Minimal Conditions on Intrinsic Parameters for Euclidean Reconstruction *

Anders Heyden, Kalle Åström

Dept of Mathematics, Lund University Box 118, S-221 00 Lund, Sweden email: {heyden,kalle}@maths.lth.se

Abstract In this paper will be investigated what constraints on the intrinsic parameters that are needed in order to reconstruct an unknown scene from a number of its projective images. Two such minimal cases are studied in detail. Firstly, it is shown that it is sufficient to know the skew parameter, even if all other parameters are unknown and varying, to obtain a Euclidean reconstruction. Secondly, the same thing can be done for known aspect ratio, again when all other intrinsic parameters are unknown and varying. In fact, we show that it is sufficient to know any of the 5 intrinsic parameters to make Euclidean reconstruction.

An algorithm, based upon bundle adjustment techniques, to obtain Euclidean reconstruction in the above mentioned cases are presented. Experiments are shown on the slightly simpler case of both known aspect ratio and skew.

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 transformation (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.

A priori information about the object can be used in a fairly straight-forward manner, see [3]. Several researchers have dealt with the problem of using additional information about the extrinsic parameters, i.e., the camera orbit, see for example [12].

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 [9], [2]. In [5] 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.

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

Another constraint, called the *modulus constraint* have been used in [11], 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 [10]. 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 the 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 and/or the skew are constant during the whole sequence (at least if the same camera is used), since they are determined by the geometry of the light sensitive array. It is furthermore possible to measure, once and for all, the aspect ratio and/or skew for a camera and then use these values for all image sequences captured by that camera. When these quantities are known, they can be compensated for and we may assume that the aspect ratio is equal to 1 and/or the skew is equal to 0.

It has been shown, see [7] that it is possible to make Euclidean reconstruction when both the skew and aspect ratio are known. In this paper we will extend that result and show theoretically that Euclidean reconstruction is possible in two minimal cases. Firstly, when the aspect ratio is known and secondly, when the skew is known. In both cases are all other intrinsic parameters unknown and allowed to vary between the different imaging instants. Although of less practical importance, it is also shown that it is possible to make Euclidean reconstruction if only one coordinate of the principal point is known or if the focal length is known. In all these cases, not only a Euclidean reconstruction of the object are obtained, but also the intrinsic camera parameters for each camera, i.e. auto-calibration is performed. The proof is based on the assumption of generic camera motion. The theoretical result is verified by experiments on simulated data.

2 The Camera Model

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

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

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 and (x_0, y_0) is called the principal point. 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 mostly deal with cameras where s = 0 and/or $\gamma = 1$.

Definition 1. A camera that can be modeled as in (1), with s = 0 is called a **non-skew** camera. When $\gamma = 1$ it is called an **aspect-free camera** and when both s = 0 and $\gamma = 1$ it is called a **camera with Euclidean image plane**. Internal calibration matrices *K* of type

$$K_{\rm ns} = \begin{bmatrix} \gamma f \ 0 \ x_0 \\ 0 \ f \ y_0 \\ 0 \ 0 \ 1 \end{bmatrix}, \quad K_{\rm af} = \begin{bmatrix} f \ s \ x_0 \\ 0 \ f \ y_0 \\ 0 \ 0 \ 1 \end{bmatrix} \quad \text{and} \quad K_{\rm Eip} = \begin{bmatrix} f \ 0 \ x_0 \\ 0 \ f \ y_0 \\ 0 \ 0 \ 1 \end{bmatrix}$$
(2)

are called **non-skew calibration matrices**, **aspect-free calibration matrices** and **Euclidean calibration matrices** respectively.

Observe that it is not necessary that $\gamma = 1$ and/or s = 0 in order to use the subsequent results. It is sufficient that they are known, since then the image coordinates can be transformed to new ones obeying the constraints on γ and/or s, for details see [8].

The following results, one of them shown in [6] and one shown in [4], will be needed later. For a proof, see [8].

Lemma 2. A camera matrix $P = \begin{bmatrix} u^T \\ v^T \\ w^T \end{bmatrix} t$, normalised such that w.w = 1, represents a non-skew camera if and only if

$$(u \times w).(v \times w) = 0 \tag{3}$$

an aspect-free camera if and only if

$$|(u \times w) \times (u \times w)| = (v \times w).(v \times w)$$
(4)

and a camera with Euclidean image plane if and only if

$$(u \times w).(v \times w) = 0 \quad and \quad (u \times w).(u \times w) = (v \times w).(v \times w) \quad . \tag{5}$$

Here 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.

It is possible to state similar conditions for the cases of known f, x_0 and y_0 . By making coordinate changes in the images similar to the previous cases it is no restriction to assume that f = 1, $x_0 = 0$ and $y_0 = 0$ respectively. The conditions on the camera matrices are in turn $|v \times w| = w.w$, u.w = 0 and v.w = 0 respectively.

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 known skew and/or aspect ration, 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, ..., m that fulfills

$$\lambda_i \mathbf{x}_i = P_i \mathbf{X}, \quad i = 1, \dots, m \quad , \tag{6}$$

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, ..., m, and a reconstruction, \mathbf{X} , also P_iH , i = 1, ..., 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.

The next step is to show that given a sequence of camera matrices that solves the projective reconstruction problem and represents cameras with known skew and/or aspect ratio, i.e. fulfills one of the conditions in Lemma 2, 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 $\mathcal{M}_{\rm P}$ the manifold of all 3 × 4 projection matrices, i.e., the set of all 3 × 4 matrices defined up to scale. Denote by $\mathcal{M}_{\rm ns}$, $\mathcal{M}_{\rm af}$ and $\mathcal{M}_{\rm Eip}$ the manifold of all camera matrices that represents non-skew cameras, aspect-free cameras and cameras with Euclidean image planes, respectively, i.e., all 3 × 4 matrices that can be written as in (2), and thus obeying one of the conditions in Lemma 2. Denote the group of all projective transformations, represented by 4 × 4 matrices, by $\mathcal{G}_{\rm P}$. The subclass of transformations that preserves the properties in Lemma 2 is denoted by $\mathcal{G}_{\rm ns}$, $\mathcal{G}_{\rm af}$ and $\mathcal{G}_{\rm Eip}$ respectively, e.g.

$${\mathcal{G}}_{\mathrm{ns}} = \{ \, H \in {\mathcal{G}}_{\mathrm{P}} \, | (P \in {\mathcal{M}}_{\mathrm{ns}}) \Rightarrow PH \in {\mathcal{M}}_{\mathrm{ns}} \, \} \,$$

Finally, the group of all similarity transformations will be denoted by \mathcal{G}_S and will be represented by

$$\mathcal{G}_S = \{ H = \begin{bmatrix} \lambda R \ t \\ 0 \ 1 \end{bmatrix} \mid RR^T = I, \ 0 \neq \lambda \in I\!\!R \} \quad .$$
(7)

It can easily be seen that the group of similarity transformations is contained in \mathcal{G}_{ns} as well as in \mathcal{G}_{af} and \mathcal{G}_{Eip} . Thus $\mathcal{G}_S \subseteq \mathcal{G}_{ns} \subseteq \mathcal{G}_P$, $\mathcal{G}_S \subseteq \mathcal{G}_{af} \subseteq \mathcal{G}_P$ and $\mathcal{G}_S \subseteq \mathcal{G}_{Eip} \subseteq \mathcal{G}_P$. Note that \mathcal{G}_{ns} , \mathcal{G}_{af} and \mathcal{G}_{Eip} , are precisely the transformation groups that groups of interest. If, for example, $\mathcal{G}_{ns} = \mathcal{G}_P$ then it is only possible to make reconstruction up to a projective transformation in the non-skew camera case. If $\mathcal{G}_{ns} = \mathcal{G}_S$ then it is possible to make reconstruction up to a similarity transformation.

Theorem 3. Let \mathcal{G}_{ns} , \mathcal{G}_{af} and \mathcal{G}_{Eip} respectively denote the class of transformations in 3D-space that preserves the conditions in Lemma 2 and \mathcal{G}_S the group of similarity transformations in 3D-space. Then

$${\cal G}_{
m ns} = {\cal G}_{
m af} = {\cal G}_{
m Eip} = {\cal G}_S$$
 .

Proof. Consider first the case of a aspect-free camera. From the discussion above we have $\mathcal{G}_S \subseteq \mathcal{G}_{Eip}$.

Observe that the constraints on the camera matrices in Lemma 2 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 an aspect-free camera, H a projective transformation and

$$PH = K[R \mid t] \begin{bmatrix} A & b \\ c & d \end{bmatrix} = [K(RA + tc) \mid K(Rb + td)] \in \mathcal{M}_{af} \quad . \tag{8}$$

Then K(RA + tc) can be factorised K(RA + tc) = K'R' where K is an aspect-free calibration matrix and R' denotes an orthogonal matrix. Since (8) is valid for any P that represents a camera matrix, i.e., for any K, R and t, we first choose t = 0.

Assume that A has the property that for every aspect-free calibration matrix K and orthogonal R, it is possible to factorise KRA according to KRA = K'R', for some aspect-free 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''' denote orthogonal matrices. Now, using the singular value decomposition of A we can write

$$D_1 = U_1 A V_1 = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \quad \text{and} \quad D_2 = U_2 A V_2 = \begin{bmatrix} b & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & a \end{bmatrix}$$

where D_2 is obtained by a simple permutation of the rows and columns in U_1 and V_1 respectively. Replacing A by D_1 and choosing R = I in (8), Lemma 2 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 aspect-free calibration matrix K, every orthogonal R and every t, K(RA + tc) can be factorised as K(RA + tc) = K'R' for some aspect-free 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 $cV = [s \, 00]$, then choose $R = (AV)^{-1}$ and $t = [100]^T$. These choices give

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

and according to Lemma 2, 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 $\frac{1}{d}H \in \mathcal{G}_S$. Thus $\mathcal{G}_{af} \subseteq \mathcal{G}_S$ from which the first part of the theorem follows.

Consider now the case of a non-skew camera. Again from the discussion above we have $\mathcal{G}_S \subseteq \mathcal{G}_{Eip}$. The proof is analogous to the above until $D_1 = U_1 A V_1 =$ diag(a, b, c) is obtained. Then according to (8)

$$R = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix} \implies RD_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} a & b & 0 \\ -a & b & 0 \\ 0 & 0 & \sqrt{2}c \end{bmatrix} .$$

Now, according to Lemma 2, RD_2 is a skew-free calibration matrix if and only if

$$((a, b, 0) \times (0, 0, c)) \cdot ((-a, b, 0) \times (0, 0, c)) = (bc, -ac, 0) \cdot (bc, ac, 0) = c^2(b^2 - a^2) = 0$$

i.e. if and only if a = b. Using a permutation of the singular values as before gives a = b = c and then we can proceed as in the aspect-free case again until the choice of cV. This time we chose cV = [0 s 0], which gives

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

which according to Lemma 2 gives s = 0 and so on.

Finally, the case of a camera with Euclidean image plane follows from either of the two cases above and this completes the proof.

A glimpse at the proof gives that the three remaining minimal cases can be proven similarly:

- If only the focal length, f is known (assumed to be equal to 1), it is possible to make Euclidean reconstruction. (A similar proof as in the aspect-free case.)
- If only the *x*-coordinate or the *y*-coordinate of the principal point is known, it is possible to make Euclidean reconstruction. (Similar to the proof in the skew-free case, but use $\begin{bmatrix} 1 & 0 & 1 \end{bmatrix}$

$$R = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1\\ 0 & \sqrt{2} & 0\\ -1 & 0 & 1 \end{bmatrix} \text{ and } cV = \begin{bmatrix} 0 & 0 & s \end{bmatrix}$$

in the case of $x_0 = 0$ and similarly in the case of $y_0 = 0$.

Thus reconstruction up to a similarity transformation is possible if any of the internal calibration parameters f, s, γ , x_0 or y_0 are known.

Observe 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.

Finally, it can be shown that these reconstruction problems can be formulated as polynomial equations in the intrinsic parameters in the first image and the parameters describing the plane at infinity, see [6]. The polynomial equations are exactly the ones in Lemma 2, i.e. two equations per image for the case of Euclidean image planes and one equation per image in the other cases. Since the equations arises from different images they are in general independent. This means that in the case of Euclidean image planes at least 4 images are needed since we have 6 parameters (3 for the plane at infinity and 3 for the unknown intrinsic parameters) and 8 equations (6 equations will not give a unique solution). In the other cases we have 7 parameters and one equation per images, which means that at least 8 images are needed in general.

4 Experiments

The method was tested on simulated data in the case of a camera with Euclidean image plane. The numerical computations has been made using a so called *bundle adjustment technique*. Briefly parameters are introduced for all object coordinates and camera matrices. Then the squared difference between the image coordinates obtained from these parameters and the true image coordinates are minimised with respect to the parameters. Apart from this least squares solution also an estimate of the accuracy of the parameters can be obtained. For further details see [8]. Note that the calculated parameters not only gives a Euclidean reconstruction, but also the intrinsic parameters for all cameras.

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. The standard deviation σ 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 are presented in Table 1 for different levels of noise, $\sigma[e]$.

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. The standard deviation σ 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 are presented in Table 1 for different levels of noise, $\sigma[e]$.

σ	f	ma	A 10	Δ	σ	f_1	x_0	y_0	Δ
	J	x_0	y_0		0.0	1010.752	4.435	1.355	0.000
0	2112.191	25.433	8.250	0.000					
0.1	2096.894	33 305	6.853	0.371	0.1	1010.787	4.460	1.385	0.017
				0.0.2	0.2	1012.072	4.723	1.271	0.135
0.2	2107.966	43.571	5.061	2.193					
05	2143.423	56 123	31 375	1 7 2 7		1008.164			
					1	1010.795	5.023	2.970	0.251
1	1982.302	9.773	-16.357	3.611	2	1014.648	7 070	2 205	0 257
2	2057.016	352 815	-22 979	11 247					
					5	1007.924	12.647	-1.364	0.669
5	1974.814	314.814	32.671	18.755		1020.446			
					10	1020.440	-0.339	-7.934	2.055

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

It is 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).

5 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/or 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 other two minimal cases, aspect ratio equal to 1, called an aspect-free camera, and skew equal to 0, called a non-skew camera, are also treated and shown to give reconstructions up to an unknown similarity transformation. Although of less practical importance, it is also shown that it is possible to make Euclidean reconstruction if only one coordinate of the principal point is known or if the focal length is known.

The paper contains a theoretical proof of these facts as well as an experimental validation using simulated data in the case of a camera with Euclidean image planes. 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 and the intrinsic parameters at the different imaging instants.

References

- 1. American Society for Photogrammetry, *Manual of Photogrammetry, Ed. C.C. Slama, 4th edition,* 1984.
- Faugeras, O. D., Luong, Q.-T., Maybank, S. J., Camera Self-Calibration: Theory and Experiments, ECCV'92, Lecture notes in Computer Science, Vol 588. Ed. G. Sandini, Springer-Verlag, 1992, pp. 321-334.
- Faugeras, O., D., Stratification of three-dimensional projective, affine and metric representations, *Journal of the Optical Society of America A.*, vol. 12, No. 3/March 1995, pp. 465-484.
- 4. Faugeras, O., D., Three-Dimensional Computer Vision, MIT Press, Cambridge, Mass., 1993.
- Hartley, R., I., Euclidean Reconstruction from Uncalibrated Views, Applications of Invariance in Computer Vision, Lecture notes in Computer Science, Vol 825. Ed. Joseph L. Mundy, Andrew Zisserman and David Forsyth, Springer-Verlag, 1994, pp. 237-256.
- 6. Heyden, A., Geometry and Algebra of Multiple Projective Transformations, *Doctoral Thesis, CODEN:LUFTD2/TFMA--95/5002--SE, ISBN 91-628-1784-1, Lund, Sweden,* 1995.
- Heyden, A., Åström, K., Euclidean Reconstruction from Image Sequences with Varying and Unknown Focal Length and Principal Point, *Proc. CVPR'97, IEEE Computer Society Press*, 1997, pp. 438–443.
- Heyden, A., Åström, K., Minimal Conditions on Intrinsic Parameters for Autocalibration and Euclidean Reconstruction, *Internal Report, Dept. of Mathematics, Lund University,* CODEN:LUFTD2/TFMA--97/???-SE, Lund, Sweden, 1997.
- Luong, Q.-T., 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.
- Pollefeys, M., Van Gool, L., Oosterlinck, A., Euclidean Reconstruction from Image Sequences with Variable Focal Length, *ECCV'96, Lecture notes in Computer Science, Vol 1064. Ed. B. Buxton, R. Cipolla, Springer-Verlag,* 1996, pp. 31-44.
- Pollefeys, M., Van Gool, L., Oosterlinck, A., The modulus constraint: a new constraint for self-calibration, *Proc. ICPR'96, IEEE Computer Society Press, vol. 1*, 1996, pp. 349-353.
- 12. Zisserman, A., Beardsley, P., Reid, I., Metric Calibration of a Stereo Rig, *Proc. IEEE Workshop on Representation of Visual Scenes*, 1995.