International Conference on Computing, Networking and Communications, Multimedia Computing and Communications Symposium

Robust shape matching using global feature space representation of contours

Shehzad Khalid
Department of Computer and Software Engineering
Bahria University, Islamabad, Pakistan
Email: shehzad_khalid@hotmail.com

***Abstract*—The fundamental ingredient of content-based image retrieval is the selection of appropriate features to describe the content of the image. Shape of an object, represented by its contour, is the most important visual feature that is thought to be used by humans to determine the similarity of objects. In this paper, we present an effective representation of shape, using its boundary information, that is robust to arbitrary distortions and affine transformation. The contour representation of shape is converted into time series and is modeled using orthogonal basis function representations. Encoding contour representation of shapes in this manner leads to efficiency gains over existing approaches such as structural shape representation and techniques that use discrete point-based flow vectors to represent the contour. Experimental evaluation demonstrates that the proposed shape representation and matching mechanism is effective, efficient and robust to different arbitrary and affine distortions.**

I. Introduction

In recent years, there has been a growth of research activity aimed at the development of sophisticated content-based searching and retrieval of digital images. This development is now especially timely given the rapid increase in the number of systems that can capture and store information as digital objects resulting in generation of more and more digital images. A fundamental feature that determines the content-based similarity of images is the similarity of object's shape in the image.

Shape matching generally looks for effective and perceptually important shape representation and distance measures that are invariant to many distortions including noise, rotation, articulation, scale, jag etc. Rotation invariance is relatively difficult to handle as compared to noise, translation and other distortions. Much of the earlier research on shape matching achieve rotation invariance at the cost of accuracy [1] or efficiency [2],[3],[4], [5]. Our ultimate goal is to generate efficient and accurate rotation-invariant features to compare shapes that are represented using closed planar contours.

In this paper, we apply time series modeling of contour-based shape representation to the problem of shape matching. Contours are modeled using function approximation technique and matching of shape is carried out in the coefficient subspace. An efficient matching technique is proposed that can match shapes which are not rotationaly aligned. The proposed technique is also invariant to noise and different affine transformations. The remainder of the paper is organized as follow. We review some relevant background material in section 2. In section 3 we present our function approximation approach to contour-based shape representation. Section 4 addresses the issue of rotation invariant shape matching without compromising on efficiency and accuracy. In section 5, experiments have been reported to show the effectiveness of proposed shape matching approach as compared to competitors. The last section summarizes the paper.

II. Background and related work

Shape descriptors are known to be useful candidates for content-based image indexing and retrieval schemes. Previous work has sought to represent shapes through shape context, shape signature, integral invariants, curvature, moments etc. Broadly speaking, these shape representation and matching techniques are classified into two classes: 1) contour-based that only exploit contour information and 2) region-based that incorporate all the pixels within the shape to generate shape descriptor.

The contour-based approaches are more common in literature as studies on human perception have shown that humans can recognize and discriminate shapes mainly by their contour features. We therefore consider only contour-based approaches here. Most of the contour-based shape representation generates a global representation of contour. Some of the global representations include shape context [6], shape signature [7], integral invariants [8] and differential invariants [9]. Some approaches [5][10]segments the contour into pieces and generate a piecewise representation of shapes. Piecewise approaches differ in the segmentation criteria to break contours into pieces and the modeling mechanism used to represent contour segments. Some of the piecewise approaches include polygon decomposition [5], smooth curve decomposition [10] and curvature decomposition [11]. The advantage of piecewise approaches is its ability to support partial matching and as a result dealing with the problem of partial occlusion. However, this merit of piecewise approaches comes with the disadvantage of complex and inefficient matching. Piecewise approaches do not capture global features of shape which is extremely important for shape recognition and discrimination.

Rotation invariance is critical for accurate shape matching and is hard to achieve as compared with invariance to other distortions [2][12]. There exists a variety of techniques that has been used for rotation invariant shape matching. Some approaches [13][1] make use of rotation invariant features

978-1-4673-0009-4/12/$26.00 ©2012 IEEE

including features associated to curvature and centroid distances, ratio of perimeter to area, circularity, convexity etc. These approaches achieve rotation invariance by compromising on the accuracy as selection of only rotation invariant features may result in discarding other features that may be sensitive to rotation but is important for fine discrimination between different shapes. Approaches using 1D time series representation of shapes has also been proposed [2][14][15]. Some of these approaches [16] achieve rotation invariance by selecting very few starting point to obtain 1D time series representation of 2D shape. Normally shapes are aligned with respect to the major axis. However, such alignment are very unreliable specially when there is no well defined major-axis and slight articulation in shape may have significant impact on rotation alignment. Some researchers proposed to use brute-force search over all possible rotation to identify the true alignment of shapes [2][3][4][14][17]. We need to shift one contour n times ($n \gg 100$) where *n* is the number of points on the contour. The matching of one shape is carried out with *n* different alignments of the other shape which jeopardises the efficiency requirement of content-based image search and retrieval systems.

The contribution of this paper is to show that a low dimensional coefficient-based contour encoding scheme can be used more efficiently for image searching than previous approaches that rely on high dimensional representation of shapes. The parameter subspace representation of contours is also robust to the presence of different levels of noise and other distortions. An efficient mechanism to achieve rotation invariance in shape matching without compromising on efficiency and accuracy is proposed.

III. Low Dimensional Shape representation

In this section, we present our global contour-based shape representation scheme based on time series representation. Formally, the object contour *C(O)* is defined by the point sequence:

$$C(O) = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\} \tag{1}$$

where (x_1, y_1) represents the location of the left-top point on the contour and *n* is the number of points on the contour.

The 2-D raw point-based feature vector *C* is then converted into 1-D centroid distance *CD* based time series by mapping each point on the contour to the distance between the point and the shape's centroid as:

$$CD_t = \{\sqrt{((x_t - x_c)^2 + (y_t - y_c)^2)}\} \qquad t = 1, 2, 3, ..., n \tag{2}$$

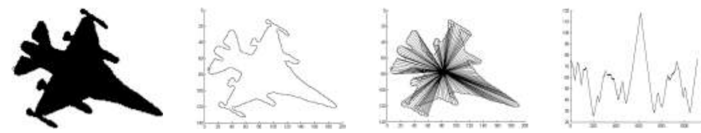
where

$$x_c = \frac{1}{n}\sum_{t=1}^{n} x_t\,,\ \ y_c = \frac{1}{n}\sum_{t=1}^{n} y_t \tag{3}$$

The scale variance in shape representation is taken care of by normalizing the distance vector *CD* with respect to the standard deviation σ of the centroid distances as:

$$CD_t = \frac{CD_t}{\sigma} \qquad t = 1, 2, 3, ..., n \tag{4}$$

The process of conversion of shape into 1-D time series representation is depicted graphically in Fig. 1.



(a) (b) (c) (d)

Fig. 1. Extracting 1D time-series representation of shapes. (a) Projection of shape in image plane. (b) Projection in *C* space with 'o' marker specifying the starting point (c) Mapping of contour from 2D *C*-space to 1D *CD*-space (d) Projection in *CD*-space.

For realistic scenarios, $n >> 100$ and this renders direct manipulation of point sequence impractical for retrieval purposes. The key to implementing efficient trajectory matching is dimensionality reduction. The idea is to determine a feature extraction function *F* that reduces the dimensionality of the data from n to m such that $m \ll n$. Similarity search and retrieval is then conducted in the reduced feature space. We model the projection of shape in CD based time series representation using DFT based coefficient feature space representation. The most salient features of the time series are captured by the low frequency DFT coefficients, i.e. the first few terms of the DFT series. The *n*-point DFT of $\{CD_i\}$, defined as a sequence $\{\overline{CD_f}\}$ of *n* complex numbers $(f = 1, ..., n)$, is given as:

$$\overline{CD_f} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} CD_i exp(-j2\pi fi/n) \quad f = 1, 2, ..., n \tag{5}$$

where j is the imaginary unit $j = \sqrt{-1}$, and $\overline{CD_f}$ are complex numbers with the exception of $\overline{CD_0}$ which is real. Typically, the DFT sequence is truncated after *m* terms, $f = 0, ..., m-1$. In this case, the feature vector consists of $2m-1$ entries (from real and imaginary parts). More formally, let a_i and $\widehat{a_i}$ be the real and imaginary part of $\overline{CD}$. Shapes can be represented in the coefficient feature space by a $2m-1$ dimensional vector of DFT coefficients $\mathbf{F}_{DFT}$, where

$$\mathbf{F}_{DFT} = [a_0, a_1, \widehat{a}_1, ..., a_{m-1}, \widehat{a}_{m-1}] \tag{6}$$

IV. Rotation Invariant Shape Representation

The mechanism specified in section 3 generates a contour distance (CD) based time series representation of shapes. It then applies the distance measure in low dimensional feature space representation of contours. This method produces good results if CD-based time series representation of two shapes are rotation aligned. However, this method can produce extremely poor results if the two shapes are not rotation aligned. We propose a Critical-point based approach for Rotational Invariant Shape Matching (CRISM). Instead of calculating the distance of fixed shape with all possible rotations of other shape, we calculate the distance with only the critical rotations of other shape. A set of critical rotations is generated by using a limited number of critical points on contour which are extracted by identifying local maximas in *CD*-space. We have employed k-beam search to identify local maximas.

The algorithm for selection of critical points using *CD*-space representation of contours is specified as follows:

1) Select k indexes $C = c_1, c_2, ..., c_k$ at equal distances on the contour in *CD*-space.
2) Update indexes of critical points C as:

$$c_i = \begin{cases} c_i + \alpha & iff \;\; CD_{(c_i+\alpha)} > CD_{(c_i)} \\ c_i - \alpha & iff \;\; CD_{(c_i-\alpha)} > CD_{(c_i)} \\ c_i & otherwise \end{cases} \quad \forall c_i \in C \tag{7}$$

where α is a constant which is used to skip local maximas caused by the presence of local noise. The value of α is determined empirically at $\alpha = 4$.
3) Iterate through step (2) till no index is updated for a given iteration.
4) Identify the closest pair of critical points whose difference between their corresponding indexes is less than a threshold β. Select the critical point whose corresponding CD value is less than the other member and remove it from C. The value of β is determine empirically at $\beta = A_{desired}/600$.
5) Iterate through step (4) till the difference between the indices of closest pair of remaining critical points is greater than β.
The process of identifying critical points on the contour is highlighted in Fig. 2.

(a) (b)

Fig. 2. Identification of critical points using local-maxima heuristic. Critical points are highlighted using '△' marker (a) in 1D *CD*-space representation (b) in 2D contour-space representation

The contour of the fixed shape is extracted by starting from the left-top point on the contour, as specified in eq. (1). The centroid distance of left-top point on a contour represents one of the local maxima in the CD space. The fixed shape therefore has a default alignment w.r.t. one of the local maxima. The distance between the fixed shape is then calculated with a set of only critical rotations of the other shape. The critical rotation that gives the minimum distance with the fixed shape will result into correct alignment of two shapes and will return the rotation invariant distance. Comparison of proposed critical-point alignment with computationally expensive brute-force alignment is presented in Fig. 3. For different pair of shapes, the alignment obtained using brute-force alignment is no better than the alignment obtained using critical-point approach. Brute force alignment requires shifting one of the curves *n* times where *n* is the number of points on the curve. On the other hand, critical-point alignment requires shifting the curve *nc* times where *nc* is the number of critical points on the curve. As $nc \ll n$, we have managed to achieve efficient rotation invariance without compromising the accuracy of rotation invariance shape matching.

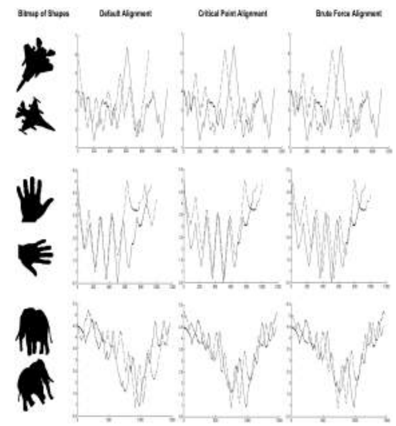


Fig. 3. Comparison of rotational invariant shape alignment using critical-point and brute-force alignment.

More formally, suppose we have two time series *A* and *B* representing two shapes in CD-space as:

$$\begin{aligned} A &= a_1, a_2, a_3, \ldots, a_p \\ B &= b_1, b_2, b_3, \ldots, b_q \end{aligned} \tag{8}$$

where *p* and *q* are length of time series *A* and *B* respectively. Assuming that *A* represents a fixed shape, we identify a set of nc critical points $C = \{c_1, c_2, .., c_{nc}\}$ using time series *B*. We achieve rotation invariance by expanding *B* into a matrix **B** of nc time series as:

$$\mathbf{B} = \begin{bmatrix} b_{c_1}, \cdots, b_{q-1}, b_q, b_1, \cdots, b_{c_1-1} \\ b_{c_2}, \cdots, b_{q-1}, b_q, b_1, \cdots, b_{c_2-1} \\ \vdots \\ b_{c_{nc}}, \cdots, b_{q-1}, b_q, b_1, \cdots, b_{c_{nc}-1} \end{bmatrix} \tag{9}$$

Each row in matrix **B** is a time series representing contour in *CD*-space aligned w.r.t. one of the critical point. To make our distance measure invariant to mirror images, we pad our matrix B with the reverse of all the time series as:

$$\mathbf{B} = \begin{bmatrix} b_{c_1}, \cdots, b_{q-1}, b_q, b_1, \cdots, b_{c_1-1} \\ b_{c_2}, \cdots, b_{q-1}, b_q, b_1, \cdots, b_{c_2-1} \\ \vdots \\ b_{c_{nc}}, \cdots, b_{q-1}, b_q, b_1, \cdots, b_{c_{nc}-1} \\ b_{c_1-1}, \cdots, b_1, b_q, b_{q-1}, \cdots, b_{c_1} \\ b_{c_2-1}, \cdots, b_1, b_q, b_{q-1}, \cdots, b_{c_2} \\ \vdots \\ b_{c_{nc}-1}, \cdots, b_1, b_q, b_{q-1}, \cdots, b_{c_{nc}} \end{bmatrix} \tag{10}$$

The proposed framework for rotational invariant shape matching also supports partial rotational invariance where we only want to allow limited rotation in shape matching. This can be achieved by rotating one of the shape along limited number of critical points. Low dimensional coefficient space representations of time series A and each time series in $\mathbf{B}$ are generated by applying one of the dimensionality reduction technique as specified in section 3.1. Let B_{c_i} represents the time series that is aligned w.r.t. critical point c_i and $F(B_{c_i})$ is the coefficient feature space representation of B_{c_i} using dimensionality reduction function $F(.)$, the rotation invariant distance ($RID(.,.)$) can then be specified as:

$$RID(A, \mathbf{B}) = \min_{1 \leq i \leq nc} (DIST(F(A), F(B_{c_i}))) \quad (11)$$

where $DIST(.,.)$ is the Euclidean distance function. The time complexity of querying a shape database using CRISM algorithm is $O(2 * nc * m * N)$ where N is the number of samples in a shape dataset. However, the value of nc rarely exceeds 15 even for complex shapes and the value of m is on the order of 8 to 32.

V. Experimental Results

This section analyzes the performance of proposed approach for rotational invariant shape matching in the presence of noise and other shape distortions. Experiments are conducted on noisy shapes from a silhouette dataset. We have randomly selected six samples from each of the shape categories from the dataset.

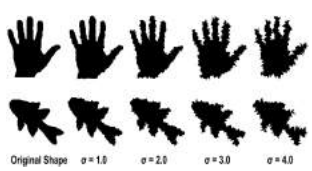
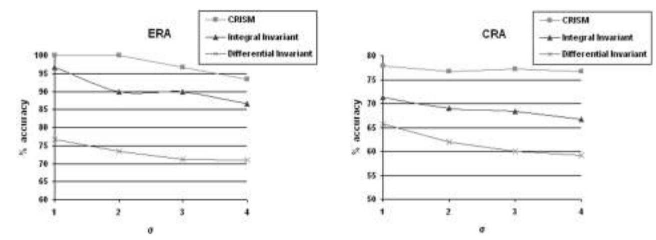


Fig. 4. Effect of simulated noise using increasing value of σ on sample shapes.

The purpose of the experiment is to compare the performance of proposed CRISM-based approach for shape matching with competitive techniques. The evaluation metrics used for the comparison of various techniques are Exact Retrieval Accuracy (ERA) and Class Retrieval Accuracy (CRA). In the context of the current experimental evaluation, ERA can be defined as the ratio of the number of 1-NN queries that retrieve the desired result to the total number of queries. CRA can be defined as the ratio of the number of correct closest matches in 6-NN queries that retrieve the desired result to the total number of retrievals.

The silhouette dataset provides ground truth (i.e. manually labelled) shapes and so we try to simulate the effects of noise. A dataset is corrupted by moving all points on the contour in the normal direction by a certain distance d which determines the amount of noise that is induced in the shape. The value of d is generated from a zero-mean gaussian distribution and standard deviation of σ. If $\mathbf{S}$ represents the original data, a noise corrupted dataset $\mathbf{S_C}$ is produced by adding the term $N[0, \sigma]$ to each (x, y) coordinate on the contour in the normal direction. We set $\sigma = \{1.0, 2.0, 3.0, 4.0\}$ to simulate different noise levels. Simulation of different level of noise on two shapes is presented in Fig. 4. Each corrupted trajectory in $\mathbf{S_C}$ is then selected as an example query Q_C and we search for a set of k nearest matches in the original dataset $\mathbf{S}$. This is defined as:

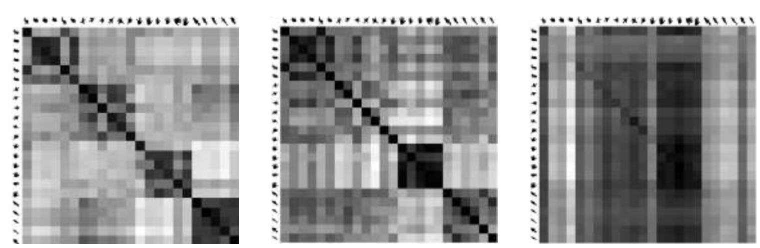
$$k\text{-NN } (Q_C, \mathbf{S}, k) = \{R \in \mathbf{S} | \forall A \in \mathbf{C}, \mathbf{B} \in \mathbf{S} - \mathbf{C} \\ \text{RID}(A, Q_C) \leq RID(B, Q_C) \wedge |R| = k\} \quad (12)$$



(a) (b)

Fig. 5. Effect of different level of simulated noise on shape retrieval accuracy using (a) ERA metric (b) CRA metric

For ERA-based evaluation metric, we set $k = 1$ in eq. (12). A set of rankings $\forall Q_C \in \mathbf{S_C}$ is produced. The closest match to Q_C should be its corresponding uncorrupted version in $\mathbf{S}$ which produces a rank value of unity. For ease of comparison we record the proportion of times the query shape is ranked correctly as unity when taken over all S_C. For CRA-based evaluation metric, we set $k = 6$ in eq. (12) as there are six members in each shape class. A set of rankings $\forall Q_C \in \mathbf{S_C}$ is produced. A rank value for each 6-NN query is calculated as a percentage of the correct closest matches in nearest six matches. ERA percentage is then the average rank values for $\forall Q_C \in \mathbf{S_C}$.



(a) (b) (c)

Fig. 6. Shape distances calculated using (a) CRISM (b) integral invariant (c) differential invariant. Lighter shades represent high distances and vice versa.

To establish a base case, we have implemented two different systems for comparison including integral invariants and differential invariants. The experiment has been conducted on silhouette dataset. For the proposed approach, shapes from the silhouette dataset are modeled using DFT-based coefficient feature vector. We assume $m = 5$ in eq. (6). The implementation of integral-invariants and differential invariants is based on the approaches presented in [8] and [9] respectively. Average ERA and CRA based retrieval accuracies for varying amount of noise in silhouette dataset are presented in Fig. 5. The results from Fig. 5 shows that CRISM gives the highest retrieval accuracies in the presence of different noise levels followed by integral invariants. Differential invariants performs worst in the presence of noisy shapes and its performance deteriorates with increasing noise levels.

To highlight the robustness of CRISM to noise and other distortions as compared to competitors, shape distances are presented in graphical format in Fig. 6 for ease of visualization. Noisy query shapes are presented in the left column and the dataset is presented in the top row. The diagonal entries show the distance of noisy query shape with its corresponding non-noisy shape and the non-diagonal entries presents the distance of noisy shapes with other samples. Lower gray levels represent low distances and vice versa. Ideally, the diagonal entries should represent the best match (minimum distance) for the noisy query sample. We would like to have a a block diagonal structure with low gray levels along the diagonal. CRISM presents the best block diagonal structure followed by integral-invariant. Differential invariant presents high distances on the diagonals and does not present a block diagonal structure. Based on these results, we see that the proposed shape matching approach is more robust to the presence of noise in shapes as compared to competitive techniques. We now compare CRISM with some other competitors using different publicly available datasets. Table 2 shows the 1-NN classification accuracies of CRISM (obtained using leave-one-out cross-validation) using different datasets.

Dataset	# of Instances	# of classes	% Accuracy
Diatom	781	37	72.73
Chicken	446	5	80.71
MixedBag	160	9	95.625

TABLE I
ACCURACIES OF CRISM USING VARIOUS SHAPE DATASETS.

Experiment on chicken dataset enable us to compare directly with to [18] and [2] who report the classification accuracy of 79.5% and 80.04% respectively. The shape matching approach specified in [18] takes around a minute for one-one shape matching whereas [2] takes 0.0039 seconds. On the other hand, shape matching using CRISM takes around 2.709×10^{-4} seconds. Similarly, [2] reported the classification accuracy of 72.47% for Diatom datasets whereas CRISM achieved the classification accuracy of 72.73% although CRISM-based shape matching is 2 orders of magnitude faster than the shape matching approach proposed in [2].

VI. Discussion and conclusions

In this paper, we have presented a detailed discussion on shape matching in the presence of noise and other distortions such as articulation and rotation. A CRISM algorithm has been proposed that exploits the contour information for shape matching. Contours are converted into normalized centroid distance based time series and is modeled using orthogonal basis coefficient feature space representation. A critical-point based approach to support efficient rotation-invariant shape matching is presented. The proposed algorithm is robust to affine transformations and other arbitrary distortions. Experimental results are presented to show that CRISM-based shape matching gives better retrieval accuracies than competitive techniques such as differential invariants and integral invariants. The proposed approach demonstrates good discrimination capability than the competitors as reflected in the results presented in Figs. 5-6.

References

[1] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACM Transactions on Graphics*, vol. 21, no. 4, pp. 807–832, October 2002.

[2] E. Keogh, L. Wei, X. Xi, S.-H. Lee, and M. Vlachos, "Lb-keogh supports exact indexing of shapes under rotation invariance with arbitrary representations and distance measures," *VLDB*, 2006.

[3] T. Adamek and N. E. Connor, "A multiscale representation method for nonrigid shapes with a single closed contour," *IEEE CSVT*, vol. 14, no. 5, pp. 742–753, 2004.

[4] T. Adamek and N. E. O. Connor, "Efficient contour-based shape representation and matching," *MIR*, pp. 138–143, 2003.

[5] E. Attalla and P. Siy, "Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching," *Pattern Recognition*, vol. 38, no. 12, pp. 2229–2241, 2005.

[6] G. Mori, S. Belongie, and J. Malik, "Efficient shape matching using shape contexts," *IEEE Transactions on PAMI*, vol. 27, no. 11, pp. 1832–1837, November 2005.

[7] E. R. Davies, *Machine Vision: Theory, Algorithms, Practicalities*. New York: Academic Press, 1997.

[8] S. Manay, D. Cremers, B.-W. Hong, A. J. Yezi, and S. Soatto, "Integral invariants for shape matching," *IEEE Transactions on PAMI*, vol. 28, no. 10, pp. 1602–1618, October 2006.

[9] M. Boutin, "Numerically invariant signature curves," *IJCV*, vol. 40, no. 3, pp. 235–248, 2000.

[10] S. Berretti, A. Bimbo, and P. Pala, "Retrieval by shape similarity with perceptual distance and effective indexing," *IEEE Transaction on Multimedia*, vol. 2, no. 4, pp. 225–239, 2000.

[11] G. Dudek and J. Tsotsos, "Shape representation and recognition from multiscale curvature," *Journal of Computer Vision and Image Understanding*, vol. 68, no. 2, pp. 170–189, 1997.

[12] D. Li and S. Simske, "Shape retrieval based on distance ratio distribution," *HP Tech Report HPL-2002-251*, 2002.

[13] G. Dudek and J. Tsotsos, "Shape representation and recognition from multiscale curvature," *Journal of Computer Vision and Image Understanding*, vol. 68, no. 2, pp. 170–189, 1997.

[14] T. Adamek and N. E. OConnor, "A multiscale representation method for nonrigid shapes with a single closed contour," *IEEE Circuits and Systems for Video Technology*, vol. 14, no. 5, pp. 742–753, 2004.

[15] X. Wang, L. Ye, E. Keogh, and C. Shelton, "Annotating historical archives of images," in *Proceedings of 8th ACM/IEEE-CS Joint Conference on Digital libraries*, 2008, pp. 341–350.

[16] B. Bhanu and X. Zhou, "Face recognition from face profile using dynamic time warping," in *Proceedings of International Conference on Pattern Recognition*, 2004, pp. 499–502.

[17] Z. Wang, Z. Chi, D. Feng, and Q. Wang, "Leaf image retrieval with shape features," in *Proceedings of the 4th International Conference on Advances in Visual Information Systems*, 2000, pp. 477–487.

[18] R. A. Mollineda, E. Vidal, and F. Casacuberta, "Cyclic sequence alignments: Approximate versus optimal techniques," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 16, no. 3, pp. 291–299, 2002.