PRE-MARKING METHODS FOR 3D OBJECT RECOGNITION

R. Safaee-Rad Graduate Student
B. Benhabib Assistant Professor
K.C. Smith Professor
Z. Zhou Graduate Student

Computer Integrated Manufacturing Laboratory, Department of Mechanical Engineering
University of Toronto, Toronto, Ontario, Canada M5S 1A4

Abstract

Two pre-marking methods are proposed for a new 3D object recognition system under development at the University of Toronto. In this system, an object is modeled using only a small number of 2D distinct perspective views (standard-views) predefined with the help of markers placed on the object. During the recognition process, a standard-view is acquired by first determining its surface normal (standard-view-axis), and then aligning the camera's optical axis with it. Standard-view-axes are obtained by analysing the images of the markers. A morphological skeleton transform (MST) is used for the extraction of required marker features. This paper presents the analytical solutions for the two proposed pre-marking schemes, based on circular markers, which can be used in acquiring standard-views of objects. Specific issues addressed include: the determination of the perspective distortion and its relative importance, the determination of the transformation parameters required for camera alignment, and the use of a class of MST, pseudo-Euclidean skeleton, for feature extraction.

1. Introduction

Flexibility of robotic workcells can be significantly increased through the integration of visual sensors, reducing the need for special and often complex tooling. The research program currently underway at the University of Toronto focusses on developing an efficient and flexible 3D object-recognition method for vision-based robotic assembly. The main incentives in pursuing this research are the lack of generalized computer vision theories applicable to robotic vision, and the inefficiency and inflexibility of existing 3D object-recognition methods for industrial environments [1].

Object recognition for assembly requires the identification of workpieces or sub-assemblies, determination of their locations (position and orientation), and extraction of their salient features for visual servoing. Volumetric methods based on the exact specification of an object, and multi-view feature representation based either on the characteristic views of an object or on the discrete view-sphere representation have been developed [2]. However, these techniques are complex and computationally expensive due to their need for a 3D matching process. The new 3D object recognition technique under development is based on a pre-marking scheme, as a consequence of which the matching process can be performed in 2D space.

In the machine-vision literature, the use of pre-marking in various fields, including biomedical, space, and robotic research, has been reported, [3,4,5]. Although the concept of pre-marking of industrial parts was proposed a decade ago [6], systematic and extensive research has not taken place since then on the development of mathematical tools and other related aspects. Specific applications of pre-marking, such as the use of metal-marking dyes for industrial vision inspection reported in [7], have not been generalized. One of the basic mathematical tools essential for a pre-marked-3D-object-recognition method, is an algorithmic procedure for object-orientation determination.

2. Problem Statement

In the proposed 3D object recognition technique, an object is modeled using only a small number of 2D distinct perspective views. The number of views depends partly on the degrees of symmetry of the object. These are referred to as ***standard-views***, each with a corresponding ***standard-view-axis***. For successful recognition purposes, the input image of an object should be one of its standard perspective views. Thus, a mobile camera is used, such that the camera's optical axis can be aligned with one of the standard-view-axes of the object in order to acquire a standard-view. Then, the matching process is performed between the acquired 2D standard-view of the object under consideration and the library of 2D standard-views of a set of objects.

To enable the vision system to acquire standard-views, standard-view-axes must be pre-defined. This can be accomplished by defining a local surface normal for each distinct view of an object. These local surface normals can be defined by adding non-functional features to an object, preferably during the manufacturing process. This process is referred to as pre-marking. Thus, in the context of the proposed object recognition method, pre-marking is used for defining local surface normals and the corresponding standard-view-axes.

The basic problem associated with such a scheme, is the development of mathematical algorithms for extracting surface normals from the markers' images. This paper presents two solution methods to this problem. These methods have been studied and analytical solutions have been developed for them. The first method can be defined as: given an image of a circular marker, determine its orientation in 3D space. The second method can be defined as: given an image of three circular markers (equal in size), where their centers create an equilateral triangle, find the orientation of the equilateral triangle in 3D space.

Circular geometry has been selected for markers for the following reasons: 1) many industrial parts have holes, which can be used as "natural" circular markers; 2) a circle has symmetry with respect to its norm; furthermore, the perspective projection of a circle in any arbitrary orientation in 3D space is always a pseudo-ellipse; and 3) circles have been shown to have the property of high image location accuracy [8].

3. Method I – Single Circular Marker

In the case of a single circular marker on an object, the standard-view is acquired by aligning the camera's optical axis with the surface normal passing through the center point of the marker. This surface normal is considered as the standard-view-axis, and it can be determined given that a camera calibration procedure has been carried out prior to any calculations [9]. Thus, it is assumed that the, internal and external camera parameters are known. Based on these parameters, the ideal central projection model is used for deriving equations for the calculation of the perspective distortion, and the required rotations and translation of the camera in order to align its optical axis with the surface normal of a circular marker in 3D space.

CH2809-2/89/0000-0592 $1.00 © 1989 IEEE

The relative position and orientation of a camera with respect to a circular marker may be classified in four cases as follows:

(1) The camera's optical axis passes through the center of the marker, and it is also perpendicular to the circular marker plane (C-plane).

(2) The camera's optical axis is perpendicular to the C-plane, but it is at an offset from the center of the marker.

(3) The camera's optical axis passes through the center of the circular marker, but it is not perpendicular to the C-plane.

(4) The camera's optical axis is not perpendicular to the C-plane, and it is at an offset from the circular marker center.

Case (4) was investigated, and the derived equations are presented in this paper, since it is the most general case. The other three cases are sub-sets of case (4). Figure 1 represents case (4). There exist three planes: image plane (I-plane), circular marker plane (C-plane), and a plane which is parallel to the I-plane and passes through the center of the circular marker (D-plane). Based on geometrical reasoning, it can be shown that Figure 1 can be simplified schematically to Figure 2.

3.1. Perspective Distortion of a Circular Marker

The perspective projection of a circular marker is not an exact ellipse, but an ellipse with some symmetrical distortion with respect to the major axis – a quasi-ellipse. This perspective distortion for case (4) can be determined using Figure 2. In this figure, the distance d is the offset of the camera's optical axis CC' from the circular marker's center point O; f is the effective focal length of the camera (CL); h' is the distance between the camera lens and the circular marker's center; h is the projection of h' onto the camera's optical axis; r is the radius of the marker; α is the angle between the C-plane and the D-plane; and, d_1 is the offset, in the image plane, of the image center from the center of the quasi-ellipse (CO_1). Perspective distortion is then defined as,

$$\Delta = O_1A_1 - O_1B_1. \tag{1}$$

Using the property of similar triangles, the perspective distortion can be expressed as follows,

$$\Delta = f\frac{\sin 2\alpha}{\frac{h^2}{r^2} - \sin^2\alpha} + f\frac{\sin^2\alpha}{\frac{h^2}{r^2} - \sin^2\alpha}\left[\frac{2d}{h}\right]. \tag{2}$$

The relative importance of the distortion can be defined as,

$$\epsilon = \frac{O_1A_1 - O_1B_1}{O_1A_1 + O_1B_1} = \frac{\Delta}{O_1A_1 + O_1B_1} \tag{3}$$

which can be simplified to,

$$\epsilon = \frac{r\sin\alpha}{h}. \tag{4}$$

Equation (4) implies that ϵ is only a function of r, h, and α, thus, it is independent of the camera offset distance d, (i.e., it is the same for cases (3) and (4)). Furthermore, in an actual situation, for example for $r = 5$ mm and $h = 500$ mm, the maximum value of ϵ occurs at $\alpha = 90°$ and it is less than or equal to 1%. Therefore, it can be concluded that distortion under an actual situation is negligible.

On the other hand, equation (2) shows that in the general case, Δ is a function of f, h, r, α, and d. Thus, for $\frac{h}{r} \geq 10$, which is generally true in actual situations, equation (2) can be approximated as,

$$\Delta \cong \frac{r^2}{h^2}\left[f(\sin 2\alpha + \frac{2d}{h}\sin^2\alpha)\right]. \tag{5}$$

Furthermore, for no camera offset ($d = 0$), equation (5) becomes,

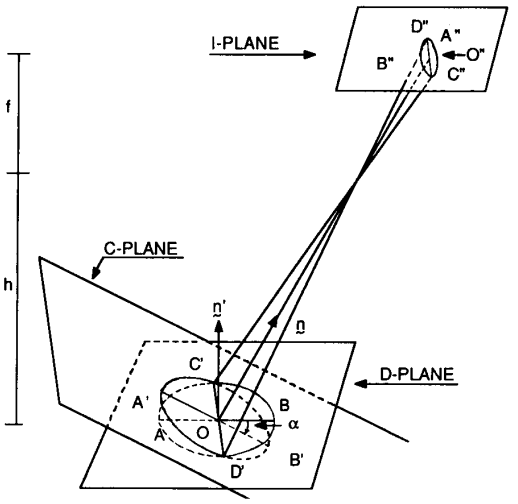

Figure 1. The relative positions of the marker and the image planes for case (4).

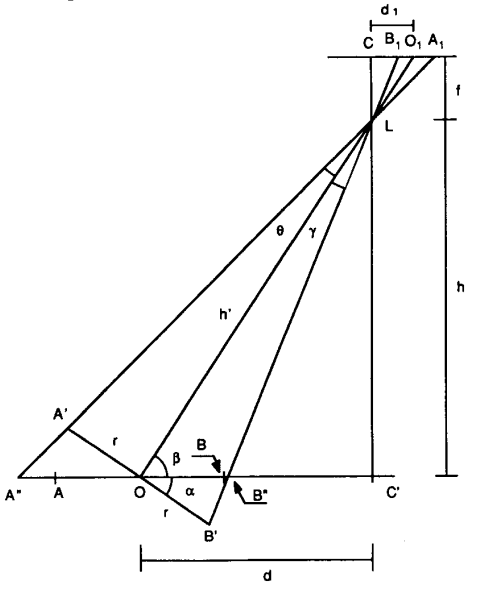

Figure 2. The simplified schematic representation of case (4).

$$\Delta \cong \frac{r^2}{h^2} f\sin 2\alpha. \tag{6}$$

Thus, the amount of distortion for case (3) is less than case (4). Furthermore, in cases (1) and (2) ($\alpha = 0$), Δ would be zero, (i.e., no perspective distortion in these two cases, as expected). For the case of $\frac{h}{r} \geq 10$, assuming $f = 25$ mm, the maximum perspective distortion occurs at $\alpha = 45°$, and it would be 0.25 mm.

3.2. Determination of the Angle α

The semi-major axis of the quasi-ellipse, R_{max}, and the semi-minor axis of the quasi-ellipse, R_{min}, are expressed as,

$$R_{max} = f\frac{r}{h} \qquad R_{min} = \frac{1}{2}(O_1A_1 + O_1B_1). \tag{7}$$

For $\frac{h}{r} \geq 10$, the ratio (R min/R max) can be approximated as,

$$\frac{R_{min}}{R_{max}} \cong \cos\alpha + \frac{d}{h}\sin\alpha. \tag{8}$$

For case (3), where the camera offset is zero, equation (8) becomes,

$$\frac{R_{min}}{R_{max}} \cong \cos\alpha.$$

For cases (1) and (2), where $\alpha = 0$, the ratio woul be,

$$\frac{R_{min}}{R_{max}} = 1.$$

In order to determine the angle α from equation (8), the camera offset d can be calculated, using Figure 2, as, $d = (h\ d_1)/f$ where d_1 is estimated from the acquired image. Thus, the offset d can be calculated if h is known. Substituting this equation into equation (8), and solving for the unknown angle α yields

$$\sin\alpha = \frac{\frac{R_{min}}{R_{max}}\frac{d_1}{f} \pm \sqrt{1+\left[\frac{d_1}{f}\right]^2 - \left[\frac{R_{min}}{R_{max}}\right]^2}}{1+\left[\frac{d_1}{f}\right]^2} \tag{9}$$

It can be shown that equation (9) will yield a positive and a negative value for sinα, where only one solution is acceptable. If it is possible to detect the perspective distortion Δ, the sign of the angle α, and the corresponding direction of the camera rotation can be concluded to be in the direction of the shorter semi-minor axis. Otherwise, information about the relative change of the minor-axis obtained from the two images that are used for depth calculation, should be used.

3.3. Procedure for Acquiring a Standard-View Image

The basic assumption of the development is $\frac{h}{r} \geq 10$. Since the radius of the marker is a controlled parameter, this assumption can easily be satisfied. The approximate solutions previously derived, for the general case (4), are used to establish the following procedure to acquire standard-view images:

(a) Applying morphological skeleton transformation, the minor and major axes of the quasi-ellipse are determined. Then, the center of the quasi-ellipse, R_{min}, and R_{max}, are estimated [10].

(b) The camera is moved in the direction of the minor-axis in order to acquire a second image of the circular marker. Then, the center, R_{min}, R_{max}, and d_1 of the new quasi-ellipse are determined.

(c) Based on the calibration method, the depth of the circular marker's center, h, is estimated. Then the camera offset, d, is calculated.

(d) Using equation (9), based on the knowledge of R_{min}, R_{max}, f, and d_1, the two solutions for the angle α are calculated. Since during the process of acquiring a second image (step (b)) the minor axis R_{min} has changed, the direction of this change can be used for selecting the acceptable solution for α (either the negative or the positive angle).

(e) If the angle between the x-axis of the camera frame and the major axis of the quasi-ellipse, acquired in step (b), is referred to as δ, then, the process of alignment goes through two rotations, δ and α, and one translation, d, as follows. The camera's x–y axes are rotated by an angle δ to be aligned with the quasi-ellipse's major-minor axes; the new x–y axes are then rotated by an angle α in the direction selected in step (d) with respect to the camera's new x-axis; and finally, the camera is translated by a distance d in the direction opposite to the direction of $\overrightarrow{CO_1}$. These two rotations and one translation of the camera frame, would align the camera's optical axis with the surface normal of the circular marker.

4. Method II – Three Circular Markers

An important aspect to consider in using a single circular marker is the degree of accuracy in obtaining R_{min} and R_{max}, and the corresponding angle α. If the circular marker is very small relative to the distance between the camera and the marker, small changes in the angle α may not be detected, and thus, result in reduction in accuracy. An alternative method to the use of a single marker is the use of three markers, where the relative positions of the three markers' centers with respect to each other are of interest.

If three circular markers are located on an object, such that their centers create an equilateral triangle, then, an image of this equilateral triangle will not be equilateral except when the camera's optical axis is aligned with the surface normal of the equilateral triangle. Thus, the problem can be defined as the determination of the orientation of the equilateral triangle plane in 3D space. Figure 3 presents this problem schematically.

The coordinates of points O, A, B, and C, in Figure 3, are known. Thus the angles α, β, and γ can be estimated using the direction cosines of lines OA, OB and OC. Note that, since all the coordinates (x_i, y_i) and f have unique values, then cosα, cosβ, and cosγ also have unique values. In order to determine the orientation of a plane that intersects the three lines OA, OB, and OC, and creates an equilateral triangle with the intersection points A', B', and C', the following expressions are derived,

$$c^2 - 2bc\cos\alpha = a^2 - 2ab\cos\gamma \tag{10}$$

and,

$$c^2 - 2ac\cos\beta = b^2 - 2ab\cos\gamma \tag{11}$$

where $OA' = a$, $OB' = b$, $OC' = c$, and $A'B' = A'C' = B'C'$. Equations (10) and (11) can be re-expressed respectively as,

$$T^2 - 2T\cos\alpha = S^2 - 2S\cos\gamma \tag{12}$$

$$S = \frac{T^2 - 1}{2T\cos\beta - 2\cos\gamma} \tag{13}$$

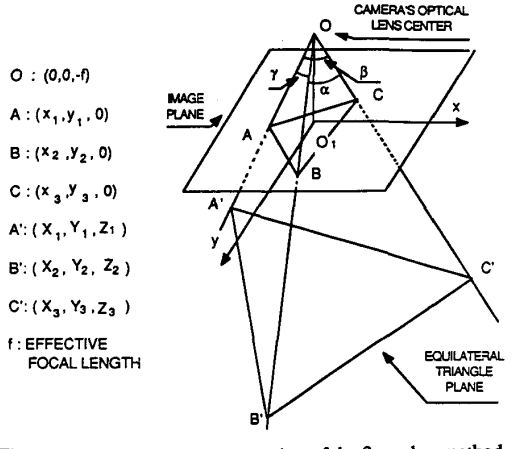

Figure 3. The schematic presentation of the 3-marker method.

where $S = a/b$ and $T = c/b$. Substituting equation (13) into equation (12), and simplifying it into a fourth-order equation, the following expression is obtained,

$$T^4 + NT^3 + PT^2 + QT + R = 0 \qquad (14)$$

where

$$N = \frac{-4\cos\beta\cos\gamma - 8\cos\alpha\cos^2\beta}{4\cos^2\beta - 1}$$

$$P = \frac{16\cos\alpha\cos\beta\cos\gamma + 2}{4\cos^2\beta - 1}$$

$$Q = \frac{-8\cos\alpha\cos^2\gamma - 4\cos\beta\cos\gamma}{4\cos^2\beta - 1}$$

$$R = \frac{4\cos^2\gamma - 1}{4\cos^2\beta - 1}.$$

Note that, if $(4\cos^2\beta = 1)$, the fourth-order term in equation (14) is eliminated yielding a third-order equation. As mentioned earlier, $\cos\alpha$, $\cos\beta$, and $\cos\gamma$ all have unique and real values, thus N, P, Q, and R also have unique and real values. Based on this, equation (14) can have at most four solutions. Furthermore, since a/b and c/b are both positive, S and T also must be positive. Thus, only positive solutions are acceptable for S and T in equations (12) and (13).

Equation (14) can be solved by using the resolvant cubic equation method [11], for which the standard solution of cubic equations is used. After solving equation (14), a set of at most four positive solutions for T, and a set of at most four positive solutions for S will be obtained.

Since there exists a unique surface normal for the equilateral triangle plane, only one solution for T and S is acceptable. In order to determine this acceptable solution, the relative sizes of the marker images in the image plane can be used (assuming all circular markers have the same size). The relative sizes of the markers can be determined by estimating the size of the major axis of each quasi-ellipse. Based on this information, and corresponding to it, the actual relative sizes of a, b, and c can be determined.

On the other hand, as previously defined, given a set of S and T values (at most four sets), the corresponding values a and c (as a function of b) can be determined, based on the following equations, $a = bS$ and $c = bT$. Then, one of the following six possible cases may occur:

$$a \geq b \geq c \quad a \geq c \geq b \quad b \geq a \geq c$$

$$b \geq c \geq a \quad c \geq a \geq b \quad c \geq b \geq a.$$

The estimated cases, based on the S and T values, are compared with the one based on the relative sizes of the three markers, and thus the unique, acceptable, and accurate solution is determined.

The remaining problem is, then, the determination of the 3D coordinates of points A', B', and C'. Based on the knowledge of the 3D coordinates of points O, and A, and the distance OA' which is equal to a, Figure 3, the following expressions hold,

$$X_1 = \frac{Z_1 + f}{f} x_1 \qquad Y_1 = \frac{Z_1 + f}{f} y_1. \qquad (15)$$

Also based on the knowledge that, $a^2 = X_1^2 + Y_1^2 + (Z_1 + f)^2$, the following expression can be obtained by using equations (15)

$$Z_1 = -f \pm \frac{af}{\sqrt{x_1^2 + y_1^2 + f^2}}. \qquad (16)$$

Since the equilateral triangle plane is in the positive side of the z-axis, only a positive value for Z_1 is acceptable.

Finally, the direction cosines of the line perpendicular to the equilateral triangle plane can be expressed using the X, Y, and Z coordinates. Note that, all the 3D coordinates are expressed in terms of b, however, since the third direction cosine is dependent on the other two, the surface normal can still be determined. The camera's optical axis can then be aligned with the standard view-axis based on the knowledge of depth information from the calibration procedure.

5. Conclusions

A new 3D-object recognition technique based on pre-marking is proposed. This technique allows one to model an object with a small number of 2D distinct perspective views. Markers are located on the surfaces of an object to define local surface normals thus, defining the standard-views of the object. As a result, based on the pre-marking scheme, it is possible to perform the matching process in 2D space, though the object may have any orientation in 3D space. Two methods of pre-marking, using circular markers, are presented. The first one is based on a single circular marker, where a standard-view of an object can be acquired by aligning the camera's optical axis with the surface normal of the marker passing through its center. The second method is based on the use of three markers, whose centers form an equilateral triangle. An analytical solution for the calculation of the direction cosines of the surface normal is formulated.

References

[1] Pavlidis, T., "A Critical Survey of Image Analysis Methods", *Proc., International Conference on Pattern Recognition*, pp. 502-511, 1986.

[2] Basel, P.J., and Jane, R.C., "Three-Dimensional Object Recognition", *Computing Surveys*, Vol. 17, No. 1, pp. 75-145, March 1985.

[3] Safaee-Rad, R., Shwedyk, E., and Quanbury, A.O., "Functional Human Arm Motion Study with a New 3D Measurement System (VCR-PIPEZ-PC)", *Proc., IEEE-EMBS Ninth Annual Conference*, November 1987.

[4] Bales, J.W., and Barker, L.K., "Marking Parts to Aid Robot Vision", *NASA, Technical Paper 1819*, 1981.

[5] Kabuka, M.R., and Arenas, A.E., "Position Verification of a Mobile Robot Using Standard Pattern", *IEEE, J. of Robotics and Automation*, Vol. RA-3, No. 6, pp. 505-516, Dec. 1987.

[6] Baird, M.L., "Future Directions of Industrial Applications of Pattern Recognition (Panel Discussion)", *Proc., Fourth International Joint Conference on Pattern Recognition*, pp. 1146, Kyoto, 1978.

[7] Bodziak, T.A., "Using UV Metal-Marking Dyes in Industrial Vision Inspection of Complex Assemblies", *Proc., Vision '86*, pp. 9-31 to 9-39, Detroit, 1986.

[8] McVey, E.S., and Jarvis, G.L., "Ranking of Patterns for Use in Automation", *IEEE, Trans. on Industrial Electronics and Control Instrumentation*, Vol. IECI-24, No. 2, pp. 211-213, May 1977.

[9] Tsai, R.Y., "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", *IEEE, J. of Robotics and Automation*, Vol. RA-3, No. 4, pp. 323-344, August 1988.

[10] Zhou, Z., Smith, K.C., Benhabib, B., and Safaee-Rad, R., "Morphological Skeleton Transforms for Determining Position and Orientation of Pre-Marked Objects", *Proc., IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, pp. 301-305, Victoria, Canada, June 1989.

[11] Beyer, W.H., *Standard Mathematical Tables*, CRC Press Inc., Florida, 1981.