Structurally Gated Pairwise Geometric Histograms for Shape Indexing

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

Abstract

This paper presents a new method for shape indexing from large data-bases of line-patterns. The basic idea is to exploit both geometric attributes and structural information to construct a shape similarity measure. We realise this goal by computing the N-nearest neighbour graph for the lines-segments for each pattern. The edges of the neighbourhood graphs are used to gate contributions to a two-dimensional pairwise geometric histogram. Shapes are indexed by searching for the line-pattern that minimises the cross-correlation of the normalised histogram bin-contents. We evaluate the new method on a data-base containing 1000 line-patterns each composed of hundreds of lines. Here we demonstrate that the structural gating of the histogram not only improves recognition performance, but that it also overcomes the problem of saturation when large patterns are being recalled.

1 Introduction

Structural indexing has been demonstrated as an effective tool for object recognition in the domain of CAD-based vision [1, 12]. Specifically, is has been used to successfully recall object-models from large data-bases of engineering drawings The basic idea is to use the relational structure of object primitives to construct an index that can be used to discriminate object shape. Since efficiency is invariably a key issue, conventional structural matching methods are too computationally demanding to be used as search-engines. This rules out many promising computational techniques such as graph matching, structural hashing and relaxation labelling. Instead, the goal is to construct easily computed structural summaries that can be used to make a rapid comparison on the basis of salient object structure. Concrete examples include the relational indexing scheme of Costa and Shapiro [1]. In this approach, each model view is represented by sub-graphs of fixed size derived from the full relational graph. Each subgraph is then encoded into a hash-table for use subsequent use in indexing. The major drawback of this graph encoding scheme are the computational overheads which grow rapidly with the number of feature types and the cardinality of the relations.

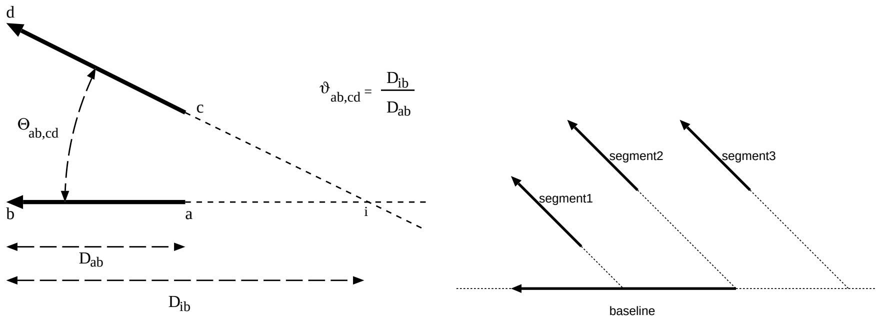
Outside the CAD domain, it is the attribute histogram that has been used as an indexation device. This idea was originally popularised by Swain and Ballard [14] for retrieving colour images from data-bases. The idea has been extended to both texture [5] and orientation representations [3]. A closely related representation is the pairwise geometric histogram [4]. Rather than using raw image attributes, this representation uses relative angles and relative perpendicular distances to recognise line patterns. In a recent study we have shown how a variant of the pairwise geometric histogram can be used to index into a data-base of line-patterns [7]. One of the key contributions of this study was to focus on the issue of how best to make a comparison of the histogram for accurate retrieval. The main conclusion drawn from this study was that statistical distance measures offer significant advantages over the more conventional L1 and L2 norms.

Although both structural indexing and attribute histograms have been separately used to recall image data from data-bases, they have not been used in conjunction. The aim in this paper is to describe a structurally gated pairwise attribute histogram representation that can be used to recall complex line-patterns. Typically, we aim to index patterns composed of hundreds of lines in data-bases of thousands of patterns. We draw our motivation from our experiences in the matching of large relational graphs [16]. Here we have explored the use of various relational structures to represent line adjacency. We have also demonstrated that the idea of comparing the edge structure of graphs can deliver acceptable matching performance when occlusion is a limiting factor and that the nearest neighbour graph is only marginally inferior to the Delaunay graph. In other words, acceptable measures of both relational structure and relational consistency can be computed inexpensively.

Our construction of the structurally gated pairwise histogram commences by extracting line-segments from the raw image data. We next construct the N-nearest neighbour graph for the line-centres. A sensitivity experiment reveals that the optimal order of this graph is six edges per node. With line-segments to hand we construct our pairwise attributes. Here we offer a novel departure by exploiting the directed relative angle and the directed relative position. The two attributes are binned by excluding from the histogram those entries that do not originate from line segments connected by an edge of the nearest-neighbour graph. This gating of the entries effectively allows us to encode structural information in the pairwise attribute histogram. This offers certain storage advantages over the related method of Evan's et al [4] where there is a single histogram for each line-segment.

The outline of this paper is as follows. We commence by describing the pairwise geometric attributes encoding the shape content of image. Specifically, we outline how the directed relative angle attribute is computed. Then, a novel attribute, which unambiguously describes pairwise relative position based on the ratio of the distance to intersection and the base segment length is introduced. Section 3 details the construction of the histograms representing the content of the image. Experimental evaluation of the technique is presented in Section 4. This takes the form of a comparative study aimed at establishing the effectiveness of the proposed algorithm on a data-base of over 1000 line-patterns. Finally, Section 5 presents some conclusions and suggest directions for future investigation.

2 The Geometric Attributes



(a) Computing the geometric attributes (b) Relative position of segments with respect to a baseline

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(a)). To illustrate how the Euclidean invariant 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.* [4], 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. This second attribute allows the representation to capture the structural information of the shape. Figure 1(b) shows a configuration where the relative angle attribute alone does not allow unambiguous representation. In this figure, the segments labelled 1, 2 and 3 all have the same relative angle with respect to the base line. The configuration may be unambiguously defined if the relative angle attribute is coupled with a directed relative position attribute that uniquely defines the local structure.

The directed relative position $\vartheta_{\underline{x}_{ab},\underline{x}_{cd}}$ is represented by the 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).

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

Both the directed relative angle and the directed relative position attributes are invariant to changes of scale, rotation and translation. This is an important advantage over the representation proposed by Evans *et al.* [4] if the technique is to be used for shape retrieval from real world image data-bases where the size of objects or, more generally shapes, is not known in advance.

3 Histogram Representation

The two dimensional histogram representation is based on the relative pairwise attributes described in Section 2. In this section we provide details about the construction of pairwise histograms from two invariant attributes together with the structural constraints provided a N-nearest neighbour graph. We also describe how the statistical distance measure is used to compare the histogram bin contents.

3.1 Pairwise Attributes

The 36 vertical bins of our histogram represent the directed relative angle varying from $-\pi$ to π. The 12 horizontal bins represent the directed relative position ranging from 0.5 to ∞. We use a non-uniform binning for this attribute to take account of three specific cases of line-geometry. The leftmost bin is reserved for line-segment pairs which intersect each other within the baseline. The second bin is reserved for line-segment pairs that intersect at the origin of the baseline segment. These line-pairs are potential junction candidates and convey important perceptual cues. The third reserved bin is rightmost in the histogram and contains parallel lines. The nine remaining horizontal bins uniformly represent line-segment pairs for which the intersection is outside the baseline with increasing distance from the baseline segment.

3.2 Relational Constraints

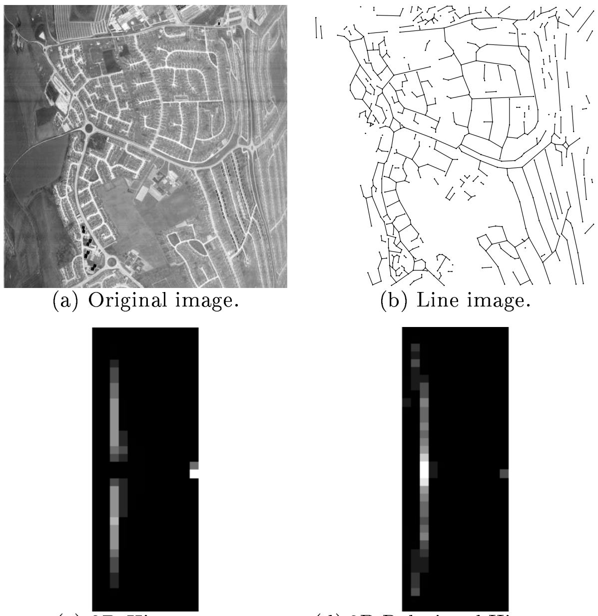
The novel idea introduced in this paper is to gate the entries in the pairwise geometric histogram using the edge-set of the N-nearest neighbour graph. Although our histograms integrate information over the complete line-pattern, contributions are gated by the local structure conveyed by a neighbourhood graph. The motivation here is that that local object representations are more robust to occlusion, missing or extra features and noise. Moreover, by integrating both geometric attributes and structural information into a single histogram we provide a compact relational representation of line-structure. This gating process greatly reduces the number of entries in each histogram, hence diminishing the risk of histogram saturation.

Our idea contrasts with a number of related contributions in the literature. For instance, Evans *et al* [4] 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. Di Mauro *et al's.* [2] 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.

Suppose that 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. With the edge-set of the nearest neighbour graph to hand, we can construct the structurally gated geometric histogram. The bin-incrementing process can be formally described as follows. Let i and j be two segments extracted from the raw image. The angle and position attributes θ_{ij} and ϑ_{ij} are binned provided the two segments are connected by an edge, i.e. $(i,j) \in E$. If this condition is met then the bin $H(\theta_{ij}, \vartheta_{ij})$ spanning the two attributes is incremented as follows

$$H(\theta_{ij}, \vartheta_{ij}) = \begin{cases} H(\theta_{ij}, \vartheta_{ij}) + 1 & \text{if } (i,j) \in E \\ H(\theta_{ij}, \vartheta_{ij}) & \text{otherwise} \end{cases}$$

Figure 2 shows the sequence of processing steps from an infra-red aerial image leading to its two dimensional histogram representation. Figure 2(a) is the raw image. Figure 2(b) is the result of applying straight-line detection to the output of a probabilistic relaxation line detector [6]. Finally, figure 2(c) shows the computed histogram based on 6 neighbour segment pair attributes. The horizontal axis corresponds to directed relative positions and the vertical axis encodes directed relative angles.



(a) Original image. (b) Line image.

(c) 2D Histogram. (d) 2D Relational Histogram.

Figure 2: A typical infra-red image going through the processing step leading to the histogram representation.

3.3 Histogram Comparison

In a recent study we have evaluated the relative performance of various distance measures for histogram comparison [7]. The main conclusion was that the Bhattacharyya distance performed noticeably better that the standard L_1 and L_2 norms when used for histogram indexation in a large data-base of line-patterns. We therefore adopt this distance measure in our experiments with the structurally gated histogram. Suppose that $h_Q(\theta_{ij}, \vartheta_{ij})$ is the *normalised* histogram for a query line-pattern and $h_D(\theta_{ij}, \vartheta_{ij})$ represents the corresponding normalised histogram contents of one of the histograms contained within the data-base. The Bhatacharyya distance between the two normalised histograms is given by

$$B(Q, D) = -\ln \sum_i \sum_j \sqrt{h_Q(\theta_{ij}, \vartheta_{ij}) \times h_D(\theta_{ij}, \vartheta_{ij})}$$

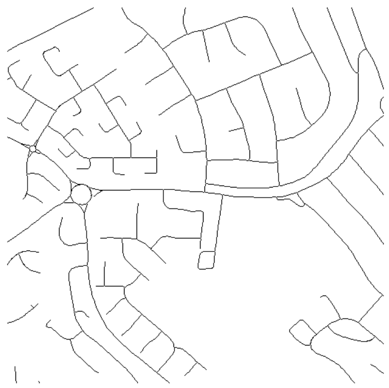


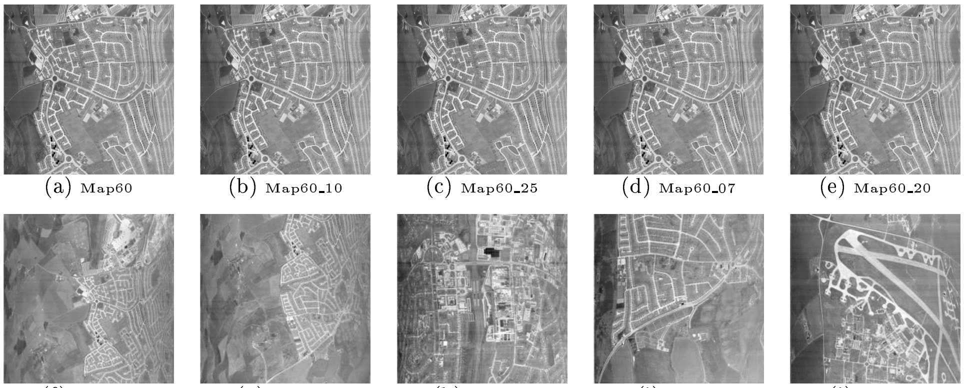
Figure 3: The digital map used for the experiments presented in this paper

4 Experiments

The examples included in this paper illustrate the effectiveness of the two dimensional relational histogram representations for performing shape based indexing into a large data-base of images. This task may be seen as indexing line patterns according to their shape similarity. The data-base used for the experiments has been constructed as described in the previous section. The data-base consists 28 infra-red aerial images and over 1000 trademarks and logos. Our first experiment aims to demonstrate the retrieval power and accuracy of the proposed methodology. The basic aim of such systems is to retrieve from a large data-base of images [9, 10, 14, 11], the n images that resemble most closely a query image. Here, the query image (see figure 3) is a *Digital Map*. This *Digital Map* corresponds to the area depicted on some of the infra-red aerial images. The second experiment aims to investigate the sensitivity of recall to the line segmentation process. We have therefore augmented the data-base with several sets of over and under segmented line-segments for the target infra-red images.

4.1 Data-Base Indexing

The line-patterns contained in the cartographic section of the data-base are infra-red aerial images of urban and semi-urban areas. The *Digital Map* (or cartographic model) used to query the data-base represents a road network known to appear in a subset of the images contained within the data-base. The images labelled *Map060* and *Map170* both contain the road network area depicted in the cartographic model. However, there is known to be a significant barrel distortion in the horizontal direction of the infra-red images caused by the rotating mirror optics underlying the line-scan process used during sensing. Additionally, the viewing angle, altitude and direction is not similar to the *Digital Map*. Ideally the result of a data-base query using the *Digital Map* depicted in figure 3 is a retrieval where both the images *Map060* and *Map170* are recalled as the first and second best matches.



(a) Map60 (b) Map60_10 (c) Map60_25 (d) Map60_07 (e) Map60_20

(f) Map170 (g) Map160 (h) Map100 (i) Map50 (j) Map120

Figure 4: The 10 Best match for retrieval based on the *Digital Map*.

Figure 4 shows the original images for the best 10 sets of line-segments that were recalled when we queried the complete data-base with the *Digital Map*. It can be seen that various segmentations of the infra-red image labelled *Map060* have been ranked first. The second best match is *Map170*. Various version of *Map060* can be found in the data-base (i.e.*Map060_07*). These correspond to different segmentations of the same original image (*Map060*). The result of controlled over and under segmentation of the image results in significant variation in the number of segments used to construct the histogram representation. The retrieval is not specifically affected by these segmentation systematics as can be seen in figure 4 [7].

Other retrieval experiments where conducted. However, their results are not presented here mostly because of lack of space. For example each image in the infra-red cartographic data-base was used to query the data-base and for all 28 images the correct match were positively identified.

4.2 Structural Sensitivity

Our second experiment is concerned with the identification of the most suitable relational structure for efficient and robust indexing regardless of the size of the data-base. The study presented in [15] suggested that nearest neighbour graphs of order six provided the best performance when full relational matching is performed. In this subsection, we aim to examine whether this assertion holds for shape indexing using structurally gated pairwise attribute histograms. Our concern here, is to determine the minimum order of the neighbourhood graph that will give the most robust recall when noise and clutter are limiting factors. Since we are dealing with a large number of images and each of these images may contain several hundred extracted segments, it is imperative that the order should be kept as small as possible.

4 neighbours		6 neighbours		8 neighbours		10 neighbours		N-1 neighbours	
Map60	0.053	Map60	0.038	Map60	0.029	Map60_20.0	0.036	56-1079.tif	0.005
Map60_10.0	0.059	Map60_25.0	0.038	Map60_20.0	0.029	Map60	0.038	56-1165.tif	0.005
Map60_20.0	0.060	Map60_20.0	0.039	Map60_25.0	0.029	Map60_25.0	0.039	44-1123.tif	0.008
Map60_07.0	0.062	Map60_10.0	0.041	Map60_10.0	0.034	Map60_10.0	0.040	Map060_25.0	0.008
Map60_25.0	0.064	Map60_07.0	0.047	Map60_07.0	0.037	Map50	0.043	54-404.tif	0.008
Map100	0.072	Map170	0.050	Map170	0.038	Map60_07.0	0.044	Map060_20.0	0.009
Map170	0.075	54-1187.tif	0.052	Map50	0.039	Map170	0.047	54-1195.tif	0.009
54-1187.tif	0.075	Map50	0.053	54-1187.tif	0.039	54-1187.tif	0.049	56-1301.tif	0.009
Map200	0.078	Map100	0.054	Map120	0.043	Map70	0.052	56-1078.tif	0.009
56-1121.tif	0.080	Map120	0.057	44-1154.tif	0.043	Map100	0.054	56-1109.tif	0.009
Map120	0.081	44-1154.tif	0.058	Map100	0.044	Map120	0.055	46-308.tif	0.009
Map50	0.081	56-1209.tif	0.058	56-906.tif	0.044	56-1209.tif	0.055	44-405.tif	0.009
56-1209.tif	0.082	44-1155.tif	0.058	54-113.tif	0.045	Map200	0.056	44-1063.tif	0.009

Table 1: Effect of the number of N-nearest neighbour on the retrieval performance

Table 1 presents the result of querying the image data-base using the *Digital Map* for a number of N-nearest neighbour graphs whose order ranges from $N = 4$ to $N = M - 1$ where M is the number of linear segments extracted from the raw image. Each entry shows the code-word assigned to the image together with its Bhatacharyya distance to the query image. The results in this table are ordered from the top-down in increasing distance. The top line indicates the image that resembles most closely the query image. The smallest recall distance is obtained when the structure is of order six. In the case when the orders of the graph are six and eight, then the alternative segmentations of *Map60* rank in the top five places of the table and are closely followed by *Map170* in rank six. The right-most column shows the retrieval obtained when all the segment pairs in the images are used to compute the histograms. This corresponds to removing the structural gating process from the histogram. In this case the highest ranks for the target image are five and seven. In other words, discarding the gating process results in poor recall.

In figure 5 we show the effect of the size of the data-base on the accuracy of recall. The plot shows the fraction of times the correct image is top ranked when 20 trials are repeated as a function of the size of the data-base. The solid curve shows the accuracy of recall for the structurally gated histogram when the neighbourhood graph of order six. The dashed curve is the corresponding result for the ungated histogram. Over, the range of data-base size, the performance of the gated histogram never falls below 95%. For the ungated histogram, the performance is consistently 30-40% poorer.

From these results, we can conclude that using a nearest neighbour graph of order six provides the most time efficient and accurate means of line-pattern recall.

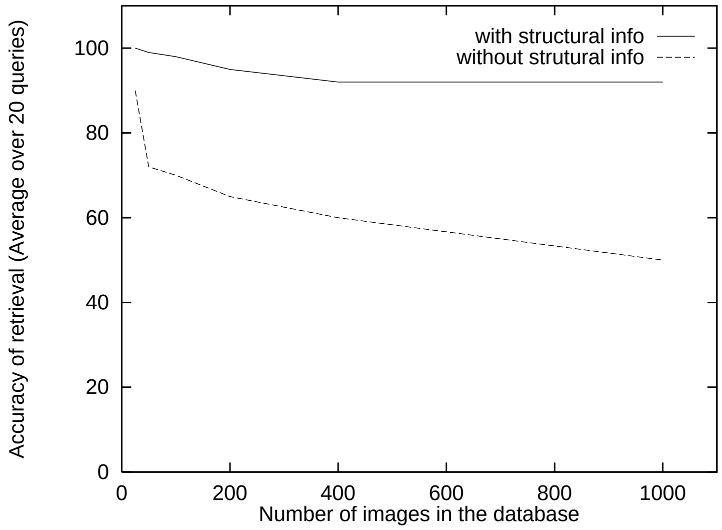


Figure 5: Effect of data-base size on accurate retrieval.

5 Discussion and Conclusion

The main contribution of this paper has been to describe a simple and compact histogram representation which combines geometrical and structural information for line-patterns. This representation can be used effectively used to index into a large database according to shape similarity. The retrieval is effected by computing the Bhatacharyya distance between histograms. For a data-base of over 1000 line-patterns, recall accuracy of over 95% is achievable. Moreover, we show that optimal performance is obtained when the nearest neighbour graph used to gate the histogram is of order six. There are also indications that the method is relatively insensitive to the under and over segmentation of the line-patterns.

The work presented in this paper may be regarded as an important step in a more ambitious programme aimed at developing a hierarchical indexing system for large image database [12]. Our next goal is to investigate the possibility of using high order statistics to measure attribute similarity. This would constitute an alternative to the histogram representation and would reduce the complexity of the search process. A second goal is to refine the rather crude index of structural similarity. In particular, we aim to derive fine-grain similarity measures from the recently reported Bayesian framework for relational graph matching of Wilson and Hancock [16].

References

[1] M. Costa and L. Shapiro. Scene analysis using appearance-based models and relational indexing. *Int. Symp. on Comp. Vision*, pages 103–108, 1995.

[2] E. DiMauro, T. Cootes, C. Taylor, and A. Lanitis. Active shape model search using pairwise geometric histograms. *BMVC'96*, 1:353–362, Sept 1996.

[3] C. Dorai and A. Jain. View organisation and matching of free-form objects. *Int. Symp. on Comp. Vision*, pages 25–30, 1995.

[4] A. Evans, N. Thacker, and J. Mayhew. The use of geometric histograms for model-based object recognition. *BMVC'93*, pages 429–438, 1993.

[5] G. Gimelfarb and A. Jain. On retrieving textured images from an image database. *Pattern Recognition*, 29(9):1461–1483, 1996.

[6] E. R. Hancock. Resolving edge-line ambiguities using probabilistic relaxation. *CVPR'93*, pages 300–306, 1993.

[7] B. Huet and E. R. Hancock. Cartographic indexing into a database of remotely sensed images. *WACV'96*, pages 8–14, Dec 1996.

[8] B. Huet and E. R. Hancock. Structural indexing of infra-red images using statistical histogram comparison. *IWISP'96*, pages 653–656, Nov 1996.

[9] W. Niblack *et al.* The QBIC project: Querying images by content using color, texture and shape. *Int. Symp. on Elect. Imaging*, pages 173–187, 1993.

[10] A. Pentland, R. Picard, and S. Scarloff. Photobook: tools for content-based manipulation of image databases. *Int. Symp. on Elect. Imaging*, pages 34–47, February 1994.

[11] R. Picard. Light-years from lena: Video and image libraries of the future. *ICIP'95*, 1:310–313, 1995.

[12] K. Sengupta and K. Boyer. Organising large structural databases. *PAMI*, 17(4):321–332, 1995.

[13] M. Stricker and M. Swain. The capacity of color histogram indexing. *CVPR'94*, pages 704–708, 1994.

[14] M. Swain. Interactive indexing into image databases. *Int. Symp. on Elect. Imaging*, pages 95–103, 1993.

[15] R. Wilson, A. Cross, and E. Hancock. Sensivity analysis for structural matching. *ICPR'96*, 1:62–66, 1996.

[16] R. Wilson and E. R. Hancock. Structural matching by discrete relaxation. *PAMI*, vol. 19, no. 6, pp. 634-648. 1997.