APPLICATION OF SHAPE RECOGNITION ALGORITHMS FOR A ROBOTIC TEST BED

ALAN SANTUCCI

FEBRUARY 1989

U. S. ARMY ARMAMENT RESEARCH, DEVELOPMENT AND ENGINEERING CENTER
FIRE SUPPORT ARMAMENT CENTER
PICATINNY ARSENAL, NEW JERSEY

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.
The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

The citation in this report of the names of commercial firms or commercially available products or services does not constitute official endorsement by or approval of the U.S. Government.

Destroy this report when no longer needed by any method that will prevent disclosure of contents or reconstruction of the document. Do not return to the originator.
APPLICATION OF SHAPE RECOGNITION ALGORITHMS FOR A ROBOTIC TEST BED

Machine vision has become an important area of research in the application of robotics to commercial and military systems. Progress in this area offers an economic benefit to the automated factory as well as the automated battlefield of the future. The algorithms discussed here are the baseline technologies that will be expanded to provide real time machine vision capability.
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Discussion</td>
<td>1</td>
</tr>
<tr>
<td>Shape Recognition Algorithm</td>
<td>1</td>
</tr>
<tr>
<td>Moments</td>
<td>2</td>
</tr>
<tr>
<td>Fourier Descriptors</td>
<td>4</td>
</tr>
<tr>
<td>Polygonal Approximation</td>
<td>6</td>
</tr>
<tr>
<td>Communications</td>
<td>9</td>
</tr>
<tr>
<td>Calibration</td>
<td>9</td>
</tr>
<tr>
<td>Conclusion</td>
<td>10</td>
</tr>
<tr>
<td>References</td>
<td>11</td>
</tr>
<tr>
<td>Distribution List</td>
<td>13</td>
</tr>
</tbody>
</table>
INTRODUCTION

Machine vision is an integral part of any robotic environment. The shape recognition algorithms, implemented on a Vicom image processor, and interfaced to a flexible robotic testbed, must be fast, accurate, and rotationally invariant. The Vicom image processing system contains special purpose hardware and software that is used in the image processing task. The shape recognition algorithms were used to provide object identification, position, and orientation data. This position and orientation data are then passed to the robot controllers by means of the system supervisor.

The elements of the shape recognition system have been implemented in a modular fashion. This method allows maximum flexibility when configuring the system to handle various shape recognition tasks. This approach will also simplify future expansion as more shape recognition techniques are added to the library. This modular approach will also be used in the optimization of the vision supervisory system which uses a set of rules to decide upon the most suitable combination of modules for a given task.

DISCUSSION

Shape Recognition Algorithm

A machine vision system can be broken down into the following series of subtasks.

```
   digitization  pre-processing  segmentation
                 |               |
                 |               |
                 v               v
    feature extraction  classification
```

The first subtask of the system is digitization of the video data. This is standard RS-170 output from a CCD camera. The Vicom uses its special purpose hardware to digitize the video into a 512 x 512 pixel frame with 256 gray levels. These gray level data are then stored for use as the original state of the workspace. A subsequent frame of video
is then sampled and digitized and used for a comparison to determine if the workspace is in a static or dynamic state. This is done by subtracting the second frame from the first and performing a linear transformation about the zero gray level. If the number of changed pixels is below a given threshold, then the workspace is said to be static and the frame of data is passed to the segmentation subtask.

In the segmentation subtask, the digitized gray level data are transformed into a binary image which consists of objects and background. The procedure developed for this purpose is a histogram based thresholding technique. It is based on the assumption that the background intensities are concentrated in one portion of the gray scale range while the object intensities are concentrated in another region. A histogram of the gray level image is obtained quickly by use of a Vicom special purpose image processing routine. A threshold is determined by locating the valley between the two nodes of the histogram. This value is then used by another Vicom library call to apply the threshold to the image. This function will set all pixels to 0 while setting all object pixels to 255. After the objects have been separated from the background, the number of pixels that represent objects are counted and checked against a given range. If the pixel count is outside this range, the data are considered invalid and the sampling procedure is restarted. After the objects have been separated from the background, they are then separated from each other.

At this point in the system, there are three different methods of feature extraction: low order moments, polygonal approximation for feature extraction and Fourier descriptors for classification, and a polygonal approximation for feature extraction and classification.

Moments

Up to third-order moments can be used to determine area, center of mass, and aspect ratio (refs 1 through 4). The basic moment equation is

\[
M_{pq} = \sum_{i=0}^{N} X_i^p Y_i^q f(X_i, Y_i)
\]

where

\[
f(X_i, Y_i) = \begin{cases} 1 & \text{if and object} \\ 0 & \text{if it is not an object} \end{cases}
\]

the midpoint of the object is located at (a,b) where

\[
a = \frac{M_{10}}{M_{00}} \quad b = \frac{M_{01}}{M_{00}}
\]
The orientation of the object may be represented by the angle of its major axis which is calculated by the central moments of the object (refs 3, 5, and 6). These are the moments for which the centroid is zero. The central moments may be obtained from the basic moments by a translation of the object by the midpoint (a,b). This can be performed directly on the moments as follows:

Letting \( m \) be the central moments

\[
 m_{pq} = \sum_{r=0}^{p} \sum_{s=0}^{q} (\begin{pmatrix} p \\ r \end{pmatrix}) (\begin{pmatrix} q \\ s \end{pmatrix}) a^{p-r} b^{q-s} M_{rs}
\]

the angle of the major axis may be calculated (refs 5 and 6)

Letting \( t \) be the angle

\[
 \tan (2t) = \frac{2 M_{11}}{M_{20} - M_{02}}
\]

However, this angle is ambiguous. The inverse tangent generates one ambiguity of 180 degrees. A second ambiguity arises because one may rotate an object the wrong way to line it up with a given line. These two rotations also differ by 180 degrees. After the factor of two is taken into account in the above expression, the result is four possible angles \( t, t+90, t+180, \) and \( t+270 \) degrees. Some additional constraints are necessary to assure a unique and consistent orientation. The angle for which

\[
 m_{20} > m_{02} \quad \text{and} \quad m_{30} > 0
\]

will be used. This angle will align the major axis of the object with the \( x \)-axis and force the larger side of an asymmetric object to the right. This approach will only work when the objects under test have only one major axis.

A rotation may be applied directly to a set of moments to determine the rotated moments. Letting \( M' \) be the new moments which result from a rotation by an angle \( t \)

\[
 M'_{pq} = \sum_{r=0}^{p} \sum_{s=0}^{q} (\begin{pmatrix} p \\ r \end{pmatrix}) (\begin{pmatrix} q \\ s \end{pmatrix}) (-1)^{q-s} (\cos t)^{p-r+s} (\sin t)^{q-r-s} M_{p+q-r-s,r+s}
\]

The correct orientation angle may then be found as follows: The initial angle \( t \) is applied to the central moment set. If both conditions are satisfied, then the angle has been found. If the first condition is satisfied, but the second is not, then the correct angle is \( t+180 \) degrees. If the first condition is not met, an additional 90-degree rotation is
applied to the moments. The first condition will now be met, and the second condition is examined. If the second condition is met, then $t+90$ degrees is correct angle; otherwise, $t+270$ degrees is the correct angle.

Position and orientation have been obtained from the calculated moments as described above. The last subtask is the classification of the objects extracted. The objects can be identified using a simple decision tree. The objects in question were shells, rack endpoints, and a cannon breech. These three objects could be distinguished on the basis of area and aspect ratio. Area is obtained directly from $M_{xx}$. The aspect ratio is found as the square root of the ratio of $m_{20}$ to $m_{02}^2$. This simple decision approach is fine for this simple set of objects but is no easily extended.

**Fourier Descriptors**

The second approach to feature extraction is a Fourier descriptor computation (ref 8) based on a polygonal approximation to the contour (ref 9). This method calculates the Freeman chain code (ref 10) of each object that is separated from the background. The chain codes are then handled one at a time. A polygonal approximation of the boundary defined by the chain code is used instead of the actual code itself. The advantages of the polygonal approximation are that the points used to calculate the Fourier descriptors are drastically reduced and the results are less subject to rotation and noise effects. The polygonal approximation algorithm starts by converting the chain code to a sequence of points in the $x$-$y$ plane. This set of points is then split into two open curves. The two most distant points on the contour are located and used for this splitting. The use of these extreme points will introduce some rotational invariance to the initial split. Each of the two open contours is processed individually. A line is drawn between the two endpoints of the curve, and the point on the curve most distant from this line is located. If the distance to this point exceeds a goodness-of-fit threshold, then this point is taken to be a breakpoint. This point and one of the endpoints are taken and processed similarly. This continues recursively until the distance of the line between the breakpoints is no further than the goodness-of-fit threshold from any point on the segment of the curve that the endpoints enclose. The result is a sequence of breakpoints that describe the vertexes of the polygon which approximate the curve.

The goodness-of-fit threshold serves to govern the tightness of the polygonal fit. A small threshold will give a very tight approximation which uses a larger number of segments. The potential drawback is that this representation will be very likely to be affected by the edge artifacts which the earlier stages of image processing may cause. In addition, a greater number of polygonal segments will require more computation in the matching process. A larger threshold, however, will lead to a looser fit and therefore to a potential loss of detail. The larger threshold will not be as sensitive to noisy edge effects. The smaller number of segments which result from the use of a larger threshold will decrease the amount of computation required for each match. Therefore, a trade-
off is encountered between the computation required for a match and the amount of
detail in the description, with sensitivity to edge effects an an additional consideration.

Now that the polygon is formed, it is u-,,I to generate the Fourier descriptors. If the
unknown polygon is defined by Z(t), it can be expanded in a Fourier series:

\[ \lambda_i(t) = \sum_{n=-\infty}^{+\infty} C_n \left( -i \frac{2\pi n}{\tau} \right)^t \]

where

\[ C_n = \frac{1}{\tau} \int_0^\tau \lambda(t) \left( -i \frac{2\pi n}{\tau} \right)^t \, dt \]

if a polygon has K sides. The increment in position generated by each side will be the
complex number \( \Delta \lambda_i \). The position at each point is

\[ \lambda_i = \lambda_{i-1} + \Delta \lambda_i \quad \lambda_0 = 0 \]

the increment in time for each side will be \( \Delta t_i \) where

\[ \Delta t_i = |\Delta \lambda_i| = (\Delta x_i^2 + \Delta y_i^2)^{1/2} \]

the time at each point is

\[ t_p = \sum_{i=1}^p \Delta t_i \quad t_0 = 0 \]

the period \( \tau = t_k \)

\[ C_n = \frac{\tau}{4\pi^2 n^2} \sum_{p=1}^k \frac{\Delta \lambda_p}{\Delta t_p} \left[ e^{-i \frac{2\pi n}{\tau} t_p} - e^{-i \frac{2\pi n}{\tau} t_{p-1}} \right] \]

for \( n \neq 0 \).
\[ C_0 = \frac{1}{\tau} \sum_{p=1}^{k} \left( \frac{1}{2} \Delta \lambda_p + \lambda_{p-1} \right) \Delta t_0 \]

These equations have generated the Fourier coefficients. To normalize the descriptors, \( C_0 \) must be set to zero to normalize orientation of best fitting ellipse.

\[ C_n = \frac{C_n}{|C_0|} e^{i(nt_0 + \alpha)} \]

where

\[ t_0 = \frac{C_1 - C_{-1}}{2} \]

\[ \alpha = -\frac{C_1 + C_{-1}}{2} \]

The unknown must now be compared with the library of known objects. Consider an unknown object \( X \) with Fourier series \( X(t) \) by making use of Parseval's theory for Fourier series

\[ P = \frac{1}{\tau} \int_{-\tau}^{\tau} |X(t)|^2 \, dt = \sum_{n=\infty}^{\infty} |X(t)|^2 \]

if you then compare it to a known library object \( Y \) with a Fourier series \( Y(t) \), then

\[ d^2 = \sum_{i=-N}^{N} |X(i) - Y(i)|^2 \]

where \( d \) is considered to be an approximation to the average power in the error between the two objects. The threshold that the value \( d \) is compared to is the acceptance level of a successful search. If the value exceeds this threshold for each entry in the library, the unknown is considered an undefined object.

**Polygonal Approximation**

The third method of feature extraction is an extension of the polygonal approximation algorithm. After the polygonal approximation has been calculated from the chain code, these data must be stored to build the library of known objects. The polygon will be described by each of the line segments. The lengths of the line segments will be
invariant to both rotation and translation, as will the angles between each pair of line segments. This way the polygon will be described as a sequence of distance and angle pairs. The first entry will be the length of the first segment. The second entry will be the angle between the first and second segments. This angle is obtained by calculating the angle of each of the line segments in the x-y plane. The first angle is then taken to be the angle of the second segment minus the angle of the first segment. A distance/angle pair is then calculated for each segment in the polygon. For the last segment, the angle is the difference in orientation between the first segment and the last segment. In this fashion the polygon may be reconstructed from this description at any location and orientation.

The first step of the shape matching algorithm (ref 11) is to identify a pair of segments on the known and unknown contours at which to start the comparison. This application scale is considered to be constant and an important factor for correct classification. Therefore, both distance and angle features will be used in starting point determination. Each segment of the unknown object is compared to each segment of the known objects to find potential starting point pairs. Thresholds for distance and angle comparisons are set based on the assumption that the known features of the library entries are of high quality. If the difference between the angle for the unknown segment and that for the known segment is within the specified tolerance, then the angle passes the potential starting point test. The difference of the lengths must be within a specified fraction of the length of the known segment to be acceptable. If both the angle and the length are within bounds, then the two segments are considered to be a potential starting point pair. The choice of thresholds determines the number of potential starting points that will be considered. A trade-off is encountered between the amount of computation which will be performed and the possibility of missing a valid starting point pair.

Given a pair of starting segments, the matching process is initiated by reconstructing the polygonal vertexes from the distance/angle descriptors. The starting segment of each object is taken to start at the origin and proceed down the positive x-axis. The first segment has an angle of zero in the x-y plane. After all segments are considered, the known and unknown contours are each represented by a sequence of points. If the objects are similar, these point sequences should also be similar.

The vertexes are then compared using the Euclidean distance between points as an indication of goodness of fit. If there are more than one unknown point matched to a single point of the known contour, then the unknown yielding the smallest distance to the common known point is taken to be matched. The other points are labeled as non-matching. Upon completion, the number of matched points is tallied, as is the sum of the distances between matched pairs of points. This total distance is taken to represent the quality of the match and will be dependent on the total size of the polygon. A
normalized distance measure in terms of a fraction of the total polygon distance will be more accurate. Therefore, the sum of the distances will be divided by the total length of the polygon.

A minimum distance classifier is used to find the best match for an unknown object. The unknown is compared to each known object with the smallest distances recorded. If this distance falls below a threshold value, then the unknown is declared to match the corresponding known object. A problem arises when there is a very small distance on a match over a small portion of polygon. This distance might be smaller than one obtained over a large portion of a polygon simply because it involves fewer points. A measure that will produce a weighted distance (ref 12) will resolve this problem.

\[ d' = d \cdot \left( \frac{1}{T^2} - \frac{1}{T} + 1 \right) \]

After the object has been matched, determining the midpoint is a straightforward operation. The midpoint is calculated by averaging the x, y points of the curve that represent the boundary of the unknown object. Recall that these x, y points were obtained by converting the chain code to the x, y plane.

The calculation of the orientation difference between the known an unknown objects requires the examination of a set of orientation differences, once associated with each pair of matched polygonal sides. Angles for each of the sides of the polygon are calculated. The angles for the unknown are generated assuming that the orientation of the first side is zero. For the known, however, an initial orientation value is used. This initial value is derived from the angle of the first segment of the known polygon when the known object is learned. It is assumed that the orientation of the entire known object is zero, but that does not mean that the orientation of the first segment is zero. The starting angle may be viewed as a constant angle offset which is added to the absolute angle of every side of the polygon, regardless of the side from which the reconstruction was started. For each pair of sides which is declared a match, the difference between the two angles is recorded. A nonparametric clustering algorithm (ref 13) is applied to the collection of angle differences to find the largest group. The values in this group are then averaged to obtain the orientation estimate.

This last method of feature extraction and classification was proven to work well in the object identification and midpoint calculation tasks, but the orientation estimates were not acceptable. It was determined that the problem arose from the polygonal approximation not being accurate enough to distinguish between the two ends of one of the object.
Communications

The communications module is the interface between the Vicom and the top level supervisor. All data to and from the Vicom must pass through this link which is made up of an I/O routine on the Vicom and another compatible routine on the supervisor. The supervisor module sends text strings across an RS-232 serial line that are then parsed into internal Vicom commands that invoke the proper image recognition commands for that particular scenario. After the command has been acted on, the Vicom returns the results of the process. These results will indicate the position and orientation of all objects of interest in the field of view. If the workspace is empty, there are no objects of interest in the field of view. In this manner, the supervisor builds and maintains the world model of the entire robot workspace.

Calibration

Generally, the coordinate system of a vision system and the robot system will not be identical. Differences that may exist are scale, rotation, and position. These differences must be identified and estimated (ref 14) so that a position obtained in vision coordinates may be used as a position for a robot. Also, for robot cooperation, conversions must be made from one robot coordinate system to the other. The physical configuration of the robots, workspace, and cameras will be such that the robots will be placed on a surface that is parallel to the workspace. Also, the image plane of the camera will be parallel to the workspace.

Letting \((x,y)\) represent a point on the image plane, the corresponding point in the workspace \((x',y')\) can be obtained by the transformation.

\[
\begin{bmatrix}
x' \\
y' \\
1
\end{bmatrix} = \begin{bmatrix}
a & b & c \\
d & e & f \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
x \\
y \\
1
\end{bmatrix}
\]

The terms \(c\) and \(f\) describe independent shifts in the \(x\) and \(y\) directions. The terms \(a, b, c,\) and \(d\) combine to represent both independent scaling in the \(x\) and \(y\) directions along with a rotation between the two planes. The calibration procedure requires that these six parameters be estimated. This requires the use of three points, using the corresponding values in both coordinate systems.

The solution is to use three points \((x_1, y_1)\), \((x_2, y_2)\), and \((x_3, y_3)\) and their corresponding points in a second coordinate system \((x'_1, y'_1)\), \((x'_2, y'_2)\), and \((x'_3, y'_3)\). The equations relating the \(x\) coordinates are
\[ X_1' = aX_1 + bY_1 + c \]
\[ X_2' = aX_2 + bY_2 + c \]
\[ X_3' = aX_3 + bY_3 + c \]

which yields

\[
\begin{align*}
    a(X_2 - X_1) + b(Y_2 - Y_1) &= X_2' - X_1' \\
    a(X_3 - X_1) + b(Y_3 - Y_1) &= X_3' - X_1' \\
    a(X_3 - X_2) + b(Y_3 - Y_2) &= X_3' - X_2'
\end{align*}
\]

Using two equations to form a matrix

\[
\begin{bmatrix}
    X_2 - X_1 & Y_2 - Y_1 \\
    X_3 - X_1 & Y_3 - Y_1
\end{bmatrix}
\begin{bmatrix}
    a \\
    b
\end{bmatrix}
= 
\begin{bmatrix}
    X_2' - X_1' \\
    X_3' - X_1'
\end{bmatrix}
\]

which yields

\[
\begin{bmatrix}
    a \\
    b
\end{bmatrix}
= 
\begin{bmatrix}
    X_2 - X_1 & Y_2 - Y_1
\end{bmatrix}^{-1}
\begin{bmatrix}
    X_2' - X_1' \\
    X_3' - X_1'
\end{bmatrix}
\]

the values for \( a \) and \( b \) are substituted back into the original equation to solve for \( c \). A similar solution for \( d, e, \) and \( f \) using the equation relating to the \( y \) coordinates is easily obtained.

This method was implemented on the Vicom by having the robots place blocks in various predetermined places in the workspace to collect all required data.

**CONCLUSION**

The work done to this point includes two fully functional shape recognition methods and a third method that requires additional study. A completely automated calibration system and a complete communications interface with the system supervisor have been installed. Long term plans include additional shape recognition algorithms that will deal with partial shape and distorted contours as well as various methods of collecting and analyzing depth data. In addition to an enhanced image processing capability, an electronic blackboard will be developed using shared memory between the Vicom and the system supervisor. This will enable the Vicom to process the workspaces independently of any external supervisor and then post the results on the blackboard. This will make full use of the idle time that currently exists waiting for commands from the supervisor.
REFERENCES


DISTRIBUTION LIST

Commander
Armament Research, Development and Engineering Center
U.S. Army Armament, Munitions and Chemical Command
ATTN: SMCAR-IMI-I (5)
      SMCAR-FSF-RC (15)
Picatinny Arsenal, NJ 07806-5000

Commander
U.S. Army Armament, Munitions and Chemical Command
ATTN: AMSMC-GCL(D)
Picatinny Arsenal, NJ 07806-5000

Administrator
Defense Technical Information Center
ATTN: Accessions Division (12)
Cameron Station
Alexandria, VA 22304-6145

Director
U.S. Army Materiel Systems Analysis Activity
ATTN: AMXSYY-MP
Aberdeen Proving Ground, MD 21005-5066

Commander
Chemical Research, Development and Engineering Center
U.S. Army Armament, Munitions and Chemical Command
ATTN: SMCCCR-MSI
Aberdeen Proving Ground, MD 21010-5423

Commander
Chemical Research, Development and Engineering Center
U.S. Army Armament, Munitions and Chemical Command
ATTN: SMCCCR-RSP-A
Aberdeen Proving Ground, MD 21010-5423

Director
Ballistic Research Laboratory
ATTN: AMXBR-OD-ST
Aberdeen Proving Ground, MD 21005-5066
Chief
Benet Weapons Laboratory, CCAC
Armament Research, Development and Engineering Center
U.S. Army Armament, Munitions and Chemical Command
ATTN: SMCAR-CCB-TL
Watervliet, NY 12189-5000

Commander
U.S. Army Armament, Munitions and Chemical Command
ATTN: SMCAR-ESP-L
Rock Island, IL 61299-6000

Director
U.S. Army TRADOC Systems Analysis Activity
ATTN: ATAA-SL
White Sands Missile Range, NM 88002