MVA'94 IAPR Workshop on Machine Vision Applications Dec. 13-15, 1994, Kawasaki

Directional Voting with Geometric Hashing for Three Dimensional Object Recognition

S.Kaneko, M.Shibata and T.Honda
Tokyo University of Agriculture and Technology

2-24-16, Naka-cho, Koganei
Tokyo 184, JAPAN

ABSTRACT

A new method for recognition of three dimensional objects by use of a monocular image is proposed. It utilizes sample monocular images of objects as aspect models and a geometrical transformation and a directional voting procedure are essential for robust recognition against feature defects such as aspect changes or occlusion. With the method, the sensitivity of recognition of the correct model can be kept high while the amount of models (a dictionary) is reduced. Experimental results with real objects show an effectiveness of the proposed method.

1 INTRODUCTION

Model-based approaches to three dimensional object recognition with respect to shape and posture are expected to be effective in the case that information of objects can be available at learning or turning[1][2][3]. Three dimensional objects change their aspects according to viewing directions and their qualitative changes should be handled systematically[4][5]. Furthermore, extraction procedures of salient features and recognition algorithms which are robust against lack of a part of object features, existence of noise or occlusions are very important for the success of the model-based approaches and strongly requested for real applications.

We propose a new method for three dimensional object recognition with a monocular image and some model images which is based on the geometric hashing technique [6] and a directional voting procedure onto a feature space structured as a hash table. Our method has the following merits:

(1) no needs of range data at both learning and recognition phases,
(2) needs of only monocular image without any correspondence procedure,
(3) reduction of registered models,
(4) a small feature space (hash table) of reasonable size.

2 GEOMETRIC HASHING

The geometric hashing is divided into the two phases: learning and recognition[4]. The details of the extended algorithm are described in section 4. In the learning phase, sample monocular images of objects are registered into the hash table (HT) as feature point sets detected by a SLES algorithm[11]. In this step, all possibilities of geometrical transformations (rotation, translation, affine distortion, scaling, etc.) can be considered by selecting all the combinations of triplets of basis points from the set of feature points and then the HT can include normalized information of feature distribution on the image plane. In the recognition phase, a set of feature points from the image of an object is transformed with respect to a triplet of basis points and then according to coordinates of transformed points the model which gets the maximum voting score is selected as the corresponding model.

To apply the geometric hashing technique to three dimensional (3D) objects, 3D information processing has been necessary to construct a HT, which involves a method based on 3D shape data (CAD data) and quadruplets of basis points[4] or another method based on two (stereoscopic) images and a corresponding procedure [12].

Our method utilize only monocular images and triplets of three basis points in both phases. In learning phase, the same HT as the Lamdan's simple one is constructed with sample images. In recognition phase, monocular features are utilized in two passes: the uniform voting and the directional voting. The latter procedure enables to recognize 3D objects with monocular images which might be assumed "aspects" and to reduce an amount of aspects necessary to handle changes of appearances. Because the dimension of basis (number of basis points) influences an order of the size of the HT, a small basis is preferred in implementation. We utilize a triplet of basis points in order to keep the HT size reasonable.

3 TRANSFORMED POINTS

Since model-based recognition systems have finite models to be matched to objects, in a matching procedure solving differences between models and an object is one of the principal issue. We can concentrate our focus into distortions of transformed points on the HT according to changes of object postures and locations by three dimensional transformations: translation and rotation with respect to the x, y and z axes respectively and in both case of the orthogonal projection and the perspective projection. Distortion of feature points in the image plane are mapped onto the HT as transformed points with a triplet of basis points. We denote a triplet of basis points as $\{P_a, P_b, P_c\}$ (P_a is the origin) and their 3D coordinates and ones of other points as:

$$\boldsymbol{P}_i=\left(X_i,Y_i,Z_i\right)^T \quad i=a,b,c,n, \tag{1}$$

where n is the point index. A projected image point is described as follows:

$$\boldsymbol{p}_i=\left(x_i,y_i\right)^T \quad i=a,b,c,n. \tag{2}$$

In the image plane, the basis three points are assumed not

to be collinear each other. Then a transformed point has coordinates (α_n, β_n) as

$$\begin{pmatrix}\alpha_n\\ \beta_n\end{pmatrix}=\frac{1}{\left|p_b-p_a\ p_c-p_a\right|}\begin{pmatrix}\left|p_n-p_a\ p_c-p_a\right|\\ \left|p_b-p_a\ p_n-p_a\right|\end{pmatrix}. \quad (3)$$

We define a distortion vector as a difference vector $(\alpha_n'-\alpha_n,\ \beta_n'-\beta_n)^T$ between a registered point and the corresponding transformed point on the HT.

We consider 3D translations and rotations individually. The x, y and z axes denote the horizontal axis, the vertical axis on the image plane and then the normal axis to the plane, respectively.

3.1 CASE 1:ORTHOGONAL PROJECTION

The perspective projection is well approximated to an orthogonal projection, when the distance from objects to the camera is much greater than the extent of the object. In this case, a transformed point is obtained as follows:

$$\begin{pmatrix}\alpha_n\\ \beta_n\end{pmatrix}=\frac{1}{f(a,b,c)}\begin{pmatrix}f(a,n,c)\\ f(a,b,n)\end{pmatrix}, \quad (4)$$

where

$$f(i,j,k)\equiv\begin{vmatrix}X_i & X_j & X_k\\ Y_i & Y_j & Y_k\\ 1 & 1 & 1\end{vmatrix}.$$

[1] Translation along the z axis: Image points and so transformed points are invariant under this translation. This means the scale invariance of 2D affine transformation.

[2] Translation along the x or y axis: Since these translations are described by 2D affine transformations, transformed points are invariant (location invariance).

[3] Rotation about the z axis: Since this rotation is also described by a 2D affine transformation, transformed points are invariant (rotation invariance).

[4] Rotation about the x or y axis: We can consider the rotation angle θ_y about the y axis without any loss of generality. After the rotation, a transformed point is obtained as follows:

$$\begin{pmatrix}\alpha_n'\\ \beta_n'\end{pmatrix}=\frac{1}{g(a,b,c)}\begin{pmatrix}g(a,n,c)\\ g(a,b,n)\end{pmatrix}, \quad (5)$$

where

$$g(i,j,k)\equiv\begin{vmatrix}X_i & X_j & X_k\\ Y_i & Y_j & Y_k\\ 1 & 1 & 1\end{vmatrix}\cos\theta_y-\begin{vmatrix}1 & 1 & 1\\ Y_i & Y_j & Y_k\\ Z_i & Z_j & Z_k\end{vmatrix}\sin\theta_y.$$

Then the distortion on the HT of transformed points, $\alpha_n'-\alpha_n$ and $\beta_n'-\beta_n$, are not invariant and given by

$$\alpha_n'-\alpha_n=\frac{(Y_a-Y_c)\sin\theta_y\left(\left|P_a\,P_b\,P_c\right|-\left|P_n\,P_b\,P_c\right|-\left|P_a\,P_n\,P_c\right|-\left|P_a\,P_b\,P_n\right|\right)}{f(a,b,c)\cdot g(a,b,c)}, \quad (6)$$

$$\beta_n'-\beta_n=\frac{(Y_b-Y_a)\sin\theta_y\left(\left|P_a\,P_b\,P_c\right|-\left|P_n\,P_b\,P_c\right|-\left|P_a\,P_n\,P_c\right|-\left|P_a\,P_b\,P_n\right|\right)}{f(a,b,c)\cdot g(a,b,c)}, \quad (7)$$

respectively. The direction of distortion on the HT: S is represented by

$$S\equiv\frac{\beta_n'-\beta_n}{\alpha_n'-\alpha_n}=\frac{Y_b-Y_a}{Y_a-Y_c}. \quad (8)$$

This equation gives us the fact that any direction of distortion is independent of the coordinates of the transformed points, that is, *once a triplet of basis points is selected, all of the distortion vectors are parallel to each other.*

3.2 CASE 2:PERSPECTIVE PROJECTION

When the distance from objects to the camera is rather small or objects are not so small, we must deal with the perspective projection as the model of images. We denote the imaging distance as D.

[1] Translation along the z axis: Directions of distortion are not parallel to each other, however, we can expect that the norm of a distortion vector on the HT: r_n will be very small.

$$r_n=\sqrt{\left(\alpha_n'-\alpha_n\right)^2+\left(\beta_n'-\beta_n\right)^2} \quad (9)$$

[2] Translation along the x or y axis: When the object is translated along the x axis, a direction of distortion on the HT is represented by

$$\frac{\beta_n'-\beta_n}{\alpha_n'-\alpha_n}=\frac{\left(D-Z_c\right)\left(\left(D-Z_a\right)Y_b-\left(D-Z_b\right)Y_a\right)}{\left(D-Z_b\right)\left(\left(D-Z_c\right)Y_a-\left(D-Z_a\right)Y_c\right)}. \quad (10)$$

Therefore, distortion vectors are parallel to each other.

[3] Rotation about the z axis: This rotation is described by a 2D affine transformation. Then transformed points are stationary or invariant.

[4] Rotation about the x or y axis: We can denote the origin of a basis as $P_a = (0, 0, 0)^T$ and a rotation is done about the y axis without any loss of generality. Distortion on the HT is given by

$$\alpha_n'-\alpha_n=\frac{\left(D+X_b\sin\theta_y-Z_b\cos\theta_y\right)\left(\begin{vmatrix}X_n & X_c\\ Y_n & Y_c\end{vmatrix}\cos\theta_y-\begin{vmatrix}Y_n & Y_c\\ Z_n & Z_c\end{vmatrix}\sin\theta_y\right)-\left(D-Z_b\right)\begin{vmatrix}X_n & X_c\\ Y_n & Y_c\end{vmatrix}}{\left(D+X_n\sin\theta_y-Z_n\cos\theta_y\right)\left(\begin{vmatrix}X_b & X_c\\ Y_b & Y_c\end{vmatrix}\cos\theta_y-\begin{vmatrix}Y_b & Y_c\\ Z_b & Z_c\end{vmatrix}\sin\theta_y\right)-\left(D-Z_n\right)\begin{vmatrix}X_b & X_c\\ Y_b & Y_c\end{vmatrix}}, \quad (11)$$

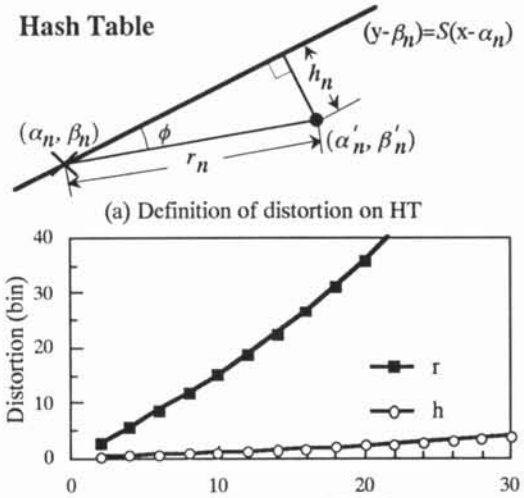
$$\beta_n'-\beta_n=\frac{\left(D+X_c\sin\theta_y-Z_c\cos\theta_y\right)\left(\begin{vmatrix}X_b & X_n\\ Y_b & Y_n\end{vmatrix}\cos\theta_y-\begin{vmatrix}Y_b & Y_n\\ Z_b & Z_n\end{vmatrix}\sin\theta_y\right)-\left(D-Z_c\right)\begin{vmatrix}X_b & X_n\\ Y_b & Y_n\end{vmatrix}}{\left(D+X_n\sin\theta_y-Z_n\cos\theta_y\right)\left(\begin{vmatrix}X_b & X_c\\ Y_b & Y_c\end{vmatrix}\cos\theta_y-\begin{vmatrix}Y_b & Y_c\\ Z_b & Z_c\end{vmatrix}\sin\theta_y\right)-\left(D-Z_n\right)\begin{vmatrix}X_b & X_c\\ Y_b & Y_c\end{vmatrix}}. \quad (12)$$

We define the values: r_n and h_n, as illustrated in Fig.1(a). The r_n is the norm of the distortion vector on the HT shown in (9), and h_n is the distance between a transformed point and a line which has the direction S shown in (8) and is represented by

$$h_n=\frac{\left|S\left(\alpha_n'-\alpha_n\right)-\left(\beta_n'-\beta_n\right)\right|}{\sqrt{S^2+1}}. \quad (13)$$

In the case of the orthogonal projection, (α'_n, β'_n) should be located on the line, but in the case of the perspective projection, h_n is normally not zero. We can characterize the behavior of distortion by the value of h_n. Fig.1(b)

shows a numerical example of it, where a magnitude distortion is represented in terms of "bin" of the HT, and from the figure since h is rather small it is shown that the direction of distortion might be estimated.



(a) Definition of distortion on HT

(b) Numerical example of distortion

Fig.1 Distortion on HT.

4 DIRECTIONAL VOTING

An efficient algorithm for visual recognition is proposed according to the analysis in the previous section. The basic approach is to restrict votes onto better possibilities on the HT by utilizing the characteristics that directions of the distortion of transformed points on the HT might be assumed to be similar to each other in some range of changes in posture and location. The most important step is the estimation of the direction which depends on a particular triplet of basis points.

Fig.2 shows the basic principle of our technique. A uniform voting is performed and then several candidates of model bases whose scores are rather high. A distortion vector is estimated for each candidate and the directional voting is performed along an estimated direction. Finally we evaluate scores of models and select the model of the sufficient high score. A distortion vector is simply estimated by the largest distortion for canceling quantization error on the HT. Fig.3 shows a schematic diagram of the proposed algorithm.

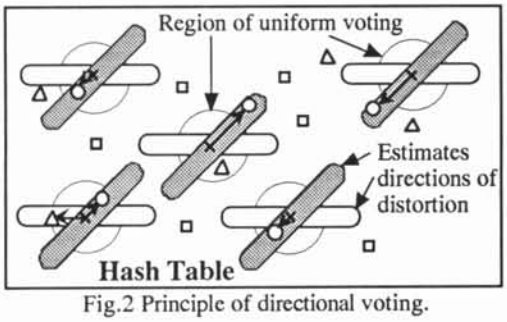


Fig.2 Principle of directional voting.

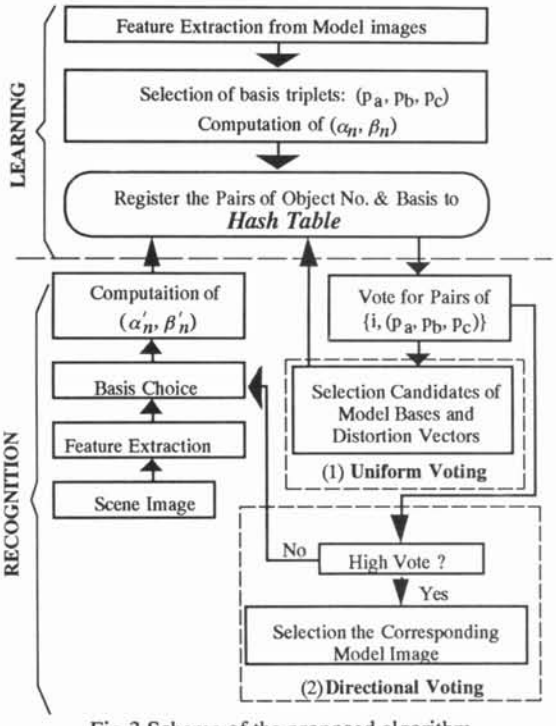
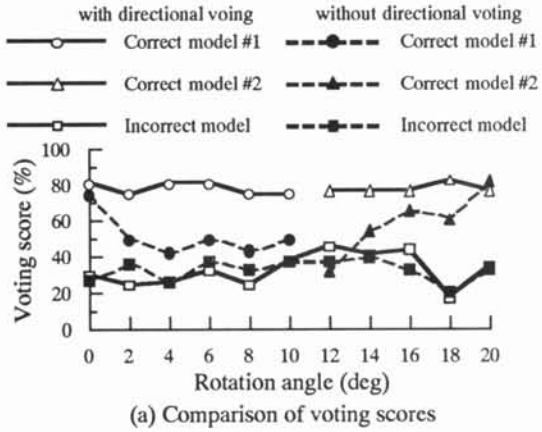


Fig.3 Scheme of the proposed algorithm

5 EXPERIMENTAL RESULTS

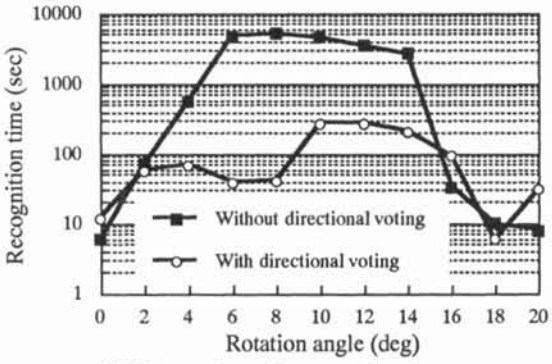
Twenty model images of 4 postures for 5 objects were used in experiments. The imaging distance is about 900 mm and the threshold on the voting score is v=75%. The stop condition is that the difference between the highest score and the second highest one is more than 30%. The HT is quantized into quadruled squares of side of 1 bin which is a unit of quantization on HT and set to 0.02mm width.

Fig.4 shows results for occluded polyhedra. Fig.4(a) shows the differences in voting scores between the conventional uniform voting[5][10] and the proposed directional voting. The model with the angle 0° and 20° both of which have high scores. As several incorrect models are usually selected, the highest one of them is



(a) Comparison of voting scores

Fig.4 Experimental results (occluded polyhedra).

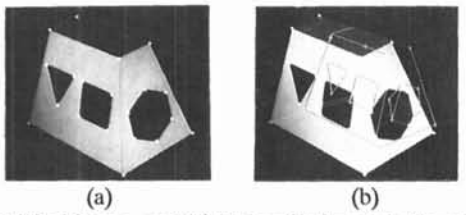


(b) Comparison of computation times

Fig.4 Experimental results (continued).

illustrated with ■ and □. In uniform voting, the larger the object is rotated, the lower the score for the correct model is, while with directional voting, the correct score keeps high over the wide range of distortion (rotation angle). Fig.4(b) is results of comparison of computation times. As may be seen in the figure, the directional voting is several or a few hundred times faster than the uniform one especially in marginal ranges of models because of a small number of re-tries waiting for a sufficient score.

Fig.5 shows results of recognition of the object which is translated along the z axis in order to investigate the case of 3.2 [1] in the perspective projection. Fig.5(a) is the original scene that the object is located at 300mm nearer to the camera than the corresponding model. Fig.5(b) is the result of recognition, which includes the original object and the recognized correct model overlapped.



(a) (b)

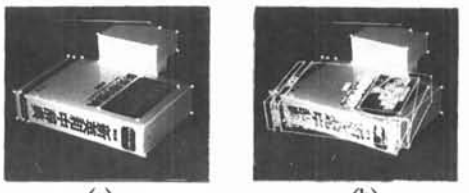
(a) Original image, and (b) Recognized model. The object is at the distance 300mm apart from the model overlapped.

Fig.5 Result of recognition (translation along the z-axis)

Fig.6 represent results of recognition of the object which is rotated about the y axis in the case of 3.2 [4]. Then, we set the voting width corresponding to h to $\pm$ 1bin and the voting length corresponding to r to $\pm$ 20bin. Fig.6(c) shows a tolerance of recognition, that is, the relation between voting scores and the magnitude of distortion which can be assumed to be proportional to rotation angle. The object is rotated by every 2° apart from the corresponding model and is recognized correctly until the posture of 16° rotation. It is assumed that a small set of model views is only required to implement the recognition system with the proposed algorithm.

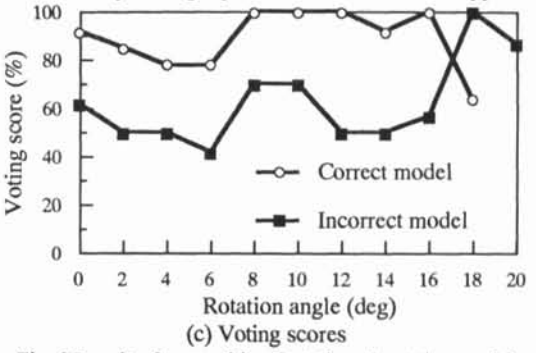
6 CONCLUSIONS

The robust method for recognition of three dimensional



(a) (b)

(a) Original image, and (b) Recognized model .The object is rotated by 16 deg. apart from the model overlapped.



(c) Voting scores

Fig.6 Result of recognition (rotation about the y axis)

objects has been proposed. The method uses only monocular images both in the learning phase and the recognition phase. With the proposed method, the total amount of aspects or views can be reduced into a small set and the high reliability can be obtained.

REFERENCES

[1]W.A.Perkins:"A Model-based Vision System for Industrial Parts", IEEE Trans., Vol.C-27, pp.126-143 (1978).

[2]W.E.L.Grimson and T.Lozano-Perez:"Localizing Overlapping Parts by Searching the Interpretation Tree", IEEE Trans., Vol.PAMI-9, No.4, pp.469-482 (1987).

[3]K.Ikeuchi and K.Koshikawa:"Determining Object Attitude and Position for Bin-Picking Tasks Guided by an Interpretation Tree Derived from a Geometrical Modeler", Trans. of IEICE, Vol.J70-D, No.1, pp.127-138 (1987) (in Japanese).

[4]Y.Lamdan and H.J.Wolfson:"Geometric Hashing:A General and Efficient Model-Based Recognition Scheme", Proc. of the 2nd ICCV, pp.238-249 (1988).

[5]W.E.L.Grimson and D.P.Huttenlocher:"On the Sensitivity of Geometric Hashing", Proc. of the 3rd ICCV, pp.334-338 (1990).

[6]Y.Lamdan, J.T.Schwartz and H.J.Wolfson:"Affine Invariant Model-Based Object Recognition", IEEE Trans., Vol.RA-6, No.5, pp.578-589 (1990).

[7]W.E.L.Grimson and D.P.Huttenlocher:"On the Verification of Hypothesized Matches in Model-Based Recognition", IEEE Trans., Vol.PAMI-13, No.12, pp.1201-1213 (1991).

[8]Y.Lamdan and H.J.Wolfson:"On the error analysis of 'Geometric Hashing'", Proc. of the CVPR, pp.22-27 (1991).

[9]S.J.Dickinson, A.P.Pentland and A.Rosenfeld:"3-D Shape Recovery Using Distributed Aspect Matching", IEEE Trans., Vol.PAMI-14, No.2, pp.174-198 (1992).

[10]Y.Kawanishi, K.Deguchi and I.Morishita:"An Improvement of Robustness of Geometric Hashing for Model-Based Matching", SIGCV-JSIP, Vol.CV-80, pp.161-168 (1992) (in Japanese).

[11]S.Miyanabe, S.Kaneko, T.Honda:"Estimating 3D Motion of Solid Objects based on Stability of Features", SIGCV-JSIP, Vol.CV-80, pp.231-238 (1992) (in Japanese).

[12]S.Vinther and R.Cipolla:"Active 3D Object Recognition using 3D Affine Invariants", ECCV, Vol.801, pp.15-24 (1994).

[13]M.Shibata, S.Kaneko and T.Honda:"Object Posture Recognition Based on Directional Voting with Geometric Hashing", MIRU'94, Vol.2, pp.81-87 (1994) (in Japanese).