Self-Calibration of An Affine Camera from Multiple Views

Long Quan

Self-calibration of an Affine Camera from Multiple Views

Long QUAN

LIFIA - CNRS - INRIA,
46, avenue Felix Viallet,
38031 Grenoble, France
Long.Quan@imag.fr

Abstract

A key limitation of all existing algorithms for shape and motion from image sequences under orthographic, weak perspective and para-perspective projection is that they require the calibration parameters of the camera. We present in this paper a new approach that allows the shape and motion to be computed from image sequences without having to know the calibration parameters. This approach is derived with the affine camera model, introduced by Mundy and Zisserman [18], which is a more general class of projections including orthographic, weak perspective and para-perspective projection models. The concept of self-calibration, introduced by Maybank and Faugeras in [16] for the perspective camera and by Hartley for the rotating camera in [10], is then applied for the affine camera.

This paper introduces the 3 intrinsic parameters that the affine camera can have at most. The intrinsic parameters of the affine camera are closely related to the usual intrinsic parameters of the pin-hole perspective camera, but different in general case. Based on the invariance of the intrinsic parameters, the methods of self-calibration of the affine camera are proposed. It is shown that with at least four views, an affine camera may be self-calibrated up to a scaling factor. It turns out Euclidean (similarity) shape reconstruction only up to a global scaling factor. Another consequence of the introduction of intrinsic and extrinsic parameters of the affine camera is that all existing algorithm using calibrated cameras can be assembled into the same framework and some of them can be easily extented to a batch solution.

Experimental results are presented and compared with other methods using calibrated affine cameras.

Key words: *affine camera, orthographic, weak perspective, para-perspective, self-calibration, intrinsic parameters, affine shape.*

R.T. Imag-Lifia 26, Nov. 1994

1 Introduction

Three dimensional metric reconstruction from image sequence is probably one of the most important themes in computer vision. Very recently, a lot of work has been developed for invariant reconstruction such as projective reconstruction from multiple uncalibrated images [4, 8, 22, 17], this class of methods assumes generally a full perspective camera model and the invariant shape is generally projective shape. Parallelly, there is also a lot of work [27, 25, 13, 23, 14] proposed for reconstruction from the simplified camera models such as orthographic, weak perspective and para-perspective. These simplified projection models provide a good approximation to the perspective projection model when the depth of an object is small compared to the viewing distance. Among those, one of the most

remarquable method is the one proposed by Tomasi and Kanade called *factorization method* which gives an elegant batch solution. In this case, it is known that the intrinsic invariant shape that may be recovered is the affine shape. This affine shape can be transformed into its usual Euclidean shape by intergrating the so called *metric constraints*. These metric constraints are essentially the intrinsic parameters of the camera. It is the purpose of this paper to propose methods which allow to get Euclidean reconstruction without having to know the intrinsic parameters of the camera from the affine reconstruction, therefore avoid the tedious and numerically non stable calibration.

To get Euclidean structure from uncalibrated affine cameras, the key idea is to use the self-calibration idea originally proposed by Maybank and Faugeras in [16] for the perspective camera. More practical self-calibration methods have been developed in [9, 10, 15], especially that for self-calibrating a rotating camera.

We will develop the self-calibration method in the context of the affine camera model, introduced by Mundy and Zisserman [18]. The affine camera generalises the orthographic, weak perspective and para-perspective models. *Parallelism is still preserved by the affine camera but the image plane shapes are potentially more distorted since the image coordinates can undergo anisotropic scaling. It is not clear what viewing process leads to the affine camera approximation.* ([18, page 514]). A number of interesting results has been achieved by Shapiro et al. in [24] based on this concept.

In the context of the affine camera, we will first introduce the intrinsic and extrinsic parameters for the affine camera by analogy to the perspective camera. It will be shown that the affine camera can have at most 3 intrinsic parameters and the intrinsic parameters of the affine camera are generally different from those of the perspective camera. Based on the invariance of these intrinsic parameters, several self-calibration methods which allow the shape and motion to be computed from uncalibrated affine camera will be developed.

The contributions of this paper are mainly twofold. The introduction of of the intrinsic parameters for the affine camera and the algorithms of self-calibration for the affine camera. One of the important consequences is that Euclidean shape and motion are computed from uncalibrated affine images. Another one is that all algorithms using calibrated affine cameras can be nicely assembled into the same framework in terms of the intrinsic calibration matrix of the affine camera. Therefore batch solution for shape and motion is available for all different projection models included in the affine camera.

The paper is organized as follows. In Section 2, the affine camera model is briefly reviewed. Then in Section 2.1, the intrinsic and extrinsic parameters of the affine camera are introduced, and some special examples of the affine camera are also given based on this decomposition. Later in Section 3, the affine shape and motion from image sequences by factorization method is reviewed following Tomasi and Kanade with the extension to the affine camera. Next in Section 4, several self-calibration methods for the affine camera are proposed. The related work on shape and motion from image sequences under orthographic, weak perspective and para-perspective projection is discussed and reformulated in terms of the intrinsic parameters of the affine camera in Section 5. The experimental results are presented in Section 6. Finally, some concluding remarks are given in Section 7.

2 Affine camera

The concept of the affine camera is reviewed in this section following Mundy and Zisserman [18] and Shapiro *et al.* [24].

A pin-hole camera can be modeled as a linear mapping from $\mathcal{P}^3$ to $\mathcal{P}^2$. This mapping is represented by a 3×4 matrix if homogeneous coordinates are used both for object and image points.

$$w \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = P_{3\times 4} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{1}$$

The camera represented by $P_{3\times 4}$ is usually called the perspective camera.

If p_{ij} denote the entries of the matrix $P_{3\times 4}$, then by introducing the constraints $p_{31} = p_{32} = p_{33} = 0$ to the perspective camera, we obtain the affine camera introduced by Mundy and Zisserman in [18]:

$$w \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = P_{3\times 4} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & p_{34} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}. \tag{2}$$

Affine camera represents a class of projections more general than orthographic, weak perspective and para-perspective projections and less general than the full perspective projection.

In terms of non-homogeneous coordinates, by removing the common scalar factor, the equation (2) is rewritten as

$$\begin{pmatrix} u \\ v \end{pmatrix} = M_{2\times 3} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} t_u \\ t_v \end{pmatrix}$$

where $m_{ij} = p_{ij}/p_{34}$ and $t_u = p_{14}/p_{34}$, $t_v = p_{24}/p_{34}$.

One of the very attractive properties of the class of affine cameras is that it is independent of translations if the relative coordinates, with respect to a given reference point, are used both in space and in image. For any given reference point $(u_r, v_r)^T$ in image and $(x_r, y_r, z_r)^T$ in space, the relative coordinates $(\tilde{u}, \tilde{v})^T$ in image and $(\tilde{x}, \tilde{y}, \tilde{z})^T$ in space of a point are

$$\begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix} = \begin{pmatrix} u - u_r \\ v - v_r \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} x - x_r \\ y - y_r \\ z - z_r \end{pmatrix}.$$

As $(u_r, v_r)^T = M_{2\times 3}(x_r, y_r, z_r)^T + (t_u, t_v)^T$, it follows that the basic projection equation for the affine camera in terms of relative coordinates is

$$\begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix} = M_{2\times 3} \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix}.$$

An affine camera $M_{2\times 3}$ has $6 = 2 \times 3$ independent parameters.

It can be easily verified that the centroid of a set of points in space is also projected into the centroid of the points in the image plane by the affine camera. Therefore, in practice, the centroid is taken as the reference point for the relative coordinates. We assume also in this paper that this is always done.

2.1 Intrinsic and extrinsic parameters of the affine camera

This section introduces the intrinsic and extrinsic parameters for the affine camera $M_{2\times 3}$ following the same idea for the perspective camera.

2.2 The general perspective camera: review

For a perspective camera $P_{3\times4}$, it has at most 11 independent parameters which are separated into two parts: the intrinsic and extrinsic parameters. This separation was achieved by uniquely decomposing $P_{3\times4}$ [5, 26] into

$$P_{3\times4} = A_{3\times3}D_{3\times4} = \begin{pmatrix} \alpha_u & \cot\theta & u_0 \\ 0 & \alpha_v/\sin\theta & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R_{3\times3} & t_{3\times1} \end{pmatrix}.$$

$A_{3\times3}$ is an upper triangular matrix and its 5 entries are so called the 5 intrinsic parameters of the perspective camera which have the following interpretation: α_u and α_v are respectively the focal length in horizontal and vertical pixels; (u_0, v_0) is the principal point; θ is the angle between the two camera axes. For convenience, we can also use two other equivalent parameters f, the focal length in pixels and $\xi = \alpha_u/\alpha_v$, called the aspect ratio of the camera instead of α_u and α_v.

$R_{3\times3}$ and $t_{3\times1}$ denote respectively a rotation and a translation in space. They count for the $6 = 3+3$ extrinsic parameters of the perspective camera.

For CCD cameras, it is generally assumed there is no skew, *i.e.* $\theta = \pi/2$. Without loss of generality and for simplicity of the development, we use 4 parameters for the perspective camera, f, ξ, u_0 and v_0.

2.3 The affine camera

By analogy to this intrinsic and extrinsic decomposition for the perspective camera, let consider the affine camera represented by $M_{2\times3}$. by invoking QR decomposition theorem ([7, 8, 5]), then $M_{2\times3}$ is decomposed as follows

$$M_{2\times3} = A_{2\times2}R_{2\times3} = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \begin{pmatrix} r_1^T \\ r_2^T \end{pmatrix},$$

this decomposition is unique if $\text{rank}(M_{2\times3}) = 2$. $A_{2\times2}$ is a 2×2 lower triangular matrix that we call the (intrinsic) calibration matrix of the affine camera. $A_{2\times2}$ encodes therefore at most 3 intrinsic parameters of the affine camera. $R_{2\times3}$ is a 2×3 matrix containing two rows of a 3×3 rotation matrix. The third missing row of the rotation matrix can always be recovered if necessary by the constraint of orthonormality. $R_{2\times3}$, though 2×3, represents a full $3D$ rotation whose 3 degrees of freedom[1] count for the 3 extrinsic parameters of the affine camera.

The intrinsic calibration matrix $A_{2\times2}$ can be rewritten as

$$A_{2\times2} = k \begin{pmatrix} \xi_a & 0 \\ s_a & 1 \end{pmatrix}.$$

k can be interpreted as the scaling factor of the affine camera, ξ_a the aspect ratio of the affine camera, s_a the skew of the affine camera. These intrinsic parameters are generally different from those of the perspective camera, except in some special simple cases. We are now to give some examples of intrinsic parameters for the special cases of the affine camera.

[1] A 2×3 matrix with three constraints $r_1^T r_2 = 0$, $\|r_1^T\|^2 = \|r_2^T\|^2 = 1$.

2.4 Some special cases of the affine camera

For orthograhic projection, the camera matrix can be taken to be $M_{2\times 3} = \begin{pmatrix} \xi r_1^T \\ r_2^T \end{pmatrix}$, then

$$A_{ortho} = \begin{pmatrix} \xi & 0 \\ 0 & 1 \end{pmatrix}.$$

In this case, $k = 1$, $\xi_a = \xi$ and $s_a = s = 0$. The unique affine intrinsic affine camera parameter is the aspect ratio which is also that of the perspective camera.

For weak perspective projection, the camera matrix is $M_{2\times 3} = k \begin{pmatrix} \xi r_1^T \\ r_2^T \end{pmatrix}$, the calibration matrix is

$$A_{weak} = k \begin{pmatrix} \xi & 0 \\ 0 & 1 \end{pmatrix}.$$

In this case, $s_a = 0$ and there are two intrinsic parameters: the aspect ratio $\xi_a = \xi$ which is still the same as the perspective camera and the scaling factor k including the scaling effect of both the focal length and the average depth of the object. As $A_{weak} = kA_{ortho}$, the weak perspective projection is also called the scaled orthographic projection.

The para-perspective projection, (*cf.* [19, 1, 24, 12]) generalises the weak perspective by projecting parallel to a given projection direction. If the projection direction is described by two angles θ_x and θ_y, where θ_x lies in the xz plane with the x axis and θ_y in the yz plane with the y axis. The camera matrix of the para-perspective projection takes the following form according to [1, 24]:

$$M_{2\times 3} = \begin{pmatrix} \xi & 0 & -\xi\cot\theta_x \\ 0 & 1 & -\cot\theta_y \end{pmatrix} \begin{pmatrix} r_1^T \\ r_2^T \\ r_3^T \end{pmatrix}.$$

A two-step QR factorization of $M_{2\times 3}$ by using Householder transformation [7] turns out the following calibration matrix:

$$A_{para} = k \begin{pmatrix} \xi & 0 \\ \frac{\cot\theta_x \cot\theta_y}{1+\cot\theta_x^2} & \frac{\sqrt{1+\cot\theta_x^2+\cot\theta_y^2}}{1+\cot\theta_x^2} \end{pmatrix}$$

The scaling factor k describes the various scaling effects introduced in the para-perspective projection. $\xi_a = \frac{1+\cot\theta_x^2}{\sqrt{1+\cot\theta_x^2+\cot\theta_y^2}}\xi$, and $s_a = \frac{\cot\theta_x\cot\theta_y}{\sqrt{1+\cot\theta_x^2+\cot\theta_y^2}}$. From this para-perspective example, we can see that the intrinsic parameters of the affine camera are closely related to those of the perspective camera, but are different.

It should not be surprised that θ_x and θ_y are also part of intrinsic parameters, they are playing the similar role as the principal point (u_0, v_0) for the perspective camera.

The calibration matrix of the para-perspective and that of the weak perspective are related as follows:

$$A_{para} = k \begin{pmatrix} \xi & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1+\cot\theta_x^2 & 0 \\ \cot\theta_x\cot\theta_y & \sqrt{1+\cot\theta_x^2+\cot\theta_y^2} \end{pmatrix} = A_{weak} \begin{pmatrix} 1+\cot\theta_x^2 & 0 \\ \cot\theta_x\cot\theta_y & \sqrt{1+\cot\theta_x^2+\cot\theta_y^2} \end{pmatrix}.$$

Remark that we may need up to $7 = 5+2$ parameters (5 intrinsic for the perspective camera and 2 angles defining the projection direction) to completely describe the calibration of the para-perspective projection, however they are not independent in terms of the intrinsic parameters of the affine camera, as we can have no more than 3 independent parameters.

Also, we should notice that once we know the intrinsic parameters of the affine camera, we can not extract those of the perspective camera since we have only 3 degrees of freedom for the affine camera instead of 5 degrees of freedom for the perspective camera, except for the case of the weak perspective case.

3 Affine shape from uncalibrated affine cameras

It is known [13, 23, 14, 25] that affine shape can be recovered from affine cameras. The elegant *factorization method* to this problem is proposed by Tomasi and Kanade [25] in orthographic projection case, and extended to weak perspective by Weinshall and Tomasi [30] and para-perspective case by Poelman and Kanade [20].

Following Tomasi and Kanade [25], suppose n points are tracked over v distinct views, we can write

$$\begin{pmatrix} \tilde{u}_{11} & \dots & \tilde{u}_{1n} \\ \dots & \dots & \dots \\ \tilde{u}_{v1} & \dots & \tilde{u}_{vn} \\ \tilde{v}_{11} & \dots & \tilde{v}_{1n} \\ \dots & \dots & \dots \\ \tilde{v}_{v1} & \dots & \tilde{v}_{vn} \end{pmatrix} = \begin{pmatrix} M_1 \\ \dots \\ M_f \end{pmatrix} \begin{pmatrix} s_1 & \dots & s_n \end{pmatrix}$$

where $s_i = (\tilde{x}_i, \tilde{y}_i, \tilde{z}_i)^T$.

This equation is noted as

$$W_{2v\times n} = M_{2v\times 3} S_{3\times n}$$

where $W_{2v\times n}$ is the $2v \times n$ measurement matrix, $M_{2v\times 3}$ the $2v \times 3$ camera matrix and $S_{3\times n}$ the $3 \times n$ shape matrix.

It is evident that $W_{2v\times n}$ still have the rank at most 3. Hence, the same factorization method applies. This is also noted in [24] that the same process applies for the general affine camera.

To find Euclidean structure from the affine one, one needs to determine a non singular 3×3 matrix D such that $M'_{2v\times 3} = M_{2v\times 3} D$ and $S' = D^{-1} S$ represent the real affine camera matrix and Euclidean shape, since

$$W_{2v\times n} = M_{2v\times 3} D D^{-1} S_{3\times n} = (M_{2v\times 3} D)(D^{-1} S_{3\times n}) = M'_{2v\times 3} S'_{3\times n}.$$

To determine D, the so-called *metric constraints* were used. The different solutions for different special cases have been proposed by Tomasi and Kanade [25], Poelman and Kanade [20], Weinshall and Tomasi [30], Weishall [29], Shapiro et al. [24], Ullman and Basri [28]. These methods will be assembled into the same framework in terms of calibration matrix in Section 5. All these methods share a common point that it requires the knowledge of the intrinsic parameters of the camera, though it may have only one, the aspect ratio, for weak perspective case. It is the main purpose of this paper to determine this D in order to get Euclidean shape and motion without having to know the calibration of the camera. This will be achieved by a self-calibration step which follows.

4 Self calibration for the affine camera

4.1 Basic idea

In this section, we will introduce the concept of self-calibration idea for the affine camera following the original idea proposed for the perspective camera by Maybank and Faugeras in [16]. More practical methods have been reported in [6, 15, 9, 11, 3, 17] for self-calibration.

The basic idea of self-calibration is to use the invariance of unknown intrinsic parameters based on the point correspondences through the distinct views.

Although the concept of self-calibration is the same as that of Maybank and Faugeras [16], the practical approach of self-calibration that will be developed for the affine camera is quite different from that used by Maybank and Faugeras. It is more inspired by the recent work of Hartley in [10] for calibrating a rotating camera.

4.2 Formulation

For the given v views and tracked n points, from the previous section, the camera matrix $M_{2v\times n}$ and shape matrix $3 \times n$ are obtained up to an unknown 3×3 matrix D by factorizing the measurement matrix $W_{2v\times n}$.

For each view i, the 2×3 camera matrix M_i is actually defined only up to an unknown linear transformation D. The constraint that may be imposed on D is that M_iD should correspond to the real camera matrix for each view i, that is, M_iD has to be decomposed into its real intrinsic A_i and extrinsic R_i parameters matrices such as

$$M_iD = A_iR_i. \tag{3}$$

Multiply each side of (3) by its transpose, and observe that for the 2×3 orthogonal matrix R_i, $R_iR_i^T = I$. (3) becomes

$$M_iDD^TM_i^T = A_iR_iR_i^TA_i^T = A_iA_i^T. \tag{4}$$

Now, let $X = DD^T$ and $M_i = \begin{pmatrix} m_i^T \\ n_i^T \end{pmatrix}$, the equation (4) is

$$\begin{pmatrix} m_i^TXm_i & m_i^TXn_i \\ m_i^TXn_i & n_i^TXn_i \end{pmatrix} = A_iA_i^T. \tag{5}$$

In the self-calibration context, if we assume that the images are taken by the same affine camera, therefore the intrinsic parameters of the camera except the individual scaling factor for each view remain invariant during the acquisition. This means that we have the following expression:

$$k_1A_1 = k_2A_2 = \ldots = k_vA_v.$$

It follows,

$$k_1^2A_1A_1^T = k_2^2A_2A_2^T = \ldots = k_v^2A_vA_v^T.$$

In terms of X and M_i according to the equation (5),

$$k_1^2\begin{pmatrix} m_1^TXm_1 & m_1^TXn_1 \\ m_1^TXn_1 & n_1^TXn_1 \end{pmatrix} = \ldots = k_v^2\begin{pmatrix} m_v^TXm_v & m_v^TXn_v \\ m_v^TXn_v & n_v^TXn_v \end{pmatrix}.$$

For each pair of views (i, j), by taking the ratios of the matrix entries to cancel the different scaling factor k_i and k_j, we obtain

$$\begin{cases} \frac{m_i^T X m_i}{n_i^T X n_i} & = & \frac{m_j^T X m_j}{n_j^T X n_j} \\ \frac{m_i^T X n_i}{n_i^T X n_i} & = & \frac{m_j^T X n_j}{n_j^T X n_j} \end{cases}$$

Each such equation is quadratic and homogeneous in the 6 entries of X (X is symmetric).

For v views, $2(v-1)$ such independent equations are obtained. The solution for X can only be determined up to a common scalar due to the homogeneity of the equations. X counts therefore only for 5 independent parameters. With a minimum of 4 views, it provides $2 \times (4-1) = 6$ equations for the 5 unknowns. A numerical solution can be expected.

4.3 Cholesky parameterization

If we compute first X, then to get D, Cholesky decomposition should be applied to X to get D such as $X = DD^T$. This has been widely used to derive linear algorithms in previous work on shape and motion from the calibrated cameras, except for [25] in which a non linear algorithm is preferred. However, this decomposition is possible and unique if and only if the computed X is positive-definite. In case of noised data, no matter what method is used to solve for X, it may fail to be decomposable into DD^T due to the non positive-definiteness of the computed X.

In the context of self-calibration, even less information than calibrated case is available. In practice, the experiments show that the computed X by self-calibration is rarely positive-definite. The constraint of positive-definiteness should be explicitly imposed on X. This constraint is essentially non linear. Without loss of generality, a positive-definite X can always be parameterized (by invoking Cholesky decomposition [7]) by ZZ^T where Z is a lower triangular matrix,

$$Z = \begin{pmatrix} z_1 & 0 & 0 \\ z_2 & z_3 & 0 \\ z_4 & z_5 & z_6 \end{pmatrix}.$$

A multiplicative constant on Z makes X multiplied by the square of this constant. The homogeneity of X is transmitted to Z. Therefore we assume $z_6 = 1$ without losss of generality ($z_6 = 0$ will make Z singular). This Cholesky reparameterization, although non linear, maintains the minimum of 5 independent parameters.

If z denotes the six vector $(z_1, z_2, \ldots, z_5, 1)$ of the entries of Z, the self-calibration is formulated as the following unconstraint minimization problem.

$$\min f(z) = \sum_{i=1}^{v-1} \left(\frac{m_i^T X m_i}{n_i^T X n_i} - \frac{m_{i+1}^T X m_{i+1}}{n_{i+1}^T X n_{i+1}}\right)^2 + \left(\frac{m_i^T X n_i}{n_i^T X n_i} - \frac{m_{i+1}^T X n_{i+1}}{n_{i+1}^T X n_{i+1}}\right)^2, \quad \text{with} \quad X = ZZ^T. \qquad (6)$$

This can be iteratively solved by, for instance, Levenberg-Marquart method [21]. The initial solution can be either provided by hand (typically Z can be taken to be the identity) or by other linear methods. This is still discussed in Section 6.

4.4 Shape and motion from Cholesky parameterization

Once Z is computed as above, it remains to determine the D such that each $M_i D$ turns out to be decomposed into the real camera matrix for each view i. By definition, $X = ZZ^T = DD^T$, where Z

is lower triangular and D is any non singular. Applying QR decomposition to D, $D = ZR$, where R is orthonormal. Therefore, D is determined up to a rotation R for the given Z. This rotation matrix R can be fixed by taking any one of the views as the reference view, that is, if the view i is taken as the reference view, QR decomposes $M_i Z$ into AR_i, then the unknown rotation can be taken to be $R = R_i^T$. Finally,

$$D = ZR_i^T.$$

4.5 Some special cases of self-calibration

The above self-calibration method is developed for a general affine camera. It can be further simplified if additional simplistic assumptions are taken for the affine camera model.

4.5.1 Weak perspective

If the affine camera is assumed to be one of the special cases, the weak perspective camera, the intrinsic parameters matrix for the weak perspective has a diagonal form. This suggests a special method for the weak perspective camera.

Recall for weak perspective,

$$AA^T = k^2 \begin{pmatrix} \xi^2 & 0 \\ 0 & 1 \end{pmatrix}$$

For each individual view i, observing the off-diagonal entries are zero, we have

$$m_i^T X n_i = 0. \tag{7}$$

For each pair of views (i, j),

$$\frac{m_i^T X m_i}{n_i^T X n_i} = \frac{m_j^T X m_j}{n_j^T X n_j}. \tag{8}$$

With at least 3 views, we obtain 3 linear equations of type (7) from each individual view and 2 quadratic equations from two pairs of the 3 views. An algebraic solution with at most 4 solutions can be expected from this system of equations.

Generally, when v views are available, the self-calibration of the weak perspective camera can be formulated as the following minimization problem.

$$\min \phi(z) = \sum_{i=1}^{v-1} \left(\frac{m_i^T X m_i}{n_i^T X n_i} - \frac{m_{i+1}^T X m_{i+1}}{n_{i+1}^T X n_{i+1}}\right)^2 + \sum_{i=1}^{v} (m_i^T X n_i)^2, \quad \text{with} \quad X = ZZ^T. \tag{9}$$

It is also interesting to remark that with at least 5 views, a linear solution is possible by using only the linear equations of type (7).

4.5.2 Fixed (unknown) scaling factor

If we make a simplistic assumption that all views have the same unknown scaling factor (this is possible when the motion is small or mainly rotational aroud the object), that means that all three intrinsic parameters including the individual scaling factor are invariant. We have the following expression in terms of calibration matrices,

$$A_1 = \ldots = A_v.$$

The self-calibration equations are therefore the following expressions:

$$\begin{pmatrix} m_1^T X m_1 & m_1^T X n_1 \\ m_1^T X n_1 & n_1^T X n_1 \end{pmatrix} = \ldots = \begin{pmatrix} m_v^T X m_v & m_v^T X n_v \\ m_v^T X n_v & n_v^T X n_v \end{pmatrix}.$$

For each pair of views (i, j),

$$\begin{cases} m_i^T X m_i &= m_j^T X m_j \\ m_i^T X n_i &= m_j^T X n_j \\ n_i^T X n_i &= n_j^T X n_j \end{cases}$$

Each such equation is homogeneous and linear in the entries of X. With $v \geq 3$ views, X can be solved by linear least squares method. For instance, with $\|x\|^2 = 1$, a solution is obtained by SVD.

5 Related work

A lot of work has been reported in the literature for shape and motion from image sequences under orthographic, weak perspective and para-perspective projection. All existing algorithms require the intrinsic parameters of the camera to be known.

Using the intrinsic parameters of the affine camera introduced in this paper, the previous work can all be reformulated as solving the following equations for X with known A_i for each view,

$$\begin{pmatrix} m_i^T X m_i & m_i^T X n_i \\ m_i^T X n_i & n_i^T X n_i \end{pmatrix} = A_i A_i^T. \tag{10}$$

The above equation system is the same as the equation (5) except that in (5) $A_i A_i^T$ is unknown.

- Tomasi and Kanade [25] initially used orthographic projection, it is equivalent to have

  $$A_i A_i^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

  thus the metric constraints are, for each view,

  $$\begin{cases} m_i^T X n_i = 0 \\ m_i^T X m_i = 1 \\ n_i^T X n_i = 1. \end{cases}$$

- Weinshall and Tomasi [30] and Weinshall [29] worked with the weak perspective camera. The method proposed in these papers needs to first select 4 reference points, then computes their Gramian. In fact, this can be easily extended to a batch solution for general n points without need of explicitly selecting 4 reference points. In the calibrated weak perspective case, the knowledge of the intrinsic parameters for each view is coded in

  $$A_i A_i^T = k_i^2 \begin{pmatrix} \xi_i^2 & 0 \\ 0 & 1 \end{pmatrix}, \tag{11}$$

  it suffices to solve the following equations derived for each view,

  $$\begin{cases} m_i^T X n_i = 0 \\ \frac{m_i^T X m_i}{n_i^T X n_i} = \xi_i^2. \end{cases} \tag{12}$$

- For the general para-perspective, the known intrinsic calibration matrix is given as

$$A_i A_i^T = k_i^2 \begin{pmatrix} \xi_i^2(1+\cot\theta_{x_i}^2) & \xi_i \cot\theta_{x_i}\cot\theta_{y_i} \\ \xi_i \cot\theta_{x_i}\cot\theta_{y_i} & 1+\cot\theta_{y_i}^2 \end{pmatrix}.$$

  Poelman and Kanade [20] used a special para-perspective model due to Otha [19], by projecting along the direction from the camera's focal point to the object's centroid. First the normalized coordinates are computed by applying the calibration matrix in each image plane for image points. Then, if (x_i, y_i) denotes the normalized coordinates of the centroid in the view i. The direction of projection are estimated by

$$\cot\theta_{x_i} = x_i \quad \text{and} \quad \cot\theta_{y_i} = y_i,$$

  and the intrinsic matrix of the para-perspective camera is simplified into

$$A_i A_i^T = k_i^2 \begin{pmatrix} 1+x_i^2 & x_i y_i \\ x_i y_i & 1+y_i^2 \end{pmatrix}.$$

  It remains to solve the following equations:

$$\begin{cases} \frac{m_i^T X n_i}{n_i^T X n_i} = \frac{x_i y_i}{1+y_i^2}, \\ \frac{m_i^T X m_i}{n_i^T X n_i} = \frac{1+x_i^2}{1+y_i^2}. \end{cases}$$

Remark All these equations can be linearly solved for X, but they suffer from the problem of non positive-definiteness of X in presence of noise. Non-linear methods can be applied using Cholesky parameterization of X introduced in Section 4. Note that the initial method using metric constraints of Tomasi and Kanade [25] did not use linear algorithm. They explicitly solved for DD^T instead of X, however how D is parameterized is not clear in [25].

As for self-calibration, as we have mentioned in Section 4 that there have been several practical methods available in the literature. All of them are developed for the perspective camera with more or less free intrinsic parameters.

6 Experimental results

The self-calibration methods developed in this paper are implemented and then applied to real image sequences. The reconstruction results obtained from the self-calibration are also compared with those obtained by the calibrated cameras.

Poelman and Kanade at CMU kindly provided the hotel image sequence that they have used in [20]. In this sequence, the camera motion included substantial translation away from the camera and across the field of view. 197 points throughout the sequence of 181 images are automatically identified and tracked. For a more detailed description of this set-up, one can consult [20].

In our experiments, not all images of the sequence are used, only 10 out of 181 images are selected from an interval of 20 images. Four of them are illustrated in Figure 1 to have an idea of the image sequence. The shape reconstruction by self-calibration is compared with that of calibrated para-perspective provided by Poelman and Kanade [20].

The self-calibration algorithm uses the minimization equation (6). The starting point for the matrix Z is always set to be the identity.

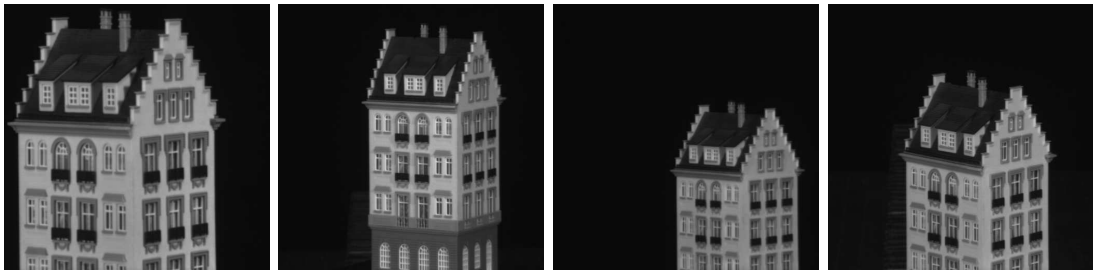

Figure 1: Image 1, 101, 151 and 181 of the hotel model image sequences.

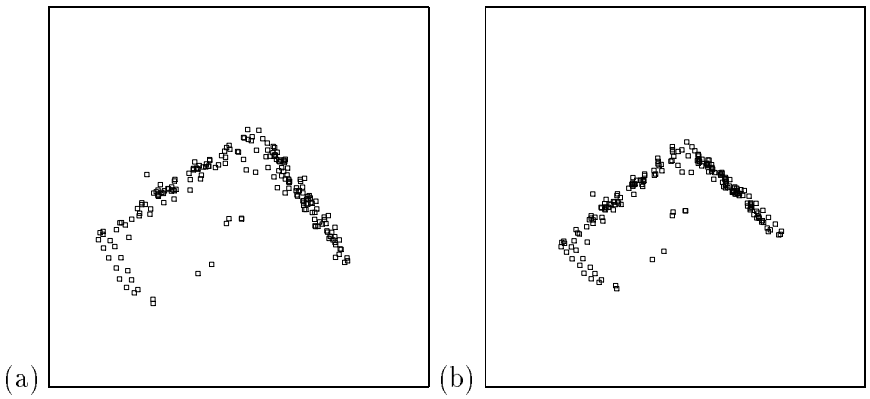

Figure 2: Top view of the reconstructed shape: (a) self-calibrated affine camera and (b) calibrated para-perspective.

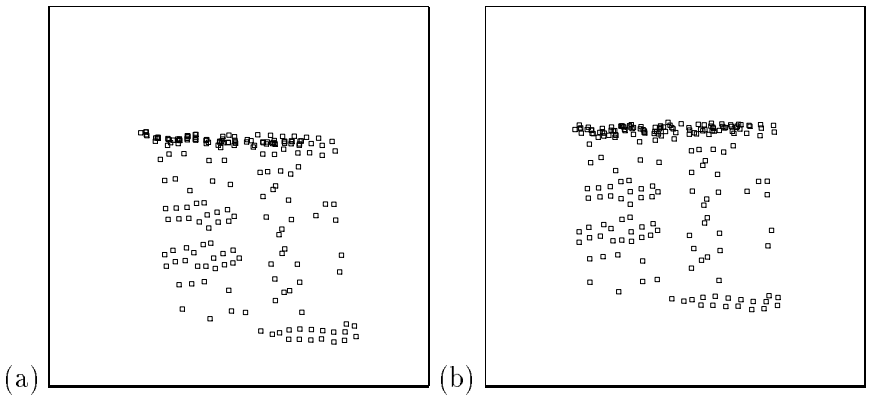

Figure 3: Front view of the reconstructed shape: (a) self-calibrated affine camera and (b) calibrated para-perspective.

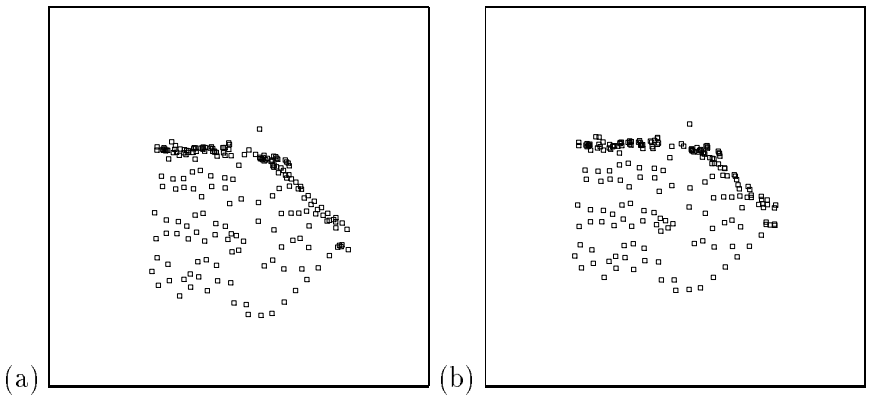

Figure 4: Side view of the reconstructed shape: (a) self-calibrated affine camera and (b) calibrated para-perspective.

In Figures 2, 3 and 4, the top, front and side views of the reconstructed shape are displayed and compared with the calibrated para-perspective solution. From the front view, it looks less accurate than the calibrated para-perspective solution, since the right angle of the wall is still slightly deformed. However the top view looks slightly better and also the side view tells us the depth of the hotel model is correctly recovered.

Another experiment is conducted on the sequence of images of a cube, kindly provided by Boufama and used in [2]. The sequence is designed to cover a small field of view so that the weak perspective approximation is appropriate [2]. To compare the self-calibration method with the calibrated method, the batch solution for the calibrated weak perspective using the equations (12) is implemented and computed, it actually turned out very good results illustrated in Figure 6.(c).

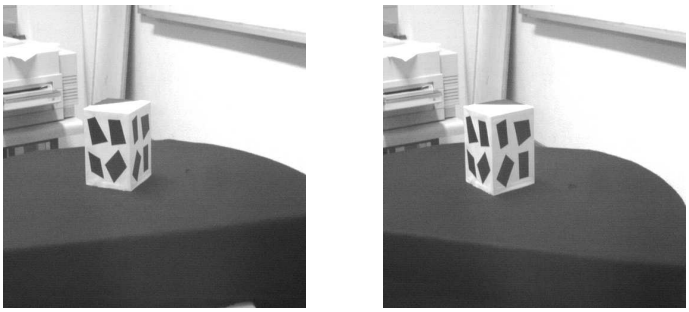

Figure 5: Image 1 and 5 of the image sequences of the cube.

For self-calibration, the weak perspective projection model is still assumed, hence the self-calibration algorithm uses the minimization equation (9). The results of the self-calibrated weak perspective are presented in Figure (6) in which one observes that the self-calibration gives less good results than the calibrated weak perspective, however Euclidean shape is greatly improved with respect to the initial affine shape illustrated in Figure 6.(a).

From the experimentation that we have performed, the self-calibration algorithms performed well and greatly improved the intial affine reconstruction. Compared with the calibrated cases, the reconstruction is slightly less accurate, it is not surprising as the self-calibration algorithm uses much less information than the algorithms for the calibrated cases.

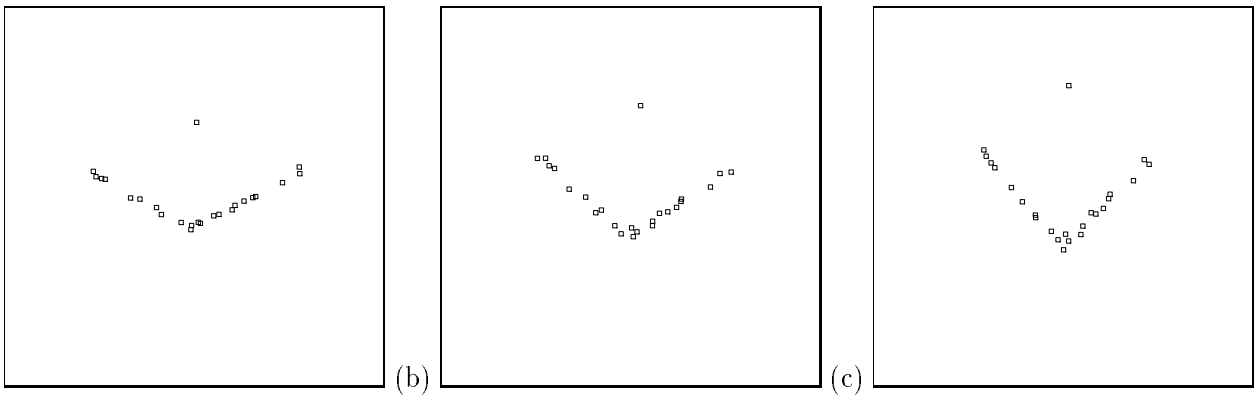

Figure 6: Top view of the reconstructed shape (a) the initial affine shape, (b) self-calibrated weak perspective and (c) calibrated weak perspective.

7 Discussion

A new method for shape and motion from image sequences taken by an uncalibrated affine camera is proposed in this paper. The method differs from all others in that it operates with uncalibrated affine camera.

The method is developed with the affine camera model. The self-calibration of the affine camera is based on the introduction of the three intrinsic parameters for the affine camera. It was also shown that these intrinsic parameters of the affine camera are closely related to those of the perspective camera, but generally different. The previous methods requiring the knowledge of calibration parameters can also be nicely assembled into the same framework in terms of the intrinsic calibration matrix of the affine camera, therefore can be easily integrated into the same numerical process for the affine camera. Therefore, The experimental results show the good performance of the self-calibration method.

Acknowledgement

This work is partly supported by European Esprit BRA projects Viva which is gratefully acknowledged. This paper benefits from the discussions with R. Hartley.

References

[1] J.Y. Aloimonos. Perspective approximation. *Image and Vision Computing*, 8(3):179–192, 1992.

[2] B. Boufama, D. Weinshall, and M. Werman. Shape from motion algorithms: a comparative analysis of scaled orthography and perspective. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 199–204. Springer-Verlag, May 1994.

[3] F. Du and M. Brady. Self-calibration of the intrinsic parameters of cameras for active vision systems. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 477–482, Los Alamitos, California, 1993. IEEE Computer Society Press.

[4] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, May 1992.

[5] O. Faugeras. *Three-Dimensional Computer Vision - A Geometric Viewpoint.* Artificial intelligence. M.I.T. Press, Cambridge, MA, 1993.

[6] O.D. Faugeras, Q.T. Luong, and S.J. Maybank. Camera Self-Calibration: Theory and Experiments. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 321–334. Springer-Verlag, May 1992.

[7] G.H. Golub and C.F. Van Loan. *Matrix Computation.* The Johns Hopkins University Press, Baltimore, 1989.

[8] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.

[9] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 579–587. Springer-Verlag, 1992.

[10] R.I. Hartley. An algorithm for self calibration from several views. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA*, pages 908–912, 1994.

[11] R.I. Hartley. Self-calibration from multiple views with a rotating camera. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 471–478. Springer-Verlag, 1994.

[12] R. Horaud, S. christy, and F. Dornaika. Object pose: the link between weak perspective, para perspective, and full perspective. Technical report, INRIA, 1994.

[13] J.J. Koenderink and A. J. van Doorn. Affine structure from motion. Technical report, Utrecht University, Utrecht, The Netherlands, October 1989.

[14] C.H. Lee and T. Huang. Finding point correspondences and determining motion of a rigid object from two weak perspective views. *Computer Vision, Graphics and Image Processing*, 52:309–327, 1990.

[15] Q.T. Luong. *Matrice Fondamentale et Autocalibration en Vision par Ordinateur.* Thèse de doctorat, Université de Paris-Sud, Orsay, France, December 1992.

[16] S.J. Maybank and O.D. Faugeras. A theory of self calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.

[17] R. Mohr, L. Quan, and F. Veillon. Relative 3D reconstruction using multiple uncalibrated images. *The International Journal of Robotics Research*, 1995. to appear.

[18] J.L. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision.* MIT Press, Cambridge, Massachusetts, USA, 1992.

[19] Y. Otha, K. Maenobu, and T. Sakai. Obtaining surface orientation from texels under perspective projection. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 746–751, 1981.

[20] C. J. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 97–108, May 1994.

[21] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling W.T. *Numerical Recipes in C.* Cambridge University Press, 1988.

[22] L. Quan. Invariants of six points and projective reconstruction from three uncalibrated images. IEEE *Transactions on PAMI*, 17(1), January 1995.

[23] L. Quan and R. Mohr. Affine shape representation from motion through reference points. *Journal of Mathematical Imaging and Vision*, 1:145–151, 1992. also in IEEE Workshop on Visual Motion, New Jersey, pages 249–254, 1991.

[24] L.S. Shapiro, A. Zisserman, and M. Brady. Motion from point matches using affine epipolar geometry. *International Journal of Computer Vision*, 1994.

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

[26] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE *Journal of Robotics and Automation*, 3(4):323–344, 1987.

[27] S. Ullman. *The Interpretation of Visual Motion*. The MIT Press, 1979.

[28] S. Ullman and R. Basri. Recognition by linear combinations of models. IEEE *Transactions on PAMI*, 13(10):992–1006, 1991.

[29] D. Weinshall. Model-based invariants for 3-D vision. *International Journal of Computer Vision*, 10(1):27–42, 1993.

[30] D. Weinshall and C. Tomasi. Linear and incremental acquisition of invariant shape models from image sequences. In *Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany*. IEEE, 1993.