A STUDY ON THE FEASIBILITY OF A GLOBAL DESCRIPTION OF 3-D OBJECTS IN RANGE DATA FOR RECOGNITION PURPOSE

STEPHEN H.Y. HUNG

National Research Council of Canada
Ottawa, Ontario, Canada K1A 0R8

Abstract: As an alternative to the common approach in 3-D object recognition, the images of an object from all possible viewing position were considered. The collection of images thus acquired formed a global descriptor of an object and was used to identify the object. This study showed that characterizing the image data in certain way allowed easy handling of the large storage requirement and made fast search possible. This approach is indeed feasible and worth further pursuit..

Key Words: 3-D Object Recognition, Global Descriptor, Range Data, Random Problem.

I. Introduction

The common approach in object recognition is to extract local features such as edges, corners and surfaces from the data as a basis for describing an object. The major drawbacks of this kind of approach are the difficulty and large computational effort required to extract the local features and their relations, as well as the complicated processing of the comparison. The combination of these drawbacks usually makes such methods slow, difficult to implement in realtime and, in some cases, they lead to a combinatorial explosion. An alternative approach is to consider an object from all possible viewing positions. The collection of images thus acquired forms the model or a global descriptor of an object and will be used to identify the object.

In this approach, we are actually treating the object recognition as a random problem whose solution requires knowledge of essentially every possible state of a system. Solving such a problem entails memorizing the set of all possible solutions and quickly selecting the best one from the set, given the input data. The goal of this study is to see whether such an approach is feasible for object recognition problems. The focus will be on overcoming the large storage requirement and developing a fast search algorithm.

II. The Basic Assumption

The objects considered in this study are objects that can be isolated, either from the background or from other objects; there is no occlusion.

For simplicity, we will consider only the geometrical description of objects and ignore shading, colour and texture. Thus, range data instead of intensity grey level data will be more suitable for our purpose.

We assume that the scene is scanned with a laser range finder scanner, described in [1], which provides a two-dimensional image of distances Z(x,y) from a zero reference plane (x,y) orthogonal to the line of sight from the scanner. The scanner makes measurements from a point above the scene, but the readings are the actual Z-values obtained using a mechanical synchronization of two scanners (as in Fig. 1). Such an arrangement provides a reference plane that can be set at any distance along the Z-axis. The readings of X,Y and Z are given in millimetres. Since there is no guarantee that the object is on the reference plane, the Z-value cannot be seen as an absolute measurement of the object. Such an assumption has an immediate drawback in that two objects which have the same visible surfaces but are different in height cannot be distinguished if viewed from certain positions (see Fig. 2). However, this is reasonable and matches the real situation.

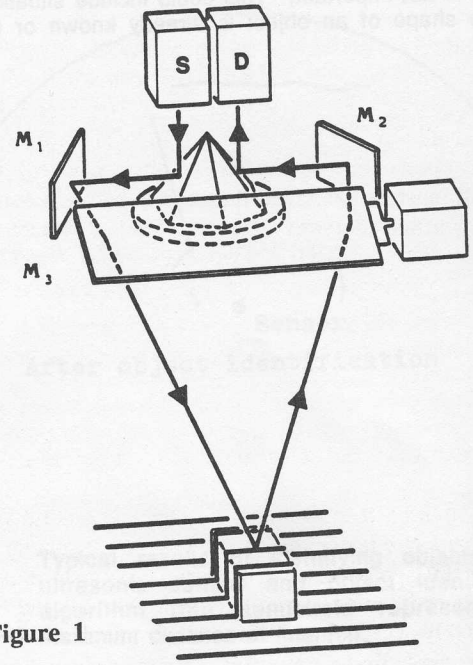

Figure 1

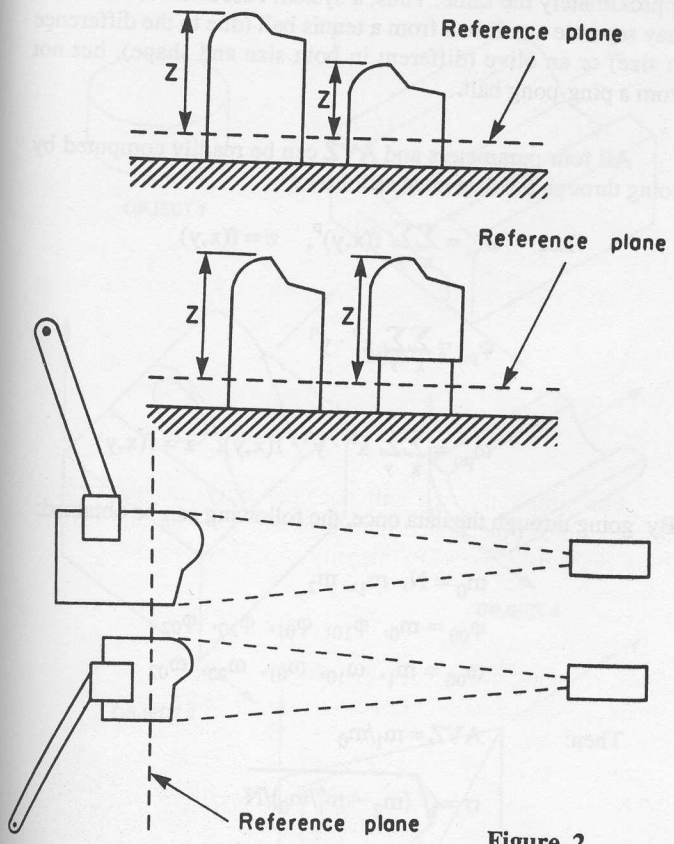

Figure 2

If the difference between two images is an in-plane rotation (along the Z-axis), an in-plane translation (along the X or Y axes) or any combination of these, then the two images are considered to be the same, i.e., all images are invariant under in-plane rotations and in-plane translations, but not when there is a change in scale. Since all measurements are in millimetres, any change in scale should be considered as different images.

We are not intending to build a vision system that can precisely distinguish a particular object from a set of similar objects. Our goal is to examine the feasibility of our approach by building a vision system that can identify instantly a sufficiently unambiguous object in a set of objects that do not resemble each other. It should also overcome one of the weakest points of the traditional approach in which even two obviously very different objects still must be examined thoroughly before a conclusion can be drawn that they are different. Such a system may be simple and trivial; however, if it works properly, it will be quite useful as a preliminary subsystem for a more powerful system in that it can quickly eliminate large numbers of obviously unsuitable candidates.

III. The Features for Characterizing the Images

It is obvious that the current level of general purpose computer hardware is not able to store the large quantity of data required by this approach. Even if we can manage to store all the data sets, there is no easy way to search them all and find the best set quickly. The key point for overcoming such a burden is that are we able to find a simplified method of characterizing an image of an object, designed so that the constraints defining the image can be checked quickly when an observed image is presented. If this can be done in some way, then only the features that characterize the images, rather than the images, must be stored; the searching among the features is certainly much easier than searching among the image data.

There are many features which can be chosen for characterizing the images to any degree of precision. In our case, representing an image precisely is not necessary. Since each object will be described by the collection of images from all possible viewing angles, if two different objects are similar from one viewing angle, then we can look at them from another angle rather than relying on the precision to separate two objects from one angle. Thus, we only need to characterize an image to "approximately right", but the chosen features must allow the easy development of a fast search algorithm.

The features thus chosen are the approximate size and shape of the silhouette, the histogram of Z and the distribution of Z in the (x,y) plane.

A. The size of the silhouette: Since all data are sampled at 0.5 millimetre intervals along the X and Y axes, the size of a silhouette is simply the number of points in an image.

B. The shape of the silhouette: We are not interested in matching the actual shapes, but distinguishing the obviously incompatible ones. We will not extract the shape from the data, but use a parameter that is associated with a shape to characterize the shape of the silhouette. The parameter chosen is the first moment invariant in Refs. 3 or 4. It is IV2D:

$$\phi_{pq} = \sum_x \sum_y (x - \bar{x})^p \cdot (y - \bar{y})^q \quad \text{(central moment)}$$

where $\bar{x}$ and $\bar{y}$ are the mean values of x and y, respectively

$$\psi_{pq} = \phi_{pq}/\phi_{00}^{\gamma}, \text{ where } r = \frac{1}{2}(p+q) + 1$$

(normalized central moment)

$$IV2D = (\psi_{20} + \psi_{02})$$

C. The standart deviation of Z-values: The histogram of Z-values is a good indicator of how volatile the visible surfaces are. Obviously, the best parameter to describe the histogram is its standard deviation, i.e.,

$$\sigma = \sqrt{\frac{\sum_x \sum_y (z - \bar{z})^2}{N}}, \quad z = f(x,y)$$

where: $\bar{z}$ is the mean value of $z = f(x,y)$,
N is the number of points.

Since the standard deviation is the square root of the central moment of the second order of Z-values, it is a 3-D quantity that is invariant under in-plane rotations and translations as well as

translations along the Z axis, i.e., it is independent of the height of the object. This is very important to us. Otherwise, it will be difficult to associate an observed image with the possible candidates in the library of models, because they may be different in height.

D. The distribution of Z-values in the (x,y) plane: This is a good descriptor of the visible portion of a 3-D object at a certain viewing angle, i.e., a 3-D image. The quantity used to represent the distribution of Z is the first invariant of moments in Refs. 3 or 4 (indicated as IV3D), namely:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p \cdot (y - \bar{y})^q \cdot f(x,y), \quad z = f(x,y)$$

(central moment)

$\bar{x}$ and $\bar{y}$ are the mean values of x and y, respectively

$$\eta_{pq} = \mu_{pq}/\mu_{00}^{\gamma}, \quad \text{where } \gamma = \frac{1}{2}(p+q) + 1$$

(normalized central moment)

$$IV3D = (\eta_{20} + \eta_{02})$$

However, IV3D is not independent of the Z-values. The comparison of IV3Ds of two images will be meaningful only if the two images are the same distance from the reference plane. However, this does not become a problem. Since if IV3D of an image at a certain height, AVZ_0, is known, then IV3D of this image at another height, AVZ_1, can be computed as follows:

$$IV3D_1 = \frac{AVZ_0^2 \cdot IV3D_0 + Z' \cdot IV2D_0}{AVZ_1^2}$$

where: $IV3D_1$ = IV3D of the image at the new height AVZ_1
$IV3D_0$ = IV3D of the image at the original height AVZ_0
$IV2D_0$ = IV2D of the image at the original height AVZ_0
$Z' = AVZ_1 - AVZ_0$ (assuming that $AVZ_1 > AVZ_0$)

So the IV3D of an image at any height can be readily obtained. Thus, the IV3Ds can always be compared at the same height. (The IV2D is a special case (2-D version) of IV3D in which the Z-values are always equal to 1.)

Because the values of IV2D and IV3D are too small, we multiply them by 1000. Thus, the four parameters for characterizing an image are denoted as:

P_1 = Number of Points of the Silhouette (No.Pt.)

P_2 = IV2D × 1000.0

P_3 = σ, Standard Deviation of Z–values (SDV)

P_4 = IV3D × 1000.0

The images can be characterized by the combination of them. This means that if two images are approximately the same in all aspects considered, then they should actually be approximately the same. Thus, a system based on such features may separate a golf ball from a tennis ball (due to the difference in size) or an olive (different in both size and shape), but not from a ping-pong ball.

All four parameters and AVZ can be readily computed by going through the data once as follows:

Let: $m_p = \sum_x \sum_y f(x,y)^p, \quad z = f(x,y)$

$$\varphi_{pq} = \sum_x \sum_y x^p \cdot y^q$$

$$\omega_{pq} = \sum_x \sum_y x^p \cdot y^q \cdot f(x,y), \quad z = f(x,y)$$

By going through the data once, the following can be obtained:

$m_0 = N, \; m_1, \; m_2$

$\varphi_{00} = m_0, \; \varphi_{10}, \; \varphi_{01}, \; \varphi_{20}, \; \varphi_{02}$

$\omega_{00} = m_1, \; \omega_{10}, \; \omega_{01}, \; \omega_{20}, \; \omega_{02}$

Then: $AVZ = m_1/m_0$

$$\sigma = \sqrt{(m_2 - m_1^2/m_0)/N}$$

$$\phi_{20} = \varphi_{20} - \varphi_{10}^2/\varphi_{00}$$

$$\phi_{02} = \varphi_{02} - \varphi_{01}^2/\varphi_{00}$$

$$\mu_{20} = \omega_{20} - \omega_{10}^2/\omega_{00}$$

$$\mu_{02} = \omega_{02} - \omega_{01}^2/\omega_{00}$$

$$\mu_{00} = \omega_{00}$$

IV. Establish The Library of Models

Because our equipment does not allow us to scan an object from all the possible angles as desired, in this study, the library of models is actually built with the simulating data that were generated by a graphics package [2] especially written for this purpose. Five objects have been chosen for this experiment (Fig. 3). Each object is analyzed manually to discover all the necessary input data for the computer and also scanned by the scanner at some different angles. The scanning data will be used to calibrate the models and used as input data.

All objects that are intended to be handled by the system are sampled at nominal 5° intervals of rotation along both the X and Y axes. The end result of N° rotation along the X axis first and then M° rotation along the Y axis is different from that of M° rotation along the Y axis first and then N° rotation along the X axis. We always rotate along the X axis first and then along the Y axis.

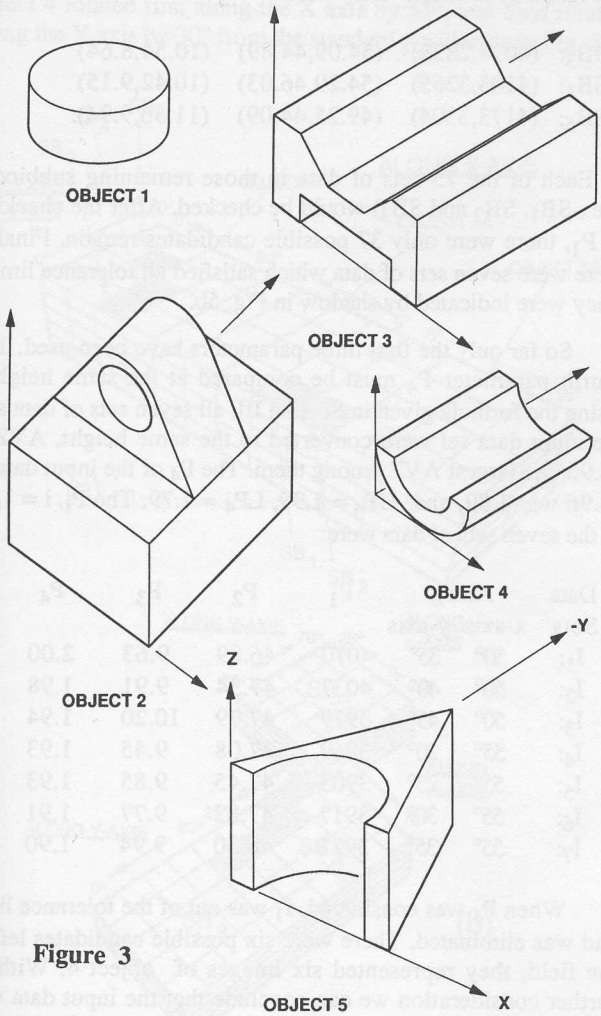

Figure 3

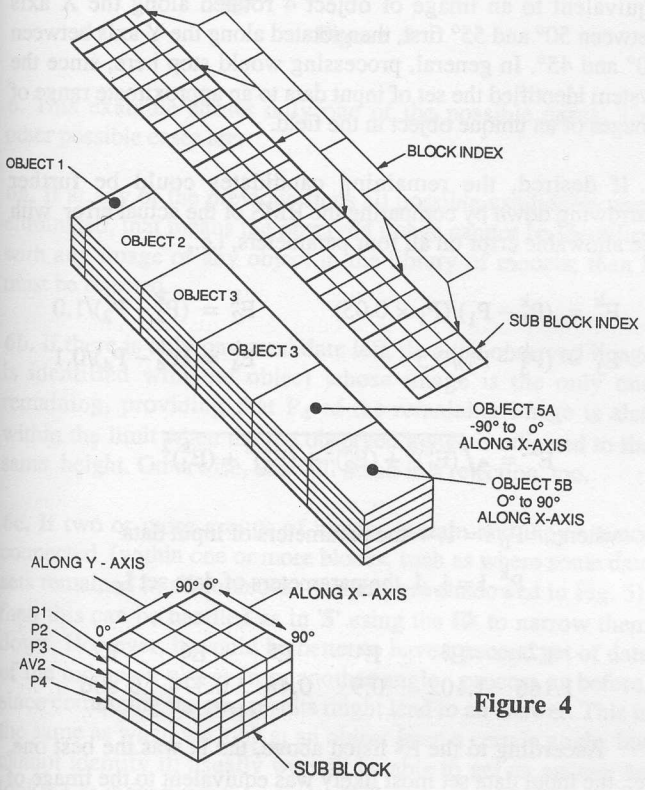

Figure 4

Because of symmetry, object 1 is only rotated along the X axis from 0° to 90°. For objects 2, 3 and 4, we do not consider the bottom-up cases, so both of them are modeled from 0° to 90° along both X and Y axes. Object 5 will be treated as two objects: first from –90° to 0° along both X axis and 0° to 90° along the Y axis; then from 0° to 90° along the X and Y axes.

Each image thus obtained will be gone through once to compute the P_i, i = 1, 4, and AVZ and then the image data will be discarded and each set of these parameters is the model of the image of the object from a particular viewing position. The four parameters of an object and its AVZ will be stored as a three-dimensional matrix (called a block) and each parameter will occupy one layer (see Fig. 4). Henceforth, such a block is the model of the corresponding object. The collection of all blocks is the library of models.

In order to facilitate the search, the maximum and minimum values of each of the first three parameters will be stored and combined as a "block index". For example, the following are the block indices of the five objects:

	BLOCK INDEX					
Block	UBI_1^j	LBI_1^j	UBI_2^j	LBI_2^j	UBI_3^j	LBI_3^j
1	4381	2101	57.51	39.79	7.82	0.00
2	13020	7398	50.98	40.89	12.44	5.20
3	12426	5053	60.87	41.68	15.44	0.00
4	5059	1233	100.97	41.67	11.70	0.00
5	6184	2906	78.83	43.77	12.51	0.00
6	6188	2906	78.83	43.74	13.71	0.00

where: $UBI_i^j = \text{Max}\{P_i \text{ of Block } j\}$

$LBI_i^j = \text{Min}\{P_i \text{ of Block } j\}$

Each block is further divided into 16 subblocks and the maximum and minimum values of these three parameters in each subblock will also be stored and then combined as the "subblock index".

All the processing described in this section is done offline and only the blocks of parameters, the block index and subblock indices are stored in the database. Since there are no image data to be stored the memory requirement is no longer a problem.

V. The Basic Principle of the Searching Algorithm

All four parameters are associated with the images but do not uniquely determine the images. For images with the same or similar parameters they might or might not be the same or similar. However, if they are the same or similar, then their parameters should be approximately the same. Consequently, if they are quite different in any one of their parameters, then they certainly cannot be the same or even similar. This is the principle on which our searching will be based, i.e., we cannot use this set of parameters to pick out the right candidates; all we can do is

use it to eliminate large numbers of obviously unsuitable candidates quickly.

When a set of observed data of an unknown object is fed into this system as input data, the image data will be gone through once to compute the parameters. The search procedure will start from here and will be described best by an example as follows:

The set of input data was obtained by rotating object 4 from the standard position (as shown in Fig. 3) 45° along the Y axis first and then 45° along the X axis. We were trying to find the equivalent image produced by rotating along the X axis first and then along the Y axis for this set of input data. After going through the data once, the following parameters were obtained for this set of data:

P_1	P_2	P_3	P_4	AVZ
3937	47.67	9.70	1.89	24.92

1. According to the possible error of the sensor, each parameter was given a tolerance limit as follows:

For P_1: ± 5%, $UP_1 = P_1 \times 1.05$, $LP_1 = P_1 \times 0.95$

For P_2: ± 1.0, $UP_2 = P_2 + 1.0$, $LP_2 = P_2 - 1.0$

For P_3: ± 0.5, $UP_3 = P_3 + 0.5$, $LP_3 = P_3 - 0.5$

For P_4: ± 0.1, $UP_4 = P_4 + 0.1$, $LP_4 = P_4 - 0.1$

where UP_i and LP_i are the upper and lower limits of P_i

In this example, the UP_i and LP_i, i = 1, 2, 3 were obtained as:

	P_1	P_2	P_3
UP_i =	4133	48.67	10.20
LP_i =	3740	46.67	9.20

2. After the tolerance limits were set, the first three parameters (i.e., P_1, P_2 and P_3) of the observed data were compared with the block index first. Those blocks whose index of any one parameter was out of the limit, that means:

$$R_i^j = \{r \mid UBI_1^j \geq r \geq LBI_i^j\},\ S_i = \{s \mid UP_i \geq s \geq LP_i\}$$

$$R_i^j \cap S_i = \varnothing,\ \text{for any } i = 1, 2, 3$$

were discarded immediately. Blocks whose indices were all within the limit were kept for further consideration.

In this example, Blocks 1, 2 and 3 were rejected quickly because one or more block indices were out of the limits. The possible candidates were narrowed down to object 4 or object 5.

3. A similar procedure was carried out in the subblock level. In Blocks 5 and 6 there were only two subblocks for which the P_1 was within the tolerance limit of the input data as indicated by shadow in Fig. 5a. Their subblock indices were:

SB_1 (in Block 5): (5058,2906) (78.83,56.74) (7.80,0.00)
SB_2 (in Block 6): (5056,2906) (78.83,56.72) (8.87,0.00)

But both were failed in the other two parameters and were eliminated. In Block 4 there were three subblocks; their subblock indices were all within the tolerance limits as indicated in Fig. 5a by shadow. They were:

SB_3: (4098,2838) (54.09,44.89) (10.54,8.64)
SB_4: (4103,3265) (54.29,46.03) (10.42,9.15)
SB_5: (4173,3304) (49.25,44.09) (11.56,9.94)

4. Each of the 75 sets of data in those remaining subblocks (i.e., SB_1, SB_2 and SB_3) would be checked. After the checking of P_1, there were only 37 possible candidates remain. Finally, there were seven sets of data which satisfied all tolerance limits. They were indicated by shadow in Fig. 5b.

So far only the first three parameters have been used. The fourth parameter P_4 must be compared at the same heights. Using the formula given in Section III, all seven sets of data and the input data set were converted to the same height, AVZ = 24.96, the largest AVZ among them. The P_4 of the input data at 24.96 was 1.89, and $UP_4 = 1.99$, $LP_4 = 1.79$. The P_i, i = 1, 4, of the seven sets of data were:

Data Sets	Along x-axis	Along y-axis	P_1	P_2	P_3	P_4
I_1:	50°	35°	4070	46.89	9.63	2.00
I_2:	50°	40°	4037	47.34	9.91	1.98
I_3:	50°	45°	3979	47.79	10.20	1.94
I_4:	55°	20°	3869	47.08	9.45	1.93
I_5:	55°	25°	3903	47.45	9.85	1.93
I_6:	55°	30°	3917	47.82	9.77	1.91
I_7:	55°	35°	3922	48.30	9.94	1.90

When P_4 was considered, I_1 was out of the tolerance limit and was eliminated. There were six possible candidates left in the field, they represented six images of object 4. Without further consideration we can conclude that the input data was equivalent to an image of object 4 rotated along the X axis between 50° and 55° first, then rotated along the Y axis between 20° and 45°. In general, processing would stop here, since the system identified the set of input data to an approximate range of images of an unique object in the field.

5. If desired, the remaining candidates could be further narrowing down by comparing the RMS of the actual error with the allowable error on all four parameters, i.e.,

$$E_1^k = (P_1^k - P_1)/(P_1 \times 0.05) \qquad E_2^k = (P_2^k - P_2)/1.0$$

$$E_3^k = (P_3^k - P_3)/0.5 \qquad E_4^k = (P_4^k - P_4)/0.1$$

$$E^k = \sqrt{(E_1^k)^2 + (E_2^k)^2 + (E_3^k)^2 + (E_4^k)^2}$$

where: P_i, i = 1, 4 the parameters of input data

P_i^k, i = 1, 4 the parameters of data set I_k

E2	E3	E4	E5	E6	E7
1.166	1.102	0.9	0.48	0.374	0.806

According to the E^k listed above, the I_6 was the best one, i.e., the input data set most likely was equivalent to the image of

object 4 rotated first along the X axis by 55°, and then rotated along the Y axis by 30° from the standard position as in Fig. 3.

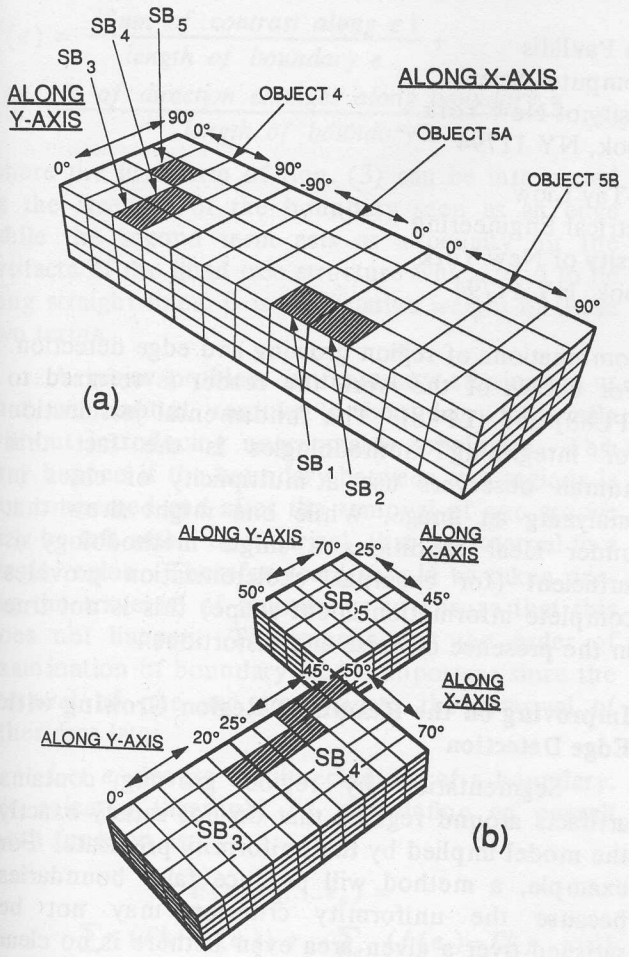

Figure 5

6. This example shows only one of the possible cases. The other possible cases are:

6a. If at any of the previous steps all possible candidates were eliminated, that means the observed image cannot be identified with any image of any object in the library of models; then it must be rejected.

6b. If there is only one candidate left, then the observed image is identified with the object whose image is the only one remaining, providing that P_4 of the remaining image is also within the limit when the the observed image is adjusted to the same height. Otherwise, this will result in a rejection, too.

6c. If two or more groups of images remain and they are not connected (within one or more blocks, such as where some data sets remained in all subblocks which were shadowed in Fig. 5), then this can be handled as in '**5**' using the E^k to narrow them down. However, it would be better to have a second set of data of the unknown object from another angle, process as before, since combining the two results might lead to an answer. This is the same as when we look at an object from a certain angle, but cannot identify it; usually we will be able to get an answer by looking at it from another angle. If "looking at the object from other viewing angles" still cannot narrow down the choice, then the limitation of such an approach is reached and another method must be used to get a clear answer. Even so, our approach has fulfilled its mandate to eliminate large numbers of obviously unsuitable candidates. Any method taking over from here must face only a very few possible candidates that are still left in the field.

VI. Comments and Discussion

We have only considered four different aspects of the images of a 3-D object; the choice of parameters for representing those features is also very preliminary. It is certain that much must be done before such an approach will become practically useful. However, we did show that such an approach is indeed a feasible one and worth further consideration. The key way for such a system to improve its ability is also quite clear; the limitation and the efficiency of the system are determined by the precision of the scanner. If the possible error of the scanner is less than ±1%, then the tolerance limits can be set at ±1% and ±2%. Firstly, the elimination of unsuitable candidates will be much faster; secondly, such a system will be able to distinguish two images with any parameter differing by more than 2%. In such a case, we might be able to separate a ping-pong ball from a golf ball because of the slight difference in size or the small concave patterns on the surface of the golf ball.

One of the advantages of this approach over the traditional method is in the way the computational effort can be allocated. Due to the way we choose to describe the images of objects, the major part of the computation can be done offline. Online computing is kept to a minimum and simple; this is a major concern for making a realtime operational system. Considering a method of traditional approach: even it extracts only the major surfaces and then uses the adjacency or other relation between the chosen surfaces to characterize the object. It still must divide its effort almost equally between building the model and online processing of observed data. Usually the latter will require more attention, since extracting local features from the unknown observed data tends to be more difficult than extracting from a known object (when building the model).

This investigation is still ongoing. We would like to build a workable system based on this approach, to work as a preliminary subsystem for a more sophisticated system to eliminate the major portion of unsuitable candidates and allow the main system to concentrate its effort on a very few finalists.

VII. References

1. Rioux, M., "Laser Range Finder Based on Synchronized Scanners". Appl. Opt. 23(21): 3837-3844; 1984.
2. Hung, S.H.Y., "A Graphics Package for Simulating the Data of Range Finders". Proceeding of this Conference.
3. Hu, M.K., "Visual Pattern Recognition by Moment Invariants". Trans. Inform. Theory, Vol. IT-8: 179-187, Feb. 1962.
4. Dudani, S.A., Breeding, K.J. and McGhee, R.B., "Aircraft Identification by Moment Invariants". IEEE Trans. Compt. Vol. C-26(1): 39-45, January 1977.