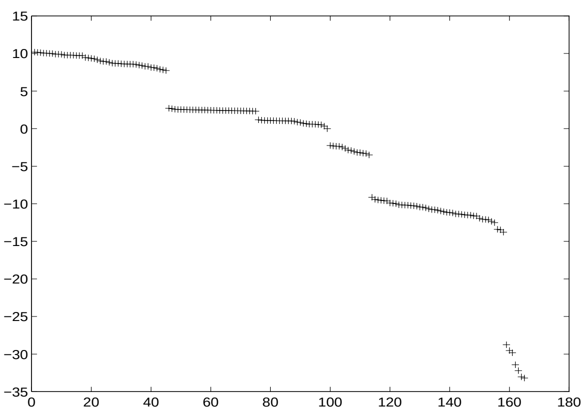

Figure 1. The logarithm of the singular values of the Hessian, $A^T A/2$.

The matrix $A^T A$ has an interesting almost block-diagonal structure. This structure can be used to simplify the solution of (10), see for example [1, 3].

The standard deviation σ, the estimate $\hat{\sigma}$ are presented in Table 1, together with the focal length f and the position (x_0, y_0) of the principal point of the first camera and the RMS of reconstructed object positions in percent of overall scale. This was done for different levels of noise, 0.1 pixels up to 5 pixels in standard deviation $\sigma[e]$. Table 2 presents some estimates, obtained from (11), of the mean or standard deviations of the corresponding entities in Table 1.

Second simulation

Second an experiment was performed with 50 points in 20 images. The points were taken as random points with coordinates between -500 and $+500$ units. The camera positions were chosen at random approximately 1000 units away. All point- and camera-parameters were estimated with the method described above.

σ	$\hat{\sigma}$	f	x_0	y_0	Δ
0	0.000	2112.191	25.433	8.250	0.000
0.1	0.099	2096.894	33.395	6.853	0.371
0.2	0.198	2107.966	43.571	5.061	2.193
0.5	0.558	2143.423	56.123	31.375	1.727
1	0.887	1982.302	9.773	-16.357	3.611
2	1.907	2057.016	352.815	-22.979	11.247
5	4.825	1974.814	314.814	32.671	18.755

Table 1. Some estimated parameters and the reconstruction error in the first simulation.

σ	$E[\hat{\sigma}]$	$\sigma[f_1]$	$\sigma[x_0]$	$\sigma[y_0]$	$E[\Delta]$
0.1	0.100	4.070	2.480	2.514	0.459
0.2	0.200	8.140	4.961	5.029	0.918
0.5	0.500	20.351	12.402	12.572	2.296
1	1.000	40.701	24.804	25.144	4.591
2	2.000	81.402	49.609	50.288	9.183
5	5.000	203.506	124.021	125.720	22.957

Table 2. Some estimated standard deviations and mean values in the first simulation.

The standard deviation σ, the estimate $\hat{\sigma}$ are presented in Table 3, together with the focal length f and the position (x_0, y_0) of the principal point of the first camera and the RMS of reconstructed object positions in percent of overall scale. This was done for different levels of noise, 0.1 pixels up to 10 pixels in standard deviation $\sigma[e]$. Table 4 presents some estimates, obtained from (11), of the mean or standard deviations of the corresponding entities in Table 3.

Comments

The proof that it is possible to obtain Euclidean reconstruction up to scale using uncalibrated cameras with zero skew and aspect ratio equal to one, was based upon the assumption that the camera motion was sufficiently general. In the

σ	$\hat{\sigma}$	f_1	x_0	y_0	Δ
0.0	0.000	1010.752	4.435	1.355	0.000
0.1	0.102	1010.787	4.460	1.385	0.017
0.2	0.198	1012.072	4.723	1.271	0.135
0.5	0.510	1008.164	4.416	1.959	0.225
1	1.047	1010.795	5.023	2.970	0.251
2	2.009	1014.648	7.878	2.285	0.357
5	4.885	1007.924	12.647	-1.364	0.669
10	10.121	1020.446	-6.559	-7.934	2.033

Table 3. Some estimated parameters and the reconstruction error in the second simulation.

σ	$E[\hat{\sigma}]$	$\sigma[f_1]$	$\sigma[x_0]$	$\sigma[y_0]$	$E[\Delta]$
0.0	0.000	0.000	0.000	0.000	0.000
0.1	0.100	0.122	0.118	0.131	0.036
0.2	0.200	0.245	0.236	0.261	0.073
0.5	0.500	0.612	0.589	0.654	0.182
1	1.000	1.225	1.178	1.307	0.364
2	2.000	2.450	2.355	2.614	0.728
5	5.000	6.124	5.888	6.536	1.821
10	10.000	12.248	11.775	13.072	3.642

Table 4. Some estimated standard deviations and mean values in the second simulation.

case where the camera is stationary but the object is translating along a line and the camera is stationary and the object is revolving around a fixed point or around a fixed axis, simulations shows that there are 10 zero singular values in each case. This indicates that reconstruction is only possible up to a 10 dimensional manifold involving the similarity transformation group.

A comparison between Table 1 and Table 2 and between Table 3 and Table 4 shows that the estimated quantities, from (11), are in compliance with the experimental data; at least for small levels of noise. It is also important to note that many points are needed in many images since there are so many unknown parameters. The first simulation with 10 points in 15 images with 300 equations and 158 unknown degrees of freedom is much less stable than the second simulation with 50 points in 20 images, (2000 equations and 323 unknown degrees of freedom). Notice, however, that this information is obtained directly from the estimate of the covariance matrix (11).

A crucial step in the algorithm is the initialisation. In order to obtain a good convergence the initial data have to be sufficiently accurate. This can be achieved if the focal length and principal point is approximatively known for each image. Once the initial data is close to the correct solution the convergence of the bundle adjustment algorithm is very fast.

It can be argued that at least 4 images are needed in order to make a Euclidean reconstruction from Euclidean image planes, see [4].

6. Conclusions

In this paper we have shown that it is possible to reconstruct an unknown object from a number of its projective images up to similarity transformations, i.e. angles and ratios of lengths can be calculated. This is possible even when the focal distance and the principal point change between the different imaging instants. The only thing we need to know about the cameras is the aspect ratio and the skew. These parameters are defined by the geometry of the light sensitive area and need only be measured once for each camera. In many cases it is reasonable to assume that the skew is 0 and the aspect ratio is 1. This is called a camera with Euclidean image plane.

The paper contains a theoretical proof of this fact as well as an experimental validation using simulated data. In these experiments a bundle adjustment technique has been used to estimate all undetermined parameters, i.e. the reconstructed object, the relative position of the cameras, the focal lengths and principal points at the different imaging instants. Using this optimisation procedure, the Hessian gives valuable information about the stability of the solution. For example, it is clearly seen that only 7 parameters, corresponding to choice of origin, orientation and global scale, can not be estimated. These corresponds to the unknown similarity transformation.

References

[1] American Society for Photogrammetry. *Manual of Photogrammetry*, 4:th edition, 1984.

[2] O. D. Faugeras, Q.-T. Luong, and S. J. Maybank. Camera self-calibration: Theory and experiments. In G. Sandini, editor, *ECCV'92*, volume 588 of *Lecture notes in Computer Science*, pages 321–334. Springer-Verlag, 1992.

[3] R. I. Hartley. Euclidean reconstruction from uncalibrated views. In J. L. Mundy, A. Zisserman, and D. Forsyth, editors, *Applications of Invariance in Computer Vision*, volume 825 of *Lecture notes in Computer Science*, pages 237–256. Springer-Verlag, 1994.

[4] A. Heyden. *Geometry and Algebra of Multiple Projective Transformations*. PhD thesis, ISBN 91-628-1784-1, Lund University, Sweden, 1995.

[5] A. Heyden and K. Åström. Euclidean reconstruction from constant intrinsic parameters. In *ICPR'96*, volume 1, pages 339–343. IEEE Computer Society Press, 1996.

[6] Q.-T. Luong. *Matrice Fondamentale et Calibration Visuelle sur l'Environnement-Vers une plus grande autonomie des systèmes robotiques*. PhD thesis, Université de Paris-Sud, Centre d'Orsay, 1992.

[7] M. Pollefeys, L. Van Gool, and A. Oosterlinck. Euclidean reconstruction from image sequences with variable focal length. In B. Buxton and R. Cipolla, editors, *ECCV'96*, volume 1064 of *Lecture notes in Computer Science*, pages 31–44. Springer-Verlag, 1996.

[8] M. Pollefeys, L. Van Gool, and A. Oosterlinck. The modulus constraint: a new constraint for self-calibration. In *ICPR'96*, volume 1, pages 349–353. IEEE Computer Society Press, 1996.

[9] C. R. Rao. *Linear Statistical Inference and its Application*. John Wiley & Sons, Inc., 1965.