International Journal of Computer Vision, **19**(2): 169–180, 1996.

Paraperspective ≡ Affine

Ronen Basri
Dept. of Applied Math
The Weizmann Institute of Science
Rehovot 76100, Israel
ronen@wisdom.weizmann.ac.il

Abstract

It is shown that the set of all paraperspective images with arbitrary reference point and the set of all affine images of a 3-D object are identical. Consequently, all uncalibrated paraperspective images of an object can be constructed from a 3-D model of the object by applying an affine transformation to the model, and every affine image of the object represents some uncalibrated paraperspective image of the object. It follows that the paraperspective images of an object can be expressed as linear combinations of any two non-degenerate images of the object. When the image position of the reference point is given the parameters of the affine transformation (and, likewise, the coefficients of the linear combinations) satisfy two quadratic constraints. Conversely, when the values of parameters are given the image position of the reference point is determined by solving a bi-quadratic equation.

Key words: affine transformations, calibration, linear combinations, paraperspective projection, 3-D object recognition.

1 Introduction

It is shown below that given an object $O \subset \mathcal{R}^3$, the set of all images of O obtained by applying a rigid transformation followed by a paraperspective projection with arbitrary reference point and the set of all images of O obtained by applying a 3-D affine transformation followed by an orthographic projection are identical. Consequently, all paraperspective images of an object can be constructed from a 3-D model of the object by applying an affine transformation to the model, and every affine image of the object represents some paraperspective image of the object. This implies in particular that all paraperspective images of O can be expressed as linear combinations of two images I_1 and I_2 for every non-degenerate pair of images I_1 and I_2. This result improves over a previous result by Sugimoto [S95], who showed that the paraperspective images of objects can be represented by linear combinations of three images. The related problem of recovering the structure of objects from a set of their paraperspective images was discussed by Poelman and Kanade [PK94].

Allowing the reference point to translate to any arbitrary position corresponds to the situation in which the focal length of the camera and the center of the image are unknown. This is of particular value when we attempt to represent the set of images of an object obtained from an uncalibrated camera. For calibrated images it is often assumed that the reference point depends on the 3-D position of the object's points. In particular, the reference point may be set at the centroid of the observed object. Clearly, setting the reference point at any specific location can only reduce the set of images of the object, and so the set of paraperspective images of an object with the reference point set at some predetermined location forms a subset to the set of the affine images of the object. Consequently, whenever the image position of the reference point is given the image may still be obtained by applying an affine transformation to the model, but now the parameters of this transformation satisfy two quadratic constraints. Conversely, when the parameters of the affine transformation are given the image position of the reference point can be derived from the transformation by solving a bi-quadratic equation.

The rest of this paper is divided as follows. The definitions of the paraperspective projection and the sets of paraperspective and affine images of objects are given in Section 2. The main result of this paper, namely, that the sets of paraperspective and affine images of objects are identical, is derived in Section 3. Section 4 discusses the relation between the position of the reference point and the calibration parameters, and Section 5 shows how these results imply that the paraperspective images of an object can be expressed by linear combinations of two non-degenerate images of the object.

2 Definitions

This section reviews the definition of the paraperspective projection and defines the set $\mathcal{P}$ of all paraperspective images and the set $\mathcal{A}$ of all affine images of an object.

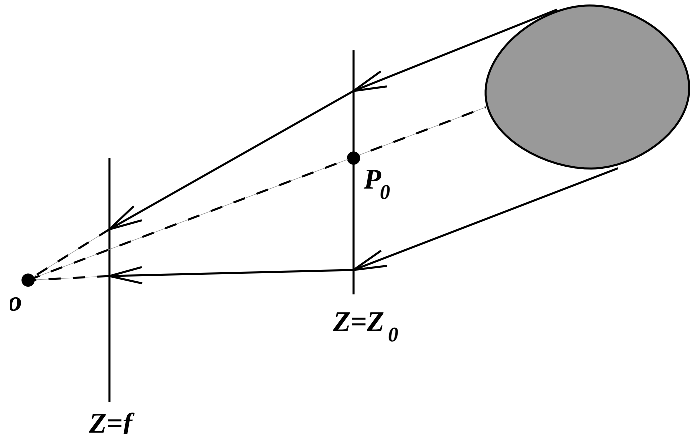

Figure 1: The paraperspective projection (a side view). The reference point is set at P_0. The object is first projected to the reference plane $Z = Z_0$ in a direction parallel to the line connecting P_0 with the focal center (denoted by o), and then to the image plane ($Z = f$) by a central projection.

2.1 Paraperspective projection

The paraperspective projection, introduced by Ohta *et al.* [OMS81] (see Aloimonos [A90] for a discussion of different approximations to perspective projection), is defined as follows (see Figure 1). Given a reference point $\vec{P}_0 = (X_0, Y_0, Z_0)^T$, an object point $\vec{P} = (X, Y, Z)^T$ is first projected in the direction parallel to $\vec{P}_0$ onto the reference plane $Z = Z_0$ (which is the plane through $\vec{P}_0$ parallel to the image plane). From the reference plane the object is then projected onto the image plane by a central projection. Since the reference plane and the image plane are parallel this last projection is equivalent to applying a uniform scaling to the object. Formally, under paraperspective projection $\vec{P}$ is projected to an image point $\vec{p} = (x, y)^T$, with coordinates given by

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{f}{Z_0} \left[\begin{pmatrix} 1 & 0 & -\frac{X_0}{Z_0} \\ 0 & 1 & -\frac{Y_0}{Z_0} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} \right]. \tag{1}$$

Sugimoto and Murota [SM93] and Poelman and Kanade [PK94] showed that paraperspective projection is the first order approximation of perspective projection when the object is relatively distant from the camera (so that $X, Y \ll Z$).

Below, we shall use the following notation:

$$s = \frac{f}{Z_0}, \qquad u = -\frac{X_0}{Z_0}, \qquad v = -\frac{Y_0}{Z_0}. \tag{2}$$

(u, v) denotes the image position (in focal length units) of the reference point, and s represents the amount of scaling applied to the object following the projection from the plane $Z = Z_0$ to the image plane. Since the paraperspective projection involves four parameters, the three coordinates of the reference point P_0 and the focal length f we choose to keep the parameter

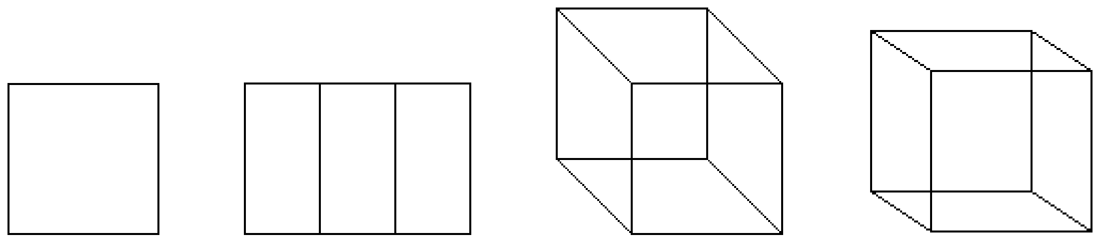

Figure 2: Paraperspective views of a frontal-parallel cube under different translations. The reference point is set at the centroid of the cube. The image position of the reference point (u, v) is, from left to right: (0,0) (this is also the orthographic view of the cube), (0.5,0), (0.5,0.5) and (0.375,0.25).

Z_0, the distance between the reference plane and the focal center. Hence, Eq. 1 becomes

$$\begin{pmatrix} x \\ y \end{pmatrix} = s \left[\begin{pmatrix} 1 & 0 & u \\ 0 & 1 & v \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} - Z_0 \begin{pmatrix} u \\ v \end{pmatrix} \right]. \tag{3}$$

The weak-perspective (= scaled-orthographic) projection is a special case of the paraperspective projection obtained by setting the reference point along the optical axis (so that $(u, v) = \vec{0}$). The paraperspective projection extends the scaled-orthographic projection by considering the perspective deformation an object undergoes when it appears in the periphery of the image. Examples for paraperspective views of a frontal-parallel cube are shown in Fig. 2.

2.2 Paraperspective and affine images

Given an object $O \subset \mathcal{R}^3$, we consider two sets of images of O, the set of paraperspective images and the set of affine images of O. *The set of paraperspective images of* O, denoted by $\mathcal{P}$, contains the images of O obtained by applying a rigid transformation to O followed by a paraperspective projection with the reference point set in any position. Suppose O is transformed by a rotation $R \in SO(3)$ and translation $\vec{t} \in \mathcal{R}^3$ followed by a paraperspective projection Π_p with the reference point at P_0. Then, the projected coordinates of a point $\vec{P}_i \in O$ are given by

$$\vec{q}_i = \Pi_p(R\vec{P}_i + \vec{t}). \tag{4}$$

Following Equations 3 and 4 $\vec{q}_i = (x_i, y_i)^T$ is given by

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = s \left[\begin{pmatrix} 1 & 0 & u \\ 0 & 1 & v \end{pmatrix} \left(\begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \right) - Z_0 \begin{pmatrix} u \\ v \end{pmatrix} \right], \tag{5}$$

where r_{ij} $(1 \leq i, j \leq 3)$ are the components of the rotation matrix, R, and t_x, t_y, and t_z are the components of the translation vector, $\vec{t}$. Eq. 5 implies that

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = s \left[\begin{pmatrix} r_{11} + ur_{31} & r_{12} + ur_{32} & r_{13} + ur_{33} \\ r_{21} + vr_{31} & r_{22} + vr_{32} & r_{23} + vr_{33} \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} + \begin{pmatrix} t_x + u(t_z - Z_0) \\ t_y + v(t_z - Z_0) \end{pmatrix} \right], \tag{6}$$

and this can be presented more concisely as

$$\begin{aligned} x_i &= s[(\vec{r}_1 + u\vec{r}_3)^T \vec{P}_i + t_x + u(t_z - Z_0)] \\ y_i &= s[(\vec{r}_2 + v\vec{r}_3)^T \vec{P}_i + t_y + v(t_z - Z_0)], \end{aligned} \tag{7}$$

where $\vec{r}_1$, $\vec{r}_2$, and $\vec{r}_3$ denote the three row vectors of R.

The set of affine images of O, denoted by $\mathcal{A}$, contains the images of O obtained by applying a 3-D affine transformation to O followed by an orthographic projection. We shall restrict $\mathcal{A}$ to include only images that are obtained by a non-degenerate affine transformations (that is, an affine transformation with a non-singular linear part). Suppose O is transformed by a 3×3 non-singular linear matrix $A \in GL(3)$ and translation $\vec{t}' \in \mathcal{R}^3$ followed by an orthographic projection Π_o. Then, the projected coordinates of a point $P_i \in O$ are given by

$$\vec{q}_i = \Pi_o(A\vec{P}_i + \vec{t}'). \tag{8}$$

This can be written in the form

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} + \begin{pmatrix} t'_x \\ t'_y \end{pmatrix}, \tag{9}$$

which in vector form becomes

$$\begin{aligned} x_i &= \vec{a}^T \vec{P}_i + t'_x \\ y_i &= \vec{b}^T \vec{P}_i + t'_y, \end{aligned} \tag{10}$$

where $\vec{a} = (a_1, a_2, a_3)^T$ and $\vec{b} = (b_1, b_2, b_3)^T$ are the two top row vectors of A, and t'_x and t'_y are the first two components of the translation vector, $\vec{t}'$. Since A was required to be non-singular then the vectors $\vec{a}$ and $\vec{b}$ must be linearly independent. Section 3 below establishes that the two sets of images, $\mathcal{P}$ and $\mathcal{A}$, are identical.

3 Paraperspective $\equiv$ affine

This section shows that the set of all images of an object obtained by applying a rigid transformation followed by a paraperspective projection with arbitrary reference point and the set of all images of the object obtained by applying a 3-D affine transformation followed by an orthographic projection are identical. It will be shown later (in Section 4) that placing the reference point in arbitrary position is equivalent to considering the paraperspective approximations of the uncalibrated images of the object. We begin by showing that the set of paraperspective images of an object are contained within the set of affine images of the object.

Theorem 1: $\mathcal{P} \subseteq \mathcal{A}$.

Proof: Given a paraperspective image $I \in \mathcal{P}$ of O, we will show that $I \in \mathcal{A}$. Suppose I is obtained from O by applying a rotation R and translation $\vec{t}$ followed by a paraperspective projection Π_p with the reference point at P_0. Given a point $P_i \in O$ its image position $\vec{q}_i$ is given by Eq. 7:

$$\begin{aligned} x_i &= s[(\vec{r}_1 + u\vec{r}_3)^T \vec{P}_i + t_x + u(t_z - Z_0)] \\ y_i &= s[(\vec{r}_2 + v\vec{r}_3)^T \vec{P}_i + t_y + v(t_z - Z_0)]. \end{aligned} \tag{11}$$

We can represent $\vec{q}_i$ as in Eq. 8 by

$$\begin{aligned} x_i &= \vec{a}^T \vec{P}_i + t'_x \\ y_i &= \vec{b}^T \vec{P}_i + t'_y \end{aligned} \tag{12}$$

by setting

$$\begin{aligned} \vec{a} &= s(\vec{r}_1 + u\vec{r}_3) \\ \vec{b} &= s(\vec{r}_2 + v\vec{r}_3) \\ t'_x &= s(t_x + u(t_z - Z_0)) \\ t'_y &= s(t_y + v(t_z - Z_0)). \end{aligned} \tag{13}$$

Intuitively, the production of a paraperspective image can be expressed as a sequence of three transformations which are included in the affine group: a rigid transformation followed by a parallel projection and then by a uniform scaling. The result of these three transformations obviously is an affine image of the object. □

Theorem 1 demonstrates that any paraperspective image of an object can be constructed by applying a 3-D affine transformation to the object followed by an orthographic projection. Notice that the parameters of this affine transformation depend only on the rotation and translation parameters and on the location of the reference point, and are independent of the coordinates of the considered point, $\vec{P}$. The same affine transformation, therefore, is applied to all object points. Applying a different rotation or translation to the object, or changing the reference point, would change the corresponding affine transformation. Thus, we have shown that the set of paraperspective images of an object is contained in the set of affine images of the object.

The converse, namely, that every affine image of an object can be interpreted as some paraperspective image of the object, also holds. This is established in Theorem 2 below. One way to show this is by noticing that by allowing the reference point to appear in arbitrary positions we can obtain any parallel projection. It then remains to be shown that every affine image can be produced by a sequence of a rigid transformation followed by a parallel projection and uniform scaling. Below we present a different, algebraic proof. Although somewhat lengthy, this proof also provides the parameters of the rigid transformation (which determine the position and orientation of the object) and the location of the reference point in closed-form.

Theorem 2: $\mathcal{A} \subseteq \mathcal{P}$.

Proof: Given an affine image $I \in \mathcal{A}$ of O, we will show that $I \in \mathcal{P}$. Suppose that I is obtained from O by applying a non-singular linear transformation A and translation $\vec{t}'$ followed by an orthographic projection Π_o. Given a point $P_i \in O$ its image position $\vec{q}_i$ is given by Eq. 10. We will show that $I \in \mathcal{P}$ by proving that there exists a paraperspective view (determined by a rotation, translation, and the location of the reference point) in which O appears as in I.

To find the corresponding paraperspective parameters, we shall look at Eq. 13 as a system of equations, where the parameters of the affine transformation are known, and the parameters of the paraperspective view are the unknowns. We will show that this system of equations has a solution for any non-degenerate affine transformation.

Consider first the translation parameters

$$\begin{array}{rcl} t'_x & = & s(t_x + u(t_z - Z_0)) \\ t'_y & = & s(t_y + v(t_z - Z_0)). \end{array} \tag{14}$$

These two equations are linear in the three translation parameters of the paraperspective view, t_x, t_y, and t_z. Obviously, this is an under-constrained system of equations, and so it has infinite different solutions. We pick the solution

$$t_z = Z_0, \tag{15}$$

which implies that

$$\begin{array}{rcl} t'_x & = & st_x \\ t'_y & = & st_y. \end{array} \tag{16}$$

Since the affine transformation is non-degenerate s cannot vanish, and so t_x and t_y are given by

$$\begin{array}{rcl} t_x & = & \frac{1}{s} t'_x \\ t_y & = & \frac{1}{s} t'_y, \end{array} \tag{17}$$

where the value of s still needs to be recovered.

We now turn to recovering the rest of the parameters of the paraperspective view. The rest of Eq. 13 contains the equations

$$\begin{array}{rcl} \vec{a} & = & s(\vec{r}_1 + u\vec{r}_3) \\ \vec{b} & = & s(\vec{r}_2 + v\vec{r}_3), \end{array} \tag{18}$$

where s, u, v are the unknown parameters of the reference point and $\vec{r}_1$, $\vec{r}_2$, $\vec{r}_3$ are the three unknown row vectors of the rotation matrix R. We first attempt to recover the values of the parameters s, u, and v, which determine the location of the reference point. Since R is unitary ($RR^T = I$), we have that

$$\begin{array}{rcl} \vec{a}^T\vec{a} & = & s^2(1+u^2) \\ \vec{b}^T\vec{b} & = & s^2(1+v^2) \\ \vec{a}^T\vec{b} & = & s^2 uv. \end{array} \tag{19}$$

In the Appendix for this paper we show that for any pair of non-degenerate vectors $\vec{a}$ and $\vec{b}$ a solution for Eq. 19 exists and is unique (up to a reflection). The proof involves turning Eq.

19 to a bi-quadratic equation in u and eliminating those solutions that are not geometrically realizable. Using the solution for u the values of v and s can also be recovered. The complete derivation is given in the Appendix.

Finally, given the values of s, u, and v we can recover the rotation parameters. Consider the three unit vectors $\vec{i}$, $\vec{j}$, and $\vec{k}$ in the directions of the main axes. Define

$$\begin{aligned} \vec{a}' &= s(\vec{i}+u\vec{k}) \\ \vec{b}' &= s(\vec{j}+v\vec{k}). \end{aligned} \tag{20}$$

The triangle $\vec{o}$ (the origin), $\vec{a}'$, and $\vec{b}'$ and the triangle $\vec{o}$, $\vec{a}$, and $\vec{b}$ are congruent. This can be noticed by the fact that

$$\begin{aligned} \vec{a}'^T\vec{a}' &= s^2(1+u^2) = \vec{a}^T\vec{a} \\ \vec{b}'^T\vec{b}' &= s^2(1+v^2) = \vec{b}^T\vec{b} \\ \vec{a}'^T\vec{b}' &= s^2uv = \vec{a}^T\vec{b}. \end{aligned} \tag{21}$$

Since the two triangles are congruent, and since they both extend from the origin, there exists a rotation matrix R such that

$$\begin{aligned} \vec{a}' &= R\vec{a} \\ \vec{b}' &= R\vec{b}. \end{aligned} \tag{22}$$

This matrix, R, is the desired rotation matrix, since

$$\begin{aligned} \vec{a} = R^T\vec{a}' = R^T s(\vec{i}+u\vec{k}) = s(\vec{r}_1+u\vec{r}_3) \\ \vec{b} = R^T\vec{b}' = R^T s(\vec{j}+v\vec{k}) = s(\vec{r}_2+v\vec{r}_3). \end{aligned} \tag{23}$$

□

To illustrate how the paraperspective parameters can be derived from the affine parameters consider the following example. Suppose an object is stretched horizontally by some factor $\alpha > 1$. In this case $\vec{a} = (\alpha, 0, 0)$ and $\vec{b} = (0, 1, 0)$. From the appendix we see (Eq. 72) that this implies that $s = 1$, $u = \pm\sqrt{\alpha^2-1}$, and $v = 0$. Solving for the rotation matrix (Equations 22-23) we find that

$$R = \begin{pmatrix} \frac{1}{\alpha} & 0 & \pm\frac{\sqrt{\alpha^2-1}}{\alpha} \\ 0 & 1 & 0 \\ \mp\frac{\sqrt{\alpha^2-1}}{\alpha} & 0 & \frac{1}{\alpha} \end{pmatrix}, \tag{24}$$

which corresponds to a rotation about the vertical axis. If for example $\alpha = 2$ then the angle of rotation is $\pm 60°$. Thus, stretching the object horizontally is obtained by rotating the object by a rotation R about the Y-axis and setting the reference point at $(u, v) = (\pm\sqrt{\alpha^2-1}, 0)$. Correspondences between other affine transformations and paraperspective views can be obtained in a similar way.

The result that paraperspective is equivalent to affine transformation implies in particular that under paraperspective projection an object can be modeled by linear combinations of a small number of images. Ullman and Basri [UB91] (see also [B93] and [P90]) showed that the affine views of an object can in general be expressed as linear combinations of two images.

This implies that the paraperspective images of an object can also be expressed by linear combinations of two images. This issue is discussed in detail in Section 5.

The equivalence of paraperspective and affine holds also when we extend the paraperspective set of images of an object to include all the images obtained by applying an affine (rather than rigid) transformation followed by a paraperspective projection. Denote by $\mathcal{P}_A$ the set of images of an object O obtained by applying an affine transformation to O followed by a paraperspective projection with arbitrary reference point. Clearly, $\mathcal{P}_A \supseteq \mathcal{P}$ (since any rigid transformation is in particular an affine transformation). On the other hand, since Theorem 1 does not require R to be a rotation matrix then $\mathcal{P}_A \subseteq \mathcal{A}$. This, together with the equality $\mathcal{P} = \mathcal{A}$, implies that $\mathcal{P}_A = \mathcal{A}$.

This last equivalence implies that the set of images of an object obtained by applying an affine transformation to the object followed by a paraperspective projection and the set of images of that object obtained by applying an affine transformation followed by an orthographic projection are identical. It follows also that the result of taking a series of paraperspective images (that is, an image of an image) is also an affine image of the object. In fact, a paraperspective image of a plane is equivalent to a 2-D affine transformation. This property generalizes the result shown by Jacobs [J94] that a series of orthographic images of a rigid object is an affine image of the object.

4 Calibration and the location of the reference point

The previous section showed that the set of paraperspective images with arbitrary reference point and the set of affine images of an object are identical. In most applications of the paraperspective projection the reference point is set at the centroid of the object. This section derives the constraints determined by setting the reference point at any specific location. It shows further that when a paraperspective projection is used to approximate the appearance of an object obtained using an uncalibrated camera the reference point can in fact appear in arbitrary position.

Assume without the loss of generality that the model centroid is set at the origin of the model (denoted by $\vec{o}$). Then, the observed position of the reference point is

$$\vec{P}_0 = R\vec{o} + \vec{t} = \vec{t}. \tag{25}$$

Eq. 2 now becomes

$$s = \frac{f}{t_z}, \qquad u = -\frac{t_x}{t_z}, \qquad v = -\frac{t_y}{t_z}, \tag{26}$$

implying that

$$t_x = -\frac{fu}{s}, \qquad t_y = -\frac{fv}{s}, \qquad t_z = \frac{f}{s}. \tag{27}$$

Combining this with Eq. 7, the position of a point $\vec{q_i} = (x_i, y_i)^T$ is given by

$$\begin{aligned} x_i &= s(\vec{r}_1 + u\vec{r}_3)^T \vec{P}_i - fu \\ y_i &= s(\vec{r}_2 + v\vec{r}_3)^T \vec{P}_i - fv. \end{aligned} \tag{28}$$

As has been shown in Theorem 1, Eq. 28 can be represented as an affine transformation with eight parameters

$$\begin{aligned} x_i = \vec{a}^T \vec{P}_i + t_x \\ y_i = \vec{b}^T \vec{P}_i + t_y. \end{aligned} \tag{29}$$

These parameters are related to the parameters of Eq. 28 by

$$\begin{aligned} \vec{a} = s(\vec{r}_1 + u\vec{r}_3) \\ \vec{b} = s(\vec{r}_2 + v\vec{r}_3) \\ t_x = -fu \\ t_y = -fv. \end{aligned} \tag{30}$$

From the proof for Theorem 2 we know that all the parameters s, u, v, $\vec{r}_1$, $\vec{r}_2$, and $\vec{r}_3$ are determined uniquely (up to a reflection) from $\vec{a}$ and $\vec{b}$. At the same time, the translation parameters determine the projected location of the reference point,

$$\begin{aligned} u &= -\tfrac{t_x}{f} \\ v &= -\tfrac{t_y}{f}. \end{aligned} \tag{31}$$

Consequently, the parameters of the affine transformation must satisfy two quadratic constraints, which are derived from Eq. 19:

$$\frac{\vec{a}^T\vec{a}}{1+u^2} = \frac{\vec{b}^T\vec{b}}{1+v^2} = \frac{\vec{a}^T\vec{b}}{uv}. \tag{32}$$

(Notice that these two constraints are identical to the metric constraints derived by Poelman and Kanade [PK94].) In term of the affine parameters we obtain the following two constraints (by plugging the values of u and v from Eq. 31 into Eq. 32)

$$\frac{\vec{a}^T\vec{a}}{f^2+t_x^2} = \frac{\vec{b}^T\vec{b}}{f^2+t_y^2} = \frac{\vec{a}^T\vec{b}}{t_x t_y}. \tag{33}$$

For an image obtained by a scaled-orthographic projection $u = v = 0$, and the constraints become

$$\begin{aligned} \vec{a}^T\vec{a} = \vec{b}^T\vec{b} \\ \vec{a}^T\vec{b} = 0 \end{aligned} \tag{34}$$

(see, e.g., [UB91, B93]).

In vision applications it is sometimes desired to recognize objects from their uncalibrated images. In such images the location of the center of the image (the intersection of the optical axis with the image plane), the focal length, and sometimes also the inter-pixel distances may be unknown. Varying these calibration parameters is equivalent under the paraperspective approximation to applying an arbitrary affine transformation to the image. Consequently, the displacement of the centroid in the image is no longer sufficient to determine the projected position of the reference point, (u, v), and the quadratic constraints (Eq. 32) cannot be resolved.

Suppose first that only the focal length is unknown. Then, the translation parameters (Eq. 31) determine the projected location of the reference point only up to a scale factor, so that

$$\frac{u}{v} = \frac{t_x}{t_y}. \tag{35}$$

Thus, the system now has seven degrees of freedom, and the affine parameters satisfy only one constraint. This constraint is obtained by eliminating f^2 from Eq. 33

$$t_x t_y(\vec{a}^T\vec{a} - \vec{b}^T\vec{b}) = (t_x^2 - t_y^2)\ \vec{a}^T\vec{b}. \tag{36}$$

Suppose now that the camera center is unknown. Denote the unknown image location of the camera center by (x_c, y_c), then Eq. 28 becomes

$$\begin{aligned} x_i &= s(\vec{r}_1 + u\vec{r}_3)^T\vec{P}_i - fu + x_c \\ y_i &= s(\vec{r}_2 + v\vec{r}_3)^T\vec{P}_i - fv + y_c. \end{aligned} \tag{37}$$

In this case u and v are not determined by the translation parameters because

$$\begin{aligned} t_x &= -fu + x_c \\ t_y &= -fv + y_c. \end{aligned} \tag{38}$$

Consequently, the system now has eight degrees of freedom and every affine image of the object represents a possible paraperspective image of the object. If in addition to the camera center also the focal length or the inter-pixel distances are unknown they cannot increase the degrees of freedom of the system further since, as has been shown in Theorem 1, all paraperspective images of an object are included in its affine set of images.

To see the effect of varying the calibration parameters on the location of the reference point consider the following examples. Changing the focal length (which results in a scaling of the image) is equivalent to sliding the reference point along the ray connecting it with the focal center along with a translation of the object in depth. Suppose that we set the focal length to αf for some $\alpha > 0$. By moving the reference point to $\alpha P_0 = (\alpha X_0, \alpha Y_0, \alpha Z_0)$ we keep s, u, and v from Eq. 2 unchanged. For the new focal length and reference point Eq. 7 holds with the exception that Z_0 is replaced by αZ_0, namely

$$\begin{aligned} x_i &= s[(\vec{r}_1 + u\vec{r}_3)^T\vec{P}_i + t_x + u(t_z - \alpha Z_0)] \\ y_i &= s[(\vec{r}_2 + v\vec{r}_3)^T\vec{P}_i + t_y + v(t_z - \alpha Z_0)]. \end{aligned} \tag{39}$$

By setting $t_z = \alpha Z_0$ (instead of $t_z = Z_0$, as we did in the proof of Theorem 2) the original image is obtained. Hence, changing the focal length has the effect of sliding the reference point along the ray connecting it with the focal center followed by a translation in depth.

Similarly, changing the location of the camera center (which results in a translation of the image) is equivalent to translating the object with respect to the reference point in a direction parallel to the image plane. Suppose the camera center is located at (x_c, y_c), then all perceived image coordinates translate by (x_c, y_c). Eq. 7 now becomes

$$\begin{aligned} x_i &= s[(\vec{r}_1 + u\vec{r}_3)^T \vec{P}_i + t_x + u(t_z - Z_0)] + x_c \\ y_i &= s[(\vec{r}_2 + v\vec{r}_3)^T \vec{P}_i + t_y + v(t_z - Z_0)] + y_c. \end{aligned} \tag{40}$$

Let

$$\begin{aligned} t'_x &= t_x + \tfrac{x_c}{s} \\ t'_y &= t_y + \tfrac{y_c}{s}. \end{aligned} \tag{41}$$

Then Eq. 40 can be rewritten as

$$\begin{aligned} x_i &= s[(\vec{r}_1 + u\vec{r}_3)^T \vec{P}_i + t'_x + u(t_z - Z_0)] \\ y_i &= s[(\vec{r}_2 + v\vec{r}_3)^T \vec{P}_i + t'_y + v(t_z - Z_0)]. \end{aligned} \tag{42}$$

The effect of moving the camera center therefore is captured by translating the object in a plane parallel to the image plane while keeping the reference point in a fixed position (relative to the center of the camera). Alternatively, this can be interpreted as a translation of the reference point (and the camera center) relative to the object.

These two examples demonstrate that setting the location of the camera center in arbitrary position (which results in a translation of the image) is equivalent to translating the reference point within the reference plane, and that changing the focal length (which results in a scaling of the image) is equivalent to sliding the reference point away or toward the focal center. Consequently, the set of paraperspective images of an object with arbitrary reference point is equivalent to the set of arbitrarily translated paraperspective images of the object with the centroid as the reference point. Thus, the set of paraperspective images of an object with an arbitrary reference point is equivalent to the set of uncalibrated paraperspective images of the object. Figure 3 shows the effect of shifting the reference point in a direction parallel to the image plane. The camera center in these example is kept fixed. It can be seen that such a shift results in new images of the object which involve stretching of the object.

As a consequence, attempting to recognize objects from uncalibrated images under the assumption of a paraperspective projection is equivalent to recognizing the objects under an affine transformation followed by an orthographic projection. Requiring the reference point to appear in a particular position (e.g., the centroid of the object) implies recognition from calibrated images. A similar distinction can be drawn in shape reconstruction from motion sequences. Shape reconstruction from uncalibrated sequences (when each of the images in a sequence may be obtained under a different calibration) is equivalent to reconstruction assuming the object may undergo general affine transformation in space. Examples for algorithms for affine reconstruction under this assumption were provided by Koenderink and van Doorn [KV91] and Weinshall and Tomasi [WT92]. Reconstruction from calibrated sequences requires the enforcement of the rigidity constraints (Eq. 32), as in Poelman and Kanade's algorithm [PK94]. This distinction is

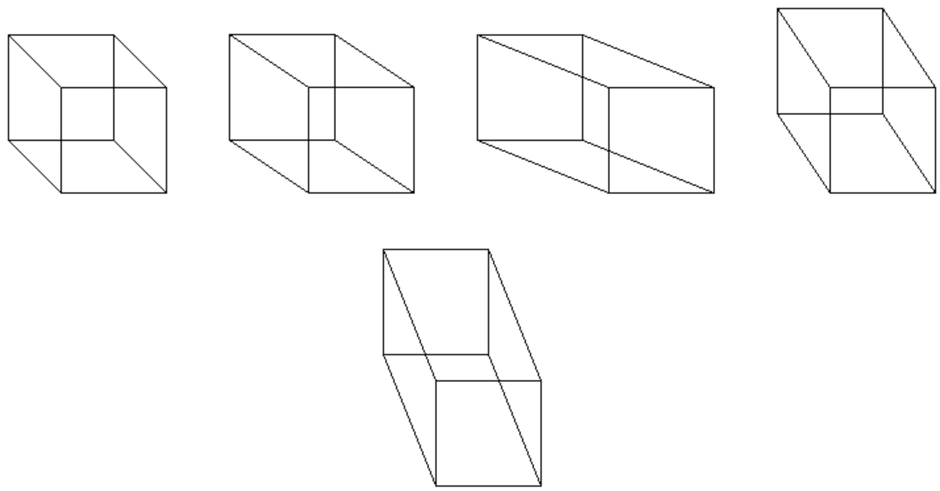

Figure 3: Shifting the location of the reference point in a direction parallel to the image plane results in transformations that involve stretching of the object. From left to right, the original cube from Fig. 2 viewed with (u,v)=(0.5,0.5), and the same cube viewed with (u,v)=(0.75,0.5), (1.25,0.5), (0.5,0.75), and (0.5,1.25).

analogous to the distinction between projective reconstruction from uncalibrated images [F92] and perspective reconstruction from calibrated images (e.g., [L81, TH84]). Loosely speaking, affine is to paraperspective what projective is to perspective.

5 Linear combinations of paraperspective views

The set of paraperspective images of an object and the set of affine images of the object are identical. Ullman and Basri showed [UB91] (see also [B93, P90]) that the affine images of an object can be expressed by linear combinations of two images of the object. This result therefore applies also to the set of paraperspective images of the object. Given two non-degenerate images $I_1 = \{\vec{x}_1, \vec{y}_1\}$ ($\vec{x}_1, \vec{y}_1 \in \mathcal{R}^n$ contain the x and y coordinates of the feature points in I_1) and $I_2 = \{\vec{x}_2, \vec{y}_2\}$, when the points in I_1 and I_2 are ordered in correspondence, any other paraperspective image $I = \{\vec{x}, \vec{y}\}$ of the object can be expressed by

$$\begin{aligned} \vec{x} &= \alpha_1\vec{x}_1 + \alpha_2\vec{y}_1 + \alpha_3\vec{x}_2 + \alpha_4\vec{1} \\ \vec{y} &= \beta_1\vec{x}_1 + \beta_2\vec{y}_1 + \beta_3\vec{x}_2 + \beta_4\vec{1}. \end{aligned} \tag{43}$$

Notice that $\vec{y}_2$ is redundant since it can be expressed as a linear combination of the four basis vectors, $\vec{x}_1$, $\vec{y}_1$, $\vec{x}_2$, and $\vec{1}$. (So in fact we use only "1.5 views" to model the object.)

Given the image position of the reference point we obtain two constraints on the coefficients of the linear combinations. These constraints are similar to Eq. 32, namely,

$$\frac{\vec{a}^T\vec{a}}{1+u^2} = \frac{\vec{b}^T\vec{b}}{1+v^2} = \frac{\vec{a}^T\vec{b}}{uv}, \tag{44}$$

where now $\vec{a}$ and $\vec{b}$ are expressed in terms of the coefficients of the linear combination and of the parameters of the model images. Let s_1, u_1, and v_1 determine the coordinates of the reference point in the first model view, I_1 (which we choose to be the frame of reference for the parameters of the affine transformation, $\vec{a}$ and $\vec{b}$). Let s_2, u_2, and v_2 determine the coordinates of the reference point in the second model view, and let R be the rotation matrix relating between the two model images, then $\vec{a}$ and $\vec{b}$ are given by

$$\begin{aligned} \vec{a} &= \alpha_1 s_1(\vec{i} + u_1\vec{k}) + \alpha_2 s_1(\vec{j} + v_1\vec{k}) + \alpha_3 s_2(\vec{r}_1 + u_2\vec{r}_3) \\ \vec{b} &= \beta_1 s_1(\vec{i} + u_1\vec{k}) + \beta_2 s_1(\vec{j} + v_1\vec{k}) + \beta_3 s_2(\vec{r}_1 + u_2\vec{r}_3), \end{aligned} \tag{45}$$

where $\vec{i}$, $\vec{j}$, and $\vec{k}$ are the three unit vectors corresponding to the three main axes. Therefore,

$$\begin{aligned} \vec{a}^T\vec{a} &= \alpha_1^2 s_1^2(1+u_1^2) + \alpha_2^2 s_1^2(1+v_1^2) + \alpha_3^2 s_2^2(1+u_2^2) + 2\alpha_1\alpha_2 s_1^2 u_1 v_1 + \\ &\quad 2\alpha_1\alpha_3 s_1 s_2(r_{11} + u_1 r_{13} + u_2 r_{31} + u_1 u_2 r_{33}) + \\ &\quad 2\alpha_2\alpha_3 s_1 s_2(r_{12} + v_1 r_{13} + u_2 r_{32} + v_1 u_2 r_{33}), \end{aligned} \tag{46}$$

$$\begin{aligned} \vec{b}^T\vec{b} &= \beta_1^2 s_1^2(1+u_1^2) + \beta_2^2 s_1^2(1+v_1^2) + \beta_3^2 s_2^2(1+u_2^2) + 2\beta_1\beta_2 s_1^2 u_1 v_1 + \\ &\quad 2\beta_1\beta_3 s_1 s_2(r_{11} + u_1 r_{13} + u_2 r_{31} + u_1 u_2 r_{33}) + \\ &\quad 2\beta_2\beta_3 s_1 s_2(r_{12} + v_1 r_{13} + u_2 r_{32} + v_1 u_2 r_{33}), \end{aligned} \tag{47}$$

and

$$\begin{aligned} \vec{a}^T\vec{b} &= \alpha_1\beta_1 s_1^2(1+u_1^2) + \alpha_2\beta_2 s_1^2(1+v_1^2) + \alpha_3\beta_3 s_2^2(1+u_2^2) + 2(\alpha_1\beta_2 + \alpha_2\beta_1) s_1^2 u_1 v_1 + \\ &\quad 2(\alpha_1\beta_3 + \alpha_3\beta_1) s_1 s_2(r_{11} + u_1 r_{13} + u_2 r_{31} + u_1 u_2 r_{33}) + \\ &\quad 2(\alpha_2\beta_3 + \alpha_3\beta_2) s_1 s_2(r_{12} + v_1 r_{13} + u_2 r_{32} + v_1 u_2 r_{33}). \end{aligned} \tag{48}$$

Plugging these values into Eq. 32 we obtain two constraints on the coefficients that corresponds to the image, I, with respect to the two model images, I_1 and I_2.

6 Conclusions

We have shown that the set of all images of an object obtained by applying a rigid transformation followed by a paraperspective projection with arbitrary reference point and the set of all images of the object obtained by applying a 3-D affine transformation followed by an orthographic projection are identical. Consequently, all uncalibrated paraperspective images of an object can be constructed from a 3-D model of the object by applying an affine transformation to the model, and every affine image of the object represents some uncalibrated paraperspective image of the object. This implies in particular that the paraperspective images of an object can be expressed as linear combinations of any two non-degenerate images of the object. When the image position of the reference point is given the parameters of the affine transformation (and, likewise, the coefficients of the linear combinations) satisfy two quadratic constraints.

Conversely, when the values of parameters are given the image position of the reference point is determined by solving a bi-quadratic equation. These results, combined with previous results by Sugimoto and Murota [SM93] and Poelman and Kanade [PK94], imply that the affine images of an object are the first order approximations of the uncalibrated perspective images of the object.

Appendix

In this appendix we show that a solution for the equation system

$$\begin{aligned} \vec{a}^T\vec{a} &= s^2(1+u^2) \\ \vec{b}^T\vec{b} &= s^2(1+v^2) \\ \vec{a}^T\vec{b} &= s^2uv. \end{aligned} \tag{49}$$

exists and is unique (up to a reflection) for any non-degenerate affine transformation (that is, for every pair of linearly independent vectors $\vec{a}$ and $\vec{b}$).

We shall solve Eq. 49 by turning it to a bi-quadratic equation in u. Denote by

$$A = \vec{a}^T\vec{a}, \qquad B = \vec{b}^T\vec{b}, \qquad C = \vec{a}^T\vec{b}. \tag{50}$$

Then,

$$s^2 = \frac{A}{1+u^2} = \frac{B}{1+v^2} = \frac{C}{uv}. \tag{51}$$

From

$$\frac{A}{1+u^2} = \frac{C}{uv} \tag{52}$$

we obtain that

$$v = \frac{C}{A}\frac{(1+u^2)}{u}. \tag{53}$$

Substituting this into

$$\frac{A}{1+u^2} = \frac{B}{1+v^2} \tag{54}$$

we obtain that

$$\frac{B}{A}(1+u^2) = 1 + \frac{C^2}{A^2}\frac{(1+u^2)^2}{u^2}. \tag{55}$$

This results in a bi-quadratic equation in u:

$$(AB - C^2)u^4 + (AB - A^2 - 2C^2)u^2 - C^2 = 0. \tag{56}$$

Notice, first, that $\vec{a}$ and $\vec{b}$ are linearly dependent if and only if

$$C^2 = AB. \tag{57}$$

The case that $A = 0$ or $B = 0$ is a special case obtained when $\vec{a} = 0$ or $\vec{b} = 0$ respectively. For all non-degenerate affine transformations the following condition is satisfied:

$$AB - C^2 > 0. \tag{58}$$

This is because

$$\frac{C^2}{AB} = \frac{(\vec{a}^T\vec{b})^2}{(\vec{a}^T\vec{a})(\vec{b}^T\vec{b})} = \cos^2\theta \leq 1, \tag{59}$$

where θ denotes the angle between the vectors $\vec{a}$ and $\vec{b}$.

We now proceed to solving Eq. 56. Eq. 56 is quadratic in u^2, and so its solution is given by

$$u^2 = \frac{-(AB - A^2 - 2C^2) \pm \sqrt{(AB - A^2 - 2C^2)^2 + 4(AB - C^2)C^2}}{2(AB - C^2)}, \tag{60}$$

which can be simplified to

$$u^2 = \frac{A^2 - AB + 2C^2 \pm A\sqrt{(A - B)^2 + 4C^2}}{2(AB - C^2)}. \tag{61}$$

Define

$$\Delta^2 = A^2[(A - B)^2 + 4C^2]. \tag{62}$$

Clearly,

$$\Delta^2 \geq 0, \tag{63}$$

and so there exist either one or two solutions for u^2.

A single solution for u^2 is obtained when $\Delta = 0$. This happens in either of the two cases. When $A = 0$, which occurs only when the affine transformation is degenerate, or when $(A - B)^2 + 4C^2 = 0$. In the latter case $A = B$ and $C = 0$, which implies that the affine transformation is equivalent to a rigid transformation followed by a scaled-orthographic projection. The values of the parameters u, v, and s in this case are given by

$$s = \sqrt{A} = \sqrt{B}, \qquad u = 0, \qquad v = 0. \tag{64}$$

When $\Delta^2 > 0$ two solutions for u^2 are obtained. These solutions are given by

$$u^2 = \frac{A^2 - AB + 2C^2 \pm \Delta}{2(AB - C^2)}. \tag{65}$$

Denote the two solutions by u_+^2 and u_-^2, we show below that one of these solutions is non-negative, whereas the other solution is non-positive. Notice that the denominator in Eq. 61 is always positive (due to Eq. 58). To figure the sign of the numerator we notice that the inequality

$$(A^2 - AB + 2C^2)^2 \leq A^2((A - B)^2 + 4C^2). \tag{66}$$

always holds. This is evident since by simplifying this inequality we obtain that

$$C^2(C^2 - AB) \leq 0. \tag{67}$$

Clearly, Eq. 67 holds since $C^2 \geq 0$ and, due to Eq. 58, $C^2 - AB < 0$. Eq. 66 implies that one of the solutions to Eq. 56 must be non-negative, while the other must be non-positive. To see that, take the square root of Eq. 66

$$-A\sqrt{(A-B)^2 + 4C^2} \leq A^2 - AB + 2C^2 \leq A\sqrt{(A-B)^2 + 4C^2}. \tag{68}$$

The left inequality implies that for one of the two solutions the numerator in Eq. 61 is non-negative:

$$A^2 - AB + 2C^2 + A\sqrt{(A-B)^2 + 4C^2} \geq 0. \tag{69}$$

Therefore, $u_+^2 \geq 0$. The right inequality implies that the numerator in Eq. 61 for the other solution is non-positive:

$$A^2 - AB + 2C^2 - A\sqrt{(A-B)^2 + 4C^2} \leq 0. \tag{70}$$

Consequently, $u_-^2 \leq 0$.

So far we have shown that Eq. 56, which is quadratic in u^2, has two solutions for u^2, one is always non-negative, whereas the other is non-positive. These solutions may vanish only when $C = 0$. This is evident from Eq. 67. The case that also $A = B$ has been handled in Eq. 64. The case that $C = 0$ and $A \neq B$ yields either one of the two following solutions:

$$s = \sqrt{A}, \qquad u = 0, \qquad v = \pm\sqrt{\frac{B}{A} - 1} \tag{71}$$

if $A \leq B$, or

$$s = \sqrt{B}, \qquad u = \pm\sqrt{\frac{A}{B} - 1}, \qquad v = 0 \tag{72}$$

if $B < A$.

Whenever $C \neq 0$ we obtain that u_+^2 must be positive, whereas u_-^2 must be negative. Thus, the solution u_-^2 is not geometrically realizable. u_+^2, however, provides two solutions for u that differ only in their sign. Using Eq. 49 we can recover also s and v:

$$s^2 = \frac{A}{1 + u^2}. \tag{73}$$

For each value of u we obtain two solutions to s which differ in signs. We generally assume that only the positive solution is feasible. Finally,

$$v^2 = \frac{B}{s^2} - 1, \tag{74}$$

and the sign of v is determined by

$$C = s^2 uv. \tag{75}$$

To summarize, in all cases (including when $C = 0$) we obtain a single solution for $s > 0$ and exactly two solutions for u, and v, (u, v) and $(-u, -v)$. These two solutions are related by a reflection.

Acknowledgment

I thank David Jacobs for insightful discussions and in particular for his contribution to proving Theorem 2.

References

[A90] Aloimonos, J., 1990. Perspective approximations. *Image and Vision Computing, 8(3)*: 177–192.

[B93] Basri, R., 1993. Viewer-Centered Representations in Object Recognition: A Computational Approach. In C.H. Chen, L.F. Pau, and P.S.P. Wang (Eds), *Handbook of Pattern Recognition and Computer Vision*. World Scientific Publishing Company, Singapore. Chapter 5.4: 863–882.

[F92] Faugeras, O.D., 1992. What can be seen in three dimensions with an uncalibrated stereo rig? *European Conf. on Computer Vision (ECCV-92)*: 564–578.

[J94] Jacobs, D., 1994. "Matching 3-D Models to 2-D Images". *International Journal of Computer Vision*, forthcoming.

[KV91] Koenderink, J. and van Doorn, A., 1991. "Affine structure from motion", *Journal of the Optical Society of America*, **8**(2): 377–385.

[L81] Longuet-Higgins, H.C., 1981. A computer algorithm for reconstructing a scene from two projections. *Nature*, **293**: 133–135.

[OMS81] Ohta, Y., Maenobu, K., and Sakai, T., 1981. Obtaining surface orientation from textels under perspective projection. *Proc. of the 7th International Joint Conference on Artificial Intelligence*: 746–751.

[PK94] Poelman, C.J. and Kanade, T., 1994. A paraperspective factorization method for shape and motion recovery. *European Conf. on Computer Vision (ECCV-94)*.

[P90] Poggio, T., 1990. 3D object recognition: on a result by Basri and Ullman, *TR 9005-03, IRST, Povo, Italy.*

[S95] Sugimoto, A., 1995. Object recognition by combining paraperspective images. *International Journal of Computer Vision*, forthcoming.

[SM93] Sugimoto, A. and Murota,, K., 1993. 3D object recognition by combination of perspective images. *Proc. of SPIE*, **1904**: 183–195.

[TH84] Tsai, R.Y. and Huang, T.S., 1984. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **6**(1): 13–27.

[UB91] Ullman, S. and Basri, R., 1991. Recognition by linear combinations of models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **13**(10): 992–1006.

[WT92] Weinshall, D. and Tomasi, C. (1992). Linear and incremental acquisition of invariant shape models from image sequences. *International Journal of Computer Vision*, forthcoming.