TR－A－0157

Recognition by Combinations of Paraperspective Images

Akihiro SUGIMOTO　Kazuo MUROTA

1992.11. 4

ATR 視聴覚機構研究所

〒619-02 京都府相楽郡精華町光台 2-2 ☎07749-5-1411

ATR Auditory and Visual Perception Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1411
Facsimile: +81-7749-5-1408

Recognition by Combinations of Paraperspective Images*

Akihiro SUGIMOTO† Kazuo MUROTA‡

†ATR Auditory and Visual Perception Research Laboratories
Soraku-gun, Kyoto 619-02, Japan
e-mail: sugimoto@atr-hr.atr.co.jp

‡Research Institute for Mathematical Sciences
Kyoto University, Kyoto 606, Japan

Abstract

This paper studies an object recognition problem under paraperspective projection, that is, the problem of determining whether a given paraperspective image is obtained from a 3-D object to be recognized or not. It is found that any paraperspective image of an object can be expressed as a linear combination of three appropriate paraperspective images of the same object. We show that any image of an object with not only a rigid 3-D transformation but also a nonrigid transformation has this property. In order to recognize a 3-D object, we have only to store three paraperspective images and, whenever a new paraperspective image is given, determine whether it can be expressed as a combination of the three images. This implies that we no longer need to recover the 3-D information of an object explicitly under paraperspective projection. Our investigation shows that three paraperspective images have sufficient information to recognize a 3-D object.

*Submitted to the 8th Scandinavian Conference on Image Analysis, 1993.

Contents

1 Introduction

Consider a 3-D object to be recognized and suppose that a number of 2-D images of the object from various viewpoints are already stored. We investigate an object recognition problem, that is, the problem of determining whether a given new image is obtained from the same object or not. A typical approach [2] to the problem is divided into three procedures: To recover the 3-D information of the object, to determine the relative position between the object and the viewpoint, and to determine whether or not it is obtained from the same object based upon the results of the two previous procedures. Whereas the first procedure can be done before being given a new image, the latter two are executed each time a new image is given.

Recently Ullman–Basri [4], [6] showed that three images are sufficient to describe any other image of the same object under orthographical projection and that any image can be described as linear combination of the three images. Their results led to an approach which does not explicitly recover 3-D information of the object and which recognizes the object if the new image can be expressed as linear combination of the three stored images. Sugimoto–Murota [5] extended their results to the case of perspective projection, showing that four images are sufficient to recognize an object under perspective projection, and that an image can be described as a certain nonlinear combination of the four images. However, it is not an easy computational task to determine whether or not an image can be described as a nonlinear combination of the stored images.

Orthographical projection is convenient, being a very rough approximation to the projection of light on the retina. On the other hand, perspective projection, which is the true model, often leads to complicated equations for many problems and makes the subsequent analysis hard. As a compromise, Ohta–Maenobu–Sakai [3] proposed a new model, named paraperspective projection by Aloimonos [1], to approximate the distortion of a texel pattern under perspective projection. Paraperspective projection stands in complexity between the orthographical and the perspective. It is a good approximation to perspective projection when the size of an object is sufficiently small, compared with the distance between the object and the viewpoint.

This paper is a study on the object recognition under paraperspective projection. As in the case of orthographical projection, any image can be expressed as a certain combination of several images of the same object. Three images are found to be sufficient, though the number of the required images for such description depends upon the representation of admissible transformations. The object recognition problem under paraperspective projection is thus reduced to the problem of determining whether or not the image is described as a combination of the three stored images. The approximation of perspective projection by paraperspective one makes the object recognition problem solvable with a computationally simple procedure.

The outline of this paper is as follow. In Section 2 we introduce paraperspective projection and show that it is the first order approximation to perspective projection. In Sections 3 and 4, we formulate the problem to solve and give a mathematical description to paraperspective images. In Section 5 we consider three representations of admissible transformations and show that, in either case, any image of the same object can be described as combinations of the several images. First we discuss a simple representation, linear combination (in the ordinary sense) representation of admissible transformations, and then exploit other representations to reduce the number of required images for the description of other images. In Section

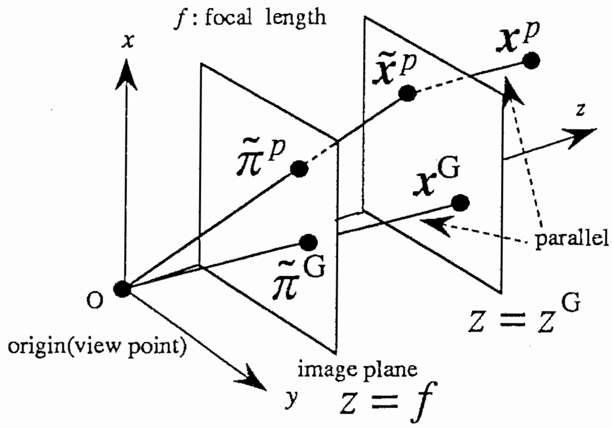


Fig. 1: Principle of paraperspective projection

6 we present an algorithm for recognition under paraperspective projection and show some experimental results. In this paper, we assume that an object moves around a fixed viewpoint, and that the motion is described by an affine transformation or by a rigid transformation.

2 Paraperspective projection

2.1 Definition of paraperspective projection

The notion of paraperspective projection was introduced by Y. Ohta, K. Maenobu and T. Sakai (see [3]) and named by J. Aloimonos (see [1]). It globally preserves the properties of perspective projection and locally realizes orthographical projection. Suppose that the center of a lens whose focal length is f coincides with the origin and that the z axis is aligned with the optical axis. Let[1] $\boldsymbol{x}^{\mathrm{G}} = (x^{\mathrm{G}}, y^{\mathrm{G}}, z^{\mathrm{G}})^{\mathrm{T}}$ be the coordinates of a reference point[2] G under paraperspective projection. Then a point p (with coordinates $\boldsymbol{x}^p$) in 3-D space is paraperspectively projected to $\tilde{\boldsymbol{\pi}}^p$ in the image plane ($z = f$) as follows (see Fig. 1):

1. $\boldsymbol{x}^p$ is first projected to[3] $\tilde{\boldsymbol{x}}^p \, (\in \mathbf{R}^3)$ on the plane $z = z^{\mathrm{G}}$, which is parallel to the image plane. The projection is performed by using the ray that is parallel to the ray OG going through the origin O and the reference point G.

2. $\tilde{\boldsymbol{x}}^p$ is then projected perspectively to $\left((\tilde{\boldsymbol{\pi}}^p)^{\mathrm{T}}, f\right)^{\mathrm{T}}$ in the image plane, where $\tilde{\boldsymbol{\pi}}^p \in \mathbf{R}^2$.

For $\boldsymbol{x}^p = (x^p, y^p, z^p)^{\mathrm{T}}$, we get

$$\begin{aligned} \tilde{\boldsymbol{\pi}}^p &= (\tilde{\pi}_1^p, \tilde{\pi}_2^p)^{\mathrm{T}} && (2.1) \\ &= f \begin{pmatrix} \frac{1}{z^{\mathrm{G}}} & 0 & -\frac{x^{\mathrm{G}}}{(z^{\mathrm{G}})^2} \\ 0 & \frac{1}{z^{\mathrm{G}}} & -\frac{y^{\mathrm{G}}}{(z^{\mathrm{G}})^2} \end{pmatrix} \boldsymbol{x}^p + f \begin{pmatrix} \frac{x^{\mathrm{G}}}{z^{\mathrm{G}}} \\ \frac{y^{\mathrm{G}}}{z^{\mathrm{G}}} \end{pmatrix}. && (2.2) \end{aligned}$$

[1] We use a column vector and denote by $\boldsymbol{x}^{\mathrm{T}}$ the transposition of a vector $\boldsymbol{x}$.

[2] We take it that the centroid of the feature points is a reference point (see (2.3)).

[3] $\mathbf{R}$ means the set of real numbers.

Since $\boldsymbol{x}^{\mathrm{G}}$ is the centroid of the feature points, we easily obtain

$$\boldsymbol{x}^{\mathrm{G}} = \frac{1}{P}\sum_{p=1}^{P}\boldsymbol{x}^{p}. \tag{2.3}$$

It is clear that paraperspective projection decomposes the image distortions into two parts: Step 1 captures the foreshortening distortion and part of the position effect, and Step 2 captures both the distance and the position effect. Paraperspective projection coincides with perspective projection for points in the plane $z = z^{\mathrm{G}}$.

Remark 2.1 When we let $f \to \infty$ in (2.2), $\tilde{\boldsymbol{\pi}}^{p}$ does not tend to the orthographical image of p. Instead, we should first shift the coordinate system by $-f$ along the z axis and then take the limit $f \to \infty$. We choose in this paper the coordinate system where the viewpoint and the origin coincide so that a rotation can be expressed as a 3×3 orthogonal matrix . □

Remark 2.2 In perspective projection, $\boldsymbol{x}^{p} = (x^{p}, y^{p}, z^{p})^{\mathrm{T}}$ is projected to $\boldsymbol{\pi}^{p} = (\pi_1^{p}, \pi_2^{p})^{\mathrm{T}}$ as follows:

$$\pi_1^{p} = \frac{x^{p}}{z^{p}}f, \qquad \pi_2^{p} = \frac{y^{p}}{z^{p}}f. \tag{2.4}$$

The coordinates of a point is not linearly related to the coordinates of its perspective image, whereas they are linear for its paraperspective image (see (2.2)). □

2.2 Meaning of paraperspective projection

Paraperspective projection is realized by the procedure explained above. Here we show that paraperspective projection is the first order approximation of perspective projection. In accordance with the notation introduced before, suppose a point p, with coordinates $\boldsymbol{x}^{p}$, is paraperspectively projected to $\tilde{\boldsymbol{\pi}}^{p}$ and perspectively projected to $\boldsymbol{\pi}^{p}$. The coordinates of the reference point under paraperspective projection are denoted as $\boldsymbol{x}^{\mathrm{G}}$. Let $\delta\boldsymbol{x}^{p} = (\delta x^{p}, \delta y^{p}, \delta z^{p})^{\mathrm{T}}$ be defined by

$$\begin{pmatrix} \delta x^{p} \\ \delta y^{p} \\ \delta z^{p} \end{pmatrix} := \begin{pmatrix} x^{p} - x^{\mathrm{G}} \\ y^{p} - y^{\mathrm{G}} \\ z^{p} - z^{\mathrm{G}} \end{pmatrix}. \tag{2.5}$$

From (2.4) we have

$$\pi_1^{p} = \frac{x^{\mathrm{G}} + \delta x^{p}}{z^{\mathrm{G}} + \delta z^{p}}f, \quad \pi_2^{p} = \frac{y^{\mathrm{G}} + \delta y^{p}}{z^{\mathrm{G}} + \delta z^{p}}f. \tag{2.6}$$

We assume

$$\left|x^{\mathrm{G}}\right| \gg |\delta x^{p}|, \ \left|y^{\mathrm{G}}\right| \gg |\delta y^{p}|, \ \left|z^{\mathrm{G}}\right| \gg |\delta z^{p}| \tag{2.7}$$

and take up to the first order terms in the Taylor expansion of (2.6) around $\boldsymbol{x}^{\mathrm{G}}$, then we get

$$\pi_1^{p} = f\left[\frac{x^{\mathrm{G}}}{z^{\mathrm{G}}} + \frac{1}{z^{\mathrm{G}}}\delta x^{p} - \frac{x^{\mathrm{G}}}{(z^{\mathrm{G}})^2}\delta z^{p} + \cdots\right], \tag{2.8}$$

$$\pi_2^{p} = f\left[\frac{y^{\mathrm{G}}}{z^{\mathrm{G}}} + \frac{1}{z^{\mathrm{G}}}\delta y^{p} - \frac{y^{\mathrm{G}}}{(z^{\mathrm{G}})^2}\delta z^{p} + \cdots\right]. \tag{2.9}$$

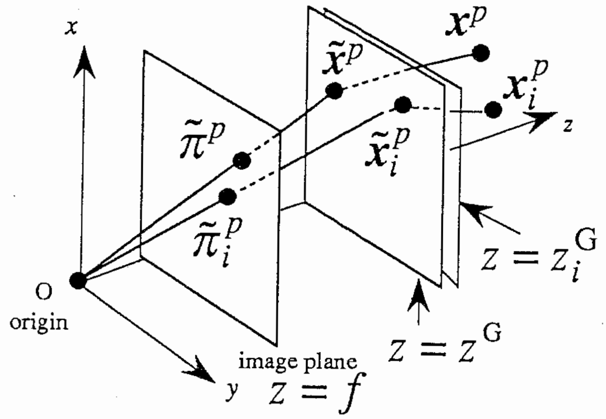


Fig. 2: A transformation i of a point p

On the other hand, from (2.2) we obtain

$$\tilde{\pi}_1^p = f\left[\frac{x^G + \delta x^p}{z^G} - \frac{x^G}{(z^G)^2}\delta z^p\right], \tag{2.10}$$

$$\tilde{\pi}_2^p = f\left[\frac{y^G + \delta y^p}{z^G} - \frac{y^G}{(z^G)^2}\delta z^p\right]. \tag{2.11}$$

It is clear that $\tilde{\boldsymbol{\pi}}^p$ is the first order approximation of $\boldsymbol{\pi}^p$. Therefore, when the distance between the object and the view point is sufficiently large, compared with the size of the object (see (2.7)), paraperspective projection will be a good approximation to perspective projection.

3 Formulation of the problem

In this section we formulate our problem in a well-defined form. The following are assumed:

- Any image is paraperspectively obtained.
- Feature points in an image are correctly extracted.
- The set of the points of which the object consists uniquely corresponds to the set of the feature points in images, and these two sets are independent of any transformation of the object.
- The set of the feature points in images is fixed, and the correspondence among the feature points is known.

Now suppose that a point p (with coordinates $\boldsymbol{x}^p$) moves to $\boldsymbol{x}_i^p$ with a transformation i and that it is paraperspectively projected to $\tilde{\boldsymbol{\pi}}_i^p$ (see Fig. 2). When p is subject to an affine transformation, the transformation i is characterized as follows:

$$\boldsymbol{x}_i^p = R_i \boldsymbol{x}^p + \boldsymbol{t}_i, \tag{3.1}$$

where

$$R_i \in \mathrm{GL}(3), \quad \boldsymbol{t}_i \in \mathrm{R}^3. \tag{3.2}$$

Here GL(3) denotes the general linear group of degree 3 over $\mathbf{R}$. In case that p is subject to a rigid transformation, (3.2) is to be replaced by

$$R_i \in \mathrm{SO}(3), \quad \boldsymbol{t}_i \in \mathbf{R}^3. \tag{3.3}$$

Note that SO(3) denotes the special orthogonal group of degree 3 over $\mathbf{R}$. It is clear that

$$\boldsymbol{x}_i^{\mathrm{G}} = R_i \boldsymbol{x}^{\mathrm{G}} + \boldsymbol{t}_i \tag{3.4}$$

follows from (2.3). Put

$$U_i := f \begin{pmatrix} 1/z_i^{\mathrm{G}} & 0 & -x_i^{\mathrm{G}}/(z_i^{\mathrm{G}})^2 \\ 0 & 1/z_i^{\mathrm{G}} & -y_i^{\mathrm{G}}/(z_i^{\mathrm{G}})^2 \end{pmatrix}, \tag{3.5}$$

$$V_i := f \begin{pmatrix} x_i^{\mathrm{G}}/z_i^{\mathrm{G}} \\ y_i^{\mathrm{G}}/z_i^{\mathrm{G}} \end{pmatrix}, \tag{3.6}$$

then we obtain

$$\tilde{\boldsymbol{\pi}}_i^p = U_i \boldsymbol{x}_i^p + V_i. \tag{3.7}$$

Let $\{\tilde{\boldsymbol{\pi}}_i^p\}_{p=1}^P$ denote the image of an object to be recognized with a transformation $i\,(\,i \in \{1, 2, \cdots, I\})$ and $\{\tilde{\boldsymbol{\pi}}_*^p\}_{p=1}^P$ denote a new image. The problem we investigate here is to determine whether or not the image $\{\tilde{\boldsymbol{\pi}}_*^p\}_{p=1}^P$ is obtained from the same object with a certain transformation. We assume a class of admissible transformations of a 3-D object is specified. This is because the decision whether the new image is obtained from the same object or not depends upon the class of admissible transformations. In this paper, we consider two classes of admissible transformations: affine transformations $\mathcal{A}_{\mathrm{a}}$ and rigid transformations $\mathcal{A}_{\mathrm{r}}$. Both form a group and are expressed respectively as follows:

$$\mathcal{A}_{\mathrm{a}} = \{(R,\ \boldsymbol{t}) \mid R \in \mathrm{GL}(3),\ \boldsymbol{t} \in \mathbf{R}^3\}, \tag{3.8}$$

$$\mathcal{A}_{\mathrm{r}} = \{(R,\ \boldsymbol{t}) \mid R \in \mathrm{SO}\,(3)\,,\ \boldsymbol{t} \in \mathbf{R}^3\}. \tag{3.9}$$

Since $\mathrm{SO}(3) \subset \mathrm{GL}(3)$, elements of $\mathcal{A}_{\mathrm{r}}$ are characterized as those elements of $\mathcal{A}_{\mathrm{a}}$ which satisfy some conditions. Therefore, (3.9) can be rewritten as

$$\mathcal{A}_{\mathrm{r}} = \{(R,\ \boldsymbol{t}) \in \mathcal{A}_{\mathrm{a}} \mid R \in \mathrm{SO}(3)\}. \tag{3.10}$$

In this paper, we regard $\mathcal{A}_{\mathrm{r}}$ as part of $\mathcal{A}_{\mathrm{a}}$ with the conditions. We write $i \in \mathcal{A}$ as a shorthand notation for $(R_i,\ \boldsymbol{t}_i) \in \mathcal{A}$. For a class of admissible transformations $\mathcal{A}$, put

$$\tilde{\Pi}^p := \{\tilde{\boldsymbol{\pi}}_i^p \mid \tilde{\boldsymbol{\pi}}_i^p = U_i \boldsymbol{x}_i^p + V_i,\ \exists i \in \mathcal{A}\}\,. \tag{3.11}$$

The problem is formulated as follow.

Problem 3.1 Suppose a class $\mathcal{A}$ of admissible transformations is specified. Find out a procedure which, treating $\{\tilde{\boldsymbol{\pi}}_i^p\}_{p=1}^P$ ($i \in \{1, 2, \cdots, I\}$) directly, determines whether or not $\tilde{\boldsymbol{\pi}}_*^p \in \tilde{\Pi}^p$ for all $p \in \{1, 2, \cdots, P\}$ every time $\{\tilde{\boldsymbol{\pi}}_*^p\}_{p=1}^P$ is given. □

We assume a representation of admissible transformations for further investigation. This is because a procedure to be constructed depends on the representation of admissible transformations.

In Section 5, we consider three representations, all of which are linear equations in the elements of admissible transformations.

4 Mathematical description of images

4.1 Coordinates in the image plane

Since

$$\tilde{\pi}_i^{\mathrm{G}} = \frac{1}{P}\sum_{p=1}^{P} \tilde{\pi}_i^p \tag{4.1}$$

follows from (2.2) and (2.3), we can calculate $\tilde{\pi}_i^{\mathrm{G}}$ easily from $\tilde{\pi}_i^p\,(p = 1, 2\cdots, P)$. We denote the increment of $\tilde{\pi}_i^p$ from $\tilde{\pi}_i^{\mathrm{G}}$ by

$$\rho_i^p := \tilde{\pi}_i^p - \tilde{\pi}_i^{\mathrm{G}}, \tag{4.2}$$

and put

$$\check{\Pi}^p := \left\{\rho_i^p \mid \rho_i^p = \tilde{\pi}_i^p - \tilde{\pi}_i^{\mathrm{G}},\ \tilde{\pi}_i^p \in \tilde{\Pi}^p\right\}. \tag{4.3}$$

Since ρ_i^p is easily calculated from $\tilde{\pi}_i^p$, we may concentrate on ρ_i^p instead of $\tilde{\pi}_i^p$. In other words, we may regard $\{\rho_i^p\}_{p=1}^P\,(i = 1, 2, \cdots, I)$ and $\{\rho_*^p\}_{p=1}^P$ as the stored image and a new image respectively.

From (4.2) we have

$$\rho_i^p = f\begin{pmatrix} \frac{\delta x_i^p}{z_i^{\mathrm{G}}} - \frac{x_i^{\mathrm{G}}}{(z_i^{\mathrm{G}})^2}\delta z_i^p \\ \frac{\delta y_i^p}{z_i^{\mathrm{G}}} - \frac{y_i^{\mathrm{G}}}{(z_i^{\mathrm{G}})^2}\delta z_i^p \end{pmatrix} \tag{4.4}$$

$$= \frac{f}{(z_i^{\mathrm{G}})^2} Q\left[\boldsymbol{x}_i^{\mathrm{G}}\right]\delta\boldsymbol{x}_i^p, \tag{4.5}$$

where

$$Q := \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} \tag{4.6}$$

and for $\boldsymbol{x} = (x_1,\ x_2,\ x_3)^{\mathrm{T}}$ in general, $[\boldsymbol{x}]$ is defined by

$$[\boldsymbol{x}] := \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix}. \tag{4.7}$$

The operation $[\boldsymbol{x}]$ has the following properties.

Lemma 4.1 *Let*[4] $\boldsymbol{x},\ \boldsymbol{y} \in \mathrm{R}^3$ *and* $R \in \mathrm{GL}(3)$.

$$[\boldsymbol{x} + \boldsymbol{y}] = [\boldsymbol{x}] + [\boldsymbol{y}], \tag{4.8}$$

$$R^{\mathrm{T}}[R\boldsymbol{x}]R = (\det R)\,[\boldsymbol{x}]. \tag{4.9}$$

□

[4] det$R$ means the determinant of a square matrix R.

4.2 Images

Assume that, for $i \in \{1, 2, \cdots, I\}$, the stored image $\{\boldsymbol{\rho}_i^p\}_{p=1}^P$ is obtained from $\{\boldsymbol{x}_i^p\}_{p=1}^P$ that satisfies

$$\boldsymbol{x}_i^p = R_i \boldsymbol{x}^p + \boldsymbol{t}_i, \quad p \in \{1, 2, \cdots, P\}. \tag{4.10}$$

This implies

$$\delta \boldsymbol{x}_i^p = R_i \delta \boldsymbol{x}^p, \tag{4.11}$$

$$\boldsymbol{x}_i^{\mathrm{G}} = R_i \boldsymbol{x}^{\mathrm{G}} + \boldsymbol{t}_i, \tag{4.12}$$

from which it follows[5]

$$[\boldsymbol{x}_i^{\mathrm{G}}]\, \delta \boldsymbol{x}_i^p = (\det R_i)\, R_i^{-\mathrm{T}} [\boldsymbol{x}^{\mathrm{G}} + R_i^{-1} \boldsymbol{t}_i]\, \delta \boldsymbol{x}^p \tag{4.13}$$

by Lemma 4.1. Substituting this into (4.5) we obtain

$$\boldsymbol{\rho}_i^p = f\left(\check{R}_i [\boldsymbol{x}^{\mathrm{G}}] + \check{T}_i\right) \delta \boldsymbol{x}^p, \tag{4.14}$$

where

$$\check{R}_i := \frac{\det R_i}{(z_i^{\mathrm{G}})^2} Q R_i^{-\mathrm{T}}, \tag{4.15}$$

$$\check{T}_i := \frac{\det R_i}{(z_i^{\mathrm{G}})^2} Q R_i^{-\mathrm{T}} [R_i^{-1} \boldsymbol{t}_i]. \tag{4.16}$$

Here $\check{R}_i$ and $\check{T}_i$ are both 2×3 matrices. The stored images are expressed as (4.14).

Remark 4.1 In case of $R_i \in \mathrm{GL}(3)$, both $\check{R}_i$ and $\check{T}_i$ can be any 2×6 matrices, whereas in case of $R_i \in \mathrm{SO}(3)$, the conditions

$$\det R_i = 1, \quad R_i^{-\mathrm{T}} = R_i \tag{4.17}$$

are satisfied. Hence the first row vector of $\check{R}_i$ is equal to the second row vector of R_i multiplied by a constant c and the second row vector of $\check{R}_i$ is equal to the first row vector of R_i multiplied by $-c$. In other words, the two row vectors of $\check{R}_i$ are orthogonal and have the same norm. □

As for a new image $\{\boldsymbol{\rho}_*^p\}_{p=1}^P$, suppose similarly that

$$\boldsymbol{x}_*^p = R_* \boldsymbol{x}^p + \boldsymbol{t}_*, \quad p \in \{1, 2, \cdots, P\} \tag{4.18}$$

are satisfied. Then by putting

$$\check{R}_* := \frac{\det R_*}{(z_*^{\mathrm{G}})^2} Q R_*^{-\mathrm{T}}, \tag{4.19}$$

$$\check{T}_* := \frac{\det R_*}{(z_*^{\mathrm{G}})^2} Q R_*^{-\mathrm{T}} [R_*^{-1} \boldsymbol{t}_*], \tag{4.20}$$

we obtain

$$\rho_*^p = f\left(\check{R}_* [\boldsymbol{x}^{\mathrm{G}}] + \check{T}_*\right) \delta \boldsymbol{x}^p. \tag{4.21}$$

[5] $R^{-\mathrm{T}}$ means $(R^{\mathrm{T}})^{-1}$.

5 Representations of images

Here we consider three representations of admissible transformations and investigate how an image can be described by a combination of the stored images. We define the following 2×3 matrices:

$$\check{C} := \left(\check{R} \mid \check{T}\right), \check{C}_i := \left(\check{R}_i \mid \check{T}_i\right), \check{C}_* := \left(\check{R}_* \mid \check{T}_*\right). \tag{5.1}$$

5.1 Linear combination I

The matrix $\check{C}_*$ in (5.1), being a 2×6 matrix, can also be thought of as a vector in $\mathbf{R}^{12}$. Therefore, if $\{\check{C}_i\}_{i=1}^{I}$ spans $\mathbf{R}^{12}$, any $\check{C}_*$ can be expressed as

$$\check{C}_* = \sum_{i=1}^{I} \lambda_i \check{C}_i \tag{5.2}$$

in terms of the coefficient set $\{\lambda_i\}_{i=1}^{I}$. This is equivalent to

$$\check{R}_* = \sum_{i=1}^{I} \lambda_i \check{R}_i, \quad \check{T}_* = \sum_{i=1}^{I} \lambda_i \check{T}_i, \tag{5.3}$$

which yields a representation of $\check{R}_*$ and $\check{T}_*$. Substituting (5.3) into (4.21) we obtain

$$\begin{aligned}
\rho_*^p &= f\left\{\sum_{i=1}^{I} \lambda_i \check{R}_i[\boldsymbol{x}^{\mathrm{G}}] + \sum_{i=1}^{I} \lambda_i \check{T}_i\right\} \delta \boldsymbol{x}^p & (5.4)\\
&= \sum_{i=1}^{I} \lambda_i f\left(\check{R}_i[\boldsymbol{x}^{\mathrm{G}}] + \check{T}_i\right) \delta \boldsymbol{x}^p & (5.5)\\
&= \sum_{i=1}^{I} \lambda_i \rho_i^p. & (5.6)
\end{aligned}$$

Theorem 5.1 *Suppose $\mathcal{A}_{\mathrm{a}}$ (affine transformations) is the class of admissible transformations and that $\{\check{C}_i\}_{i=1}^{12}$ is linearly independent. Then for $\forall \rho_*^p \in \check{\Pi}^p$, there exists $\{\lambda_i\}_{i=1}^{12}$,independent of p, such that*

$$\rho_*^p = \sum_{i=1}^{12} \lambda_i \rho_i^p. \tag{5.7}$$

□

In case that $\mathcal{A}_{\mathrm{r}}$ (rigid transformations) is the class of admissible transformations, the two row vectors of $\check{R}_*$ are orthogonal and have the same norm (see Remark 4.1). Putting

$$\check{R}_i = \begin{pmatrix} (\check{\boldsymbol{r}}_1^i)^{\mathrm{T}} \\ \hline (\check{\boldsymbol{r}}_2^i)^{\mathrm{T}} \end{pmatrix} \tag{5.8}$$

we see the conditions[6]

$$\begin{aligned}
\|\sum_{i=1}^{12} \lambda_i \check{\boldsymbol{r}}_1^i\| &= \|\sum_{i=1}^{12} \lambda_i \check{\boldsymbol{r}}_2^i\|, & (5.9)\\
(\sum_{i=1}^{12} \lambda_i \check{\boldsymbol{r}}_1^i) \cdot (\sum_{i=1}^{12} \lambda_i \check{\boldsymbol{r}}_2^i) &= 0 & (5.10)
\end{aligned}$$

on $\{\lambda_i\}_{i=1}^{12}$ in Theorem 5.1.

[6] $\|\boldsymbol{x}\|$ means the Euclidean norm of a vector $\boldsymbol{x}$.

Theorem 5.2 *In case that $\mathcal{A}_{\rm r}$ is the class of admissible transformations, $\{\lambda_i\}_{i=1}^{12}$ in Theorem* 5.1 *is subject to* (5.9) and (5.10). □

Remark 5.1 Theorem 5.1 states that all the feature points in the images obtained from the same object should satisfy (5.7) with a common coefficient set $\{\lambda_i\}_{i=1}^{I}$. The converse is not true, namely, an image in which all the feature points satisfy (5.7) is not necessarily obtained from the same object. However, when the number of the feature points are sufficiently large, the probability is almost equal to zero that all the feature points of an image of a different object happen to satisfy (5.7). This remark applies also to the theorems below. □

5.2 Linear combination II

Put

$$\check{C}_i = \left(\frac{(\check{\boldsymbol{c}}_1^i)^{\rm T}}{(\check{\boldsymbol{c}}_2^i)^{\rm T}} \right). \tag{5.11}$$

If both $\{\check{\boldsymbol{c}}_1^i\}_{i=1}^{I}$ and $\{\check{\boldsymbol{c}}_2^i\}_{i=1}^{I}$ span $\mathbf{R}^6$ respectively, there exist $D_i := \begin{pmatrix} \mu_i & 0 \\ 0 & \nu_i \end{pmatrix}$ $(i = 1, 2 \cdots I)$ such that

$$\check{C}_* = \sum_{i=1}^{I} D_i \check{C}_i, \tag{5.12}$$

which is equivalent to

$$\check{R}_* = \sum_{i=1}^{I} D_i \check{R}_i, \quad \check{T}_* = \sum_{i=1}^{I} D_i \check{T}_i. \tag{5.13}$$

This gives another representation of $\check{R}_*$ and $\check{T}_*$. Then we obtain

$$\boldsymbol{\rho}_*^p = f\left\{ \sum_{i=1}^{I} D_i \check{R}_i [\boldsymbol{x}^{\rm G}] + \sum_{i=1}^{I} D_i \check{T}_i \right\} \delta \boldsymbol{x}^p \tag{5.14}$$

$$= \sum_{i=1}^{I} D_i \boldsymbol{\rho}_i^p. \tag{5.15}$$

Theorem 5.3 *Suppose that $\mathcal{A}_{\rm a}$ is the class of admissible transformations and that $\{\check{\boldsymbol{c}}_1^i\}_{i=1}^{6}$ and $\{\check{\boldsymbol{c}}_2^i\}_{i=1}^{6}$ are linearly independent respectively. Then for $\forall \boldsymbol{\rho}_*^p \in \check{\Pi}^p$, there exists $\{\mu_i, \nu_i\}_{i=1}^{6}$, independent of p, such that*

$$\boldsymbol{\rho}_*^p = \sum_{i=1}^{6} \begin{pmatrix} \mu_i & 0 \\ 0 & \nu_i \end{pmatrix} \boldsymbol{\rho}_i^p. \tag{5.16}$$

□

Theorem 5.4 *In case that $\mathcal{A}_{\rm r}$ is the class of admissible transformations, $\{\mu_i, \nu_i\}_{i=1}^{6}$ in Theorem* 5.3 *is subject to the following two conditions*:

$$\|\sum_{i=1}^{6} \mu_i \check{\boldsymbol{r}}_1^i\| = \|\sum_{i=1}^{6} \nu_i \check{\boldsymbol{r}}_2^i\|, \tag{5.17}$$

$$(\sum_{i=1}^{6} \mu_i \check{\boldsymbol{r}}_1^i) \cdot (\sum_{i=1}^{6} \nu_i \check{\boldsymbol{r}}_2^i) = 0. \tag{5.18}$$

□

We have shown that any image can be described as a combination of six appropriate images under the representation (5.13).

Remark 5.2 The representation (5.13) is similar to that of Ullman–Basri [6]. □

5.3 Linear combination III

If $\{\check{c}_1^i,\ \check{c}_2^i\}_{i=1}^3$ spans $\mathbf{R}^6$, there exists $\{a_i,\ b_i,\ c_i,\ d_i\}_{i=1}^3$ which satisfies

$$\check{C}_* = \sum_{i=1}^{3} \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \check{C}_i. \tag{5.19}$$

Putting

$$M_i := \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix}, \tag{5.20}$$

we rewrite (5.19) as

$$\check{R}_* = \sum_{i=1}^{3} M_i \check{R}_i, \quad \check{T}_* = \sum_{i=1}^{3} M_i \check{T}_i, \tag{5.21}$$

which gives a third representation of $\check{R}_*$ and $\check{T}_*$. Then,

$$\rho_*^p = f\left\{\sum_{i=1}^{3} \left(M_i\check{R}_i\right)[\boldsymbol{x}^{\mathrm{G}}] + \sum_{i=1}^{3} \left(M_i\check{T}_i\right)\right\} \delta\boldsymbol{x}^p \tag{5.22}$$

$$= \sum_{i=1}^{3} M_i \rho_i^p. \tag{5.23}$$

Theorem 5.5 *Suppose that $\mathcal{A}_{\mathrm{a}}$ is the class of admissible transformations and that $\{\check{c}_1^i,\ \check{c}_2^i\}_{i=1}^3$ is linearly independent. Then for $\forall \rho_*^p \in \check{\Pi}^p$, there exists $\{a_i,\ b_i,\ c_i,\ d_i\}_{i=1}^3$, independent of p, such that*

$$\rho_*^p = \sum_{i=1}^{3} \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \rho_i^p. \tag{5.24}$$

□

Theorem 5.6 *In case that $\mathcal{A}_{\mathrm{r}}$ is the class of admissible transformations, $\{a_i,\ b_i,\ c_i,\ d_i\}_{i=1}^3$ in Theorem 5.5 is subject to the following two conditions:*

$$\left\|\sum_{i=1}^{3} (a_i\check{r}_1^i + b_i\check{r}_2^i)\right\| = \left\|\sum_{i=1}^{3} (c_i\check{r}_1^i + d_i\check{r}_2^i)\right\|, \tag{5.25}$$

$$\sum_{i=1}^{3} (a_i\check{r}_1^i + b_i\check{r}_2^i) \cdot \sum_{i=1}^{3} (c_i\check{r}_1^i + d_i\check{r}_2^i) = 0. \tag{5.26}$$

□

We have demonstrated that for the two classes of admissible transformations (affine and rigid), any image can be expressed as a linear combination of three appropriate images under the representation (5.21).

6 Algorithm and experiments

6.1 Algorithm

In this section we describe an algorithm for object recognition based on the last representation considered in Section 5. It has the advantage that it requires the smallest number of images. Similar algorithms could be made for the other representations.

In Subsection 5.3 we proved that $\boldsymbol{\rho}_*^p$ can be expressed as a combination of $\{\boldsymbol{\rho}_i^p\}_{i=1}^3$ for all $p\,(p = 1, 2, \cdots, P)$. When a new image $\{\tilde{\boldsymbol{\pi}}_*^p\}_{p=1}^P$ is given, we first calculate $\{\boldsymbol{\rho}_*^p\}_{p=1}^P$ and then regard (5.24) as an overdetermined system of linear equations in $\{a_i, b_i, c_i, d_i\}_{i=1}^3$. Then we apply the method of least squares to see whether the residual is (almost) equal to zero or not. To be more specific, we define

$$f_p \;:=\; \|\sum_{i=1}^{3} \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \boldsymbol{\rho}_i^p - \boldsymbol{\rho}_*^p\|^2 \tag{6.1}$$

for $p \in \{1, 2, \cdots, P\}$ and

$$g_1 \;:=\; \left(\|\sum_{i=1}^{3}(a_i\check{\boldsymbol{r}}_1^i + b_i\check{\boldsymbol{r}}_2^i)\| - \|\sum_{i=1}^{3}(c_i\check{\boldsymbol{r}}_1^i + d_i\check{\boldsymbol{r}}_2^i)\|\right)^2, \tag{6.2}$$

$$g_2 \;:=\; \left(\sum_{i=1}^{3}(a_i\check{\boldsymbol{r}}_1^i + b_i\check{\boldsymbol{r}}_2^i) \cdot \sum_{i=1}^{3}(c_i\check{\boldsymbol{r}}_1^i + d_i\check{\boldsymbol{r}}_2^i)\right)^2. \tag{6.3}$$

In addition, in case that the rigid transformations are considered admissible, we define

$$h \;:=\; \sum_{p=1}^{P} f_p + g_1 + g_2, \tag{6.4}$$

whereas, in case of affine transformations, we put

$$h \;:=\; \sum_{p=1}^{P} f_p. \tag{6.5}$$

The following procedure determines, for a given paraperspective image, whether it is obtained from an object to be recognized or not.

$\boxed{\text{Algorithm}}$

1. Calculate $\tilde{\boldsymbol{\pi}}_*^{\mathrm{G}}$ (see (4.1)).
2. Calculate $\boldsymbol{\rho}_*^p$ for all $p\,(p \in \{1, 2, \cdots, P\})$ (see (4.2)).
3. Determine whether or not there exists
   $\{a_i,\ b_i,\ c_i,\ d_i\}_{i=1}^3$ such that $h(\{a_i, b_i, c_i, d_i\}_{i=1}^3)$ is (almost) equal to zero.
   - Exist $\Longrightarrow$ the same object.
   - Not exist $\Longrightarrow$ a different object.

□

Remark 6.1 The decision to be made in Step 3 should be "$h(\{a_i, b_i, c_i, d_i\}_{i=1}^3) = 0$" from theoretical point of view, whereas it should be "$h(\{a_i, b_i, c_i, d_i\}_{i=1}^3) \approx 0$" from practical point of view. This is due to the rounding errors in numerical computation. □

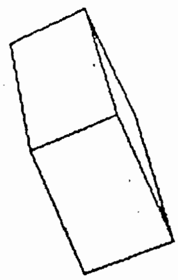


Fig. 3: The parallelepiped to be recognized

6.2 Experimental results

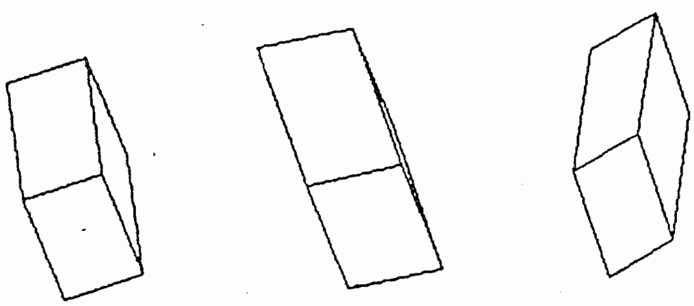
On the basis of the algorithm above, our experimental results are shown. Paraperspective images are artificially generated. Note that we fixed the focal length $f = 1.0$.

The object to be recognized is a parallelepiped (see Fig.3) with eight vertices: $(5.00, 6.50, 10.00)$, $(5.80, 5.96, 9.73)$, $(6.16, 7.04, 9.37)$, $(5.36, 7.58, 9.64)$, $(5.45, 6.05, 11.35)$, $(6.25, 5.51, 11.08)$, $(6.61, 6.59, 10.72)$, $(5.81, 7.13, 10.99)$. We regard the seven visible vertices as the feature points. Three stored images for the object are shown in Fig. 4. Each image is obtained with a transformation[7] in Table 1. The algorithm was applied to the three images in Fig. 5. The results are shown in Table 2 (first column). Since the values of ρ_i^p is $O(10^{-2})$, we set the

Table 1: Transformations of the stored images

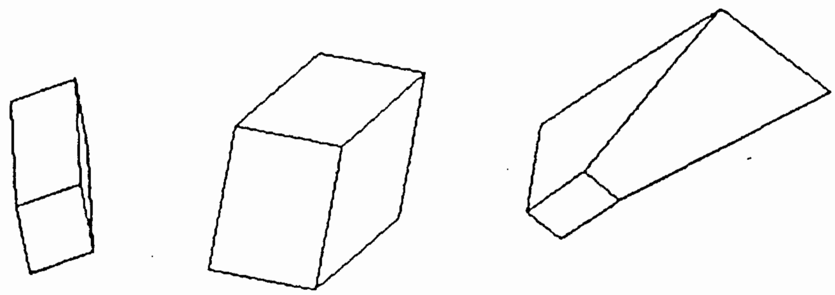
	rotation		translation
	axis	degree	(x, y, z)
(a)	x	10°	$(-0.50, 2.00, 0.00)$
(b)	x	15°	$(2.50, 4.00, -1.00)$
(c)	y	-10°	$(-4.50, 2.50, 0.50)$

[7]All the transfomations consist of a rotation followed by a translation.



(a) (b) (c)

Fig. 4: The stored images of the object in Fig. 3



(d) (e) (f)

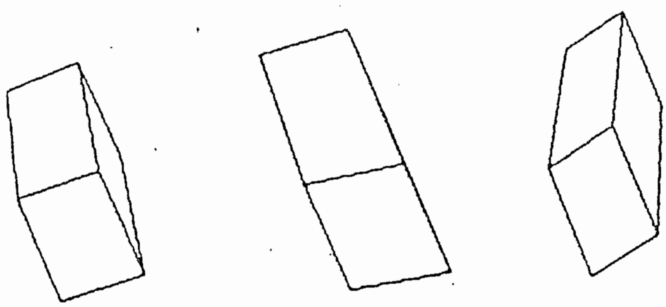
Fig. 5: New images

threshold for the decision "$h \approx 0$" to 1.0×10^{-5}(rather arbitrarily). Then Table 2 shows that both (d) and (e) are obtained from the same object, and that (f) is obtained from a different object. Actually in Fig. 5, (d) was obtained by rotating the parallelepiped by 30° around the x axis and then by $-30°$ around the y axis and then translating it by $(1.00, 4.50, -0.50)$; (e) was obtained by rotating the parallelepiped by 5°, 20°, and 30° around the x, the y and the z axes, respectively and then translating it by $(-3.00, -5.00, 0.00)$; whereas (f) was obtained by a frustum of pyramid.

Table 2: Minimum values of the cost function h

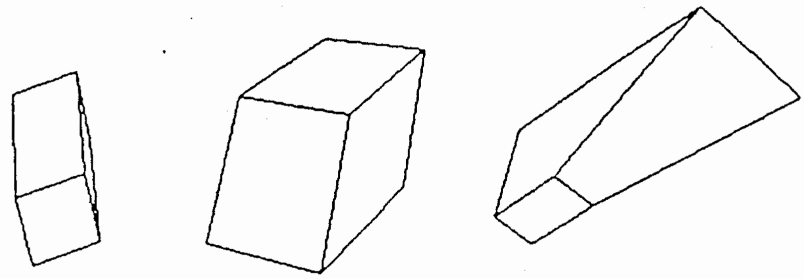
	h (paraperspective)	h (perspective)
(d)	5.63×10^{-14}	2.86×10^{-7}
(e)	5.99×10^{-13}	4.75×10^{-6}
(f)	1.14×10^{-2}	2.32×10^{-3}

Table 2 (second column) also gives the results of the algorithm applied to the perspective images under the same conditions. Note that the stored images and the new images are shown in Fig. 6 and in Fig. 7 respectively. Table 2 also shows that both (d) and (e) are obtained from the same object, and that (f) is obtained from a different object under the threshold



(a) (b) (c)

Fig. 6: The stored images of the object in Fig. 3 (perspective)



(d) (e) (f)

Fig. 7: New images (perspective)

1.0×10^{-5}.

Our experimental results indicate that our algorithm correctly determines whether a given image is obtained from the same object or not.

7 Conclusion

It is found that several images are sufficient to recognize any object under paraperspective projection. Any image can be described as a certain combination of the three images under the condition that the class of admissible transformations for an object is affine or rigid. This implies that we no longer need pre-procedure and that, when a new image is given, we have only to determine whether or not the cost function can be almost nullified by a suitable set of parameter values.

Left for future investigations are (1) the analysis of the errors incurred by the approximation of perspective projection by paraperspective projection, and (2) the analysis of the rounding errors in the actual implementation of the proposed algorithm.

Acknowledgments

The authors are grateful to Eiji Yodogawa of ATR Auditory and Visual Perception Research Laboratories for providing the opportunity for this research and to Keiichi Ueno there for his encouragement. They also appreciate the discussion with Shinjiro Kawato of ATR Human Information Processing Research Laboratories.

References

[1] J. Aloimonos: Shape from Texture, *Biological Cybernetics*, 58, 5, 345-360 (1989).

[2] P. J. Besl and R. C. Jain: Three-Dimensional Object Recognition, *ACM Computing Surveys*, 1, 17, 75-145 (1985).

[3] Y. Ohta, K. Maenobu and T. Sakai: Obtaining Surface Orientation from Texels under Perspective Projection, *Proc. of the 7th IJCAI*, 746-751, 1981.

[4] T. Poggio: *3D Object Recognition: On a Result of Basri and Ullman*, IRST Technical Report, 9005-03, Trento, Italy, 1990.

[5] A. Sugimoto and K. Murota: 3D Object Recognition by Combination of Perspective Images, *Proc. of SPIE'93*, to appear.

[6] S. Ullman and R. Basri: Recognition by Linear Combinations of Models, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-13, 10, 992-1006 (1991).