Tina Memo No. 1992-003
Proc(I), ICPR, The Hague August 1992.

Pairwise Representations of Shape

A.C.Evans, N.A.Thacker and J.E.W.Mayhew.

Last updated
5 / 12 / 1992

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Pairwise Representations of Shape

Alun C. Evans, Neil A. Thacker, John E. W. Mayhew

Artificial Intelligence Vision Research Unit,
University of Sheffield, Sheffield S10 2TN, England.
email: alun@aivru.sheffield.ac.uk

Abstract

Representation of arbitrary shape for the purposes of visual object recognition is an unsolved problem. This paper outlines some theoretical properties of shape representations based on stored geometric constraints that make them suitable for input to a 3D object recognition system. The construction of such representations is detailed together with an evaluation of their efficacy based on experimental findings.

1 Object Recognition

The basic requirement of a 3D recognition system is that it be capable of consistently associating the wide range of possible views of an object with a single, unique object classification. This requires the ability to deal with variations in projected shape arising from variable lighting conditions, rotations and translations in three space and occlusion by other objects.

One approach to this problem [5] involves using a regularization network to provide view interpolation between a number of stored 2D views. The present work is an attempt to provide a more robust 2D shape descriptor suitable for use in such a system. Taking a single grey level image as input we require a representational scheme which makes explicit the relevant features of an objects' projected shape and also possesses the following desirable characteristics.

Firstly the representation must have sufficient descriptive power to allow discrimination between all dissimilar shapes while capturing any large scale similarities that may exist. Secondly the representation must be applicable to as wide a class of shapes as possible, including those containing curved sections. Thirdly the representation should be robust to the types of noise expected in vision data from any real, unconstrained environment.

The requirements of later processes always impose constraints on the nature of the representation that should be produced at a particular stage. Given that the shape descriptions are to be classified by a neural network, transformations in the projected position, scale and 2D orientation of the object should be excluded as they are largely irrelevant for the purposes of recognition. Finally, the smooth, exclusively qualitative changes that occur in the visible feature set as an object rotates between catastrophic events should produce corresponding changes in the shape description.

2 Local Geometric Constraints

Descriptions constructed within the proposed representational scheme should be based on a set of local image features that somehow define the shape of the object and which can be reconstructed at all rotations, translations and scales. A sensible choice is 2D oriented line segments. These are obtained by performing a linear approximation of the edge contours detected in the image.

Comparisons between pairs of model and image features on the basis of geometric constraints have frequently been used to prune the interpretation tree inherent in model based recognition strategies [4] [3]. We propose using the values of these constraints to construct a robust descriptor of 2D shape. There are a number of possible constraints that can be used to define the geometric relationship between a pair of oriented line segments [1]. Various combinations of these constraints can be used to form shape representations with differing invariance characteristics. The constraints chosen here are based on relative angle and perpendicular distance.

Relative angle is defined as the angle between two line segments, oriented in direction away from their point of intersection. This is a useful constraint as it

0-8186-2910-X/92 $3.00 © 1992 IEEE

is un-affected by line fragmentation and is invariant over 2D transformations in shape. However, it is not a sufficient basis for constructing shape descriptions as the constraint values do not uniquely specify the full geometric relationship between lines. This can result in recognition ambiguities [6].

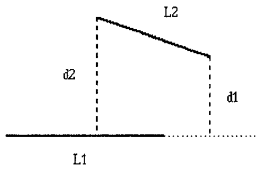

Figure 1: The Perpendicular Distance Constraint.

For better discriminability we require a constraint based on some notion of the distance between a pair of line segments. The particular measure used is the perpendicular distances from the extension of the first line to the endpoints of the second line, **Fig 1**. The values d_1 and d_2 define a range of distances over which the second line extends. The power of this constraint is based on the fact that the range of distances computed for a sub-segment of a fragmented line is *bounded* by that for the original line. This is not true of most distance constraints and is essential if representations based on these values are to be robust to noise.

3 Form of Representation

The use of the shape descriptions as input to a neural network classifier places certain constraints on the way the relationships are recorded. Such systems perform best if the components of their input vectors provide some form of statistical measure of the presence of a specific feature. The proposed representational scheme involves computing the geometric constraints that hold between pairs of lines in the image and storing the values in a histogram. The resulting structure resembles the 'R-Table' of the Generalised Hough Transform[2], though the geometric relationships are recorded in a voting fashion rather than being listed explicitly.

Given that we have 2 constraints we use a 2 dimensional histogram with axes for relative angle and perpendicular distance. Each axis is divided into a number of bins determined by the expected accuracy of the feature extraction and linear approximation processes, **Fig 2**. Information is stored in the histogram at a position determined by the geometric relationship between pairs of lines. The entry on the angle axis is computed straightforwardly from the measured relative angle. For non-parallel lines there will be a minimum and a maximum perpendicular distance between which entries are made on the distance axis.

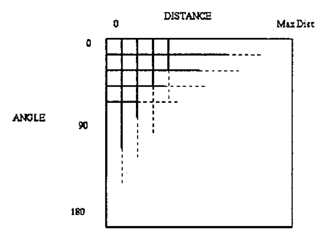

Figure 2: The Pairwise Histogram.

By recording constraint values in a histogram we accumulate evidence for the existence of pairs of lines within the shape at particular geometric relationships. The size of entries should thus be related to the importance of the lines in forming the shape to be described. This is achieved if the total entry is made equal to the product of the lengths of the two lines.

A number of extra features have to be added to the basic scheme to deal with certain complications. The case of intersecting lines is dealt with by splitting the lines at the crossover point and handling each segment separately. The effect of errors in the computation of the absolute position and orientation of lines is taken into account by blurring the entries along both axes. A gaussian blur, centred on the computed value, is used on the angle axis while a linear ramp is used on the distance axis.

Individual pairwise histograms are constructed for every line in the shape. Each histogram is constructed by taking a line as a reference and computing all orientation and distance measures relative to it. Entries for comparisons with all other lines are then combined in the histogram. The additive nature of this scheme provides a degree of invariance to line fragmentation and allows the representation to robustly describe curves. A complete shape description is formed by combining the histograms of every line. The effect of each line in defining the shape is thus taken into account many times. It is expected that the local nature and redundancy of this representation will have advantages when used as part of a recognition system. Example histograms for a square and circle are shown in **Fig 3**.

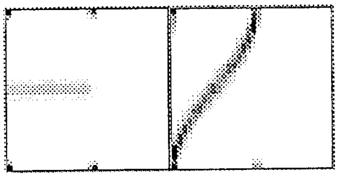

Figure 3: Example Histograms for a Square and Circle

4 Evaluation

Recognition in the neural network will be performed by comparing the pairwise histogram(s) for the current shape with templates of previously encountered shapes stored at nodes within the network. This comparison will be based on the strength of a dot product correlation between the input and stored pairwise *vectors*. Therefore we are able to evaluate the shape representation independently of the recognition module by analysing the effect of various shape perturbations on the strength of this correlation. Individual histograms for each line are summed and normalised. This provides a single shape description and ensures that the correlations are based purely on shape similarity.

4.1 Discriminability

To test the discriminability of the descriptions produced by the representation, an arbitrary curved shape was successively smoothed by applying an averaging operator to its contour. By smoothing in discrete steps the ten shapes in **Fig 4** were produced.

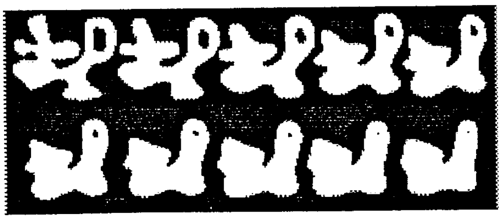

Figure 4: The smoothed shapes, 0-9 clockwise from top left

Correlations between each shape were computed and are shown in **Fig 5**. As required the peak correlations lie along the diagonal ridge corresponding to each shapes' correlation with itself. The correlations then fall away smoothly as the difference between the shapes increases. Thus any recognition system based

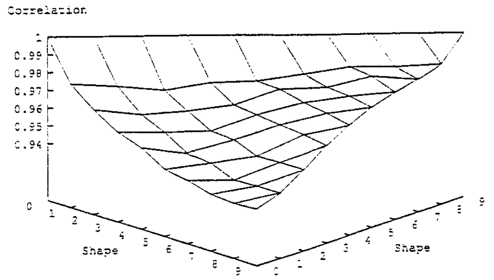

Figure 5: Graph of cross correlation between shapes

on correlating pairwise histograms with stored templates would be able to correctly identify each of the experimental shapes.

4.2 Noise

To be useful in a practical vision system the above result must be robust to the types of noise found in image data. To simulate the effects of such noise the shapes were increasingly degraded by the random addition and removal of a proportion of their lines. Example of the shapes produced are shown in **Fig 6**.

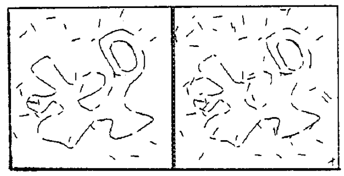

Figure 6: Example of shape at 20% & 40% noise levels

The noise levels were increased in 5% steps and the discrimination experiment repeated for the shapes produced at each stage. Thus each noisy shape was correlated with every noise free example and the highest match noted. The experiment was repeated over 40 trials to average out the random nature of the noise. **Table 1** shows the percentage of correct identifications for each shape as the noise levels increase. It can be seen that near perfect identification is preserved up to the 20% noise level. Above this level the identification degrades quite smoothly until at 40% the correlations can no longer support recognition.

Thus recognition based on pairwise descriptions of shape should be quite robust to changes in data caused by image noise. These results are to be contrasted with alternative shape descriptions, such as

	Shapes									
% Noise	0	1	2	3	4	5	6	7	8	9
0	100	100	100	100	100	100	100	100	100	100
5	100	100	100	100	100	100	100	100	100	100
10	100	100	100	100	100	100	100	100	100	100
15	100	100	100	100	100	100	100	100	100	100
20	100	100	100	100	92	97	94	97	94	100
25	100	100	100	92	76	64	51	82	71	66
30	100	97	89	56	35	38	7	51	38	35
35	100	82	61	5	7	2	0	17	12	10
40	100	64	12	2	0	0	0	5	0	0

Table 1: Effect of noise on accuracy of identification

those based on fourier methods, which are adversely affected by the loss or addition of small amounts of data.

4.3 Change in Projected Shape

Since the geometric constraints that form the basis of the representations are based on relationships *within* the shape, the descriptions produced will be invariant to 2D transformations in position and orientation. However, the use of an absolute distance measure means that the descriptions are not invariant to scale. This could be countered by scaling the distance measures with a reliable estimate of the scale of the object.

In order for the representation to form the basis of a visual recognition system which interpolates between characteristic views we must show that the similarity measure also varies smoothly with object rotation, or the equivalent change in view angle. This was tested by comparing the correlation between the pairwise description of a fixed view of a cube and that of its rotated counterparts.

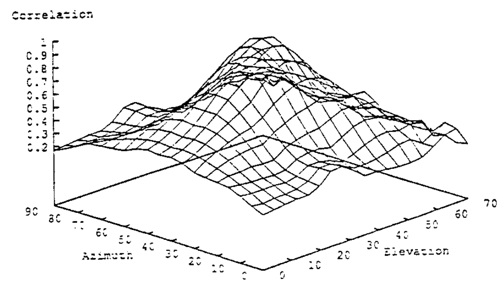

Figure 7: Correlation between view angles

Views were taken within a single octant of the view sphere, varying from $0-90^o$ azimuth and $0-70^o$ elevation. The fixed view is from the centre of this *aspect* at $45^o,35^o$. The graph of view angle against correlation is shown in **Fig7**. It can be seen that the correlation between the fixed angle and successively rotated views falls smoothly as the angular distance increases. This is exactly the behaviour required for views falling within an aspect and should allow the recognition system to group views on the basis of visible features. The degree of variation in the similarity measure would determine the number of characteristic views required to completely represent an object in a practical vision system.

5 Summary

This paper has introduced the notion of using shape representations based on stored pairwise geometric relationships. Experimental results suggest that such descriptions exhibit many of the characteristics desirable of a compact, invariant, 2D shape representation. In particular the representation has the capacity to deal with discontinuous, arbitrary curves and is robust to the kinds of noise encountered in real vision data. The structure of the descriptions, together with the fact that they change smoothly with shape, mean that they are ideally suited as input to neural networks. This provides a representational scheme with potentially far more scope than those previously used in adaptive, view based recognition systems.

References

[1] Bray, A. & Hlavac, V. 1991. 'Properties of Local Geometric Constraints.' BMVC91, pp96-103.

[2] Ballard, DH & Brown, C. 1982. Computer Vision. Prentice Hall.

[3] Goad, C. 1983. 'Special purpose automatic programming for 3D model-based vision' DARPA83, pp94-104.

[4] Grimson, WEL. & Lozano-Perez, T. 1984. 'Model based recognition from sparse range or tactile data' Int. J. Robotics Res, 3/3 pp3-35.

[5] Poggio, T. & Edelman, S. 1990. 'A network that learns to recognise three-dimensional objects.' Nature, 343 pp263-266.

[6] Thacker, NA. & Mayhew, JEW. 1989. 'Designing a layered network for context sensitive pattern classification.' Neural Networks, 3/3 pp291-300.