Paper Submitted in Korea-Japan Joint Conference

3-D Object Recognition Using a New Invariant Relationship by Single-View

Kyoung Sig Roh and In So Kweon

Dept. of Automation and Design Eng.,
Korea Advanced Institute of Science and Technology.
207-43, Cheongryangri-dong, Dongdaemoon-gu, Seoul, Korea.

Internet Address:
iskweon@cais.kaist.ac.kr
ksroh@cais.kaist.ac.kr

Keyword : Object Recognition, One-viewed Invariant Relationship, Hashing

3-D Object Recognition Using a New Invariant Relationship by Single-View

Kyoung Sig Roh and In So Kweon

Dept. of Automation and Design Eng.,
Korea Advanced Institute of Science and Technology.
207-43, Cheongryangri-dong, Dongdaemoon-gu, Seoul, Korea.
Internet Address: iskweon@cais.kaist.ac.kr

Abstract : *We propose a new method for recognizing three-dimensional objects using a three dimensional invariant relationship and geometric hashing by single-view. We develop a special structure consisting of four co-planar points and any two non-coplanar points with respect to the plane. We derive an invariant relationship for the structure, which is represented by a plane equation. For the recognition of 3-D objects using the geometric hashing, a set of points on the plane, thereby satisfying the invariant relationship, are mapped into a set of points intersecting the plane and the unit sphere. Since the structure is much more general than the previous structures proposed by Rothwell et al. [1] and Zhu et al. [2,3], it gives enough many voting to generate hypotheses. We also show that from the proposed invariant relationship, an invariant for the structure by two-view and an invariant for a structure proposed by Zhu et al. [2,3]can also be derived.*

Experiments using 3-D polyhedral objects and an outdoor building scene are carried out to demonstrate the feasibility of our method for 3-D object recognition

Keyword : 3-D Object Recognition, One-Viewed Invariant Relationship, Geometric Hashing

1. Introduction

Most of the invariants used so far in computer vision applications are based on plane-to-plane mappings. These invariants of the plane projective group have been very well studied and many forms have been known. They have also been successfully applied in working vision systems [4-6].

Constructing invariants for 3D structures from their 2D perspective images is much more difficult and represents the major goal of current research in the application of invariant theory to vision. Burns et al. [7] show that invariants can not be measured for 3D point sets in general position from a single view, that is, for sets that contain absolutely no structure.

There are three categories to solve this problem. In the first category, one basically deals with space projective invariants from two images, provided that the epipolar geometry of the two images is determined a priori [8,10,17]. Secondly, without computing the epipolar geometry space projective invariants from three images can be determined [11,12]. Thirdly, some special structures can provide projective invariants by one view [1-3].

Among three categories, the third approach does not need the correspondence information between features in each image. Rothwell et al. [1] proposed two special structures from which one-viewed projective invariant can be derived : One is for points that lie on the vertices of polyhedron, from which invariants are computed by using an algebraic framework of constraints between points and planes. The other is for objects that are bilateral symmetric. For the first class of objects, a minimum of seven points, that lie on the vertices of a six-sided polyhedron, are required in order to recover the projective structure. For the second class of objects, a minimum of eight points, or four points and two lines that are bilateral symmetric, are needed.

Zhu et al. [2,3] proposed an algorithm to compute an invariant based on a structure of six points on adjacent planes which provided two sets of four coplanar points. The invariant is less constrained than the invariant proposed by Rothwell et al. [1], because it needs only six points instead of seven.

In this paper, we propose a new invariant relationship for a structure that is even more general than one by Zhu et al. [2,3]. The structure consists of a set of six points ; four coplanar points and two non-coplanar points with respect to the plane. In general, this structure provides an invariant by two viewed images for which a priori epipolar geometry is not required [10]. However, we derive an invariant relationship for the structure using just one-view. The relationship can be represented as an orthogonal plane for a vector that is computed uniquely from the structure. To recognize three-dimensional objects, we propose a model-base using the geometric hashing in order to use the invariant relationship.

This paper is organized as follows. In section 2, the invariant relationship is derived for a structure which consists of four coplanar points and two non-coplanar points. In section 3, we present a method to construct the model-base by using the invariant relationship and the geometric hashing. In section 4, we present experimental results for real three dimensional objects.

2. Invariant relationship from a single view

In this section, we present a three dimensional projective invariant relationship from a single view, which is based on a structure with six points: four coplanar points and two non-coplanar points. We derive the invariant for the structure using a canonical frame concept [13].

Fig.1 shows the structure and the projection of the structure. $\mathbf{X}_i$, i=1~4 are four coplanar points and $\mathbf{X}_5$, $\mathbf{X}_6$ denote two non-coplanar points. And $\mathbf{x}_i$ are the corresponding image points.

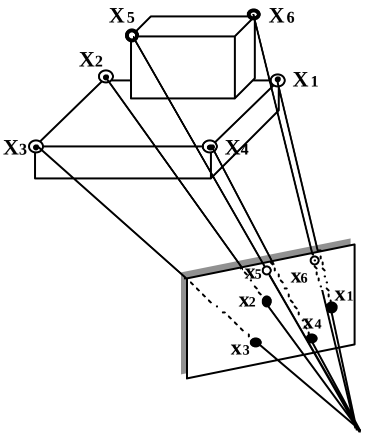


Fig.1 Projection of a set of six points

We assign canonical projective coordinates to the six points as follows:

$$
\begin{aligned}
\mathbf{X}_1 &= (X_1,\ Y_1,\ Z_1,\ 1) \rightarrow (1,\ 0,\ 0,\ 0)\\
\mathbf{X}_2 &= (X_2,\ Y_2,\ Z_2,\ 1) \rightarrow (0,\ 1,\ 0,\ 0)\\
\mathbf{X}_3 &= (X_3,\ Y_3,\ Z_3,\ 1) \rightarrow (0,\ 0,\ 1,\ 0)\\
\mathbf{X}_4 &= (X_4,\ Y_4,\ Z_4,\ 1) \rightarrow (\boldsymbol{a},\ \boldsymbol{b},\ \boldsymbol{g},\ 0)\\
\mathbf{X}_5 &= (X_5,\ Y_5,\ Z_5,\ 1) \rightarrow (0,\ 0,\ 0,\ 1)\\
\mathbf{X}_6 &= (X_6,\ Y_6,\ Z_6,\ 1) \rightarrow (1,\ 1,\ 1,\ 1)
\end{aligned}
\quad . \tag{1}
$$

Thus, $\mathbf{X}_i$, i=1~3 and $\mathbf{X}_5$ $\mathbf{X}_6$, form a canonical basis. We can obtain a unique space collineation $\mathbf{A}_{4x4}$, $\det(\mathbf{A}_{4x4}) \neq 0$, which transforms the original five points into the canonical basis. The fourth point is transformed into its projective coordinates $(\alpha, \beta, \gamma, 0)^T$ by $\mathbf{A}_{4x4}$. For the projections of these six points onto an image, we take $\mathbf{x}_i$, i=1, ..., 4 as the canonical projective coordinates in the image plane space. Then we can obtain a unique plane collineation $\mathbf{A}_{3x3}$, $\det(\mathbf{A}_{3x3}) \neq 0$. And $\mathbf{A}_{3x3}$ transforms the fifth and sixth points into $(u_5, v_5, w_5)^T$ and $(u_6, v_6, w_6)^T$.

$$
\begin{aligned}
\mathbf{x}_1 &= (x_1,\ y_1,\ 1) \rightarrow (1,\ 0,\ 0)\\
\mathbf{x}_2 &= (x_2,\ y_2,\ 1) \rightarrow (0,\ 1,\ 0)\\
\mathbf{x}_3 &= (x_3,\ y_3,\ 1) \rightarrow (0,\ 0,\ 1)\\
\mathbf{x}_4 &= (x_4,\ y_4,\ 1) \rightarrow (1,\ 1,\ 1)\\
\mathbf{x}_5 &= (x_5,\ y_5,\ 1) \rightarrow (u_5,\ v_5,\ w_5)\\
\mathbf{x}_6 &= (x_6,\ y_6,\ 1) \rightarrow (u_6,\ v_6,\ w_6)
\end{aligned}
\tag{2}
$$

The relationship between the object points and the corresponding image points is

$$
\begin{bmatrix} 1\ 0\ 0\ 1 & u_5\ u_6 \\ 0\ 1\ 0\ 1 & v_5\ v_6 \\ 0\ 0\ 1\ 1 & w_5\ w_6 \end{bmatrix} = r_i \mathbf{T} \begin{bmatrix} 1\ 0\ 0\ \boldsymbol{a}\ 0\ 1 \\ 0\ 1\ 0\ \boldsymbol{b}\ 0\ 1 \\ 0\ 0\ 1\ \boldsymbol{g}\ 0\ 1 \\ 0\ 0\ 0\ 0\ 1\ 1 \end{bmatrix}, \tag{3}
$$

where

$$
\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix}
$$

The right hand side of Eq.(3) is arranged to

$$
\begin{bmatrix}
r_1 t_{11} & r_2 t_{12} & r_3 t_{13} & r_4(\boldsymbol{a} t_{11} + \boldsymbol{b} t_{12} + \boldsymbol{g} t_{13}) & r_5 t_{14} & r_6(t_{11} + t_{12} + t_{13} + t_{14}) \\
r_1 t_{21} & r_2 t_{22} & r_3 t_{23} & r_4(\boldsymbol{a} t_{21} + \boldsymbol{b} t_{22} + \boldsymbol{g} t_{23}) & r_5 t_{24} & r_6(t_{21} + t_{22} + t_{23} + t_{24}) \\
r_1 t_{31} & r_2 t_{32} & r_3 t_{33} & r_4(\boldsymbol{a} t_{31} + \boldsymbol{b} t_{32} + \boldsymbol{g} t_{33}) & r_5 t_{34} & r_6(t_{31} + t_{32} + t_{33} + t_{34})
\end{bmatrix} \tag{4}
$$

Therefore, from Eqs. (3) and (4), we can obtain each element of transformation matrix $\mathbf{T}$ as follows:

$$
\begin{aligned}
& t_{11} = 1/r_1,\ t_{22} = 1/r_2,\ t_{33} = 1/r_3,\ t_{12} = t_{13} = t_{21} = t_{23} = t_{31} = t_{32} = 0 \\
& t_{14} = \frac{u_5}{r_5},\ t_{24} = \frac{v_5}{r_5},\ t_{34} = \frac{w_5}{r_5},\ \text{and} \\
& \frac{1}{r_1} = \frac{\bar{a}}{r_4},\ \frac{1}{r_2} = \frac{\bar{b}}{r_4},\ \frac{1}{r_3} = \frac{\bar{g}}{r_4},\ \text{where } \bar{a} = \frac{1}{a},\ \bar{b} = \frac{1}{b},\ \bar{g} = \frac{1}{g}.
\end{aligned}
\tag{5}
$$

We can define the invariant relationship from the sixth column in Eq.(4) and the elements computed in Eq.(5),

$$
\begin{bmatrix} \bar{a} & u_5 & -u_6 \\ \bar{b} & v_5 & -v_6 \\ \bar{g} & w_5 & -w_6 \end{bmatrix} \begin{bmatrix} 1/r_4 \\ 1/r_5 \\ 1/r_6 \end{bmatrix} = 0 \tag{6}
$$

From the condition for a non-trivial solution for the equation, we obtain the relationship,

$$\begin{vmatrix} \bar{a} & u_5 & -u_6 \\ \bar{b} & v_5 & -v_6 \\ \bar{g} & w_5 & -w_6 \end{vmatrix} = -(\mathbf{V}_1 \times \mathbf{V}_2) \bullet \mathbf{V}_3 = 0 \tag{7}$$

$$\text{or} \quad -(V_{41}\bar{a} + V_{42}\bar{b} + V_{43}\bar{g}) = 0 \,,$$

where $\mathbf{V}_1 = (u_5, v_5, w_5)$, $\mathbf{V}_2 = (u_6, v_6, w_6)$, $\mathbf{V}_3 = (\bar{a}, \bar{b}, \bar{g})$, and $\mathbf{V}_4 = (V_{41}\ V_{42}\ V_{43}) = \mathbf{V}_1 \times \mathbf{V}_2$

If the fifth and the sixth points exchange each other, Eq.(3) becomes

$$\begin{bmatrix} 1\ 0\ 0\ 1 & u_5\, u_6 \\ 0\ 1\ 0\ 1 & v_5\, v_6 \\ 0\ 0\ 1\ 1 & w_5\, w_6 \end{bmatrix} = r_i \mathbf{T} \begin{bmatrix} 1\ 0\ 0\ a\ 1\ 0 \\ 0\ 1\ 0\ b\ 1\ 0 \\ 0\ 0\ 1\ g\ 1\ 0 \\ 0\ 0\ 0\ 0\ 1\ 1 \end{bmatrix} . \tag{8}$$

In a similar way, we can obtain the invariant relationship,

$$\begin{vmatrix} \bar{a} & u_6 & -u_5 \\ \bar{b} & v_6 & -v_5 \\ \bar{g} & w_6 & -w_5 \end{vmatrix} = (\mathbf{V}_2 \times \mathbf{V}_1) \bullet \mathbf{V}_3 = 0 \tag{9}$$

$$\text{or} \quad V_{41}\bar{a} + V_{42}\bar{b} + V_{43}\bar{g} = 0 \,.$$

Thus, we can discriminate the fifth and the sixth points by Eq.(7) and (9). From the invariant relationship defined by Eq.(7) or (9), $\mathbf{V}_3$, designed from the structured object points, is orthogonal to the cross product of $\mathbf{V}_1$ and $\mathbf{V}_2$ which are extracted from the image. Therefore, all the vectors on the plane orthogonal to $\mathbf{V}_3$ satisfy the above relationship.

If the sixth point $\mathbf{X}_6$ is on the plane constructed by ($\mathbf{X}_3$, $\mathbf{X}_4$, $\mathbf{X}_5$), the structure becomes the same one proposed by Zhu et al.[2,3]. We can easily derive the invariant for the structure by adding the invariant relation to the coplanar condition.

The coplanar condition becomes

$$\left| \mathbf{X}_3 \quad \mathbf{X}_4 \quad \mathbf{X}_5 \quad \mathbf{X}_6 \right| = \begin{vmatrix} 0 & a & 0 & 1 \\ 0 & b & 0 & 1 \\ 1 & g & 0 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} = 0 \ \text{ or } \ a = b. \tag{10}$$

By substituting Eq.(10) into Eq.(7), the invariant is represented as

$$\bar{a} = \bar{b} = \pm \frac{V_{43}}{\sqrt{(2-(V_{41}+V_{42})^2)}} \,, \quad \bar{g} = \pm \frac{(V_{42}-V_{41})}{\sqrt{(2-(V_{41}+V_{42})^2)}} \tag{11}$$

Thus, in this case we can obtain a unique invariant from Eq.(11), which is the same form as one proposed by Zhu et al.[2,3].

Also we consider the construction of the invariant by two-view proposed by A.Zisserman and S.J.Maybank [10]. For each view, the invariant relationship can be written from Eq.(7) as follows:

$$\begin{aligned} V^1_{41}\bar{a} + V^1_{42}\bar{b} + V^1_{43}\bar{g} &= 0 \\ V^2_{41}\bar{a} + V^2_{42}\bar{b} + V^2_{43}\bar{g} &= 0 \end{aligned} \tag{12}$$

From Eq.(12), we can obtain the invariant in terms of $\bar{a}, \bar{b}, \bar{g}$;

$$\bar{a} = \frac{(V^1_{42}V^2_{43} - V^1_{43}V^2_{42})}{(V^1_{41}V^2_{42} - V^1_{42}V^2_{41})}\bar{g} \,, \quad \bar{b} = \frac{(V^1_{41}V^2_{43} - V^1_{43}V^2_{41})}{(V^1_{41}V^2_{42} - V^1_{42}V^2_{41})}\bar{g} \,. \tag{13}$$

3. Recognition algorithm

3.1 Database structure for indexing

To use the invariant relationship obtained in the previous section for the recognition of three dimensional polyhedral objects, we must construct an efficient database or model-base.

Given the invariant for a set of points on a structured object $(\bar{a}, \bar{b}, \bar{g})^{\mathrm{T}}$, we must record the information about the structure, a model number and a plane number, and another two points. But, it is very inefficient to consider all positions on the plane. Thus, we consider a surface on an unit sphere as a structure of a model-base. Fig.2 shows the proposed model-base structure, where $(\bar{a}, \bar{b}, \bar{g})^{\mathrm{T}}$ is the normalized vector for the invariant for object points and the invariant circle(χ) represents the group of vectors that are orthogonal to $(\bar{a}, \bar{b}, \bar{g})^{\mathrm{T}}$.

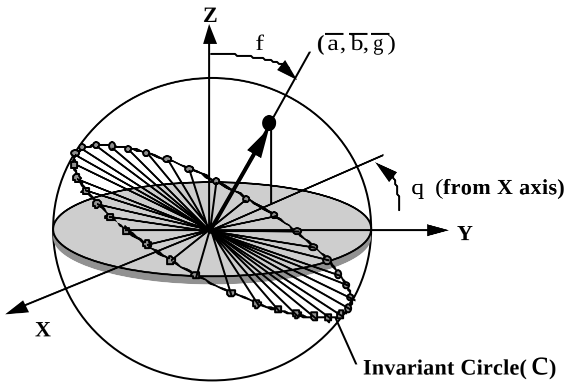


Fig.2 An unit sphere as a structure of a model-base

A vector in the model-base structure can be represented by two parameters (θ, φ) as follows:

$$(\bar{a}, \bar{b}, \bar{g}) = (\sin f \cos q,\ \sin f \sin q,\ \cos f) \tag{14}$$

From Eq.(14), we can convert the model-base space to (θ, φ)-space :

$$q = \tan^{-1}\!\left(\frac{\bar{b}}{\bar{a}}\right), \quad f = \cos^{-1}(\bar{g}) \,. \tag{15}$$

We can compute vectors on the invariant circle by a coordinate transformation : Z-axis of the new coordinate system is aligned with old Z-axis and X-axis is placed on the X-Y plane of the old coordinate system. We then obtain

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} (\cos\phi\cos\theta) & (-\sin\theta) & (\sin\phi\cos\theta) \\ (\cos\phi\ \sin\theta) & (\cos\theta) & (\sin\phi\sin\theta) \\ (-\sin\phi) & 0 & (\cos\phi) \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} . \quad (16)$$

Fig.3 shows the coordinates systems, where (X, Y, Z) is the old coordinate system and (X',Y',Z') is the new coordinate system.

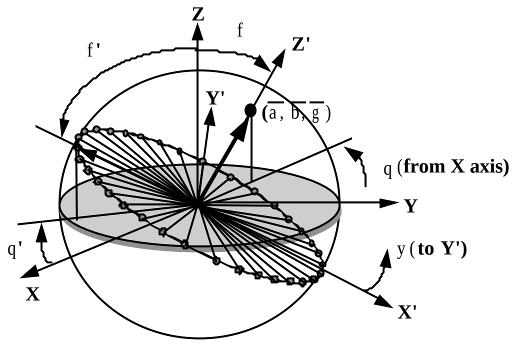


Fig.3 The coordinates system

Then vectors on the invariant circle are

$(X', Y', Z') = (\cos\varphi, \sin\varphi, 0)$, for $\varphi = 0 \sim 180$

or

$$\begin{aligned} X &= (\cos\phi\cos\theta)(\cos\varphi) - \sin\theta(\sin\varphi) \\ Y &= (\cos\phi\ \sin\theta)(\cos\varphi) + (\cos\theta)(\sin\varphi) \\ Z &= (-\sin\phi)(\cos\varphi) \end{aligned} \quad . \quad (17)$$

Here, we only consider $\varphi = 0 \sim 180$, because of the symmetric property of Eq.(11).

These vectors are represented in the (θ, ϕ)–space as

$$\theta' = \tan^{-1}\left(\frac{Y}{X}\right), \quad \phi' = \cos^{-1}(Z) \quad . \quad (18)$$

3.2 Geometric hashing

Geometric invariant provides an indexing function for an efficient model-based object recognition, in which the time complexity is rarely affected by the number of models. Geometric hashing idea has been introduced by Y.Lamdan etc.[14] as a method of the indexing-based object recognition. And its importance and efficiency has been emphasized in many recognition systems [15,16,18].

Pre-processing

For each model **M** and for every feasible basis **b** consisting of four coplanar points **(X_1 X_2, X_3, X_5)** and one non-coplanar point $\mathbf{X_6}$, we :

i) compute the canonical coordinates of points $\mathbf{X_4}=(\alpha, \beta, \gamma, 0)$, whose basis points are $\mathbf{X_1}$, $\mathbf{X_2}$, $\mathbf{X_3}$, $\mathbf{X_6}$;

ii) compute the orthogonal vectors and (θ', ϕ') from Eq.(18); and

iii) record a node (**M**, **b**) in the hash table entries indicated by (θ', ϕ').

Recognition

Given a scene with **n** point features extracted, we

i) choose a feasible set of five points consisting of four points as a basis of two-dimensional canonical frame and another one point;

ii) compute canonical coordinates (or vectors) of another one point and all the remaining points, (u_5, v_5, w_5) and (u_6, v_6, w_6);

iii) compute the cross product of the two vectors and use the resulting vector to index the hash table and hit all ($\mathbf{M_i}$, $\mathbf{b_j}$)'s that are stored in the entry retrieved;

iv) histogram all ($\mathbf{M_i}$, $\mathbf{b_j}$)'s with the number of hits received;

v) establish a hypothesis of the existence of an instance of model $\mathbf{M_i}$ in the scene if ($\mathbf{M_i}$, $\mathbf{b_j}$), for some j, peaks in the histogram with sufficiently many hits.

4. Experiments

4.1 Preliminary test using an 3-D polyhedral object

As a feasibility test for 3-D object recognition, we select a simple three dimensional object. Fig.4 shows the object and Table 1 represents the coordinates in Euclidean space.

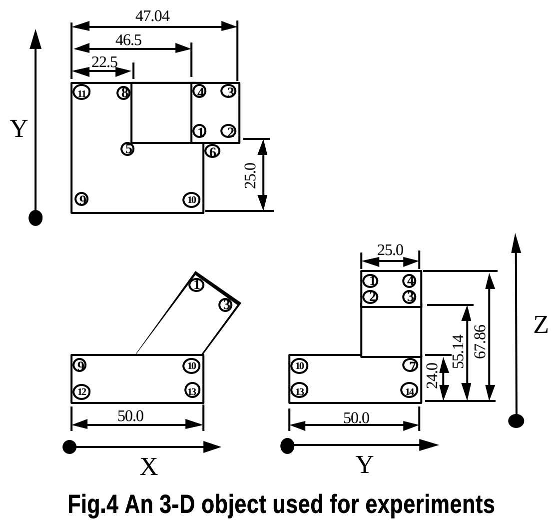
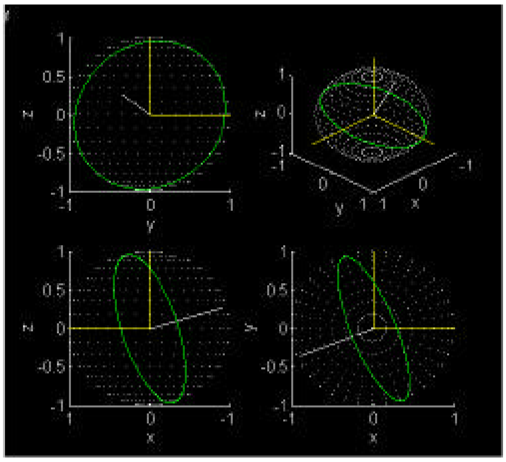


Fig.4 An 3-D object used for experiments

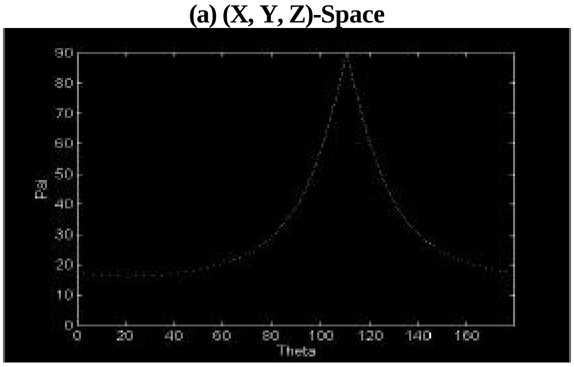
Table 1 The coordinates of object points(unit:mm)

No.	Coordinate (X, Y, Z)	No.	Coordinate (X, Y, Z)
1.	(46.50, 25.00, 67.86)	9.	(0.00, 0.00, 24.00)
2.	(67.04, 25.00, 55.14)	10.	(50.00, 0.00, 24.00)
3.	(67.04, 50.00, 55.14)	11.	(0.00, 50.00, 24.00)
4.	(46.50, 50.00, 67.86)	12.	(0.00, 0.00, 0.00)
5.	(22.50, 25.00, 24.00)	13.	(50.00, 0.00, 0.00)
6.	(50.00, 25.00, 24.00)	14.	(50.00, 50.00, 0.00)
7.	(50.00, 50.00, 24.00)	15.	(0.00, 50.00, 0.00)
8.	(22.50, 50.00, 24.00)		
No.	**Points on Plane**	**No.**	**Points on Plane**
1	1,2,3,4	5	9,11,15,12
2	1,4,8,5	6	2,6,7,3
3	1,2,5,6	7	10,13,14,7
4	9,10,12,13		

Fig.5(a) shows the (X,Y,Z)-space of the model-base constructed for the structure consisting of four coplanar points(1,2,3,4) and two non-coplanar points(9,12). Fig.5(b) shows (θ, φ)–space. For this structure, $(\bar{a}, \bar{b}, \bar{g})^T$ is (-0.8966, -0.3472, 0.2747) .



(a) (X, Y, Z)-Space



(b) (θ, φ)–space.

Fig.5 The model-base for a structure consisting of four coplanar points(1,2,3,4) and two non-coplanar points space(9,12) in (X,Y,Z) and (θ, φ) space, respectively.

Fig.6 shows seven images of the object from different viewing directions.

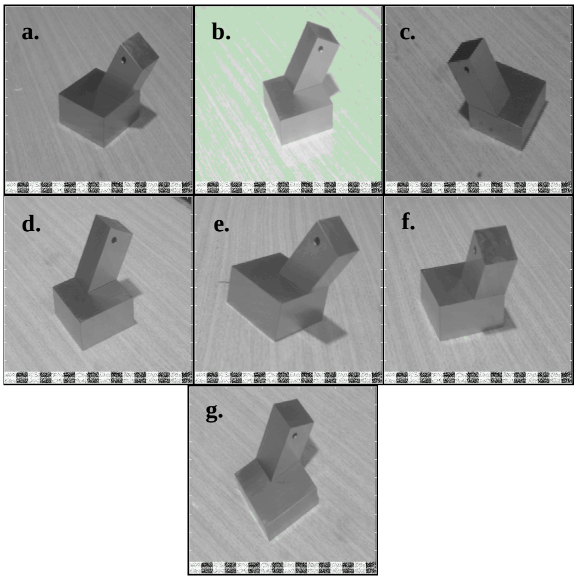


Fig.6 Seven images of the object from different view.

Table 2 represents the cross products of two canonical coordinates (or vectors) computed in each image, and the dot product between the cross product vector and $(\bar{a}, \bar{b}, \bar{g})^T$, which is computed in advance by using the stored 3-D coordinate values of the object. In this table, error denotes the angle difference between the computed and the true $(\bar{a}, \bar{b}, \bar{g})^T$.

Fig.7 shows indexing by the invariant vector computed by the corresponding points on each image.

Table 2. Extracted indexing vector

Known $\mathbf{V}_3 = (\bar{a}, \bar{b}, \bar{g})$ =(-0.8966, -0.3472, 0.2747)				
	Computed $\mathbf{V}_4 (= \mathbf{V}_1 \times \mathbf{V}_2)$	Index Values (q', f')	$\cos^{-1}(\mathbf{V}_3 \bullet \mathbf{V}_4)$ deg.	Error deg.
a	-0.42, 0.57,-0.70	126.66,45.25	90.65	0.65
b	-0.43, 0.67,-0.60	122.87,53.45	90.50	0.50
c	-0.43, 0.60,-0.67	125.55,47.68	90.48	0.48
d	-0.45, 0.71,-0.53	122.55,57.70	89.23	0.77
e	-0.43, 0.56,-0.71	127.76,44.80	90.09	0.09
f	-0.44, 0.49,-0.75	131.50,41.06	89.29	0.71
g	-0.45, 0.68,-0.58	123.67,54.64	89.39	0.61

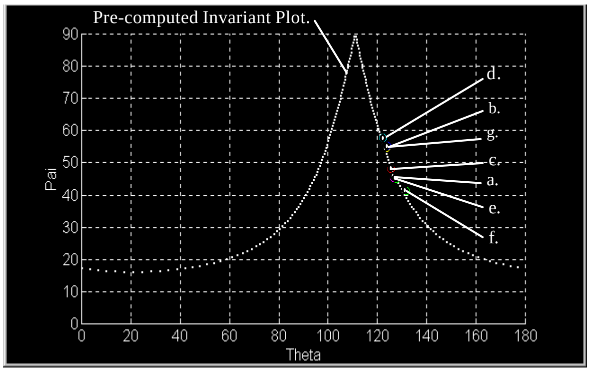


Fig.7 Indexing by the invariant vector

If we have two views, it is possible to compute a unique invariant vector from Eq.(9). Table 3 represents the examples of the invariant computed from any two-view.

Table 3. Invariant computed by any two-view

Known $\mathbf{V}_3 = (\bar{a}, \bar{b}, \bar{g})$ =(-0.90, -0.35,0.27)		
Images	computed $\tilde{\mathbf{V}}_3 = (\tilde{\bar{a}}, \tilde{\bar{b}}, \tilde{\bar{g}}) = \mathbf{V}_4^1 \times \mathbf{V}_4^2$	error $\cos^{-1}(\mathbf{V}_3 \bullet \tilde{\mathbf{V}}_3)$
a. and **b.**	-0.90,-0.35,0.26	1.4421 deg.
a. and **c.**	-0.89,-0.39,0.22	4.0469 deg.
b. and **c.**	-0.90,-0.34,0.27	0.8156 deg.
b. and **e.**	-0.90,-0.31,0.30	2.6951 deg.
d. and **e.**	-0.89,-0.39,0.23	3.6934 deg.
d. and **g.**	-0.89,-0.35,0.29	1.0397 deg.

For the structure, (3, 4, 5, 6) are coplanar points, we can extract the invariant from Eq.(10). Table 4,5 represent the known and computed invariant for the structures.

Table 4. Invariant for the structure consisting of points, (1,4,8,5) and (9,10)

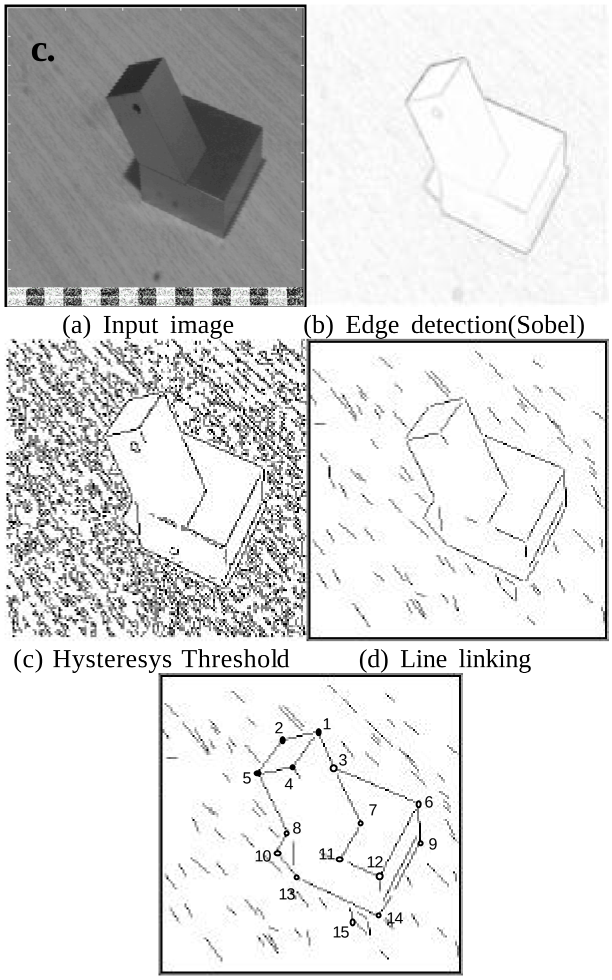
Known $\mathbf{V}_3 = (\bar{a}, \bar{b}, \bar{g}) = (0.6667\ \ 0.6667\ \ 0.3333)$		
Image	Computed $\tilde{\mathbf{V}}_3 = (\tilde{\bar{a}}, \tilde{\bar{b}}, \tilde{\bar{g}})$	Error $\cos^{-1}(\mathbf{V}_3 \bullet \tilde{\mathbf{V}}_3)$
b.	(0.6703, 0.6703,0.3183)	0.9339
d.	(0.6802, 0.6802,-0.2733)	3.5581

Table 5. Invariant for the structure consisting of points, (5,11,9,10) and (12,13)

Known $\mathbf{V}_3 = (\bar{a}, \bar{b}, \bar{g}) = (0.0990\ \ 0.0990\ \ 0.9901)$		
Image	Computed $\tilde{\mathbf{V}}_3 = (\tilde{\bar{a}}, \tilde{\bar{b}}, \tilde{\bar{g}})$	Error $\cos^{-1}(\mathbf{V}_3 \bullet \tilde{\mathbf{V}}_3)$
a.	(0.0598, 0.0598, 0.9964)	3.2658 deg.
b.	(0.0518, 0.0518, 0.9973)	3.9010 deg.

4.2 Preprocessing and Hypotheses Generation

To reduce time complexity of hypotheses generation, we search for corner points as well as closed polygons in preprocessing. We use an algorithm proposed by A.Etemadi[19] to extract corners and polygons.



(a) Input image (b) Edge detection(Sobel)

(c) Hysteresys Threshold (d) Line linking

(e) Corner and polygon detection

(●: Polygon, ○: Corner)

Fig.8 Preprocessing

In Fig.8, if we select point features 1, 2, 5, 4 and 7 as a feasible set, it is the structure proposed by Rothwell[1], consisting of three adjacent planes, (1,2,5,4), (4,5,8,11) and (1,4,11,7). Also the structure proposed by Zhu[2,3] can be constructed by two adjacent planes, (1,2,5,4) and (1,4,11,7). Unfortunately, they do not provide sufficient invariants for object recognition. For this particular scene, however, our proposed invariant can be defined up to nine different structures, which can be used to generate many hypotheses for object recognition.

We compute invariants for points set consisting of the (1,2,5,4) and 7, and 3, 6, 8, 9, ..., 15. And we vote the information in the model-base indexed by these invariants, which include information for the plane number and another one point

Then, hypotheses are generated if the voted number is greater than a predefined threshold. Table 6 represents ten generated hypotheses for scene features 1,2,5,4, and 7. The plane means the number of plane defined in Table 1. And the point represents the point stored in model-base as a basis, which is explained in section 3.2.

Table 6. The result of hypotheses generation.

	Plane	Point	Vote		Plane	Point	Vote
1^{st}	**1**	**5**	**9**	6^{th}	5	1	7
2^{nd}	1	9	8	7^{th}	5	2	8
3^{rd}	1	13	8	8^{th}	5	3	7
4^{th}	3	12	8	9^{th}	6	1	6
5^{th}	3	13	8	10^{th}	6	2	7

4.3 Verification and Registration

For each generated hypothesis, we compute a transformation between the image and the model, and project the model onto the image plane. Then, we count points within an error bound, i.e. matching points. We select a hypothesis with a maximum number of matching points.

Fig.10 shows the results of transformation for the 1^{st} and 2^{nd} hypotheses. The stars(*) represent detected corner points and the circles(O) represent the transformed model corners. Table 7 shows the number of matching points obtained by verification.

From the result of verification, the first hypothesis is selected as the true hypothesis with 12 matching points.

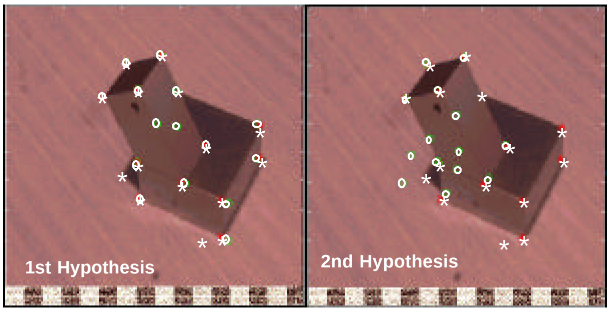


Fig.10 The result of verification for 1^{st} and 2^{nd} hypotheses

Table 7. The result of Verification.

	# of Matching Points		# of Matching Points
1^{st}	**13**	6^{th}	6
2^{nd}	8	7^{th}	7
3^{rd}	5	8^{th}	5
4^{th}	4	9^{th}	4
5^{th}	5	10^{th}	5

Fig.11 shows a registration of the three dimensional model overlaid onto the third image.

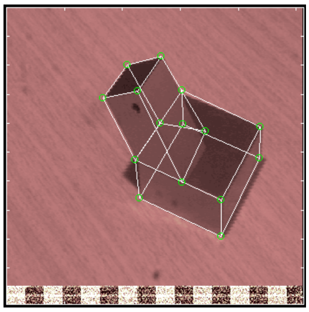
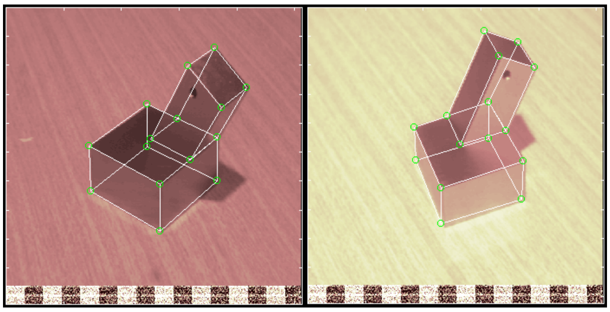
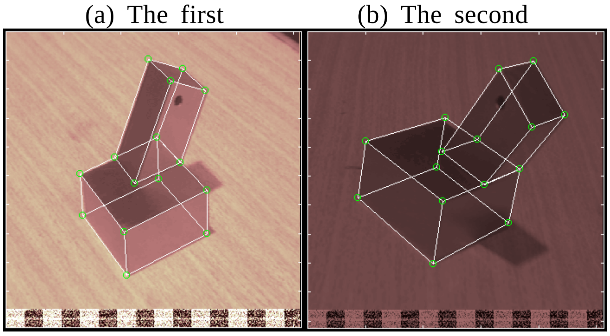


Fig.11 The registration of 3-D object onto the image

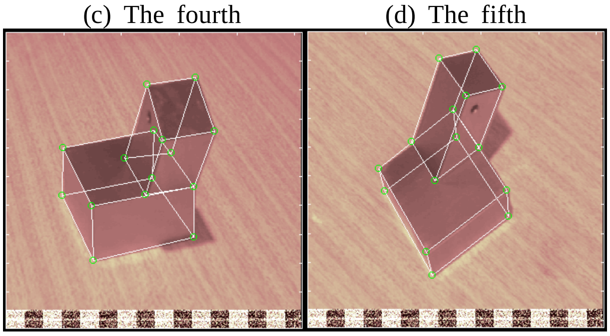
Fig.12 shows the reconstruction of the three dimensional model onto the other six images after recognition. Fig.13 shows the result of recognition and registration for occlusion cases.



(a) The first (b) The second



(c) The fourth (d) The fifth



(e) The sixth (f) The seventh

Fig.12 The result of recognition for the other images

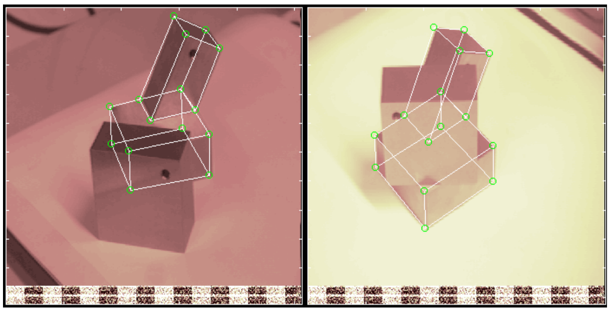


Fig.13 The result of recognition for the occlusion case

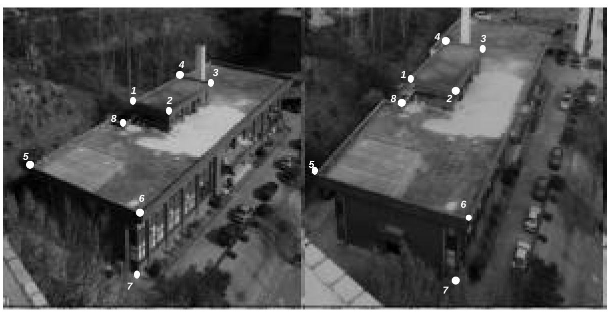
4.4 Experiments using an outdoor building scene

We can construct the normal vector $\mathbf{V}_3$ by two vectors, $\mathbf{V}_1$**,** $\mathbf{V}_2$ which are computed from two views of an 3-D object (Eq.(7)). Then, we use the vector to recognize the 3-D object in other views.

Fig.14 shows two images for a building scene, which are used to compute the normal vector. Fig.16 shows a hash table constructed by $\mathbf{V}_3$ computed from the two views. Two circles represents $\mathbf{V}_1$**,** $\mathbf{V}_2$ and the dotted curves represent the normal vector of $\mathbf{V}_1$x$\mathbf{V}_2$.

Fig.14 Any two views of a building for constructing the model-base

Fig.15 shows four input images captured from very different viewing directions. Indexing invariant computed in these images are represented as stars in Fig.16. We can observe that the invariant relationship of six points satisfying our proposed invariant structure well corresponds to the predicted invariant relationship, which is represented by a dotted curve in Fig.16.



(a) Input image 1. (b) Input image 2.

(c) Input image 3. (d) Input image 4.

Fig.15 Four input images

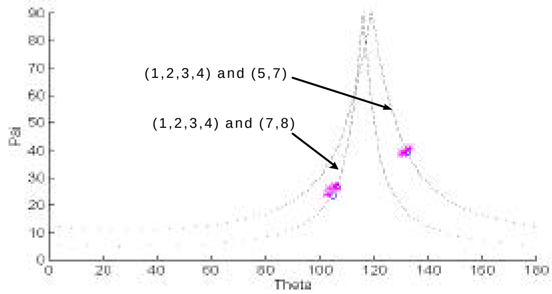


Fig.16 Model-base constructed by two views and indexing of the model-base by four input images

In order to compare the proposed invariant with the well-known five-point plane invariant, we compute five point invariants using the same point features. Fig.17 shows the canonical coordinates of points when we use 1,2,3, and 4 as a basis. It shows that the canonical coordinates or five-point invariants are distributed over much larger areas than our six-points invariant. In other words, five-point invariants vary too much to be used for object recognition in this particular scene.

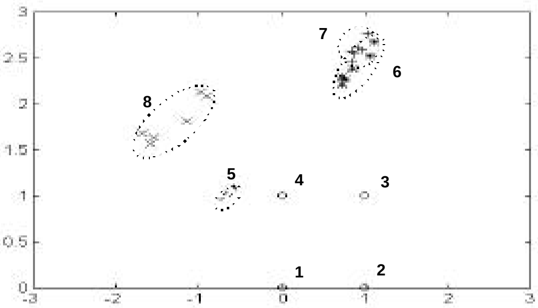


Fig.17. The canonical coordinates of the other points when the basis points are 1,2,3, and 4.

5. Conclusion

In this paper, we proposed a new 3-D invariant relationship of a special structure consisting of four coplanar points and any two non-coplanar points using only single-view. For some structures, Zisserman and Maybank [10] showed that the invariant can be constructed by two-view without computing the epipolar geometry. However, we derived an invariant relationship by one-view, which is represented as a form of plane equation. Based on this plane equation, we proposed a method for combining the relationship with geometric hashing concept for recognizing three-dimensional objects.

We showed that the invariant for the structure proposed by Zhu et al.[2,3] can be easily derived from the invariant relationship. With two-view for the structure, we can also derive the invariant from the relationship.

Since the structure is more general than the previously proposed structures, a hashing based method was feasible for 3-D object recognition.

Experiments using real scenes demonstrate that the proposed invariant relationship can be further extended to a real 3-D object recognition.

References

[1] Rothwell, C.A., Forsyth, D.A., Zisserman, A. and Mundy, J.L. "Extracting Projective Structure from Single Perspective Views of 3D Point Sets," *Proceedings of the 4th ICCV*, Berlin, Germany, pp.573-582, May, 1993.

[2] Y.Zhu, L.D.Seneviratne and S.W.E.Earles, "A New Structure of Invariant for 3D Point Sets from a Single View," *Proceedings 12th ICRA*, Nagoya, Japan, pp.1726-1731, 1995.

[3] Y.Zhu, L.D.Seneviratne and S.W.E.Earles, "Three Dimensional Object Recognition Using Invariants," *Proceedings 12th IROS*, Pittsburgh, USA, pp.354-359, 1995.

[4] D.Forsyth, J.L.Mundy, A.Zisserman, C.Coelho, C.Rothwell, "Invariant Descriptors for 3-D Object Recognition and Pose," *IEEE Transaction on PAMI*, Vol.13, No.10, pp.971-991, Octorber, 1991.

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

[6] M.H.Brill and A.B.Barrett, "Closed-Form Extension of the An-harmonic Ratio to N-Space," *Computer Vision Graphics Image Process*, Vol.23, pp92- 98, 1983.

[7] Burns, J.B., Weiss, R.S. and Riseman, E.M. "The Non-existence of General-case View-Invariants," *Geometric Invariance in Computer Vision*, Mundy, J.L. and Zisserman, A. editors, MIT Press, Cambridge, Massachusetts, USA, 1992.

[8] O.Fauseras, "What can be seen in three dimensions with an uncalibrated stereo rig?," *Proceedings of the 2nd ECCV*, Santa Margherita, Italy, G.Sandini, Ed., Springer-Verlag, pp.563-578, May, 1992.

[9] E.B.Barrett, G.Gheen, and P.Payton, "Representation of Three-Dimensional Object Structure as Cross-Ratios of Determinants of Stereo Image Points," *Applications of Invariance in Computer Vision*, edited by J.L.Mundy, A.Zisserman, and D.Forsyth, Springer-Verlag, Belin, pp.47 -68,1993.

[10] Zisserman, A. and Maybank, S.J. "A Case Against Epipolar Geometry," *Applications of Invariance in Computer Vision*, edited by J.L.Mundy, A.Zisserman, and D.Forsyth, Springer-Verlag, Belin, pp.69 -88, 1993.

[11] L.Quan, "Invariants of Six Points from 3 Uncalibrated Images," *Proceedings of the 4th ECCV*, Stocholm, Sweden, pp.459-470, 1994.

[12] L.Quan, "Invariants of Six Points and Projective Reconstruction From Three Uncalibated Images," *IEEE Transaction on PAMI*, Vol.17, No.1, pp.34-46, January, 1995.

[13] J.G.Semple and G.T.Kneebone, *Algebraic Projective Geometry*, Oxford Sceince Publication, 1952.

[14] Y.Lamdan, J.T.Schwartz, and H.J.Wolfson, "Object Recognition by Affine Invariant Matching," *In Proceedings of CVPR*, pp.335-344, 1988.

[15] H.J.Wolfson, "Model-Based Object Recognition by Geometric Hashing," *Proceedings of the 2nd ECCV*, Santa Margherita, Italy, G.Sandini, Ed., Springer-Verlag, pp.526-536, May, 1992.

[16] F.C.Tasi, "A Probability Approach to Geometric Hashing Using Line Features," Ph.D Thesis, Computer Science Department, New York University, 1993.

[17] E.B.Barrett and P.M.Payton, "General Methods for Determining Projective Invariants in Imagery," *CVGIP: Image Understanding*, Vol.53, No.1, January, pp.46-65, 1991.

[18] F.Stein and G.Medioni, "Structural Indexing: Efficient Three Dimensional Object Recognition," *IEEE Transaction on PAMI*, Vol.14, No.2, pp.125-146, February, 1992.

[19] A.Etemadi, "Low-level Grouping of Straight Line Segments," U.S.G. Tech. Report., United Kingdom.