Object Recognition from Large Structural Libraries

Benoit Huet and Edwin R. Hancock

Department of Computer Science
University of York, York, Y01 5DD, UK

Abstract. This paper presents a new similarity measure for object recognition from large libraries of line-patterns. The measure commences from a Bayesian consistency criterion which as been developed for locating correspondence matches between attributed relational graphs using iterative relaxation operations. The aim in this paper is to simplify the consistency measure so that it may used in a non-iterative manner without the need to compute explicit correspondence matches. This considerably reduces the computational overheads and renders the consistency measure suitable for large-scale object recognition. The measure uses robust error-kernels to gauge the similarity of pairwise attribute relations defined on the edges of nearest neighbour graphs. We use the similarity measure in a recognition experiment which involves a library of over 2000 line-patterns. A sensitivity study reveals that the method is capable of delivering a recognition accuracy of 94%. A comparative study reveals that the method is most effective when a Gaussian kernel or Huber's robust kernel is used to weight the attribute relations. Moreover, the method consistently outperforms Rucklidge's median Hausdorff distance.

1 Introduction

Object recognition from large libraries of images holds the key to the automatic manipulation of massive volumes of visual information. The overall goal is the rapid indexation of images according to their contents. The problem can be viewed as having two distinct ingredients. The first of these is a compact image representation that is robust to noise, occlusion and changes in imaging geometry. The second ingredient is a means of comparing descriptions. Ideally, the distance measure should have a degree of robustness to outliers.

The first of these issues, i.e. of efficient object representation, has recently stimulated considerable interest in the literature. Examples include both geometric [1] and structural hashing [2], a variety of invariants [3,4] and pairwise geometric histograms [5]. However, the second issue of how to compare representations has received less attention. One exception is the recent work of Rucklidge [6] which has shown how the Hausdorff distance can be used for relatively robust object recognition and location.

Despite offering an interesting and effective strategy for comparing image representations, there are a number of criticisms that can be leveled at the

use of Hausdorff distance. In the first instance, the measure is crisply defined over the max-min tests between the elements of the sets of object-primitives being compared. Although this offers a certain degree of robustness to noise and outliers, it fails to adequately capture uncertainties in the image attributes being compared. The second shortcoming, is the failure to impose relational structure on the arrangements of object-primitives. In other words, a considerable wealth of contextual information is overlooked. The aim in this paper is to address these two issues in a more critical manner. Specifically, we aim to draw on our recently reported work on graph-matching [7] to develop a probabilistic similarity measure for object recognition from large structural data-bases.
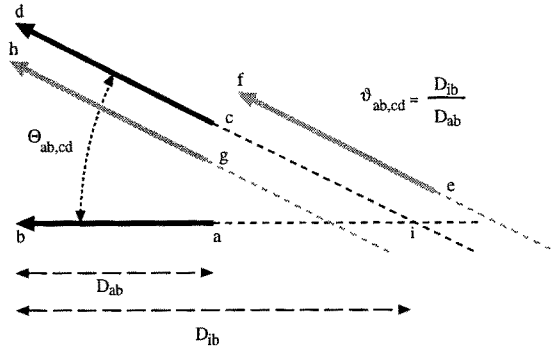
2 Pairwise Geometric Attributes

Fig. 1. Geometry for shape representation

The raw information available for each line segment are its orientation (angle with respect to the horizontal axis) and its length (see figure 1). To illustrate how the pairwise feature attributes are computed suppose that we denote the line segments indexed (ab) and (cd) by the vectors $\underline{x}_{ab}$ and $\underline{x}_{cd}$ respectively. The vectors are directed away from their point of intersection. The relative angle attribute is given by

$$\theta_{\underline{x}_{ab},\underline{x}_{cd}} = \arccos\left[\frac{\underline{x}_{ab} \cdot \underline{x}_{cd}}{|\underline{x}_{ab}||\underline{x}_{cd}|}\right]$$

From the relative angle we compute the directed relative angle. This is an extension to the attribute used by Evans *et al.* [5], that consists of giving the relative angle a positive sign if the direction of the angle from the baseline $\underline{x}_{ab}$ to its pair $\underline{x}_{cd}$ is clockwise and a negative sign if it is counter-clockwise. This allows us to extend the range of angles describing pairs of segments from $[0,\pi]$ to $[-\pi,\pi]$ and therefore, reduce indexation errors associated with angular ambiguities.

In order to describe the relative position between a pair of segments and resolve the local shape ambiguities produced by the relative angle attribute we

introduce a second attribute. The directed relative position $\vartheta_{\underline{x}_{ab}, \underline{x}_{cd}}$ is represented by the normalised length ratio between the oriented baseline vector $\underline{x}_{ab}$ and the vector $\underline{x}_{ib}$ joining the end (b) of the baseline segment (ab) to the intersection of the segment pair (cd).

$$\vartheta_{\underline{x}_{ab}, \underline{x}_{cd}} = \frac{1}{\frac{1}{2} + \frac{D_{ib}}{D_{ab}}}$$

The physical range of this attribute is $(0, 1]$. A relative position of 0 indicates that the two segments are parallel, while a relative position of 1 indicates that the two segment intersect at the middle point of the baseline.

3   Relational Constraints

We aim to augment the pairwise attributes with constraints provided by the edge-set of the N-nearest neighbour graph. The conventional Hausdorff distance explores the complete set of associations between the set of tokens constituting the model and the data. Here our aim is to limit the set of associations to those that are consistent with the local structure of the neighbourhood graph. The motivation here is that that local object representations are more robust to occlusion, missing/extra features and noise.

   We represent the sets of line-patterns as 4-tuples of the form $G = (V, E, U, B)$. Here the line-segments extracted from an image are indexed by the set V. More formally, the set V represents the nodes of our nearest neighbourhood graph. The edge-set of this graph $E \subset V \times V$ is constructed as follows. For each node in turn, we create an edge to the N line-segments that have the closest distances. Associated with the nodes and edges of the N-nearest neighbour graph are unary and binary attributes. The unary attributes are defined on the nodes of the graph and are represented by the set $U = \{(\phi_i, l_i); i \in V\}$. Specifically, the attributes are the line-orientation ϕ_i and the line-length and l_i. By contrast, the binary attributes are defined over the edge-set of the graph. The attribute set $B = \{(\theta_{i,j}, \vartheta_{i,j}; (i, j) \in E \subseteq V \times V\}$ consists of the set of pairwise geometric attributes for line-pairs connected by an edge in the N-nearest neighbour graph.

   We are concerned with attempting to recognise a single line-pattern $G_m = (V_m, E_m, U_m, B_m)$, or model, in a data-base of possible alternatives. The alternative data-patterns are denoted by $G_d = (V_d, E_d, U_d, B_d)$, $\forall d \in \mathcal{D}$ where $\mathcal{D}$ is the index-set of the data-base.

4   Pairwise Attribute Consistency

The aim in this paper is to draw on recent work on relational graph matching to develop a similarity measure for rapidly comparing relational descriptions of line-patterns which are represented in the manner outlined in the previous section. Although the framework furnishes a principled Bayesian measure of relational

consistency, it has hitherto been used exclusively for graph-matching using itera-
tive relaxation operations. It is hence unsuitable for rapid recognition of objects
from large object libraries on two counts. In the first instance, graph matching
is concerned with detailed correspondence matching rather than global object
recognition. Secondly, the since relaxation algorithms are iterative in nature,
they are too computationally demanding to be used when large object-libraries
are being considered. The aim here is to provide a simplified relational consis-
tency measure which can be used for recognition without the need to iteratively
establish correspondence matches.

4.1 Global Pattern Similarity

We take as our starting point the the weak-context version of the average con-
sistency measure developed for evidence combination by Kittler and Hancock
[10]. Following Christmas, Kittler and Petrou [11] we measure the compatibility
of the graphs being compared using pairwise attribute relations defined on the
edges of the nearest-neighbour graph. To be more formal, suppose that the set of
nodes connected to the model-graph node I is $C_I^m = \{J|(I,J) \in E_M\}$. The corre-
sponding set of data-graph nodes connected to the node i is $C_i^d = \{j|(i,j) \in E_d\}$.
With these ingredients, the consistency criterion which combines evidence for the
match of the graph G_m onto G_d is

$$Q(G_d, G_m) =$$
$$\frac{1}{|V_M| \times |V_d|} \sum_{i \in V_d} \sum_{I \in V_M} \frac{1}{|C_i^d|} \sum_{j \in C_i^d} \frac{1}{|C_I^m|} \sum_{J \in C_I^m} P\left((i,j) \to (I,J)|\underline{v}_{I,J}^m, \underline{v}_{i,j}^d\right) \quad (1)$$

The probabilistic ingredients of the evidence combining formula need further
explanation. The *a posteriori* probability $P\left((i,j) \to (I,J)|\underline{v}_{I,J}^m, \underline{v}_{i,j}^d\right)$ represents
the evidence for the match of the model-graph edge (I,J) onto the data-graph
edge (i,j) provided by the corresponding pair of attribute relations $v_{I,J}^m$ and $v_{i,j}^d$.
Based upon our discussion of the qualitative properties of the Hausdorff
model, we would like to use the Bayesian consistency criterion as the basis of
a similarity measure for graph-based object recognition. To commence this de-
velopment, we consider a very simple form for the structural error process. We
assume that the conditional prior can be modelled as follows

$$P\left((i,j) \to (I,J)|\underline{v}_{I,J}^m, \underline{v}_{i,j}^d\right) = \Gamma_\sigma(||\underline{v}_{I,J}^m - \underline{v}_{i,j}^d||) \quad (2)$$

where $\Gamma_\sigma(||\underline{u}_I^m - \underline{u}_i^d||)$ is a distance weighting function.
We now consider how to simplify the computation of relational consistency.
We commence by considering the inner sum over the nodes in the model-graph
neighbourhood C_I^M. Rather than averaging the edge-compatibilities over the en-
tire set of feasible edge-wise associations, we limit the sum to the contribution
of maximum probability. Similarly, we limit the sum over the node-wise associ-
ations in the model graph by considering only the matched neighbourhood of

maximum compatibility. With these restrictions, the process of maximising the Bayesian consistency measure is equivalent to that of maximising the following relational-similarity measure

$$Q(G_d, G_m) = \sum_{i \in V_d} \max_{I \in V_m} \sum_{j \in C_i^d} \max_{J \in C_I^m} \left(\Gamma_\sigma(\| \underline{v}_{I,J}^m - \underline{v}_{i,j}^d \|) \right) \qquad (3)$$

With the similarity measure to-hand, the best matched line pattern is the one which satisfies the condition

$$Q(G_d, G_m) = \arg \max_{d' \in \mathcal{D}} Q(G_d', G_m) \qquad (4)$$

4.2   Robust Weighting Kernels

We will consider several alternative robust weighting functions. The most appealing of these is a Gaussian of the form

$$\Gamma_\sigma(\rho) = \exp\left(-\frac{1}{2} \frac{\rho^2}{\sigma} \right)$$

We will also consider several alternatives suggested by the robust statistics literature. These include

- the sigmoidal derivative $\Gamma_\sigma(\rho) = \rho^{-1} \tanh\left(\frac{\pi \rho}{\sigma}\right)$
- Huber's kernel $\Gamma_\sigma(\rho) = \begin{cases} 1 & \text{if } \rho < \sigma \\ \frac{\sigma}{|\rho|} & \text{otherwise} \end{cases}$
- Huber's narrow-band kernel $\Gamma_\sigma(\rho) = \left(1 + \frac{|\rho|}{\sigma}\right)^{-1}$

Stated in this way, the recognition metric has much in common with the graph-matching criterion recently reported by Wilson and Hancock [7]. However, rather than being used for primitive-by-primitive correspondence matching, in the work reported here we use the criterion for recognising primitive ensembles.

4.3   Hausdorff Distance

In our experimental evaluation of the new recognition measure, we will provide some comparison with the Hausdorff distance used by Rucklidge [6]. However in order to make the comparison meaningful, in this section we describe how the Hausdorff distance can be extended to graph-based object representations.

The idea underpinning the Hausdorff distance is to compute the distance between two sets of unordered observations when the correspondences between the individual items are unknown. In object recognition, this problem presents itself when sets of unlabelled image primitives are being compared.

The distance is computed by exploring the entire space of possible model-data associations between two sets of unstructured measurement vectors. The metric gauges the distance between the two sets of observations using the maximum

value of the minimum pairwise data associations. More formally, with the graph-based notation introduced in Section 3, the distance is defined to be

$$H_G(G_d, G_m) = \max_{(i,j) \in E_d} \min_{(I,J) \in E_m} ||\underline{v}^m_{(I,J)} - \underline{v}^d_{(i,j)}||$$

Rucklidge [6] has reported a further modification of the standard Hausdorff distance which produces tangible performance improvements. His idea is to replace the max operator by an operator that selects from the set of attribute distances using a median test or f-th quantile test. His version of the Hausdorff distance can be written as

$$H_G^f(G_d, G_m) = F_{(i,j) \in E_d}^f \left\{ \min_{(I,J) \in E_m} ||\underline{v}^m_{(I,J)} - \underline{v}^d_{(i,j)}|| \right\}$$

where the operator $F_{(i,j) \in E_d}^f$ selects the f-th quantile value from the set of edge-wise attribute distances.

Rucklidge's median operator represents a robust procedure for selecting pairwise attribute differences. Our motivation in the previous section was slightly different, since we used robust error kernels to weight the attributes. Rather than performing a quantile test we sum the attribute weights. As a final primitive-based distance-measure, we have therefore considered applying the Hausdorff tests to the summed complement of the weighting function. The new distance measure is defined to be

$$H^P(G_d, G_m) = \sum_{i \in V_d} \max_{(i,j) \in E_d} \sum_{j \in C_i^d} \min_{(I,J) \in E_m} \left(1 - \Gamma_\sigma(||\underline{v}^m_{I,J} - \underline{v}^d_{i,j}||) \right) \qquad (5)$$

Finally, when recognition is being attempted over a large data-base of patterns, the model is taken to associate with the minimum Hausdorff distance set of data. The data item associated with the model is

$$\theta_m = \arg \min_{d \in \mathcal{D}} H(G_d, G_m)$$

5 Experiments

We have conducted our recognition experiments with a data-base of 2000 line-patterns each containing over a hundred lines. The line-patterns have been obtained by applying line/edge detection algorithms to the raw grey-scale images followed by polygonisation. For each line-pattern in the data-base, we construct the six-nearest neighbour graph.

The recognition task is posed as one of recovering the line-pattern which most closely resembles a digital map. The original images from which our line-patterns have been extracted have been obtained from a number of diverse sources. However, a subset of the images are aerial infra-red line-scan views of southern England. Two of these infra-red images correspond to different views of the area

(a) Digital Map   (b) Target 1   (c) Target 2
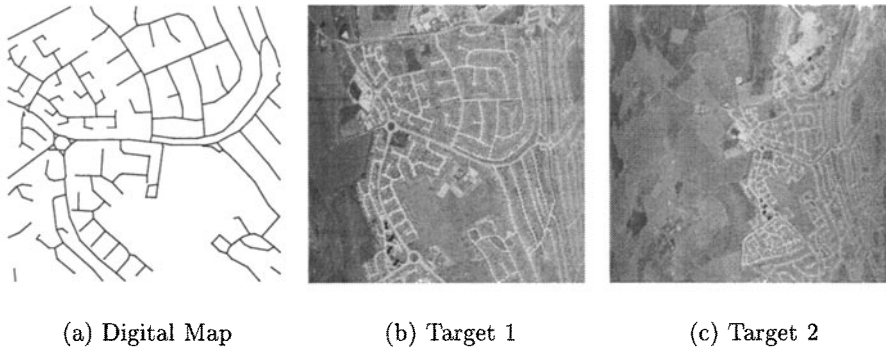
Fig. 2. Images from the data-base

covered by the digital map. These views are obtained when the line-scan device is flying at different altitudes. The line-scan device used to obtain the aerial images introduces severe barrel distortions.

In order to explore the sensitivity of our recognition method to segmentation systematics, we have introduced multiple segmentation of the target images into the data-base. These different segmentations have been obtained by maliciously adjusting the control parameters of the feature extraction algorithm. In total their are 10 different segmentations for each of the two target images. The digital map, the target infra-red images and some sample segmentations are shown in figure 2.
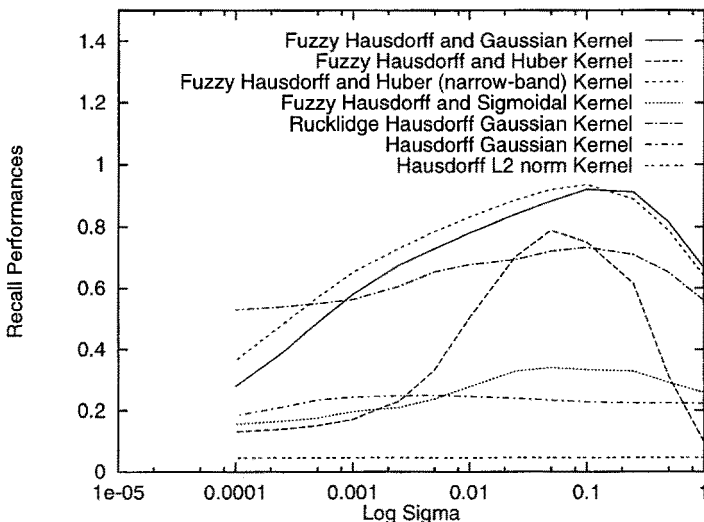
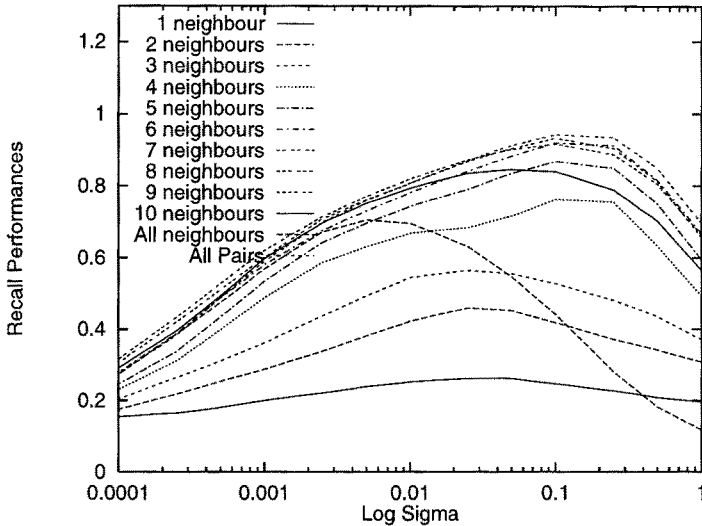**Fig. 3.** Relative recognition performance for various distance measures

Fig. 4. Relative recognition performance for various relational structure

Our first set of experiments aim to illustrate the relative recognition performance of the different distance measures. The performance of the system in terms of retrieval accuracy are assessed using the standard normalised IAVRR/AVRR recall metric [12] which is equal to 1 for perfect retrieval accuracy. For this experiment a database composed of 850 line patterns is used and the result shown represent the average retrieval accuracy of 100 distinct queries. Figure 3 shows the recognition performance as a function of the control parameter σ for each of the distance measures presented in section 4 in turn. From the figure it is clear that the best performance is obtained when the weighting kernel is either Gaussian or a modified narrow-band Huber. The poorest performance is obtained with the crisp Hausdorff distance coupled with the $L2$ norm. Rucklidge's modified Hausdorff distance (using median instead of max comparator [6] and a Gaussian kernel) does not provide an optimal recall performance for this particular task but presents an obvious improvement over the standard Hausdorff distance. It is important to note that the x-axis of the plot is logarithmic and therefore that recognition performance is not particularly sensitive to the kernel width parameter σ. From this graph it can also be seen that an average correct retrieval rate of 94% is achievable.

In the next set of experiments we illustrate the effect of relational structure on the recognition process. Figure 4 shows the recognition accuracy as a function of the width parameter σ for a number of different graph-structures. In this experiment we have used the similarity measure of equation (3) with the pairwise attributes weighted using the Gaussian error-kernel. a Gaussian kernel. Here we compare the performance obtained with N-nearest neighbour graphs of various orders. We also provide results for the recognition performance obtained when the relational constraints are weakened. In the first such example we relax the requirement for neighbourhood structure, and evaluate the similarity measure

over the complete space of edge-wise associations. In the second example we remove the edge-structure and compute the similarity measure over the complete space of pairwise associations. The first observation that can be drawn from this set of experiments is that the best recognition performance is obtained when the order of the nearest neighbour graph is seven. However, even when the order of the graph is small (i.e. one neighbour) or large (i.e. ten neighbours), then the recognition performance exceeds that obtained when either the neighbourhood structure or the edge-structure is ignored.
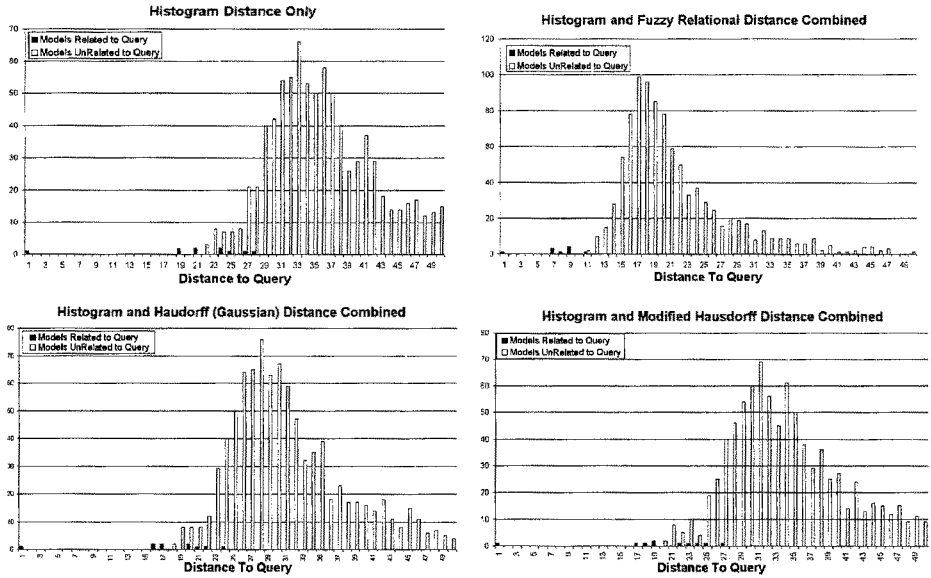
Fig. 5. Distribution of the distance measures during retrieval using the digital map

The final set of experiments focuses on the distribution of the distance measures. We have extracted from the data-base the 1000 best histogram matches for the digital map query and have used these for more detailed recognition experiments. Figure 5 compares histograms of the various distance measures for the target images and for the remaining line-patterns. Figure 5(a) shows the distribution of Bhatacharyya distance between the histograms in the data-base and the query image. The remaining three plots show the distribution of the primitive-based distance measures. Figure 5(b) shows the distribution of Bayesian similarity measure, Figure 5(c) shows the distribution of standard Hausdorff distance while Figure 5(d) shows the distribution of Rucklidge's median-based Hausdorff distance. The main feature to note from these distributions is that the Bayesian measure gives the best separation between the various segmentations of the two aerial images (shown in black) and the remainder of the data-base. In this case the 10 different aerial image segmentations are ranked above the rest of the data-base. In other words, there is no overlap with the remaining patterns in the data-

base. In each of the remaining three cases (i.e. Figures 5a,c and d) only five of the ten segmentations are top ranked. In other words, the primitive-based Hausdorff distances fair only as well as histogram-based comparison which overlooks the primitive structure of the line-patterns. Moreover, the Bayesian recognition process does not appear to be sensitive to the segmentation and polygonisation process used to extract the line-patterns from the two aerial images.

6 Conclusion

In this paper we have presented a new similarity measure for comparing relational object descriptions. The idea underpinning the measure is to gauge the similarity of the pairwise attributes residing on the edges of a graph-structure that represents the proximity structure of a set of object-primitives. The measure exploits the neighbourhood structure to limit the set of comparisons required.

For a database of 2000 objects (or line-patterns) we have shown that a recall accuracy of 94% is achievable when the weighting function is Gaussian. We have presented a number of experiments demonstrating the performance of the proposed methodology. Moreover, the results obtained indicate that the method is relatively insensitive to the under and over segmentation of the line-patterns. Moreover, the method consistently outperforms the Hausdorff distance in terms of its recognition performance.

References

1. Y. Lamdan and H. Wolfson, "Geometric hashing: A general and efficient model-based recognition scheme," *In Proc. of ICCV*, pp. 238–249, 1988.
2. M. Costa and L. Shapiro, "Scene analysis using appearance-based models and relational indexing," *In Proc. of ISCV*, pp. 103–108, 1995.
3. C. A. Rothwell, A. Zisserman, D. Forsyth, and J. Mundy, "Canonical frames for planar object recognition," *In Proc. of ECCV*, pp. 757–772, 1992.
4. Y. Landam, J. T. Schwartz, and H. J. Wolfson, "Object recognition by affine invariant matching," in *Proc. of CVPR Conf.*, pp. 335–344, 1988.
5. A. Evans, N. Thacker, and J. Mayhew, "The use of geometric histograms for model-based object recognition," *In Proc. of BMVC*, pp. 429–438, 1993.
6. W. Rucklidge, "Locating ojects using the Hausdorff distance," *In Proc. of ICCV*, pp. 457–464, 1995.
7. R. Wilson and E. R. Hancock, "Structural matching by discrete relaxation," *PAMI*, vol. 19, pp. 634–648, June 1997.
8. A. Cross and E. Hancock, "Recovering perspective pose with dual step EM algorithm," *Advances in NIPS*, vol. 10, 1998. MIT Press, to appear.
9. B. Huet and E. R. Hancock, "Cartographic indexing into a database of remotely sensed images," *In Proc. of WACV*, pp. 8–14, Dec 1996.
10. J. Kittler and E. Hancock, "Combining evidence in probabilistic relaxation," *Int. J. of Pattern Recognition and Artificial Intelligence*, vol. 3, no. 1, pp. 29–51, 1989.
11. W. Christmas, J. Kittler, and M. Petrou, "Structural matching in computer vision using probabilistic relaxation," *PAMI*, vol. 17, pp. 749–764, 1995.
12. A. Pentland, R. Picard, and S. Sclaroff, "Photobook: Content-based manipulation of image databases," *Int. J. of Computer Vision*, vol. 18, no. 3, pp. 233–254, 1996.