Tina Memo No. 1996-005

Proc. IEEE Image Processing 1995.

Multiple Shape Recognition Using Pairwise Geometric Histogram Based Algorithms.

A. P. Ashbrook, N. A. Thacker and P. I. Rockett.

Last updated
24 /8 / 2006

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

MULTIPLE SHAPE RECOGNITION USING PAIRWISE GEOMETRIC HISTOGRAM BASED ALGORITHMS

A P Ashbrook, N A Thacker and P I Rockett

University of Sheffield, UK

INTRODUCTION

The scope for object recognition systems is vast, with applications ranging from industrial inspection to image interpretation. Many systems have already been demonstrated with limited success but frequently the problems that they attempt to solve are highly constrained, Hu (1), Zahn (2), Lanitis (3), Hough (4) and Rothwell (5). In particular, recognition is often limited to a very small number of known shapes. To satisfy the requirements of any arbitrary object recognition application a general solution to this problem is needed.

Pairwise Geometric Histogram (PGH) based algorithms have been shown to be a robust solution for the recognition of arbitrary, 2D shape in the presence of occlusion and scene clutter, Evans (6). The method is both statistically founded and complete in the sense that a shape may be reconstructed from its PGH representation, Riocreux (7). The generality of this method has been further reinforced by a recent analysis of its scaleability which concludes that, if used appropriately, it is suitable for the recognition of very large numbers of objects, Ashbrook (8). In this paper we demonstrate the application of PGHs to recognition tasks involving very large model training sets.

PAIRWISE GEOMETRIC HISTOGRAMS

PGHs are a representation used for the recognition of rigid, 2D shape. The complete algorithm comprises a number of stages:

1. Model image data (during training) and scene image data (during recognition) is processed by an edge extraction process and the edge data is approximated by line segments.
2. Model histograms (during training) and scene histograms (during recognition) are constructed for each line segment (reference line) by comparing the reference line to all other lines and making histogram entries according to the measured relative angles and perpendicular distances, see figure 1. To account for errors in the measurement processes and to compensate for the variability in the way a shape may be segmented into lines, entries are first blurred appropriately before being placed into the histogram. This representation encodes local shape geometry in a manner which is invariant to rotation and translation and is robust to missing data, line fragmentation and clutter.
3. Scene line labelling is performed by finding good matches between scene histograms and model histograms using the Bhattacharyya metric. This statistical metric is appropriate as PGHs are pdfs of local shape geometry.
4. Object classifications are confirmed by finding consistent labelling within a scene using a probabilistic Hough transform. This stage also determines object locations and orientations in a way which is equivalent to the result of conducting a robust least-squares fit of the model line data to the scene line data.

A SOLUTION FOR VERY LARGE MODEL DATABASES

The maximum useable size of a model database for any particular shape recognition method is determined by a combination of limiting factors:

- Diminishing reliability
- Insufficient pattern storage
- Excessive computation

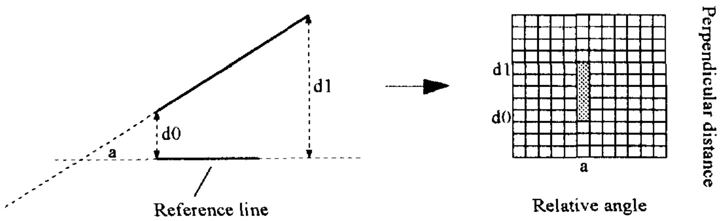


Figure 1: A histogram entry for a single line comparison

Image Processing And Its Applications, 4–6 July 1995
Conference Publication No. 410, © IEE 1995.

In a recent analysis of PGHs (8) we have shown that, if used appropriately, this method provides a solution which is applicable to very large model databases. In particular:

- The reliability of recognition using pairwise geometric histograms is unaffected by the size of the model database.
- Pairwise geometric histograms can represent very large numbers of distinct shapes (typically $>10^8$ unique histograms may be stored).
- The processing requirement for recognition scales in a manner which is no worse than linear with the number of stored models.

In this analysis a number of additions to the existing algorithm have been prescribed.

Soft Line Labelling

Previously, matching scene PGHs to model PGHs resulted in each scene line being labelled with a *single* model line hypothesis. This winner-takes-all approach can be shown to be unreliable for very large model databases.

When labelling a scene line in the presence of noise there is a probability, P_r, that a random model line will be selected as the best candidate match in preference to the correct model line. This can happen because the score for the correct match is reduced by the presence of scene noise. To determine P_r we need to consider both the effect of noise on the correct match scores and the distribution of match scores when random histograms are compared. This is demonstrated in figure 2 (Δ_{dp} is the reduction in match score due to scene noise).

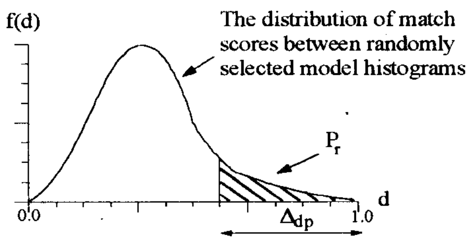


Figure 2: The probability of a random line winning

The major sources of noise encountered in a scene are missing data (possibly due to occlusion) and clutter. If D is the match score in the absence of noise, k is the proportion of missing data and a is the proportion of scene clutter then the reduced match score D' is:

$$D' \rightarrow D\frac{(1-k)}{(1+a)}$$

For reliable recognition it is important to ensure that most of the correct scene line to model line correspondences are identified. The probability, P_c, that a scene line will be labelled correctly when compared to all N of the stored model histograms is:

$$P_c = (1 - P_r)^N$$

For large databases, where N is large, this probability becomes vanishingly small so it is unlikely that the correct model lines will always win, even for small values of P_r.

To minimise the possibility that the correct model line is discarded it is necessary to label scene lines with *multiple* hypotheses. Because of the robustness of the Hough transform, entering a limited amount of noise (the incorrect labels) along with the correct labels does not unduly degrade the recognition. Typically, the number of labels to maintain for each scene line, N_{labels}, should be the mean number of random winners:

$$N_{labels} = NP_r$$

Consequently, to maintain reliable recognition the number of labels per scene line should be scaled in accordance with the number of stored histograms. Although this results in the total amount of noise scaling with the database size, if these labels have no bias towards any particular model the noise per model will remain fixed and the performance will continue to be reliable.

The Model Classification Stage

After scene line labelling is complete a probabilistic, generalised Hough transform is usually constructed for *every* known model to determine whether or not that model is present in the scene (and to determine its location if it is present). An additional stage is proposed which accumulates evidence for each model based upon the scene line labelling. Computationally expensive Hough transforms are then only constructed for models which have significant evidence at this stage.

The measure of evidence we have used is the accumulated length of scene lines which have been labelled for a particular model shape, normalised by the total length of lines in the model. This measure corresponds to the proportion of the model which is present in the scene so a principled threshold can be used to select candidate models accordingly.

Because there is a limit to the number of models shapes which may appear in any scene, the number of hough transforms which need to be constructed is bounded. Consequently, this stage of recognition can be performed in some fixed time irrespective of the number of stored models.

DEMONSTRATION

To demonstrate the effectiveness of PGH's for the recognition of large numbers of shapes we have chosen a training set of over 100 models which comprises both shape silhouettes and 3 dimensional objects, see figure 6. Although our analysis of the algorithm suggests that many more models could reliably be recognised, the size of this training set is significantly larger than has been demonstrated by other algorithms.

For reliable performance with this size of model set we have extended the algorithm to maintain multiple labels as discussed above. By considering the distribution of match scores between randomly selected histograms and allowing the correct match scores to fall by 30% due to scene noise we have found that the probability of a random histogram winning is 0.002. The models in the training set comprise a total of 6933 lines (so 6933 histograms are stored). Therefore the number of hypothesised labels to maintain should, typically, be:

$$N_{labels} = 0 \cdot 002 \times 6933 \approx 14 \text{ labels per scene line}$$

Simple 2D Templates

Figure 3(a) depicts a scene containing four of the models from the training set. This particular example embodies a number of problems encountered in object recognition, namely: occlusion, clutter and missing data. The algorithm is able to recognise and locate all of these models successfully. The figure shows one of the recognised models and demonstrates how accurately it has been located (this is equivalent to a robust least square fit of the model to the appropriate sections of the scene).

Figure 4(a) depicts the output of the classification stage constructed from the labelled scene lines. The four highest peaks correspond to the four models in the scene and the height of each peak relates to the proportion of each model present. Because of the severe occlusion in the scene the level of evidence is relatively low but still sufficient to identify them as likely candidates.

If the classification stage had not been introduced, over 100 generalised Hough transforms would need to have been constructed to identify model occurrences in the scene. By using this stage to filter out unlikely candidates only four Hough transforms needed to be used.

Characteristic Views of 3D Models

Many 3 dimensional object recognition applications are simplified if the objects of interest can only ever present a limited number of views to the camera. For example, machine parts lying on a flat surface when viewed from above have a view for each of the stable positions in which the part may lie. These objects can be completely represented and later recognised using the set of possible views.

PGH based algorithms are a particularly effective solution for this type of application because of their ability to recognise relatively complex and arbitrary shape.

Figure 3(b) depicts a cluttered scene containing 3 dimensional objects from the training set viewed from above. These models are not simple outline templates but contain internal line structure which is more characteristic of real objects. Again recognition is successful as demonstrated in the figure.

View Base 3D and Pose Estimation

Any 3 dimensional object can be represented by a large set of 2 dimensional views taken at intervals around it. In fact there is evidence which suggests that this is how the human perception of 3 dimensional form operates. Bulthoff (9) and Koenderink and van Doorn (10).

There is scope for extending PGH based algorithms to the recognition of 3 dimensional objects using this approach because:

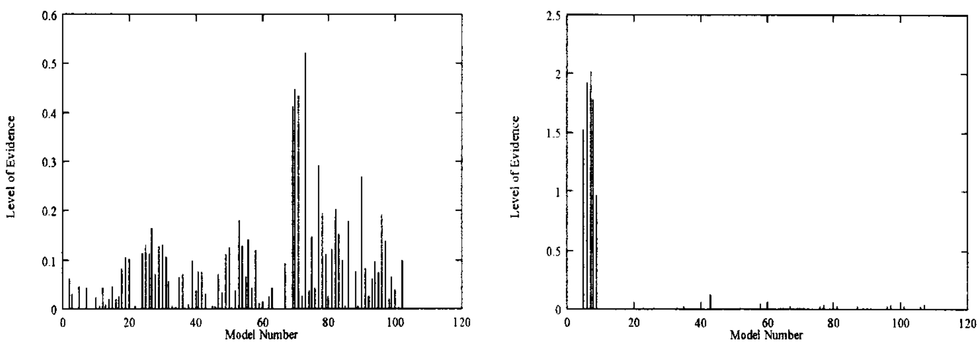
- The algorithm is scaleable to very large numbers of models
- The representation is sufficiently descriptive to discriminate between similar views of the same 3 dimensional model.
- The representation changes gradually for small changes in view direction This property is necessary if a 3 dimensional object is to be represented by a *finite* number of 2 dimensional views.

To demonstrate the latter two points we have used PGHs to identify the occurrence of a wire-frame aircraft model in a scene. Figure 5(a) contains a set of 2 dimensional views of the aircraft, each taken at a slightly more tilted angles. The aircraft in the scene, shown in figure 5(b), is tilted by the same amount as the central model.

As we would expect, the aircraft is successfully recognised and located by the PGH algorithm. What is of interest in this example, however, is the way the level of evidence varies across the different aircraft views. The classification layer for this scene is depicted in figure 4(b). The five peaks in the plot correspond to the five stored views of the aircraft. As predicted, the highest peak corresponds to the correct view and the evidence falls off slowly either side of this view. Using this information it is possible to estimate the pose of *any* 3 dimensional object by interpolation provided enough views of the object have been stored.

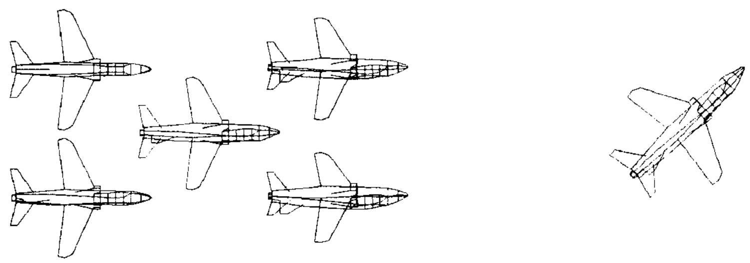
(a) (b)

Figure 3: The recognition and location of (a) a simple 2D silhouette and (b) a 3D object in complex scenes



(a) (b)

Figure 4: The classification layers for (a) the dinosaur scene and (b) the aircraft scene



(a) (b)

Figure 5: (a) 5 views of an aircraft. (b) A single view of the aircraft located in a simple scene

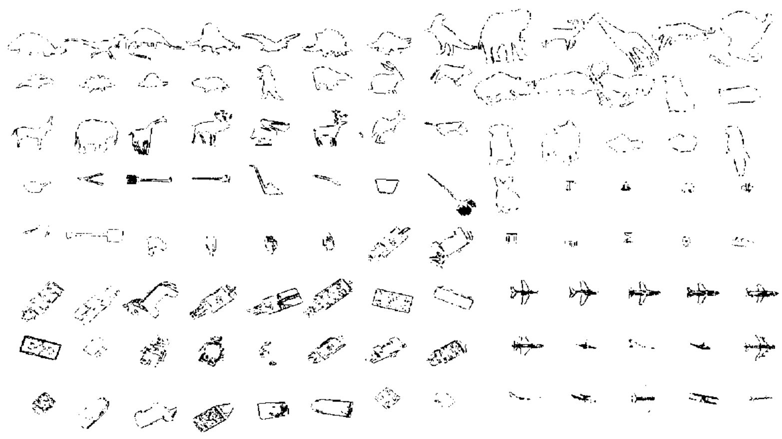


Figure 6: The training set used for the demonstrations in this paper

CONCLUSIONS

We have identified the need for a general solution to the object recognition problem and in particular, a solution which is scaleable. The use of PGHs has already been demonstrated for the recognition of small numbers of models and a recent analysis has shown that this method is applicable to very large model databases.

In this paper we have used PGHs in the manner prescribed by this analysis and demonstrated that it is an effective solution for the recognition of very large numbers of model shapes. We have also shown that the algorithm may be extended to the recognition of 3 dimensional objects using large numbers of 2 dimensional views.

Model databases of the size used in this paper are beginning to reach a limit imposed by the computing power of general purpose workstations. To allow this line of research to be continued we intend to map the algorithms described above onto custom hardware using specialised array processors designed in house, Thacker (11), Yates (12) and Brown (13). This will also allow object recognition to be used in applications demanding real-time performance.

REFERENCES

1. Hu M K, 1962, "Visual Pattern Recognition by Moment Invariants", IRE Trans. Inf. Theory, 8, 179-187

2. Zahn C T and Roskies R Z, 1972, "Fourier Descriptors for Plane Closed Curves", IEEE Trans. Comp., 21, 269-281

3. Lanitis A, Taylor C J and Cootes T F, 1993, "A Generic System for Classifying Variable Objects using Flexible Template Matching", Proc. BMVC93, 1, 329-338

4. Hough P V C, 1962, "Method and Means for Recognizing Complex Patterns", US Patent 3069654

5. Rothwell C A, Zissermann A, Forsyth D A and Mundy J L, 1992, "Canonical Frames for Planar Object Recognition", Proc. ECCV, 757-772.

6. Evans A C, Thacker N A and Mayhew J E W, 1993, "The use of Geometric Histograms for Model-Based Object Recognition", Proc. BMVC93, 429-438

7. Riocreux P A, Thacker N A and Yates R B, 1994, "An Analysis of Pairwise Geometric Histograms for View-Based Object Recognition", Proc. BMVC94, 1, 75-84

8. Ashbrook A P, Thacker N A and Rockett P I, 1995, "Pairwise Geometric Histograms: A Scaleable Solution for the Recognition of 2D Rigid Shape", Proc. SCIA95.

9. Bulthoff H H and Edelman S, 1992, "Psychophysical Support for a Two-dimensional View Interpolation Theory of Object Recognition", Proc. Natl. Acad. Sci. USA, 89, 60-64

10. Koenderink J J and Doorn A J, 1979, "The Internal Representation of Solid Shape with Respect to Vision", Biological Cybernetics, 32, 211-217

11. Thacker N A, Courtney P, Walker S N, Evans S J and Yates R B, 1994, "Specification and Design of a General Purpose Image Processing Chip", 12th ICPR94, 3, 268-273

12. Yates R B, Thacker N A, Evans S J, Walker S N and Ivey P A, 1995, "An Array Processor for General Purpose Digital Image Compression", IEE JSSC, 30, 224-249

13. Brown C I, Thacker N A and Yates R B, 1995, "A VLSI Architecture for the Wavelet Transform", Proc. IPA95