Three-Dimensional Object Recognition: Statistical Approach

Rosalina Abdul Salam
School of Computer Science
University Science of Malaysia
11900 Penang
Malaysia
E-mail: rosalina@cs.usm.my

Marcos Aurelio Rodrigues
School of Computing and Management Sciences, Sheffield Hallam University
City Campus, Howard Street.
Sheffield, S1 1WB, United Kingdom
E-mail: m.rodrigues@shu.ac.uk

ABSTRACT

The design of a general purpose artificial vision system capable of recognizing arbitrarily complex three-dimensional objects without human intervention is still a challenging task in computer vision. Experiments have been conducted to test the ability of incorporating the knowledge of how human vision system works in a three-dimensional object recognition system. Firstly, the process of shape outline detection and secondly, the use of multiple viewpoints of object. Shape outline readings are put through normalization and dimensionality reduction process using an eigenvector based method to produce a new set of readings. Through statistical analysis, these readings together with other key measures, namely *peak measures* and *distance measures,* a robust classification and recognition process is achieved. Tests show that the suggested methods are able to automatically recognize three-dimensional objects from multiple viewpoints. Finally, experiments also demonstrate the system invariance to rotation, translation, scale, reflection and to a small degree of distortion.

Keywords
Shape outline, viewpoint dependent, multiple viewpoints, classification, recognition, three-dimensional.

1. INTRODUCTION
How human recognize objects is always a question to anybody in this world and among researchers, this is a very huge area to explore. There is more information on how our visual system works compared to how our brain functions. The functionality of our brain is not very clear and there are a number of disagreements between different research groups. One of an example is how objects are represent in our brain. There are two main approaches that are the viewpoint dependent and the viewpoint independent. The viewpoint dependent approach was chosen in the research conducted.

Visual processing encompasses one of the most complex operations performed by living organisms. Research in computer vision, uses knowledge of how biological vision works to emulate its performance on computers. Humans are capable of recognizing objects regardless of their size, variations in shape, position and orientation. Endowing machines with similar recognition behavior is a very difficult task acknowledged in the computer vision literature.

The first part of this research is based on our early vision system. Early vision system plays an important part of our earlier stage of perception. Our studies were inspired by the front end visual system. At this stage the basic visual information that is the edges is available for perception. The visual information is then carry for further processing in our extra-striate visual cortex. Recognition and motion processing happen at this stage. Zenon Pylyshyn [Ply01a], in his paper concluded that the output of early vision system consists of shape representations involving at least surface layouts and occluding edges.

Perceptions are partly based on our visual system as well as knowledge and expectations. Different viewpoints of objects are important aspects of our recognition system. A view-point dependant representation consists of multiple models of the same object, each corresponding to a different set of views. Shape outline of different viewpoints of a three-dimensional object can be sufficient to use as a

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WSCG Poster Proceedings
WSCG'2003, February 3-7, 2003, Plzen, Czech Republic.
Copyright UNION Agency – Science Press

measure for a recognition system. The use of shape outline is not a new idea and it has shown a significant results for a recognition system [Bie01a], [Hay01a], [Tay01a], [Roc01a], [Hab01a], [Cro01a].

Object representation is another main task in this research. Studies and experiments done by a number of researchers demonstrate that the human visual system is based on multiple viewpoints [Ull01a], [Ede01b], [Hay01b], [Beu01a], [Low01a]. This research has thus been motivated by the human biological visual system, which uses multiple views in recognizing objects.

This paper reports results of experiments to recognize objects through multiple viewpoints. It begins with the features extraction that is, the shape outline, which go through a normalization and dimensionality reduction process using an eigenvector based method to produce a new set of readings. Through statistical analysis, together with other key measures, namely *peak measures* and *distance measures*, a robust classification and recognition process is achieved. Tests show that the suggested methods are able to automatically recognize three-dimensional objects from multiple viewpoints.

2. METHODS AND ASSUMPTIONS

2.1.1 Shape Outlines

An edge following technique was used for acquiring shape outline readings and storing them in a list format. The initial point of the outline is determined by firing a number of simulated range finders sensors from random positions at the border of the display window pointing to its centre until a point on the outline is encountered. As soon as an object is encountered by at least two nearby simulated sensors the pixel co-ordinates *(x,y)* will be returned.

In the next stage, three virtual sensors are configured. These three sensors are configured to be at least two pixels apart. These three sensors follow the object's outline, recording a list of *(x,y)* positions. The first thing that these three sensors will do is to rotate until the next reading is obtained. All three sensors must hit the shape for valid readings.

Rotation angles for all three sensors are recorded. The rotation angle θ can be computed as:

$$\theta = \frac{1}{3}\int_0^3 \theta_i$$

$$\theta_i = \tan^{-1}(\Delta y_i / \Delta x_i)$$

The above procedure is repeated until the initial point, that is $(x_0 y_0)$ is reached, where this process will be automatically stopped. The ordered list of all angle rotation values (the average), that is, the n measurements, θ , is constructed as:

$$\theta = [\theta_1, \theta_2, ..., \theta_n]$$

$d\theta$ the difference from one angle to the next one. The list of the differential angles $d\theta$ will go through a process of dimensionality reduction and normalization before being used for training and testing.

2.1.2 Normalization and Dimensionality Reduction

Outline readings went through a process of transformation which involved normalization and dimensionality reduction. This transformation used eigenvector based methods. The list of $d\theta$ described earlier is computed during the feature extraction process, is further filtered by calculating the average of every three readings. Each current value is substituted with the average reading. Therefore, a new set of $d\theta$ is obtained. Let the list of $d\theta$ be transformed into list of vectors **V**, where **V** = $\{v_1, v_2, \ldots, v_n\}$. The new co-ordinates after the transformation can be constructed as follows:

$$C = E^T V$$

Where C is the new set of co-ordinates after the transformation, V is the set of vectors computed from rotation angles and E is the set of eigenvectors. The eigenvector e_i is computed as:

$$e_i = \begin{pmatrix} k_x(x_0 + i) \\ k_y\left(\dfrac{d\theta_i}{|d\theta_{\max}|}\right) \end{pmatrix}$$

where x_0 is 0 and $|d\theta_{\max}|$ is the largest absolute value in the list. k_x and k_y are arbitrary constant factors in the x and y axis. These constants are determined experimentally and play a very important role in determining the new co-ordinates.

Normalization is carried out on C, where three of the values are added up and the average obtained. The new set of readings after normalization is Z, that is $Z = \{z_1, z_2, ..., z_n\}$ and represents the new set of vectors.

2.1.3 Multiple Views

As soon as new readings is obtained, it is stored in a database. There are a number of different views for each object. Multiple views of objects were acquired assuming that the objects were rigid body objects, although these assumptions can be relaxed at higher computational costs. The mean and the standard deviation for each view is kept as follows:

$$view_i = [[\mu_1, \sigma_1], [\mu_2, \sigma_2], ..., [\mu_{120}, \sigma_{120}]]$$

Where μ is the mean and σ is the standard deviation. These means and standard deviations form a single view of an object and basically is an outline of a shape. Recognizing an object from different views, require the grouping of available views. This can be done by using the *distance measure.*

2.1.4 Distance Measures

The *distance measure* technique used in this paper is similar to the approach of (Edelman, 1997), that is, representation by similarity. However, it is different in concept and implementation since different objects with similar features were grouped together, while in this paper, a *distance measure* was used to group multiple views on an object. The distance measure is calculated as follows:

$$D = 100 - \tfrac{1}{n}\sum_{i=1}^{n} C_i$$

where D is the *distance measure,* n is the number of points in the view and $C_i (0 \le C_i \le 100)$ is the measure of the confidence interval for feature i where $i = \{1,2,...,n\}$. *Distance measure* can also be used between different objects. If the objects are different, the *distance measure* will be higher. A value, 5 that was based on experiments, was assigned for differentiating one view from another. This *distance measure* will be accepted if its value is equal to or less than 5.

2.1.5 Pattern Matching

Readings, obtained from the earlier stage, were subject to statistical analysis, through the use of the *z*-scores method for the classification of each point in the list. Matching was accomplished together with the peaks and distance measures for more accurate results.

Assuming that the list of points of each signature is normally distributed [Mul01a] and since it is a continuous probability density function, the probability that a point y lies between two specified values a and b of a point in the database is given as follows:

$$\Pr(a \le y \le b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy$$

where y can assume all values from $-\infty$ to $+\infty$ and the parameter μ and σ represent respectively the mean and the standard deviation of the distribution. The above equation can be simplified [Mul01a] by a transformation:

$$z_i = \frac{y_i - \mu}{\sigma}$$

where z is the z-scores of observation y_i. The most suitable value for z was determined based on the results of the experiments. The value of z was between -1.96 and 1.96, that is 5 per cent of the distribution (2.5 percent on each side). If z lies outside this range, then the point is rejected.

3. EXPERIMENTAL RESULTS

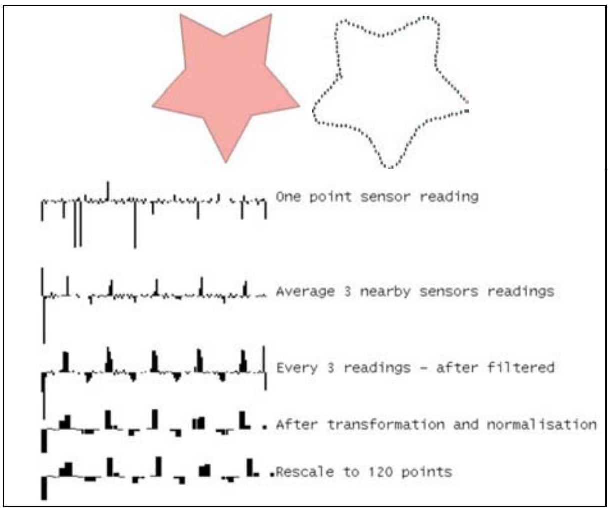

Figure 1. Results of a simple object and the reconstruction of the shape outline.

Experiments were conducted to test the shape outline reading on set of objects. Shape outline taken from a simple object and the reconstruction of the shape outline can be seen in Figure 1. The graph shows the results of the outline readings for normalization and dimensionality reduction.

Further experiments were conducted to investigate that the system is invariant to rotation, translation, size and reflection and to a certain degree of distortion. Results obtained from the test for 15 objects rotated at 30 degrees, shows that the method used is invariant to rotation. The accuracy level for all objects were above 95%. Same results were obtained for testing the invariance in sizes and on a small degree of distortion and translation. Mirror effect of an object can easily be created by using the reverse list. This is a very useful feature for a recognition system.

Further experiments were conducted on three-dimensional objects. Each object consists of at least thirty to forty different views. Figure 2 shows the *distance measure* for an object called toy boat.

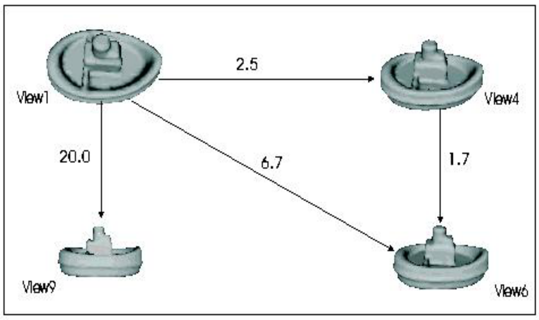

Figure 2 Distance measure between $View_1$ towards $View_4$, $View_6$ and $View_9$.

The *distance measure* between $View_1$ and $View_6$ is 6.7 that is greater than 5, however it is classified as the same object as $View_1$. This is because the *distance measure* between $View_4$ and $View_6$ is 1.7. Therefore, there exists a link between $View_1$ and $View_6$. When $View_6$ was introduced, the distance measure was also calculated for each other view. This new view also helped to establish a relationship between $View_1$ and $View_9$.

4. CONCLUSION AND FUTURE RESEARCH

Experiments showed that the system is invariant to rotation, translation, size, mirror effect and to a certain degree of distortion. The system is capable of recognizing three-dimensional objects from different views. The distance measure allows us to group different views of an object, so they will be classified as the same object. The system is capable of grouping new views that are views that do not belong to any existing category by putting it into a new category. This is achieved without any human interference.

In this paper, the shape outline reading, which shows relationship to biological human vision and cognition was chosen as the main feature. This can be extended by incorporating other features, such as depth information, color and texture and the possibility is worth to explore. Number of views used can be reduced with this extra information, hence minimize the computational cost and increase the system performance. In comparison with other methods such as neural networks, the next stage of the research could carry out a real comparison with the same data for both methods. Another possibility is the combination of both methods.

5. REFERENCES

[Bie01a] Bierderman, I., & Ju, G. (1988). Surface vs. Edge-based Determinants of Visual Recognition. *Cognitive Psychology,* 20, 38-64.

[Bue01a] Buelthoff, H. H. & Edelman, S. (1992). *Psychophysical Support for a Two-Dimensional View Interpolation Theory of Object Recognition*. Vol. 89. USA: Proceedings of The National Academic of Science.

[Cro01a] Crowder, R. G. (1982). *The Psychology of Reading*. Oxford University Press.

[Ede01a] Edelman, S. (1997). Representation is Representation of Similarities. Technical Report. Center fro Biological and Computational Learning, Department of Brain and Cognitive Sciences,http://www.kybele.psyh.cornell.edu/~edelman/archive.html (1st. December 1999).

[Ede01b] Edelman, S. & Weinshall, D. (1991). A Self-Organizing Multiple Views Representation of 3D Objects. *Biological Cybernetics*, 64, 209-219.

[Hab01a] Haber, R. N. & Haber, L. R. (1981). Visual components of the Reading Process. *Visible Language*, 15, 147-182.

[Hay01a] Hayward, W. G. (1988). Effects of Outline Shape in Object Recognition. *Journal of Experimental psychology: Human Perception and Performance*, 24(2), 427-440.

[Hay01b] Hayward, W. G. & Tarr, M. J. (1997). Testing Conditions for Viewpoint Invariance in Object Recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 23(5), 1511-1521.

[Low01a] Lowe, D. G. (1986). *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, Boston, MA.

[Mul01a] Mulholand, H. & Jones, C. R. (1968). *Fundamental of Statistics*. London Butterworths, London.

[Pyl01a] Pylyshyn, Z. (1998). *Is Vision Continuous with Cognition? - The Case for Cognitive Impenetrability of Visual Perception*. Technical Report TR-38 , Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ, http: //ruccs.rutgers.edu/publicationsreports.html .

[Roc01a] Rock, I., Halper, F. & Clayton, T. (1972). *The* Perception and Recognition of Complex Figures. *Cognitive Psychology*, 3, 655-673.

[Tay01a] Taylor, I. & Taylor, M. M. (1983). *The Psychology of Reading*. London and New York Academic Press.

[Ull01a]Ulman, S. & Basri. R. (1991). Recognition by Linear Combinations of Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 992-1005.