Fuzzy Relational Distance for Large-scale Object Recognition

Benoit Huet and Edwin R. Hancock
Department of Computer Science
University of York, York Y01 5DD, UK.
huetb@cs.york.ac.uk erh@cs.york.ac.uk

Abstract

This paper presents a new similarity measure for object recognition from large libraries of line-patterns. The measure draws its inspiration from both the Hausdorff distance and a recently reported Bayesian consistency measure that has been sucessfully used for graph-based correspondence matching. 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 1000 line-patterns. A sensitivity study reveals that the method is capable of delivering a recognition accuracy of 98%. 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. Our first observation is that since the object primitives under study are subject to both measurement uncertainty and segmentation error, fuzzy or probabilistic distance measures may be more appropriate to the comparison task. In particular, recent interest in the matching of relational graphs has furnished a methodology for exploiting contextual constraints [7, 8]. However, this methodology has not been exploited in large-scale object recognition tasks.

Viewed from these dual perspectives, the main contribution of this paper is to investigate a synergy of methodology. We aim to both offer alternatives to the Hausdorff distances as a recognition metric and to incorporate relational constraints into the matching process.

This graph-based recognition process can be viewed as an intermediate step in a coarse-to-fine object retrieval system. As we envisage the process, a set of candidate images is retrieved using a coarse-grained representation. In fact, in our previously reported work we have provided a tangible example of how such a set of candidates can be located by comparing pairwise geometric histograms [9]. Once overall object recognition has been achieved then detailed correspondences may be recovered. Here, techniques such as graph-matching [7] and pose estimation [8] can be used to verify recognition hypotheses and initiate new searches if necessary.

2 Object Representation

In this section we review our object representation. There are two aspects. The first of these is is the attribute or measurement content. Here we describe a set of Euclidean invariant pairwise geometric attributes based on relative angles and relative position. This representation has been successfully exploited in histogram-based object retrieval [10]. The second aspect of representation is structural. Here we use the six-nearest neighbour graph, since a recent sensitivity study reveals that it offers the best compromise between robustness to structural errors and computational overheads [11]. It is the edges of this graph-structure which provide simple contextual constraints on the recognition process.

2.1 Pairwise Geometric Attributes

In a recent study [9] we have shown that the angular relation between pairs of line segments is very robust to both Euclidean transformation and segmental clutter when line-pattern recall is being attempted.

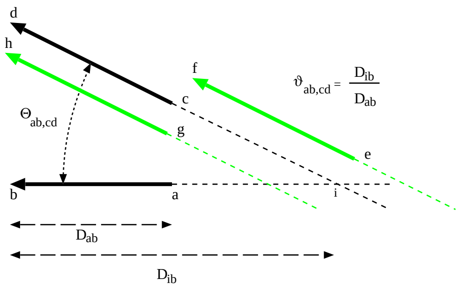


Figure 1: Geometry for shape representation

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]. It differs by 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 segments intersect at the middle point of the baseline.

2.2 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 or extra features and noise.

Our idea contrasts with a number of related contributions in the literature. For instance, Evans *et al* [5] effectively have a local representation of line-pattern structure which is less compact than ours since it employs one histogram per line-segment. Moreover, their attributes are not scale invariant. The idea of using a region-of-interest is close to that of using a neighbourhood graph. However, their regions are controlled by a scale parameter. By virtue of their structural character, neighbourhood graphs are scale-invariant.

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.

3 Recognition Metrics

In this section we commence by reviewing how the Hausdorff distance can be used for object recognition. Next we indicate how the distance can be extended to incorporate binary measurement relations. This leads us to consider two developments of the idea. The first of these is to show how relational constraints can be embedded in the measure to limit the set of potential associations. The second development is to develop a fuzzy variant of the Hausdorff distance where we soften the max-min operations which are used to make associations across the sets of observations being compared. second development is observations being compared.

3.1 Hausdorff distance

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. In other words, it provides a means of avoiding the computationally demanding problem of attempting to find correspondence matches between individual primitives whilst performing recognition.

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. by exploring the set of possible pairwise associations. In the case of pairwise attributes, the Hausdorff distance is given by

$$H_B(G_d, G_m) = \max_{(i,j) \in V_d \times V_d} \min_{(I,J) \in V_m \times V_m} ||\underline{\mathrm{v}}^m_{(I,J)} - \underline{\mathrm{v}}^d_{(i,j)}||$$

From the computational standpoint, this represents an increase in the number of associations that have to be compared. Whereas in the unary case there are $|V_m| \times |V_d|$ comparisons, in the pairwise case there are $|V_m|^2 \times |V_d|^2$ comparisons. Moreover, if recognition is being attempted then the large number of possible pairwise associations is likely to render the object representation highly ambiguous.

One way of overcoming these problems is to confine our attention to those pairwise measurement relations that are defined on the edges of the graphs representing the adjacency structure of the object primitives. With this goal in mind we can redefine the Hausdorff distance over the edge-sets of the model and data graphs. The modified object-distance is

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

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

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

3.2 Fuzzy distance measures

The feature which endows the Hausdorff distance with a degree of robustness is the idea of using the max and min tests. The max test selects on the basis of saliency while the min test selects on the basis of closeness. However, the manner in which the distance associates unordered elements is crisp. As a result it may become ineffective if there is either lack of closeness due to excessive noise or lack of saliency due to overlap.

One way of overcoming the problems of crispness is to make fuzzy associations. Rather than taking the inner min operation, we would like to weight associations according to their closeness. Unfortunately, if this is attempted using a standard distance norm, then when the outer max operation is performed, there will be a tendency to loose saliency. In other words, we need to choose a weighting function which saturates for large attribute distances rather than growing monotonically. The robust statistics literature furnishes several weighting functions which meet this requirement.

The graph-distance measure is developed from a Bayesian consistency criterion that has recently been used to locate correspondence matches by relaxation labelling. However, rather than iteratively updating correspondence matches, our aim is to use the consistency criterion in a non-iterative manner to the purposes of object recognition. The Bayesian consistency criterion measures the the similarity of the pairwise-attribute relations on the neighbourhoods of nodes in the model and data graphs. To formalise this first step suppose that the set of nodes connected to the model-graph node I is $C_I^m = \{J|(I,J) \in E_M\}$. The corresponding 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

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

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 $\underline{v}_{I,J}^m$ and $\underline{v}_{i,j}^d$.

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 development, we consider a very simple form for the structural error process. We assume that the conditional prior can be modelled as follows

$$P\left((i,j) \to (I,J)|\underline{v}_{I,J}^m, \underline{v}_{i,j}^d\right) = \begin{cases} \Gamma_\sigma(||\underline{v}_{I,J}^m - \underline{v}_{i,j}^d||) & \text{if } J = \arg\min_{J' \in C_I^M} ||\underline{v}_{I,J}^m - \underline{v}_{i,j'}^d|| \\ 0 & \text{otherwise} \end{cases}$$

where $\Gamma_\sigma(||\underline{u}_I^m - \underline{u}_i^d||)$ is a distance weighting function. The role of the min operator is to order the associations between the two edges in the two neighbourhoods under comparison. This removes the need to perform and explicit rotational permutation of the of the edges. This permutation operation has proved to provide the major computational overhead in our work on graph-matching [7].

The process of maximising the Bayesian consistency measure is equivalent of minimising the following relational-distance distance measure

$$S(G_d, G_m) = \sum_{i \in V_d} \max_{I \in V_m} \sum_{j \in C_i^d} \min_{J \in C_I^m} \left(1 - \Gamma_\sigma(||\underline{v}_{I,J}^m - \underline{v}_{i,j}^d||)\right)$$

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

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

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

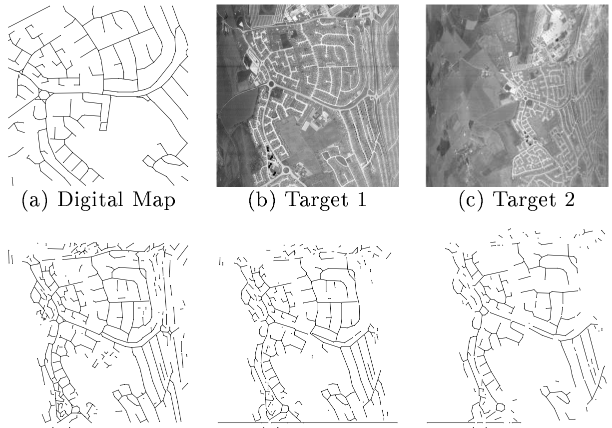
$$\Gamma_\sigma(\rho) = \left(1 + \frac{|\rho|}{\sigma}\right)^{-1}$$

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 Recognition

The practical goal in this paper is to incorporate the distance measures defined in the previous section into a hierarchical object recognition system. The overall goal is the recognition of complex line patterns from large data-bases of alternatives. The architecture that we have in mind is as follows: Recognition commences by comparing a compact shape representation to deliver a set of candidate objects. In our present work this compact object representation is a multidimensional histogram of pairwise Euclidean invariant geometric attributes. The candidates are then subjected to a more detailed comparison based on the attributes of the individual object primitives. It is here that we need a distance measure that captures the closeness and saliency of the pairwise attributes for the line patterns. Once the closest distance pattern has been identified, then detailed verification can be attempted by looking for individual correspondences between the primitives and assessing their relational consistency. Here we envisage using our recently reported framework for graph-matching.

Here we are interested in assessing the first two levels of this architecture. We use our recently reported work on compact line-pattern indexing to deliver a set of candidate patterns [9]. Here the patterns are represented using a two-dimensional pairwise geometric histogram of the directed relative angles and lengths described in Section 2. The contributions to the histogram are gated using relational constraints. In other words, a histogram entry is only made if a line-segment pair is connected by an edge in the nearest neighbour graph [10]. We compare the histogram bin contents using the Bhatacharyya distance [9]. Based on the ordered set of histogram distances we select the N-best matches for more detailed comparison.



(a) Digital Map (b) Target 1 (c) Target 2

(d) $\sigma = 15$ (e) $\sigma = 20$ (f) $\sigma = 25$

Figure 2: Images from the data-base and alternative segmentations

5 Experiments

We have conducted our recognition experiments with a data-base of 1000 line-patterns each containing over a hundred lines. The line-patterns have been obtained by applying feature/edge detection algorithms to the raw grey-scale images. 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 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 segmentations 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 there 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.

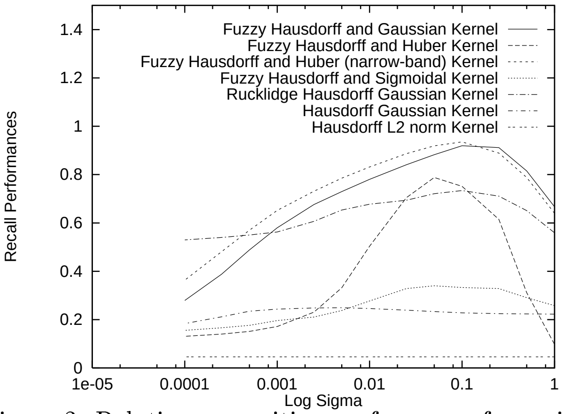


Figure 3: Relative recognition performance for various distance measures

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 3.2 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 98% is achievable.

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 4 compares the distribution of distances using our recently reported histogram-based recognition method [10] and the best fuzzy distance measure (using a gaussian ker-

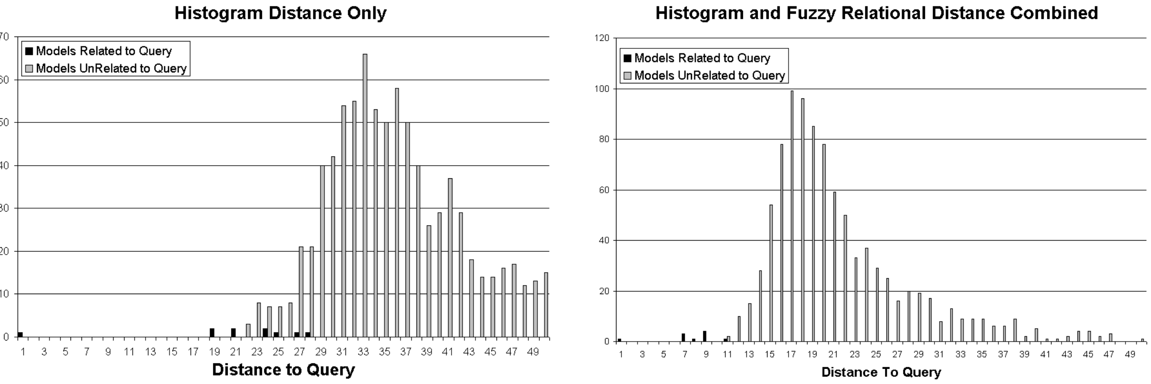


Figure 4: Distribution of the distance measures during retrieval using the digital map

nel). Here the target images are drawn in black while the remaining entries are in grey. In the case of the single histogram-based method there is a greater overlap between the two components than in the case of the fuzzy distance measures. It fact there is no overlap at all between the distances of the target images and the remainder of the database after using the fuzzy relational distance. Since the data-base was augmented with various segmentations of the target images (or line-patterns) and all these data-base entries have been ranked best, it can be noted that our approach is robust to changes in the segmentation and polygonisation process.

6 Conclusion

In this paper we have presented the intermediate level of a hierarchical large scale object recognition system. While the low level is aimed at rapidly indexing the model database using pairwise geometric histogram comparison, the intermediate level is devised to improve the accuracy of retrieval by performing correspondance matching. We have demonstrated how a modified Hausdorff distance measure can be applied to the problem of exploring the pairwise associations between model and data images.

For a database of 1000 objects (or line-patterns) we have shown that a recall accuracy of over 98% is achievable when the weighting function used by the fuzzy Hausdorff distance is is Gaussian. We have presented a number of experiments demontrating the performance of the proposed methodology. Moreover, the result obtained indicates that the method is relatively insensitive to the under and over segmentation of the line-patterns. Furthermore, the fuzzy relational distance used as a recognition metric provides a very effective way to reject or accept hypotheses provided by the less-refined lower level according to their local geometric and structural properties.

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," *International Symposium on Computer Vision*, pp. 103–108, 1995.

[3] C. A. Rothwell, A. Zisserman, D. Forsyth, and J. Mundy, "Canonical frames for planar object recognition," *Proc. of ECCV Conf.*, 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," *Proc. of BMVC Conf.*, pp. 429–438, Sept 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," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 634–648, June 1997.

[8] A. Cross and E. Hancock, "Recovering perspective pose with dual step EM algorithm," *Advances in Neural Information Processing Systems*, vol. 10, 1998. MIT Press, to appear.

[9] B. Huet and E. R. Hancock, "Cartographic indexing into a database of remotely sensed images," *IEEE Workshop on Applications of Computer Vision*, pp. 8–14, Dec 1996.

[10] B. Huet and E. Hancock, "Relational histograms for shape indexing," *In Proc. of ICCV*, pp. 563–569, Jan 1998.

[11] R. Wilson, A. Cross, and E. Hancock, "Sensitivity analysis for structural matching," *Proc. of ICPR Conf.*, vol. 1, pp. 62–66, August 1996.

[12] A. Pentland, R. Picard, and S. Sclaroff, "Photobook: Content-based manipulation of image databases," *International Journal of Computer Vision*, vol. 18, no. 3, pp. 233–254, 1996.