

Matching and Retrieval of Distorted and Occluded Shapes Using Dynamic Programming*

Euripides G.M. Petrakis[†] *Aristeidis Diplaros*[‡] *Evangelos Milios*[§]

March 25, 2002

Abstract

We propose an approach for matching distorted and possibly occluded shapes using Dynamic Programming (DP). We distinguish among various cases of matching such as cases where the shapes are scaled with respect to each other and cases where an open shape matches the whole or only a part of another open or closed shape. Our algorithm treats noise and shape distortions by allowing matching of merged sequences of consecutive small segments in a shape with larger segments of another shape, while being invariant to translation, scale, orientation and starting point selection. We illustrate the effectiveness of our algorithm in retrieval of shapes on two datasets of two-dimensional open and closed shapes of marine life species. We demonstrate the superiority of our approach over traditional approaches to shape matching and retrieval based on Fourier descriptors and moments. We also compare our method with SQUID, a well known method which is available on the Internet. Our evaluation is based on human relevance judgments following a well-established methodology from the information retrieval field.

*A preliminary version of this work was presented at the Int. Conf. on Pattern Recognition, Barcelona, Spain, pages 67-71, Vol. 4, Sept. 2000.

[†]Corresponding author. Department of Electronic and Computer Engineering, Technical University of Crete, Chania, Crete, GR-73100, Greece, E-mail: petrakis@ced.tuc.gr, URL: <http://www.ced.tuc.gr/~petrakis>,

[‡]Intelligent Sensory Information Systems, Faculty of Science, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands. E-mail: diplaros@science.uva.nl. This work is part of the author's student dissertation at the Department of Electronic and Computer Engineering of the Technical University of Crete.

[§]Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, Canada B3H 1W5. E-mail: eem@cs.dal.ca, URL: <http://www.cs.dal.ca/~eem>.

Index Terms: image database, shape retrieval, query by example, dynamic programming, relevance judgments.

1 Introduction

A wide range of shape recognition methods have been proposed and many of them have been implemented into commercial systems and experimental prototypes [1, 2]. Recently, the increasing amounts of image data in many application domains has generated additional interest for real-time management and retrieval of shapes from large collections of shapes referred to as shape or image databases [3, 4, 5]. The effectiveness of an image database system supporting retrievals by shape content depends on the types of shape representations used, the types of queries allowed and the efficiency of the shape matching techniques implemented. Image database systems must be capable of handling even unknown shapes and to respond to queries of arbitrary shape complexity such as queries by example (i.e., by providing an example shape or by drawing a sketch on the screen). To determine which shapes are similar to a given query, representations extracted from all stored shapes are matched with a similar representation computed to the query shape.

Below are some criteria for shape representation for reliable shape matching and retrieval: (a) *Uniqueness*: A representation must uniquely specify a shape; otherwise, a query may retrieve shapes which are not similar to it, although they have similar representations. (b) *Robustness*: A representation must be resistant to moderate amounts of distortion and noise, which are typical of real images of natural shapes. Because it is not possible to guarantee that a shape representation is not affected by such factors, distortions and noise should (at least) result in variations in the representations of similar magnitude. (c) *Invariance*: A representation must be invariant to viewing conditions that is, it must be invariant to translation, scale, rotation, viewing angle changes and symmetric transformations of the shapes. (d) *Scalability*: It must contain information about the shape at many levels of detail so that similar shapes can be recognized even if they appear at different view-scales (resolution). (e) *Efficiency*: A representation must be computationally efficient.

A shape matching algorithm must take advantage of the properties of its underlying representation and, in addition, it must be *accurate* (i.e., finds the similar objects with as few errors as possible). Shape representations alone are often insufficient to support accurate retrievals and are combined with color and texture features [6, 4, 5]. However, shape representation and matching remains a central problem in retrieving images by shape content.

This work focuses on shape matching. We propose a methodology for shape matching based on Dynamic Programming (DP). We assume that the shapes have already been extracted from images and are represented by their bounding contours. The basic idea behind our approach is to represent each shape by a sequence of convex and concave segments and to allow the matching of merged sequences of small segments in a noisy shape with larger segments in the other shape. Individual small or noisy segments are more likely to correspond to corrupted larger segments. Merging has a similar effect to that of smoothing several short segments in a shape to produce a single longer segment, but without actually performing the costly smoothing operation. The algorithm selects the most promising merges (the equivalent of degree of smoothing) based on local information. This is determined at run-time, as the algorithm searches for the least cost match (path) in a DP table.

Existing shape matching approaches address complex issues such as matching with occlusion (e.g., [7]), matching under excessive noise conditions (e.g., [8]) etc. However, most approaches suffer from one or more of the following drawbacks: (a) They work only for closed shapes (e.g., [8, 9, 10, 11]) assuming that whole shapes can always be extracted from images or, (b) They are sensitive to geometric transformations (e.g., [12, 7]) or, (c) They are not always optimal in that they may fail to find they may fail to find the least cost (correct) match and, even worse, they may fail to find a match, although one exists [10, 11]. Our proposed approach addresses all these issues and, in addition, through merging, treats noise and shape distortions or shapes at different levels of detail without performing the costly smoothing operation of the actual scale-based approaches [8, 10, 9]. The contributions of our shape matching approach are summarized in the following:

- It is always optimal, in that it always computes the least cost match.
- Handles both open and closed shapes uniformly (i.e., it doesn't require reconfiguration of the basic algorithm) and cases where one shape is open and matches only a part of the other (open or closed) shape. This is considered to be a far more difficult problem than whole-to-whole matching treated by most other methods (e.g., [10, 9, 11]) but also a much more interesting one, as it is unlikely in general to know in advance which parts of the two shapes are similar.
- Matching is independent of shape translation, scaling, rotation and starting point selection.

Our shape matching algorithm handles occlusion assuming that occlusion boundaries have been identified and segments belonging to the contours of single objects are available as open

shapes. Our method cannot handle the case of shapes consisting of multiple objects occluding each other or cases where parts of one shape may match parts of another shape. This is the most complex case of occlusion and it is still an unsolved problem.

A recent contribution by Gdalyahu and Weinshall [7] shares some commonalities with our approach in the use of dynamic programming to deal with shape matching and handles overlapping objects. Their formulation (based on polygonal approximations) and their matching approach are inherently different to ours. Their algorithm is not intrinsically invariant to scale and rotation and their merging scheme might be weak for noisy or very smooth shapes. A more elaborate discussion on commonalities and differences between this and our approach is presented in the next section.

We study the effectiveness of our proposed shape matching approach in retrieving similar objects from a shape database. Given a query shape, we want to find the n (e.g., 50) most similar shapes. We performed extensive performance comparisons using two data sets, one of 1,100 closed shapes and one of 1,500 open shapes of marine life species, already available as boundary contours.

Regarding retrieval by shape content, the contributions of our work are as follows:

- We establish the superiority of our approach for matching and retrieval of shapes with moderate amounts of noise and distortions over traditional methods such as, methods based on Fourier descriptors [13] and moments [14, 15, 16]. We also establish the superiority of our proposed method over our previous method [11] which, however, works only for closed shapes.
- We compare results obtained by our method with similar results obtained by SQUID [9]. SQUID is a well-established and well-researched approach to shape matching, and it is becoming accepted as a standard for whole shape matching. Our method demonstrates improved performance over SQUID. Although the improvement is not dramatic, our method has the advantage that is designed naturally for open shapes, which is not the case with SQUID. Extending SQUID to the matching of open or occluded shapes is non-trivial.
- We introduce a well established method from information retrieval for the empirical evaluation of retrieval results obtained by many competing methods [17]. The evaluations are based on human relevance judgments by four independent referees.

The rest of this work is organized as follows: A review of the work on object recognition and shape retrieval is presented in Section 2. The main idea behind our proposed method along with basic definitions of the DP table and cost functions are presented in Section 3. Our shape matching

algorithm is presented in Section 4. Finally, the evaluation method along with experimental results are presented in Section 5 followed by conclusions and issues for future research in Section 6.

2 Related Work

A wide range of shape recognition methods have been published [1, 2]. They are classified into *structural* (e.g., methods organizing local features into graphs [18, 19], trees [20] or strings [21]), *fuzzy* or *probabilistic* (e.g., relaxation methods [22]), *statistical* (e.g., methods based on moments [23]), methods that work in a transform domain (e.g., Fourier [13] or Hough [24]), methods based on Neural Networks [25] etc. Shape recognition methods are also classified into *local*, emphasizing local shape features (e.g., [18]) or *global*, representing the shape as a whole (e.g., [13]). Global methods are usually easy to compute and robust against noise and shape distortions; local methods are more complicated requiring sophisticated implementations and are slow but, are more suitable than global methods for recognizing occluded or partially visible objects.

An important class of contour tracking and matching methods relies on physical models of the deformation and is based on minimization of an energy function, without first extracting a symbolic representation of the shapes [26, 27, 28, 29]. Contour matching has been also addressed with dynamic programming together with detection of contours in image sequences. In [12, 30] dynamic programming is used to minimize a cost function that accounts for displacement of a contour in a pair of images from an image sequence. In [31, 32], dynamic programming is used to fit a closed curve template to an image (deformable template matching). However, these methods have been designed mainly for contour tracking over space or time and are not particularly well suited for shape retrieval. They assume that the curves which are matched be close to each other, they are sensitive to geometric transformations (i.e., scale, rotations, translations), they don't handle occlusion and, in certain cases, they fail to find the optimal match.

Another class of matching methods relies on symbolic entities extracted from shape contours [33, 34, 35]. Dynamic programming has been a popular approach for matching such symbolic entities [36, 37, 38, 39]. In [36], the inability of dynamic programming to combine contour segments is mentioned, as well as the fact that differing resolutions in the matched contours may lead to reduced performance. In [38], deletions and insertions of features (corners in a polygonal representation) as well as smoothing of features (i.e. dropping corners) is incorporated in the dynamic programming scheme. This type of smoothing lends a primitive multiple-scale character to

the method. In [37], dynamic programming is used to guide the application of grammar rules that transform one shape into another, in the spirit of [40]. In [39], matching proceeds both forward and backward from a support match between two features (landmark points) that are maximally similar. Features are extracted based on their persistence across scales. However, there is no matching of features at multiple scales. Therefore, multiple scales are used as a preprocessing stage only.

Multiscale methods are considered the most promising for shape matching because they are resistant to moderate amounts of deformation and noise, typical of real images of natural shapes. Different forms of scale space descriptions have been proposed [41, 42, 43]. In an earlier approach [8], matching is performed through “*interval trees*” which are computed by tracking the “*Curvature Scale Space*” (CSS) representation [41] from coarser to finer scales. In [35], which describes the matching mechanism of SQUID, only the maxima of the CSS curves are used. In [10], it is demonstrated that small shape changes may cause major structural changes in the interval tree and this may lead to matching errors. Recently, multiscale methods have been combined with dynamic programming [10, 11].

Building upon the previously mentioned work, [10] is a sophisticated dynamic programming algorithm which can group segments together in order to come up with appropriate correspondences. This algorithm uses the scale space representation of [41] to constrain the possible merges (i.e., it accepts merges that are only present at coarser scales of the scale space representation). The algorithm in [44, 11] is a substantial extension of the above algorithm to perform k -best search as it searches for best matches in the DP framework while avoiding the expensive computation of zero crossings in scale space. Algorithms such as [10, 44, 11] work only for closed shapes and, as observed in [45], they are not optimal, in that they may miss the optimal match, or even worse, they may fail to find a valid match altogether.

A recent contribution by Gdalyahu and Weinshall [7] shares some commonalities with our approach, in that it is motivated by an effort to avoid the high computational complexity of the true scale-space approaches. That algorithm, treats occlusion, while remaining computationally efficient. That paper uses a polygonal approximation, which may be weak in representing smooth shapes. The entities that are being associated are not convex and concave segments, but line segments of the polygonal approximation, defined by curvature extrema or extra points added to improve the approximation. This implies a finer-grained representation than ours. Gdalyahu and Weinshall use line segments as primitives with length and orientation as attributes. Merging is defined as simple vector addition, and is not constrained by an underlying shape grammar, therefore

it is less principled, and might lead to excessive flattening for smooth shapes. Their method seems to work well for noiseless shapes which are at the same scale and orientation and for moderate amounts of deformation. They do not use the edit distance for computing the overall similarity measure, but instead compute the residual distance between points after the matching. We instead have taken the approach of normalizing the edit distance for curve length.

Gdalyahu and Weinshall [7], like us, generalize string-based dynamic programming to account for merging, and introduce to the algorithm the notion of a gap to deal with occlusion. Their algorithm is not intrinsically invariant to scale and rotation but it can handle objects overlapping each other. Extending our algorithm to handle such occlusion is possible (e.g., by introducing a cost of deleting a sequence of segments in one of the two shapes in our cost formulas) but non-trivial. An issue with [7] involves the use of a heuristic to locate the most promising starting segments for matching which also, determines the orientation and scale that align the shapes which are matched. In case the heuristic fails, for example if it selects two inappropriate starting segments, the computation of orientation and scale will be inaccurate and the result of matching will be inaccurate too. Our algorithm is intrinsically invariant to rotation and, similarly to theirs, invariant to translation and scaling.

In our work, we propose a mechanism for computing the attributes for the merged segments, which allows us to bypass the explicit and expensive computation of the scale-space representation. We introduce operators on invariant attributes that are equivalent to smoothing and we define merging operations on attributes (by introducing an analytical relationship between the attributes of the segments being merged to the attributes of the merged segment). This is considered “impossible” by Gdalyahu and Weinshall [7] and as a serious drawback of the method by Ueda and Suzuki [10]. Finally, we present extensive performance comparisons with other methods using much larger datasets than the sets used in [7] and in [10].

It is an interesting future project to have a thorough comparative experimental evaluation with the above two methods. However, the underlying matching algorithms of these methods are quite complicated and their implementation is non-trivial. In [46] we present performance comparisons of our previous non-optimal DP-matching approach with the method by Ueda and Suzuki [10].

3 Proposed Methodology

The shape matching algorithm that lies at the core of our methodology takes in two shapes, determines whether the shapes are open or closed, and computes: (a) Their distance; the more similar the shapes are, the lower the value of the distance function and (b) The correspondences between similar parts of the two shapes. In retrievals, only distances between shapes are used. However, the correspondences help assess the plausibility of the distance computation, if necessary.

In matching two shapes A and B , the algorithm builds a Dynamic Programming (DP) table (Figure 1), where rows and columns correspond to inflection points of A and B respectively. Starting at a cell at the bottom row and proceeding upwards and to the right, the table is filled with the cost of the partial match containing the segments between the inflection points (rows and columns) swept so far. Because convex segments cannot match concave ones [37], only about half the cells are assigned cost values, in a checkerboard pattern. Merges, where a segment sequence of one shape matches a single segment of the other shape, can occur. Merges introduce “jumps” in the traversal of the DP table. Reaching the top row implies a complete match, where all inflection points of shape A have been swept. Additional information is stored in each cell to allow the tracing of a path starting from that cell and working backwards. The tracing of a path reveals segment associations between the two shapes. Dynamic Programming is used to find the minimum cost path from a cell in the initialization area to one in the termination area. In the remainder of this section we will define more precisely the algorithm and associated data structures.

3.1 Shape Representation

We assume that all objects are segmented into closed contours which are approximated by polygons. Automatic shape extraction from images (e.g., via region segmentation or edge following) is a non-trivial problem, and it is outside the scope of this paper. For our purposes, the images are already available in the desired polygonal form. The curvature of the shape polygons is computed via smooth approximations obtained by local cubic B-splines [47]: Inflection points are computed and the shape is segmented into convex and concave segments. The smoothness of this approximation is controlled by a parameter (called “tension”) which, in this work is set to 1 (i.e., this value seems to be a good compromise between loss of shape information and noise smoothing). Noise and distortions remaining after this smoothing can still be treated by the matching algorithm through merging. Notice that, inflection points computed on such B-spline approximated curves

are more stable than inflection points computed on raw contour data.

Let A and B be the two shapes to be matched. Elements of A and B are indexed by i and j respectively ($i, j \geq 1$); inflection points are denoted by p_i and p_j . $A = a_1, a_2, \dots, a_M$ and $B = b_1, b_2, \dots, b_N$ denote the sequences of N and M convex (C) and concave (V) segments of the two shapes respectively, with a_i being the segment between inflection points p_i and p_{i+1} and b_j the segment between inflection points q_j and q_{j+1} . Henceforth, $a(i - m|i)$, $m \geq 0$, denotes the sequence of segments $a_{i-m}, a_{i-m+1}, \dots, a_i$; similarly for $b(j - n|j)$, $n \geq 0$. If shape A (or B) is closed, then $p_1 = p_{M+1}$ (or $q_1 = q_{N+1}$). This implies that the number of inflection points in closed shapes equals the number of segments. If shape A (or B) is open, then $p_1 \neq p_{M+1}$ (or $q_1 \neq q_{N+1}$) and the number of inflection points equals $M + 1$ (or $N + 1$).

3.2 Matching Cases

Let A and B be the two shapes to be matched. We distinguish between the following two cases of matching:

Global: The algorithm will find the *best* mapping between segments in A and segments in B so that, no segments remain unassociated in either shape.

Local: The algorithm will find the *best* association of all segments of A to all or to a subsequence of segments of B (i.e., part of B may be left unmatched) and vice versa. Computing an appropriate scale for matching and finding which part of a shape matches the other shape, are the key issues in this case.

We focus our attention on the more general case of shape matching, that of local matching. However, it is more difficult to handle: The two shapes may be scaled with respect to each other and it is not possible to know in advance which shape is included within the other. Our proposed algorithm resolves both these issues.

Shapes A and B can be either open or closed. Based on this information, we consider the following matching cases as being of more practical interest:

Both shapes are open: Matching is local. Because we cannot know in advance which shape is included within the other one, we run the algorithm twice (i.e., once for each possibility) and we take the matching with the minimum cost. Local matching will also consider the case where all segments from both shapes are matched (global matching).

Shape A is open and shape B is closed: Matching is local. Shape A may be contained within shape B , but not the other way around (part of B may be left unmatched). Again, this includes the case where shape A matches the whole shape B (global matching).

Both shapes are closed: Matching is global. This case reduces to the previous one by pretending that A is open (B is closed), repeating the algorithm for global open and closed shape matching for each possible starting point on A , and by taking the least cost match as the cost of matching (see also Section 4). Notice that, regardless of cost, local matching would make no sense here.

3.3 Dynamic Programming (DP) Table

The DP table has \mathcal{M} rows and \mathcal{N} columns, where \mathcal{M} and \mathcal{N} are defined as follows:

Both shapes are open: $\mathcal{M} = M + 1$ and $\mathcal{N} = N + 1$.

Shape A is open and shape B is closed: $\mathcal{M} = M + 1$ and $\mathcal{N} = 2N$. Shape B is traversed twice to force the algorithm consider all possible starting points on B . If A is closed and B is open, we switch the roles of A and B .

Both shapes are closed: This case reduces to the previous one.

The rows of a DP table are indexed by i , $1 \leq i \leq \mathcal{M}$ and its columns are indexed by j , $1 \leq j \leq \mathcal{N}$ where, i, j are indices to inflection points of A and B respectively. If shape B is closed, its indices are taken modulo N . The cell at the intersection of rows i and column j is referred to as $cell(i, j)$. A link between cells (i_{w-1}, j_{w-1}) and (i_w, j_w) denotes the matching of the merged sequence of segments $a(i_{w-1}|i_w)$ with $b(j_{w-1}|j_w)$. $cell(i_{w-1}, j_{w-1})$ is called *parent* of $cell(i_w, j_w)$.

A *path* is a linked sequence of cells $((i_0, j_0), (i_1, j_1), \dots, (i_t, j_t))$, not necessarily adjacent, indicating a partial match, where $1 \leq i_0 < i_1 < \dots < i_t \leq \mathcal{M}$ and $1 \leq j_0 < j_1 < \dots < j_t \leq \mathcal{N}$. Index t denotes number of associations (between segments or between groups of segments) along the path. This path begins at inflection point p_{i_0} of shape A and at inflection point q_{j_0} of shape B and tries to match sequences of segments $a(i_{w-1}|i_w)$ of A with sequences $b(j_{w-1}|j_w)$ of B for $w = 1, 2, \dots, t$.

Each $cell(i_w, j_w)$ contains the following values: $g(i_w, j_w)$, i_{w-1} , j_{w-1} , u_w , v_w and ρ_w where $g(i_w, j_w)$ is the partially accumulated match cost up to that cell, u_w and v_w denote number of

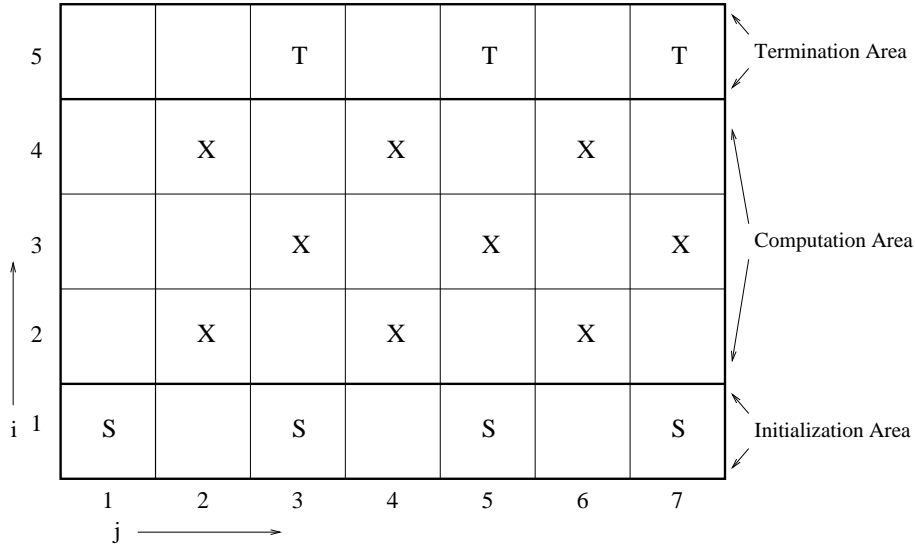


Figure 1: Example of a DP table with $M = 5$ (shape A) and $N = 7$ (shape B). S , X and T denote cells in the initialization, computation and termination areas respectively.

unmatched segments of A and B respectively, i_{w-1} and j_{w-1} are the indices of the parent cell of $cell(i_w, j_w)$ and are used to trace back a complete path. Finally, ρ_w denotes the scale factor corresponding to the parts of A and B which have been matched up to $cell(i_w, j_w)$ and it is defined in Section 3.4.

Figure 1 illustrates an example of a DP table. The DP table consists of three distinct areas:

Initialization area: It is the first row of the DP table. All paths start from cells in this area.

Matching starts always at the first segment a_1 of A ($i_0 = 1$). Matching may start at any segment b_{j_0} of B , where $1 \leq j_0 \leq N$. If a_1 and b_{j_0} have the same polarity, then $g(1, j_0)$, i_{w-1} , j_{w-1} , u_w , v_w , ρ_w are $0, 0, 0, M, N, 1$ respectively; otherwise, $g(1, j_0) = \infty$.

Computation area: It is the area between the first and last row of the DP table. Cells in this area correspond to incomplete paths.

Termination area: It is the last row of the DP table. All complete paths end at cells in this area.

The best match corresponds to the path with the least cost.

Notice that about half of the cells of the above DP table are empty; this is because associations between opposite type segments (i.e., C and V) are not allowed [37]. By convention, the cost of matching C with V segments is infinite. Matching always starts at the first inflection point of A while any point of B is a candidate starting point. Figure 1 implies that the first segments of A and B have the same polarity; otherwise matching will start from the second segment of B .

3.4 Distance Function

A *complete match* is a correspondence between sequences of segments in order, such that no segments are left unassociated in shape A and there are no crossovers or omissions. A complete match is characterized by a *complete path* $((i_0, j_0), (i_1, j_1), \dots, (i_T, j_T))$ in the DP table, i.e. a path that starts at the initialization and ends at the termination area. The cost $D(A, B)$ of matching shape A with shape B is defined as

$$D(A, B) = \min_T \{g(i_T, j_T)\}, \quad (1)$$

where $g(i_T, j_T)$ is the cost of a complete match. In turn, $g(i_T, j_T)$ is defined as follows:

$$g(i_T, j_T) = \min_{(i_w, j_w)} \sum_{w=1}^T \psi(a(i_{w-1}|i_w), b(j_{w-1}|j_w)). \quad (2)$$

Function $\psi(a(i_{w-1}|i_w), b(j_{w-1}|j_w))$ represents the dissimilarity cost of its two arguments and consists of three additive components:

$$\begin{aligned} \psi(a(i_{w-1}|i_w), b(j_{w-1}|j_w)) = & \\ & \lambda \text{MergingCost}(a(i_{w-1}|i_w)) + \\ & \lambda \text{MergingCost}(b(j_{w-1}|j_w)) + \\ & \text{DissimilarityCost}(a(i_{w-1}|i_w), b(j_{w-1}|j_w)). \end{aligned} \quad (3)$$

The first two terms in Equation 3 represent the cost of merging segments $a(i_{w-1}|i_w)$ in shape A and segments $b(j_{w-1}|j_w)$ in shape B respectively while, the last term is the cost of associating the merged sequence $a(i_{w-1}|i_w)$ with the merged sequence $b(j_{w-1}|j_w)$.

Each allowable merging should be a recursive application of the grammar rules $CVC \Rightarrow C$ and $VCV \Rightarrow V$ [37]. This is enforced by the DP algorithm. Constant λ represents the relative importance of the merging and dissimilarity costs. Low values of λ encourage merging and, conversely, high values of λ inhibit merging. For example, matching shapes with much detail must employ low values of λ . A method for the experimental specification of an appropriate value for λ is discussed in Section 5.2.1.

3.5 Geometric Quantities

We now define geometric quantities (features), as illustrated in Figure 2, that are required in the definition of the cost functions.

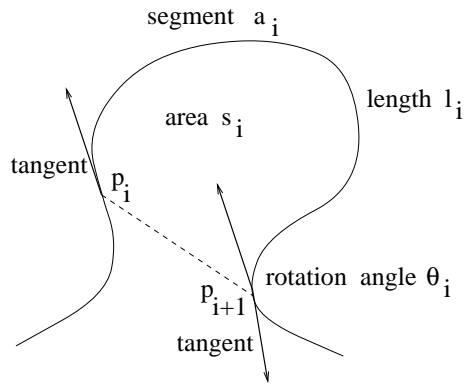


Figure 2: *Geometric quantities for defining the importance of a segment*

Rotation Angle θ_i is the angle traversed by the tangent to the segment from inflection point p_i to inflection point p_{i+1} and shows how strongly a segment is curved. θ_i is positive for convex and negative for concave segments.

Length l_i is the length of segment a_i .

Area s_i is the area enclosed between the chord and the arc between the inflection points p_i and p_{i+1} .

3.6 Scale Factor

If one of the two shapes is scaled with respect to the other, then the length of one of the two shapes (i.e., shape B) has to be multiplied by an appropriate *scale factor*. This scale factor can be computed as the ratio of the lengths of the matched parts of shapes A and B respectively. The definition of scale factor depends on the type of matching as follows:

Global matching: Shape A matches the whole shape B . The algorithm consumes all segments from both shapes. The scale factor is constant and is computed as

$$\rho = \frac{\text{length of } A}{\text{length of } B}. \quad (4)$$

Equivalently, we can normalize initially both shapes with respect to their perimeter. It is easy to accommodate this in our method by setting all scale terms ρ to 1 in the algorithm.

Local matching: Shape A may match either the whole or only a part of shape B . This case is more difficult to handle but it is more general and includes the previous one (i.e., when matching the whole shape B yields the least cost). Although we know that A matches

completely, the matched portion of shape B is unknown before the algorithm terminates. To handle this problem, we introduce a scale factor ρ_t , that is estimated for each partial match $((i_0, j_0), (i_1, j_1), \dots, (i_{t-1}, j_{t-1}))$, corresponding to matched parts so far (i.e., up to $t - 1$):

$$\rho_t = \frac{\sum_{w=1}^{t-1} \sum_{i=i_{w-1}}^{i_w-1} l_i(A)}{\sum_{w=1}^{t-1} \sum_{j=j_{w-1}}^{j_w-1} l_j(B)}, \quad (5)$$

where $2 \leq t \leq T$ and $l_i(A)$ and $l_j(B)$ are the lengths of a_i and b_j respectively. This value is an approximation of the actual scale factor of a complete match. Notice that ρ_1 is undefined since the total matched length is 0 for both shapes. In this work ρ_1 is set to 1.

3.7 Dissimilarity Cost

The dissimilarity cost of associating a group of segments from shape A with a group of segments from shape B is computed as

$$DissimilarityCost = U \max_{all\ features\ f} \{d_f\}. \quad (6)$$

The term d_f is the cost associated with the difference in feature f (i.e., length, area or angle). The intuition behind the use of *max* is that it tends to emphasize large differences on any feature. We choose the max operation instead of product [10] because in the product, a small cost in terms of one feature can cancel the effect of a high cost in terms of another feature, something that may lead to a visually implausible outcome. The max operation addresses this problem.

U is a weight term associated with the importance of this partial match. U emphasizes the importance of matching large parts from both shapes similarly to the way humans pay more attention on large shape parts when judging the quality of matching. The proportion of the matched shape length with respect to total length is used to define U :

$$U = \max \left\{ \frac{\sum_{i=i_{w-1}}^{i_w} l_i(A)}{length\ of\ A}, \frac{\sum_{j=j_{w-1}}^{j_w} l_j(B)}{length\ of\ B} \right\}. \quad (7)$$

The term d_f is defined as

$$d_f = \frac{|F_A - S_w(f)F_B|}{F_A + S_w(f)F_B}, \quad (8)$$

where, $F_A = \sum_{i=i_{w-1}}^{i_w} |f_i|$, $F_B = \sum_{j=j_{w-1}}^{j_w} |f_j|$ and $S_w(f)$ is a parameter depending on the feature f . Specifically $S_w(f) = \rho_w$ for f being length and ρ_w^2 for f being area. ρ_w is computed according to Equation 5 or Equation 4 for local and global matching respectively. For f being rotation angle, $S_w(f) = 1$, since angle measurements do not depend on a scale factor.

3.8 Merging Cost

Let the types of the segments being merged be $CVC \dots C$, leading to a single merged convex segment C by absorbing the concave segments in between. The opposite case is obtained by switching C and V in the formulas. The merging cost is defined as follows:

$$MergingCost = \max_{all\ features\ f} \{U_f C_f\}, \quad (9)$$

where subscript f refers to a feature (length, area or rotation angle).

For all features:

$$C_f = \frac{\sum_{V\ segs\ of\ group} |f|}{\sum_{all\ segs\ of\ group} |f|}. \quad (10)$$

where the sum of the numerator is over the absorbed concave segments, whereas the sum of the denominator is over all segments of the group. The intuition behind this formula is to measure the importance of the absorbed segments (of type V) relative to the whole matched consecutive segments of the group.

For f being any feature (length, area, rotation angle) the weight term of the merging cost is defined as

$$U_f = \frac{\sum_{V\ segs\ of\ group} |f|}{\sum_{V\ segs\ of\ shape} |f|}. \quad (11)$$

where the sum of the numerator is over the absorbed concave segments, while the sum of the denominator is over all concave segments of the shape. The intuition behind this weight term is to measure the importance of the absorbed segments (of type V) within the shape as a whole.

We choose the maximization formula in Equation 9 instead of sum of products of terms comparing consecutive segments [10], because in a product, a small cost in terms of one feature can cancel the effect of a high cost in terms of another. Another drawback of the use of a sum is that the merging cost increases with the number of segments merged, even if several very short segments are being merged into a large one.

4 Algorithm

Let A and B be the two shapes to be matched. A is assumed to be open; B can be either open or closed. If both shapes are closed, we assume that A is open and we attempt to match the open A on the closed B . Each point of A is a candidate starting point for matching. Matching starts with segments having the same polarity (C or V). There are $M/2$ such segments (i.e. potential starting

points) on A and the algorithm is repeated $M/2$ times. In the following, we assume that the first segment of B has the same polarity (C or V) with the first segment of A ; otherwise, matching starts at the second segment of B . The last matched segments of A and B must have the same polarity too.

The following summarizes the above discussion (“S”, “X” and “T” denote cells in the initialization, computation and termination areas respectively as in Figure 1):

Global matching: The algorithm consumes all segments from both shapes. Equivalently, the algorithm starts at the left-most cell (marked “S”), of the DP table, proceeds upwards and to the right through cells of computation area (marked “X”) and terminates at the right-most filled (marked “T”) cell of the DP table corresponding to same polarity segments of A and B . This cell contains the overall cost of matching (i.e., there is not need to search the top row for the least cost match). The scale factor is computed according to Equation 4.

Local matching: Any segment on shape B is a candidate starting segment for matching provided that it has the same polarity with the first segment of A . Only half of the cells (marked “S”) in the initialization area are candidate cells of starting a path. The algorithm consumes some or all segments of B and may end at any segment of B having the same polarity with the last segment of A . Therefore, half of the cells (marked “T”) at the termination area are candidate termination cells of a complete match path. All these cells must be searched to select the least cost match (best match). The scale factor is computed according to Equation 5.

Figure 3 outlines the matching algorithm. The algorithm computes the distance $D(A, B)$ of its two input shapes. The *for* loop for j_w does not run over all the indicated values, as convex to concave matches are not possible (only half of the cells are used) At each cell, the algorithm computes the optimum cost of the incomplete path ending at this cell:

$$g(i_w, j_w) = \min \{g(i_{w-1}, j_{w-1}) + \psi(a(i_{w-1}|i_w), b(j_{w-1}|j_w))\}, \quad (12)$$

where the minimum is over all possible values of (i_{w-1}, j_{w-1}) . Merging always involves an odd number of segments that is, $(i_{w-1}, j_{w-1}) = (i_w - 2m_w - 1, j_w - 2n_w - 1)$, where $m_w \geq 0$ and $n_w \geq 0$. Equation 12 determines the minimum cost transition from cell $cell(i_{w-1}, j_{w-1})$ to $cell(i_w, j_w)$ for all possible values of i_{w-1} and j_{w-1} . Indices i_{w-1} and j_{w-1} are stored in $cell(i_w, j_w)$ and can be used to retrace the path from $cell(i_w, j_w)$ back to its starting point.


```

Input: Shapes  $A = a_1, a_2, \dots, a_M, B = b_1, b_2, \dots, b_N$ ;
Output: Distance  $D(A, B)$  and correspondences between segments;
// Initialization: Fill the first row
for  $j_0 = 1, 2, \dots, \mathcal{N}$  do
    if  $a_1$  and  $b_{j_0}$  are both  $C$  or  $V$  then  $cell(1, j_0) = (0, 0, 0, M, N, 1)$ ;
    otherwise  $cell(1, j_0) = (\infty, 0, 0, M, N, 1)$ ;
end for
// Fill from the 2nd to the  $\mathcal{M}$ -th row
for  $i_w = 2, 3, \dots, \mathcal{M}$  do
    for  $j_w = 2, 3, \dots, \mathcal{N}$  do
        if  $a_{i_w}$  and  $b_{j_w}$  are both  $C$  or  $V$  then fill  $cell(i_w, j_w)$  using Equation 12;
        compute  $\rho_w$  using Equation 4 or Equation 5;
    end for
end for
// Select the least cost complete path
    select the least cost path from the  $\mathcal{M}$ -th row;
    retrace path using  $i_{w-1}, j_{w-1}$  cell values;

```

Figure 3: *Outline of the algorithm.*

The algorithm of Figure 3 is optimal, in that it always finds the path with the least cost from the initialization to the termination area that fully matches shape A with a portion of shape B [45]. It is worth noting that the algorithm for closed shape matching is optimal too.

Equation 12 implies that the algorithm computes the minimum cost transition from each allowable cell (i_{w-1}, j_{w-1}) to cell (i_w, j_w) . However, the algorithm may become very slow especially on large DP tables. Notice that, transitions on the DP table correspond to merges of segments. The algorithm examines all merges, even the less plausible ones, such as merges involving all segments. It seems reasonable to restrict the maximum number of segments which are allowed to merge to a constant K ($i_w - i_{w-1} = K, j_w - j_{w-1} = K$). K is always an odd number and for global matching has to be $K \geq M/N$ ($M \geq N$). The results in Section 5.2 indicates that the algorithm exchanges a small loss of accuracy for much faster retrieval: The algorithm may miss the least cost match if this involves merging of more than K segments.

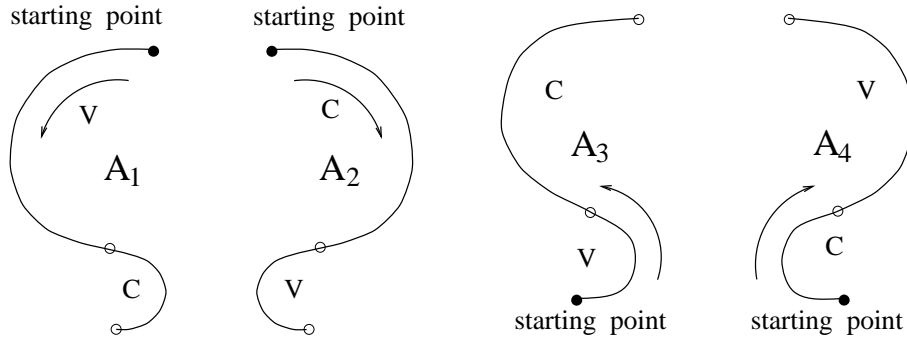


Figure 4: *Curve representation cases: A_1 is the original curve, A_2 is its mirror image, A_3 illustrates curve traversal in the opposite direction and A_4 is the mirror image of A with opposite traversal.*

4.1 Invariance to Shape Transformations

The matching algorithm must be capable of handling symmetric shapes and alternative curve traversals. Figure 4 illustrates all these cases for an open curve A : A_1 is the original curve, A_2 is its mirror image, A_3 corresponds to the opposite curve traversal of A_1 (i.e., selection of starting point) and A_4 corresponds to a combination of A_2 and A_3 . A complete representation of A consists of the representations of all A_1 , A_2 , A_3 and A_4 ; this is denoted as $A = (A_1, A_2, A_3, A_4)$. The same holds for A closed. Notice that, A_2 , A_3 and A_4 need not be computed from raw curve data but they are derived from A_1 : A_2 is derived from A_1 by switching C to V and vice versa; A_3 is derived from A_1 by taking its sequence of segments in reverse order and by switching C and V and, finally, A_4 is obtained by taking the segments of A_1 in reverse order.

We handle the cases with symmetric shapes or different starting points by repeating the algorithm of Figure 3 for each possible case and by taking the least cost match as the cost of matching. For open shapes and local matching, because we cannot know in advance which shape is included within the other one, we repeat the algorithm twice, once for each possibility. Table 1 summarizes the necessary distance computations. \mathcal{D} denotes the overall distance between A and B . \mathcal{D}_k denotes global distance between the open $A_{i,k}$ and the closed shape B ; $A_{i,k}$ denotes the open A_i which is produced starting at the k -th inflection point of the closed A_i .

4.2 Complexity

The run-time complexity of the algorithm depends on the time of computing ψ , the cost of matching two sequences of segments. This is the basic operation of the algorithm. From Equation 12,

	local matching	global matching
A open B open	$\mathcal{D}(A, B) = \min\{\min_{i \in [1,2,3,4]} \{D(A_i, B_1)\}, \min_{j \in [1,2,3,4]} \{D(B_j, A_1)\}\}$	$\mathcal{D}(A, B) = \min_{i \in [1,2,3,4]} \{D(A_i, B_1)\}$
A open B closed	$\mathcal{D}(A, B) = \min_{i \in [1,2,3,4]} \{D(A_i, B_1)\}$	$\mathcal{D}_k(A, B) = \min_{i \in [1,2,3,4]} \{D(A_{i,k}, B_1)\}$
A closed B closed	undefined	$\mathcal{D}(A, B) = \min_{1 \leq k \leq M} \{\mathcal{D}_k(A, B)\}$

Table 1: *Distance computations to achieve invariance with respect to shape transformations.*

the cost computation at each $cell(i, j)$ takes $\frac{ij}{2}$ time (i.e., equals the number of filled cells up to $cell(i_w, j_w)$). Therefore, the time complexity for filling a DP table of size $\mathcal{M} \times \mathcal{N}$ is $\mathcal{O}(M^2N^2)$. This is the time complexity of the algorithm when at least one of the shapes is open. If both shapes are closed, the algorithm is repeated M times (i.e., for all starting points of A) so the time complexity of the algorithm becomes $\mathcal{O}(M^3N^2)$. By restricting merging to K segments (usually $K \ll M, N$), the complexity becomes $\mathcal{O}(K^2MN)$ for open shapes and $\mathcal{O}(K^2M^2N)$ for closed shapes.

4.3 Matching Examples

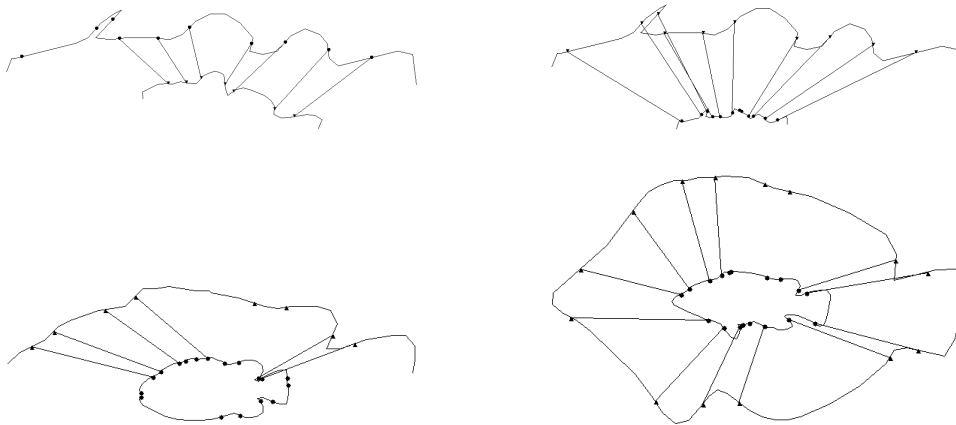


Figure 5: Segment associations reported by the matching algorithm.

Figure 5 illustrates segment correspondences (indicated by consecutive lines connecting the

starting and ending points of the associated segments) obtained by matching fish silhouettes. Inflection points on the two shapes are denoted by dots and triangles respectively. Figure 5 illustrates original polygonal shapes. Inflection points are computed on their B-spline approximation and are back-projected on the original (polygonal) shapes. One of the two shapes has been shrunk, rotated, and translated to better illustrate the associations between matched parts of the two shapes. The top left figure illustrates local matching between open shapes (i.e., part of the bigger shape has been left unmatched). The figure on its right illustrates global matching. The bottom left figure illustrates local matching between an open and a closed curve while figure on its right corresponds to global matching between closed shapes.

5 Shape Retrieval

In our experiments¹ we used the following datasets

CLOSED: It is the dataset of SQUID² and consists of 1,100 closed shapes of marine life species.

OPEN: Consists of 1,500 open shapes which have been generated from the CLOSED dataset by editing (i.e., by deleting manually about half of each shape).

To evaluate the effectiveness of each method we also created 20 query shapes for each data set. In all our experiments with open shapes we focus on the most general case of matching, that is local matching. In our experiments, each measurement is the average over 20 queries. Each query retrieved the 50 most similar shapes.

The experiments are designed to illustrate the superiority of our approach over traditional methods for shape matching and retrieval based on Fourier descriptors [13] and moments [14, 15, 16]. We also establish the superiority of our method over our previously proposed (non-optimal) method [11]. Finally, we compare results obtained by our method with similar results obtained by the method of SQUID for closed shapes. The same queries and the same measurements are used with all methods. Each method computes a distance for each pair of matched shapes (e.g., for a query and a stored shape). The shape database is searched sequentially and the retrieved shapes are ranked by descending similarity with the query. Our method has the additional advantage of re-

¹The algorithm, the results and the datasets are available on the internet at: <http://www.ced.tuc.gr/~petrakis>.

²<http://www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html>.

porting all associations between similar segments choosing (possibly) different scales for different parts of the shapes depending on noise or shape detail.

5.1 Evaluation Method

Two shapes (open or closed) are considered similar if they represent the same figure. In particular, an open shape is considered similar to another open or closed shape if the former is similar (at least) to a part of the later. We used human relevance judgments to compute the effectiveness of each method. The evaluations have been carried-out by four different human reviewers. Each reviewer inspects the answers of a query and, for each answer, judges if it is similar to the query or not. This is a highly subjective process. Two or more methods may retrieve the same answer for the same query, but the same answer might be considered similar by one reviewer and not similar by another. Moreover, the same answer may not be recognized as similar when it is retrieved by different methods. To be fair, the evaluations must be consistent. To achieve consistency, a query and a retrieved shape are taken to be similar if at least one human reviewer considers them similar for any method tested.

To evaluate the effectiveness of retrieval, for each candidate method we computed:

Precision is the percentage of qualifying (similar) shapes retrieved with respect to the total number of retrieved shapes.

Recall is the percentage of qualifying shapes retrieved with respect to the total number of similar shapes in the database. Because we don't have the resources to visually compare each query with every database shape (this would require, for each method, 1,100 or 1,500 human relevance judgments for each one of the 20 queries) for each query, we estimate the total number of similar shapes in the database by merging the similar shapes obtained by all methods tested for this query. This is a valid sampling method known as "*pooling method*" [17]. This method does not allow for absolute judgments such as "method *A* misses 10% of the total qualifying answers in the database". It provides, however, a fair basis for comparisons between methods allowing judgments such as "method *A* returns 5% fewer correct answers than method *B*".

We present a *precision-recall plot* for each method. The horizontal axis in such a plot corresponds to the measured recall while, the vertical axis corresponds to precision. Each method in such a plot is represented by a curve. Each query retrieves the best 50 answers (best matches) and

each point in our plots is the average over 20 queries. Precision and recall values are computed from each answer set after each answer (from 1 to 50) and therefore, each plot contains exactly 50 points. The top-left point of a precision/recall curve corresponds to the precision/recall values for the best answer or best match (which has rank 1) while, the bottom right point corresponds to the precision/recall values for the entire answer set.

A method is better than another if it achieves better precision and better recall. As we shall see in the experiments, it is possible for two precision-recall curves to cross-over. This means that one of the two methods performs better for small answer sets (containing less answers than the number of points up to the cross-section) while, the other performs better for larger answer sets. The method achieving higher precision and recall for large answer sets is considered to be the better method (based on the assumption that typical users retrieve more than 10 or 20 shapes).

5.2 Experimental Results

We carried-out several groups of experiments with open and closed shapes. In all our experiments two or more candidate methods are compared.

5.2.1 Estimation of λ

The goal of this set of experiments is to select an appropriate value of λ . This value controls the amount of merging (the basic feature of our algorithm) and, through merging, effects the quality of matches and, therefore, the accuracy of retrievals. Low values of λ encourage merging; this is desirable for shapes with much shape detail or noise. For smooth shapes, high values of λ are more appropriate.

Figure 6 illustrates the precision-recall diagram for retrievals with local matching on the OPEN dataset and open queries for various values of λ . The appropriate value of λ is the one achieving better precision and better recall than any other value. Based on Figure 6 we select $\lambda = 1$. We also confirmed this value for global matching. This was expected since the value of λ depends on shape properties and not on the type of matching.

5.2.2 Response Time

Fourier descriptors and moments are pre-computed and stored in separate files in the database along with the original contours. Searching such a database typically takes less than 2 seconds

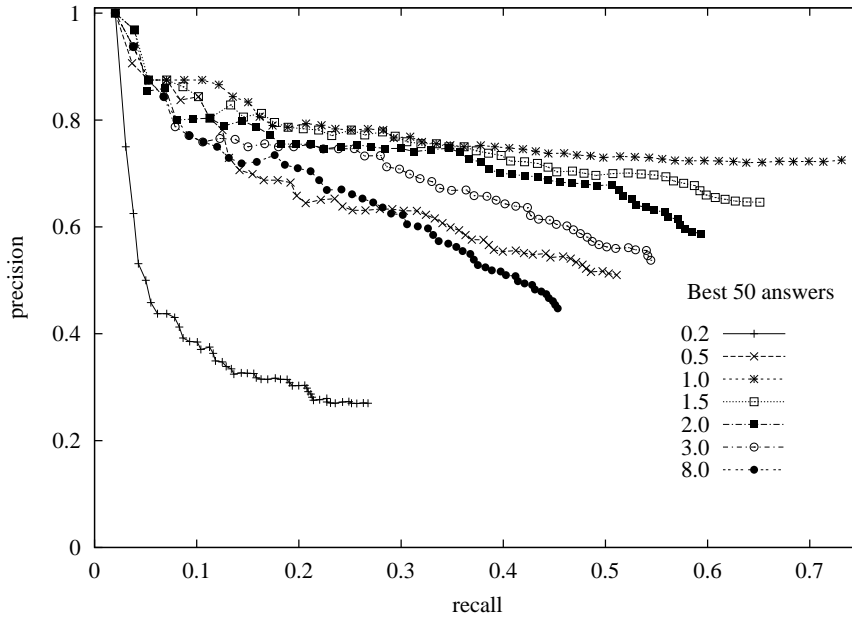


Figure 6: Precision-recall diagram corresponding to local matching and retrievals on the OPEN dataset, for various values of λ .

per query on a Pentium PC 1000MHz. For our method, no pre-computed information is stored. Instead, the actual shape contours are used to search the database. For this reason, our method is the slowest requiring (depending on K) more than 8 minutes to search the CLOSED dataset and more than 2 minutes to search the OPEN dataset. Table 2 illustrates these results. Notice the speed-up as a function of K ($K = \infty$ corresponds to the original method with no restriction on merging). Certain optimizations that could speed up our method are possible, such as the precomputation and storage of the convex and concave segments of all shapes in the database and the non consideration of symmetric shapes.

Maximum merging	$K = 3$	$K = 5$	$K = 9$	$K = \infty$
CLOSED dataset	518	916	1,884	3,202
OPEN dataset	146	159	182	183

Table 2: Average retrieval response times in seconds for the CLOSED and OPEN datasets as a function of the of allowable merging K .

5.2.3 Open Shapes

Figure 7 illustrates the precision-recall diagram for OPEN queries on the OPEN dataset. The competitors to our method for this group of experiments are:

Fourier Descriptors [48]: We take the lower order 20 coefficients (excluding the 0-th coefficient) of the Fourier transform of the of *arclength* versus *turning angle* representation of the curve. The distance between a query and model (database) curve is computed as the Euclidean distance between their vectors of coefficients.

Polygon Moments [16]: The coordinate transform that aligns the two curves is computed. This transform involves computation of cross moments and polygon moments and minimizes a squared error which is taken to be the distance between the two curves.

All methods are invariant to geometric curve transforms (i.e., translation, scale, rotation). They are also taken to be independent on starting point selection and on symmetric transformations using the method described in Section 4.1.

Figure 7 demonstrates that our method achieves at least 10% better precision and better recall for answer sets with the best 50 shapes (rightmost points of the precision/recall curves). Fourier and moments perform about the same. For small answer sets (top left points of the precision/recall curves), all methods perform approximately the same achieving precision close to 1 (i.e., their answers are almost 100% correct). Notice that, our DP local matching method is always more accurate than global matching methods such as Fourier and Moments for any K . Notice finally that, matching with $K \geq 5$ achieves almost the same precision and recall with the original method (without restriction on the size of merging), which means that merges of more than 5 segments are rare.

Figure 8 illustrates an open query (an open curve) and its 20 best retrieved curves. Notice that almost all answers (except 1 marked with “no”) may be considered similar to the query. Answer 10 looks dissimilar to the query. However, a closer look reveals that this shape matches the upper part (the fish tail) of the query.

5.2.4 Closed Shapes

Figure 9 illustrates the precision recall diagram for closed queries on the CLOSED dataset. The competitors to our method are as follows.

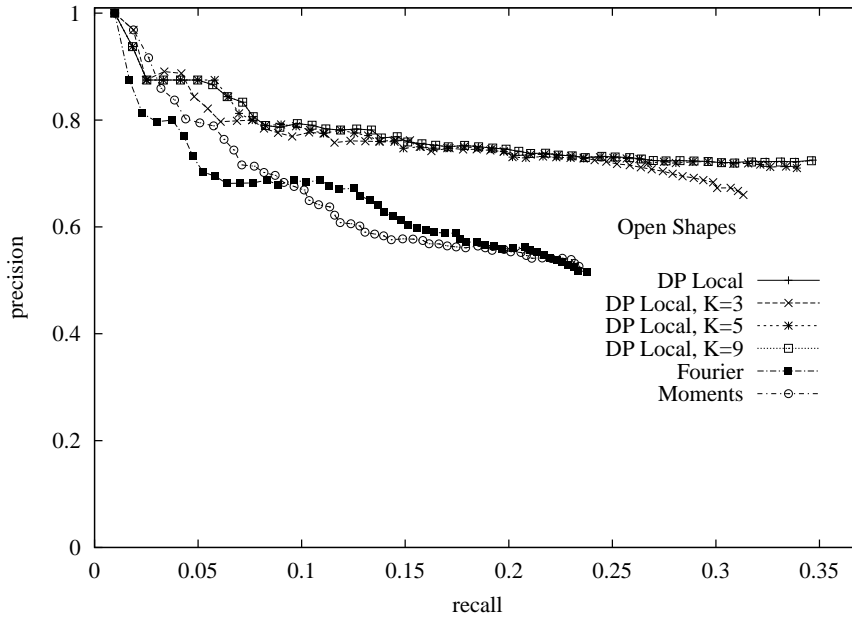


Figure 7: Precision-recall diagram for the OPEN dataset corresponding to the proposed DP local matching method, the same method for $K = 3, 5, 9$, Fourier descriptors, and Moments.

Non-Optimal DP [11]: This is our non-optimal DP method which works only for closed shapes.

Fourier Descriptors [13]: This is known to be one of the most successful methods for the recognition of closed shapes. We computed the first (lower order) 20 coefficients of the Fourier transform.

Sequential (Contour) Moments [15]: This is one of the most effective moment-based methods for closed shapes. For each shape, a representation of 4 moment coefficients is computed from its bounding contour.

Geometric (Area) Moments [14]: Known also as *invariant moments*. This is the original and the most characteristic representative of a wide class of methods based on area moments. A representation of 7 moment coefficients of the shape is computed from the area it occupies. It has been adopted by many shape-based retrieval systems such as [3, 4].

For Fourier and moments, the distance between a query and a database curve is computed as the Euclidean distance between their vectors of descriptors.

Figure 9 illustrates that our proposed method achieves approximately 30% better precision and better recall than any other method for any K . Notice the loss of accuracy for retrievals with $K \leq 5$. On this dataset, merges of more than 5 segments are very common and by restricting

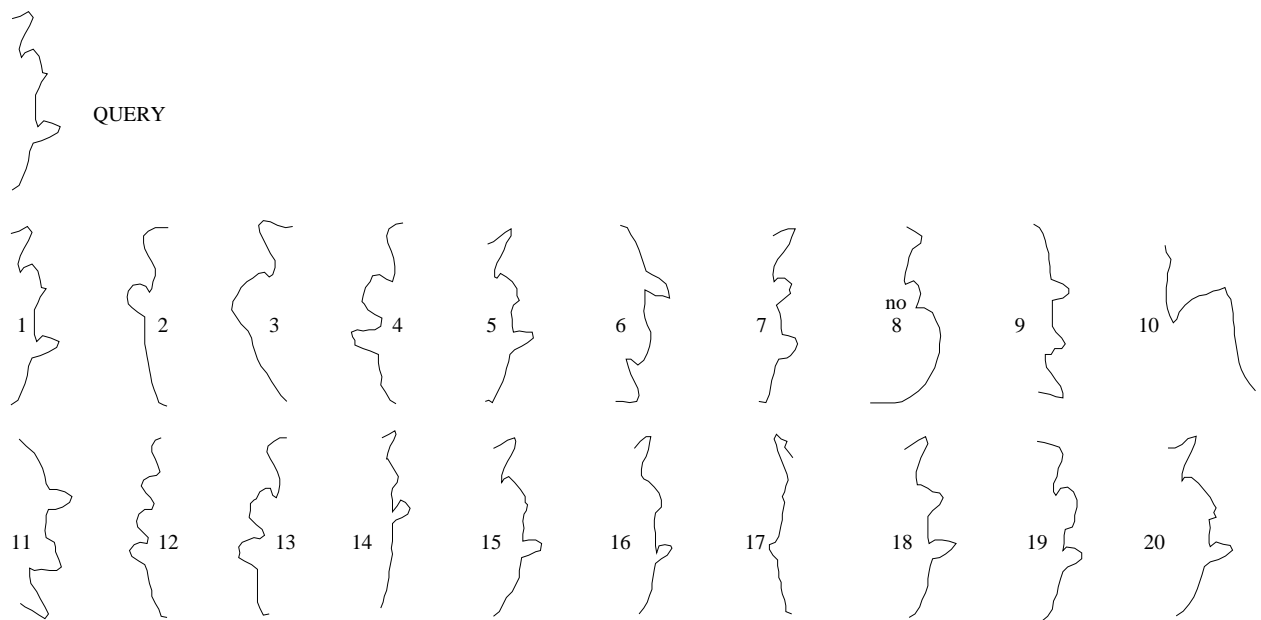


Figure 8: *Example of an open query and its 20 best matches retrieved by the optimal DP method.*

merging to $K = 5$ or $K = 3$ the algorithm becomes less accurate. Our previous non-optimal method still performs much better than Fourier descriptors, Sequential and Geometric moments. This experiment also confirms the results of [11]. Notice finally that, our method is the only method with precision close to 1 for small answer sets (i.e., most of its answers are correct).

Figure 10 illustrates a typical query (top left image) and its 20 best matches retrieved by our method. Notice that almost all shapes (except two marked with “no”) may be considered similar to the query.

The results on the marine datasets indicate that our method performs better than its competitors on shapes with moderate amounts of noise and shape detail (all other methods are more sensitive to noise and detail). The algorithm may not be the same effective on smooth shapes such as the shapes of the GESTURES dataset [11]: Either it will try to match segments one by one or, it will try less plausible merges. Figure 11 illustrates the results obtained on the above GESTURES dataset with 980 smooth hand gestures shapes. On this dataset, $\lambda = 2$ (optimal value). Our method performs approximately the same with our previous non-optimal method and with Fourier (in fact, our method performs slightly better for large answer sets and slightly worst for small answer sets).

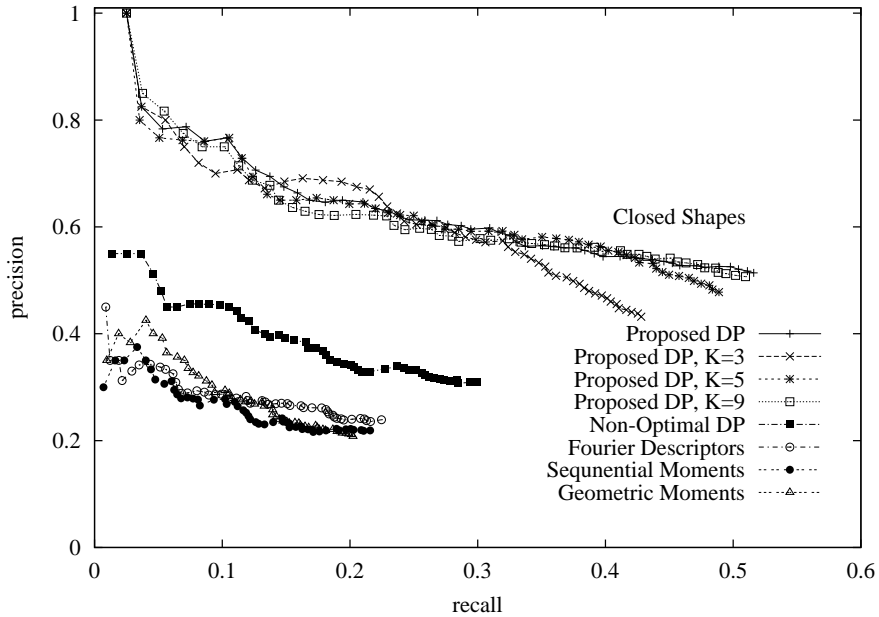


Figure 9: Precision-recall diagram for the CLOSED dataset corresponding to the Proposed DP method, the Proposed DP method for $K = 3, 5, 9$, the Non-optimal DP method, Fourier descriptors, Sequential moments, and Geometric moments.

5.2.5 Comparisons with SQUID

The purpose of this set of experiments is to compare the performance of our method with the method of SQUID [9] which is available on the Internet³. SQUID is a well-established and well-researched approach to shape matching, and it is becoming accepted as a standard for whole shape matching. Notice that SQUID treats only closed shapes. Extending SQUID for open or occluded shapes is non-trivial.

Figure 12 illustrates the precision-recall diagram for the same methods as in the previous experiment including SQUID. For SQUID, we located the same queries on its WWW interface, we applied these queries and we downloaded their results. We managed to locate all but 2 of the 20 queries of the previous experiment in Section 5.2.4. Therefore, our results are averages over 18 queries. SQUID interface supports only 18 answers. Therefore, each curve of Figure 12 contains only 18 points instead of 50.

Figure 12 demonstrates that our method performs better than SQUID for large answer sets containing more than 4 answers, achieving up to 10% better precision and better recall. Notice that, in databases, users typically retrieve more than 5-10 answers. For small answer sets both

³<http://www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html>.

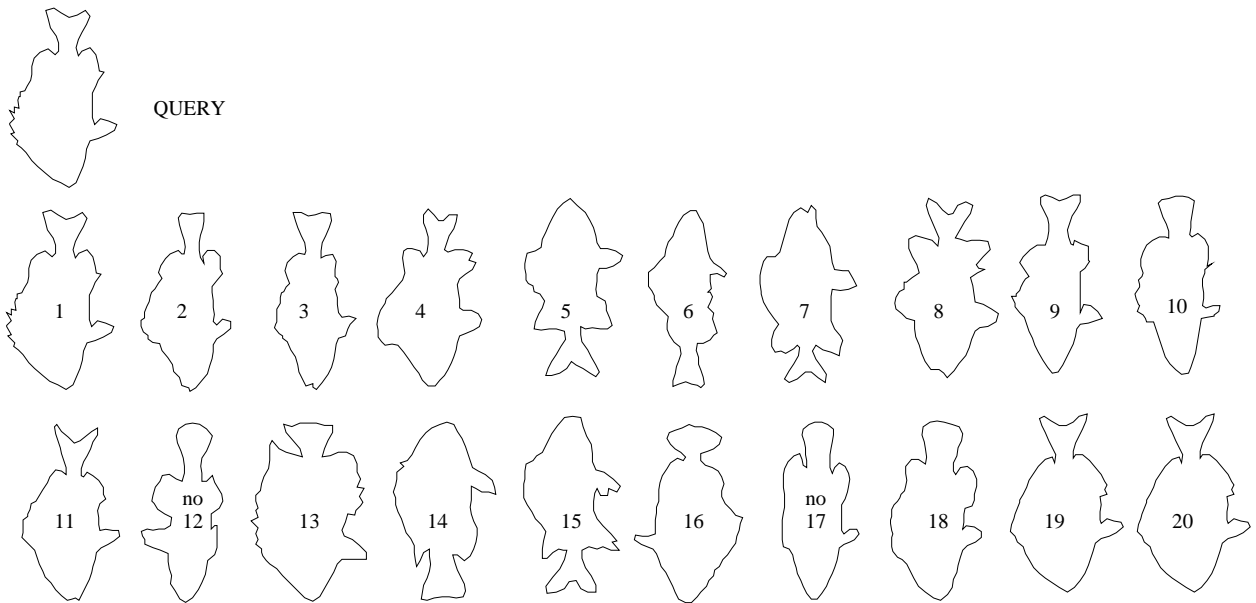


Figure 10: *Example of a closed query and its 20 best matches retrieved by the optimal DP method.*

methods achieve precision close to 1, that is, all their answers are correct.

Figure 13 illustrates the results (18 answers) obtained by SQUID by applying the same query of Figure 10. Again, all shapes (except 2 marked with “no”) may be considered similar to the query. Notice that, many shapes are common to the answers sets obtained by our method and SQUID.

6 Conclusions

We propose an approach for shape matching and shape similarity retrieval based on dynamic programming. Our approach treats open, noisy or distorted shapes and is independent of translation, scale, rotation and starting point selection. It operates implicitly at multiple scales by allowing the matching of merged sequences of consecutive segments in the shapes which are matched. This way our method maintains the advantages of previous methods (e.g., [10, 35]) utilizing smoothed versions of the shapes at various levels of detail, while avoiding the expensive computation of explicit scale-space representations.

We carried out extensive performance experiments on several datasets and our evaluations are based on human relevance judgments by 4 independent referees. The experiments indicate that our approach is well suited to shape matching and retrieval on shapes with moderate amounts of noise and distortion, achieving higher precision and recall than traditional shape matching and retrieval methods based on Fourier descriptors and moments. Our method performs better than our previous

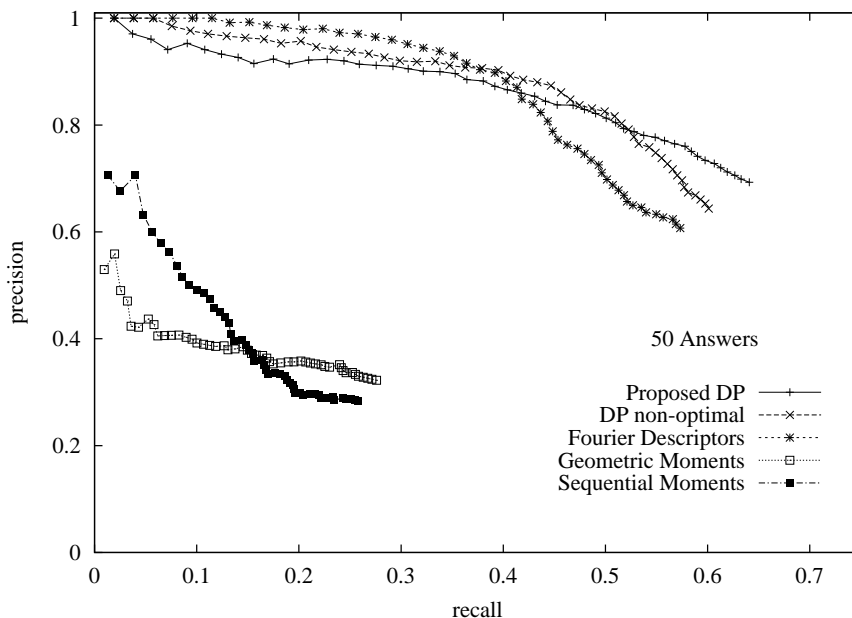


Figure 11: Precision-recall diagram for the *GESTURES* dataset corresponding to the Proposed DP method, the Non-optimal DP method, Fourier descriptors, Sequential moments, and Geometric moments.

non-optimal method and SQUID, while being the only method capable of handling both open and closed shapes at the same time.

Future work includes the extension of our method to handle gaps in shapes and partial matches as in [7] (where parts of one shape match possible many parts on the other shape). Future work also includes the experimentation with more datasets and methods, handling of combined queries involving more than one feature (e.g., shape, color, text), the development of indexing methods that could speed up retrievals and the development of a graphical user interface on the World Wide Web.

Acknowledgments

We are grateful to Z. Rao for valuable contributions to this work. Z. Rao also developed a first version the shape matching algorithm. We are also grateful to P. Kaklamanis and C. Genzis for their help in the experiments, R. Jeske who implemented the programs for creating the accuracy plots, E. Voutsakis who developed an interactive environment for the evaluation of visual queries (available from <http://www.ced.tuc.gr/~petrakis>), Prof. G. Bebis of the Department Computer Science at the University of Nevada for providing us the codes of the Fourier and moment methods

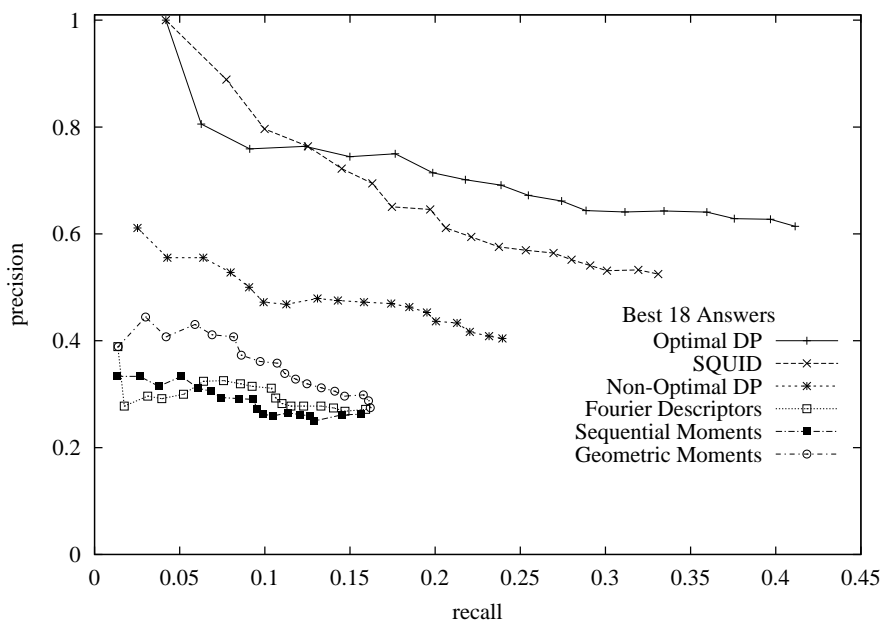


Figure 12: Precision-recall diagram for the CLOSED dataset corresponding to the Proposed DP method, SQUID, the Non-optimal DP method Fourier descriptors, Sequential moments, and Geometric moments.

for closed curves and to Prof. F. Mokhtarian of the Centre for Vision, Speech and Signal Processing laboratory at the University of Surrey, UK, for providing us the marine dataset.

This work was supported by project HIPER (BE97-5084) under programme BRIGHT-EURAM of the European Union (EU) and by a grant from the Natural Sciences and Engineering Research Council of Canada.

References

- [1] S. Loncaric. A Survey of Shape Analysis Techniques. *Pattern Recognition*, 31(8):983–1001, 1998.
- [2] P. Suetens, P. Fua, and A. J. Hanson. Computational Strategies for Object Recognition. *ACM Computing Surveys*, 24(1):5–61, March 1992.
- [3] A. K. Jain and A. Vailaya. Shape-Based Retrieval: A Case Study With Trademark Image Databases. *Pattern Recognition*, 31(9):1369–13990, 1998.
- [4] B. M. Mehtre, M. S. Kankanhalli, and W. F. Lee. Content-Based Image Retrieval using a Composite Color-Shape Approach. *Information Processing and Management*, 34(1):109–120, 1998.

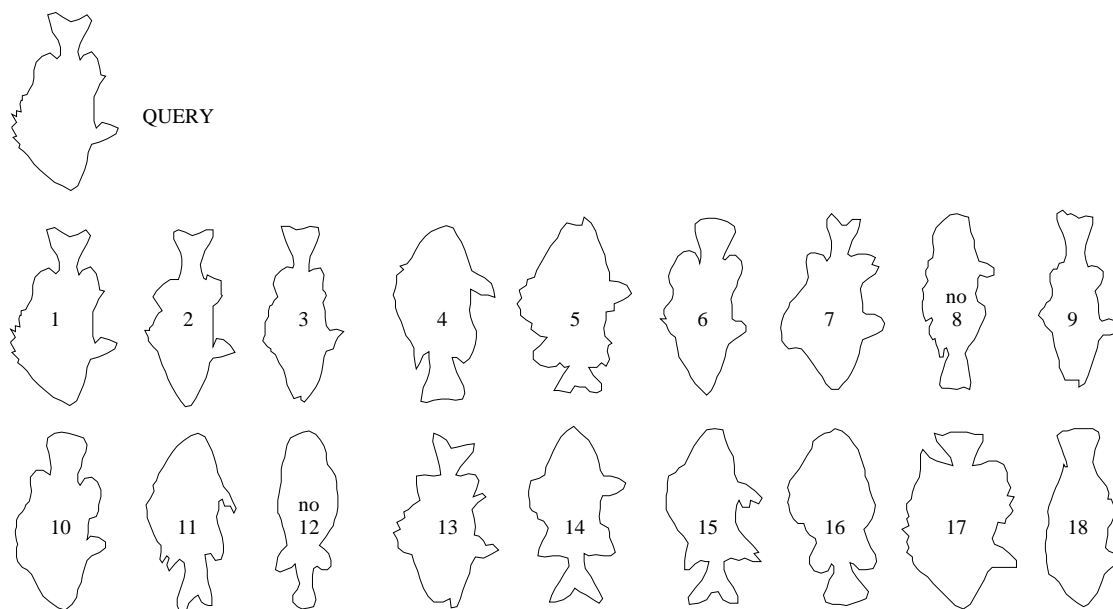


Figure 13: Example of a closed query and its 18 best matches retrieved by SQUID.

- [5] T. Gevers and A. W. M. Smeulders. PicToSeek: Combining Color and Shape Invariant Features for Image Retrieval. *IEEE Trans. on Image Processing*, 9(1):102–119, 2000. (http://zomax.wins.uva.nl:5345/ret_user/index.html).
- [6] M. Flickner et. al. Query By Image and Video Content: The QBIC System. *IEEE Computer*, 28(9):23–32, September 1995. (<http://www.qbic.almaden.ibm.com>).
- [7] Y. Gdalyahu and D. Weinshall. Flexible Syntactic Matching of Curves and its Application to Automatic Hierarchical Classification of Silhouettes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(12):1312–1328, 1999.
- [8] F. Mokhtarian and A. Mackworth. Scale-Based Description of Planar Curves and Two-Dimensional Shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(1):34–43, 1986.
- [9] F. Mokhtarian, S. Abbasi, and J. Kittler. Efficient and Robust Retrieval by Shape Content through Curvature Scale Space. In *Proc. of Intern. Workshop on Image DataBases and MultiMedia Search*, pages 35–42, Amsterdam, The Netherlands, 1996. (<http://www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html>).
- [10] N. Ueda and S. Suzuki. Learning Visual Models from Shape Contours Using Multiscale Convex/Concave Structure Matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(4):337–352, April 1993.
- [11] E. Milios and E. G.M. Petrakis. Shape Retrieval Based on Dynamic Programming. *IEEE Trans. on Image Processing*, 9(1):141–147, 2000.
- [12] D. Geiger, A. Gupta, L.A. Costa, and J. Vlontzos. Dynamic Programming for Detecting, Tracking and Matching Deformable contours. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(3):294–302, 1995.

- [13] T. P. Wallace and P. A. Wintz. An Efficient Three-Dimensional Aircraft Recognition Algorithm Using Normalized Fourier Descriptors. *Computer Graphics and Image Processing*, 13:99–126, 1980.
- [14] M.-K. Hu. Visual Pattern Recogn. by Moment Invariants. *IRE Trans. on Info. Theory*, IT-8:179–187, 1962.
- [15] L. Gurta and M. D. Srinath. Contour Sequence Moments for the Classification of Closed Planar Shapes. *Pattern Recognition*, 20(3):267–271, 1987.
- [16] M. W. Koch and R. L. Kashyap. Matching Polygon Fragments. *Pattern Recognition Letters*, 10(5):297–308, November 1989.
- [17] E.M. Voorhees and D.K. Harmann. Overview of the Seventh Text REtrieval Conference (TREC-7). In *NIST Special Publication 500-242: The Seventh Text REtrieval Conference*, pages 1–23, 1998. (http://trec.nist.gov/pubs/trec7/t7_proceedings.html).
- [18] M.-H. Han and D. Jang. The Use of Maximum Curvature Points for the Recognition of Partially Occluded Objects. *Pattern Recognition*, 23(1/2):21–33, 1990.
- [19] K. Siddiqi, A. Shokoufandeh, S. Dickinson, and S. Zucker. Shock Graphs and Shape Matching. *Intern. Journal of Computer Vision*, 35(1):13–32, 1999.
- [20] P. G. Gottschalk, J. L. Turney, and T. N. Mudge. Efficient Recognition of Partially Visible Objects Using a Logarithmic Complexity Matching Technique. *The Intern. Journal of Robotics Research*, 8(6):110–131, December 1989.
- [21] W.-H. Tsai and S.-S. Yu. Attributed String Matching with Merging for Shape Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7(4):453–462, 1985.
- [22] B. Bhanu and O. D. Faugeras. Shape Matching of Two-Dimensional Objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(2):137–156, 1984.
- [23] R. J. Prokop and A. P. Reeves. A Survey of Moment-Based Techniques for Unoccluded Object Representation and Recognition. *CVGIP: Graphical Models and Image Processing*, 54(5):438–460, September 1992.
- [24] J. L. Turney, T. Mudge, and R. A. Volz. Recognition of Partially Occluded Objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7(4):410–421, July 1985.
- [25] G. Papadourakis G. Bebis and S. Orphanoudakis. Curvature Scale Space Driven Object Recognition with an Indexing Scheme based on Artificial Neural Networks. *Pattern Recognition*, 32(7):1175–1201, 1999.
- [26] A. Witkin, D. Terzopoulos, and M. Kass. Signal Matching Through Scale Space. *Intern. Journal of Computer Vision*, pages 133–144, 1987.
- [27] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. *Intern. Journal of Computer Vision*, 1(4):321–331, 1988.
- [28] B. Vemuri and R. Malladi. Constructing Intrinsic Parameters with Active Models for Invariant Surface Reconstruction. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(7), July 1993.

- [29] F. Leymarie and M. Levine. Tracking Deformable Objects in the Plane using an Active Contour Model. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(6), 1993.
- [30] L. Floreby. A Multiscale Algorithm for Closed Contour Matching in Image Sequence. In *IEEE Intern. Conf. on Pattern Recognition*, pages 884–888, 1996.
- [31] H. Tagare. Deformable 2D Template Matching using Orthogonal Curves. *IEEE Trans. on Medical Imaging*, 16(1):108–117, February 1997.
- [32] A. Del Bimbo and P. Pala. Visual Image Retrieval by Elastic Matching of User Sketches. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(2):121–132, February 1997.
- [33] L. Shapiro. A Structural Model of Shape. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2, March 1980.
- [34] D. Wuescher and K. Boyer. Robust Contour Decomposition using a Constant Curvature Criterion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(1):41–51, 1991.
- [35] F. Mokhtarian. Silhouette-Based Object Recognition through Curvature Scale Space. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(5):539–544, May 1995.
- [36] J. Gorman, R. Mitchell, and F. Kuhl. Partial Shape Recognition using Dynamic Programming. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 10(2):257–266, March 1988.
- [37] E. Milios. Shape Matching using Curvature Processes. *CVGIP: Graphical Models and Image Processing*, 47:203–226, 1989.
- [38] R. Mehrotra and W. Grosky. Shape Matching Utilizing Indexed Hypotheses Generation and Testing. *IEEE Trans. on Robotics and Automation*, 5(1):70–77, February 1989.
- [39] N. Ansari and E. Delp. Partial Shape Recognition: A Landmark-Based Approach. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(5):470–483, May 1990.
- [40] M. Leyton. A Process Grammar for Shape. *Artificial Intelligence*, 34:213–247, 1988.
- [41] A. Witkin. Scale Space Filtering. In *Proc. of IJCAI*, pages 1019–1022, Karlsruhe, 1983.
- [42] E. Saund. Symbolic Construction of a 2D Scale-Space Image. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 817–830, August 1990.
- [43] K. Siddiqi and B. Kimia. Parts of Visual Form: Computational Aspects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(3):239–251, 1995.
- [44] J. Baid and E. Milios. Deformed Shape Matching Using Multiscale Dynamic Programming. In *Vision Interface*, Toronto, 1996.
- [45] Z. Rao. Fast Retrieval Algorithms for Shape Databases. Master’s thesis, Department of Computer Science, York University, 1999.
- [46] E. Milios and E. G.M. Petrakis. Efficient Shape Matching and Retrieval at Multiple Scales. TR CS-1998-11, Department of Computer Science, York University, Toronto, 1998. (<http://www.cs.yorku.ca/techreports/1998/CS-1998-11.html>).

- [47] A.R. Smith. Spline Tutorial Notes. Tech memo 77, Computer Division, Lucasfilm, May 1983. Also tutorial notes at SIGGRAPHs 83 and 84, (<http://www.alvyray.com/Memos/MemosPixar.htm#SplineTutorial>).
- [48] A. Kalvin, E. Schonberg, J. T. Schwartz, and M. Sharir. Two-Dimensional, Model-Based, Boundary Matching Using Footprints. *The Intern. Journal of Robotics Research*, 5(4):38–55, 1986.

Contents

1	Introduction	2
2	Related Work	5
3	Proposed Methodology	8
3.1	Shape Representation	8
3.2	Matching Cases	9
3.3	Dynamic Programming (DP) Table	10
3.4	Distance Function	12
3.5	Geometric Quantities	12
3.6	Scale Factor	13
3.7	Dissimilarity Cost	14
3.8	Merging Cost	15
4	Algorithm	15
4.1	Invariance to Shape Transformations	18
4.2	Complexity	18
4.3	Matching Examples	19
5	Shape Retrieval	20
5.1	Evaluation Method	21
5.2	Experimental Results	22
5.2.1	Estimation of λ	22
5.2.2	Response Time	22
5.2.3	Open Shapes	24
5.2.4	Closed Shapes	24
5.2.5	Comparisons with SQUID	27
6	Conclusions	28

Biographies

Euripides Petrakis was born in Chania, Crete in 1962. He received a BSc in physics from the National University of Athens, Greece in 1985 and the PhD degree in computer science from the University of Crete, Greece in 1993. He is assistant professor of Computer Science at the Department of Electronic and Computer Engineering of the Technical University of Crete (TUC) since 1998. His research interests include image and video databases, access methods for spatial and geographic data, medical image databases and computer vision. His current research activity focuses on searching for images and video by content on the internet. He is a member of the IEEE.

Aristeidis Diplaros was born in Salamina, Greece in 1976. He received a diploma from the Dept. of Electronic and Computer Engineering, Chania - Crete in 2001. He currently a PhD candidate in the Intelligent Sensory Information Systems of the University of Amsterdam. His research interests include image databases, computer vision, numerical analysis, databases and their applications to the retrieval of multimedia documents by content.

Evangelos Milios received a diploma in Electrical Engineering from the National Technical University of Athens, Greece, and Master's and Ph.D. degrees in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology. He has been a research assistant professor of Computer Science, University of Toronto, and associate professor of Computer Science at York University. Since 1998 he has been with the Faculty of Computer Science, Dalhousie University, where he is currently professor and graduate director. He is a Senior Member of the IEEE. He served as a member of the ACM Dissertation Award committee (1990-1992). He is on the organizing committee of the ACM/SIGART Doctoral Consortium. He has published on acoustic signal interpretation, shape matching, and on the processing, interpretation and use of visual and range signals for landmark-based navigation and map construction in single- and multi-agent robotics. His current research activity is centered on software agents for Web information retrieval.