Res. Rep. Fac. Eng. Mie Univ., Vol. 8, pp. 117 - 126(1983)

Review

Scene Analysis by Matching 2-D Image with 3-D Model

Fumitaka KIMURA, Mitsu YOSHIMURA*
and Yasuji MIYAKE
(Department of Electrical and Electronic Engineering)

(Received August 10, 1983)

This paper proposes a technique of matching a 2-D image taken of a 3-D object with the model. It is shown that the perspective transformation relating an image to the model can be approximated by an orthogonal projection. The formula determining the parameters of the orthogonal projection from the coordinates of three corresponding points on the model and image is developed.

1. Introduction

The matching technique proposed here is an extended 3-D version of the early proposed algorithm for the subpattern matching of 2-D line patterns(1).

The relational structural matching has often been applied to 3-D scene analysis(2)(3), where geometrical transformation relating the input and the model is not explicitly utilized since it doesn't make sense for the relational structures. Roberts(5) and Falk(6) calculate, in their scene analysis system for polyhedra, the parameters of a geometrical transformation to determine the location of the objects. But those correspondences of the image vertexes and model vertexes which are necessary to calculate the parameters are determined by topological matching. The geometrical transformation doesn't play an important role in the matching process. On the other hand, some matching techniques for intrinsically 2-D pictures such as coast lines or characters determine and utilize them in the matching process(1)(4)(7). The technique described here determines and utilizes the parameters of orthogonal projection relating the 2-D image and the 3-D model. An example is shown on determining the kind and location of the polyhedron portraied in the input image.

* Faculty of Engineering, Nagoya University

2. Orthogonal Projection Approximation

Let's consider that a given image is the one taken of the model by a camera in an arbitrary position and orientation. The image and the model is related by the perspective transformation. In this section we show that the perspective transformation can be approximated by an orthogonal projection in order to make the later analytical treatment easy.

Let O-xyz and O'-x'y'z' be a model coordinate system and a camera coordinate system respectively. Suppose that the image plane of the camera is on the x'-z' plane and the optical axis coincides the y' axis (Fig.1). Then, the model coordinates of a point P and the camera coordinates are in the relationship

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} e_{11} & e_{12} & e_{13} & u_1 \\ e_{21} & e_{22} & e_{23} & u_2 \\ e_{31} & e_{32} & e_{33} & u_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{1}$$

Where $\mathbf{e}_i = (e_{i1}, e_{i2}, e_{i3})^t$ $(i = 1,2,3)$ denote the direction consines of x', y', z' axes respectively and $\mathbf{u} = (u_1, u_2, u_3)$ denotes the origin of the model coordinate system.

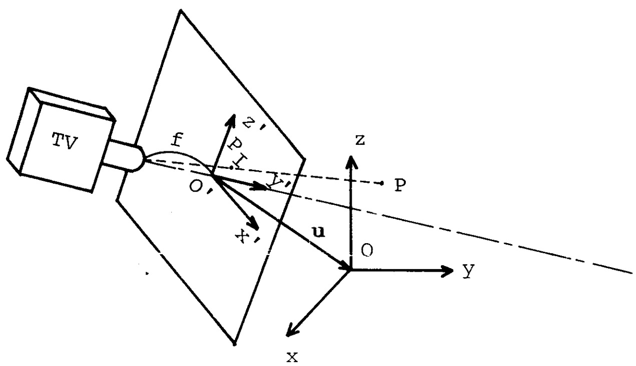


Fig.1 A model of taking images.

The coordinates of the point P in the camera coordinate system (camera coordinates) and the coordinates x_P' and z_P' of the image point P_I on the image plane are related by the perspective transformation[(8)]

$$\begin{pmatrix} w'x_p' \\ w'y_p' \\ w'z_p' \\ w' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1/f & 0 & 1 \end{pmatrix} \begin{pmatrix} wx' \\ wy' \\ wz' \\ w \end{pmatrix}. \quad (2)$$

Where f denotes the distance between the image plane and the lens, and w and w' are arbitrary constants introduced to represent the perspective transformation in the form of linear transformation. y_p' is always equel to 0 regardless of (2).

Combining (1) and (2), the following relationship between the model coordinates of the point P and the camera coordinates of the image point P_I is obtained.

$$\begin{pmatrix} w'x_p' \\ w'y_p' \\ w'z_p' \\ w' \end{pmatrix} = \begin{pmatrix} e_{11} & e_{12} & e_{13} & u_1 \\ e_{21} & e_{22} & e_{23} & u_2 \\ e_{31} & e_{32} & e_{33} & u_3 \\ e_{21}/f & e_{22}/f & e_{23}/f & (u_2+f)/f \end{pmatrix} \begin{pmatrix} wx \\ wy \\ wz \\ w \end{pmatrix} \quad (3)$$

From (3),

$$w' = w(e_{21}x + e_{22}y + e_{23}z + u_2 + f)/f. \quad (4)$$

Then, (3) can be represented as

$$\begin{pmatrix} x_p' \\ y_p' \\ z_p' \\ 1 \end{pmatrix} = \frac{f}{u_2} \begin{pmatrix} e_{11} & e_{12} & e_{13} & u_1 & x \\ e_{21} & e_{22} & e_{23} & u_2 & y \\ e_{31} & e_{32} & e_{33} & u_3 & z \\ 0 & 0 & 0 & u_2/f & 1 \end{pmatrix}, \quad (5)$$

if inequality $u_2 \gg e_{21}x + e_{22}y + e_{23}z + f$ holds. From $y_p' = 0$ and (5), we can say that the camera coordinates of the image point p_I can be approximated by multiplying coordinates of the point p projected on the image plane by f/u_2.

3. Parameter Determination

In this section, we will develop the formula determining the parameters of the orthogonal projection relating an image and the model. From (5), among the coordinates (x_k, y_k, z_k) $(k = 1, 2, 3)$ of three points on a model and the coordinates (x_k, z_k) of their image points hold the equations

$$\begin{aligned} x_k &= e_{11}'x_k + e_{12}'y_k + e_{13}'z_k + u_1' \\ z_k &= e_{31}'x_k + e_{32}'y_k + e_{33}'z_k + u_3' \quad (k = 1, 2, 3) \end{aligned} \tag{6}$$

where $e_{11}' = fe_{11}/u_2'$ $e_{12}' = fe_{12}/u_2' \ldots$, and so on. From the property of direction cosines, the equations

$$e_{i1}'e_{j1}' + e_{i2}'e_{j2}' + e_{i3}'e_{j3}' = \begin{cases} (f/u_2)^2 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \quad (i, j = 1, 2, 3) \tag{7}$$

hold. Since the number of the equations in (6) and (7) is equal to that of the unknown parameters, we can solve them to develop the formula (8) by which we can determine the transformation parameters from the coordinates of three points on a model and the camera coordinates of their image points.

$$u_1 = u_1'/\ \|\mathbf{e}_1'\| \ ,\ u_2 = f/\ \|\mathbf{e}_1'\| \ ,\ u_3 = u_3'/\ \|\mathbf{e}_3'\|$$

$$\mathbf{e}_1 = \mathbf{e}_1'/\ \|\mathbf{e}_1'\| \ ,\ \mathbf{e}_3' = \mathbf{e}_3'/\ \|\mathbf{e}_3'\| \ ,\ \mathbf{e}_2 = \mathbf{e}_3 \times \mathbf{e}_1'$$

where (8)

$$u_3' = (A_z u_1' - A_{xz})/(A_e u_1' - A_x)$$

$$\mathbf{e}_1' = (\mathbf{W}^{-1})^t(\mathbf{X} - u_1'\mathbf{E})$$

$$\mathbf{e}_3' = (\mathbf{W}^{-1})^t(\mathbf{Z} - u_3'\mathbf{E}).$$

In (8), u_1' is a root of a fourth order equation

$$A_e^{\,3}u_1'^{\,4} - 4A_e^{\,2}A_x u_1'^{\,3}$$

$$+ \quad A_e\ A_e(A_{xx} - A_{zz}) + 5A_x^{\,2} + A_z^{\,2} \quad u_1'^{\,2}$$

$$- \quad 2A_eA_x(A_{xx} - A_{zz}) + 2A_x^{\,3} + 2A_z^{\,2}A_x \quad u_1$$

$$+ A_x^2(A_{xx} - A_{zz}) + 2A_xA_zA_{xz} - A_eA_{xx}^2 \quad ,$$

where

$$A_e = \mathbf{E}^t\mathbf{W}^{-1}(\mathbf{W}^{-1})^t\mathbf{E} \;, \quad A_x = \mathbf{E}^t\mathbf{W}^{-1}(\mathbf{W}^{-1})^t\mathbf{X}, \quad A_z = \mathbf{E}^t\mathbf{W}^{-1}(\mathbf{W}^{-1})^t\mathbf{Z}$$

$$A_{xx} = \mathbf{X}^t\mathbf{W}^{-1}(\mathbf{W}^{-1})^t\mathbf{X}, \quad A_{zz} = \mathbf{Z}^t\mathbf{W}^{-1}(\mathbf{W}^{-1})^t\mathbf{Z}, \quad A_{xz} = \mathbf{X}^t\mathbf{W}^{-1}(\mathbf{W}^{-1})^t\mathbf{Z}$$

$$\mathbf{W} = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \mathbf{Z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \qquad \mathbf{E} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{9}$$

Even if the size of an object is different from that of the model, all parameters are determined correctly except for the distance u_2 between the image plane and the origin of the model coordinate system. Therefore, the matching technique described in the next section is still valid for the objects different in the size from the model.

4. Matching

In this section, a metric measuring "badness" of a match of a given image with the model is introduced. The image point corresponding to each model point is determined so as to make the metric minimum. The metric must measure the badness of the match taking into account of the perspective transformation. Hereafter, we define a matching metric for polyhedra as the objects and describe a matching technique using the metric.

The model of a polyhedron is represented by the coordinates (x_i, y_i, z_i) ($i = 1, 2, 3, \ldots, I$) of the vertexes and the edges between the vertexes. An input image is also represented by the coordinates (x_j, z_j) ($j = 1, 2, 3, \ldots, J$) of the vertexes and the edges between them if we apply some preprocessings such as edge detection and line fitting to the image. There may be false vertexes and edges, in the line drawings thus obtained, which don't correspond to those of the model.

First, we define a metric measuring the badness of matches of four pairs (associations) $a_i = (i, j_i)$ $(i = 1, 2, 3, 4)$ of the model vertexes i and the input vertexes j_i. Given four associations, three of them can be used to determine the parameters of transformation from the model to the input. The squared error between the coordinates expected by mapping the fourth model point onto the image plane according to the parameters and the actual coordinates of the fourth point can be used as a metric reflecting the badness of the matches (Fig.2).

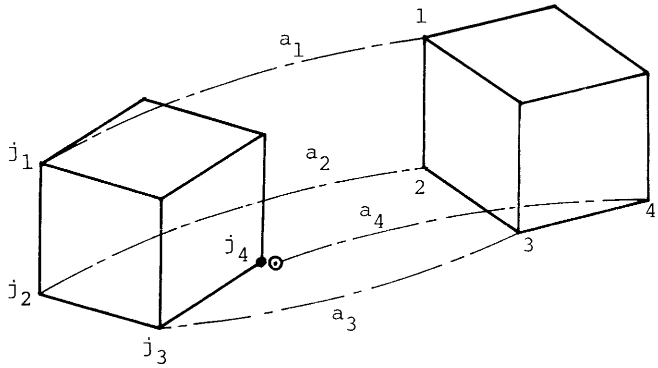


Fig.2 Actual location ● and expected location ⊙ of fourth point.

Finally, if we put the vertexes of the model into a proper sequence and if we calculate the transformation parameters using the first three points of the successive four points, a matching metric for I pairs of associations $\mathbf{a} = (a_1, a_2, \ldots, a_I)$ is given as

$$D(\mathbf{a})^2 = \sum_{i=4}^{I} [\{X_{ji} - (e_{11}^{(i)} x_i + e_{12}^{(i)} y_i + e_{13}^{(i)} z_i + u_1^{(i)})\}^2$$

$$+ \{Z_{ji} - (e_{31}^{(i)} x_i + e_{32}^{(i)} y_i + e_{33}^{(i)} z_i + u_3^{(i)})\}^2], \qquad (10)$$

where $\mathbf{e}_1^{(i)}$, $u_1^{(i)}$, ..., are calculated from the coordinates of the i-3, i-2, i-1 th model vertexes and the corresponding j_{i-3}, j_{i-2}, j_{i-1} th input vertexes. Since (10) can be rewritten in the form of

$$D(\mathbf{a})^2 = \sum_{i=4}^{I} g_i(a_{i-3}, a_{i-2}, a_{i-1}, a_i), \qquad (11)$$

the associations $\mathbf{a}$ minimizing (11) can be obtained by the third order dynamic programming.

The vertexes of the model can be serialized in some arbitrary ways and the value of the matching metric can vary with the way, but the combination of the model with serialized elements and D. P. -like matching has been shown to be effective in both the aspects of cost and reliability.[(1)(9)]

Eq. (10) has no term related to the edges. The condition that the two image vertexes respectively corresponding to two model vertexes connected by an edge must also be connected by an edge, is utilized in the decision process of the D. P. matching as a restriction. Therefore, the more adjacent model vertexes in the sequence are connected by the edges, the more the computational cost may be saved.

5. Example

Example is shown on determining the kind and location of a polyhedron by matching an input image with several models of polyhedra. A cube, a rectangular parallelepiped, a regular tetrahedron and a quadrangular pyramid made of plaster were used as the objects.

A model of a cube is shown in Fig. 3. This model consists of a sequence of six vertexes among eight vertexes of the cube. In this experiment, we represented each model in terms of the subset of vertexes which always appear in the images at a time (except critical cases) easily to solve the problem of self occlusion. Fig.4 shows an input image (in a resolution of 128×128) (a), results of edge enhancement (b), edge detection (c) and line fitting (d). Eight images (two images for each model) were used. The kind and location of each polyhedron portraied in the images were correctly determined except for one image. Fig.5 shows the image of the quadrangular pyramid misrecognized as a regular tetrahedron. The positions of the vertexes are apt to vary in the preprocessing process due to the poor resolution. Since the misrecognition seems to be caused by the inexact positions of the vertexes, the error may be avoided by taking images in higher resolution. Examples of correct matching are shown in Fig.6. Thick lines in the figure correspond to the models.

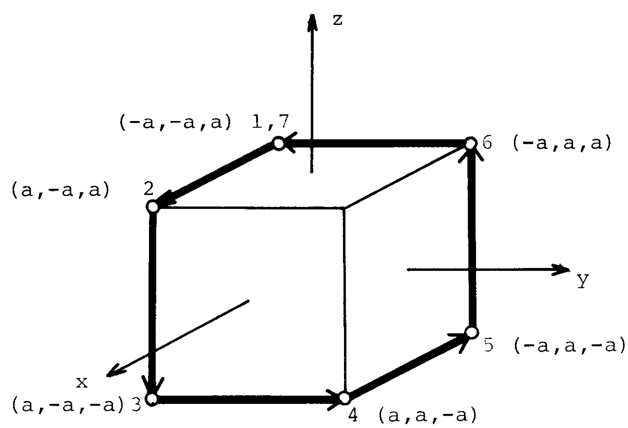
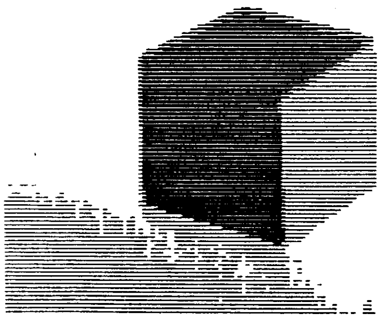
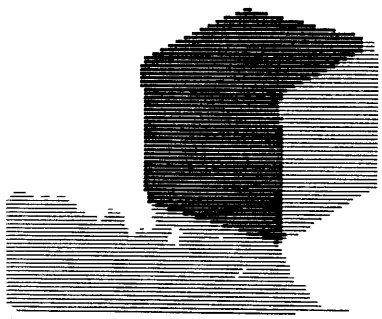


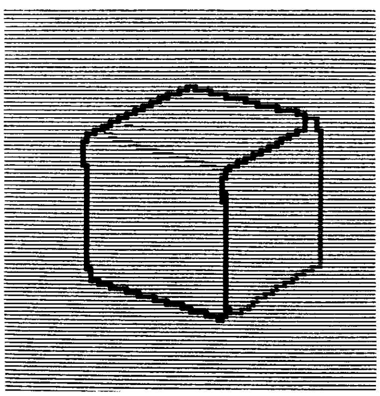
Fig.3 Model of a cube (a=6.25 cm)



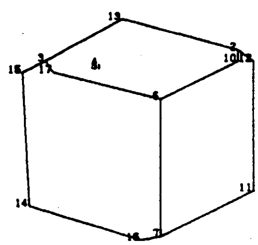
(a)



(b)



(c)



(d)

Fig.4 Input image (128 × 128) (a), results of edge enhancement (b), edge detection (c) and line fitting (d).

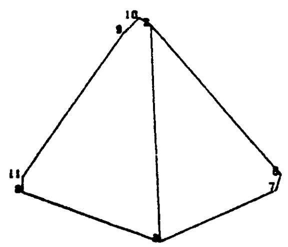


Fig.5 An example of quadrangular pyramid.

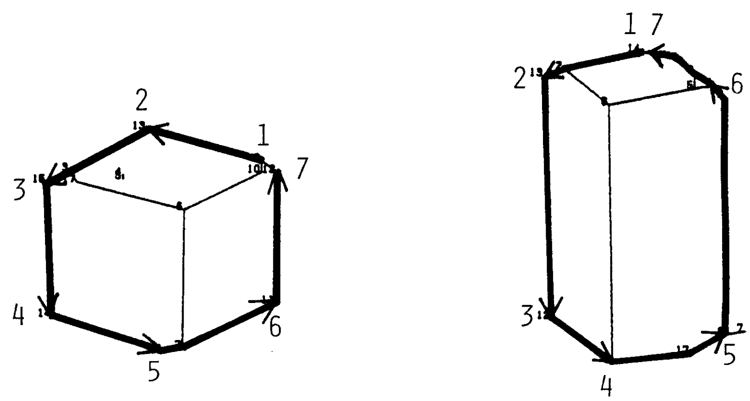


Fig.6 Examples of matching (Thick lines correspond to the model).

6. Conclusion

In this paper, (1) it was shown that the perspective transformation can be approximated by an orthogonal projection, (2) the formula determining the parameters of the orthogonal projection from the coordinates of corresponding three points on the model and image was developed, and (3) a technique of matching a 2-D image taken of a 3-D object with the model was proposed.

Our matching techinique sets the correspondences between the model vertexes and the image vertexes on the basis of their geometrical location rather than their topological properties. Therefore, it is less affected by the false vertexes and false edges in the image. Breaks of edges can be properly treated if a simple preprocessing is applied to the line drawings.[10] In the preprocessing, all probable breaks may be immediately complemented, since this matching techinique is not affected by the false edges (caused by the unnecessary miscomplementation). The formula developed can be applied to some problems other than the matching. For example, the parameters determined by this formula can be used as good initial values to determine the parameters of original perspective transformation by an iterative techinique such as hill climbing.

Remaining problems are:
(1) the error analysis and the application of parameter determination using orthogonal projection approximation, (2) decreasing the computational cost of the matching and (3) the treatment of the occlusion.

Heuristic pruning in the process of D. P. and imposing penlty on the occluded vertexes encountered in the matching process will be concidered to solve the problems (2) and (3) respectively.

References

1) Kimura, F., Yoshimura, M., Miyake, Y : An Algorithm for Subpattern Matching of Line Patterns. in Proc. 6th Int. Conf. Pattern Recognition, 461-464, October,(1982).

2) Ambler, A. P., Barrow, H. G., Brown, C. M., Burstall, R. M., Popplestone , R. J: A Versatile Computer Controled Assembly System. in Proc. 3rd Int. Joint Conf. on Artificial Intelligence, 298-307,(1973).

3) Barrow, H. G., Popplestone, R. J: Relational Descriptions in Picture Processing. in Machine Intelligence 6, 377-396,(1971).

4) Davis, L. S: Shape Matching Using Relaxation Techniques. IEEE trans, on PAMI, vol. PAMI-1, no. 1, 60-72, January,(1979).

5) Roberts, L. G: Machine perception of three-dimensional solids. in Optical and Electro-optical Information Processing, Tippett et al.(Eds.) MIT Press, Cambrige, Mass., 159-197,(1965).

6) Falk, G: Interpretation of Imperfect Line Data as a Three-Demensional Scene. Artificial Intelligence, vol. 3, no. 2, Summer,(1972).

7) Barrow, H. G., Tenenbaum, J. M., Bolles, R. C. and Wolf, H. C: Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Processing. in Proc. 5th Int. Joint Conf. Artificial Intelligence 659-663, August,(1977).

8) Duda, R., Hart, P: Pattern Classification and Scene Analysis. Wiley, New York,(1973).

9) Fischler, M., Elschlager, R: The Representation and Matching Pictorial Structures, IEEE trans. Comp., no 22, 67-92,(1973).

10) Kimura, F., Hai, T., Yoshimura, M., Miyake, Y: On Extracting Subpattern and Classification of Handprinted Chinese Characters. IECE trans., vol. J64-D, no. 11, 1037-1044, November,(1981).