Uncalibrated Factorization Using a Variable Symmetric Affine Camera

Kenichi Kanatani[1], Yasuyuki Sugaya[2], and Hanno Ackermann[1]

[1] Department of Computer Science, Okayama University, Okayama 700-8530, Japan
[2]Department of Information and Computer Sciences,
Toyohashi University of Technology, Toyohashi, Aichi 441-8480, Japan
{kanatani, sugaya, hanno}@suri.it.okayama-u.ac.jp

Abstract. In order to reconstruct 3-D Euclidean shape by the Tomasi-Kanade factorization, one needs to specify an affine camera model such as orthographic, weak perspective, and paraperspective. We present a new method that does not require any such specific models. We show that a minimal requirement for an affine camera to mimic perspective projection leads to a unique camera model, called *symmetric affine camera*, which has two free functions. We determine their values from input images by linear computation and demonstrate by experiments that an appropriate camera model is automatically selected.

1 Introduction

One of the best known techniques for 3-D reconstruction from feature point tracking through a video stream is the Tomasi-Kanade *factorization* [10], which computes the 3-D shape of the scene by approximating the camera imaging by an affine transformation. The computation consists of linear calculus alone without involving iterations [5]. The solution is sufficiently accurate for many practical purposes and is used as an initial solution for more sophisticated iterative reconstruction based on perspective projection [2].

If the camera model is not specified, other than being affine, the 3-D shape is computed only up to an affine transformation, known as *affine reconstruction* [9]. For computing the correct shape (*Euclid reconstruction*), we need to specify the camera model. For this, *orthographic*, *weak perspective*, and *paraperspective* projections have been used [7]. However, the reconstruction accuracy does not necessarily follow in that order [1]. To find the best camera models in a particular circumstance, one needs to choose the best one *a posteriori*. Is there any method for automatically selecting an appropriate camera model?

Quan [8] showed that a generic affine camera has three intrinsic parameters and that they can be determined by self-calibration if they are fixed. The intrinsic parameters cannot be determined if they vary freely. The situation is similar to the *dual absolute quadric constraint* [2] for upgrading projective reconstruction to Euclidean, which cannot be imposed unless something is known about the camera (e.g., zero skew).

In this paper, we show that minimal requirements for the general affine camera to mimic perspective projection leads to a unique camera model, which we

A. Leonardis, H. Bischof, and A. Prinz (Eds.): ECCV 2006, Part IV, LNCS 3954, pp. 147-158, 2006.

call a *symmetric affine camera*, having two free functions of motion parameters; specific choices of their function forms result in the orthographic, weak perspective, and paraperspective models.

However, we need not specify such function forms. We can determine their values directly from input images. All the computation is linear just as in the case of the traditional factorization method, and an appropriate model is automatically selected.

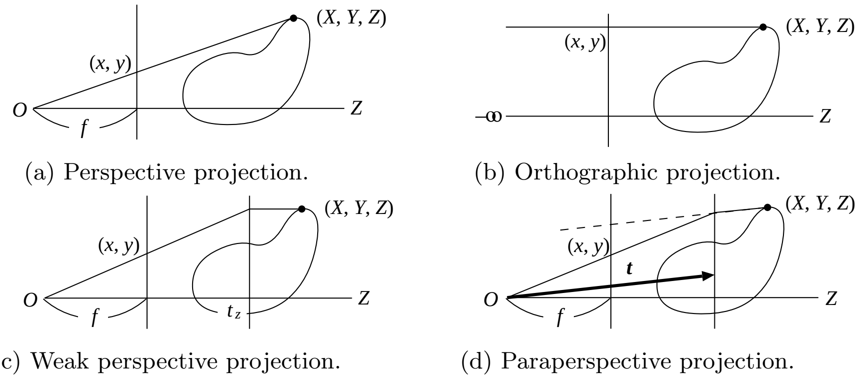
Sec. 2 summarizes fundamentals of affine cameras, and Sec. 3 summarizes the metric constraint. In Sec. 4, we derive our symmetric affine camera model. Sec. 5 describes the procedure for 3-D reconstruction using our model. Sec. 6 shows experiments, and Sec. 7 concludes this paper.

2 Affine Cameras

Consider a camera-based XYZ coordinate system with the origin O at the projection center and the Z axis along the optical axis. *Perspective projection* maps a point (X, Y, Z) in the scene onto a point with image coordinates (x, y) such that

$$x = f\frac{X}{Z}, \qquad y = f\frac{Y}{Z}, \tag{1}$$

where f is a constant called the *focal length* (Fig. 1(a)).



(a) Perspective projection.

(b) Orthographic projection.

(c) Weak perspective projection.

(d) Paraperspective projection.

Fig. 1. Camera models.

Consider a world coordinate system fixed to the scene, and let $\boldsymbol{t}$ and $\{\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\}$ be its origin and the orthonormal basis vectors described with respect to the camera coordinate system. We call $\boldsymbol{t}$ the *translation*, the matrix $\boldsymbol{R} = \begin{pmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \end{pmatrix}$ having $\{\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\}$ as columns the *rotation*, and $\{\boldsymbol{t}, \boldsymbol{R}\}$ the *motion parameters*.

If (i) the object of our interest is localized around the world coordinate origin $\boldsymbol{t}$, and (ii) the size of the object is small as compared with $\|\boldsymbol{t}\|$, the imaging can be approximated by an *affine camera* [9] in the form

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \mathit{\Pi}_{11} & \mathit{\Pi}_{12} & \mathit{\Pi}_{13} \\ \mathit{\Pi}_{21} & \mathit{\Pi}_{22} & \mathit{\Pi}_{23} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}. \tag{2}$$

We call the 2×3 matrix $\boldsymbol{\Pi} = (\mathit{\Pi}_{ij})$ and the 2-D vector $\boldsymbol{\pi} = (\pi_i)$ the *projection matrix* and the *projection vector*, respectively; their elements are "functions" of the motion parameters $\{\boldsymbol{t}, \boldsymbol{R}\}$. The intrinsic parameters are *implicitly* defined via the functional forms of $\{\boldsymbol{\Pi}, \boldsymbol{\pi}\}$ on $\{\boldsymbol{t}, \boldsymbol{R}\}$, e.g., as coefficients. Typical affine cameras are

Orthographic projection (Fig. 1(b))

$$\boldsymbol{\Pi} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \boldsymbol{\pi} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{3}$$

Weak perspective projection (Fig. 1(c))

$$\boldsymbol{\Pi} = \begin{pmatrix} f/t_z & 0 & 0 \\ 0 & f/t_z & 0 \end{pmatrix}, \qquad \boldsymbol{\pi} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{4}$$

Paraperspective projection (Fig. 1(d))

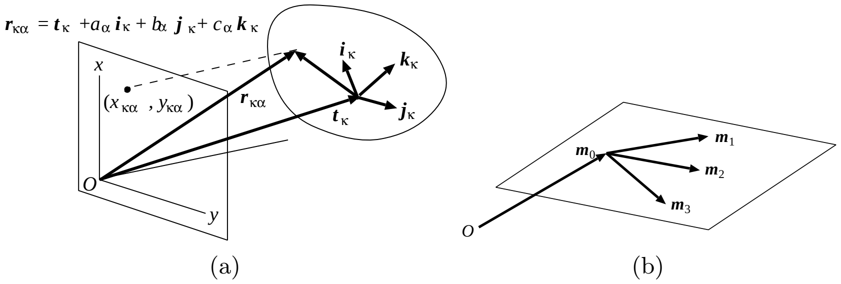
$$\boldsymbol{\Pi} = \begin{pmatrix} f/t_z & 0 & -ft_x/t_z^2 \\ 0 & f/t_z & -ft_x/t_z^2 \end{pmatrix}, \qquad \boldsymbol{\pi} = \begin{pmatrix} ft_x/t_z \\ ft_y/t_z \end{pmatrix}. \tag{5}$$

Suppose we track N feature points over M frames. Identifying the frame number κ with "time", let $\boldsymbol{t}_\kappa$ and $\{\boldsymbol{i}_\kappa, \boldsymbol{j}_\kappa, \boldsymbol{k}_\kappa\}$ be the origin and the basis vectors of the world coordinate system at time κ (Fig. 2(a)). The 3-D position of the αth point at time κ has the form

$$\boldsymbol{r}_{\kappa\alpha} = \boldsymbol{t}_\kappa + a_\alpha \boldsymbol{i}_\kappa + b_\alpha \boldsymbol{j}_\kappa + c_\alpha \boldsymbol{k}_\kappa. \tag{6}$$

Under the affine camera of eq. (2), its image coordinates $(x_{\kappa\alpha}, y_{\kappa\alpha})$ are given by

$$\begin{pmatrix} x_{\kappa\alpha} \\ y_{\kappa\alpha} \end{pmatrix} = \tilde{\boldsymbol{t}}_\kappa + a_\alpha \tilde{\boldsymbol{i}}_\kappa + b_\alpha \tilde{\boldsymbol{j}}_\kappa + c_\alpha \tilde{\boldsymbol{k}}_\kappa, \tag{7}$$


$\boldsymbol{r}_{\kappa\alpha} = \boldsymbol{t}_\kappa + a_\alpha \boldsymbol{i}_\kappa + b_\alpha \boldsymbol{j}_\kappa + c_\alpha \boldsymbol{k}_\kappa$
x
$(x_{\kappa\alpha}, y_{\kappa\alpha})$
$\boldsymbol{r}_{\kappa\alpha}$
$\boldsymbol{i}_\kappa$
$\boldsymbol{k}_\kappa$
$\boldsymbol{t}_\kappa$
$\boldsymbol{j}_\kappa$
O
y
$\boldsymbol{m}_0$
$\boldsymbol{m}_1$
$\boldsymbol{m}_2$
$\boldsymbol{m}_3$
O

(a) (b)

Fig. 2. (a) Camera-based description of the world coordinate system. (b) Affine space constraint.

where $\tilde{\boldsymbol{t}}_\kappa$, $\tilde{\boldsymbol{i}}_\kappa$, $\tilde{\boldsymbol{j}}_\kappa$, and $\tilde{\boldsymbol{k}}_\kappa$ are 2-D vectors defined by

$$\tilde{\boldsymbol{t}}_\kappa = \boldsymbol{\Pi}_\kappa \boldsymbol{t}_\kappa + \boldsymbol{\pi}_\kappa, \qquad \tilde{\boldsymbol{i}}_\kappa = \boldsymbol{\Pi}_\kappa \boldsymbol{i}_\kappa, \qquad \tilde{\boldsymbol{j}}_\kappa = \boldsymbol{\Pi}_\kappa \boldsymbol{j}_\kappa, \qquad \tilde{\boldsymbol{k}}_\kappa = \boldsymbol{\Pi}_\kappa \boldsymbol{k}_\kappa. \tag{8}$$

Here, $\boldsymbol{\Pi}_\kappa$ and $\boldsymbol{\pi}_\kappa$ are the projection matrix and the projective vector, respectively, at time κ. The motion history of the αth point is represented by a vector

$$\boldsymbol{p}_\alpha = \begin{pmatrix} x_{1\alpha} & y_{1\alpha} & x_{2\alpha} & y_{2\alpha} & \dots & x_{M\alpha} & y_{M\alpha} \end{pmatrix}^\top, \tag{9}$$

which we simply call the *trajectory* of that point. Using eq. (7), we can write

$$\boldsymbol{p}_\alpha = \boldsymbol{m}_0 + a_\alpha \boldsymbol{m}_1 + b_\alpha \boldsymbol{m}_2 + c_\alpha \boldsymbol{m}_3, \tag{10}$$

where $\boldsymbol{m}_0$, $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and $\boldsymbol{m}_3$ are the following $2M$-dimensional vectors:

$$\boldsymbol{m}_0 = \begin{pmatrix} \tilde{\boldsymbol{t}}_1 \\ \tilde{\boldsymbol{t}}_2 \\ \vdots \\ \tilde{\boldsymbol{t}}_M \end{pmatrix}, \quad \boldsymbol{m}_1 = \begin{pmatrix} \tilde{\boldsymbol{i}}_1 \\ \tilde{\boldsymbol{i}}_2 \\ \vdots \\ \tilde{\boldsymbol{i}}_M \end{pmatrix}, \quad \boldsymbol{m}_2 = \begin{pmatrix} \tilde{\boldsymbol{j}}_1 \\ \tilde{\boldsymbol{j}}_2 \\ \vdots \\ \tilde{\boldsymbol{j}}_M \end{pmatrix}, \quad \boldsymbol{m}_3 = \begin{pmatrix} \tilde{\boldsymbol{k}}_1 \\ \tilde{\boldsymbol{k}}_2 \\ \vdots \\ \tilde{\boldsymbol{k}}_M \end{pmatrix}. \tag{11}$$

Thus, all the trajectories $\{\boldsymbol{p}_\alpha\}$ are constrained to be in the 3-D affine space $\mathcal{A}$ in $\mathcal{R}^{2M}$ passing through $\boldsymbol{m}_0$ and spanned by $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and $\boldsymbol{m}_3$ (Fig. 2(b)). This fact is known as the *affine space constraint*.

3 Metric Constraint

Since the world coordinate system can be placed arbitrarily, we let its origin coincide with the centroid of the N feature points. This implies $\sum_{\alpha=1}^N a_\alpha = \sum_{\alpha=1}^N b_\alpha = \sum_{\alpha=1}^N c_\alpha = 0$, so we have from eq. (10)

$$\frac{1}{N}\sum_{\alpha=1}^N \boldsymbol{p}_\alpha = \boldsymbol{m}_0, \tag{12}$$

i.e., $\boldsymbol{m}_0$ is the centroid of the trajectories $\{\boldsymbol{p}_\alpha\}$ in $\mathcal{R}^{2M}$. It follows that the deviation $\boldsymbol{p}'_\alpha$ of $\boldsymbol{p}_\alpha$ from the centroid $\boldsymbol{m}_0$ is written as[1]

$$\boldsymbol{p}'_\alpha = \boldsymbol{p}_\alpha - \boldsymbol{m}_0 = a_\alpha \boldsymbol{m}_1 + b_\alpha \boldsymbol{m}_2 + c_\alpha \boldsymbol{m}_3, \tag{13}$$

which means that $\{\boldsymbol{p}'_\alpha\}$ are constrained to be in the 3-D subspace $\mathcal{L}$ in $\mathcal{R}^{2M}$. Hence, the matrix

$$\boldsymbol{C} = \sum_{\alpha=1}^N \boldsymbol{p}'_\alpha \boldsymbol{p}'^\top_\alpha \tag{14}$$

[1] In the traditional formulation [7, 10], vectors $\{\boldsymbol{p}'_\alpha\}$ are combined into the *measurement matrix*, $\boldsymbol{W} = \begin{pmatrix} \boldsymbol{p}'_1 & \dots & \boldsymbol{p}'_N \end{pmatrix}$, and the object coordinates $\{(a_\alpha, b_\alpha, c_\alpha)\}$ are combined into the *shape matrix*, $\boldsymbol{S} = \begin{pmatrix} a_1 & \dots & a_N \\ b_1 & \dots & b_N \\ c_1 & \dots & c_N \end{pmatrix}$. Then, eq. (13) is written as $\boldsymbol{W} = \boldsymbol{M}\boldsymbol{S}$, where $\boldsymbol{M}$, the *motion matrix*, is defined by the first of eqs. (16).

has rank 3, having three nonzero eigenvalues. The corresponding unit eigenvectors $\{\boldsymbol{u}_1,\ \boldsymbol{u}_2,\ \boldsymbol{u}_3\}$ constitute an orthonormal basis of the subspace $\mathcal{L}$, and $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and $\boldsymbol{m}_3$ are expressed as a linear combination of them in the form

$$\boldsymbol{m}_i = \sum_{j=1}^{3} A_{ji}\boldsymbol{u}_j. \tag{15}$$

Let $\boldsymbol{M}$ and $\boldsymbol{U}$ be the $2M \times 3$ matrices consisting of $\{\boldsymbol{m}_1,\ \boldsymbol{m}_2,\ \boldsymbol{m}_3\}$ and $\{\boldsymbol{u}_1,\ \boldsymbol{u}_2,\ \boldsymbol{u}_3\}$ as columns:

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{m}_1 & \boldsymbol{m}_2 & \boldsymbol{m}_3 \end{pmatrix}, \qquad \boldsymbol{U} = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \boldsymbol{u}_3 \end{pmatrix}. \tag{16}$$

From eq. (15), $\boldsymbol{M}$ and $\boldsymbol{U}$ are related by the matrix $\boldsymbol{A} = (A_{ij})$ in the form[2]:

$$\boldsymbol{M} = \boldsymbol{U}\boldsymbol{A}. \tag{17}$$

The rectifying matrix $\boldsymbol{A} = (A_{ij})$ is determined so that $\boldsymbol{m}_1$, $\boldsymbol{m}_2$ and $\boldsymbol{m}_3$ in eqs. (11) are projections of the orthonormal basis vectors $\{\boldsymbol{i}_\kappa,\ \boldsymbol{j}_\kappa,\ \boldsymbol{k}_\kappa\}$ in the form of eqs. (8). From eqs. (8), we obtain

$$\begin{pmatrix} \tilde{\boldsymbol{i}}_\kappa & \tilde{\boldsymbol{j}}_\kappa & \tilde{\boldsymbol{k}}_\kappa \end{pmatrix} = \boldsymbol{\Pi}_\kappa \begin{pmatrix} \boldsymbol{i}_\kappa & \boldsymbol{j}_\kappa & \boldsymbol{k}_\kappa \end{pmatrix} = \boldsymbol{\Pi}_\kappa \boldsymbol{R}_\kappa, \tag{18}$$

where $\boldsymbol{R}_\kappa$ is the rotation at time κ. If we let $\boldsymbol{m}^\dagger_{\kappa(a)}$ be the $(2(\kappa-1)+a)$th column of the transpose $\boldsymbol{M}^\top$ of the matrix $\boldsymbol{M}$ in eqs. (16), $\kappa = 1, ..., M$, $a = 1, 2$. The transpose of both sides of eq. (18) is

$$\boldsymbol{R}_\kappa^\top \boldsymbol{\Pi}_\kappa^\top = \begin{pmatrix} \boldsymbol{m}^\dagger_{\kappa(1)} & \boldsymbol{m}^\dagger_{\kappa(2)} \end{pmatrix}. \tag{19}$$

Eq. (17) implies $\boldsymbol{M}^\top = \boldsymbol{A}^\top \boldsymbol{U}^\top$, so if we let $\boldsymbol{u}^\dagger_{\kappa(a)}$ be the $(2(\kappa-1)+a)$th column of the transpose $\boldsymbol{U}^\top$ of the matrix $\boldsymbol{U}$ in eqs. (16), we obtain

$$\boldsymbol{m}^\dagger_{\kappa(a)} = \boldsymbol{A}^\top \boldsymbol{u}^\dagger_{\kappa(a)}. \tag{20}$$

Substituting this, we can rewrite eq. (19) as

$$\boldsymbol{R}_\kappa^\top \boldsymbol{\Pi}_\kappa^\top = \boldsymbol{A}^\top \begin{pmatrix} \boldsymbol{u}^\dagger_{\kappa(1)} & \boldsymbol{u}^\dagger_{\kappa(2)} \end{pmatrix}. \tag{21}$$

Let $\boldsymbol{U}^\dagger_\kappa$ the 3×2 matrix having $\boldsymbol{u}^\dagger_{\kappa(1)}$ and $\boldsymbol{u}^\dagger_{\kappa(2)}$ as columns:

$$\boldsymbol{U}^\dagger_\kappa = \begin{pmatrix} \boldsymbol{u}^\dagger_{\kappa(1)} & \boldsymbol{u}^\dagger_{\kappa(2)} \end{pmatrix}. \tag{22}$$

[2] In the traditional formulation [7, 10], the measurement matrix $\boldsymbol{W}$ is decomposed by the singular value decomposition into $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^\top$, and the motion and the shape matrices $\boldsymbol{M}$ and $\boldsymbol{S}$ are set to $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{A}$ an $\boldsymbol{S} = \boldsymbol{A}^{-1}\boldsymbol{\Lambda}\boldsymbol{V}^\top$ via a nonsingular matrix $\boldsymbol{A}$.

From eq. (21), we have $\boldsymbol{U}_\kappa^{\dagger\top}\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{U}_\kappa^\dagger = \boldsymbol{\Pi}_\kappa\boldsymbol{R}_\kappa\boldsymbol{R}_\kappa^\top\boldsymbol{\Pi}_\kappa^\top$. Since $\boldsymbol{R}_\kappa$ is a rotation matrix, we have the generic *metric constraint*

$$\boldsymbol{U}_\kappa^{\dagger\top}\boldsymbol{T}\boldsymbol{U}_\kappa^\dagger = \boldsymbol{\Pi}_\kappa\boldsymbol{\Pi}_\kappa^\top, \tag{23}$$

where we define the *metric matrix* $\boldsymbol{T}$ as follows:

$$\boldsymbol{T} = \boldsymbol{A}\boldsymbol{A}^\top. \tag{24}$$

Eq. (23) is the generic metric constraint given by Quan [8]. If we take out the elements on both sides, we have the following three expressions:

$$(\boldsymbol{u}_{\kappa(1)}^\dagger, \boldsymbol{T}\boldsymbol{u}_{\kappa(1)}^\dagger) = \sum_{i=1}^3 \Pi_{1i\kappa}^2, \qquad (\boldsymbol{u}_{\kappa(2)}^\dagger, \boldsymbol{T}\boldsymbol{u}_{\kappa(2)}^\dagger) = \sum_{i=1}^3 \Pi_{2i\kappa}^2,$$
$$(\boldsymbol{u}_{\kappa(1)}^\dagger, \boldsymbol{T}\boldsymbol{u}_{\kappa(2)}^\dagger) = \sum_{i=1}^3 \Pi_{1i\kappa}\Pi_{2i\kappa}. \tag{25}$$

These correspond to the *dual absolute quadric constraint* [2] on the homography that rectifies the basis of projective reconstruction to Euclidean.

We focus on the fact that *at most two* time varying unknowns of the camera model can be eliminated from eqs. (25). We show that (i) we can restrict the camera model without much impairing its descriptive capability so that it has *two* free functions and (ii) we can redefine them in such a way that the resulting $2M$ unknowns are *linearly* estimated.

4 Symmetric Affine Cameras

We now seek a concrete form of the affine camera by imposing minimal requirements that eq. (2) mimic perspective projection.

Requirement 1. *The frontal parallel plane passing through the world coordinate origin is projected as if by perspective projection.*

This corresponds to our assumption that the object of our interest is small and localized around the world coordinate origin (t_x, t_y, t_z). A point on the plane $Z = t_z$ is written as (X, Y, t_z), so Requirement 1 implies

$$\begin{pmatrix} fX/t_z \\ fY/t_z \end{pmatrix} = \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix} + t_z\begin{pmatrix} \Pi_{13} \\ \Pi_{23} \end{pmatrix} + \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}. \tag{26}$$

Since this should hold for arbitrary X and Y, we obtain

$$\Pi_{11} = \Pi_{22} = \frac{f}{t_z}, \quad \Pi_{12} = \Pi_{21} = 0, \quad t_z\Pi_{13} + \pi_1 = 0, \quad t_z\Pi_{23} + \pi_2 = 0, \tag{27}$$

which reduces eq. (2) to

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{f}{t_z}\begin{pmatrix} X \\ Y \end{pmatrix} - (t_z - Z)\begin{pmatrix} \Pi_{13} \\ \Pi_{23} \end{pmatrix}, \tag{28}$$

where f, Π_{13} and Π_{23} are arbitrary functions of $\{\boldsymbol{t}, \boldsymbol{R}\}$. In order to obtain a more specific form, we impose the following requirements:

Requirement 2. *The camera imaging is symmetric around the Z-axis.*

Requirement 3. *The camera imaging does not depend on $\boldsymbol{R}$.*

Requirement 2 states that if the scene is rotated around the optical axis by an angle θ, the resulting image should also rotate around the image origin by the same angle θ, a very natural requirement. Requirement 3 is also natural, since the orientation of the world coordinate system can be defined arbitrarily, and such indeterminate parameterization should not affect the actual observation.

Let $\mathcal{R}(\theta)$ be the 2-D rotation matrix by angle θ:

$$\mathcal{R}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{29}$$

Requirement 2 is written as

$$\mathcal{R}(\theta)\begin{pmatrix} x \\ y \end{pmatrix} = \frac{f}{t_z}\mathcal{R}(\theta)\begin{pmatrix} X \\ Y \end{pmatrix} - (t_z - Z)\begin{pmatrix} \Pi'_{13} \\ \Pi'_{23} \end{pmatrix}, \tag{30}$$

where Π'_{13} and Π'_{23} are the values of the functions Π_{13} and Π_{23}, respectively, obtained by replacing t_x and t_y in their arguments by $t_x\cos\theta - t_y\sin\theta$ and $t_x\sin\theta + t_y\cos\theta$, respectively; by Requirement 3, the arguments of Π_{13} and Π_{23} do not contain $\boldsymbol{R}$. Multiplying both sides of eq. (28) by $\mathcal{R}(\theta)$, we obtain

$$\mathcal{R}(\theta)\begin{pmatrix} x \\ y \end{pmatrix} = \frac{f}{t_z}\mathcal{R}(\theta)\begin{pmatrix} X \\ Y \end{pmatrix} - (t_z - Z)\mathcal{R}(\theta)\begin{pmatrix} \Pi_{13} \\ \Pi_{23} \end{pmatrix}. \tag{31}$$

Comparing eqs. (30) and (31), we conclude that the equality

$$\begin{pmatrix} \Pi'_{13} \\ \Pi'_{23} \end{pmatrix} = \mathcal{R}(\theta)\begin{pmatrix} \Pi_{13} \\ \Pi_{23} \end{pmatrix} \tag{32}$$

should hold identically for an arbitrary θ. According to the theory of invariants [3], this implies

$$\begin{pmatrix} \Pi_{13} \\ \Pi_{23} \end{pmatrix} = c\begin{pmatrix} t_x \\ t_y \end{pmatrix}, \tag{33}$$

where c is an arbitrary function of $t_x^2 + t_y^2$ and t_z. Thus, if we define

$$\zeta = \frac{t_z}{f}, \qquad \beta = -\frac{ct_z}{f}, \tag{34}$$

eq. (28) is written as

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\zeta}\Big(\begin{pmatrix} X \\ Y \end{pmatrix} + \beta(t_z - Z)\begin{pmatrix} t_x \\ t_y \end{pmatrix}\Big). \tag{35}$$

The corresponding projection matrix $\boldsymbol{\Pi}$ and the projection vector $\boldsymbol{\pi}$ are

$$\boldsymbol{\Pi} = \begin{pmatrix} 1/\zeta & 0 & -\beta t_x/\zeta \\ 0 & 1/\zeta & -\beta t_y/\zeta \end{pmatrix}, \qquad \boldsymbol{\pi} = \begin{pmatrix} \beta t_x t_z/\zeta \\ \beta t_y t_z/\zeta \end{pmatrix}, \tag{36}$$

where ζ and β are arbitrary *functions* of $t_x^2 + t_y^2$ and t_z. We observe:

– Eq. (35) reduces to the paraperspective projection (eqs. (5)) if we choose

$$\zeta = \frac{t_z}{f}, \qquad \beta = \frac{1}{t_z}. \tag{37}$$

– Eq. (35) reduces to the weak perspective projection (eqs. (4)) if we choose

$$\zeta = \frac{t_z}{f}, \qquad \beta = 0. \tag{38}$$

– Eq. (35) reduces to the orthographic projection (eqs. (3)) if we choose

$$\zeta = 1, \qquad \beta = 0. \tag{39}$$

Thus, eq. (35) includes the traditional affine camera models as special instances and is the *only possible form* that satisfies Requirements 1, 2, and 3.

However, we need not define the functions ζ and β in any particular form; we can regard them as *time varying unknowns* and determine their values by self-calibration. This is made possible by the fact that *at most two* time varying unknowns can be eliminated from the metric constraint of eqs. (25).

5 Procedure for 3-D Reconstruction

3-D Euclidean reconstruction using eq. (35) goes just as for the traditional camera models (see [6] for the details):

1. We fit a 3-D affine space $\mathcal{A}$ to the trajectories $\{\boldsymbol{p}_\alpha\}$ by least squares. Namely, we compute the centroid $\boldsymbol{m}_0$ by eq. (12) and compute the unit eigenvectors $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ of the matrix $\boldsymbol{C}$ in eq. (14) for the largest three eigenvalues[3].
2. We eliminate time varying unknowns from the the metric constraint of eqs. (25) and solve for the metric matrix $\boldsymbol{T}$ by least squares. To be specific, substituting eqs. (36) into eqs. (25), we have

$$(\boldsymbol{u}^\dagger_{\kappa(1)}, \boldsymbol{T}\boldsymbol{u}^\dagger_{\kappa(1)}) = \frac{1}{\zeta_\kappa^2} + \beta_\kappa^2 \tilde{t}_{x\kappa}^2, \qquad (\boldsymbol{u}^\dagger_{\kappa(2)}, \boldsymbol{T}\boldsymbol{u}^\dagger_{\kappa(2)}) = \frac{1}{\zeta_\kappa^2} + \beta_\kappa^2 \tilde{t}_{y\kappa}^2$$

$$(\boldsymbol{u}^\dagger_{\kappa(1)}, \boldsymbol{T}\boldsymbol{u}^\dagger_{\kappa(2)}) = \beta_\kappa^2 \tilde{t}_{x\kappa} \tilde{t}_{y\kappa}, \tag{40}$$

where $\tilde{t}_{x\kappa}$ and $\tilde{t}_{y\kappa}$ are, respectively, the $(2(\kappa-1)+1)$th and the $(2(\kappa-1)+2)$th components of the centroid $\boldsymbol{m}_0$. Eliminating ζ_κ and β_κ, we obtain

$$A_\kappa(\boldsymbol{u}^\dagger_{\kappa(1)}, \boldsymbol{T}\boldsymbol{u}^\dagger_{\kappa(1)}) - C_\kappa(\boldsymbol{u}^\dagger_{\kappa(1)}, \boldsymbol{T}\boldsymbol{u}^\dagger_{\kappa(2)}) - A_\kappa(\boldsymbol{u}^\dagger_{\kappa(2)}, \boldsymbol{T}\boldsymbol{u}^\dagger_{\kappa(2)}) = 0, \tag{41}$$

where $A_\kappa = \tilde{t}_{x\kappa}\tilde{t}_{y\kappa}$ and $C_\kappa = \tilde{t}_{x\kappa}^2 - \tilde{t}_{y\kappa}^2$. This is a linear constraint on $\boldsymbol{T}$, so we can determine $\boldsymbol{T}$ by solving the M equations for $\kappa = 1$, ..., M by least squares. Once we have determined $\boldsymbol{T}$, we can determine ζ_κ and β_κ from eqs. (40) by least squares.

[3] This corresponds to the singular value decomposition $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^\top$ of the measurement matrix $\boldsymbol{W}$ in the traditional formulation [7, 10].

3. We decompose the metric matrix $\boldsymbol{T}$ into the rectifying matrix $\boldsymbol{A}$ in the form of eq. (24), and compute the vectors $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and $\boldsymbol{m}_3$ from eq. (15).
4. We compute the translation $\boldsymbol{t}_\kappa$ and the rotation $\boldsymbol{R}_\kappa$ at each time. The translation components $t_{x\kappa}$ and $t_{y\kappa}$ are given by the first of eqs. (8) in the form of $t_{x\kappa} = \zeta_\kappa \tilde{t}_{x\kappa}$ and $t_{y\kappa} = \zeta_\kappa \tilde{t}_{y\kappa}$. The three rows $\boldsymbol{r}_{\kappa(1)}$, $\boldsymbol{r}_{\kappa(2)}$, and $\boldsymbol{r}_{\kappa(3)}$ of the rotation $\boldsymbol{R}_\kappa$ are given by solving the linear equations

$$\begin{array}{rrrl} \boldsymbol{r}_{\kappa(1)} & & -\beta_\kappa t_{x\kappa}\boldsymbol{r}_{\kappa(3)} & = \zeta_\kappa \boldsymbol{m}^\dagger_{\kappa(1)}, \\ & \boldsymbol{r}_{\kappa(2)} & -\beta_\kappa t_{y\kappa}\boldsymbol{r}_{\kappa(3)} & = \zeta_\kappa \boldsymbol{m}^\dagger_{\kappa(2)}, \\ \beta_\kappa t_{x\kappa}\boldsymbol{r}_{\kappa(1)} & +\beta_\kappa t_{y\kappa}\boldsymbol{r}_{\kappa(2)} & +\boldsymbol{r}_{\kappa(3)} & = \zeta_\kappa^2 \boldsymbol{m}^\dagger_{\kappa(1)} \times \boldsymbol{m}^\dagger_{\kappa(2)}. \end{array} \qquad (42)$$

   The resulting matrix $\begin{pmatrix}\boldsymbol{r}_{\kappa(1)} & \boldsymbol{r}_{\kappa(2)} & \boldsymbol{r}_{\kappa(3)}\end{pmatrix}$ may not be strictly orthogonal, so we compute its singular value decomposition $\boldsymbol{V}_\kappa \boldsymbol{\Lambda}_\kappa \boldsymbol{U}_\kappa^\top$ and let $\boldsymbol{R}_\kappa = \boldsymbol{U}_\kappa \boldsymbol{V}_\kappa^\top$ [4].
5. We recompute the vectors $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and $\boldsymbol{m}_3$ in the form of eqs. (11) using the computed rotations $\boldsymbol{R}_\kappa = \begin{pmatrix}\boldsymbol{i}_\kappa & \boldsymbol{j}_\kappa & \boldsymbol{k}_\kappa\end{pmatrix}$.
6. We compute the object coordinates $(a_\alpha, b_\beta, c_\beta)$ of each point by least-squares expansion of $\boldsymbol{p}'_\alpha$ in the form of eq. (13). The solution is given by $\boldsymbol{M}^- \boldsymbol{p}_\alpha$, using the pseudoinverse $\boldsymbol{M}^-$ of $\boldsymbol{M}$.

However, the following indeterminacy remains:

1. Another solution is obtained by multiplying all $\{\boldsymbol{t}_\kappa\}$ and $\{(a_\alpha, b_\alpha, c_\alpha)\}$ by a common constant.
2. Another solution is obtained by multiplying the all $\{\boldsymbol{R}_\kappa\}$ by a common rotation. The object coordinates $\{(a_\alpha, b_\alpha, c_\alpha)\}$ are rotated accordingly.
3. Each solution has its mirror image solution. The mirror image rotation $\boldsymbol{R}'_\kappa$ is obtained by the rotation $\boldsymbol{R}_\kappa$ followed by a rotation around axis $(\beta_\kappa t_{x\kappa}, \beta_\kappa t_{y\kappa}, 1)$ by angle 2π. At the same time, the object coordinates $\{(a_\alpha, b_\alpha, c_\alpha)\}$ change their signs.
4. *The absolute depth t_z of the world coordinate origin is indeterminate.*

Item 1 is the fundamental ambiguity of 3-D reconstruction from images, meaning that a large motion of a large object in the distance is indistinguishable from a small motion of a small object nearby. Item 2 reflects the fact that the orientation of the world coordinate system can be arbitrarily chosen. Item 3 is due to eq. (24), which can be written as $\boldsymbol{T} = (\pm \boldsymbol{AQ})(\pm \boldsymbol{AQ})^\top$ for an arbitrary rotation $\boldsymbol{Q}$. This ambiguity is inherent of all affine cameras [8, 9].

Item 4 is due to the fact that eq. (35) involves only the *relative depth* of individual point from the world coordinate origin $\boldsymbol{t}_\kappa$. The absolute depth t_z is determined only if ζ and β are given as *specific functions of* t_z, as in the case of the traditional camera models. Here, however, we do not specify their functional forms, directly determining their values by self-calibration and leaving t_z unspecified.

6 Experiments

Fig. 3 shows four simulated image sequences of 600 × 600 pixels perspectively projected with focal length $f = 600$ pixels. Each consists of 11 frames; six decimated frames are shown here. We added Gaussian random noise of mean 0 and standard deviation 1 pixel independently to the x and y coordinates of the feature points and reconstructed their 3-D shape (the frames in Fig. 3(a), (b) are merely for visual ease).

From the resulting two mirror image shapes, we choose the correct one by comparing the depths of two points that are known be close to and away from the camera. Since the absolute depth and scale are indeterminate, we translate the true and the reconstructed shapes so that their centroids are at the coordinate origin and scaled their sizes so that the root-mean-square distance of the feature points from the origin is 1. Then, we rotate the reconstructed shape so that root-mean-square distances between the corresponding points of the two shapes is minimized. We adopt the resulting residual as the measure of reconstruction accuracy.

We compare three camera models: the weak perspective, the paraperspective, and our symmetric affine camera models. The orthographic model is omitted, since evidently good results cannot be obtained when the object moves in the depth direction. For the weak perspective and paraperspective models, we need to specify the focal length f (see eqs. (4) and (5)). If the size of the reconstructed shape is normalized as described earlier, the choice of f is irrelevant for the weak perspective model, because it only affects the object size as a whole. However, the paraperspective model depends on the value of f we use.

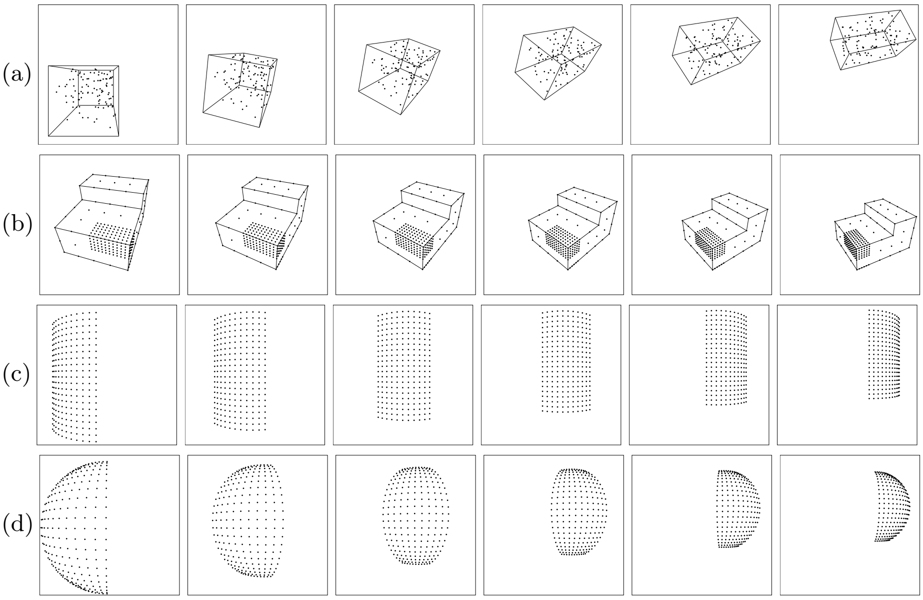
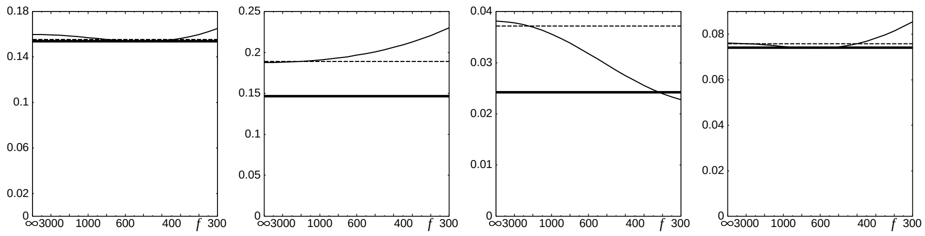


Fig. 3. Simulated image sequences (six decimated frames for each).



(a) (b) (c) (d)

Fig. 4. 3-D reconstruction accuracy for the image sequences of Fig. 3(a)∼(d). The horizontal axis is scaled in proportion to $1/f$. Three models are compared: The dashed line: weak perspective (dashed lines), paraperspective (thin solid lines), and our generic model (thick solid lines).

Fig. 4 plots the reconstruction accuracy vs. the input focal length f; the horizontal axis is scaled in proportion to $1/f$. The dashed line is for weak perspective, the thin solid line is for paraperspective, and the thick solid line is for our model. We observe that the paraperspective model does not necessarily give the highest accuracy when f coincides with the focal length (600 pixels) of the perspective images. The error is indeed minimum around $f = 600$ for Fig. 4(a), (d), but the error decreases as f increases for Fig. 4(b) and as f decreases for Fig. 4(c).

We conclude that our model achieves the accuracy comparable to paraperspective projection given an appropriate value of f, which is unknown in advance. This means that our model automatically chooses appropriate parameter values without any knowledge about f.

We conducted many other experiments (not shown here) and observed similar results. We have found that *degeneracy* can occur in special circumstances; the matrix $\boldsymbol{A}$ becomes rank deficient so that the resulting vectors $\{\boldsymbol{m}_i\}$ are linearly dependent (see eq. (15)). As a result, the reconstructed shape is “flat” (see eq. (13)). This occurs when the smallest eigenvalue of $\boldsymbol{T}$ computed by least squares is negative, while eq. (24) requires $\boldsymbol{T}$ to be positive semidefinite. In the computation, we replace the negative eigenvalue by zero, resulting in degeneracy.

This type of degeneracy occurs for the traditional camera models, too. In principle, we could avoid it by parameterizing $\boldsymbol{T}$ so that it is guaranteed to be positive definite [8]. However, this would require nonlinear optimization, and the merit of the factorization approach (i.e., linear computation only) would be lost. Moreover, if we look at the images that cause degeneracy, they really look as if a planar object is moving. Since the information is insufficient in the first place, any methods may not be able to solve such degeneracy.

7 Conclusions

We showed that minimal requirements for an affine camera to mimic perspective projection leads to a unique camera model, which we call “symmetric affine camera”, having two free functions, whose specific choices would result in the traditional camera models. We regarded them as time varying parameters and determined their values by self-calibration, using linear computation alone, so that an appropriate model is automatically selected. We have demonstrated by simulation that the reconstruction accuracy is comparable to the paraperspective model given an appropriate focal length estimate.

Acknowledgments: The authors thank Seitaro Asahara of Recruit Staffing, Co., Ltd., Japan, for participating in this project. This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under a Grant in Aid for Scientific Research C (No. 17500112).

References

1. K. Deguchi, T. Sasano, H. Arai, and H. Yoshikawa, 3-D shape reconstruction from endoscope image sequences by the factorization method, *IEICE Trans. Inf. & Syst.*, *E79-D*-9 (1996-9), 1329–1336.
2. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, U.K., 2000.
3. K. Kanatani, *Group-Theoretical Methods in Image Understanding*, Springer-Verlag, Berlin, Germany, 1990.
4. K. Kanatani, *Geometric Computation for Machine Vision*, Oxford University Press, Oxford, U.K., 1993.
5. K. Kanatani and Y. Sugaya, Factorization without factorization: complete recipe, *Mem. Fac. Eng. Okayama Univ.*, **38**-1/2 (2004-3), pp. 61-72.
6. K. Kanatani, Y. Sugaya and H. Ackermann, Uncalibrated factorization using a variable symmetric affine camera, *Mem. Fac. Eng. Okayama Univ.*, **40** (2006-1), pp. 53–63.
7. C. J. Poelman and T. Kanade, A paraperspective factorization method for shape and motion recovery, *IEEE Trans. Patt. Anal. Mach. Intell.*, *19*-3 (1997-3), 206–218.
8. L. Quan, Self-calibration of an affine camera from multiple views, *Int. J. Comput. Vision*, **19**-1 (1996-7), 93–105.
9. L. S. Shapiro, A. Zisserman, and M. Brady, 3D motion recovery via affine epipolar geometry, *Int. J. Comput. Vision*, **16**-2 (1995-10), 147–182.
10. C. Tomasi and T. Kanade, Shape and motion from image streams under orthography—A factorization method, *Int. J. Comput. Vision*, *9*-2 (1992-10), 137–154.