The Use of Geometric Histograms for Model-Based Object Recognition

— Source link ↗

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

Institutions: University of Sheffield

Related papers:

- Robust recognition of scaled shapes using pairwise geometric histograms

- Assessing the completeness properties of pairwise geometric histograms

- A Computational Approach to Edge Detection

- Structural matching by discrete relaxation

- Scene analysis using appearance-based models and relational indexing

The Use of Geometric Histograms for Model-Based Object Recognition

Alun C. Evans[1]   Neil A. Thacker[2]   John E. W. Mayhew[1]

1 Artificial Intelligence Vision Research Unit,
2 Department of Electrical and Electronic Engineering,

University of Sheffield, Sheffield. S10 2TN, England.
email: alun@aivru.sheffield.ac.uk

Abstract

We introduce a novel form of shape representation based on recording the distribution of pairwise geometric relationships between local shape features. It is shown that the *geometric histograms* used to record these distributions can be easily and robustly acquired from image data and can support recognition even when the shape extracted from the image is badly degraded by fragmentation noise and occlusion. Moreover, the processing involved in establishing correspondences between model and image features is both simple and parallel and has many advantages over previous search based methods.

1 Introduction

The problem addressed by this paper is the recognition and location of multiple rigid objects from their $2D$ projections in grey-level images. Changes in the lighting of the scene may result in sections of the projected shape being lost or fragmented. The fact that multiple, possibly overlapping, objects may be present in the scene implies that parts of their shape may be lost due to occlusion. In a very general sense we can say that any solution to this problem must have two components: a representational scheme for modeling the objects to be recognised and a matching algorithm for identifying correspondences between object and image features, **fig 1**. The possibility of occlusion effectively rules out representational schemes based on *global* characteristics of shape, such as fourier coefficients, [11], or moments, [8]. A number of researchers have attempted to overcome these difficulties by exploiting the geometric relationships between pairs of *local* shape features, [7] [9] [2]. Common to all these systems is the approach of matching model and image features on the basis of *individual* pairwise geometric relationships. The present approach is novel in that features are matched on the basis of the distribution of *multiple* pairwise geometric relationships within a shape. By correlating the *geometric histograms* used to record these distributions we are able to provide a principled metric of shape similarity. Moreover,

it can be shown that such representations degrade gracefully in conditions where features are lost through image noise or occlusion, [5]. The use of such representations as input to an adaptive, view-based $3D$ object recognition system has been previously presented in [4]. In the following sections we introduce the details of the geometric histogramming scheme and describe a simple, parallel matching strategy for performing $2D$ recognition and location.
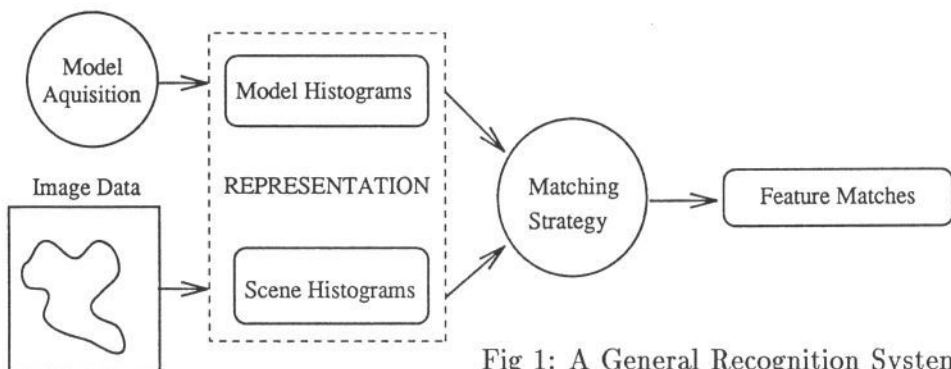

Fig 1: A General Recognition System

2 Geometric Histograms

The purpose of the representational scheme is to enable objects to be modeled in such a way that they can be quickly and reliably recognised in subsequent images. The scheme should therefore be based on some characteristic of shape which is both discriminatory and robust. Given a shape description in terms of a set of line features, obtained by performing a linear approximation of the edge strings extracted from the image, one possibility would be to record the set of pairwise geometric relationships between these line features.

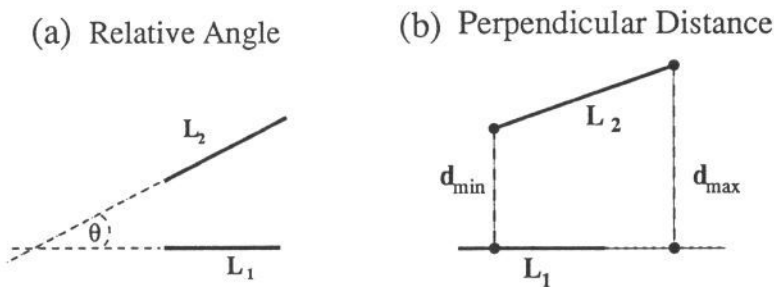(a) Relative Angle   (b) Perpendicular Distance

Fig 2: The Geometric Features

The geometric relationship between a pair of line segments can be defined to an arbitrary degree of uniqueness by the use of *geometric features* [3]. Two such features are used in the present approach. The relative angle feature is defined simply as the angle, θ, between the direction vectors of the line segments, **fig 2(a)**. This feature originally formed the basis of

the representations used in [10]. In order to improve the disambiguation power of this representation we use a second feature whose value is defined as the range of perpendicular distances from the endpoints of one line to the extension of the other. For non-parallel lines this defines a range of distances over which the second line extends, d_{min} to d_{max}, **fig 2(b)**.
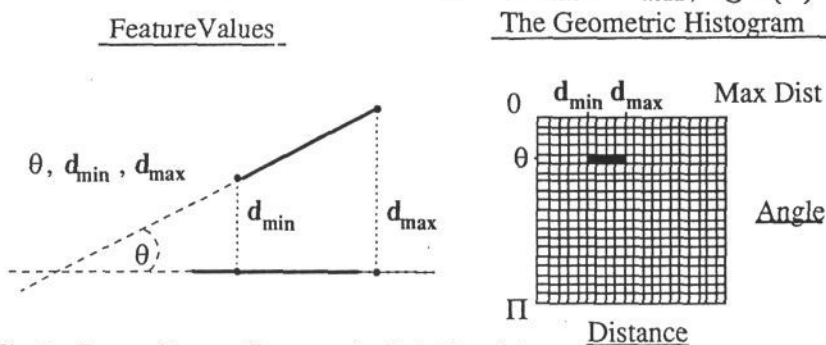
FeatureValues | The Geometric Histogram

Fig 3: Recording a Geometric Relationship

The proposed representational scheme is based on recording the distribution of values of these geometric features in a histogram. The histogram has two axes; an angle axis, whose values range from $0 \rightarrow \pi$, and a distance axis, whose values range from 0 to the maximum allowable distance between a pair of line segments. Geometric relationships are recorded in the histogram by making an entry at the measured position on the angle axis and from the minimum to maximum values on the distance axis, **fig 3**. In order to obtain the required statistical properties assumed during the matching process, the histogram contents should approximate as closely as possible the likely distribution of entries that would have been obtained if individual edge elements were considered rather than extended lines. The geometric histogram, once completed, can thus be considered as an approximation to the p.d.f for geometric co-occurrences for an edgel based description of the data. Thus, in order to encode the importance of a pair of line segments in defining the shape to be represented the value of the entry is made equal to the product of their lengths. In order to encode uncertainty regarding the true position and orientation of each line segment the entry is blurred along each axis. The scale of binning and extent of blurring define the extent of allowable differences when matching similar shapes.

Complete representations of shape are constructed by first defining a local coordinate frame around each line feature, **fig 4**. Geometric relationships, defined in this local coordinate frame, between the *base* line and all other lines are recorded in a histogram associated with the base line. This is done for each line in the shape, such that the complete shape representation is composed of n geometric histograms, where the shape is described using n line features. Restricting the range over which the geometric relationships in each histogram are measured, for example by defining a circular decision boundary around each line, **fig 4**, produces a representation which has
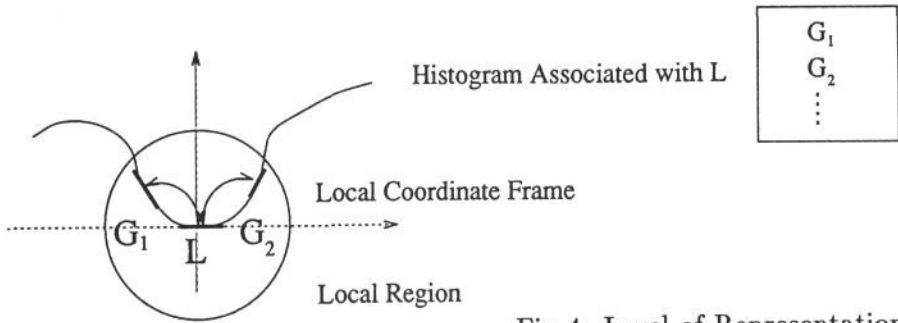
increased robustness to object occlusion.

Histogram Associated with L

G_1
G_2
⋮

Local Coordinate Frame

G_1 L G_2

Local Region

Fig 4: Level of Representation

Handling Shape Variation

One of the major difficulties in recognising objects consistently across many images is in dealing with variations in projected shape caused by fragmentation noise and variable line approximation. The more robust the representation is to these changes the easier the problem of matching becomes. Within the geometric histogramming scheme this robustness is achieved by the use of geometric features that are unaffected by, or change linearly with, the position at which a line is fragmented. The importance of this property can be demonstrated by comparing the additive effect of entries for multiple line fragments with that for the original, complete line features, **fig 5**. The use of appropriate geometric features ensures that the effect of fragmentation noise is minimised. A similar argument can be given in the case where line descriptions are changed by variations in the accuracy at which the approximation of curved shape is performed, [6].
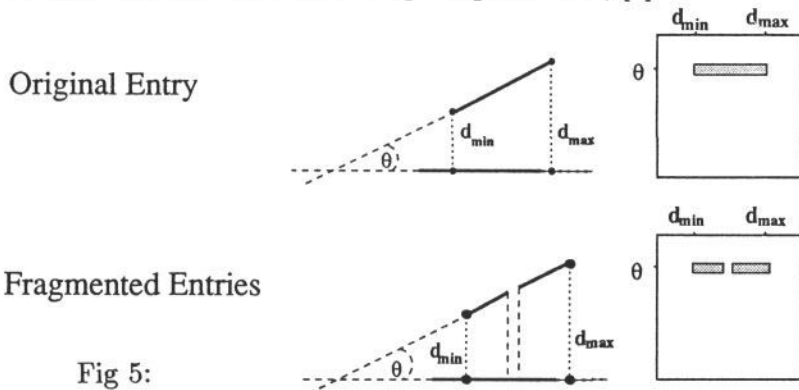
Original Entry

d_{min} d_{max}

θ

d_{min} d_{max}
θ

Fragmented Entries

d_{min} d_{max}

θ

d_{min} d_{max}
θ

Fig 5:

Assessment

Shape representations in the form of geometric histograms have many advantageous properties.

- **Redundant.** The fact that geometric relationships are recorded in more than one histogram means that the representation is quite re-

dundant. This, together with their distributed nature, implies that geometric histograms should possess considerable robustness to loss of data.

- **Local.** The fact that each geometric histogram is a measure only of local shape means that the representation is robust to partial occlusion.

- **Robust.** The use of appropriate geometric features means that representation of local shape within a histogram is quite robust to variations in line description caused by fragmentation noise and variable line approximation.

- **Easy to Match.** As we shall see in the next section, geometric histograms provide a representation of shape which is very easy to match within a parallel recognition system.

3 Matching Algorithm

The first stage in the recognition process is to construct geometric histograms for each line feature found in the image. As we have said, the purpose of the matching algorithm is to identify correspondences between image and object features. Given that stored shape representations are composed of many geometric histograms, each representing a single model feature, the degree to which an image feature matches a particular model feature can be determined by comparing their histograms. If each histogram is taken to be a vector then the degree of match between model feature M_j and image feature I is given by a dot product correlation.

$$D_j = \sum_{i=1}^{n} \sqrt{I_i . M_{ji}}$$

This can be related, via the χ^2 variable, to a maximum likelihood similarity metric [Appendix A] and can be considered as a general result for comparing two probability distributions. This statistical measure is defined on the basis of a model which assumes that each edgel has an equal probability of being detected, and thus has a simple interpretation in terms of edge detection efficiency and the possibility of object occlusion. The scale invariance of this dot product comparison metric provides a degree of invariance to line fragmentation, while the fact that spurious features will, on average, have little effect on the value of D_j means that the metric is robust to scene clutter.

Based on this idea of correlating individual histograms we can construct the following simple, parallel, competitive matching algorithm, **fig 6**. Each

image histogram is correlated with all stored model histograms. A winner-takes-all competition then takes place between the model histograms, the model feature with the highest correlation being taken as a match for the particular image feature. A threshold may be placed on the strength of the winning match to exclude image features which do not correlate strongly with any model features. Once matching is complete, each above threshold image feature is matched against a particular model feature. The set of hypothesised matches produced by the matching can be used in various ways to determine the pose of each object in the image. In the present system we make use of the Generalised Hough Transform, [1].
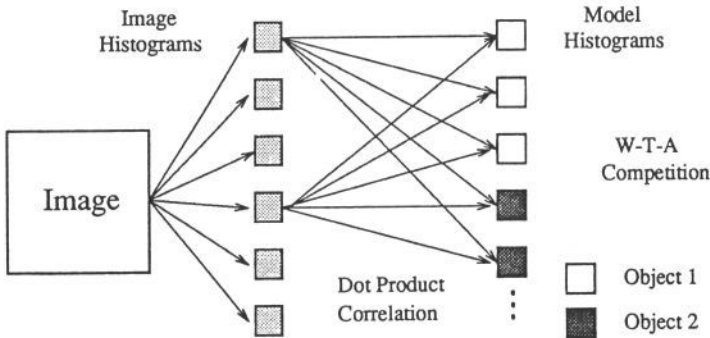

Fig 6: The Parallel Matching Strategy

This matching strategy has many desirable features. Firstly, not all model features must be matched; an important consideration given the likelihood of certain features being lost due to image noise or occlusion. This situation must be handled explicitly in search based methods, adding considerable complexity to an approach which is already computationally expensive. Secondly, an image feature may be matched by more than one model feature. This is essential if multiple occurrences of an object are to be recognised and if fragmented image lines are to be correctly matched. Finally, not all image features must be matched. If a particular feature is below threshold then it can be excluded from the segmentation as it most likely belongs to some unknown object.

Assessment
In addition to the above features there are certain advantages to the nature of the processing involved in this matching algorithm.

- **Principled.** The matching criterion is related to a maximum likelihood statistic based on the matching of two statistical distributions.

- **Simple.** The computation required to determine the degree of match between two features involves only simple array multiplication.

- **Parallel.** Histogram correlations may be computed independently, leading to a simple hardware implementation.

4 System Demonstration

The performance of the recognition system in three situations is presented: multiple occurrences of a single object, multiple occluded objects and multiple objects in scene clutter. In each case two things are shown; the lines features extracted from the image and the objects located by the system overlayed onto the image.

Multiple Occurrences of Same Object

This example demonstrates the property of the recognition system whereby an object feature may be matched by more than one image feature.
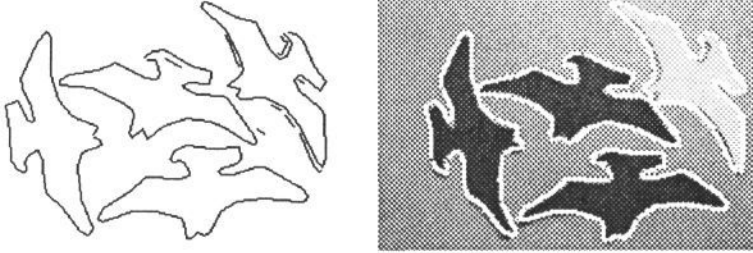

Multiple Occluded Objects

The ability of the system to recognise the objects despite this occlusion is a consequence of the fact that each geometric histogram is a local measure of shape. The correspondences established in one part of the object are therefore unaffected by loss of another part due to occlusion.
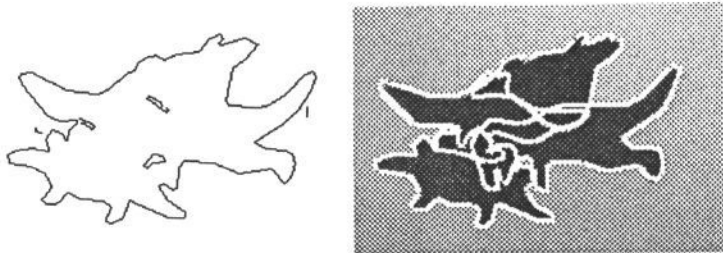

Unknown Objects

In the previous cases all image features belonged to a known object. Here we have an image of a scene containing a number of unknown objects, (humans), that partially occlude the known objects.
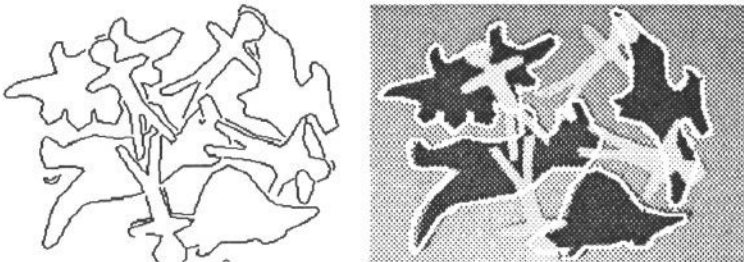
Fragmentation Noise

To simulate the effect of fragmentation noise the lines detected from the image were corrupted by removing 50% of their length. This situation is much worse than would be expected due to illumination problems and reliability of edge detection. As can be seen, in all three cases the system is still able to recognise the objects.
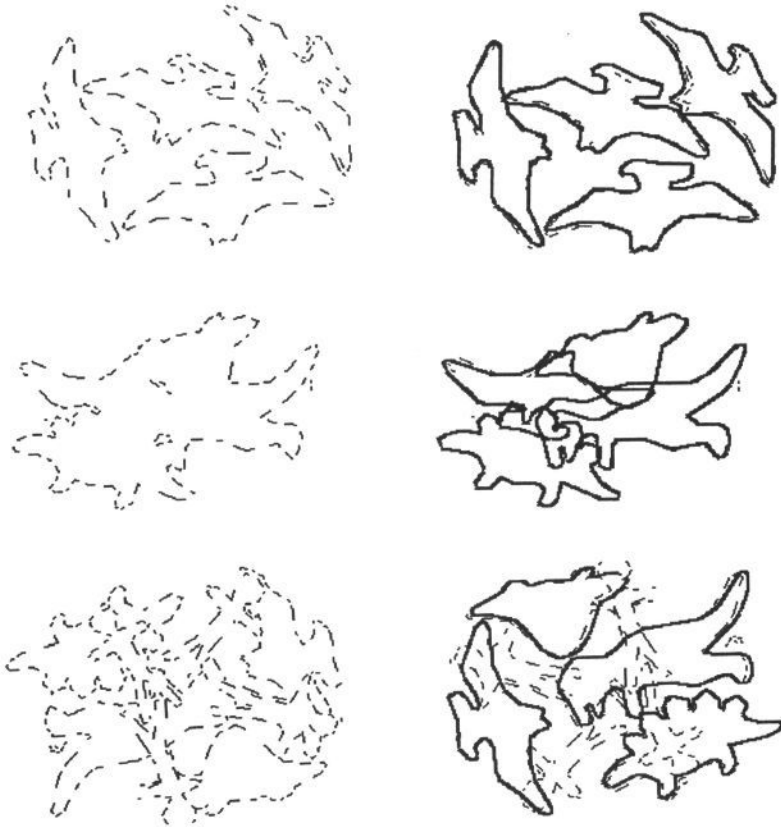
5 Conclusion

The ease with which a recognition system is able to identify and locate objects is determined to a large degree by the from of the representations used to model the objects. A novel method of shape representation has been introduced which is based on recording the distribution of pairwise geometric relationships within a shape. It has been shown that such representations can be matched using a principled, statistically based measure of shape similarity which ensures that if information is present in the histograms then matching will be correct. It has been demonstrated that a simple recognition system based on these representations is capable of establishing shape

feature correspondences even in the presence of considerable fragmentation noise, occlusion and scene clutter. These matches can either be used to segment the objects present in the image or as input to a pose estimation algorithm, such as the GHT. Moreover, the simple, parallel nature of the processing involved in recognition means that it lends itself easily to hardware implementation. This is currently the subject of further research.

Appendix A

We will show that the maximum of a dot product metric $D_j = \sum_i^n \sqrt{o_i}\sqrt{m_i}$ is the minimum of a χ^2 variable comparing two frequency distributions o_i and m_i.

The maximum likelihood statistic χ^2 for comparing two distributions o_i and m_i is defined as

$$\chi^2 = \sum_i^n (o_i - m_i)^2/m_i \qquad (1)$$

for small $(o_i - m_i)$, the first order Taylor expansion of f at m_i can be written as

$$f(o_i) \approx f(m_i) + (o_i - m_i)\frac{\partial f(m_i)}{\partial m_i} \qquad (2)$$

which gives

$$(o_i - m_i) \approx \frac{f(o_i) - f(m_i)}{\frac{\partial f(m_i)}{\partial m_i}} \qquad (3)$$

Substituting (3) in (1) gives

$$\chi^2 = \sum_i^n \frac{(f(o_i) - f(m_i))^2}{(\frac{\partial f(m_i)}{\partial m})^2 m_i} \qquad (4)$$

In the special case of $f(x) = \sqrt{x}$ we have

$$\chi^2 = 4\sum_i^n (\sqrt{o_i} - \sqrt{m_i})^2$$

which expanded gives

$$\chi^2 = 4\sum_i^n o_i + 4\sum_i^n m_i - 8\sum_i^n \sqrt{o_i}\sqrt{m_i}$$

which for normalised m gives

$$\chi^2 = const - 8\sum_i^n \sqrt{o_i}\sqrt{m_i}$$

Thus, under these assumptions, taking the maximum of the dot product metric D_j is the same as taking the minimum of the χ^2 statistic.

Acknowledgements

References

[1] Ballard. D.H. & Brown, C. 1982. Computer Vision. Prentice Hall.

[2] Bray, A. 1990. Object Recognition using Local Geometric Constraints. 1st European Conference on Computer Vision. pp.499-515.

[3] Bray, A. & Hlavac, V. 1991. Properties of Local Geometric Constraints. BMVC '91.

[4] Evans, A.C., Thacker, N.A., & Mayhew, J.E.W. 1993. A Practical View-Based 3D Object Recognition System. Proc. 3rd Int. Conf. on Artificial Neural Networks. pp. 6-10.

[5] Evans, A.C., Thacker, N.A., & Mayhew, J.E.W. 1992. Pairwise Representations of Shape. Proc. 11th Int. Conf. on Pattern Recognition. pp.135-139.

[6] Evans, A.C. 1992. Geometric Histograms. PhD Thesis (in preparation).

[7] Grimson, W.E.L. & Lozano-Perez, T. 1984. Model Based Recognition from Sparse Range or Tactile Data. Int. J. Robotics Res. 3(3).

[8] Hu, M.K. 1962. Visual Pattern Recognition by Invariant Moments. IRE Trans. on Information Theory, Vol. 8. pp.179-187.

[9] Lambdan, Y. & Wolfson, H.J. 1988. Geometric Hashing: a general and efficient model-based recognition scheme. Proc. 2nd Intnl. Conf. on Comp. Vision. pp.238-249.

[10] Thacker, NA. & Mayhew, JEW. 1989. Designing a Layered Network for Context Sensitive Pattern Classification. Neural Networks. Vol.3 No.3. pp.291-300.

[11] Zahn, C.T. & Roskies, R.Z. 1972. Fourier Descriptors for Plane Closed Curves. IEEE Trans. Comp. Vol.21, No.3. pp.269-281.