2D SHAPE CLASSIFICATION BASED ON MULTI-LEVEL BARYCENTER CONTOUR

KOSORL THOURN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF ENGINEERING IN ELECTRONICS ENGINEERING

FACULTY OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2009

KMITL-2009-EN-M-040-153

หัวข้อวิทยานิพนธ์	2D Shape Classification Based on Multi-Level Barycenter Contour
นักศึกษา	Mr. Kosorl Thourn
รหัสนักศึกษา	50060451
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมอิเล็กทรอนิกส์
พ.ศ.	2552
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผศ.ดร.ยุทธนา คิดใจเดียว

บทคัดย่อ

งานวิจัยในปัจจุบันนี้มุ่งความสนใจไปที่การใช้วิธีการ multi-resolution ซึ่งสามารถปรับปรุงความถูกต้องและทนทานต่อสัญญาณรบกวนได้ดี ในวิทยานิพนธ์นี้ได้นำเสนอวิธีการใหม่โดยใช้ multi-scale ซึ่งเรียกว่า Barycenter contour (BcC) decomposition

ในงานวิจัยนี้วิธีการ Barycenter Contour (BcC) decomposition ถูกสร้างจากตำแหน่งของเส้นขอบภาพ ซึ่งใช้การแยกรูปร่างใน 2 มิติที่มีขอบเขตปิดแบบ a multi-scale contour วิธีการนี้สามารถทำให้ขอบเขตของรูปร่างและรูปร่างที่ถูกบังบางส่วนราบเรียบ เพื่อที่จะสร้างอัลกรอริทึมที่ไม่เปลี่ยนแปลงต่อการบิดเบือนแบบแอฟไฟน์ พื้นที่รูปสามเหลี่ยมได้ถูกนำเสนอในการคำนวณหาค่า BcC ในแต่ละระดับ ต่อจากนั้นทำการแปลงโดยใช้ Discrete Fourier Transforms (DFT) เพื่อให้ทนทานต่อการสะท้อนกลับและไม่ขึ้นกับจุดเริ่มต้น

ในขั้นตอนการแยกแยะ ได้นำเสนอ 2 วิธีการ คือ Eigenface และ Fisherface การเปรียบเทียบความเหมือนกันจะถูกวัดค่าโดยวิธี the Euclidian distance หรือ the normalized cross correlation

อัลกรอลิทึมที่นำเสนอนี้ทนทานต่อการแปลงแบบแอฟไฟน์ (การเลื่อนตำแหน่ง, ขนาด, การหมุน และการเฉือน) และการเปลี่ยนแปลงจุดเริ่มต้นของรูปร่าง จากผลการทดลองแสดงให้เห็นว่าวิธีการที่นำเสนอนี้ให้ประสิทธิภาพที่สูงมากกว่างานอื่น ๆ ที่ผ่านมา โดยทดสอบกับฐานข้อมูล 3 แบบต่าง ๆ กัน คือ The affine shape database, the other two well-known databases คือ the MPEG-7 database CE-1 part B และ the Kimai's database

ACKNOWLEDGMENTS

This thesis could not have been written without the support from a large number of people. It is a pleasure to convey my gratitude to all of them in my humble acknowledgement since I came to King Mongkut's Institute of Technology, Ladkrabang (KMITL).

In the first place, I would like to record my gratitude to my supervisors, Asst. Prof. Dr. Yuttana Kitjaidure for his supervision, advice, and guidance from the very early stage of this research as well as giving me extraordinary experiences throughout the work

My sincere gratitude goes also to all Professors, lecturers and supporting staffs in Electronic Department, who always continuously encouraged, helped and gave me a guideline during the whole period of my study at King Mongkut's Institute of Technology, Ladkrabang (KMITL).

To ASEAN University Network/Southeast Asia Engineering Education Development Network (AUN/SEED-Net), I would acknowledge, for awarding me the scholarship with the financial support for my study in Thailand.

I would like to recognize all friends both in Cambodia and Thailand who always encouraged and helped me during the period of my study. Throughout my years at KMITL, they have enriched my life with their companion and friendship.

Finally, I am forever indebted to my parents for their patience, continuous encouragement, love, guidance, and support.

Table of Contents

Abstract (Thai) I

Abstract II

Acknowledgement III

Table of Contents IV

List of Tables VII

List of Figures VIII

Chapter 1 Introduction 1

1.1 Motivation 2

1.2 Research Objectives and methodologies 3

1.3 Thesis outline 4

Chapter 2 Literature Reviews 6

2.1 Shape representation and description 6

2.1.1 Global contour-based techniques 7

2.1.2 Structural contour-based technique 9

2.1.3 Global region-based techniques 10

2.1.4 Structural region-based technique 12

2.2 Similarity measure 13

2.2.1 Minkowski distance (L_p) 13

2.2.2 Cosine distance 13

2.2.3 Dynamic time warping (DTW) distance 14

2.3 Review recent shape matching algorithms 15

2.3.1 Curvature scale space 15

2.3.2 Multi-scale Wavelet-Based 17

2.3.3 Multi-scale Convexity Concavity 18

2.3.4 Triangle Area Representation (TAR) 19

2.3.5 Dynamic Programming 20

2.3.6 WARP 21

2.3.7 Inner Distance 21
2.3.8 Curve Edit Distance 22
2.3.9 Visual Parts 22
2.3.10 Beam Angle Statistics 22
2.3.11 Symbolic Representation 23
2.3.12 Shape Tree 24

Chapter 3 Feature Extraction 25
3.1 Preprocessing 25
3.2 Barycenter contour decomposition 26
3.3 Shape representation 33
3.3.1 Signed enclosed area signature 33
3.3.2 Triangle area representation signature 34
3.4 Invariant to starting point selection 37
3.4.1 Spectral domain 38
3.4.2 Spatial domain 39

Chapter 4 Classification 42
4.1 Principal component analysis (PCA) 42
4.1.1 Training 43
4.1.2 Testing 44
4.2 Fisher discriminant analysis (FDA) 45
4.2.1 Training 45
4.2.2 Testing 47
4.3 Similarity measure 47

Chapter 5 Experiment and Results 49
5.1 Affine shape database 49
5.1.1 Shape recognition 51
5.1.2 Shape matching and retrieval 55
5.2 MPEG-7 database CE-shape-1 58
5.3 Kimia's database 61
5.4 Robustness to noise 65

Chapter 6 Conclusion and Future Work 68
6.1 Conclusions and discussions 68
6.2 Future Work 70

References 71

Appendix A Barycenter 76

Biography 78

List of International Conference Proceeding Papers 79

List of Tables

Table Pages

3.1 The N-DFT WARP coefficients 41

5.1 Affine transformation's parameters 50

5.2 The recognition rates of SEA and TAR at various numbers of training set in spectral domain....51

5.3 The recognition rates of power spectrum of SEA and TAR at various numbers of training set 52

5.4 The recognition rates of SEA and TAR at various numbers of training set in spatial domain....53

5.5 The best-match accuracy of SEA and TAR tested for the affine shape database....56

5.6 The Bullseye test of the previous algorithms testing for the MPEG-7 database part B 59

5.7 The Bullseye test of our algorithm tested for the MPEG-7 database part B....59

5.8 The retrieval rate of different algorithms, evaluating for the Kimia's database 62

5.9 The retrieval rate of SEA using EBcC and FBcC method evaluating for the Kimia's database 62

5.10 The retrieval rate of TAR using EBcC and FBcC method evaluating on the Kimia's database 63

6.1 Time execution for the affine shape database....69

6.2 Time execution for the MPEG-7 shape CE-1 database 69

6.3 Time execution for the Kimia's database....70

List of Figures

Figure Pages

1.1 Examples of some applications related to shape analysis 1

1.2 Example of (a) affine distorted objects, (b) Occluded objects, and (c) complex objects. 2

1.3 The diagram of the proposed algorithm in spectral and spatial domain. 4

2.1 Classification of the shape representation and description techniques 6

2.2 Shape context computation (a) edge image of character 'A'; (b) a point *p* on character and all the vectors (bins) started from *p*; (c) a log-polar histogram bins 8

2.3 Chain codes: (a) 4-connectivity; (b) 8-connectivity; (c) example 10

2.4 (a) An original shape in polar space; (b) A polar-raster sampled image plotted in Cartesian space 11

2.5 (a) Convex hull and its concavities; (b) Concavity tree representation of convex hull 12

2.6 Medial axis of rectangular shape 12

2.7 Illustration of matching (DP) (a) two 1D sequences in [17]; (b) Two TAR representation in [19]. Both are matched by DP. 14

2.8 Curvature scale space representation 16

2.9 Multi-scale representation of shape using wavelet decomposition. 17

2.10 The MCC shape representation (a) original shape, (b) filtered version of (a) at different scale levels, and (c) MCC representation. 18

2.11 (a) A schematic of how to transform the shape into the symbolic representation; (b) The mapping obtained by the algorithm of a given contour into a string of symbols. 23

3.1 The pre-processing process (a) binary 2D shape silhouette (b) extracted contour (c) re-sampling point and (d) shifted contour. 26

3.2 Barycenter contour of bird at the first level 27

3.3 Barycenter contour of bird at the second level 27

3.4 Example of the Barycenter contour decomposition of (a) bird's; (b) horseshoe and (c) device9's contours at various scale levels 1,5,10,20,30,40, and 60. 28

Figure Pages

3.5 The Barycenter contour of bird (a) at level 1; (b) at level 125; (c) both levels are normalized in both horizontal and vertical axe in which level 125 is plotted in reversion direction 29
3.6 The SEA signature of bird in Figure3.1 at various scale levels of BcC 33
3.7 The SEA signature of bird's shape in Figure3.1 plotted in 3D format of (a) non normalization; (b) local normalization; (c) global normalization. 34
3.8 The TAR signature of bird in Figure3.1 at various TSL....35
3.9 The TAR signature at all TSL of bird's shape in Figure3.1 plotted in 3D format of (a) non normalization; (b) local normalization; (c) global normalization. 35
3.10 The TAR signature at TSL = 1 of bird in Figure3.1 at various scale levels of BcC....36
3.11 The TAR signature at TSL = 1 of bird's shape in Figure3.1 plotted in 3D format of (a) non normalization; (b) local normalization; (c) global normalization. 36
3.12 An affine transformation of a triangle *S* into another triangle *T(S)* 37
3.13 (a) The shape silhouette, its 90^0 rotation and its reflection transformation; (b) the corresponding of SEA signature and (c) DFT coefficients of SEA at all levels of BcC as 3D plot. 38
3.14 (a) The shape silhouette, its 90^0 rotation and its reflection transformation; (b) the corresponding of SEA at all scale levels of BcC before phase normalization; (c) after phase normalization....40

4.1 The computation of PCA and FDA including the training and testing steps....42
4.2 The scree plot for selecting the optimum eigenvectors....44
4.3 Subspace axes of PCA and FDA for data points of two classes. 45
4.4 Matching and retrieval shape by normalized cross correlation....47

5.1 Shape in each class of affine shape database 50
5.2 Sample of affine distorted shapes of class butterfly 50
5.3 The recognition rates of SEA and TAR signatures using EBcC and FBcC methods, tested for the affine shape database. 54
5.4 Precision-recall curve of SEA signature tested for the affine shape database 55
5.5 Precision-recall curve of TAR signature tested for the affine shape database....56
5.6 Sample shapes in each class of the MPEG-7 database CE-shape-1 part B....57
5.7 Precision-recall curve of SEA signature tested for the MPEG-7 database CE-shape-1 part B. 57
5.8 Precision-recall curve of TAR signature tested for the MPEG-7 database CE-shape-1 part B....58

Figure Pages

5.9 A comparison of the retrieval accuracy of SEA and TAR in spectral domain for each class of the MPEG-7 database CE-shape-1 part B using FBcC method.....60

5.10 A comparison of the retrieval accuracy of power spectrum of SEA and TAR for each class of the MPEG-7 database CE-shape-1 part B using FBcC method.....60

5.11 A comparison of the retrieval accuracy of SEA and TAR in spatial domain for each class of the MPEG-7 database CE-shape-1 part B using FBcC method.....61

5.12 All 99 shapes in Kimia's database.....62

5.13 Precision-recall curve of SEA signature tested for the Kimia's database.....63

5.14 Precision-recall curve of TAR signature tested for the Kimia's database.....64

5.15 Example of bird's contour at different SNR ratio.....65

5.16 The precision-recall curve of SEA at various SRN ratios.....66

5.17 The precision-recall curve of TAR at various SRN ratios.....67

Chapter 1

Introduction

Several properties obtained from the objects have been used for recognition and categorization such as shape, color, texture, brightness, and others. Of such properties, human can easily recognize and identify a type of the object using their geometries (shape) rather than intensities (colour, texture, brightness). That is why shape has been considered as an important visual feature and as the most promising in a system for object recognition, matching, registration, and analysis. In the research and application area of artificial vision, shape can be modeled by a mathematical function (points, lines, curves, planes, and so on).

In computer vision area, many researchers have developed several algorithms or computer program for artificial vision. Such a wide range of algorithms is not only applied to computer but also robot and other machines for their artificial visions which provide an ease for life and are important for industrial applications, and others.

Based on shape analysis, numerous applications of object recognition have been met in many areas — some examples as shown in Figure 1.1 include:

- Robotic control, for example, to determine its topological location and to perform tasks using specific items

Figure 1.1: Examples of some applications related to shape analysis

- Medical imaging to understand shape changes related to illness or aid surgical planning
- Virtual environments or on the 3D model market to identify objects for copyright purposes
- Security applications such as face recognition, fingerprint, iris, signature verification
- Document analysis such as WWW, OCR (optical character recognition), online character recognition, multimedia database, historical documents
- Agriculture, for example, harvest control, seed counting and quality control, species identification, fruit maturation analysis

1.1 Motivation

Now a day, even though the technologies are developed more and more, it still has problems, for example, artificial visions which inspire researchers to develop and find solution corresponding to such problems. Several new algorithms have been developed and created. Of many state-of-the-art techniques, some have been set up into commercial systems while there are still under research development to support various applications. In this thesis, two important and challenging things have motivated us to propose a new shape representation and to improve the algorithm for shape classification.

In real world situation when the 3D object is taken by a camera in multiple viewpoints, it normally results differently deformed 2D shapes. This means that one dimension of object information is lost when 3D object is projected onto 2D image plane. These 2D shapes can be modeled by affine transformation, if the viewpoints are sufficiently far away. Furthermore, the 2D image is corrupted with noise, defects, arbitrary distortion and occlusion (Examples of such image are shown in Figure 1.2).

Figure 1.2: Example of (a) affine distorted objects, (b) Occluded objects, and (c) complex objects.

This is the first motivation which encourages us and other researchers to develop shape representation and matching algorithm for increasing the classification performance of such artificial vision of the system.

The other motivation has come with recent developments in digital imaging technologies, that is, an increasing number of images are generated every day. Millions of images are available via the internet. Therefore, there is a growing interest in finding images in large collections or from remote databases. In order to achieve this task, images have to be represented by specific features.

1.2 Research Objectives and methodologies

Recently, a growing research interest has been focused on the multi-resolution technique for shape analysis in order to improve the accuracy and robustness to the moderate amount of noise. Curvature Scale Space (CSS), Wavelet descriptor, Multi-scale Convexity Concavity (MCC), and others are the examples of the multi-resolution technique. Although, the performance have been achieved by these techniques still have been under improvement. The main objective of this study is to develop a new multi-scale representation modeling from shape boundary called Barycenter Contour (BcC) decomposition.

The assumption of this study is that our proposed technique has been evaluated by three different databases (Chapter 5) in which all shapes are the binary images. Their boundaries are considered as closed shape boundary by joining the contour's end points with linear line if their contours are the opened ones.

In this study, the Barycenter Contour (BcC) decomposition (in Chapter 3) has been utilized to decompose 2D shape closed boundary into multi-scale contours for improving the performance of both recognition and retrieval accuracy. Then, the triangle area (the Signed Enclosed Area (SEA) and the Triangle Area Representation (TAR) which are used alternatively) is computed at each scale level of the Barycenter Contour (BcC) as the shape representation. Next, Discrete Fourier Transform (DFT) is employed onto the feature to resolve the starting point dependency's problem. After that, the Eigenface and Fisherface technique are utilized in the classification process because these methods are frequently applied for face recognition systems for dimensionality reduction, making discrimination and indexing of the object in each category. Finally, the normalized cross correlation or the Euclidian distance is taken into account for a similarity measure. All the steps of our algorithm are depicted in Figure 1.3.

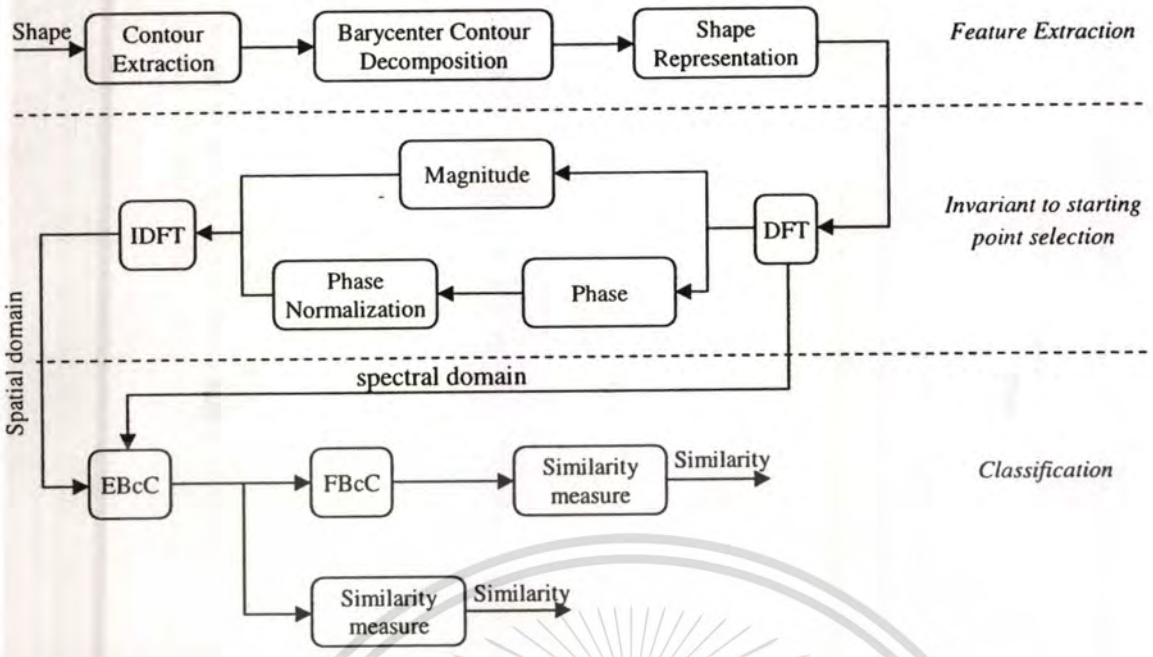


Figure 1.3: The diagram of the proposed algorithm in spectral and spatial domains

1.3 Thesis outline

The remainder of this thesis is organized as follows:

Chapter 2 presents the background information and reviews of the recent algorithms in the literature. The background includes the main methodologies for representing and describing shapes, distances for measuring the similarity between objects.

Chapter 3 describes the feature extraction. In this chapter, a new and simple multi-resolution method, called Barycenter Contour (BcC) decomposition, and shape representations are presented. Furthermore, Discrete Fourier Transform (DFT) is taken into consideration for automatically fixing the starting point selection which can affect onto the performance of classification.

Chapter 4 is concerned about the shape classification. The Principal Component Analysis (PCA) and Fisher Discriminant Analysis (FDA), called Eigen Barycenter Contour (EBcC) and Fisher Barycenter Contour (FBcC) respectively in this literature, have been done as the classifier. The Euclidian distance and Normalized Cross-Correlation are utilized for measuring the similarity between shapes.

Chapter 5 will show the experiment and the results. Further discussion is given. Two well-known databases, MPEG-7 shape CE-1 part B and Kimia's database, and another database developed by the author called affine shape database are experimented to evaluate the performance of our proposed

algorithm under affine transformation. The performance of the system is observed by the recognition rate, the precision-recall curve and the Bullseye score.

Chapter 6 summarizes the main contribution of this thesis and the suggestion of the future work.

Chapter 2

Literature Reviews

This chapter presents a background and review about the shape representation and descriptor which is in section 2.1. The last section describes a detailed review of the shape analysis techniques in the literature. For further reading and more information, the reader can refer to the books about digital image processing [1], shape analysis and classification [2], and pattern recognition [3].

2.1 Shape representation and description

Various shape representation and description methods have been proposed in the past [4], for example, shape context, Fourier descriptor, polygon decomposition, geometric moment invariants, convex hull and so on. As depicted in Figure 2.1, these methods can be classified into two classes, contour-based methods and region-based methods based on whether shape features are extracted from the contour only or from the whole shape region. In each class, the method is further divided into structural approaches and global approaches. Global approaches usually compute a numeric feature vector from a shape. Then, the matching is conducted using a metric distance, such as Euclidian distance or city block distance. On the other hand, with the structural approaches, shapes are split into segment called primitives. The similarity matching in this case is usually performed using string or graph matching.

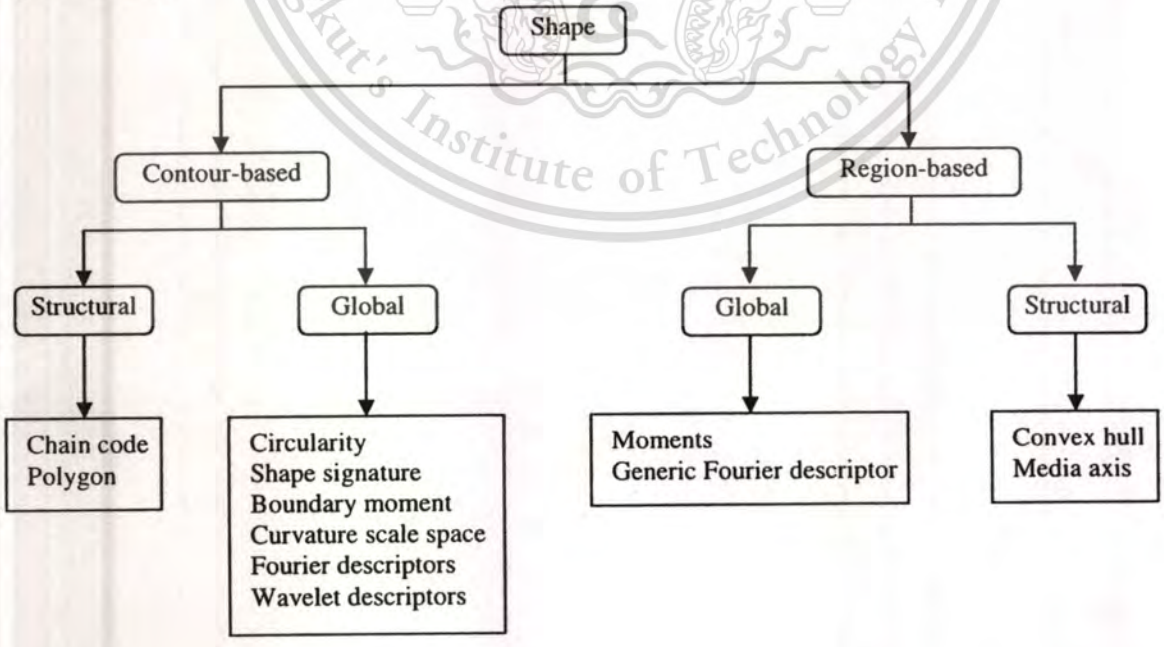


Figure 2.1: Classification of the shape representation and description techniques

The effectiveness of a shape retrieval system depends mainly on the type of the used shape descriptor and the efficiency of the shape matching function. There are some criteria that should be taken into consideration when developing a shape representation and descriptor for reliable shape matching and retrieval, which include [5]:

- *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.
- *Robustness*: a representation must be resistant to moderate amounts of distortion and noise, which are a typical of real images with 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) cause in variations of the representation of similar magnitude.
- *Invariance*: a representation must be invariant to geometric transformation such as affine transformation including translation, scale, rotation, and shear. Further invariant assumptions are a viewing angle changes, symmetric and reflection of the shapes.
- *Scalability*: a representation 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-scale (resolution).
- *Efficiency*: a representation must be efficient to compute and store. This is important since it may be necessary for an object recognition and retrieval system to perform in real-time.

2.1.1 Global contour-based techniques

2.1.1.1 Circularity

Circularity [4] reflects the compactness of the contour of an object which is defined as follows:

$$circularity = \frac{perimeter^2}{Area} \tag{2.1}$$

Circularity has a large value for elongated objects and it is roughly correlated with the complexity of the contour. It has a minimum value of 4π for circle. It is also invariant to scaling, translation and rotation.

2.1.1.2 Shape signature

A shape signature, derived from shape boundary points, is a one dimensional function for describing a two dimensional shape. Many shape signatures have been proposed including complex coordinates, centroid distance, chord-length, cumulative angle, curvature, and area signature. The two shape signatures, chord-length and area signatures, have been proposed by Zhang and Lu [4] (more information about shape signatures is in the literature [6]).

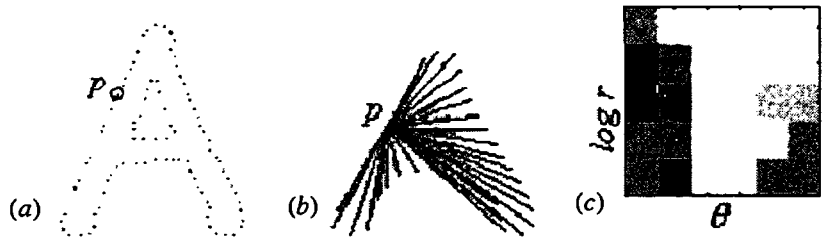


Figure 2.2: Shape context computation (a) edge image of character 'A'; (b) a point p on character and all the vectors (bins) started from p; (c) a log-polar histogram bins

Shape signatures are usually normalized into being translation and scale invariant. The area signature is invariant to affine transforms (translation, scale, rotation, and shear).

2.1.1.3 Shape context

Shape context has been introduced by Belongie et al. [7]. An object is represented by a discrete set of points sampled regularly along the contour. For every point, a log-polar histogram, the shape context as shown in Figure 2.2 is computed, approximating the distribution of adjacent point locations relative to a reference point. In order to achieve scale invariance, the outer radius for the histograms is set equal to the mean distance between all the pair points.

For a point $p_i \ (i = 1, 2, ..., N)$ of the shape, the shape context is a coarse histogram h_i of the relative coordinates of the remaining q_j $N-1$ points:

$$h_i(k) = \#\{q \neq p_i : (q - p_i) \in bin(k)\} \tag{2.2}$$

The bins are uniform in log-polar space, making the descriptor more sensitive to positions of nearby points than those of more distant points. The cost of matching a pair of points p_i and q_j from two shapes is computed as:

$$C(p_i, q_j) = \frac{1}{2}\sum_{k=1}^{K}\frac{\left[h_i(k) - h_j(k)\right]^2}{h_i(k) + h_j(k)} \tag{2.3}$$

where $h_i(k)$ and $h_j(k)$ denote the K-bin normalized histogram at p_i and q_j, respectively. This definition of shape context is not rotationally invariant, but it can be easily modified to be a completely rotational invariant descriptor. There are several ways to obtain rotational invariance. One possibility is to make the reference angle for shape context equal to the local tangent angle on the shape. Another option is to align the shape context axis with the principal axis of the shape. Shape context is not

invariant for mirror transformation; therefore when computing the shape context, the mirrored shape is also considered.

2.1.1.4 Fourier descriptor

Due to its simple computation, clarity and coarse-to-fine description capability, Fourier descriptors (FD) are one of the most widely used as shape descriptor. Three different types of FD have been presented in [4] such as conventional FD method, affine FD method, and short time FD method. For the conventional FD method, it is the complex coefficients of the Fourier series expansion of waveforms which is computed as the discrete Fourier transforms (DFT) by:

$$F(n)=\sum_{k=0}^{N-1} f(k)\exp\left(-j\frac{2\pi kn}{N}\right) \tag{2.4}$$

for $n,k=0,1,2,.....,N-1$. The complex coefficients $F(n)$ are called the Fourier descriptors of a signal $f(k)$ which may be the shape signature as presented above.

2.1.2 Structural contour-based technique

2.1.2.1 Chain code representation

Chain code, which was introduced by Freeman [8], describes an object by a sequence of unit-size line segments with a given orientation. In this approach, an arbitrary curve is represented by a sequence of vectors of unit length and a limited set of possible directions. From a selected starting point, a chain code can be generated by using 4-directional or 8-directional chain code as shown in Figure 2.3. Furthermore, it can be generated by N-dimensional chain code ($N>8$ and $N=2^k$) called general chain code. An example of representing a shape using 4-directional chain code is shown in Figure 2.3. This method is sensitive to noise and to the starting point.

2.1.2.2 Polygon decomposition

The shape boundary is decomposed into line segments by polygon approximation [9]. Then, the polygon vertices are used as primitives. Each primitive feature is a 4-element string which consists of the internal angle, distance from the next vertex and its x and y coordinates. Obviously, this feature is not invariant to translation, scale, and rotation.

For efficiency, only a fixed number of sharpest vertices are considered. In this approach, the shape is represented as string of line segments which are then organized into a tree data structure. The drawback of this method is impractical for natural objects, but it work well for artificial drawings.

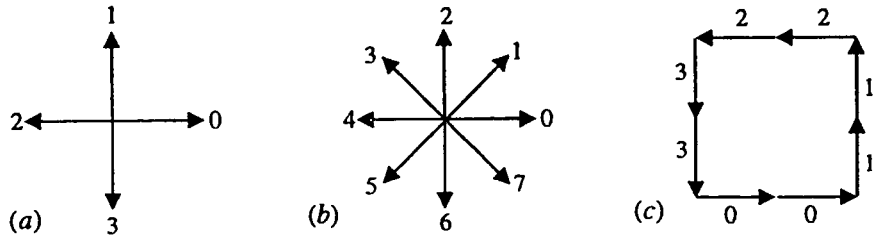


Figure 2.3: Chain codes: (a) 4-connectivity; (b) 8-connectivity; (c) example

2.1.3 Global region-based techniques

2.1.3.1 Geometric moment invariants

Given a binary image $f(x,y)$, moments of order $(p+q)$ are defined as:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \tag{2.5}$$

The central moments of $f(x,y)$ are defined as:

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y) \tag{2.6}$$

where (x,y) are the pixel coordinates and $(\bar{x},\bar{y})$ are the coordinates of the region's centre of gravity which can be obtained by $\bar{x} = \frac{m_{10}}{m_{00}}$ and $\bar{y} = \frac{m_{01}}{m_{00}}$.

In 1962, Hu [10] published the first significant paper on the use of moment invariant for 2D shape recognition application through his seven different moments which are invariant with respect to translation, scale, and rotation. Then, the normalized central moments are computed:

$$\vartheta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\lambda}} \quad \text{with} \quad \lambda = 1 + \frac{p+q}{2} \text{ and } (p+q) \geq 2 \tag{2.7}$$

The first to the seventh order moment invariant, which are derived by Hu, are defined as follows:

$$\begin{aligned}
\varphi_1 &= \vartheta_{20} + \vartheta_{02} \\
\varphi_2 &= (\vartheta_{20} - \vartheta_{02})^2 + 4\vartheta_{11}^2 \\
\varphi_3 &= (\vartheta_{30} - 3\vartheta_{12})^2 + (3\vartheta_{21} - \vartheta_{03})^2 \\
\varphi_4 &= (\vartheta_{30} + \vartheta_{12})^2 + (\vartheta_{21} + \vartheta_{03})^2 \\
\varphi_5 &= (\vartheta_{30} - 3\vartheta_{12})(\vartheta_{30} + \vartheta_{12})\left[(\vartheta_{30} + \vartheta_{12})^2 - 3(\vartheta_{21} + \vartheta_{03})^2\right] + \\
&\quad (3\vartheta_{21} - \vartheta_{03})(\vartheta_{21} + \vartheta_{03})\left[3(\vartheta_{30} + \vartheta_{12})^2 - (\vartheta_{21} + \vartheta_{03})^2\right] \\
\varphi_6 &= (\vartheta_{20} - \vartheta_{02})\left[(\vartheta_{30} + \vartheta_{12})^2 - (\vartheta_{21} + \vartheta_{03})^2\right] + 4\vartheta_{11}(\vartheta_{30} + \vartheta_{12})(\vartheta_{21} + \vartheta_{03}) \\
\varphi_7 &= (3\vartheta_{21} - \vartheta_{30})(\vartheta_{30} + \vartheta_{12})\left[(\vartheta_{30} + \vartheta_{12})^2 - 3(\vartheta_{21} + \vartheta_{03})^2\right] - \\
&\quad (\vartheta_{30} - 3\vartheta_{12})(\vartheta_{21} + \vartheta_{03})\left[3(\vartheta_{30} + \vartheta_{12})^2 - (\vartheta_{21} + \vartheta_{03})^2\right]
\end{aligned} \tag{2.8}$$

However, these representations are not invariant under general affine transformation. Flusser and Suk [11] have presented a complete set of four affine moment invariants derived from second and third order moments and computed as follows:

$$I_1 = \frac{\mu_{20}\mu_{02} - \mu_{11}^2}{\mu_{00}^4}$$

$$I_2 = \frac{\mu_{30}^2\mu_{03}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} + 4\mu_{30}\mu_{12}^3 + 4\mu_{21}\mu_{03}^3 - 3\mu_{21}^2\mu_{12}^2}{\mu_{00}^{10}}$$

$$I_3 = \frac{\mu_{20}\left(\mu_{21}\mu_{03} - \mu_{12}^2\right) - \mu_{11}\left(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}\right) + \mu_{02}\left(\mu_{30}\mu_{12} - \mu_{21}^2\right)}{\mu_{00}^7} \tag{2.9}$$

$$I_4 = \frac{\begin{pmatrix} \mu_{20}^3\mu_{03}^2 - 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} - 6\mu_{20}^2\mu_{02}\mu_{21}\mu_{03} + 9\mu_{20}^2\mu_{02}\mu_{12}^2 \\ +12\mu_{20}\mu_{11}^2\mu_{21}\mu_{03} + 6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} - 18\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} \\ -8\mu_{11}^3\mu_{30}\mu_{03} - 6\mu_{20}\mu_{02}^2\mu_{30}\mu_{12} + 9\mu_{20}\mu_{02}^2\mu_{21}^2 + 12\mu_{11}^2\mu_{20}\mu_{30}\mu_{12} \\ -6\mu_{11}\mu_{02}^2\mu_{30}\mu_{21} + \mu_{02}^3\mu_{30}^2 \end{pmatrix}}{\mu_{00}^{11}}$$

2.1.3.2 Generic Fourier descriptor

A generic Fourier descriptor (GFD) has been proposed by Zhang and Lu [4]. The GFD is a 2D Fourier transform which is applied onto a sampled shape image in a polar space. It is computed by:

$$PF(\rho,\phi) = \sum_r \sum_i f(r,\theta_i) \exp\left[j2\pi\left(\frac{r}{R}\rho + \frac{2\pi i}{T}\phi\right)\right] \tag{2.10}$$

where $f(x,y)$ is the sampled shape image in Cartesian space, $f(r,\theta_i)$ is the sampled shape image in polar space. $r = \sqrt{(x-x_c)^2 + (y-y_c)^2}$ $(0 \le r \le R)$ and $\theta_i = i\frac{2\pi}{T}$ $(0 \le i < T)$; $0 \le \rho < R$, $0 \le \phi < T$. R and T are the radial frequency resolution and angular frequency resolution respectively. The normalized coefficients are the GFD. Figure 2.4 depicts the original image in (a) and its polar-raster plotted in Cartesian space in (b).

Figure 2.4: (a) An original shape in polar space

(b) A polar-raster sampled image plotted in Cartesian space

2.1.4 Structural region-based technique

2.1.4.1 Convex hull

A region R is convex if and only if for any two points $P_1, P_2 \in R$, the whole line segment whose end points are P_1 and P_2 is also inside R. The convex hull of a region is the smallest convex region H which satisfies the condition $R \subseteq H$ [12].

The convex hull can be used to build a tree structure of region concavity called concavity tree [13]. A concavity tree is constructed recursively during the generation of the convex hull. This concept is illustrated in Figure 2.5, where the convex hull of the whole region is constructed first, and convex hull of concave residues are found next. The resulting convex hull of concave residues of the region from previous steps is searched until no concave residuum exists. Thus, the nodes in a concavity tree represent convex hulls and the root represents the convex hull of all objects in the image. Nodes in the first level represent the convex hulls of the concavities (or holes) and nodes in the second level represent convex hulls of meta-concavities, and so on. The resulting tree is a shape representation of the region.

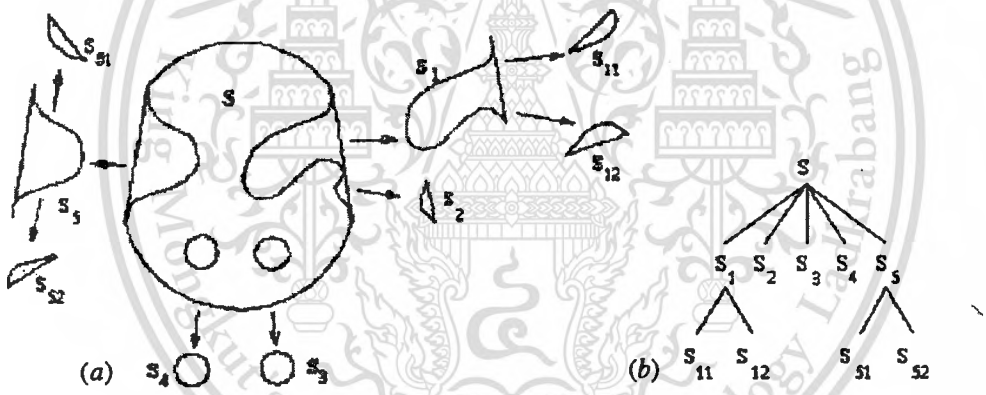


Figure 2.5: (a) Convex hull and its concavities
(b) Concavity tree representation of convex hull

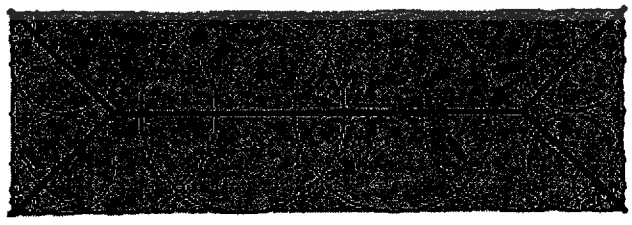


Figure 2.6: Medial axis of rectangular shape

2.1.4.2 Medial axis

Like the convex hull, the skeleton of an object can be employed for shape description. The basic idea is to eliminate redundant information while retaining only the topological structure of the object. The medial axis is the locus of the centers of maximal circles that fit within the shape [14], as illustrated in Figure 2.6. The skeleton is then segmented and represented as a graph according to certain criteria. The matching between shapes becomes graph matching problem. This method is sensitive to noise and requires high computations.

2.2 Similarity measure

2.2.1 Minkowski distance (L_p)

A large number of statistical shape analysis methods use the L_p distance [15] as a similarity measure. It works for vectors of a fixed size. Let $x, y \in \Re^n$, L_p is defined as:

$$L_p(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} \tag{2.11}$$

where $p > 0$. For $p = 1$, L_p becomes the Manhattan or sum of absolute difference (SAD) distance. When $p = 2$, L_p becomes the well-known Euclidian distance. L_p is a matrix.

2.2.2 Cosine distance

The Cosine distance is defined as:

$$D = 1 - \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} p_i^2 \sum_{i=1}^{n} q_i^2}} \tag{2.12}$$

For measuring the similarity among shapes, the minimum distance corresponds to a very similar shape. However, it is different from the normalized cross correlation, which is defined by Eq. (2.13), the very similar shape correspond to its maximum value.

$$R = \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} p_i^2 \sum_{i=1}^{n} q_i^2}} \tag{2.13}$$

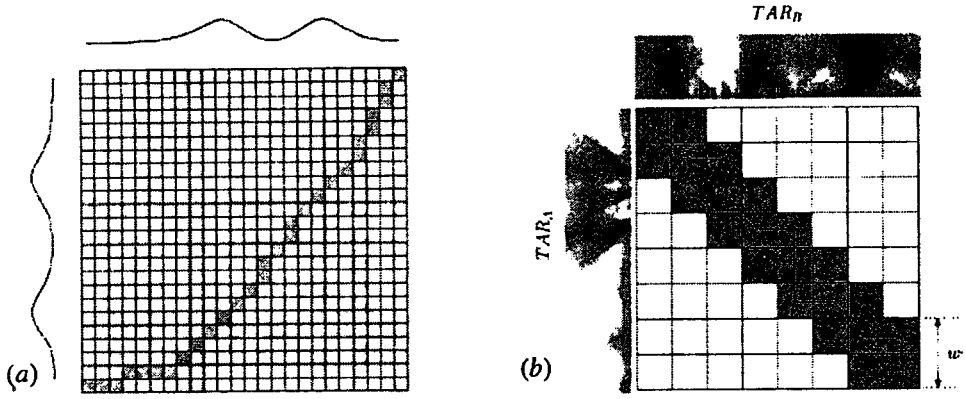


Figure 2.7: Illustration of matching (DP) (a) two 1D sequences in [17]

(b) Two TAR representation in [19]. Both are matched by DP.

2.2.3 Dynamic time warping (DTW) distance

Since a past decade, the dynamic programming (DP), called dynamic time warping (DTW), has been taken into consideration in a system of 2D shape matching [5]. The idea of using Dynamic programming for matching 1D sequence originally comes from the speech recognition community [16] where optimal alignment between two 1D sequences is searched via DP. Unlike the Euclidian distance that provides one-to-one alignment, nonlinear alignment can be achieved by the DTW, where one point on the sequence can be aligned to one or more points on another sequence. For example, in WARP method [17], DTW distance is applied for shape matching and retrieval under 1D sequence. This distance is further utilized under 2D sequence. In [5], a DP table (the example of DP tables of WARP method and TAR image are shown in Figure 2.7) is used to find the least cost match between segments of two curves. Merging of segments is allowed during the matching to facilitate a more meaningful correspondence between segments. In the MCC method [18], their DP algorithm searches for optimal correspondence between the N-points boundaries. A window (w), which limits the optimal path to be around the diagonal, is used to make the search more efficient. The details of DP algorithm in [19] is described as follows: at first, the distance between two individual contour points is defined as

$$D(n,m)=\frac{1}{T_s}\sum_{t_s=1}^{T_s}\left|TAR_A(n,t_s)-TAR_B(m,t_s)\right| \tag{2.14}$$

where $n,m\in\langle 1,N\rangle$ are the index of boundary points and $t_s\in\langle 1,T_s\rangle$ is the triangle side length.

Then, an $N\times N$ distance table, DTW, is constructed to find the optimal correspondence between the points of the two contours. The columns of DTW represent the points of one contour and the rows represent the points the other. Initially, the elements of DTW are set as:

$$DTW_{initial}(n,m)=\begin{cases}0 & \max(1,n-w+1)\le m\le \min(N,n+w-1)\\ \infty & \text{otherwise}\end{cases} \quad (2.15)$$

Finally, the DTW distance is computed as

$$DTW(n,m)=D(n,m)+\min\begin{cases}DTW(n-1,m)\\ DTW(n-1,m-1)\\ DTW(n,m-1)\end{cases} \quad (2.16)$$

The least cost path through the DTW is minimum value of $DTW(N,N)$.

2.3 Review recent shape matching algorithms

The majority of the existing techniques for shape representation and matching fall into this section. These issues include descriptor compactness, robustness to noise, scalability, invariance to affine transformations, and efficiency and effectiveness of the matching.

Since the nature of human perception on shapes is in a multi-scale formats; then the multi-scale approach for shape representation and matching is considered the most promising in the system for recognition, analysis, registration, etc. In addition, many interesting shape properties are revealed at different scale levels. Another advantage includes its invariance to moderate amounts of deformations and noise. In the following, recent shape matching methods and related works are reviewed.

2.3.1 Curvature scale space

The curvature scale space (CSS) is one of the well-known shape-based techniques for shape representation and description which has been proposed by Mokhtarian and Mackworth [20, 21] and has been selected for MPEG-7 standard [22]. First, a parameterized representation of the shape boundary is extracted into two 1D curve in term of $x(u)$ and $y(u)$. The curvature of that shape boundary can be formulated as follows:

$$\kappa(u)=\frac{\dot{x}(u)\ddot{y}(u)-\ddot{x}(u)\dot{y}(u)}{\left(\dot{x}(u)^2+\dot{y}(u)^2\right)^{\frac{3}{2}}} \quad (2.17)$$

where $\kappa(u)$ is the curvature function, $\dot{x}$ and $\ddot{x}$ are the first and second derivation of x, respectively.

If $g(u,\sigma)$ is the Gaussian kernel of standard deviation σ, the components of an evolved digital curve are:

$$\begin{cases}X(u,\sigma)=x(u)*g(u,\sigma)\\ Y(u,\sigma)=y(u)*g(u,\sigma)\end{cases} \quad (2.18)$$

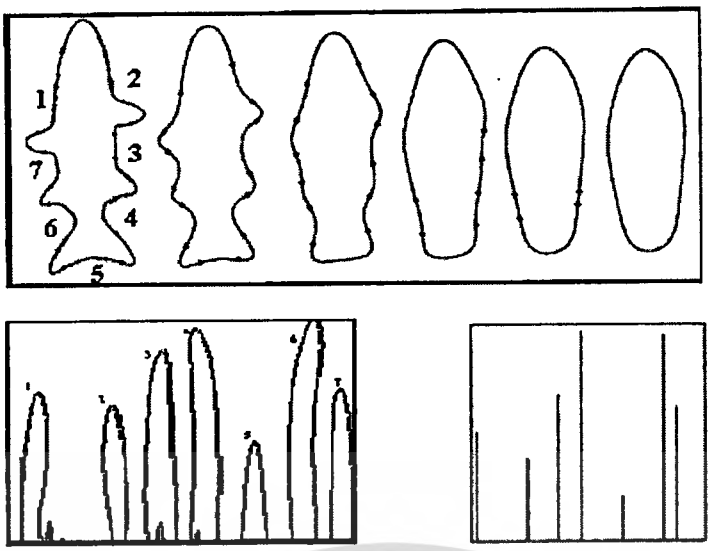


Figure 2.8: Curvature scale space representation

According to the properties of convolution, the derivatives of every component can be calculated:

$$\begin{cases} X_u(u,\sigma) = x(u) * g_u(u,\sigma) \\ Y_u(u,\sigma) = y(u) * g_u(u,\sigma) \end{cases} \text{ and } \begin{cases} X_{uu}(u,\sigma) = x(u) * g_{uu}(u,\sigma) \\ Y_{uu}(u,\sigma) = y(u) * g_{uu}(u,\sigma) \end{cases} \tag{2.19}$$

The curvature scale space of an evolved digital curve is given by:

$$\kappa(u,\sigma) = \frac{X_u(u,\sigma)Y_{uu}(u,\sigma) - X_{uu}(u,\sigma)Y_u(u,\sigma)}{\left(X_u(u,\sigma)^2 + Y_u(u,\sigma)^2\right)^{\frac{3}{2}}} \tag{2.20}$$

where $\kappa(u,\sigma)$ is the curvature scale space at location u and scale σ. By setting Eq. (2.20) to zero, the inflection points (curvature zero crossings) are located at each scale. As the scale level increases, the smoothing effect increases and the number of inflection points decreases until the contour becomes totally convex as shown in Figure 2.8. Only the maxima of the CSS images' contours are used for matching two shapes [23]. In order to achieve rotation or starting point invariance, correspondence between two sets of maxima points is established by finding the optimum horizontal shift of one CSS image with respect to the other that results in the least cost, which represents the dissimilarity function. The cost is defined in terms of the differences of the scale levels of the corresponding maxima. Many heuristics are employed to find the best correspondence efficiently. This method exhibits certain degree of robustness to affine transformation [24], where in another implementation explicit estimation of the affine parameters using the least squares approach was employed to verify the matching results [25]. The main limitations of the CSS method include its limited representation to only the concave segments

and its failure to discriminate a shallow concavity from a deep one. For instance, totally convex shapes like squares and triangles have the same CSS image, which is only the background. Another disadvantage is the requirement for a large number of scales to obtain the CSS image (may exceed 400).

2.3.2 Multi-scale Wavelet-Based

I. El Rube et al. [26, 27] have presented a multi-scale wavelet-based (MWB) representation for 2D closed shape contour. In their method, a closed-contour is represented by its x and y coordinates and 1-D dyadic wavelet transform is applied separately to each coordinate sequence in order to decompose the contour into multi-scale levels as shown in Figure 2.9. Then, the invariant moments of the wavelet approximation coefficients at all scale levels are computed. For detail coefficients, it is difficult to obtain the moments because of the nature of the detail coefficients. As a result, the areas of the triangle between any three adjacent points on the wavelet detail coefficient of shape contour are computed instead. In the matching process, it is performed following two steps. The first step, the global dissimilarities are computed based on a distance matrix which is constructed from the Euclidian distance between moment invariants. However, the detail-based similarity is calculated for globally similar shapes only. This means that the detail coefficients are used only if it is necessary to differentiate between similar shapes. This scheme provides coarse-to-fine matching in order to eliminate dissimilar shapes. It has been demonstrated to perform well under affine transformations and small boundary deformations.

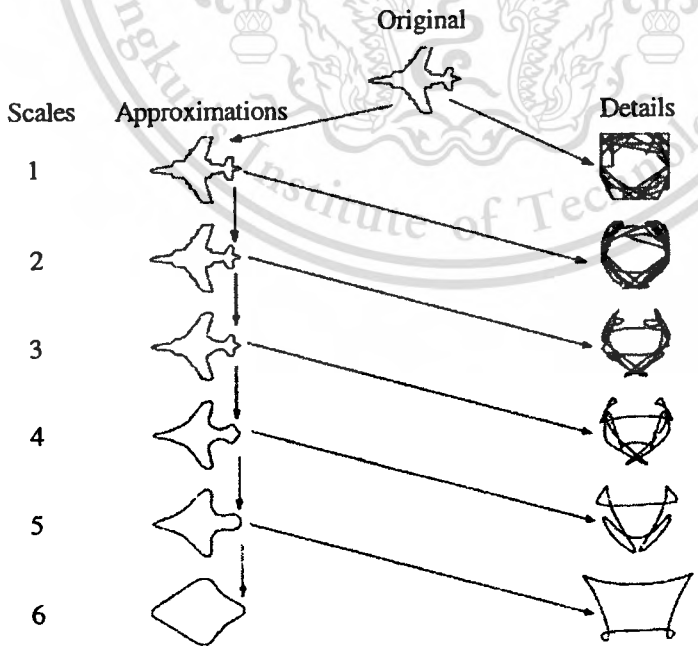


Figure 2.9: Multi-scale representation of shape using wavelet decomposition.

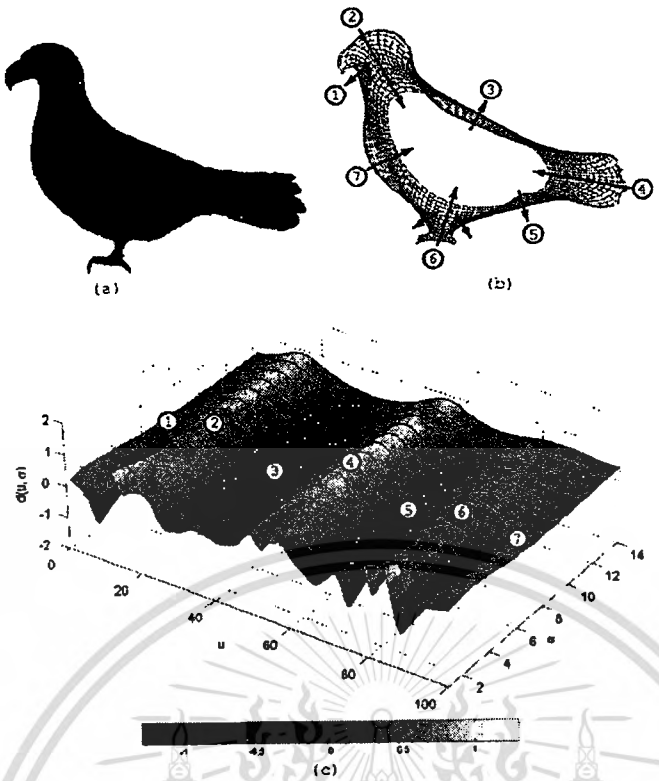


Figure 2.10: The MCC shape representation (a) original shape, (b) filtered version of (a) at different scale levels, and (c) MCC representation.

2.3.3 Multi-scale Convexity Concavity

In a literature [18], T. Adamek and N.E. O'Connor have proposed another multi-scale representation for closed contours that makes use of both concavities and convexities of all contour points. It is called multi-scale convexity concavity (MCC) representation and depicted in Figure 2.10. The MCC representation is stored in the form of a 2-D matrix where the columns correspond to contour point location and the rows correspond to the different scale levels. The simplified contours at different scale levels are obtained via a curve evolution process which is very similar to the way of extracting the CSS image [20, 21, 23-25]. However, a new measure for the curvature is used based on the relative displacement of a contour point with respect to its position in the preceding scale level. This idea is motivated by the observation when smoothing a closed contour, convex and concave points are moved inside and outside the contour, respectively. The amount of displacement reflects the curvature degree. Afterwards, the matching is done using a dynamic programming (DP) approach where a global optimal match is searched between two MCC representations using a DP table. In the DP table, each entry denotes the distance between two points (one from each contour). This distance is chosen to be the sum of absolute differences between the columns of each point in the MCC matrices. The search for the

optimal match is restricted with many heuristics; for instance, no more than two points on one contour are matched to a single point on the other. The rotation invariance is achieved in similar manner as the CSS method, that is, by repeating the algorithm for each contour point as the starting point. The MCC was able to achieve better average accuracy (84.9%) than the CSS method (80.5%) when they were tested on the MPEG-7 dataset. However, the MCC is computationally expensive, $O(N^3)$ where N is the number of contour points. The robustness of the MCC method to affine transformation was not demonstrated.

2.3.4 Triangle Area Representation (TAR)

I. El Rub et al. [27] have proposed the zero crossing of the triangle area representation (TAR) at multi-scale wavelet levels (called MTAR) to construct MTAR images. First, the wavelet transform is used for smoothing and decomposing the shape boundaries into multi-scale levels. At each scale level, then, the TAR images are computed from each three consecutive and equally apart point on the shape boundary and the corresponding Maxima-Minima lines are obtained. The matching of TAR images is based on the peaks and locations of the concavity maxima in images. This scheme shows that, for the same shape, both the MTAR image and CSS image describe concavities along the contour but measure in different way. CSS method measures the curvature as the contour is smoothed by the Gaussian kernel at different scales. On the other hand, each MTAR image describes the locations of the concavities using different triangle side length at a specific wavelet-smoothed scale levels. For testing on the MPEG-7 dataset, the results show that MTAR images outperform CSS images based on the precision-recall curve.

Another recent work, A. Alajlan et al. [19] have developed a multi-scale approach for 2D closed shape matching and retrieval. In their algorithm, TAR is utilized in order to derive the multi-scale representation for 2D closed boundary shapes. Then, this triangle area normalization is made locally per scale, which is employed in the matching via dynamic programming (DP). The DP algorithm (called dynamic space warping (DSW) in this literature) is employed to find the best alignment between two shape representations. For the MPEG-7 Core Experiment CE-shape-1 part B retrieval test, this method achieves high retrieval accuracy of 85.03% without corporation of the global parameters (circularity, eccentricity, aspect ratio). However, when combine with the global parameters, the retrieval efficiency is of 87.23%. Regarding the computational complexity, the matching complexity of this algorithm is $O(N^2)$, where N is the number of the boundary points.

In both literatures, it has been demonstrated that TAR provides useful information about shape features such as the convexity and concavity at each boundary point; therefore, TAR provides high discrimination capability and affine invariant. Besides, TAR exhibits high robustness to the affine transformation, it is also robust against noise and moderate amounts of deformations.

In our previous work [28] have presented the algorithm for 2D shape recognition under general affine transformation. In their techniques, TAR at various levels of triangle side length (TSL) is computed as the shape representation. Then, the normalized cross correlation (NCC) function is utilized for similarity measure among shapes where the achievement of being invariant to starting point is made by repeatedly calculating the NCC for each contour point as the starting point. In the work [29], the authors have developed their algorithms by using artificial neural network (ANN) as the classifier and using Fisher's discriminant technique to find the optimum TSL of the shape representation. They have demonstrated that ANN achieves high speed and high accuracy even under a moderate rage of visual transformation if comparing to NCC and the optimum level is found at level 20 (the maximum level there is 49).

2.3.5 Dynamic Programming

Milios et al. proposed one of the earliest approaches for matching shapes using dynamic programming [5, 30]. In their approach, implicit multi-scale matching takes place through matching merged contour segments in order to avoid the cost of computing the scale space explicitly. At first, the contour is approximated into polygons using cubic B-splines and inflection points are located, which segments the contour into convex and concave segments. Then, a set of geometrical features is computed for each segment to guide the merging process. The dynamic programming (DP) algorithm examines all possible merges of small segments of one shape to match with large segments of the other and selects the best merge, i.e., that results in the minimum cost. An upper limit K is imposed on the maximum number of merged segments to compromise speed with accuracy. The rotation invariance is achieved in similar way to the MCC method. In this case, both shapes are closed, the complexity of the algorithm is $O(M^3N^2)$, where M and N are the number of inflection points of the two contours, and reduces to $O(K^2M^2N)$ when K is considered. The author did not report superior performance over other existing methods. The main limitations of this method include the lack of robustness to the general affine transformations and the high computational complexity of the matching process.

2.3.6 WARP

Bartolini et al. proposed a method for shape matching and retrieval based on the Fourier descriptor that is called WARP [17]. This literature is related to our work to resolve the problem of starting point independency. In their approach, the phase of the Fourier descriptors is employed and claimed to outperform the state-of-art Fourier-based methods. At first, a shape boundary is extracted to obtain a complex discrete time periodic signal. Then, the Discrete Fourier Transform (DFT) is computed to reduce dimension, only the first N low-frequency coefficients are retained and normalized in terms of translation, scale, rotation and further starting point selection. For matching, the inverse discrete Fourier transform is used to obtain normalized versions of the original contours in the spatial domain. Finally, a dynamic programming (DP) method is employed to find the similarity between the two transformed contours. Although this technique outperformed other Fourier-based methods, the authors reported less retrieval accuracy than the CSS method [23]. Besides, this WARP is not invariant to general affine transformation.

2.3.7 Inner Distance

H. Ling and D.W. Jacobs [31] have been motivated to build new shape descriptor (called inner-distance) which are invariant to articulation and capture part structure. The inner-distance is computed as a length of the shortest path between landmark points within the shapes silhouette. The authors have claimed that this representation is articulation insensitive and more effective at capturing part structures than the well-known Euclidean distance. They have further stated that it can be used as a replacement for the Euclidian distance to build more accurate descriptors for complex shapes, especially for those with articulated parts. In order to apply the inner-distance for shape matching and retrieval, the authors have extended the inner-distance by combining with two main features. The first feature is multidimensional scaling (MDS), to build articulation invariant signatures for articulation shapes. The other one is the shape context (SC) method [7] to build another new descriptor called it inner-distance shape context (IDSC). For matching, the dynamic programming (DP) has been used. The retrieval performance on the MPEG-7 dataset is 84.35% of MDS+SC+DP method and 85.40% of IDSC+DP method. The main disadvantage of this method is the use of different parameter settings for different databases. This method has not been tested on geometrically transformed shapes.

2.3.8 Curve Edit Distance

T.B. Sebastain et al. have proposed a curve alignment approach, which is called curve edit distance (CED), for matching open and closed curves [32]. This approach has been employed to find the optimal correspondence or alignment between 2D curves relies on using the intrinsic properties of the curves. In their method, the correspondence between the points of the two curves is controlled by the relative difference in their spatial location and their curvature. Then, a matching function is defined as the minimum cost of such correspondence. The search for the optimal correspondence is efficiently made by decomposing each curve into segments, and then ideally solved using dynamic programming (DP). Moreover, merging, deletion and addition of curve segments are allowed in order to account for shape deformations. The complexity of this method is $O(N^2 \log N)$ for closed curves and the authors reported 78.17% accuracy for MPEG-7 retrieval test.

2.3.9 Visual Parts

In another research work, L.J. Latecki and R. Lakamper [33] have proposed a shape matching approach which works directly on the closed boundaries. It is based on visual parts (VP), where (part of) a database shape is simplified in the context of the query shape prior to their matching. The simplification process includes the elimination of particular points from the database shape such that the similarity to the query shape is maximized. The main disadvantage of this method is the high computational complexity of the matching algorithm, which is $O(N^3 \log N)$, where N is the number of the boundary points.

2.3.10 Beam Angle Statistics

N. Arica and F.T. Yarman-Vural have proposed another descriptor for closed contours based on the convexity and concavity information of all boundary points, which called Beam Angle Statistics (BAS) [34]. In BAS, the convexity and concavity at each boundary point are viewed as a random variable that draws its values from the angles between each equally-distant neighboring point at that point. Then, few order moments are computed for the random variable at each point. After getting the shape representation, further compression is realized by using the Spectral Coefficients of the representation. The author also has claimed that this approach can eliminate the use of a heuristic rule or empirical threshold value in representation of shape boundaries in a predefined resolution scale. For measuring the similarity, dynamic programming (DP) is used to find the best correspondence that minimizes the Euclidian distance between the signatures of two shapes. The authors have reported 82.4% of accuracy using the MPEG-7 retrieval test.

2.3.11 Symbolic Representation

A recent work, a new and different approach to object recognition and retrieval, using a string of symbol for the possibility of representing a shape (representation of computing a string of symbol of shape is shown in Figure 2.11), has been proposed by M.R. Daliri and V. Torre [35]. There are several steps of this algorithm. At first, a shape contour is recovered and mapped into a pair of N points. To reduce the cost of matching among all the points from two shapes, the shape context (SC) is considered. Then, dynamic programming (DP) is observed to obtain the best matching (correspondence) between the point sets of the two shapes. In this process, DP is not only recovers the best matching between points, but also identifies outliers (M points), that is, points in two shapes which cannot be properly matched. The authors also have stated that DP is essential for identifying partial occlusions in two shapes. After that, two contours are aligned using Procrustes analysis [36]. After alignment, each contour is transformed into a string of symbols and a modified version of edit distance is used to compute the similarity between strings of symbols. Finally, recognition and retrieval are obtained by simple nearest neighbor procedure. The authors have reported the retrieval efficiency of 83.45% ($M=5$) and 85.92% ($M=3,5,7,9,11$) for the MPEG-7 retrieval test. Regarding the computational complexity, its matching complexity is $O(N^2)$, where N is the number of the boundary points.

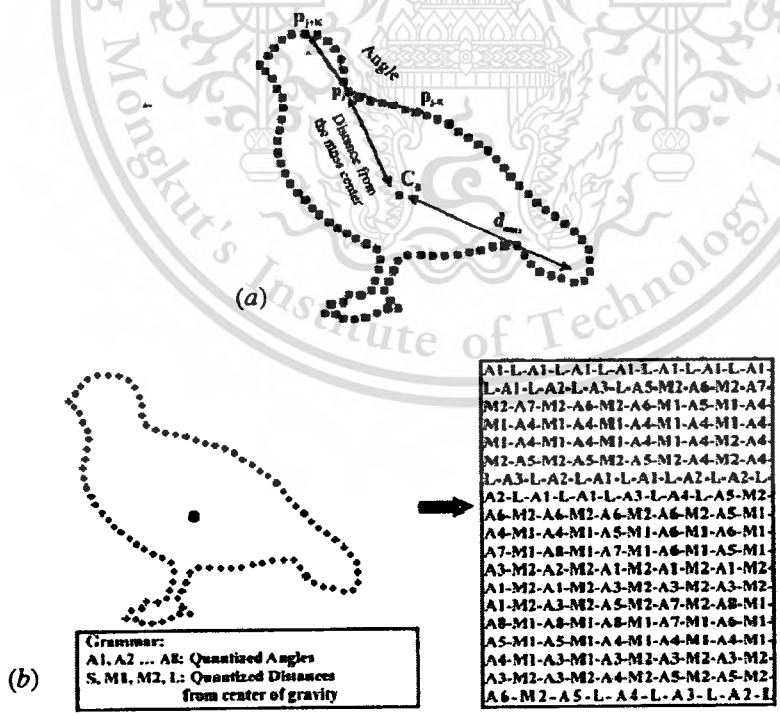


Figure 2.11: (a) A schematic of how to transform the shape into the symbolic representation (b) The mapping obtained by the algorithm of a given contour into a string of symbols.

2.3.12 Shape Tree

A recent approach for matching shapes based on a hierarchical description of their boundaries has been introduced by P.F. Felzenszwalb and J.D. Schwartz [37]. This approach can be used both for determining the similarity between two curves including open and close curves. First, the curves are broken into two halves at their corresponding midpoint and then represented by binary tree (arc-tree [38]), called shape tree in this literature. This representation can be thought of as a modified arc-tree in which the only information kept at each node is the relative position of the selected midpoint. Finally, elastic matching techniques are utilized in the matching process which can take global geometric information into account. The authors have further proven that the hierarchical shape tree representation can capture geometric properties at different levels of resolution. At the finest level these properties are related to standard local descriptions (capturing local curvature, for example). At coarser levels, the properties capture global shape aspects. For the MPEG-7 Core Experiment CE-shape-1 part B retrieval test, this method achieves 87.70% of retrieval efficiency. Regarding, the computational complexity, the matching complexity of this algorithm is $O\left(N^3K^3\right)$, where N is the number of the boundary points and K is the number of end-points.

Chapter 3

Feature Extraction

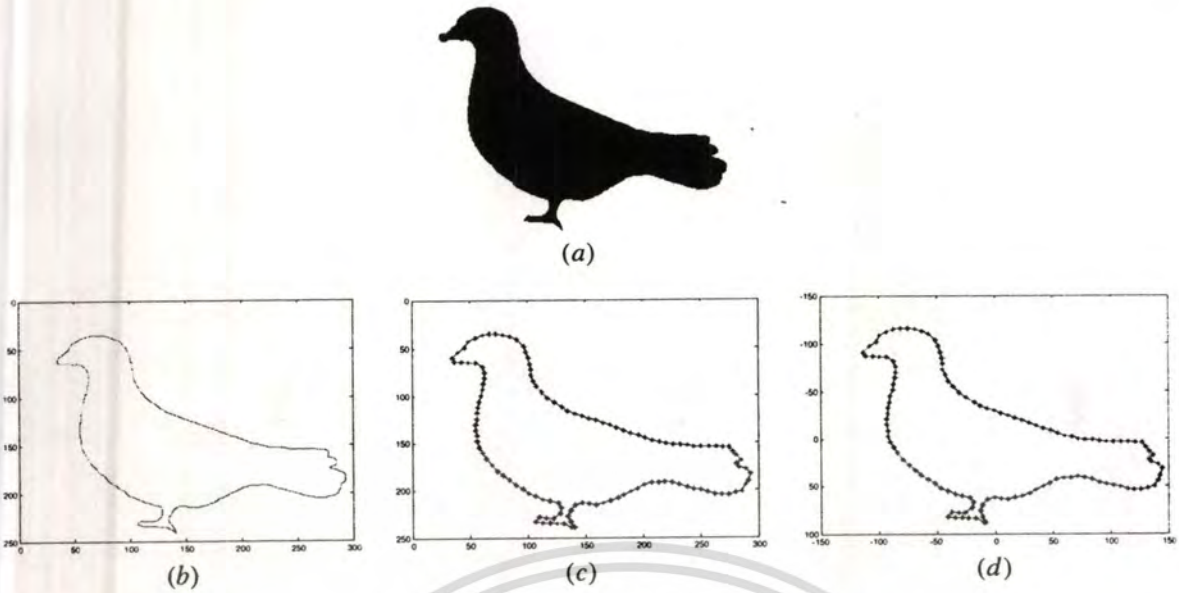
As mentioned in chapter 2, several techniques have been examined to extract the feature from the shape. A good shape representation should be robust to translation, scale, rotation, and shear which are all the parameters of affine transform. In this study, a new and simple technique, called Barycenter Contour (BcC) decomposition, is proposed to decompose the shape boundary of object into the multi-scale contours.

In order to compute the multi-scale shape representation, the features are extracted from the segmented shapes by performing three steps as follows: the preprocessing process is described in section 3.1. Then, it is followed by the BcC decomposition in section 3.2. The area of the triangle computed from the re-sampled points of the shape boundary is employed as the shape representation in section 3.3. Importantly, the triangle's area used as the shape representation is only invariant to affine transform. Although, they are not invariant to the selected starting point which can affect to the performance of the classification. Thus, this problem can be achieved using Discrete Fourier Transform (DFT) as discussed in section 3.4.

3.1 Preprocessing

To be able to classify 2D object silhouettes using contour-based technique, the pre-processing (as shown in Figure 3.1) is the first step in the classification process. In this work, the binary image is employed for testing our proposed algorithm. Firstly, the shape boundary is extracted from 2D binary object, using the conventional technique [2, 12], by starting from a selected point on the shape's boundary and moving in any direction (clockwise or counterclockwise). However, the counterclockwise direction is selected in this work. Each pair of the coordinate points of the contour are obtained and parameterized in term of $x(k)$ and $y(k)$. Then, they are re-sampled to have the same length. The appropriated number of the re-sampled points in the experiment is 128 points and denoted as (x_i, y_i), where $i = (1, 2, 3, ..., N)$. Finally, the shape boundary is shifted about its centroid (x_c, y_c) to eliminate the translation property by

$$\begin{cases} \hat{x}_i = x_i - x_c \\ \hat{y}_i = y_i - y_c \end{cases} \tag{3.1}$$



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

Figure 3.1: The pre-processing process (a) binary 2D shape silhouette (b) extracted contour (c) re-sampling point and (d) shifted contour.

where $(\hat{x}_i, \hat{y}_i)$ is a pair of coordinate of the shifted point, $(x_c, y_c) = (\mu_{10}, \mu_{01})$ is the centroid coordinate of the shape contour. The moment [39] of the shape contour is defined

$$\mu_{pq} = E\{x_i^p y_i^q\} \qquad \text{for } p,q = 0,1,2,3..... \tag{3.2}$$

For a digital shape, the moment can be written $\mu_{pq} = E\{x_i^p y_i^q\} = \frac{1}{N}\sum_{i=1}^{N} x_i^p y_i^q$. Thus, the centroid coordinate of the shape contour can be computed again by

$$(x_c, y_c) = \frac{1}{N}\left(\sum_{i=1}^{N} x_i, \sum_{i=1}^{N} y_i\right) \tag{3.3}$$

3.2 Barycenter contour decomposition

In geometry, Barycenter of a set of coordinate points is the average of this set of coordinate points (Appendix A). The proposed technique, BcC decomposition has been utilized to decompose shape contour into various scale levels of BcC [40-42] in which the first level has been adopted from [43] and it is constructed from the coordinates of Barycenter of a triangle consisting of $(0,0), (\hat{x}_i, \hat{y}_i)$, and $(\hat{x}_{i+1}, \hat{y}_{i+1})$ as depicted in Figure 3.2 and defined by

$$BcC(i,1) = \begin{cases} \tilde{x}_i^1 = \dfrac{\hat{x}_i + \hat{x}_{i+1}}{3} \\ \tilde{y}_i^1 = \dfrac{\hat{y}_i + \hat{y}_{i+1}}{3} \end{cases} \tag{3.4}$$

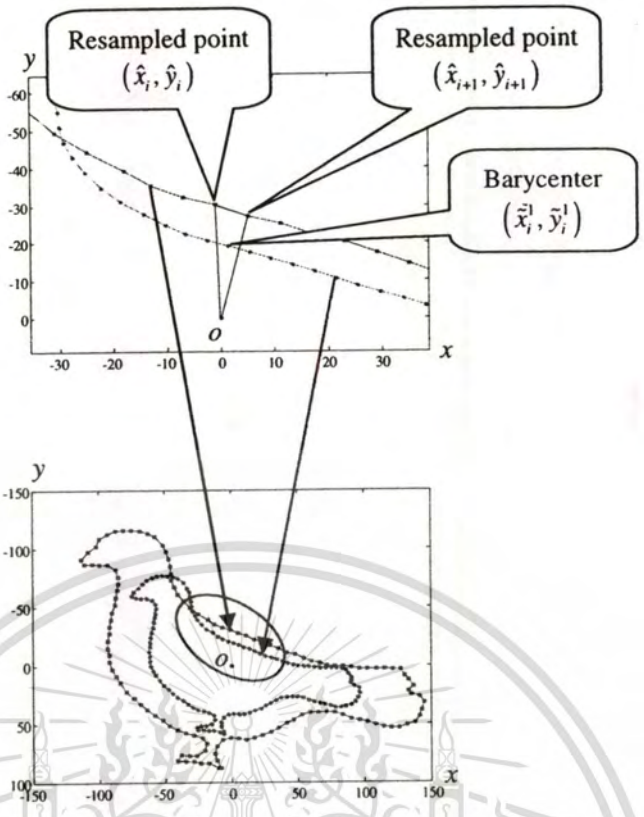


Figure 3.2: Barycenter contour of bird at the first level

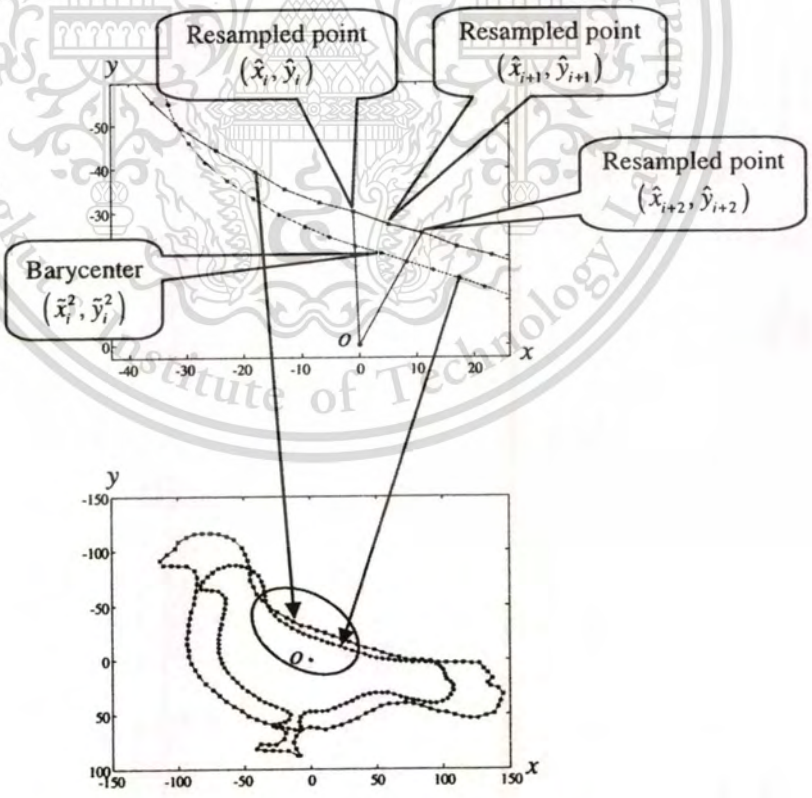


Figure 3.3: Barycenter contour of bird at the second level

The second level is constructed from the coordinates of Barycenter of points $(0,0)$, $(\hat{x}_i, \hat{y}_i)$, $(\hat{x}_{i+1}, \hat{y}_{i+1})$, and $(\hat{x}_{i+2}, \hat{y}_{i+2})$ as depicted in Figure 3.3 and computed by

$$BcC(i,2) = \begin{cases} \tilde{x}_i^2 = \dfrac{\hat{x}_i + \hat{x}_{i+1} + \hat{x}_{i+2}}{4} \\ \tilde{y}_i^2 = \dfrac{\hat{y}_i + \hat{y}_{i+1} + \hat{y}_{i+2}}{4} \end{cases} \tag{3.5}$$

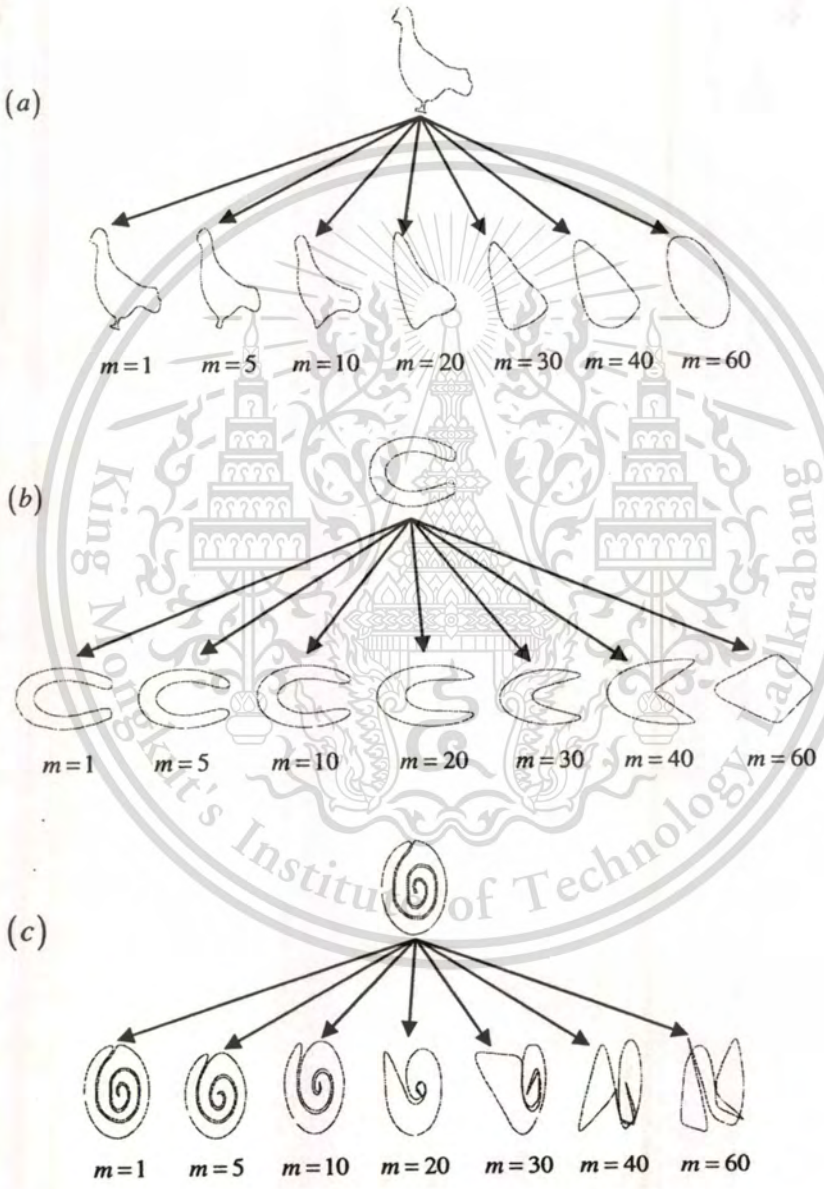


Figure 3.4: Example of the Barycenter contour decomposition of (a) bird's; (b) horseshoe's and (c) device9's contours at different scale level 1,5,10,20,30,40, and 60.

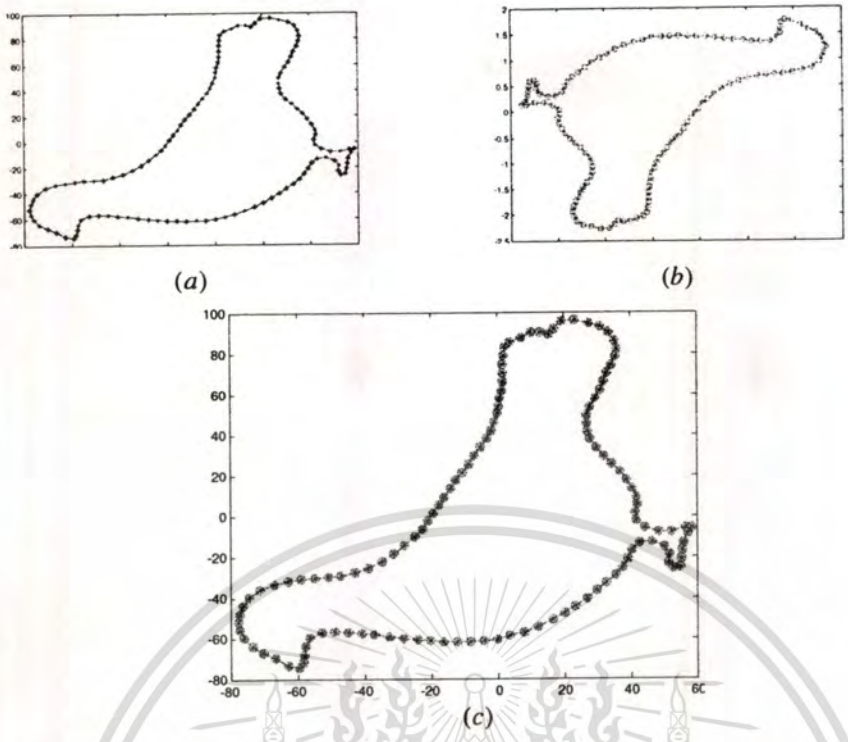


Figure 3.5: The Barycenter contour of bird (a) at level 1; (b) at level 125; (c) both levels are normalized in both horizontal and vertical axe in which level 125 is plotted in reversion direction

The next level is created by adding one adjacent point more to the previous level. Mathematically, if $(\tilde{x}_i, \tilde{y}_i)$ is a pair of the coordinates of Barycenter contour where $i = 1, 2, ..., N$, it is computed as:

$$BcC(i,m) = \begin{cases} \tilde{x}_i^m = \dfrac{\hat{x}_i + \hat{x}_{i+1} + \hat{x}_{i+2} + ... + \hat{x}_{i+m}}{m+2} \\ \tilde{y}_i^m = \dfrac{\hat{y}_i + \hat{y}_{i+1} + \hat{y}_{i+2} + ... + \hat{y}_{i+m}}{m+2} \end{cases} \tag{3.6}$$

where $(\hat{x}_i, \hat{y}_i)$ is a pair of coordinate of shifted point. $(\tilde{x}_i, \tilde{y}_i)$ is a pair of coordinate of Barycenter point. m is the level of Barycenter contour taking values from 1 until 63.

Figure 3.4 illustrates an example of Barycenter contour decomposition of bird's and device9's contours at various scale levels 1, 5, 10, 20, 30, 40 and 60. When the level of BcC is increased, it makes the contour of bird became smoother (circle or ellipse form). However, for a deep concavity and a complex shape such as horseshoe and device9, their boundaries cannot be smoothed until becoming circle or ellipse form (smoothed only at low level). In conclusion, this technique can reduce the noise caused by the distortion and deformation of shape's boundary and but limits for the complex shape. Moreover, if the shape is subjected to affine transformation, then the Barycenter contour of that shape contour at various scale levels will be affected by the same transformation as proven below.

By modeling Eq. (3.6), thus, the choice of the number of scale level of BcC is constrained by the implied periodicity of the closed boundary and the scale coefficient. More specifically, for a closed contour of N points:

$$BcC(i,m)=\begin{cases}-\dfrac{N-m}{m+2}\cdot BcC(i+m+1,N-2-m) & m=1,\ldots\ldots,\left\lfloor\dfrac{N-2}{2}\right\rfloor\\ \dfrac{1}{N}(\hat{x}_{i+N-1},\hat{y}_{i+N-1}) & m=N-2\\ 0 & m=N-1\\ \dfrac{1}{N+2}(\hat{x}_i,\hat{y}_i) & m=N\end{cases} \tag{3.7}$$

where $\left\lfloor\dfrac{N-2}{2}\right\rfloor$ is the limit level of BcC. The first line in Eq. (3.7) shows the symmetry property including the scale version of BcC versus the level of BcC m. Also, at $m=N-1$, there is only one point of BcC, i.e, at point $(0,0)$. For example, for $m=1$, we have got $BcC(i,1)=-\dfrac{127}{3}BcC(i+2,125)$. Figure 3.5 illustrates the BcC of bird at scale level 1 in (a) and level 125 in (b). From this figure, it shows that level 125 is a reverse and scale version of level 1.

Affine transformation: The general affine transform is mathematically defined by:

$$\begin{cases}x_a=ax+by+e\\ y_a=cx+dy+f\end{cases} \tag{3.8}$$

From Eq. (3.6), it can be written in a matrix form in the Euclidian space as follows:

$$\begin{bmatrix}x_a\\ y_a\\ 1\end{bmatrix}=\begin{bmatrix}a & b & e\\ c & d & f\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x\\ y\\ 1\end{bmatrix} \tag{3.9}$$

where (x,y) are the pixel coordinates of a given image, and (x_a,y_a) are the pixel coordinates of the distorted version of the given image under general affine transformation. Translation is represented by e and f, while scale, rotation, and shear are reflected in the remaining four parameters a,b,c,d.

The translation, scale, rotation and shear parameters are presented by the following matrix:

$$\mathrm{A}_{\text{translation}}=\begin{bmatrix}1 & 0 & t_x\\ 0 & 1 & t_y\\ 0 & 0 & 1\end{bmatrix},\ \mathrm{A}_{\text{scale}}=\begin{bmatrix}s_x & 0 & 0\\ 0 & s_y & 0\\ 0 & 0 & 1\end{bmatrix}$$

$$\mathrm{A}_{\text{rotation}}=\begin{bmatrix}\cos(\theta) & -\sin(\theta) & 0\\ \sin(\theta) & \cos(\theta) & 0\\ 0 & 0 & 1\end{bmatrix}\text{ and }\mathrm{A}_{\text{shear}}=\begin{bmatrix}1 & sh_x & 0\\ sh_y & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

where t_x, t_y, s_x, s_y, sh_x, and sh_y are the translation, scale and shear parameters by x and y direction respectively. θ is the angle parameter of rotation. All the parameters of affine transformation are represented in a matrix A_{aff} defined in [44] as follows:

$$\mathrm{A}_{aff} = \mathrm{A}_{\text{translation}} \cdot \mathrm{A}_{\text{scale}} \cdot \mathrm{A}_{\text{rotation}} \cdot \mathrm{A}_{\text{shear}} \tag{3.10}$$

However, the translation property can be removed by shifting the shape boundary about its center point. So the matrix A_{aff} is written again as:

$$\mathrm{A}_{aff} = \mathrm{A}_{\text{scale}} \cdot \mathrm{A}_{\text{rotation}} \cdot \mathrm{A}_{\text{shear}} \tag{3.11}$$

Property1: If $BcC_a(i,m) = \left(\tilde{x}^m_{a,i}, \tilde{y}^m_{a,i}\right)$ is the transformed version of a planar curve $BcC(i,m) = \left(\tilde{x}^m_i, \tilde{y}^m_i\right)$ under general affine transformation, there is a one-to-one correspondence between $BcC(i,m)$ and $BcC_a(i,m)$.

Proof: From Eq. (3.6), the affine version is written as follows:

$$\begin{bmatrix} \tilde{x}^m_{a,i} \\ \tilde{y}^m_{a,i} \\ 1 \end{bmatrix} = \frac{1}{m+2} \begin{bmatrix} \hat{x}_{a,i} & \hat{x}_{a,i+1} & \hat{x}_{a,i+2} & \cdots & \hat{x}_{a,i+m} \\ \hat{y}_{a,i} & \hat{y}_{a,i+1} & \hat{y}_{a,i+2} & \cdots & \hat{y}_{a,i+m} \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{3.12}$$

By Eq. (3.9), we can write

$$\begin{bmatrix} \hat{x}_{a,i} & \hat{x}_{a,i+1} & \hat{x}_{a,i+2} & \cdots & \hat{x}_{a,i+m} \\ \hat{y}_{a,i} & \hat{y}_{a,i+1} & \hat{y}_{a,i+2} & \cdots & \hat{y}_{a,i+m} \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} a & b & e \\ c & d & f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}_i & \hat{x}_{i+1} & \hat{x}_{i+2} & \cdots & \hat{x}_{i+m} \\ \hat{y}_i & \hat{y}_{i+1} & \hat{y}_{i+2} & \cdots & \hat{y}_{i+m} \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \tag{3.13}$$

Thus, Eq. (3.12) is written as

$$\begin{bmatrix} \tilde{x}^m_{a,i} \\ \tilde{y}^m_{a,i} \\ 1 \end{bmatrix} = \frac{1}{m+2} \begin{bmatrix} a & b & e \\ c & d & f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}_i & \hat{x}_{i+1} & \hat{x}_{i+2} & \cdots & \hat{x}_{i+m} \\ \hat{y}_i & \hat{y}_{i+1} & \hat{y}_{i+2} & \cdots & \hat{y}_{i+m} \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{3.14}$$

Then, we have

$$\begin{bmatrix} \tilde{x}^m_{a,i} \\ \tilde{y}^m_{a,i} \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & e \\ c & d & f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{x}^m_i \\ \tilde{y}^m_i \\ 1 \end{bmatrix} \tag{3.15}$$

By Eq. (3.15), we can conclude that BcC decomposition technique is affected by the same affine distortion when the shape is subjected by the affine transformation.

Property2: If (x_c, y_c) is the centroid of the shape contour, thus it is also the centroid of BcC at each scale level.

Proof: By Eq. (3.1) and Eq. (3.6), we have

$$\begin{cases} \tilde{x}_i^m = \dfrac{(x_i - x_c) + (x_{i+1} - x_c) + \ldots\ldots + (x_{i+m} - x_c) + x_c - x_c}{m+2} \\ \tilde{y}_i^m = \dfrac{(y_i - y_c) + (y_{i+1} - y_c) + \ldots\ldots + (y_{i+m} - y_c) + y_c - y_c}{m+2} \end{cases} \tag{3.16}$$

$$\begin{cases} \tilde{x}_i^m = \dfrac{x_i + x_{i+1} + \ldots\ldots + x_{i+m} + x_c}{m+2} - x_c \\ \tilde{y}_i^m = \dfrac{y_i + y_{i+1} + \ldots\ldots + y_{i+m} + y_c}{m+2} - y_c \end{cases} \tag{3.17}$$

Thus, by Eq. (3.17) the BcC at each scale level m of shape contour with the centroid (x_c, y_c) is

$$\begin{cases} \tilde{x}_i^m + x_c = \dfrac{x_i + x_{i+1} + \ldots\ldots + x_{i+m} + x_c}{m+2} \\ \tilde{y}_i^m + y_c = \dfrac{y_i + y_{i+1} + \ldots\ldots + y_{i+m} + y_c}{m+2} \end{cases} \tag{3.18}$$

The moment of BcC at each scale level m is

$$\begin{aligned} \mu_{pq}^m &= E\left\{ \left(\tilde{x}_i^m + x_c\right)^p \left(\tilde{y}_i^m + y_c\right)^q \right\} \\ &= \frac{1}{N} \sum_{i=1}^{N} \left(\tilde{x}_i^m + x_c\right)^p \left(\tilde{y}_i^m + y_c\right)^q \end{aligned} \tag{3.19}$$

The centroid on x-axis of BcC at each scale level m is computed as

$$\begin{aligned} \mu_{10}^m &= \frac{1}{N} \sum_{i=1}^{N} \left(\tilde{x}_i^m + x_c\right) \\ &= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i + x_{i+1} + x_{i+2} + \ldots\ldots + x_{i+m} + x_c}{m+2} \right) \\ &= \frac{1}{N(m+2)} \left(\sum_{i=1}^{N} x_i + \sum_{i=1}^{N} x_{i+1} + \ldots\ldots + \sum_{i=1}^{N} x_{i+m} + \sum_{i=1}^{N} x_c \right) \end{aligned} \tag{3.20}$$

Due to a periodicity of shape contour, we have

$$\frac{1}{N} \sum_{i=1}^{N} x_i = \frac{1}{N} \sum_{i=1}^{N} x_{i+1} = \ldots\ldots = \frac{1}{N} \sum_{i=1}^{N} x_{i+m} = \mu_{10} = x_c \tag{3.21}$$

Thus,

$$\mu_{10}^{m} = \frac{1}{(m+2)}\left(\underbrace{x_c + x_c + \ldots\ldots + x_c + x_c}_{m+2 \text{ terms}}\right)$$
$$\mu_{10}^{m} = x_c \tag{3.22}$$

The same for the centroid on y-axis of BcC at each scale level m, that is, $\mu_{01}^{m} = y_c$.

3.3 Shape representation

In order to make the algorithm invariant to affine distortion, the triangle area computed from shape boundary is utilized. In this thesis, two triangle area representations are carried out which are computed alternatively. They are the signed enclosed area and triangle area representation signatures.

3.3.1 Signed enclosed area signature

The signed enclosed area (SEA) signature, introduced by D. Zhang and G. Lu [4], called Triangle Area Representation with two points (TAR-2p) in [40], is computed as the area of the triangle formed by two adjacent points on the shape's boundary and its center of gravity. Let any two re-sampled points $(\tilde{x}_i, \tilde{y}_i)$, $(\tilde{x}_{i+1}, \tilde{y}_{i+1})$ and point $(0,0)$, be the vertices of the triangle, the area at each re-sampled point location is computed as

$$SEA(i,m) = \frac{1}{2}\left(\tilde{x}_i^m . \tilde{y}_{i+1}^m - \tilde{x}_{i+1}^m . \tilde{y}_i^m\right) \tag{3.23}$$

Figure 3.6 shows the signed enclosed area signature at some levels of BcC. These representations are further plotted in 3D format as shown in Figure 3.7.

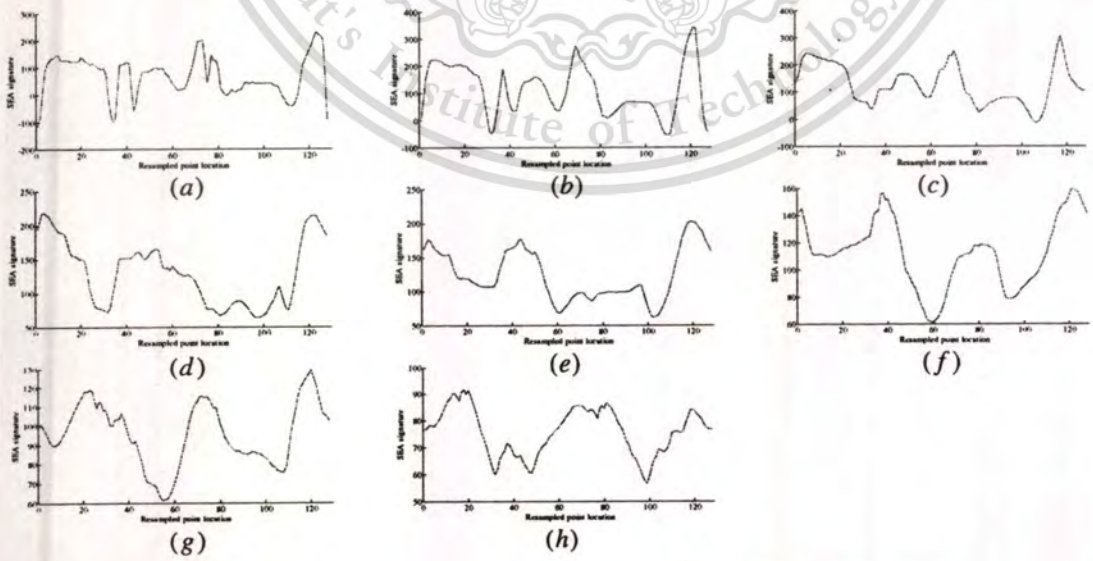


Figure 3.6: The SEA signature of bird in Figure3.1 at level 1, 5, 10, 20, 30, 40, 50, and 60 in (a), (b), (c), (d), (e), (f), (g), and (h) respectively.

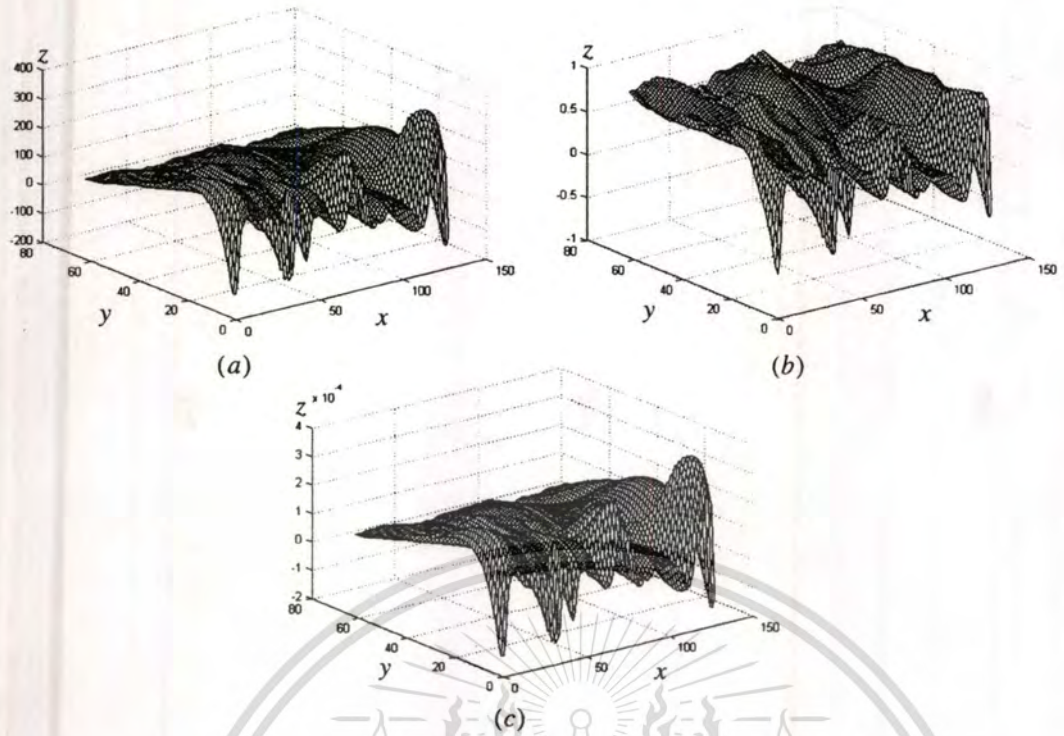


Figure 3.7: The SEA of bird's shape in Figure3.1 plotted in 3D format of (a) non normalization; (b) local normalization; (c) global normalization. x is the re-sampled point location; y is the level of Barycenter contour; z is the value of SEA.

3.3.2 Triangle area representation signature

The triangle area representation (TAR) signature is computed as the area of triangle formed by any three consecutive sample points on the shape's boundary, i.e., $(\tilde{x}_{i-l}, \tilde{y}_{i-l})$, $(\tilde{x}_i, \tilde{y}_i)$, and $(\tilde{x}_{i+l}, \tilde{y}_{i+l})$ and it is defined by:

$$TAR(i,l) = \frac{1}{2}\begin{vmatrix} \tilde{x}_{i-l} & \tilde{y}_{i-l} & 1 \\ \tilde{x}_i & \tilde{y}_i & 1 \\ \tilde{x}_{i+l} & \tilde{y}_{i+l} & 1 \end{vmatrix} \tag{3.24}$$

where $l = 1, 2, ..., \left\lfloor \frac{N-1}{2} \right\rfloor$ is the number of triangle side length. $\left\lfloor \frac{N-1}{2} \right\rfloor$ is the maximum level of TAR. The TAR signatures at all triangle side length are computed from the shifted contour by Eq. (3.24) (TARs at some triangle side length are shown in Figure 3.8 and also plotted in 3D format in Figure 3.9). However, only TAR at the triangle side length equal to 1 (called Triangle Area Representation with three points (TAR-3p) in [40] and also 1-D triangle area representation in [26, 45]) is computed from the points of BcC in this thesis.

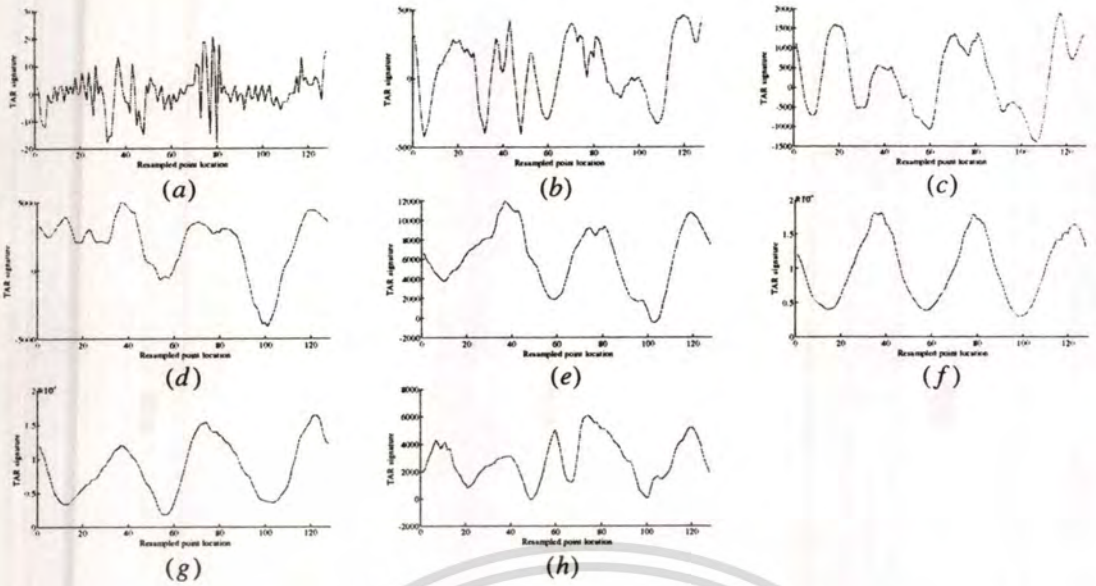


Figure 3.8: The TAR signature of bird in Figure3.1 at triangle side length 1, 5, 10, 20, 30, 40, 50, and 60 in (a), (b), (c), (d), (e), (f), (g), and (h) respectively.

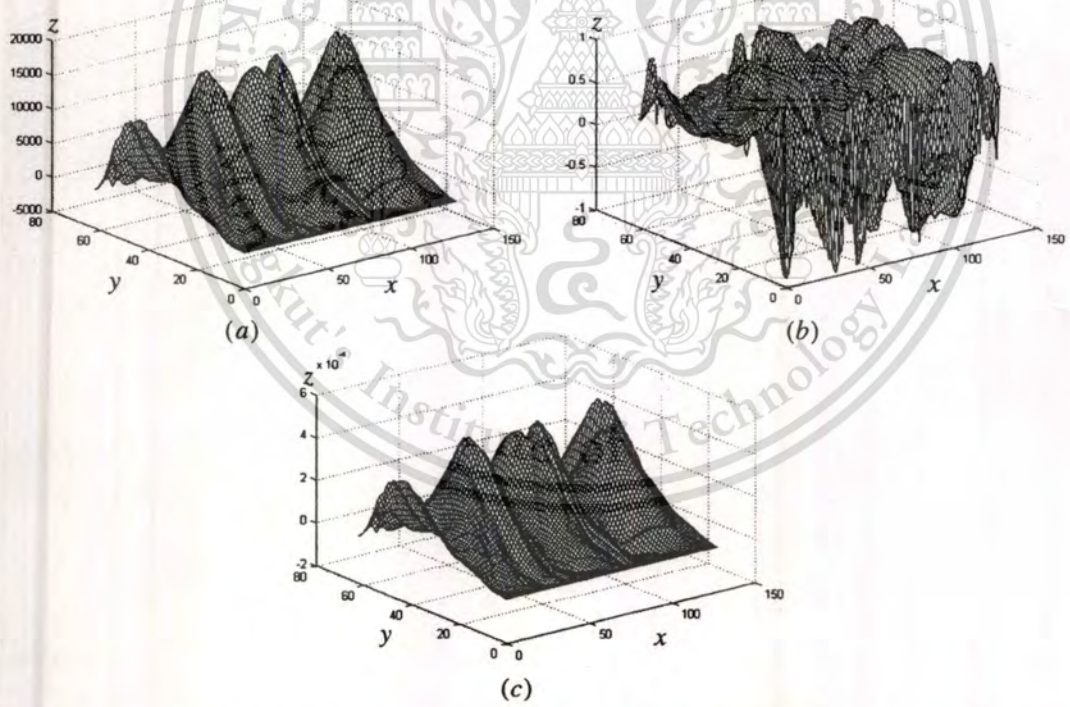


Figure 3.9: TAR at all the triangle side length of bird's shape in Figure3.1 plotted in 3D format of (a) non normalization; (b) local normalization; (c) global normalization. *x* is the resampled point location; *y* is the triangle side length; *z* is the value of that triangle area.

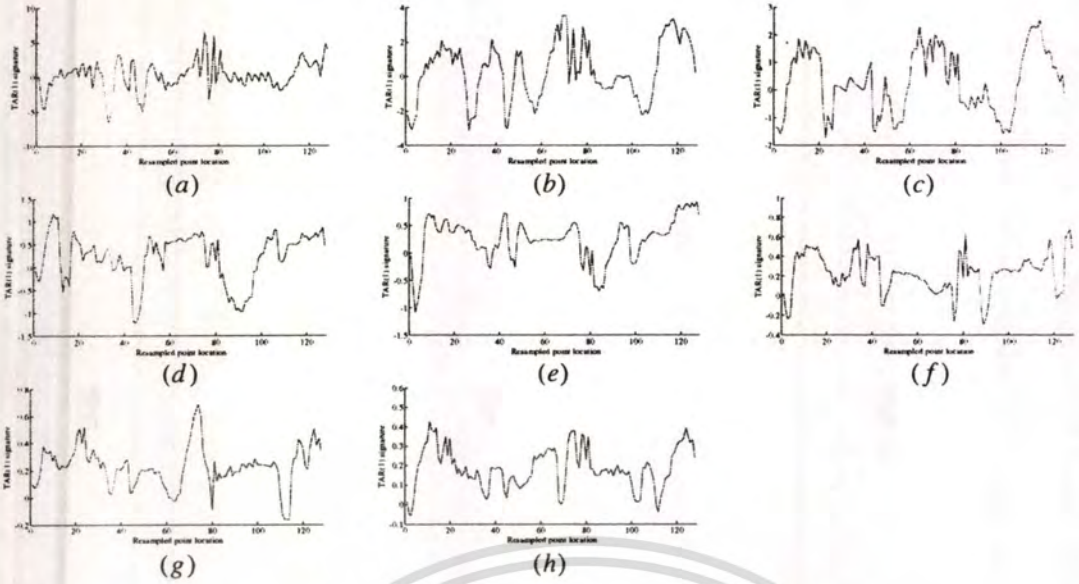


Figure 3.10: The TAR at TSL = 1 of bird in Figure3.1 at level 1, 5, 10, 20, 30, 40, 50, and 60 in (a), (b), (c), (d), (e), (f), (g), and (h) respectively.

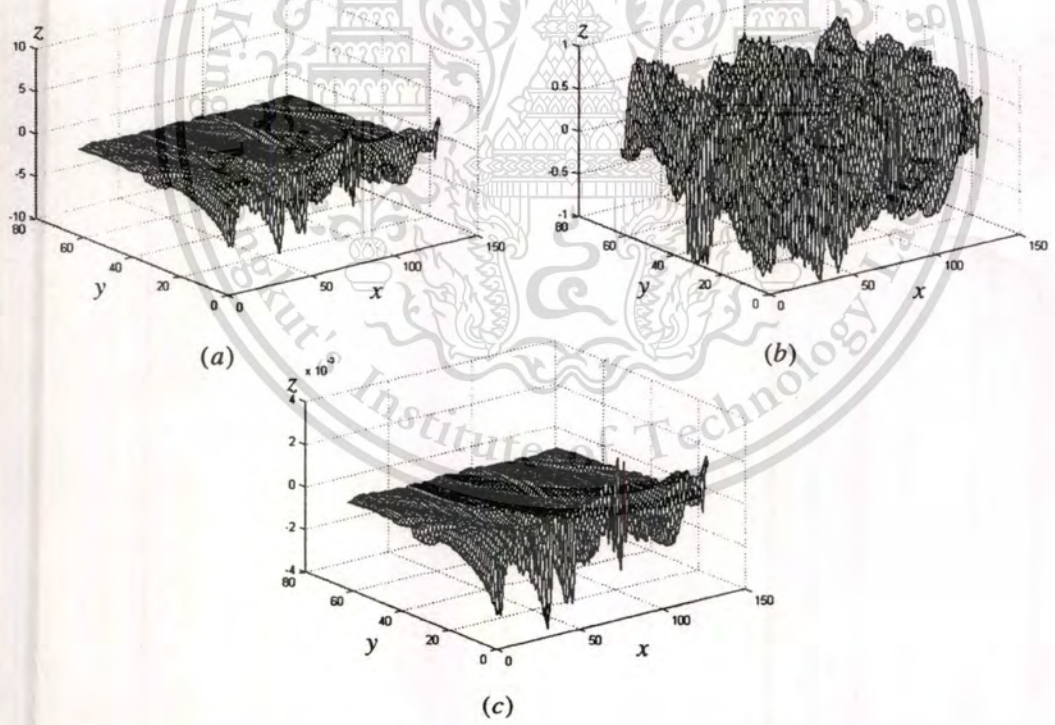


Figure 3.11: The TAR at TSL = 1 of bird's shape in Figure3.1 plotted in 3D format of (a) non normalization; (b) local normalization; (c) global normalization. x is the resampled point location; y is the level of Barycenter contour; z is the value of TAR.

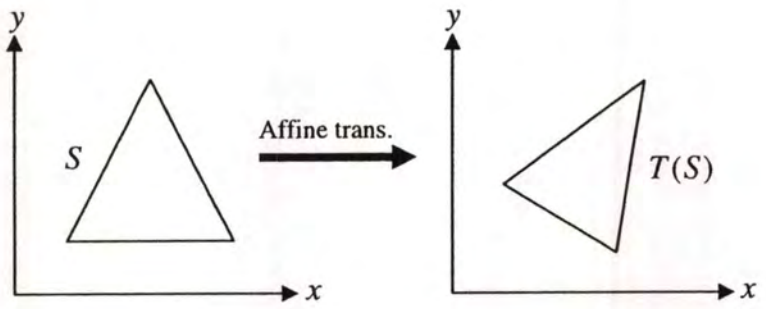


Figure 3.12: An affine transformation of a triangle *S* into another triangle *T(S)*

Thus, this triangle area representation at each level of BcC is computed by:

$$TAR(i,m) = \frac{1}{2}\begin{vmatrix} \tilde{x}_{i-1}^{m} & \tilde{y}_{i-1}^{m} & 1 \\ \tilde{x}_{i}^{m} & \tilde{y}_{i}^{m} & 1 \\ \tilde{x}_{i+1}^{m} & \tilde{y}_{i+1}^{m} & 1 \end{vmatrix} \tag{3.25}$$

Figure 3.10 illustrates the triangle area representation at triangle side length equal to one at some levels of Barycenter contour, i.e., level 1, 5, 10, 20, 30, 40, 50 and 60. The plot of TAR at all levels of BcC is plotted as 3D format in Figure 3.11.

From Figure 3.12, *S* is represented as a triangle and *T(S)* is represented as the affine transformation version of the triangle *S*. Under a general affine transformation, based on the determinant theory as stated in [45], the relation between both triangles' areas is expressed as:

$$\{\text{area of } T(S)\} = |J|.\{\text{area of } S\} \tag{3.26}$$

where J is the Jacobin between $\{x_a, y_a\}$ and $\{x, y\}$ derived from Eq. (3.6). This Jacobin is calculated as:

$$J = \begin{vmatrix} \frac{\partial x_a}{\partial x} & \frac{\partial x_a}{\partial y} \\ \frac{\partial y_a}{\partial x} & \frac{\partial y_a}{\partial y} \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = (ad - bc) \tag{3.27}$$

Thus,

$$\{\text{area of } T(S)\} = |ad - bc|.\{\text{area of } S\} \tag{3.28}$$

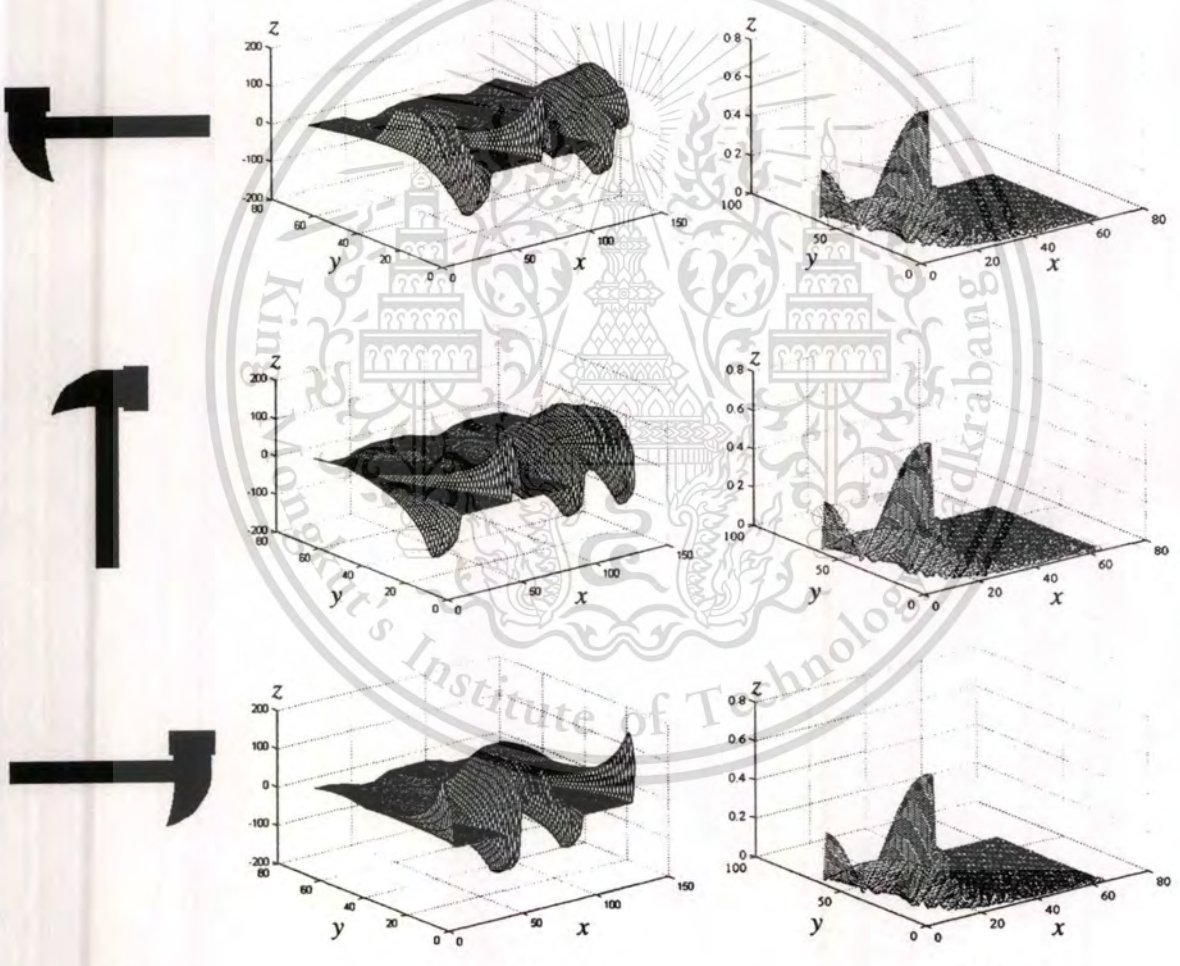
For this reason, SEA and TAR are invariant to affine transformation (translation, scaling, rotation, and shear).

3.4 Invariant to starting point selection

A main problem in getting a good result in shape classification, using contour-based method, is how to make the shape representation invariant to translation, scale, rotation, shear, reflection and starting point selection.

In this thesis, the Barycenter Contour (BcC) decomposition technique and the areas of the triangles (SEA and TAR) are utilized to extract the shape features. As proven above, the BcC technique and the shape signatures are invariant to affine transforms; its parameters are translation, scale, rotation and shear except reflection and starting point selection. These two exceptions will affect to the performance of classification. In this case, to guarantee all parameters' invariance, the Discrete Fourier Transform (DFT) is utilized to resolve this problem. Two ways of matching are considered to be invariant to the starting point selection and the reflection property by using DFT, i.e., matching in a spectral domain and the other one, matching in a spatial domain.

3.4.1 Spectral domain



(*a*) (*b*) (*c*)

Figure 3.13: (a) The shape silhouette, its 90^0 rotation and its reflection transformation; (b) the corresponding of SEA at all levels of BcC as 3D plot; (c) DFT coefficients of SEA at all levels of BcC as 3D plot. x is the resampled point location; y is the level of BcC; and z is the value of SEA and its DFT coefficients, respectively.

The sequence of N complex numbers $f(1), f(2), \ldots\ldots, f(N)$ is transformed into the sequence of N coefficients $F(1), F(2), \ldots\ldots, F(N)$ by the DFT according to the formula:

$$F(n) = \sum_{k=1}^{N} f(k) \exp\left(-j\frac{2\pi kn}{N}\right) \qquad n = 1, 2, 3, \ldots\ldots, N \tag{3.29}$$

The inverse discrete Fourier transform (IDFT) is given by:

$$f(k) = \frac{1}{N}\sum_{n=1}^{N} F(n) \exp\left(j\frac{2\pi kn}{N}\right) \qquad k = 1, 2, 3, \ldots\ldots, N \tag{3.30}$$

In frequency domain, only the DFT coefficients of the shape signatures (SEA and TAR) are utilized in the classification process as described in chapter 4. Capability of invariant to starting point selection and refection transform of hammer's shape is proven by the DFT coefficients of SEA signature in Figure 3.13.

3.4.2 Spatial domain

In contrast to the section 3.4.1, the classification process takes place in spatial domain. The important thing to resolve the problem of invariant to starting in spatial domain is examined by the phase normalization (sometime called phase correlation) on the shape representation (SEA or TAR) at each scale level of BcC. This process is in the same way as the Fourier phase normalization [17] which has been used to eliminate the starting point dependency and to be invariant to translation, scale and rotation. This can be accomplished by transforming the shape representation (SEA or TAR) at each scale level of BcC into spectral domain by DFT, resulting its magnitude and phase. Then, phase of DFT is normalized in term of the invariant properties as shown in Table 3.1. We will get the normalized version of the original shape representations by transforming them back into spatial domain. The Table 3.1 shows how to normalize the DFT coefficients $\hat{F}(n) = \hat{R}_n \exp\left(j\hat{\theta}_n\right)$ corresponding to invariant requirement.

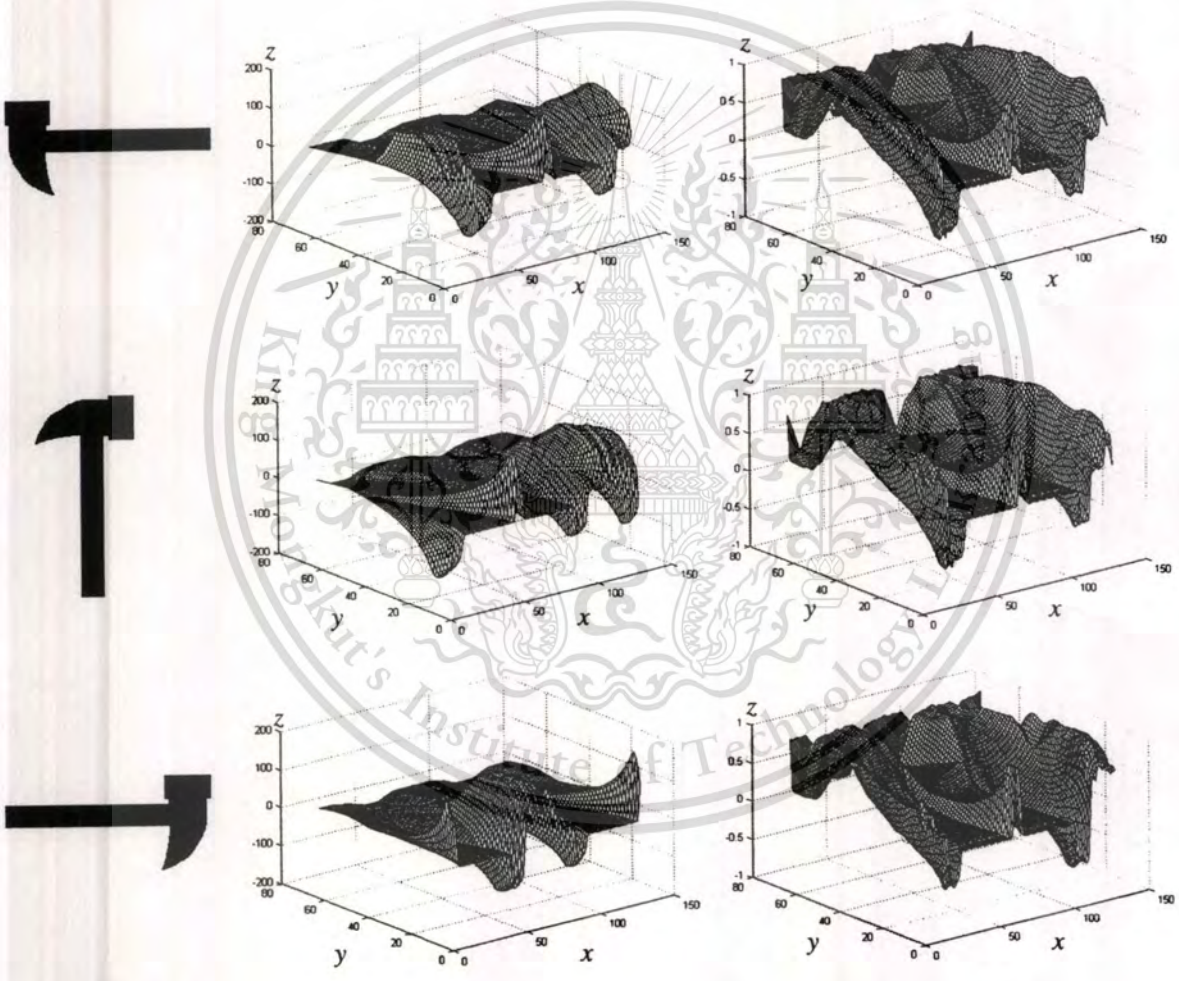
In Ref. [17], the proof of achieving both rotation and starting point invariance is presented below, whereas, the translation and scale invariance are obtained as in [46, 47]. Consider the origin-centered signals f and f', where f' is obtained from f by rotating each point counterclockwise by a constant factor $\bar{\theta}$ and by shifting the starting point by k_0 position, i.e., $f'(k) = f(k - k_0).\exp\left(j\bar{\theta}\right)$. The corresponding DFT coefficients are:

$$F(n) = \sum_{k=1}^{N} f(k) \exp\left(-j\frac{2\pi kn}{N}\right) = R_n \exp\left(j\theta_n\right) \tag{3.31}$$

$$F'(n) = F(n)\exp\left(j\bar{\theta}\right)\exp\left(-j\frac{2\pi k_0 n}{N}\right) = R_n \exp\left\{j\left(\theta_n + \bar{\theta} - \frac{2\pi k_0 n}{N}\right)\right\} = R_n \exp\left(j\theta'_n\right) \tag{3.32}$$

Thus, it is $\theta'_n = \theta_n + \bar{\theta} - \frac{2\pi k_0 n}{N}$, in particular $\theta'_1 = \theta_1 + \bar{\theta} - \frac{2\pi k_0}{N}$ and $\theta'_{-1} = \theta_{-1} + \bar{\theta} + \frac{2\pi k_0}{N}$. By referring to Table 3.1, to achieve the rotation and starting point invariance, both have to be integrated into:

$$\begin{aligned}\hat{\theta}'_n &= \theta'_n - \frac{\theta'_{-1} + \theta'_1}{2} + n\frac{\theta'_{-1} - \theta'_1}{2} \\ &= \theta_n + \bar{\theta} - \frac{2\pi k_0 n}{N} - \frac{\theta_{-1} + \theta_1}{2} - \bar{\theta} + n\frac{\theta_{-1} - \theta_1}{2} + \frac{2\pi k_0 n}{N} \\ &= \theta_n - \frac{\theta_{-1} + \theta_1}{2} + n\frac{\theta_{-1} - \theta_1}{2} \\ \hat{\theta}'_n &= \hat{\theta}_n\end{aligned} \tag{3.33}$$



(*a*) (*b*) (*c*)

Figure 3.14: (a) The shape silhouette, its 90^0 rotation and its reflection transformation; (b) the corresponding of SEA at various scale levels of BcC before phase normalization; (c) after phase normalization. *x* is the re-sampled point location, *y* is the level of BcC, and *z* is the value of SEA.

By simply performing the IDFT on DFT coefficients, then, a modified (normalized) signal $\hat{f}$ is obtained, which satisfies all the requested invariance:

$$\hat{f}(k) = \frac{1}{N} \sum_{n=-N/2}^{N/2-1} \hat{F}(n).\exp\left(j\frac{2\pi kn}{N} \right) \tag{3.34}$$

By following the phase normalization, the SEA signatures of hammers' shapes are transformed into spectral domain. Then, they are normalized to be invariant to selected starting point. Finally, they are transformed back into spatial domain as shown in Figure 3.14.

Table 3.1: The N-DFT WARP coefficients

Invariance	Modified coefficient
Translation	$\hat{f}_0 = 0$
Scale	$\hat{R}_n = R_n / R_1$
Rotation	$\hat{\theta}_n = \theta_n - \frac{\theta_{-1} + \theta_1}{2}$
Starting point selection	$\hat{\theta}_n = \theta_n + n\frac{\theta_{-1} - \theta_1}{2}$

Chapter 4

Classification

As described in chapter 3, shapes can be decomposed into various scale levels of BcC. Mathematically, the shape features are represented by 2D matrices in which each column corresponds to level of BcC and each row corresponds to the re-sampled point location on the shape contour. By using all levels in the classification process, the 2D matrices are transformed into a single column feature vectors by concatenating in the order of their columns. Owing to such feature vectors with a very high dimension subspace, in this chapter, we examine two classification techniques, Principal Component Analysis (PCA) [48, 49] and Fisher Discriminant Analysis (FDA). In this thesis, both methods are called Eigen Barycenter Contour (EBcC) and Fisher Barycenter Contour (FBcC), respectively.

4.1 Principal component analysis (PCA)

Principal Component Analysis (PCA) is a statistical technique. It is a well-known technique and also known as Karhunen-Loeve Transform, shortly KLT, which is frequently applied in face recognition system [50-52]. The common goal of PCA method is to approximate a set of data of high dimension into lower dimensional feature vectors by maximizing variance of all projected samples. Two important steps are considered in PCA method. The first step is a training step and another one is a testing step.

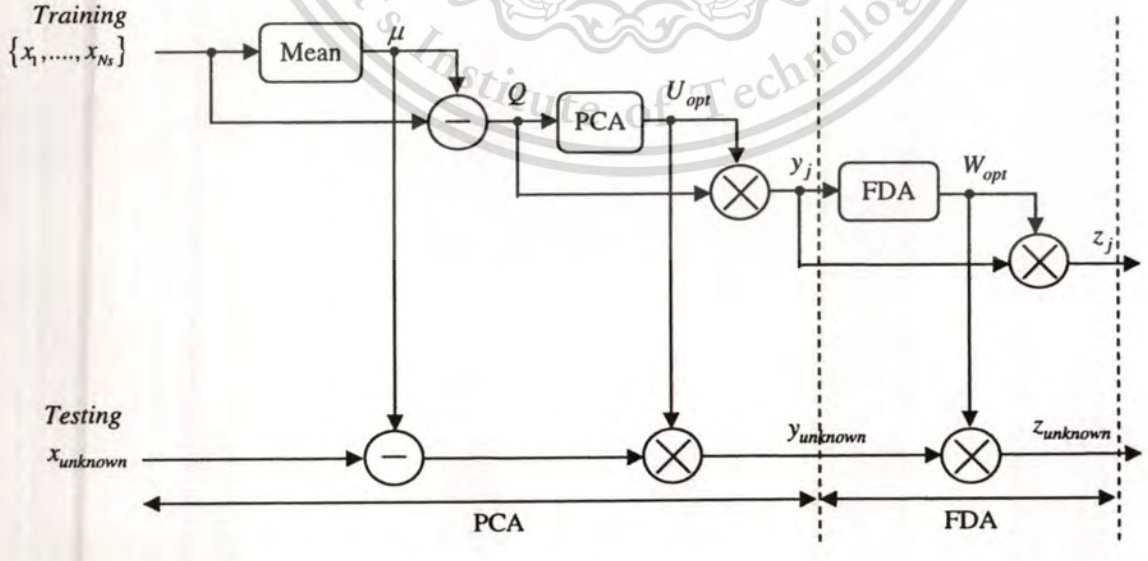


Figure 4.1: The computation of PCA and FDA including the training and the testing step.

4.1.1 Training

Let $\{x_1, x_2, x_3, ..., x_{Ns}\}$ be a set of *Ns* samples of shape features represented in column vector $(x_j \in \Re^n)$. For instance, the column vector x_j can be represented by the SEA, TAR, DFT or the power spectral coefficients of both signatures. Let us consider a linear transformation mapping the original *n*-dimensional shape space into an *r*-dimensional feature space, where $r < n$. The new feature vectors $y_j \in \Re^r$ are defined by the following linear transformation:

$$y_j = U^T(x_j - \mu) \quad , \quad j = 1, 2, 3, ..., Ns \tag{4.1}$$

where $U \in \Re^{n \times r}$ is a matrix with orthogonal column or a set of eigenvectors; μ is a mean of a set of *Ns* sample shape features.

The total scatter matrix (variance-covariance matrix) S_T is defined as:

$$S_T = \sum_{j=1}^{Ns} (x_j - \mu)(x_j - \mu)^T = QQ^T \tag{4.2}$$

Where $\mu = \frac{1}{Ns}\sum_{j=1}^{Ns} x_j$ is the mean of all sample shape features and $Q = [x_1 - \mu, x_2 - \mu, ..., x_{Ns} - \mu]$.

After applying the linear transformation U^T, the scatter of the transformed feature vectors $\{y_j \mid j = 1, 2, ..., Ns\}$ is $U^T S_T U$. In PCA, the objective is to find U_{opt} for maximizing the determinant of the total scatter matrix of the projected sample y_j, that is:

$$\begin{aligned} U_{opt} &= \arg\max_U \left| U^T S_T U \right| \\ U_{opt} &= [u_1 \; u_2 \; ..... \; u_r] \end{aligned} \tag{4.3}$$

where $\{u_k \mid k = 1, 2, ..., r\}$ is the set of *n*-dimensional eigenvectors of S_T corresponding to the *r* largest eigenvalues. There are two ways of finding the eigenvectors u_k. The first method is by directly calculating the eigenvectors of S_T and another method is by the singular value decomposition technique (SVD). Directly finding the eigenvectors from the total scatter matrix S_T is quite expensive computation and time-consuming in the process. Due to such an inconvenient, in this thesis, the SVD is applied instead. In MATLAB, we can find the eigenvector and eigenvalue by using command *eig( )* and *svd( )*. The results from SVD are the required eigenvectors $U = [u_1, u_2, u_3, ..., u_{Ns}]$ of scatter matrix S_T corresponding to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ..... \geq \lambda_{Ns}$ which are placed in decreasing order.

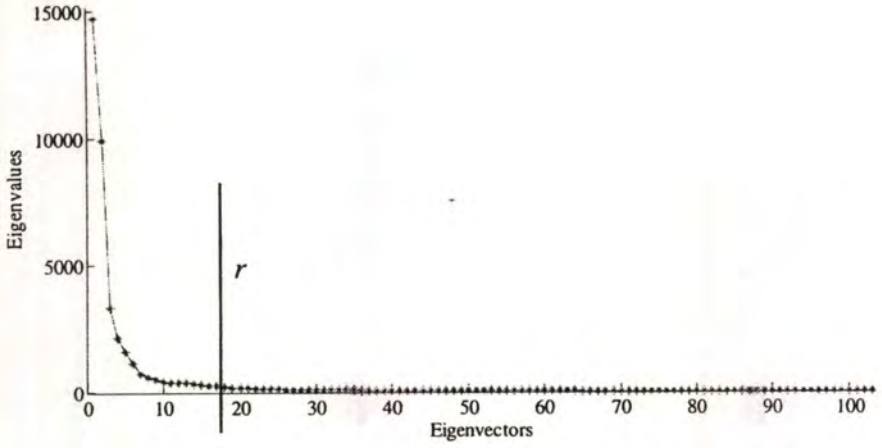


Figure 4.2: The scree plot for selecting the optimum eigenvector.

The optimum eigenvector U_{opt} (the first r principal components) is chosen corresponding to the r largest eigenvalues where the index value r is defined by a so-called scree plot as shown in Figure 4.2. A nice feature of principal components is immediately visible from such a plot. The first few principal components generally explain most of the variance (each eigenvalue is the variance of each corresponding component) that is present in the data, whereas the cumulative impact of the principal components is on the portion of the scree plot with a small slope added little explanation. This is the reason why principal components are usually an effective way of reducing the dimensionality of a feature set. We can use the first principal components (up to where contribution becomes negligible) in lieu of the whole data set. In the programming, the way of selecting the first r eigenvectors is based on a percent of variance in which the index value r is defined as the minimum index value below.

$$\frac{\lambda_1 + \lambda_2 + \lambda_3 + ... + \lambda_r}{\lambda_1 + \lambda_2 + ... + \lambda_r + ... + \lambda_{Ns}} \geq \alpha \tag{4.4}$$

where α is a threshold value which is greater than or equal to 0.90. Two value of the threshold value α are frequently used in the literatures — they are 0.90 and 0.95 [53]. In this thesis $\alpha = 0.95$ is taken into consideration.

4.1.2 Testing

In the testing process, it is not complicated as in the training process. Assume that $x_{unknown}$ is an unknown input. The projected feature $y_{unknown}$ of the unknown input (as shown in Figure4.1) is defined:

$$y_{unknown} = U_{opt}^{T}(x_{unknown} - \mu) \tag{4.5}$$

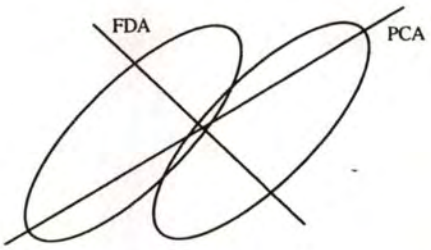


Figure 4.3: Subspace axes of PCA and FDA for data points of two classes.

4.2 Fisher discriminant analysis (FDA)

As clearly explained, PCA method is the dimensionality reduction method by a linear subspace of maximum variance for a set of shape representations. Another similar method for dimensionality reduction is Fisher Discriminant Analysis (FDA) [50, 52-55] and sometime called Linear Discriminant Analysis (LDA). When the shape representations are attached with class labels, it is often hoped that the projections of the shape representations onto the subspace are maximally separated into different classes. In that case, PCA cannot fulfill this purpose because it does not consider the class labels in estimating the subspace. The technique for learning subspace by maximizing separability is performed by FDA. Figure 4.3 shows an example of FDA as compared to PCA. In 2D space, there are two classes of Gaussian distributions; the 1D subspace learned by FDA has better separability than that learned by PCA.

4.2.1 Training

Let us consider a set of *Ns* sample shape features $\{x_1, x_2, x_3, ..., x_{Ns}\}$ represented in column vector $\left(x_j \in \Re^n\right)$ and assume that each shape feature belongs to one of c classes. $\{X_1, X_2, X_3, ..., X_c\}$. For instance, the column vector x_j can be represented by the SEA, TAR, DFT or the spectral coefficients of both signatures.

The objective of both PCA and FDA methods is to find the eigenvectors for optimizing dimensionality reduction. In FDA, however, the eigenvectors are found by the separated matrix derived from Fisher's linear discriminant function instead of the total scatter S_T (variance-covariance) matrix which is used in PCA. The separated matrices are the between-class scatter matrix S_B and the within-

class scatter matrix S_w. The FDA method selects the optimum projection W_{opt} of a set of eigenvector W in such a way that maximizes the Fisher's ratio.

Let the between-class scatter matrix be defined as:

$$S_B = \sum_{k=1}^{c} (\mu_k - \mu)(\mu_k - \mu)^T \tag{4.6}$$

and the within-class scatter matrix be defined as:

$$S_W = \sum_{k=1}^{c} \sum_{x_j \in X_k} (x_j - \mu_k)(x_j - \mu_k)^T \tag{4.7}$$

where μ_k are the mean shape representation of class X_k and μ is the mean of all sample shapes. If S_w is nonsingular, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples, that is

$$\begin{aligned} W_{opt} &= \arg\max_{W} \frac{|W^T S_B W|}{|W^T S_W W|} \\ W_{opt} &= \left[w_1 \;\; w_2 \;\; \ldots\ldots \;\; w_p \right] \end{aligned} \tag{4.8}$$

where $\{w_k \mid k = 1, 2, \ldots\ldots, p\}$ is the set of generalized eigenvectors of matrix $S_W^{-1} S_B$ corresponding to the p largest generalized eigenvalues $\{\lambda_k \mid k = 1, 2, \ldots\ldots, p\}$.

Due to very expansive computation of directly finding the eigenvector from $S_W^{-1} S_B$ matrix of too large dimensional space, a hybrid algorithm PCA and FDA is utilized instead of only FDA algorithm in classification process. In such algorithm, PCA is applied to reduce dimension first. Then, FDA is applied to improve a performance of classification by maximizing the between-class scatter matrix and minimizing the within-class scatter matrix.

After finding the eigenvectors of $S_{W(PCA)}^{-1} S_{B(PCA)}$, the first p eigenvectors are selected corresponding to the (c-1) largest eigenvalues of that matrix [53]. However, if $p < (c-1)$, all p eigenvector are selected. A new projected feature z_j of the given feature x_j is computed as follows:

$$z_j = W_{opt}^T U_{opt}^T (x_j - \mu) \tag{4.9}$$

4.2.2 Testing

Assume that $x_{unknown}$ is an unknown input. The projected feature $z_{unknown}$ of the unknown input (as shown in Figure4.1) is defined as:

$$z_{unknown} = W_{opt}^{T} U_{opt}^{T} \left(x_{unknown} - \mu \right) \tag{4.10}$$

4.3 Similarity measure

The similarity between the unknown shape (query shape) $s_{unknown}$ and all shapes in the database s_j is measured by (s can be represented by the projected features of PCA or the projected features of FDA):

- The Euclidian distance is defined as

$$d\left(s_{unknown}, s_j\right) = \left\| s_{unknown} - s_j \right\| \tag{4.11}$$

In this thesis, the Euclidian distance is utilized for shape recognition. The distances of the unknown shape to all the shape in database are computed. Then, the most similar shape corresponds to the minimum value of Euclidian distance.

- Another similarity measure, the normalized cross correlation value is defined as

$$R\left(s_{unknown}, s_j\right) = \frac{s_{unknown} \cdot s_j}{\left\| s_{unknown} \right\| \cdot \left\| s_j \right\|} \tag{4.12}$$

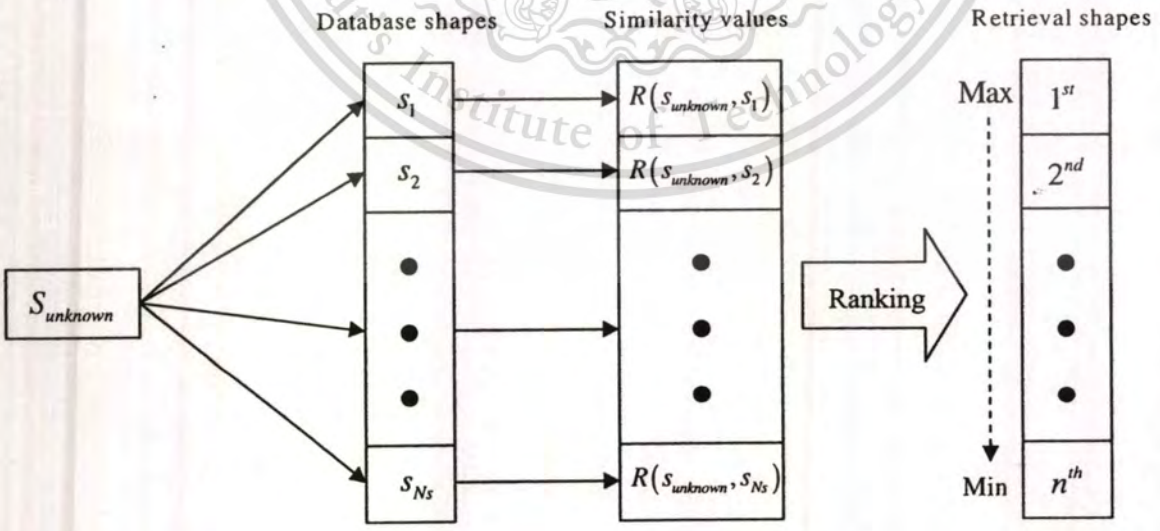


Figure 4.4: Matching and retrieval shape by normalized cross correlation.

It is employed for shape matching and retrieval in this thesis. The normalized cross correlation values of the unknown shape to all the database shape are computed. Then, these values are ranked in decreasing order (as shown in Figure 4.4) in which the most similar shape corresponds to the highest value of the normalized cross correlation.

Chapter 5

Experiment and Results

In this chapter, the performance of the proposed algorithms is demonstrated and compared using three different databases, the affine shape database, the MPEG-7 database CE-shape-1 part B and the Kimia's database. The first database is utilized to investigate the robustness of our algorithm to affine transformations such as translation, scale, rotation and shear. Then, the MPEG-7 database part B is tested for the retrieval test, which is the most comprehensive shape retrieval test in the literature so far. The last database is evaluated for the robustness to partial occlusion. The experiment of robustness to noise is carried out as well.

For the retrieval performances (accuracies), two frequently assessed methods are employed. The first one is the precision-recall curve in which the precision value at a certain recall is the average of the precision value of all database shapes at that recall. The precision and recall values can be computed as follows:

$$\text{Precision} = \frac{\text{Number of retrieved relevant shape}}{\text{Total number of retrieved shape}} \tag{5.1}$$

$$\text{Recall} = \frac{\text{Number of retrieved relevant shape}}{\text{Total number of relevant shape}} \tag{5.2}$$

The second retrieval evaluation is the Bullseye test, for example, the retrieval of shapes in the MPEG-7 database part B is counted as correct if it is in the same class as the query. The number of correct retrievals in the first top 40 ranks was counted (Section 5.2) and the retrieval accuracy is calculated as $\frac{\text{number of correct retrievals}}{1400*20}$.

5.1 Affine shape database

The affine invariant shape database [40-42] consists of 40 categories of shapes chosen from MPEG-7 CE-shape-1 database part B as shown in Figure 5.1. Each shape in that figure is transformed into 14 different distorted shapes including the original one by Eq. (3.4) with scale, rotation, and shear parameters as detailed in Table 5.1, except translation parameter which is removed by shifting shape contours about their center points. In total, there are 560 affine distorted shapes in this database. The sample affine distorted shape for class butterfly is shown in Figure 5.2.

Figure 5.1: Shapes in each class of affine shape database

Table 5.1: Affine transformation's parameters

Affine transform parameters	Shape's order													
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Scale (k)	1	1.5	0.6	1.6	1	0.75	2.5	1.35	0.5	1.4	1.8	0.8	1.7	2.1
Rotation (θ^0)	0	25	50	75	100	125	150	175	200	225	250	275	300	325
Shear-x (sh_x)	0	1	1.5	0.5	1.5	0.5	1	0	0	0	0	0	0	0
Shear-y (sh_y)	0	0	0	0	0	0	0	0.5	0.5	1.4	0.5	1.5	0.5	1

Figure 5.2: A sample of affine distorted shapes of class butterfly

5.1.1 Shape recognition

In this section, all 560 shapes are divided into two sets, a training set and a testing set. In this experiment, various numbers of sample distorted shapes in each class (1 sample shape until 13 sample shapes per class) are randomly selected as a training set and the rest are used as a testing set. Each experiment is performed many times, that is, 30 times. Thus, the recognition rates have been given in an average value over 30 times of experiment.

The recognition rates of SEA and TAR features using EBcC and FBcC methods in spectral and spatial domains, respectively, at various numbers of training distorted shapes are listed in Table 5.2 and Table 5.4. Also, Table 5.3 shows the recognition rates of power spectrum of SEA and TAR using EBcC and FBcC method respectively. To make it easy to explain the data in the tables, they are previewed by curves as plotted in Figure 5.3.

Table 5.2: The recognition rates of SEA and TAR at various numbers of training set in spectral domain

Training/Sample shapes per class	Testing/Sample shapes per class	EBcC method		FBcC method	
		SEA	TAR	SEA	TAR
40 (1)	520 (13)	85.67%	70.63%	84.83%	70.29%
80 (2)	480 (12)	92.99%	86.62%	92.50%	80.30%
120 (3)	440 (11)	95.60%	91.55%	96.67%	90.21%
160 (4)	400 (10)	96.58%	93.59%	98.36%	92.38%
200 (5)	360 (9)	97.52%	95.59%	98.84%	92.38%
240 (6)	320 (8)	98.06%	96.54%	99.03%	95.44%
280 (7)	280 (7)	98.07%	96.98%	99.50%	96.96%
320 (8)	240 (6)	98.53%	97.39%	99.47%	96.92%
360 (9)	200 (5)	98.65%	97.83%	99.37%	97.53%
400 (10)	160 (4)	98.79%	97.96%	99.37%	97.52%
440 (11)	120 (3)	98.92%	98.17%	99.69%	98.03%
480 (12)	80 (2)	98.58%	98.67%	99.63%	97.58%
520 (13)	40 (1)	99.08%	98.08%	99.17%	97.92%

The classifications in various domains have been done. Figure 5.3(a) shows a comparison of the recognition rates between EBcC and FBcC technique associating with SEA signature. For instance, the recognition rates of SEA signature using FBcC method are slightly lower than those using EBcC method in frequency domain. However, FBcC method is better than EBcC method when the number of training shapes is increased. The same for the power spectrum of SEA signature, FBcC method well outperform the EBcC method. In contrast, the recognition rates of SEA signature classified by EBcC method are higher than those classified by FBcC method in spatial domain. Figure 5.3(b) shows a comparison of TAR signature classified by EBcC and FBcC method. For the moment, the recognition rates of TAR signature classified by EBcC method are better than those classified by FBcC method in frequency domain and power spectrum (the recognition rate fall down to 84.89% at 6 training sample shapes per class).

Table 5.3: The recognition rates of power spectrum of SEA and TAR at various numbers of training set

Training/Sample shapes per class	Testing/Sample shapes per class	EBcC method		FBcC method	
		SEA	TAR	SEA	TAR
40 (1)	520 (13)	77.63%	75.88%	78.53%	77.24%
80 (2)	480 (12)	88.74%	91.10%	93.76%	90.22%
120 (3)	440 (11)	92.70%	95.19%	93.23%	93.47%
160 (4)	400 (10)	94.56%	96.63%	97.59%	94.94%
200 (5)	360 (9)	95.71%	97.83%	98.34%	95.94%
240 (6)	320 (8)	96.41%	98.16%	98.81%	84.89%
280 (7)	280 (7)	97.04%	98.79%	99.20%	94.90%
320 (8)	240 (6)	97.97%	98.86%	99.32%	96.32%
360 (9)	200 (5)	98.18%	99.25%	99.50%	97.77%
400 (10)	160 (4)	98.42%	99.31%	99.48%	97.92%
440 (11)	120 (3)	98.56%	99.75%	99.61%	98.22%
480 (12)	80 (2)	98.96%	99.75%	99.46%	98.67%
520 (13)	40 (1)	99.17%	99.58%	99.50%	99.17%

If it is classified in spatial domain, FBcC method only outperforms EBcC method at lower number of training sample shapes per class (until 6 training sample shapes per class). An alternative comparison to Figure 5.3(a) and Figure 5.3(b), a comparison between SEA and TAR signature classified by EBcC and FBcC method is shown in Figure 5.3(c) and Figure 5.3(d), respectively. From the last two figures, we found that SEA signature does better than TAR signature, except its power spectrum as classified by EBcC method and at a few number of training sample shapes per class when classified by FBcC method in spatial domain, because TAR signature is sensitive to noise rather than SEA signature as shown in Figure 3.10.

Table 5.4: The recognition rates of SEA and TAR at various numbers of training set in spatial domain

Training/Sample shapes per class	Testing/Sample shapes per class	EBcC method		FBcC method	
		SEA	TAR	SEA	TAR
40 (1)	520 (13)	74.27%	63.67%	72.97%	63.88%
80 (2)	480 (12)	84.90%	76.01%	80.44%	77.73%
120 (3)	440 (11)	88.95%	81.60%	83.06%	83.64%
160 (4)	400 (10)	91.26%	84.95%	84.36%	87.38%
200 (5)	360 (9)	93.27%	87.67%	87.03%	88.51%
240 (6)	320 (8)	94.49%	89.20%	89.10%	89.84%
280 (7)	280 (7)	94.58%	90.36%	89.62%	84.64%
320 (8)	240 (6)	95.75%	91.78%	90.61%	85.32%
360 (9)	200 (5)	96.33%	92.55%	91.13%	87.65%
400 (10)	160 (4)	96.77%	92.79%	91.92%	89.60%
440 (11)	120 (3)	97.28%	93.17%	92.28%	89.97%
480 (12)	80 (2)	96.88%	93.92%	92.75%	91.71%
520 (13)	40 (1)	98.17%	94.58%	92.92%	90.17%

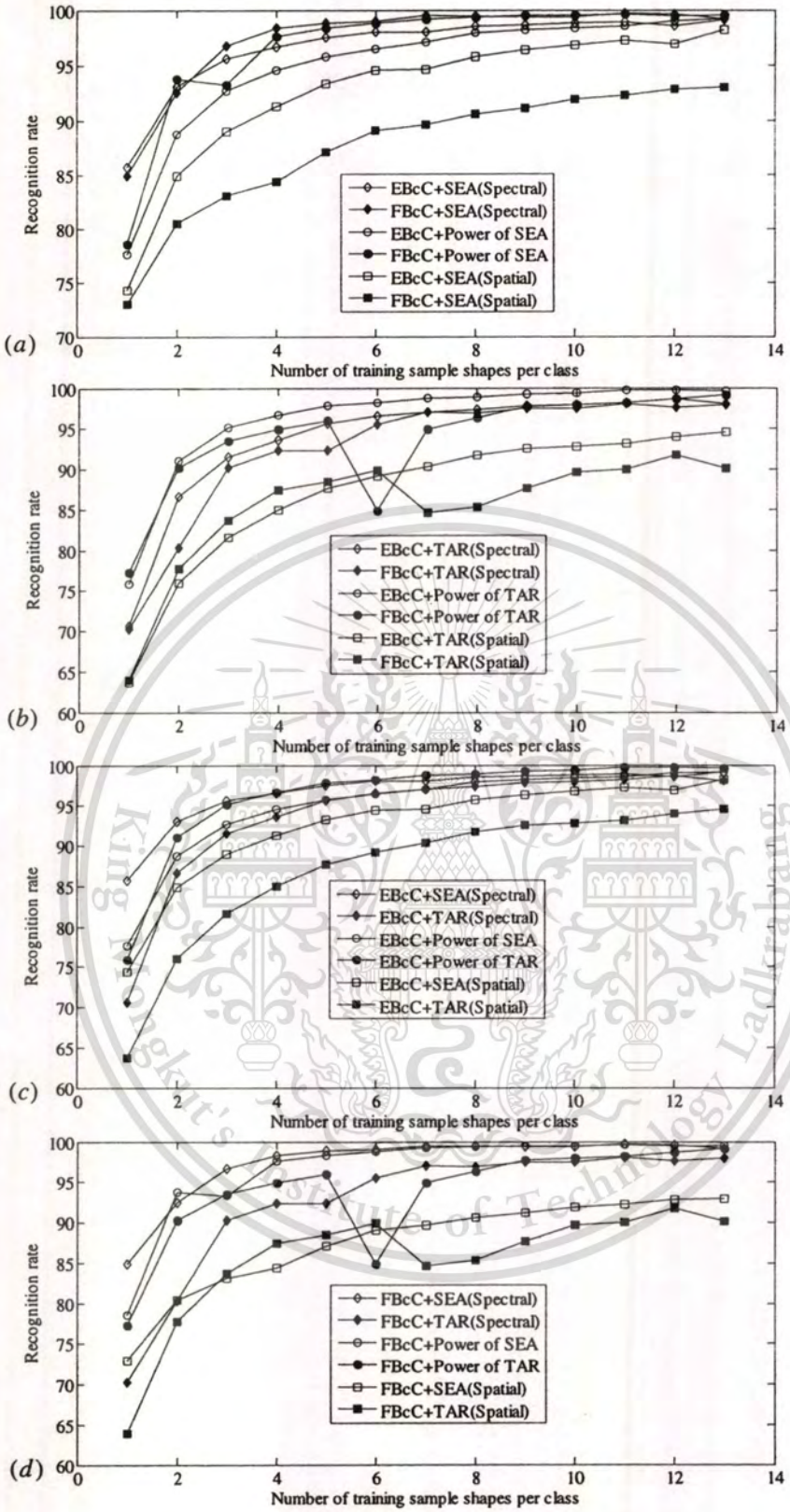


Figure 5.3: The recognition rates of SEA and TAR signatures using EBcC and FBcC methods, tested for the affine shape database.

5.1.2 Shape matching and retrieval

Another evaluation for shape classification is the shape retrieval test. In this case, all 560 shapes are stored in database and also used as the test query shapes. The retrieval performance of SEA and TAR signatures are assessed by the precision-recall curve as depicted in Figure 5.4 and Figure 5.5, respectively, except in Figure 5.5(a). We found that the FBcC method well outperforms the EBcC method for both signatures. Only the power spectrums of both signatures, when classified by EBcC and FBcC techniques, are better than the others (classified in frequency and spatial domains) as illustrated in Figure 5.4(e), Figure 5.5(d), and Figure 5.5(e). If we look at Figure 5.4(d), however, the retrieval performance in frequency domain of SEA signature is the best classified by EBcC method.

For the matching results (or called the best-match accuracy) is observed by leave-one-out procedure. If we look at the precision-recall curve, the best-match accuracy of each feature using both techniques is the second point of that curve as shown in Figure 5.4 and Figure 5.5. The accuracies of each signature are listed in Table 5.5.

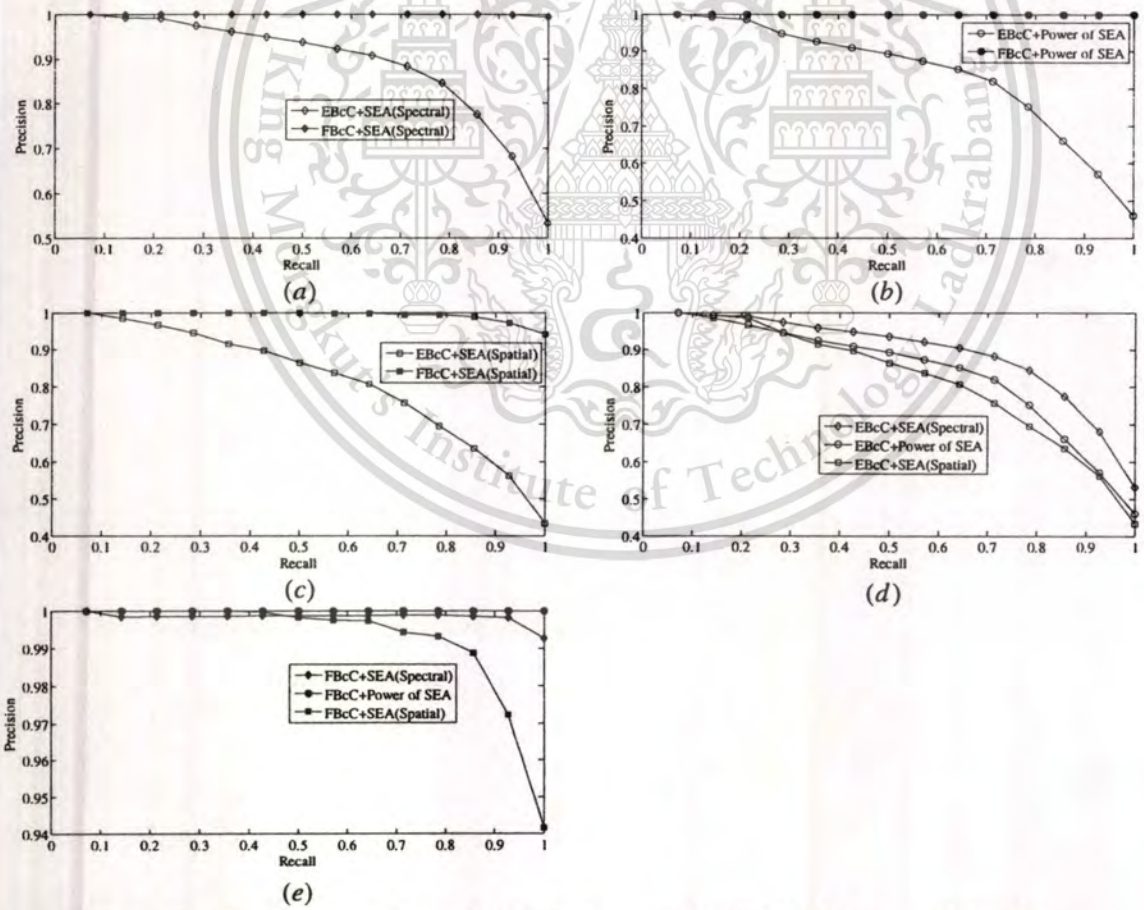


Figure 5.4: Precision-recall curve of SEA signature classified (a) in spectral domain; (b) with Power spectrum of SEA; (c) in spatial domain; (d) by EBcC; (e) by FBcC.

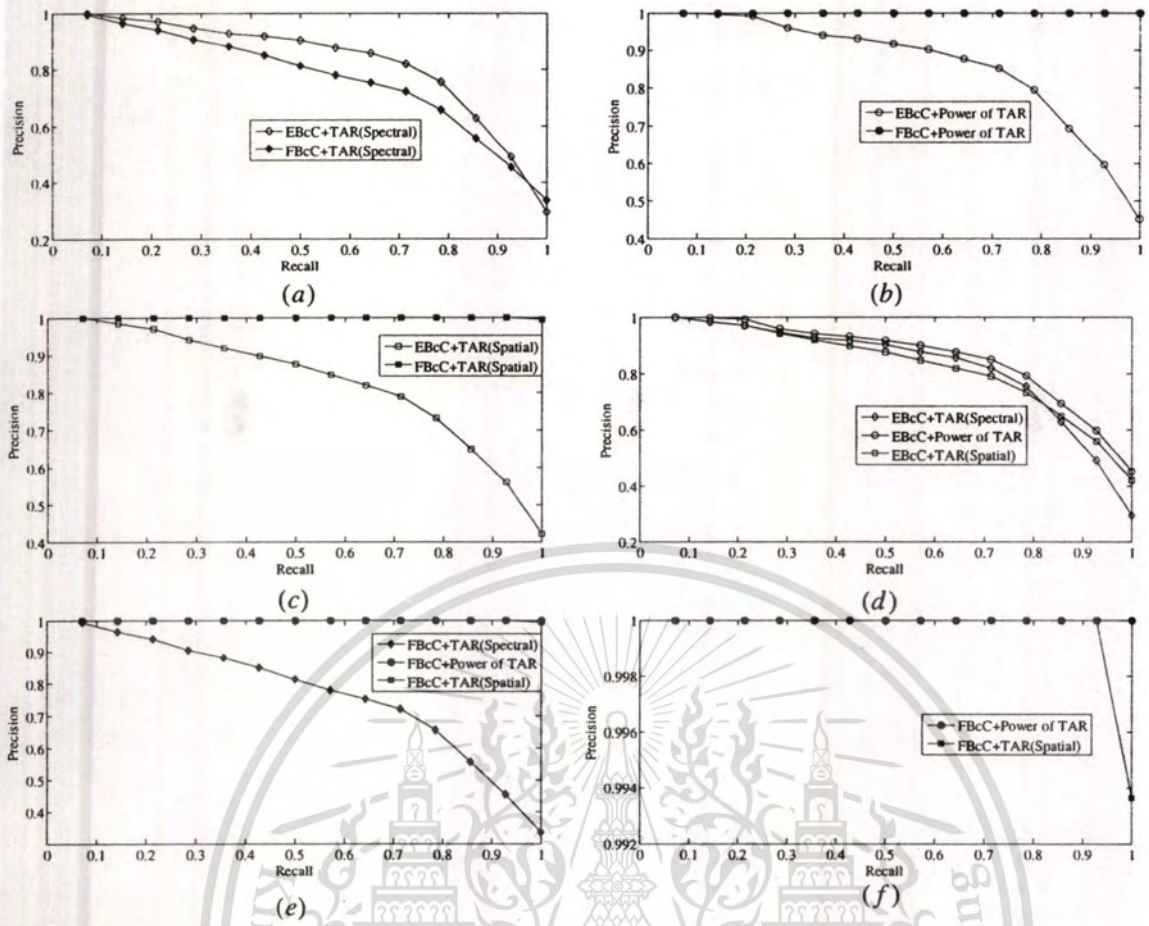


Figure 5.5: Precision-recall curve of TAR signature classified (a) in spectral domain; (b) with Power spectrum of SEA; (c) in spatial domain; (d) by EBcC; (e) by FBcC; (f) detail of (e).

Table 5.5: The best-match accuracy of SEA and TAR tested for the affine shape database

Our algorithms	SEA	TAR
EBcC (spectral domain)	99.20%	98.11%
EBcC + Power spectrum	99.24%	99.59%
EBcC (spatial domain)	98.50%	98.41%
FBcC (spectral domain)	99.85%	96.37%
FBcC + Power spectrum	100%	100%
FBcC (spatial domain)	100%	100%

Figure 5.6: Sample shapes in each class of the MPEG-7 database CE-shape-1 part B.

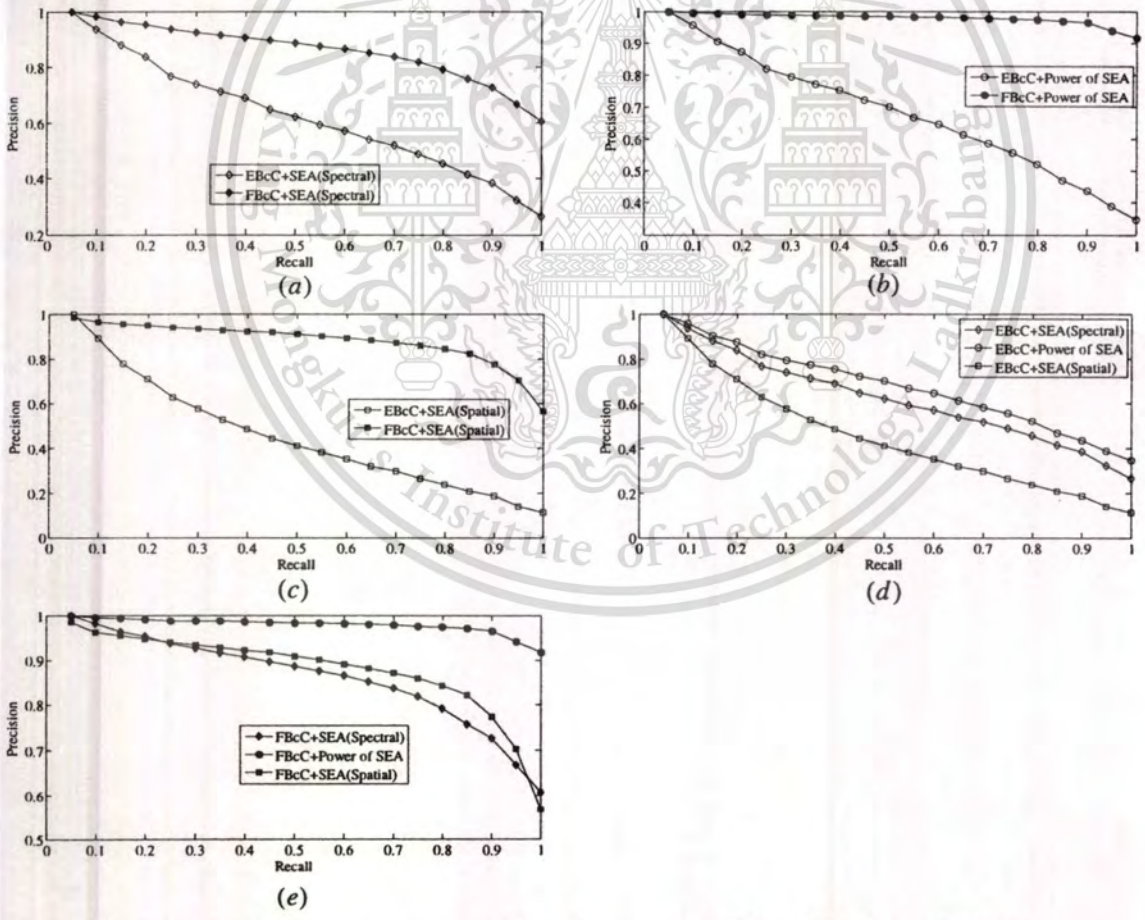


Figure 5.7: Precision-recall curve of SEA signature classified (a) in spectral domain; (b) with Power spectrum of SEA; (c) in spatial domain; (d) by EBcC; (e) by FBcC.

5.2 MPEG-7 database CE-shape-1

The MPEG-7 CE-shape-1 database has been widely used for shape matching and retrieval. It is a standard and well-known database consisting of 1400 images classified into 70 classes and containing a mixture of natural and artificial objects under various rigid and non-rigid deformations. The sample shapes of each class in MPEG-7 CE-shape-1 database part B are shown in Figure 5.6.

The retrieval performance is assessed by the precision-recall curve as depicted in Figure 5.7 and Figure 5.8. We found that FBcC method does better than EBcC method for both signatures. The power spectrum of both signatures achieves the highest rank of retrieval accuracy based on precision-recall curve as classified by EBcC method. The second rank is in spectral domain and the last one is in spatial domain as plotted in Figure 5.7(d) and Figure 5.8(d). The same for the power spectrum, it is still in the first rank of retrieval accuracy as classified by FBcC method. The second and the last rank are in spatial and spectral domains as shown in Figure 5.7(e) and Figure 5.8(e), respectively.

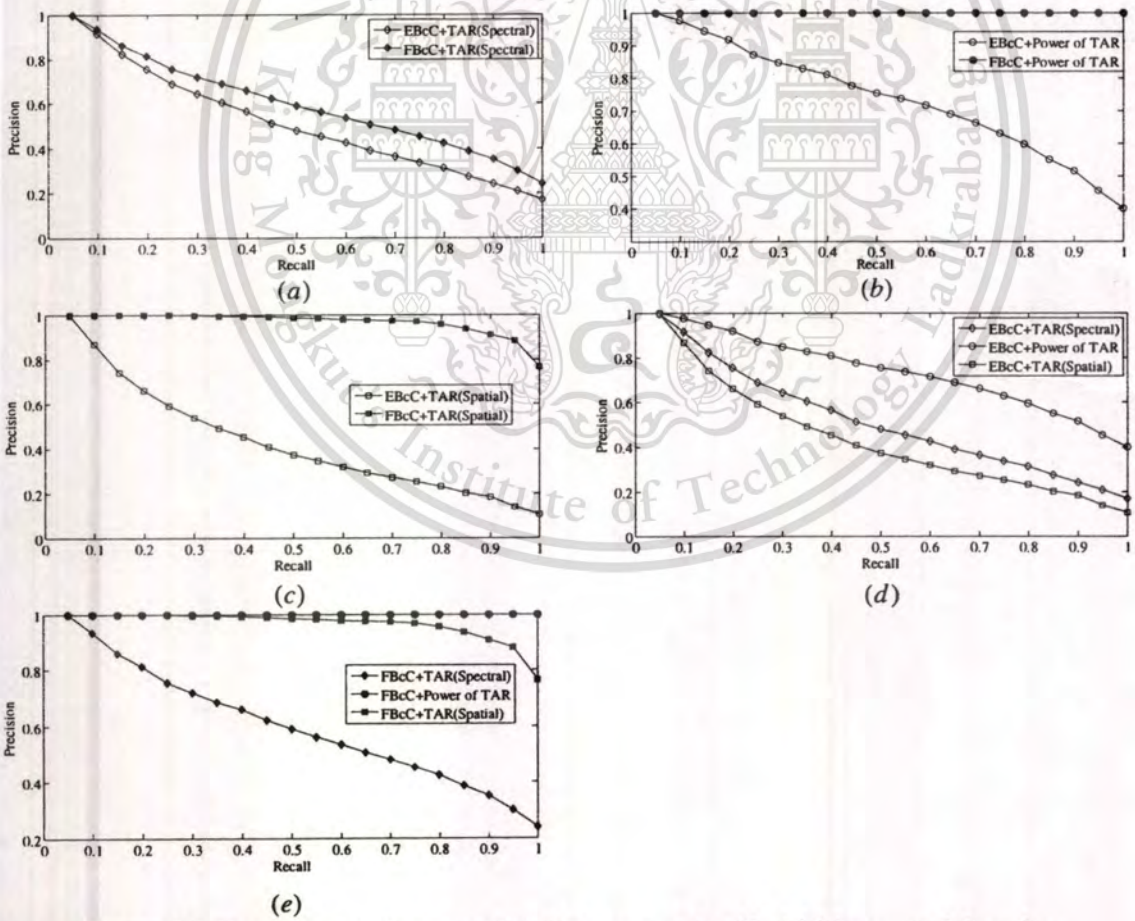


Figure 5.8: Precision-recall curve of TAR signature classified (a) in spectral domain; (b) with Power spectrum of TAR; (c) in spatial domain; (d) by EBcC; (e) by FBcC.

Additionally, we use another standard test called Bullseye test for the evaluation. For instance, the number of correct retrievals in the top 40 ranks is counted, including the self-match. We have compared our results with the published results by the MPEG-7 database part B, in which the best result has been reported in [54] using contour flexibility with the rate of 89.31%, as listed in Table 5.6. For our proposed algorithm, the Bullseye scores are listed in Table 5.7.

Table 5.6: The Bullseye test of the previous algorithms testing for the MPEG-7 database part B

Method	Retrieval accuracy (%)
WARP [17]	58.50
VP [31]	76.45
Curve Edit Distance [32]	78.17
CSS [57]	81.12
BAS [34]	82.37
MCC [18]	84.93
IDSC + DP [19]	85.40
Symbolic Representation [35]	85.92
DSW+Global [10]	87.23
Hierarchical deformable-shape tree [19]	87.70
Contour flexibility [56]	89.31

Table 5.7: The Bullseye test of our algorithm tested for the MPEG-7 database part B

Our algorithms	SEA	TAR
EBcC (spectral domain)	66.50% [58]	56.93%
EBcC + Power spectrum	72.42% [58]	78.53%
EBcC (spatial domain)	48.89%	47.18%
FBcC (spectral domain)	89.60% [58]	65.84%
FBcC + Power spectrum	98.62% [58]	100%
FBcC (spatial domain)	90.88%	97.13%

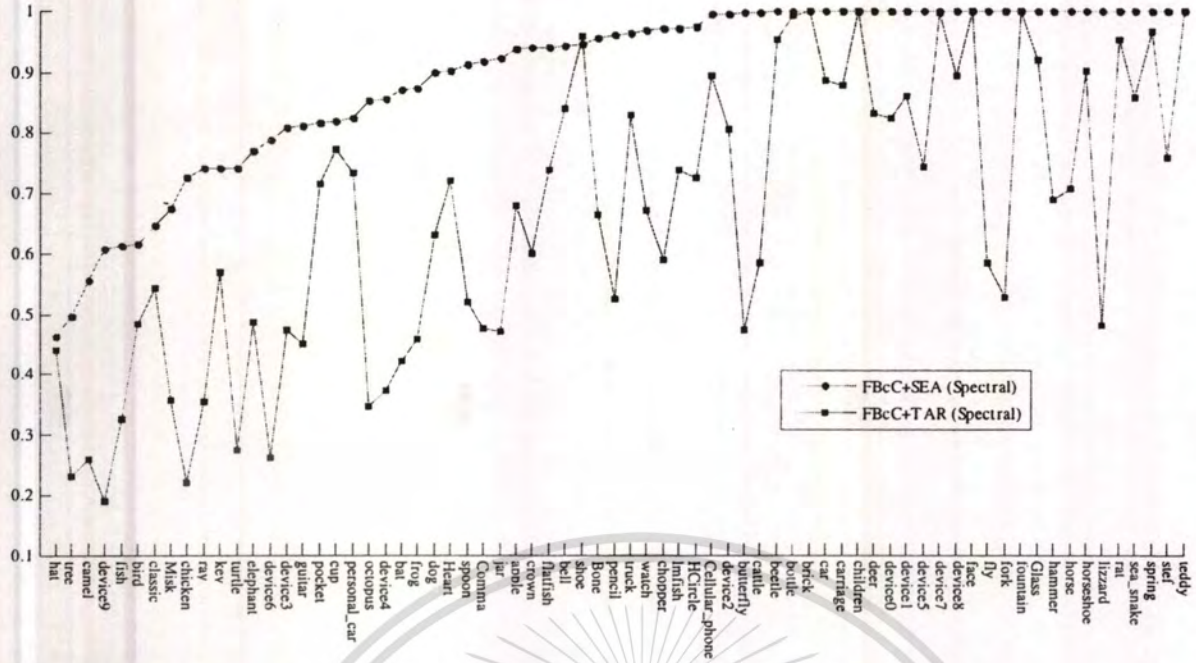


Figure 5.9: A comparison of the retrieval accuracy of SEA and TAR in spectral domain for each class of the MPEG-7 database CE-shape-1 part B using FBcC method.

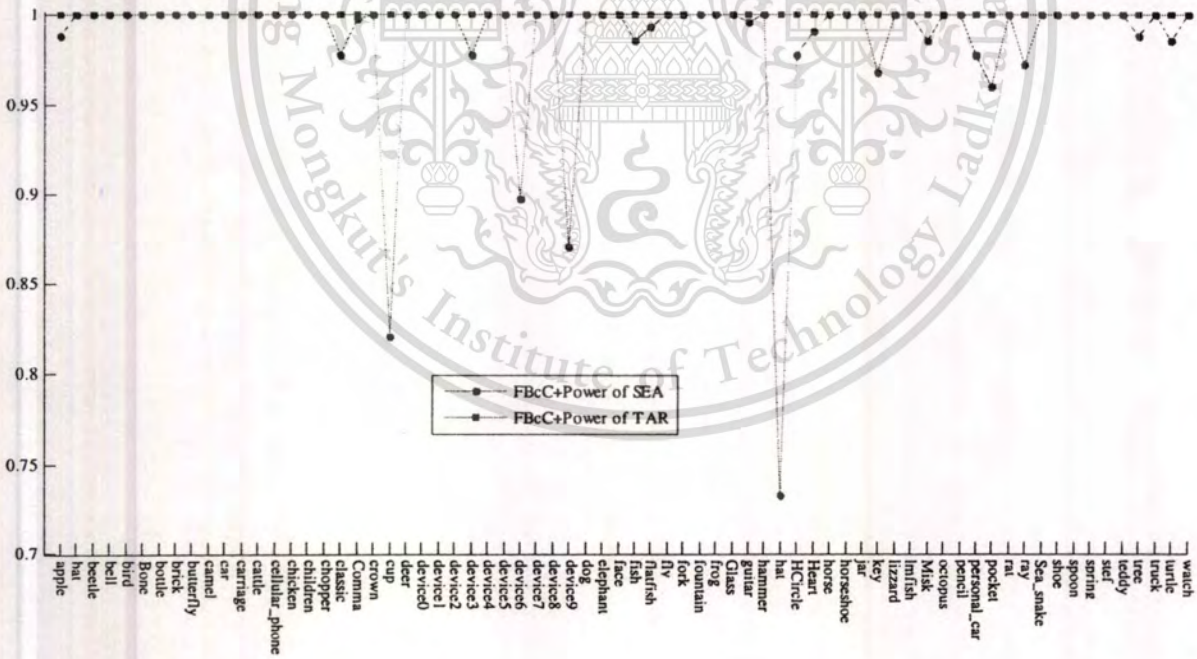


Figure 5.10: A comparison of the retrieval accuracy of power spectrum of SEA and TAR for each class of the MPEG-7 database CE-shape-1 part B using FBcC method.

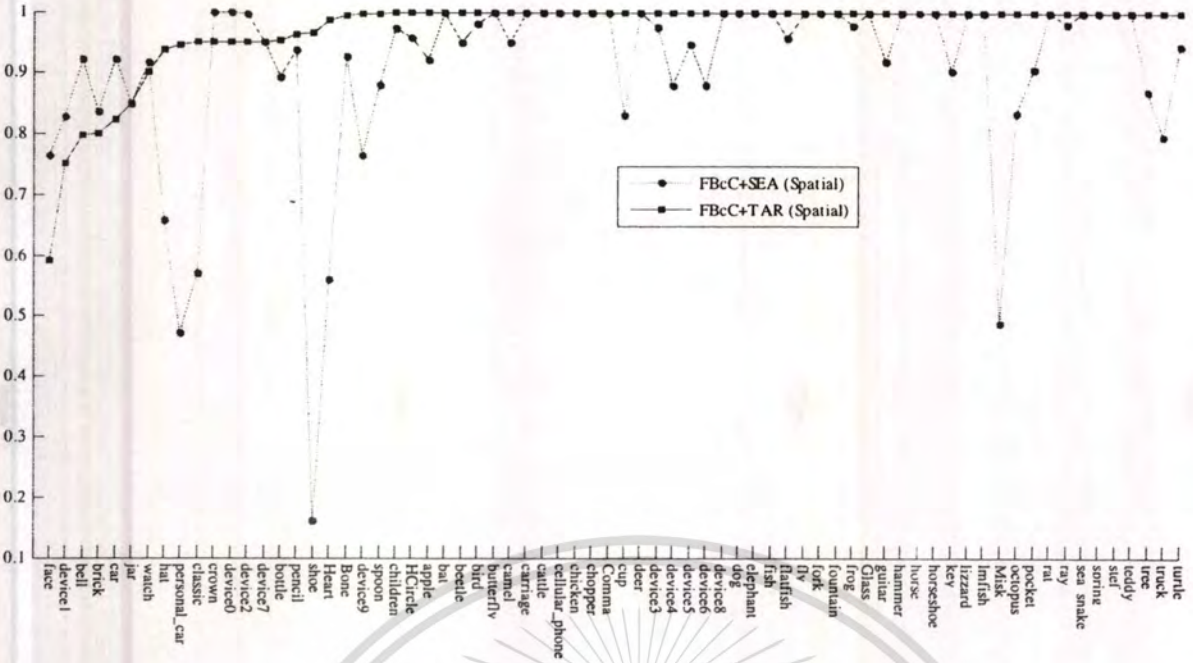


Figure 5.11: A comparison of the retrieval accuracy of SEA and TAR in spatial domain for each class of the MPEG-7 database CE-shape-1 part B using FBcC method.

EBcC method gives Bullseye score (retrieval rate) of less than 50% for both signatures (SEA and TAR) which is the lowest ones comparing with the previous algorithms. However, FBcC method when associated with the power spectrum of both signatures achieves unexpected results with very high retrieval rates of 98.62% and 100%, over all the published methods, for SEA and TAR signatures, respectively. The retrieval accuracy, based on the Bullseye test, of SEA and TAR classified by FBcC method in spectral domain, power spectrum and in spatial domain are plotted in Figure 5.9, Figure 5.10 and Figure 5.11, respectively.

5.3 Kimia's database

The Kimia's database consists of 99 shapes from nine categories in which most of shapes are partially occluded. All the shapes in this database are shown in Figure 5.12.

In the test of Kimia's database, each shape in the database is considered as a query shape and the number of the correct retrieval shapes from the first top 10 closet match is counted. Table 5.8 summarizes the results of the previous algorithms. Notice that the maximum number of correctness for each category is 99. The results of correct retrieval shapes of our algorithm are shown in Table 5.9 and Table 5.10 for SEA and TAR signatures, respectively.

Figure 5.12: All 99 shapes in Kimia's database.

Table 5.8: The retrieval rate of different algorithms, evaluating for the Kimia's database

Method	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Total
Shape Context [7]	97	91	88	85	84	77	75	66	56	37	756
Shock graph [30]	99	99	99	98	98	97	96	95	93	82	956
MDS+SC+DP [29]	99	98	98	98	97	99	97	96	97	85	964
DSW+Global [19]	99	99	99	98	98	97	98	95	93	80	956
Symbolic representation [33]	99	99	99	98	99	98	98	95	96	94	975

Table 5.9: The retrieval rate of SEA using EBcC and FBcC method evaluating for the Kimia's database

	Method	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Total
SEA	EBcC (spectral domain)	99	94	87	82	76	73	65	60	54	45	735
	EBcC + Power spectrum	99	94	94	92	82	84	82	77	67	57	828
	EBcC (spatial domain)	99	79	67	50	53	42	46	38	37	42	553
	FBcC (spectral domain)	99	99	99	99	97	98	98	95	97	94	975
	FBcC + Power spectrum	99	99	99	99	99	99	99	99	99	99	990
	FBcC (spatial domain)	99	99	99	99	99	99	99	99	99	88	979

Table 5.10: The retrieval rate of TAR using EBcC and FBcC method evaluating on the Kimia's database

	Method	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Total
TAR	EBcC (spectral domain)	11	11	12	13	17	14	10	8	12	11	119
	EBcC + Power spectrum	99	93	94	91	82	80	83	78	76	64	840
	EBcC (spatial domain)	99	84	73	62	60	44	45	41	41	38	587
	FBcC (spectral domain)	61	26	29	35	34	33	23	30	33	30	334
	FBcC + Power spectrum	99	99	99	99	99	99	99	99	99	99	990
	FBcC (spatial domain)	99	99	99	99	99	99	99	99	99	99	990

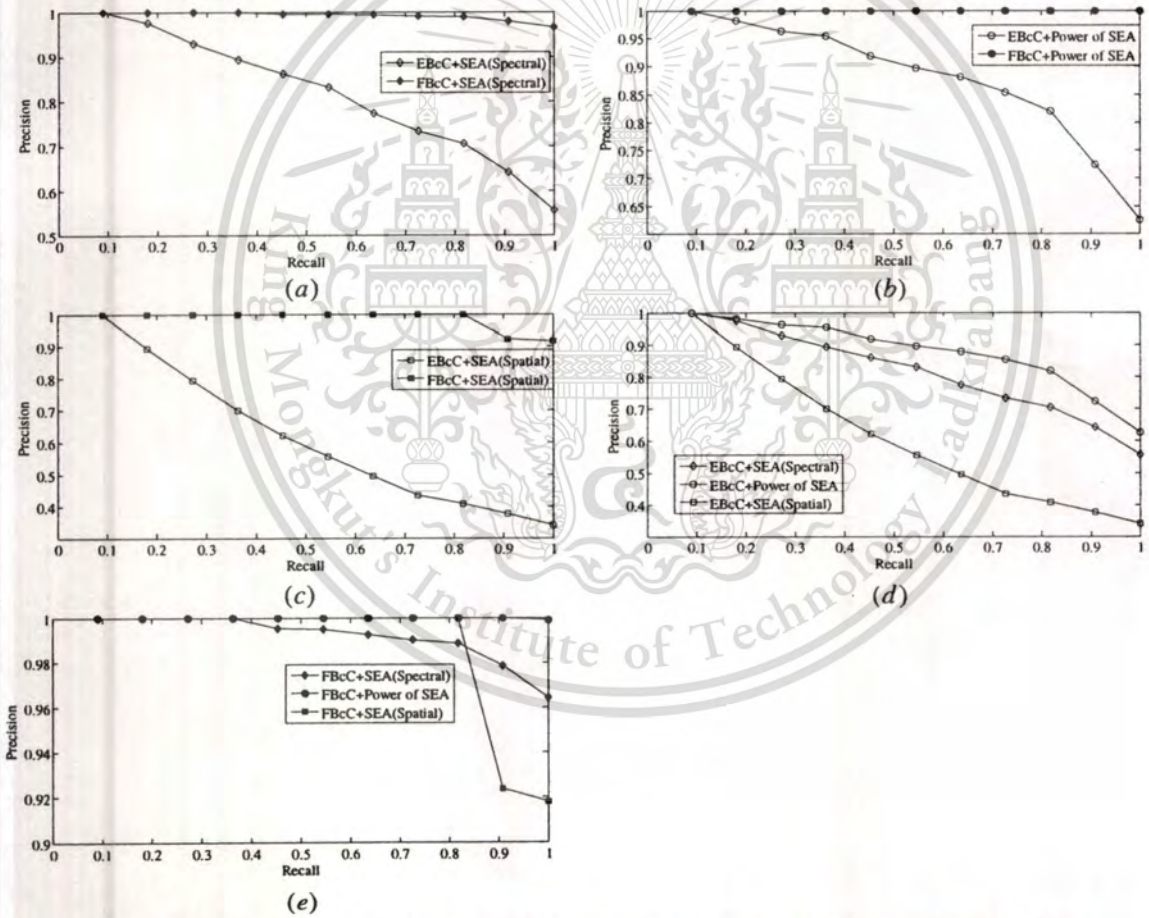


Figure 5.13: Precision-recall curve of SEA signature classified (a) in spectral domain; (b) with Power spectrum of SEA; (c) in spatial domain; (d) by EBcC; (e) by FBcC.

We found that the EBcC method for SEA signature can retrieve in 10 cases of 553 shapes which is low performance comparing with the other algorithms. However, for TAR signature, when classified in spectral domain by EBcC and FBcC methods — they can retrieve in the first 10 cases of 119 and 334 shapes, respectively, which is even lower than the rate of SEA signature classified by EBcC method in spatial domain. Of all test, only the three tests give the retrieval rate of 100% in the top 10 cases, those are the power spectrum of SEA and TAR signatures classified by FBcC method and TAR signature classified by FBcC method in spatial domain. Moreover, in each case, the retrieval rates are plotted as the precision-recall curve in Figure 5.13 and Figure 5.14 for SEA and TAR signatures, respectively.

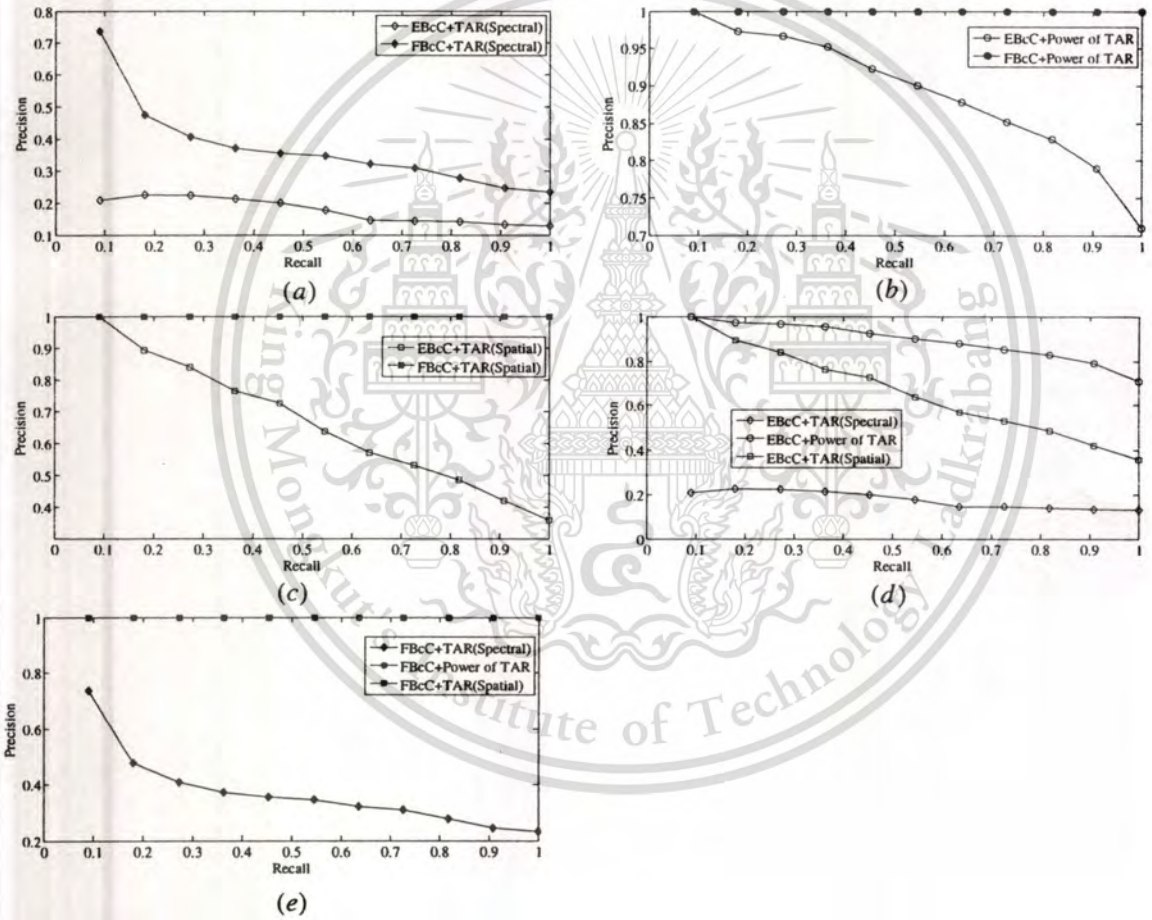


Figure 5.14: Precision-recall curve of TAR signature classified (a) in spectral domain; (b) with Power spectrum of SEA; (c) in spatial domain; (d) by EBcC; (e) by FBcC.

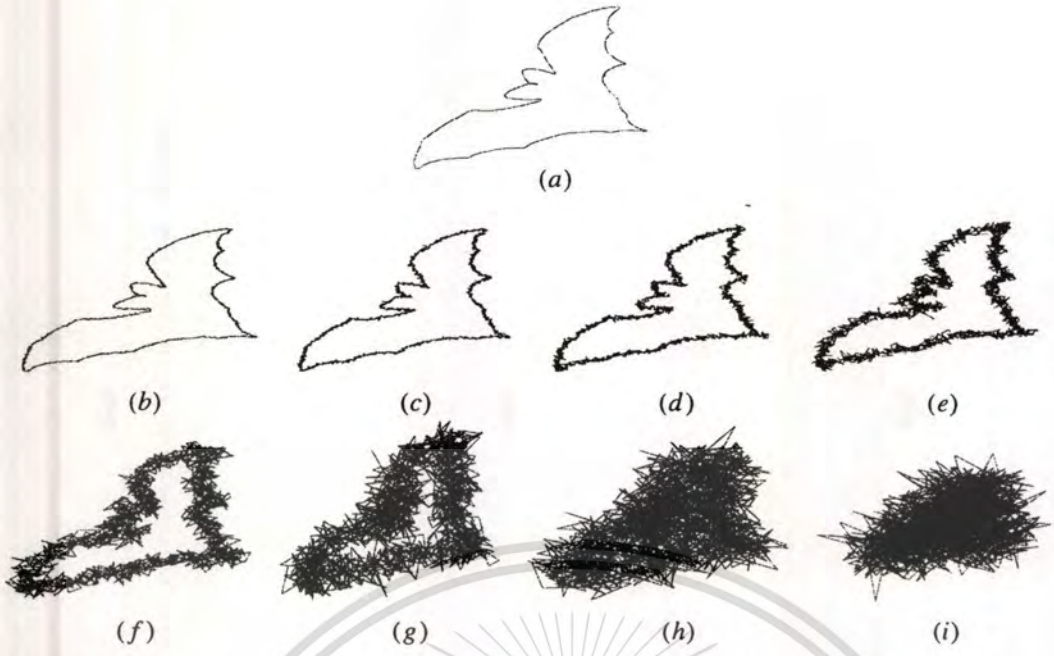


Figure 5.15: Example of bird's contour at different SNR ratio (a) original; (b) 45 dB; (c) 40 dB; (d) 35 dB; (e) 30 dB; (f) 25 dB; (g) 20 dB; (h) 15 dB; (i) 10 dB.

5.4 Robustness to noise

In this section, the performance of our proposed algorithm is evaluated in the noisy environment. Just follow [59], after the boundary extraction from each query shape, uniformly distributed noise is artificially added to the x and y coordinates of each contour point. Various noise levels or signal-to-noise ratios (SNR) are added to the coordinates of each boundary point of bird as shown in Figure 5.15. It is difficult to identify which of the class in Figure 5.15(g)-(i) even by human eyes.

The testing results of the robustness to noise are previewed by the precision-recall curve shown in Figure 5.16 and Figure 5.17 for SEA and TAR signatures, respectively. By comparing eight cases of robustness to noise testing in Figure 5.16, the power spectrum of SEA signature classified by FBcC method does better than the others except two cases at low SRN ratios of 15 dB and 10 dB. Based on a leave-all-in test, when matching SEA signature by FBcC method in spatial domain, it achieves 82.44% correct categorization with SNR of 20 dB. Then, with the results of EBcC method in spatial domain, it gives the correct categorization of 65.21% and 45.88% when testing with image corrupted by large amount of noise of 15 dB and 10 dB, respectively. If we look at Figure 5.17 for TAR signature, FBcC method in spatial domain gives better performance than the others except for the last three cases, those are, rate of 61.16% for EBcC method in spatial domain with 20 dB and EBcC method with power spectrum with the rates of 34.70 % and 27.19% in cases of 15 dB and 10 dB, respectively.

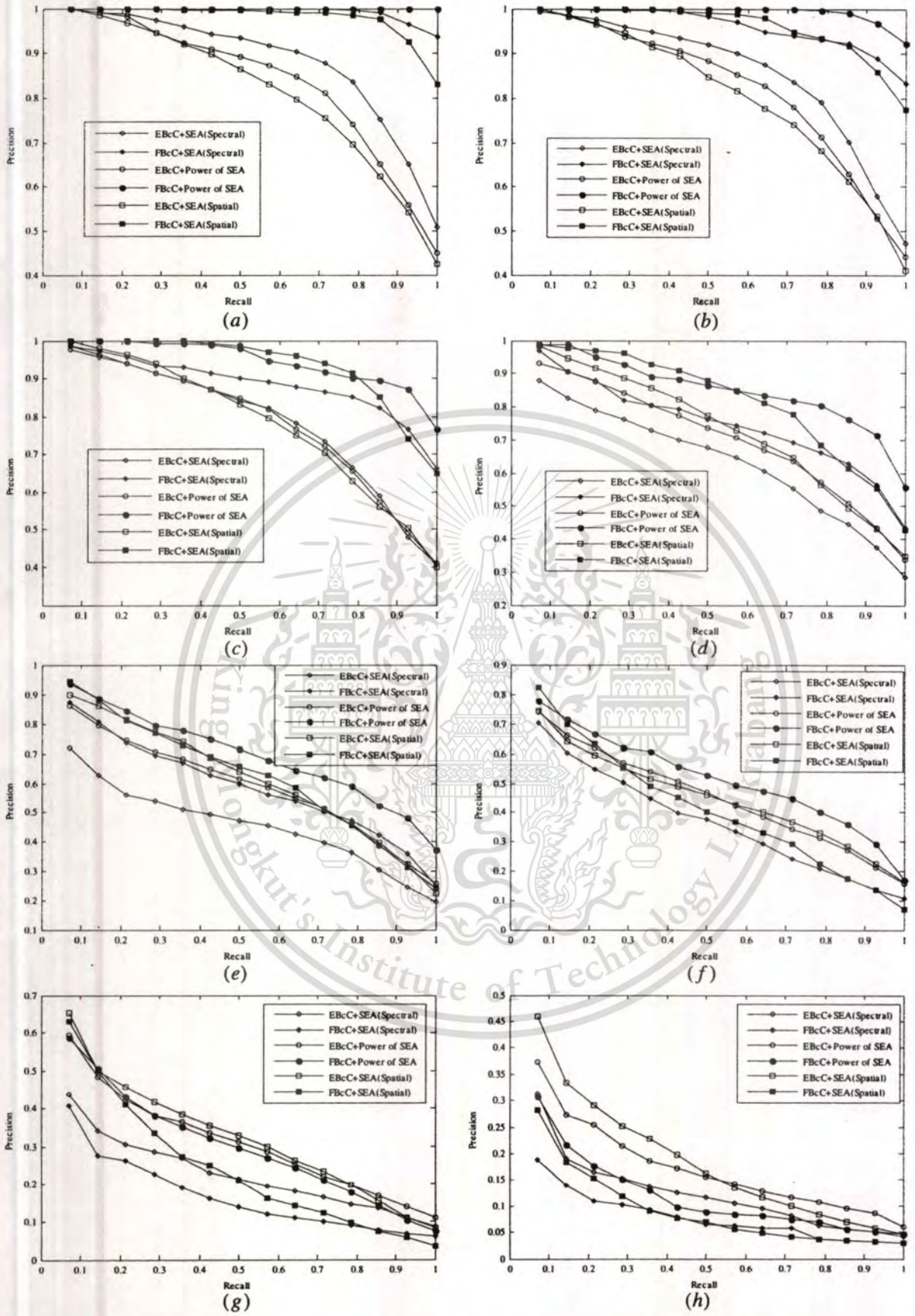


Figure 5.16: The precision-recall curve of SEA at different SRN ratio (a) 45 dB; (b) 40 dB; (c) 35 dB; (d) 30 dB; (e) 25 dB; (f) 20 dB; (g) 15 dB; (h) 10 dB.

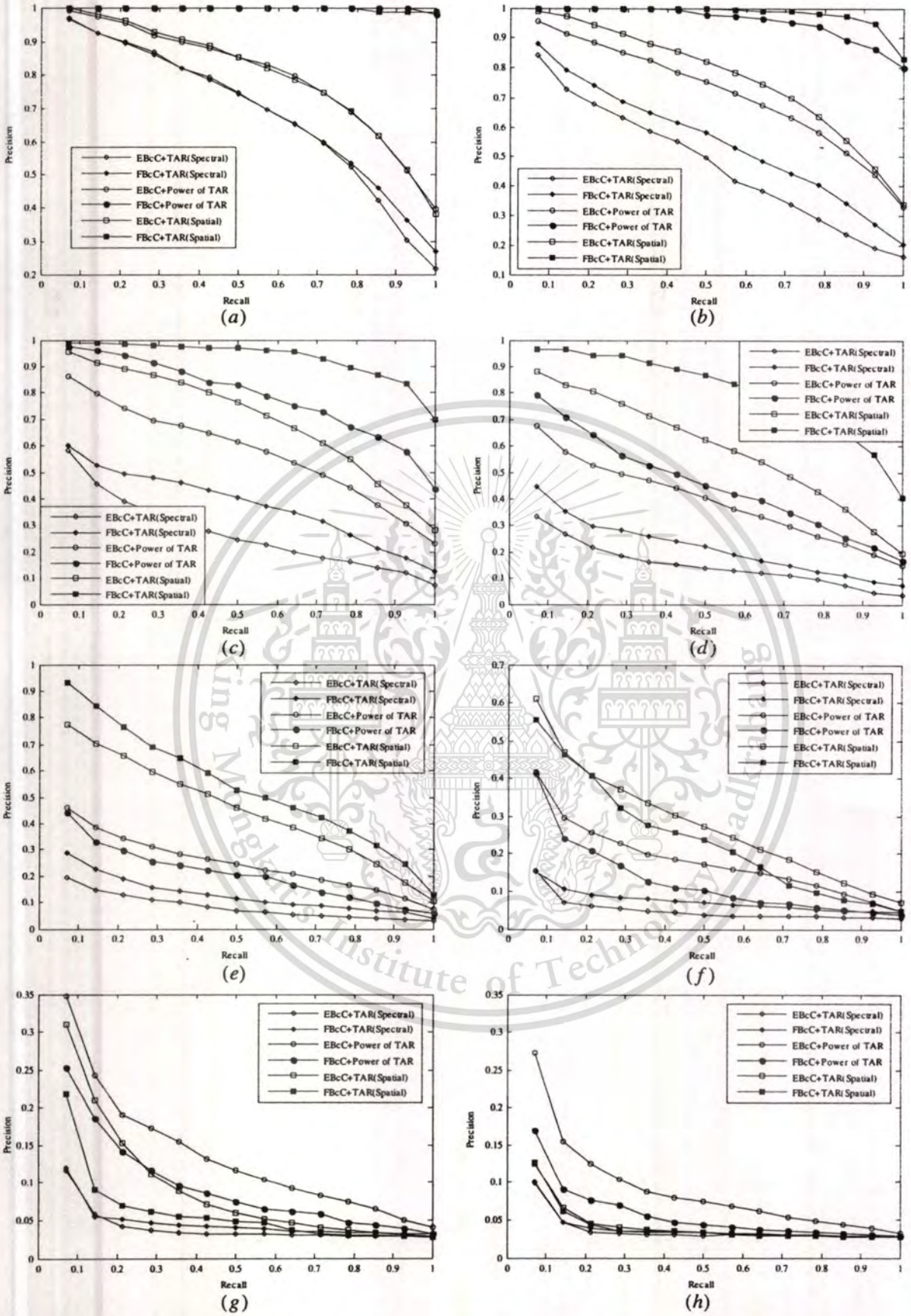


Figure 5.17: The precision-recall curve of TAR at different SRN ratio (a) 45 dB; (b) 40 dB; (c) 35 dB; (d) 30 dB; (e) 25 dB; (f) 20 dB; (g) 15 dB; (h) 10 dB.

Chapter 6

Conclusion and Future Work

6.1 Conclusions and discussions

In this thesis, the new multi-scale algorithm has been developed for 2D closed boundary shape classification. The proposed algorithm is invariant to affine transformation (translation, rotation, scale, and shear) as well as to the starting point variation of the shape boundary. In our algorithm, a new multi-resolution technique has been proposed, called Barycenter Contour (BcC) decomposition, to decompose the shape boundary into the multi-scale contours. This technique can reduce noise and small boundary distortion. However, it cannot smooth the deep concavity and complex shape contours until becoming circle or ellipse. We observe two signatures (SEA and TAR at TSL = 1) which are computed alternatively from Barycenter contour (BcC) of shape at each scale level. Both shape signatures are invariant to affine transformation as proof in Chapter 3.

Only the shape representation, it is impossible to achieve the high performance for shape classification. It is required further, that is the shape classification technique. In Chapter 4, we have used two popular methods which are frequently applied to face recognition system; Principal Component Analysis (PCA) and Fisher Discriminant Analysis (FDA). Both methods are important in the classification process. The first one, PCA is used for dimensionality reduction because of data of the high dimension. The other method, FDA is employed for dimensionality reduction as well and for improving the performance by increasing discrimination between classes of object and decreasing discrimination within classes of object. However, this method cannot use alone because of time-consuming and an expensive computation when finding the eigenvectors. It should be combined with PCA method for a convenient way and a good performance as shown in Figure 4.1.

For the evaluation of our algorithm presented in Chapter 5, four main experiments in various domains have been done with three databases. The first database called affine shape database has been used to test the robustness to affine transformation and to noise. The last two databases, the well-known databases which are popular employed for shape classification research; the MPEG-7 CE-shape-1 part B database is a large database and tested for shape matching and retrieval, and the Kimia's database is tested for robustness to partially occlusion.

The experimental results show that most of the FBcC method outperforms the EBcC method over three databases for shape matching and retrieval tests. Moreover, the performance of shape matching and retrieval have been improved using the power spectrum of the signature without considering the global features of shape (circularity, eccentricity, and aspect ratio) and shape indexing (shape complexity) which are important to improve the performance of shape retrieval as in some algorithms [18, 19, 46]. The power spectrum of signature is better than the corresponding signature when classified by EBcC and FBcC methods in both spectral and spatial domains. For example, the power spectrum of TAR signature classified by FBcC method achieves retrieval rate of 100% over those three databases except testing with noisy environment. In case of testing with noise, the power spectrum of SEA signature classified by FBcC method is better than the others while TAR signature classified by FBcC method is better than the others, at almost all levels of signal-to-noise ratio (SNR).

The time processing per query shape based on FBcC method for each database are shown in Table 6.1, Table 6.2 and Table 6.3. These results are obtained using MATLAB R2008b program running on desktop Intel Core 2 Duo 2.4 GHz and RAM 2 GB. It should be noted that the codes are not optimized and better speed can be obtained using other programming languages such as C and C++.

Table 6.1: Time execution for the affine shape database

Signature	SEA			TAR		
Domain	spectral	power	spatial	spectral	power	spatial
Starting point (avt a shape)	2.4 ms	2.4 ms	28 ms	2.4 ms	2.4 ms	28 ms
Training time by FBcC	8.4 s	8.4 s	15.8 s	8.3 s	9.8 s	17.2 s
Testing by FBcC (avt. a shape)	10 ms	10 ms	20.3 ms	10 ms	10 ms	20.5 ms
NCC (av. time shape)	12.4 ms (39 dimension of each projected feature)					

Table 6.2: Time execution for the MPEG-7 shape CE-1 database

Signature	SEA			TAR		
Domain	spectral	power	spatial	spectral	power	spatial
Starting point (avt a shape)	2.4 ms	2.4 ms	28 ms	2.4 ms	2.4 ms	28 ms
Training time by FBcC	72.94 s	74.34 s	116.76 s	71.12 s	87.36 s	148.26 s
Testing by FBcC (avt. a shape)	24.6 ms	24.7 ms	50.1 ms	24.6 ms	24.3 ms	49.7 ms
NCC (av. time shape)	36.2 ms (69 dimension of each projected feature)					

Table 6.3: Time execution for the Kimia's database

Signature	SEA			TAR		
Domain	spectral	power	spatial	spectral	power	spatial
Starting point (avt a shape)	2.4 ms	2.4 ms	28 ms	2.4 ms	2.4 ms	28 ms
Training time by FBcC	336.6 ms	336.6 ms	722.7 ms	336.6 ms	356.4 ms	722.7 ms
Testing by FBcC (avt. a shape)	1.8 ms	1.7 ms	3.8 ms	1.6 ms	1.9 ms	3.8 ms
NCC (av. time shape)	2.1 ms (8 dimension of each projected feature)					

Note: avt a shape = average time per shape

NCC = Normalized Cross Correlation

As shown in Tables for each database, the time processing in spatial domain is more than that in spectral domain and for power spectrum. Although, the computational time of the spectral domain and power spectrum are almost the same. For the classification step including training, testing and matching, the time executions are different for each database because of the size of database. This means that if the size of database increases, it requires more computational time for training, testing and similarity matching.

6.2 Future Work

As stated earlier, even though our algorithm has achieved a good performance with unexpected results, the disadvantage is the high computation which requires much time in the process. To reduce this inconvenient, we will reduce the level of BcC by selecting some best levels and find the other suitable techniques for classification but still keeping the same result or the acceptable results.

Another work needed to be studied more is the problem of partial occlusion. Although the MPEG-7 and Kimia's shape databases used in our experiments include partially occluded shapes, this problem has not been studied deeply in this thesis. Partial occlusion occurs in different forms; for instance, it can be modeled as missing parts of the image or as another unknown object occluding (part of) the image. We will work more deeply with partially occluded shapes.

References

[1] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, Prentice Hall, 2002.

[2] L.da.F. Costa and R.M. Cesar Jr, *Shape Analysis and Classification: Theory and Practice*, CRC Press LLC, 2001.

[3] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Elsevier, 2006.

[4] D. Zhang and G. Lu, "*Review of shape representation and description techniques*," *Pattern Recognition*, Vol. 37, pp. 1-19, 2004.

[5] E.G.M. Petrakis, A. Diplaros, and E. Milios, "Matching and Retrieval of Distorted and Occluded Shapes Using Dynamic Programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 11, pp. 1501-1516, November 2002.

[6] M. Seul, L. O'Gorman, and M.J. Sammon, *Practical Algorithms for Image Analysis: Description, Examples, and Code*, Cambridge University Press, 2000.

[7] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Context," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, No. 4, pp. 509-522, April 2002.

[8] H. Freeman, "On the encoding of arbitrary geometric configurations," *IRE Transaction on Electronic Computers*, EC-10(2), pp. 260-268, 1961.

[9] W.I. Groskey and R. Mehrotra, "Index-based object recognition in pictorial data management, " *Computing Vision Graphics Image Processing 52*, pp. 416-436, 1990.

[10] M.K. Hu, "Visual pattern recognition by moment invariants, " *IRE Transactions Information Theory*, Vol. 8, No. 2, pp. 179-187, 1962.

[11] J. Flusser and T. Suk, "Pattern recognition by affine moment invariant, " *Pattern Recognition 26*, pp. 167-174, 1993.

[12] M. Sonka, V. Hiavac, and R. Boyle, *Image Processing, Analysis and Machine Vison*, Thomson Learning, 2008.

[13] J. Sklansky, "Measuring concavity on a rectangular mosaic, " *IEEE Transactions on Computers*, Vol. 21, No. 12, pp. 1355-1364, 1972.

[14] S. Loncaric, "A survey of shape analysis techniques, " *Pattern Recognition 31(8)*, pp. 983-1001, 1998.

[15] R. Klette and A. Rosenfeld, *Digital Geometry: geometric method for digital picture analysis*, Morgan Kaufmann, 2002.

[16] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 26, No. 1, pp. 43-49, February 1978.

[17] I. Bartolini, P. Ciaccia, and M. Patella, "WARP: Accurate Retrieval of Shapes Using Phase of Fourier Descriptors and Time Warping Distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 27, No. 1, pp. 142-147,January 2005.

[18] T. Adamek and N.E. O'Connor, "A Multiscale Representation Method for Nonrigid Shapes with a Single Closed Contour," *IEEE Transactions on Circuits System and Video Technology*, pp. 742-753, 2004.

[19] N. Alajlan, I. El Rube, M.S. Kamel, and G. Freeman, "Shape retrieval using triangle-area representation and dynamic space warping," *Pattern Recognition 40*, pp. 1911-1920, 2007.

[20] F. Mokhtarian and A. Mackworth, "Scale Based Description and Recognition of Planar Curves and Two-Dimensional Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 1, pp. 34-43, January 1986.

[21] F. Mokhtarian and A.K. Mackworth, "A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 8, pp. 789-805, August 1992.

[22] The MPEG Home Page. www.chiariglione.org/mpeg/index.htm

[23] S. Abbasi, F. Mokhtarian, and J. Kittle, "Curvature Scale Space Image in Shape Similarity Retrieval, " *MultiMedia Systems 7(6)*, pp. 467-476, 1999.

[24] S. Abbasi and F. Mokhtarian, "Affine-Similar Shape Retrieval: Application to Multiview 3-D Object Recognition," *IEEE Transactions on Image Processing*, Vol. 10, No. 1, pp. 131-139, January 2001.

[25] F. Mokhtarian, "Silhouette-Based Isolated Object Recognition through Curvature Scale Space, " *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.15, No. 5, pp. 539-544, 1995.

[26] I.El Rube, M. Ahmed and M. Kamel, "Affine Invariant Multiscale Wavelet-base Shape Matching Algorithm," *Proceedings of the first Canadian Conference of Computer and Robot Vision*, pp. 217-224, May 2004.

[27] I. El Rube, N. Alajlan, M. Kamel, M. Ahmed, and G. Freeman, "Efficient Multiscale Shape-Based Representation and Retrieval," *International Conference on Image Analysis and Recognition*, pp. 415-422, September 2005.

[28] K. Yawichai and Y. Kitjaidure, "Multi-view Shape Matching and Retrieval," *International Conference on Robotics, Vision, Information, and Signal Processing (ROVISP)*, Penang, Malaysia, pp. 445-458, 2007.

[29] T. Kosorl, K. Yawichai, and Y. Kitjaidure, "Optimum Shape Representation Based on Fisher's Discriminant Analysis, " *International Conference in Electrical/Electronics, Computer, Telecommunications, and Information Technology (ECTI-CON)*, Krabi, Thailand, pp. 509-512, 2008.

[30] E. Milios and E.G.M. Petrakis, "Shape Retrieval Based on Dynamic Programming, " *IEEE Transactions on Image Processing*, Vol. 9, No. 1, pp. 141-147, 2000.

[31] H. Ling, and D. Jacobs, "Shape Classification Using Inner-Distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, No. 2, pp. 286-299, February 2007.

[32] T. Sebastian, P. Klein, and B. Kimia, "On Aligning Curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 25, No. 1, pp. 116-125, January 2003.

[33] L.J. Latecki and R. Lakamper, "Shape Similarity Measure Based on Correspondence of Visual Parts," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, No. 10, pp. 1185-1190, October 2000.

[34] N. Arica and F. Vural, "BAS: A Perceptual Shape Descriptor Based on the Beam Angle Statistics," *Pattern Recognition* Letter, pp. 1627-1639, 2003.

[35] M.R. Darili and V. Torre, "Robust Symbolic Representation for Shape Recognition and Retrieval," *Pattern Recognition 41*, pp. 1782-1798, 2008.

[36] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active Shape Models: Their Training and Application," *Computer Vision Image Understanding 61(1)*, pp. 38-59, 1995.

[37] P.F. Felzenszwalb and J.D. Schwartz, "Hierarchical Matching of Deformable Shapes," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007.

[38] O. Gunther and E. Wong, "The arc tree: An Approximation Scheme to Represent Arbitrary Curved Shapes, " *Computer Vision, Graphics, and Image Processing 51*, pp. 313-337, 1990.

[39] C.C. Chen, "Improved moment invariants for shape discrimination, " *Pattern Recognition* 26, pp. 683–686, 1993.

[40] K. Thourn and Y. Kitjaidure, and S. Kondo, "Affine Invariant Shape Recognition Based on Multi-Level of Barycenter contour," *International Symposium on Communication and Information Technologies*, pp. 145-149, October 2008.

[41] K. Thourn and Y. Kitjaidure, "Multi-View Shape Recognition Based on Principal Component Analysis," *International Conference on Advanced Computer Control*, pp. 265-269, January 2009.

[42] K. Thourn and Y. Kitjaidure, "Robust Multi-TAR of Barycenter Contour for Multiple View Shape Matching and Retrieval," *IEEE International Conference on Robotics and Biomimetics*, pp. 1894-1898, February 2009.

[43] P. Thumwarin and T. Matsuura, "On-Line Writer Recognition for Thai Based on Velocity of Barycenter of Pen-Point Movement," *International Conference on Image Processing*, pp. 889-892, 2004.

[44] Thomas.H. Reiss, *Recognizing Planar Object Using Invariant Image Features*, Springer-Verlag, 1991.

[45] K. Yawichai and Y. Kitjaidure, "A Simple and Efficient Algorithm for Affine Invariant Shape Matching, " *International Conference on Engineering, Applied Scieinces, and Technology (ICEAST)*, Bangkok, Thailand, pp. 246-249, 2007.

[46] Y. Rui, A.C. She, and T.S. Huang, "Modified Fourier Descriptors for Shape Representation: A Practical Approach, " *Proceeding of the First International Workshop on Image Databases and Multi Media Search*, pp. 22-23, August 1996.

[47] D. Zhang and G. Lu, "A Comparative Study of Fourier Descriptors for Shape Representation and Retrieval, " *Proceeding of the Fifth Asian Conference on Computer Vision*, pp. 646-651, January 2002.

[48] B. Wang and J.A. Bangham, "PCA Based Shape Descriptors for Shape Retrieval and the Evaluations," *International Conference on Computational Intelligence and Security*, pp. 1401-1406, 2006.

[49] M.S. Drew, T.K. Lee, and A. Rova, "Shape Retrieval with Eigen-CSS Search," *Image and Vision Computing 27*, pp. 748-755, 2009.

[50] P.M. Belhumeur, J.P. Haspanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711-720, July 1997.

[51] Marios Savvides, B.V.K. Vijaya Kumar and P.K. Khosla, "Eigenphases vs. Eigenfaces," *Proceedings of the 17th International Conference on Pattern Recognition*, pp. 810-813, 2004.

[52] A.M. Martinez and A.C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, pp. 228-233, February 2001.

[53] E. Alpaydm, *Introduction to Machine Learning*, The MIT Press, 2004.

[54] S.Y. Kung, M.W. Mak, and S.H. Lin, *Biometric Authentication: A Machine Learning Approach*, Pearson Education, Inc., 2005.

[55] M. Cheriet, N. Kharma, C.L. Liu, and C.Y. Suen, *Character Recognition Systems: A Guide for Students and Practioners*, John Wiley & Son, Inc., 2007.

[56] C. Xu, J. Liu, and X. Tang, "2D Shape Matching by Contour Flexibility, " *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 1, pp. 180-186, January 2009.

[57] F. Mokhtarian and M. Bober, *Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization*, Kluwer Academic Publishers, 2003.

[58] K. Thourn, Y. Kitjaidure, and S. Kondo, "Robust 2D Shape Matching and Retrieval Using Fisher Barycenter Contour, " ECTI-CON2009, Vol. 2, pp. 659-662, May 2009.

[59] G. Tzimiropoulos, N. Mitianoudis, and T. Stathaki, "Robust Recognition of Planar Shapes Under Affine Transforms Using Principal Component Analysis," *IEEE Signal Processing Letters*, Vol. 14, No. 10, pp. 723-726, October 2007.

Appendix A

Barycenter

In geometry, the centroid, geometric center, or Barycenter of a plane figure X is the intersection of all straight lines that divide X into two parts of equal moment about the line. Informally, it is the average of all points of X. The definition extends to any object X in n-dimensional space: its centroid is the intersection of all hyper-planes that divide X into two parts of equal moment.

In physic, the word centroid may mean the geometric center of the object's shape, as above; or its physical Barycenter, which is its center of mass or the center of gravity depending on the context. Informally, the Barycenter is the average of all points, weighted by the local density or specific weight, respectively.

A.1. Properties

The geometric centroid of a convex object always lies in the object. A non-convex object might have a centroid that is outside the figure itself. The centroid of a ring or a bowl, for example, lies in the object's central void.

If the centroid is defined, it is a fixed point of all isometries in its symmetry group. In particular, the geometric centroid of an object lies in the intersection of all its hyperplanes of symmetry. The centroid of many figures (regular polygon, regular polyhedron, cylinder, rectangle, rhombus, circle, sphere, ellipse, ellipsoid, super-ellipse, super-ellipsoid, etc.) can be determined by this principle alone.

In particular, the centroid of a parallelogram is the meeting point of its two diagonals. This is not true for other quadrilaterals.

For the same reason, the centroid of an object with translational symmetry is undefined (or lies outside the enclosing space), because a translation has no fixed point.

A.2. Centroid of a finite set of points

The centroid of a finite set of points $x_1, x_2, \ldots\ldots, x_k$ in n-dimensional space is

$$C = \frac{x_1 + x_2 + \ldots\ldots + x_k}{k} = \frac{1}{k}\sum_{i=1}^{k} x_i$$

A.3. Centroid by geometric decomposition

The centroid of a plane figure X can be computed by dividing it into a finite number of simpler figures $X_1, X_2, X_3, ..., X_n$, computing the centroid C_i and area A_i of each part, and then computing:

$$C = \frac{\sum C_i A_i}{\sum A_i}$$

The same formula holds for any three-dimensional objects, except that each A_i should be the volume of X_i, rather than its area. It also holds for any subject of $\Re^d$, for any dimension d, with the areas replaced by the d-dimensional measures of the parts. This formula holds even if the parts overlap and/or extend outside the set X, provided that the measures A_i are taken with positive and negative signs in such a way that the sum of the A_i of all parts that enclose a given point x is 1 if x belongs to X, and 0 otherwise.

A.4. Centroid of polygon

The centroid of a non-overlapping closed polygon defined by n vertices (x_i, y_i) can be calculated as follows.

The area of the polygon is: $$A = \frac{1}{2}\sum_{i=0}^{n-1}(x_i y_{i+1} - x_{i+1} y_i)$$

and its centroid is $C = (C_x, C_y)$ where,

$$C_x = \frac{1}{6A}\sum_{i=0}^{n-1}(x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

$$C_y = \frac{1}{6A}\sum_{i=0}^{n-1}(y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

In these formulas, the vertex (x_n, y_n) is assumed to be the same as (x_0, y_0).

Biography

Personal Information

Name – Surname	KOSORL THOURN
Nationality	Cambodian
Date of birth	August 28, 1985
Place of birth	Kandal province, Cambodia

Education

Bachelor degree

Field	Electrical Engineering
Duration	2002-2007
Department	Department of Electrical Engineering
University	Institute of Technology of Cambodia (ITC), Cambodia

Master degree

Field	Electronic Engineering
Duration	2007-2009
Department	Department of Electronics
University	King Mongkut's Institute of Technology, Ladkrabang (KMITL), Thailand

Research Interests

Shape Matching and Retrieval, Pattern Recognition, Computer Vision, Signal Processing, and Image Processing.

List of International Conference Proceeding Papers

1- Kosorl Thourn, Yuttana Kitjaidure, and Shozo Kondo, "Affine Invariant Shape Recognition Based on Multi-Level of Barycenter Contour," *International Symposium on Communication and Information Technologies*, October 21-23, 2008, Vientiane, Lao PDR, pp. 145-149.

2- Kosorl Thourn and Yuttana Kitjaidure, "Multi-View Shape Recognition Based on Principal Component Analysis," *International Conference on Advanced Computer Control*, January 22-24, 2009, Singapore, pp. 265-269.

3- Kosorl Thourn and Yuttana Kitjaidure, "Robust Multi-TAR of Barycenter Contour for Multiple View Shape Matching and Retrieval," *IEEE International Conference on Robotics and Biomimetics*, February 22-25, 2009, Bangkok, Thailand, pp. 1894-1898.

4- Kosorl Thourn, Yuttana Kitjaidure, and Shozo Kondo, "Robust 2D Shape Matching and Retrieval Using Fisher Barycenter Contour," *ECTI-CON2009*, May 6-9, 2009, Vol. 2, Pattaya, Thailand, pp. 659-662.

5- Kosorl Thourn, Yuttana Kitjaidure, and Shozo Kondo, "Eigen and Fisher Barycenter Contour for 2D Shape Classification," *The 2009 IEEE-RIVF International Conference on Computing and Communication Technologies*, July 13-17, 2009, Da Nang, Vietnam, pp. 290-295.

IEEE

ISCIT 2008

ISCIT 2008

INTERNATIONAL SYMPOSIUM ON COMMUNICATIONS AND INFORMATION TECHNOLOGIES 2008

October 21-23, 2008
Don Chan Palace, Vientiane, Lao PDR

PROGRAM AND ABSTRACT BOOK

TOKAI UNIVERSITY

jica

ECTI Association

EIC

RENESAS
Everywhere you imagine.

Affine Invariant Shape Recognition Based on Multi-Level of Barycenter Contour

Kosorl Thourn[1], Yuttana Kitjaidure[1], and Shozo Kondo[2]
[1]Department of Electronics, Faculty of Engineering
King Mongkut's Institute of Technology, Ladkrabang, Bangkok 10520, Thailand
E-mail: kosal_th@yahoo.com , kkyuttan@kmitl.ac.th
[2]Information Media Technology Department, Tokai University
1117 Kitakaname, Hiratsuka, Kanagawa 259-1292, Japan
E-mail: kondo@keyaki.cc.u-tokai.ac.jp

***Abstract*—In this paper, a new multiresolution created from multi-level of barycenter contour is proposed in order to reduce the moderate amount of noise and to improve the retrieval efficiency of the recognition task in computer vision. Then, the Triangle Area Representation with two points (TAR-2p) signature at each level of barycenter contour is introduced as the shape representation. Finally, the normalized cross-correlation function at each level is used for measuring the similarity among the shapes. Our experiment has been performed on database consisting of 560 affine distorted shapes, chosen from MPEG-7 contour shape database CE-1. The results illustrate that our algorithm is invariant to affine transformation, robustness to the noise. Moreover, it achieves high retrieval efficiencies when compares to those of the Triangle Area Representation with three points (TAR-3p) signature and the centroid distance signature.**

***Keywords*- Affine transformation, Multi-level of barycenter contour, Triangle Area Representation with two points (TAR-2p), Triangle Area Representation with three points (TAR-3p).**

I. INTRODUCTION

In the recognition and retrieval system, there are several tasks that require us to carry out particularly the object recognition task. The objects seen from multiple viewpoints, which approximately give the affine distorted shapes [1], is one of the fundamental problems required to resolve in last decade.

There are many shape descriptors to describe 2D shape such as Fourier Descriptor [2], [3] that is widely used by many researchers, Wavelet Transform [4], [5], B-Spline Moment [6]. These methods can be classified into contour-based and region-based. In our work, the contour-based method has been chosen because region-based method [7] uses more data involved with more computation and storage. Many shape representations have been introduced in the literature [3], [8].

Recently, growing research interest has been focused on the multiple resolutions technique in order to improve the accuracy and robustness to the moderate amount of noise. F. Mohkarian and F.K. Mackworth [9] have presented a multi-scale curvature- based descriptor technique for classification planar curve. As known early, Fourier descriptor cannot provide the multiple resolutions for shape representation. However, I. Kunttu, L. Lepisto, J. Tauhamaa, and A. Visa [10] have applied the Fourier transform to the coefficients of the multiscale wavelet transform in order to create a new Fourier descriptor that can be presented in multiple resolutions. In the literature [11], I.El Rube, M. Ahmed, and M. Kamel have examined on affine invariant shape classification based on dyadic wavelet transform. Both, Fourier-based and Wavelet-based are used to transform shape's boundary from spatial domain into spectral domain. In this work, a new multiresolution has been proposed in spatial domain that is called as the multi-level of barycenter contour. This new feature can reduce the noise caused by deformation and distortion. Then, the Triangle Area Representation with two points (TAR-2p) is used as the shape descriptor by computing the triangle area from two adjacent points on the shape's boundary and its center point. Finally, the matching strategy based on normalized cross correlation function is used to measure the similarity among the shapes. The results show that our algorithm is robust to affine transformation and moderate amount of noise, and achieves high accuracy.

The paper is organized as follows: Affine transformation is introduced in section II, then shape matching algorithm including preprocessing in section III, feature extraction in section IV, and recognition stage in section V. Next, the experiment results are illustrated in section VI. Finally, conclusion and future works are given in section VII.

II. AFFINE TRANSFORMATION

The general affine transformation can be represented mathematically with the following equations:

$$\left.\begin{aligned} x' &= ax + by + e \\ y' &= cx + dy + f \end{aligned}\right\} \quad (1)$$

The equations (1) can be written in matrix form:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \quad (2)$$

978-1-4244-2336-1/08/$25.00 © 2008 IEEE

Where x' and y' are the distorted pixel coordinates by affine transformation. x and y are the pixel coordinates of the original 2D shape. a, b, c, and d are the coefficients of scaling, rotation, and shear, e and f are the coefficients of translation. Because the translation parameters are removed by shifting the shape to its center of gravity, so equation (2) can be written in form:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{3}$$

Where the matrix $T = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is given by [12]:

$$T = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 & sh_x \\ sh_y & 1 \end{bmatrix} \tag{4}$$

Where s_x and s_y are the scaling parameters ($s_x = s_y = sc$), sh_x and sh_y are the shear parameters, and θ is the rotation angle parameter.

III. Preprocessing

In this stage, the shape is segmented from the image. Then applying technique in the literature [13] to the shape, the boundary of shape is extracted by starting from the specific selected point on the shape's boundary and moving counterclockwise, to obtain the coordinates of each point as two vectors denoted as $x(k)$ and $y(k)$. Next, these vectors are sampled to have the same length N, denoted as $P_1, P_2, P_3, \ldots, P_N$ where $P_i = (x_i, y_i)$ and $i = (1, 2, 3, \ldots, N)$. Finally, the shape's boundary is shifted to the origin by its center of gravity $P_c = (x_c, y_c)$.

$$\left.\begin{aligned} \hat{x}_i &= x_i - x_c \\ \hat{y}_i &= y_i - y_c \end{aligned}\right\} \tag{5}$$

Where $x_c = \frac{1}{N}\sum_{i=1}^{N} x_i$ and $y_c = \frac{1}{N}\sum_{i=1}^{N} y_i$ are the coordinates of center of gravity, $\hat{x}_i$ and $\hat{y}_i$ are the coordinates of shifted shape's boundary.

IV. Feature Extraction

In this section, we introduce how to create multi-level of barycenter contour. In the geometry, the barycentre of the set of coordinate points is the average of this set of coordinate points.

In this paper, the first level of barycentre contour have been adopted from [14] and it is determined from the coordinates of barycenter of triangle composed of $(0,0)$, $(\hat{x}_i, \hat{y}_i)$, and $(\hat{x}_{i+1}, \hat{y}_{i+1})$. The second level is determined from the coordinates of barycenter of points $(0,0)$, $(\hat{x}_i, \hat{y}_i)$, $(\hat{x}_{i+1}, \hat{y}_{i+1})$, and $(\hat{x}_{i+2}, \hat{y}_{i+2})$. So the next level is created by adding one adjacent point more to the previous level. This approach can reduce the noise caused by the distortion and deformation of shape's boundary. If $(\tilde{x}_i, \tilde{y}_i)$ is the coordinates of barycenter contour, then it is determined by:

$$\left.\begin{aligned} \tilde{x}_i^m &= \frac{\hat{x}_i + \hat{x}_{i+1} + \ldots + \hat{x}_{i+m}}{m+2} \\ \tilde{y}_i^m &= \frac{\hat{y}_i + \hat{y}_{i+1} + \ldots + \hat{y}_{i+m}}{m+2} \end{aligned}\right\} \tag{6}$$

Where $\hat{x}_i$ and $\hat{y}_i$ are the distorted boundary sequences. $\tilde{x}_i$ and $\tilde{y}_i$ are the multi-level of barycenter boundary sequences, and m is the level of barycenter contour. Figure 1 shows the example of multi-level of barycenter contour for class 'bat' at level 1, 5, 10, 20, 30, 40, and 49.

Shape signature is mathematically represented by one dimensional function to describe shapes. Many shape signatures have been presented in the literature [2], [8].

A. TAR-2p Signature

The TAR-2p is called in this paper instead of area signature which is introduced by D. Zhang and G. Lu [2]. It is computed as the area of the triangle formed by two adjacent points on the shape's boundary and its center of gravity. Let any two sample points $\tilde{P}_i = (\tilde{x}_i, \tilde{y}_i)$, $\tilde{P}_{i+1} = (\tilde{x}_{i+1}, \tilde{y}_{i+1})$ and point $O = (0,0)$, be the vertices of the triangle as shown in figure 2(a). It can be calculated by:

$$TAR.2p(m,i) = \frac{1}{2}\left|\tilde{x}_i^m \cdot \tilde{y}_{i+1}^m - \tilde{y}_i^m \cdot \tilde{x}_{i+1}^m\right| \tag{7}$$

B. TAR-3p Signature

In the literature [11], the 1D triangle area representation is calculated from the detail coefficient points in spectral domain. However, it is calculated in spatial domain instead [15]. In this work, it is called as TAR-3p. This area is computed as the triangle area of any three sample points on the shape's boundary $\tilde{P}_{i-1} = (\tilde{x}_{i-1}, \tilde{y}_{i-1})$, $\tilde{P}_i = (\tilde{x}_i, \tilde{y}_i)$, and $\tilde{P}_{i+1} = (\tilde{x}_{i+1}, \tilde{y}_{i+1})$ as shown in Figure 2(b). It is defined by:

$$TAR.3p(m,i) = \frac{1}{2}\begin{vmatrix} \tilde{x}_{i-1}^m & \tilde{y}_{i-1}^m & 1 \\ \tilde{x}_i^m & \tilde{y}_i^m & 1 \\ \tilde{x}_{i+1}^m & \tilde{y}_{i+1}^m & 1 \end{vmatrix} \tag{8}$$

Under affine transformation, the relation between the area of the distorted triangle and the original one can be expressed by:

$$A' = |J| \cdot A \tag{9}$$

Where A' is the distorted triangle area by affine transformation. A is the original triangle area, and J is the Jacobin between $\{x', y'\}$ and $\{x, y\}$. By equation (1), this Jacobin can be calculated by:

$$J = \begin{vmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = (ad - bc) \tag{10}$$

Then $$A' = |ad - bc| A \quad (11)$$

For this reason, the TAR-2p and TAR-3p are invariant to affine transformation (translation, scaling, rotation, and shear).

C. *Centroid Distance Signature*

The centroid distance is presented in many literatures [2]. It is invariant to translation, circularly shifted by rotation, and linearly changed by scaling. It is computed as the distance between the shape's boundary point and its center of gravity and can be calculated by:

$$r(m,i) = \sqrt{\left[\tilde{x}_i^m\right]^2 + \left[\tilde{y}_i^m\right]^2} \quad (12)$$

These three representations are normalized by dividing equation (7), (8), and (12) by its corresponding maximum value at each level of barycenter contour. Figure 3(a)-(b) illustrate the example of the shape in the same class 'bat' and Figure 3(c)-(d) illustrates TAR-2p, TAR-3p, and centroid distance of 'bat' at level 1, 5, 10, 20, 30, and 40. The horizontal axis represents the location of barycenter contour points and the vertical axis represents the TAR-2p, TAR-3p, and centroid distance by column respectively.

V. Recognition Stage

To measure the similarity between the query shape and the database shape, the normalized cross-correlation function between the shape representations (TAR-2p, TAR-3p, and centroid distance signature) of the query shape denoted as $Q = [q_1, q_2, q_3, \ldots, q_N]$ and those (TAR-2p, TAR-3p, and centroid distance signature) of the database shape denoted as $D = [d_1, d_2, d_3, \ldots, d_N]$ is taken into account.

To be invariant to the starting point, the normalized cross-correlation value at each matching pair (the query shape and one of the database shapes) is found from the maximum value calculated at j^{th} time step. It is calculated by:

$$R(Q,D) = \max_{1 \le j \le N} \frac{\sum_{i=1}^{N} q_i d_{i+j}}{\sqrt{\sum_{i=1}^{N} q_i^2} \cdot \sqrt{\sum_{i=1}^{N} d_i^2}} \quad (13)$$

Then the query shape will be assigned to one class in the database shapes corresponding to the maximum value of the normalized cross-correlation.

VI. Experimental Results

Our algorithm has been evaluated on the dataset which consists of 40 classes of shape chosen from MPEG-7 contour shape database CE-1. Each class has 14 different distorted shapes. So there are in total 560 affine distorted shapes in this dataset. The sample of affine distorted shape for class 'bat' is shown in figure 4. These distorted shapes are obtained by transforming the original 2D shape using equation (1). The parameters of affine transformation are shown in table I.

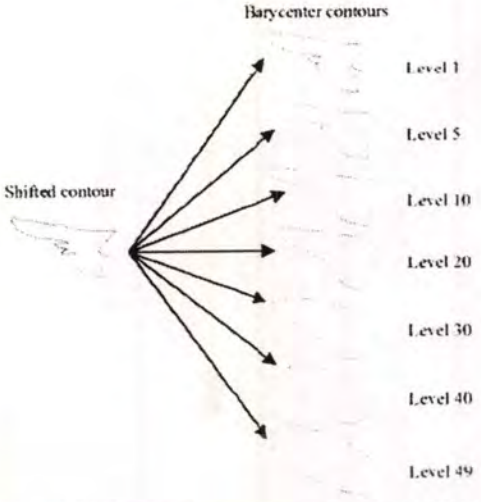


Figure 1: The sample of barycenter contour at level 1, 5, 10, 20, 30, 40, and 49 for class 'bat'

Figure 2: (a) The example of any two sample points used to calculate TAR-2p. (b) The example of any three sample points used to calculate TAR-3p.

In the preprocessing stage, after extracting the shape's contour as the sequences $x(k)$ and $y(k)$. Then these sequences are normalized to have the same length. All the shape's boundary in our experiment are normalized to have 100 points, then the multi-level of barycenter contour is applied by equation (6) where m is the level of barycenter contour ($m = 1, 2, \ldots, 49$).

The retrieval results have been performed using the precision-recall curve which is commonly used by many researchers [2].

Precision is the ability to retrieve top-ranked shapes that are mostly relevant and recall is the ability of the search to find all of the relevant shapes in the same class. The precision and recall can be defined by:

$$\text{Precision} = \frac{\text{Number of retrieved relevant shape}}{\text{Total number of retrieved shape}} \quad (14)$$

$$\text{Recall} = \frac{\text{Number of retrieved relevant shape}}{\text{Total number of relevant shape}} \quad (15)$$

The precision value at each recall is the average of the precision values over all the query shapes Q.

TABLE I: THE PARAMETERS OF AFFINE TRANSFORMATION

Affine Parameters	Shape's Number													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
sh_x	0	1	1.5	0.5	1.5	0.5	1	0	0	0	0	0	0	0
sh_y	0	0	0	0	0	0	0	0.5	0.5	1.4	0.5	1.5	0.5	1
sc	1	1.5	0.6	1.6	1	0.75	2.5	1.35	0.5	1.4	1.8	0.8	1.7	2.1
θ	0	25°	50°	75°	100°	125°	150°	175°	200°	225°	250°	275°	300°	325°

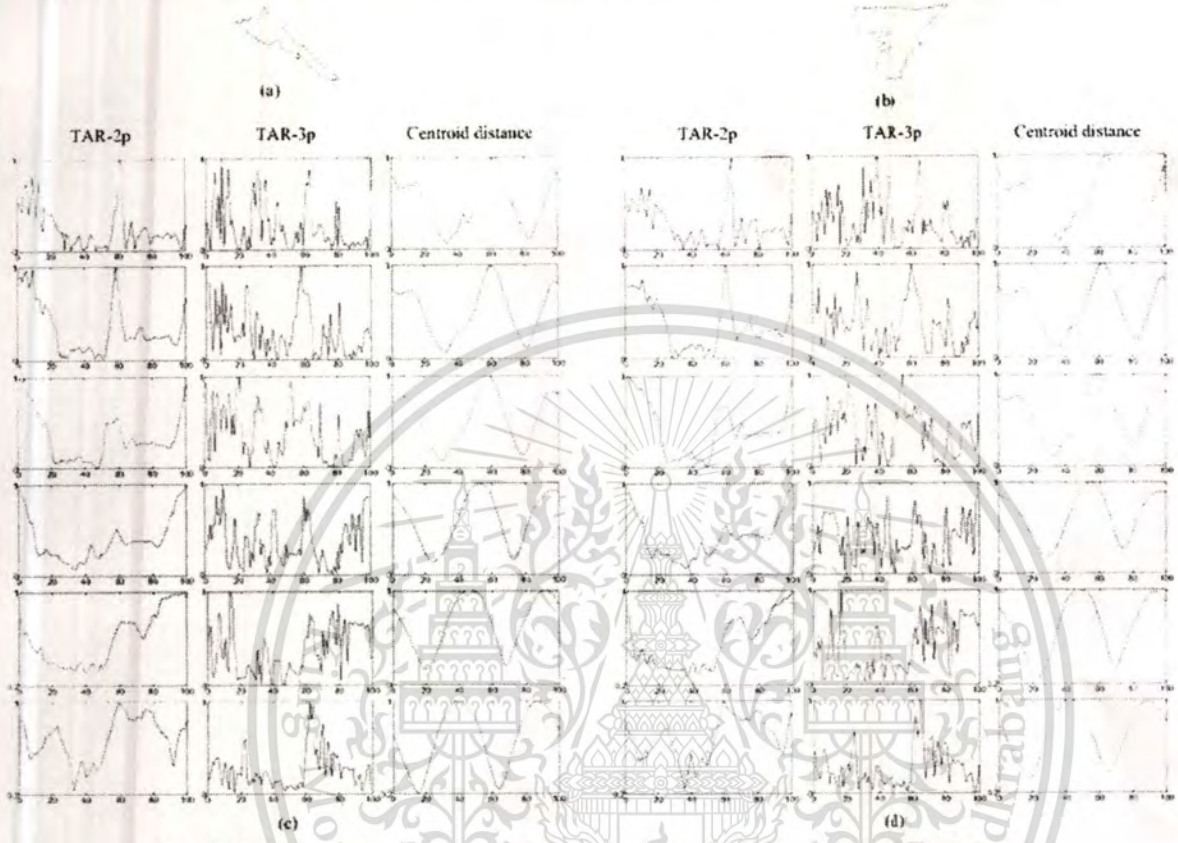


Figure 3: (a)-(b) example of the shape in the same class 'bat', (c)-(d) example of TAR-2p, TAR-3p, and centroid distance of shape (a) and (b) respectively at level 1, 5, 10, 20, 30, and 40

Figure 4: The samples of affine distorted shape for class 'bat'

In this work, the TAR-2p signature at each level from $m = 1$ until $m = 49$ is evaluated by precision-recall curve. Its performance is compared with the other two features, TAR-3p signature and centroid distance signature.

Figure 5 shows the precision-recall curve for TAR-2p, TAR-3p, and centroid distance at level 1, 5, 10, 20, 30, and 40.

From the experimental results, we observe that when the level of barycenter contour is increased, it makes the shape's boundary less sensitive to noise and becoming smoother (ellipse and circle form).

Furthermore, the centroid distance has less information while the TAR-3p has too much detail (noise) and the information of TAR-2p is in-between of the other two signatures as shown in Figure 3. That is why the performance of TAR-2p outperforms the performance of the other two signatures by the precision-recall cure as shown in figure 5.

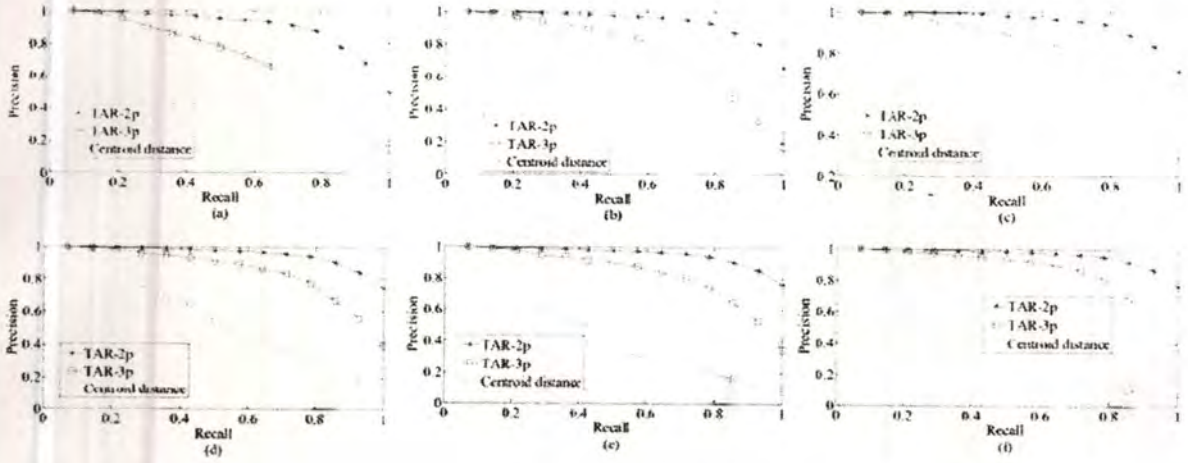


Figure 5: Precision-recall curve forTAR-2p, TAR-3p, and Centroid distance at (a) level 1, (b) level 5, (c) level 10, (d) level 20, (e) level 30, and (f) level 40

The recognition accuracy of the system is shown at the second point of the precision-recall curve of which the first point is the match to itself. By selecting the ten most significant levels of the TAR-2p signature, then combining it by averaging the normalized cross-correlation, we found that the combined TAR-2p (Multi-level) gives higher retrieval accuracy than the single level as shown in figure 6.

VII. Conclusion and Future Works

We propose the multi-level of barycenter contour-based for affine invariant shape classification. The experimental results illustrate that the proposed multi-level of barycenter contour can reduce the noise caused by small boundary distortion and deformation.

Additionally, our approach is invariant to affine transformation and achieves high retrieval efficiency. Research on this work is still going on, improving the algorithm that will further enhance its performance and capable of matching partially occluded shape.

References

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, 2000.

[2] D. Zhang and G. Lu, "Study and evaluation of different Fourier methods for image retrieval," *Image and Vision Computing 23*, 2005, pp. 33-49.

[3] F. Chaker, M.T. Bannour and F. Ghorbel, "A complete and stable set of Affine-Invariant Fourier Descriptors," *Proceedings of the 12th International Conference on Image Analysis and Processing*, 2003, pp. 578-581.

[4] H.S. Yang, S.U. Lee, and K.M. Lee, "Recognition of 2D Object Contour Using Starting-Point-Independent Wavelet Coefficient Matching," *Journal of Visual Communication and Image Representation*, June 1998, vol. 9, no. 2, pp. 171, 181.

[5] Q.M. Tieng and W.W. Boles, "Recognition of 2D object contour using the wavelet transform zero-crossing representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, August 1997, vol. 19, no. 8, pp. 910-916.

[6] Z. Huang and F.S. Cohen, "Affine-Invariant B-Spline Moment for Curve Matching," *Computer Vision and Pattern Recognition*, June 1994, pp. 490-495.

[7] H.K. Kim and J. D. Kim, "Region-based shape descriptor invariant to rotation scale, and translation," *Signal Processing, Image Communication 16*, 2000, pp. 87-93.

[8] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition 37*, 2004, pp. 1-19.

[9] F. Mokhatian and A.K. Mackworth, "A Theory of Multiscale Curvature-Based Shape Representation for Planar Curves," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Aug. 1992, vol 14, no. 8, pp. 789-805.

[10] I. Kunttu, L. Lepisto, J. Rauhamaa, and A. Visa, "Multiscale Fourier Descriptor for Shape-Based Image Retrieval," *Proceeding of the 17th International Conference on Pattern Recognition* 2004, pp. 765-768.

[11] I.El Rube, M. Ahmed and M. Kamel, "Affine Invariant Multiscale Wavelet-base Shape Matching Algorithm," *Proceedings of the first Canadian Conference of Computer and Robot Vision*, May 2004, pp. 217-224.

[12] Thomas.H. Reiss, *Recognizing Planar Object Using Invariant Image Features*, Springer-Verlag, 1991.

[13] L.da.F. Costa and R.M. Cesar Jr, *Shape Analysis and Classification Theory and Pratice*, CRC Press LLC, 2001.

[14] P. Thumwarin and T. Matsuura, "On-line writer recognition for Thai based on velocity of barycenter of pen-point movement," *International Conference on Image Processing* 2004, pp. 889-892.

[15] Kritsana Yawichai and Yuttana Kitjaidure, "A Simple and Efficient Algorithm for Affine Invariant Shape Matching," *International Conference on Engineering Applied Science and Technology*, Bangkok, Thailand, 2007, pp. 246-249.

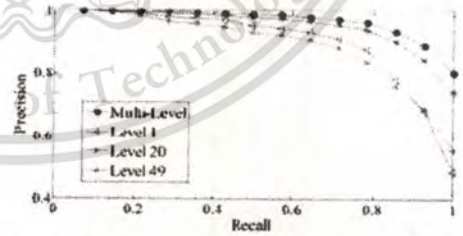
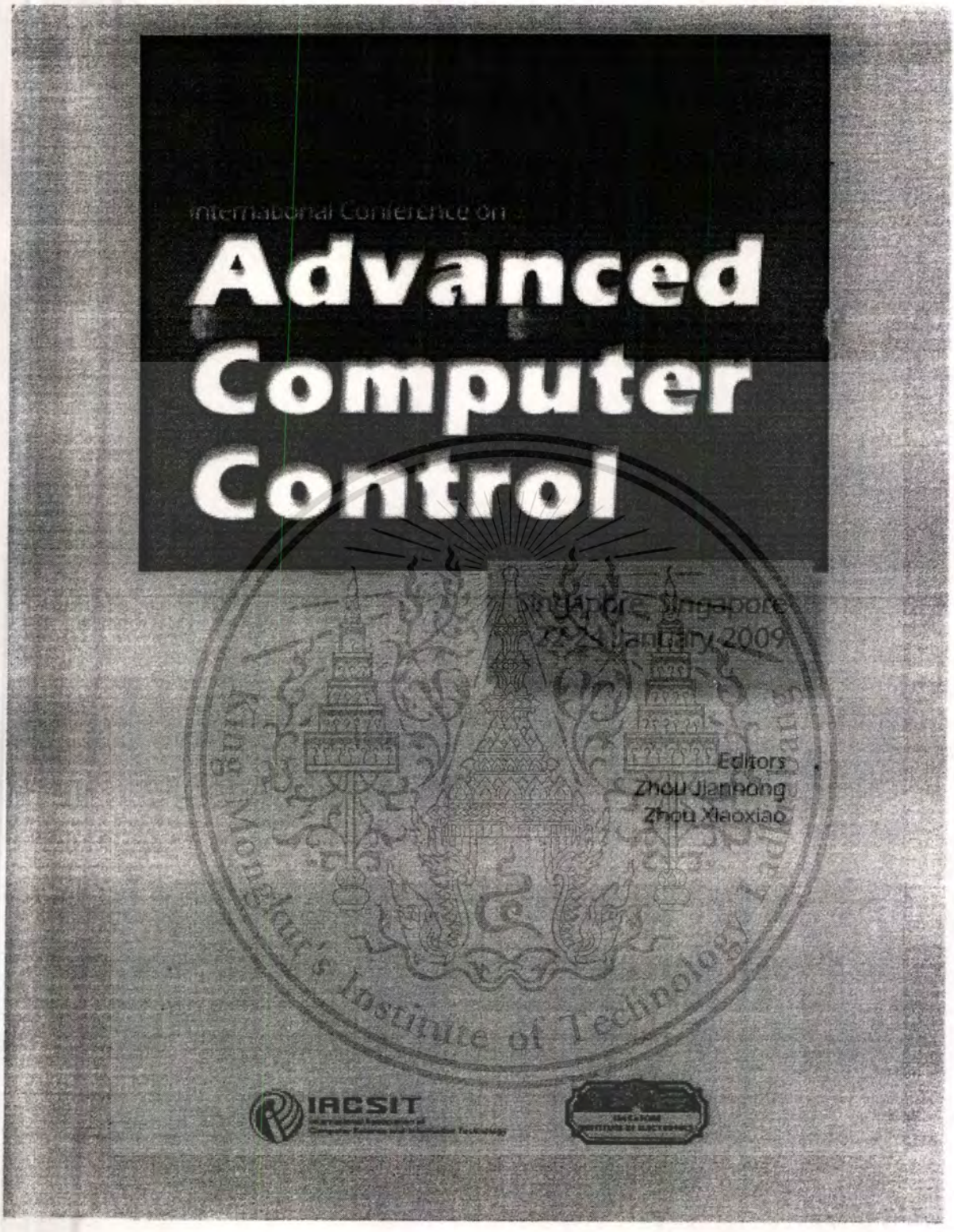


Figure 6: Precision-recall curve for TAR-2p at different level and combination of TAR-2p of the top 10 levels of barycenter contour.



International Conference on Advanced Computer Control

Multi-View Shape Recognition Based on Principal Component Analysis

Kosorl Thourn and Yuttana Kitjaidure
Department of Electronics, Faculty of Engineering
King Mongkut's Institute of Technology, Ladkrabang, Bangkok 10520, Thailand
E-mail: kosal_th@yahoo.com and kkyuttan@kmitl.ac.th

Abstract

In this paper, the principal component analysis (PCA) for multi-view shape recognition is proposed. Our algorithm presents the signed enclosed area signature as the shape representation. In our method, the barycenter contour is used for decomposing the shape boundary into multiscale level. At each scale level, the signed enclosed area signatures are obtained. After that, the principal component analysis (FCA) is used as the recognition strategy. This method is independent to starting point of the contour by exploiting the property of the discrete Fourier transform (DFT). In the experimentation, the various number of the sample shapes are used as the training set and the rest are used as the testing set. The experimental results indicate that the recognition accuracy are high enough even one sample shape per class is used as the training set. As the more training sample shapes per class are used, the higher recognition will be.

1. Introduction

Many methods for 2D shape representation and recognition have been reported. Curvature scale space (CSS) [1], dynamic programming [2], shape context [3], Fourier descriptor [4], [5], and wavelet descriptor [6] are as the example of these approaches.

The principal component analysis (PCA) or Eigen face [7] is known as the suitable data representation that can dramatically reduce the dimensionality of the original image and retain most of the information. It is frequently applied to face recognition. Mark S. Drew et al [8] have introduced shape retrieval with eigen-CSS search by creating the new feature from CSS image and named as the marginal-sum feature vector. Finally, the PCA is used as the classifier. B. Wang and J. A. Bangham [9] have investigated on shape retrieval and the evaluation by using PCA based shape descriptor. G. Tzimiropoulos et al [10] have used PCA for recognition of planar shapes under affine transforms in which PCA is used for encoding the shape boundary to its canonical form and then the enclosed area is used as the shape representation because it is invariant to affine transforms.

In this paper, the PCA is used as the classifier for shape classification and recognition. The barycenter contour is used for decomposing the shape boundary into multiscale level. At each scale level, the enclosed area signatures are obtained. Then the discrete Fourier transform onto that feature to transform from spatial domain into frequency domain and it makes our feature invariant to starting point of the contour. Based on the technique of the PCA or eigenface, the DFT of the signed enclosed area signature at each level are concatenated as a column vector. Finally, the PCA is used as the recognition strategy.

In the experimentation, the various number of the sample shape per class is used as the training set and the rest is used as the testing set. The results show that our algorithm is robust to affine transform and achieved high recognition accuracy.

The remainder of paper is organized as follows: Affine transforms is introduced in section 2, then shape recognition algorithm in section 3 including preprocessing, feature extraction, and recognition stage. Next, the experiment results and discussion are illustrated in section 4. Finally, conclusion is given in section 5.

2. Affine Transforms

The general affine transforms can be represented mathematically with the following equations:

$$\left.\begin{aligned} x' &= ax + by + e \\ y' &= cx + dy + f \end{aligned}\right\} \quad (1)$$

The equation (1) can be written in matrix form:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} = A \cdot \begin{bmatrix} x \\ y \end{bmatrix} + B \quad (2)$$

978-0-7695-3516-6/08 $25.00 © 2008 IEEE
DOI 10.1109/ICACC.2009.69

where x' and y' are the distorted pixel coordinates by affine transforms, x and y are the pixel coordinates of the original 2D shape. The translation is represented by matrix B while scaling, rotation and shear are reflected in matrix A. They are represented by the following matrices:

$$A_{Scaling} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix}, \; A_{Rotation} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix},$$

$$\text{and } A_{Shear} = \begin{bmatrix} 1 & sh_x \\ sh_y & 1 \end{bmatrix}$$

where s_x and s_y are the scaling parameters, sh_x and sh_y are the shear parameters, and θ is the rotation angle parameter. If $s_x = s_y = sc$, $A_{Scaling}$ represents a uniform scaling. A shape is not deformed under rotation, uniform scaling and translation. However, non-uniform scaling and shear contribute to the shape deformation under general affine transforms.

Matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ in equation (2) is given by:

$$A = A_{Scaling} \cdot A_{Rotation} \cdot A_{Shear} \qquad (3)$$

Matrix $B = \begin{bmatrix} e \\ f \end{bmatrix}$ in equation (2) is removed by shifting the shape contour to its center of gravity.

3. Shape Recognition Algorithm

3.1. Preprocessing

In this stage, the shape is segmented from the image. Then, applying technique in [11] to the shape, the boundary of the shape is extracted by starting from the specific selected starting point on the shape's boundary and moving counterclockwise, and the coordinates of each point are obtained denoted as the vector $x(k)$ and $y(k)$. Next, these vectors are sampled to have the same length N, denoted as $P_1, P_2, P_3, ..., P_N$ where $P_i = (x_i, y_i)$ and $i = (1, 2, 3, ..., N)$. The appropriate number of sampled points for the experimentation is 128 points. Finally, the shape's boundary is shifted to the origin by its center of gravity $P_c = (x_c, y_c)$. It is given as following:

$$\left.\begin{aligned} \hat{x}_i &= x_i - x_c \\ \hat{y}_i &= y_i - y_c \end{aligned}\right\} \qquad (4)$$

where $x_c = \frac{1}{N}\sum_{i=1}^{N} x_i$ and $y_c = \frac{1}{N}\sum_{i=1}^{N} y_i$ is the coordinate of center of gravity, $\hat{x}_i$ and $\hat{y}_i$ is the coordinate of shifted shape's boundary.

3.2. Feature Extraction

In this section, we introduce how to create multi-level of barycenter contour. In the geometry, the barycenter of the set of coordinate points is the average of this set of coordinate points.

In this paper, the first level of barycenter contour was adopted from [12] and it is determined from the coordinates of barycenter of triangle composed of $(0,0), (\hat{x}_i, \hat{y}_i)$, and $(\hat{x}_{i+1}, \hat{y}_{i+1})$. The second level is determined from the coordinates of barycenter of points $(0,0), (\hat{x}_i, \hat{y}_i), (\hat{x}_{i+1}, \hat{y}_{i+1})$, and $(\hat{x}_{i+2}, \hat{y}_{i+2})$. So the next level is created by adding one adjacent point more to the previous level. This approach can reduce the noise caused by the distortion and deformation of shape's boundary. If $(\bar{x}_i, \bar{y}_i)$ is the coordinates of barycenter contour, then it is determined by:

$$\left.\begin{aligned} \bar{x}_i^m &= \frac{\hat{x}_i + \hat{x}_{i-1} + ... + \hat{x}_{i+m}}{m+2} \\ \bar{y}_i^m &= \frac{\hat{y}_i + \hat{y}_{i+1} + ... + \hat{y}_{i+m}}{m+2} \end{aligned}\right\} \qquad (5)$$

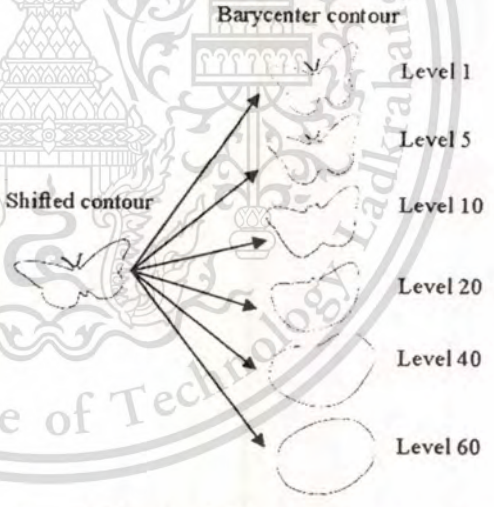


Figure 1. The sample of barycenter contour at level 1, 5, 10, 20, 40, and 60 for class 'butterfly'

where $\hat{x}_t$ and $\hat{y}_t$ are the affine distorted boundary sequences, $\bar{x}_t$ and $\bar{y}_t$ are the barycenter contour, and m is the level of barycenter contour and it take the values from 1 to 63. Figure 1 shows the example of barycenter contour for class 'butterfly' at level 1, 5, 10, 20, 40, and 60.

The signed enclosed area signature is used as the shape representation and it is given as following:

$$A(m,t)=\frac{1}{2}\left(\bar{x}_t^m.\bar{y}_{t+1}^m-\bar{x}_{t+1}^m.\bar{y}_t^m\right) \quad (6)$$

Under affine transforms, the relation between the area of the distorted triangle (A') and the original one (A) can be expressed by:

$$A'=|J|.A \quad (7)$$

where J is the Jacobin between $\{x',y'\}$ and $\{x,y\}$. By equation (1) this Jacobin can be calculated by:

$$J=\begin{vmatrix}\frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y}\\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y}\end{vmatrix}=\begin{vmatrix}a & b\\ c & d\end{vmatrix}=(ad-bc) \quad (8)$$

Then

$$A'=|ad-bc|.A \quad (9)$$

For this reason, the signed enclosed area is invariant to affine transformation (translation, scale, rotation, and shear).

Then to be invariant to starting point, the discrete Fourier transform (DFT) is applied onto the signed enclosed area signature at each level of barycenter contour. It can be calculated as following:

$$F(n)=\frac{1}{N}\sum_{k=1}^{N}f(k).e^{-j2\pi nk/N} \quad (10)$$

for $n=1,2,...,N$ and $F(n)$ are the transform coefficient of $f(k)$. By taking only the transform coefficients $|F(n)|$ and they are normalized to be invariant to scale by dividing with the DC component coefficient $|F(0)|$.

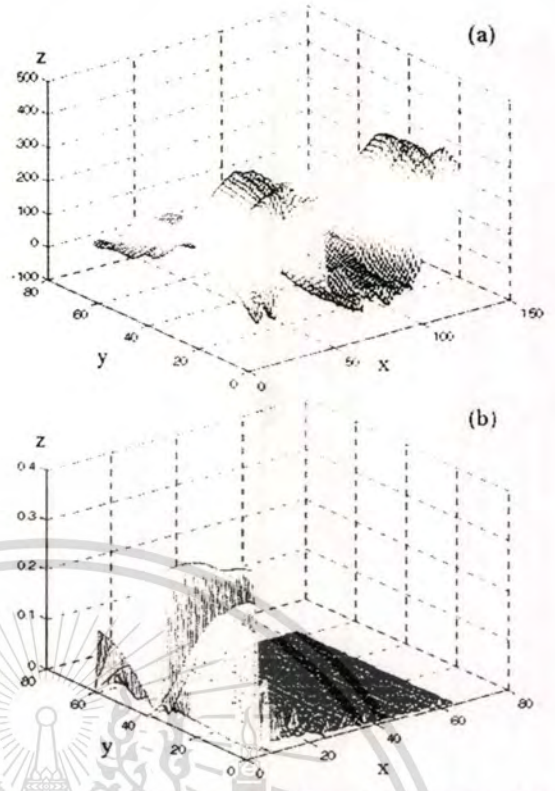


Figure 2. The signed enclosed area signature of the butterfly in figure 1 of all levels of barycenter contour and its normalized Fourier coefficients respectively where x, y, and z axis are represented as the sampling point index, the level of barycenter contour, and the value of signed enclosed area (a) and its normalized Fourier coefficients (b) respectively.

3.3. Recognition Stage

Let A_j be the column vector of shape feature and suppose that we have a set of M shape feature as the training sample set $j=(1,2,3,....,M)$. The one dimensional shape's feature column vectors A are formed by concatenating the DFT coefficients of the signed enclosed area signature of all levels of barycenter contour. The total covariance matrix C can be formed by:

$$C=\frac{1}{M}\sum_{j=1}^{M}(A_j-\mu)(A_j-\mu)^T \quad (11)$$

where $\mu = \frac{1}{M}\sum_{j=1}^{M} A_j$ is the mean of all training sample sets. To avoid the time-consuming calculation of the C's eigenvector directly, the singular value decomposition (SVD) technique is applied to determine the required eigenvector.

Let $Q = [A_1 - \mu, A_2 - \mu, ..., A_M - \mu]$ then C can be represented by:

$$C = \frac{1}{M} Q Q^T \quad (12)$$

Let us form the matrix $R = Q^T . Q$, which is $M \times M$ semi-positive definite matrix. Let us calculate the orthogonal eigenvectors $U = [u_1, u_2, ..., u_r]$ of R corresponding to r largest eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r$. Then, by SVD theorem, the orthogonal eigenvectors $w_1, w_2, ..., w_M$ of C corresponding to r largest eigenvalues $\lambda_1, \lambda_2, ... \lambda_r$ are:

$$w_j = U^T (A_j - \mu) \quad (13)$$

The parameter r is defined as the minimum value of the expression below:

$$\frac{\lambda_1 + \lambda_2 + ... + \lambda_r}{\lambda_1 + \lambda_2 + ... + \lambda_r + ... + \lambda_M} \geq 0.95 \quad (14)$$

The unknown shape will be assigned to the class C_i by the minimum Euclidian distance:

$$\min_{1 \leq j \leq M} \| w - w_j \| \quad (15)$$

4. Experimental Results and Discussion

Our algorithm has been evaluated on the dataset which consists of 40 classes of shapes chosen from MPEG-7 contour shape database CE-1. Each class has 14 different distorted shapes. So there are in total 560 affine distorted shapes in this dataset.

The sample of affine distorted shape for class 'butterfly' is shown in figure 3. These distorted shapes are obtained by transforming the original 2D shape using equation (1). The parameters of affine transforms are shown in table 1.

The shape recognition is evaluated by randomly selecting the various number of training sample sets (from 1 sample shape until 13 sample shapes) and the rest are used as the testing sets.

Figure 3. The sample of affine distorted shapes of the class of butterfly.

Table 2. The recognition rate at various training sample set of the enclosed area and the centroid distance

Training set	Testing set	Signed enclosed Area	Centroid Distance
1	13	85.51 %	35.93 %
2	12	92.96 %	48.33 %
3	11	95.54 %	57.56 %
4	10	96.55 %	63.83 %
5	9	97.44 %	68.35 %
6	8	98.04 %	72.10 %
7	7	98.06 %	75.32 %
8	6	98.51 %	78.36 %
9	5	98.58 %	80.85 %
10	4	98.73 %	81.65 %
11	3	98.75 %	83.72 %
12	2	98.80 %	88.08 %
13	1	99.08 %	88.75 %

Table 1. The parameters of affine transforms

Affine Parameters	Shape's Number													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
sh_x	0	1	1.5	0.5	1.5	0.5	1	0	0	0	0	0	0	0
sh_y	0	0	0	0	0	0	0	0.5	0.5	1.4	0.5	1.5	0.5	1
sc	1	1.5	0.6	1.6	1	0.75	2.5	1.35	0.5	1.4	1.8	0.8	1.7	2.1
θ	0	25°	50°	75°	100°	125°	150°	175°	200°	225°	250°	275°	300°	325°

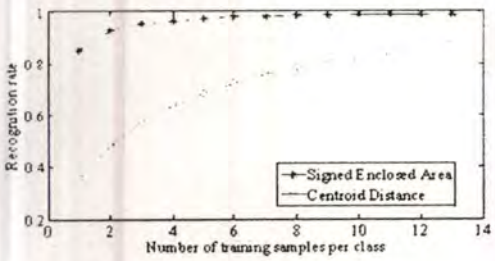


Figure 4: The comparison of the recognition rate between the signed enclosed area and the centroid distance in the number of training sample set.

To evaluate our algorithm, the comparison of the signed enclosed area signature and the centroid distance signature are used.

The recognition rate of our algorithm in term of the number of training samples are given in the table 2 and are depicted in Figure 4. The experimental results show that the recognition of the signed enclosed area is higher than the centroid distance. Even though the recognition rate of the signed enclosed is seem to be stable at 8 training samples per class. We also found that the recognition rates of both signatures are increased proportional to the number of training samples.

5. Conclusion

We proposed the principal component analysis (PCA) for multi-view shape recognition. In the method, the barycenter contour was used for reducing the noise and small boundary distortion. It was associated with the enclosed area signature which is invariant to affine transforms, to form the multi-level shape representation. The PCA or eigenface was used as the classifier. However, this classifier was not invariant to starting point on the contour so the DFT was applied.

The experimental results illustrate that the proposed method achieved high enough recognition accuracy even using one sample per class as the training set. When the number of training samples was increased, the recognition accuracy was increased as well.

References

[1] F. Mokharian and A.K. Mackworth, "A Theory of Multiscale Curvature-Based Shape Representation for Planar Curves", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Aug. 1992, vol.14, no. 8, pp. 789-805.

[2] E.G.M. Petrakis, A. Diplaros, and E. Milios, "Matching and retrieval of distorted and occluded shape using dynamic programming", *IEEE Trans. on PAMI 24*, 2002, no. 11, pp. 1501-1516.

[3] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts", *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 2002, no. 24, pp. 509-522.

[4] D. Zhang and G. Lu, "Study and evaluation of different Fourier methods for image retrieval", *Image and Vision Computing 23*, 2005, pp. 33-49.

[5] I. Kunttu, L. Lepisto, J. Rauhamma, and A. Visa, "Multiscale Fourier Descriptor for Shape-Based Image Retrieval", *Proceeding of the 17th International Conference on Pattern Recognition*, 2004, pp. 765-768.

[6] I.El Rube, M. Ahmed, and M. Kamel, "Affine Invariant Multiscale Wavelet-base Shape Matching Algorithm", *Proceeding of the first Canadian Conference of Computer and Robot Vision*, May 2004, pp. 217-224.

[7] F. Chichizola, Laura .D.G, Armando .D.G, and M. Naiouf, "Face Recognition: Reduced Image Eigenface Method", *The 47th International Symposium ELMAR-2005*, Zadar, Croatia, June 2005, pp. 159-162.

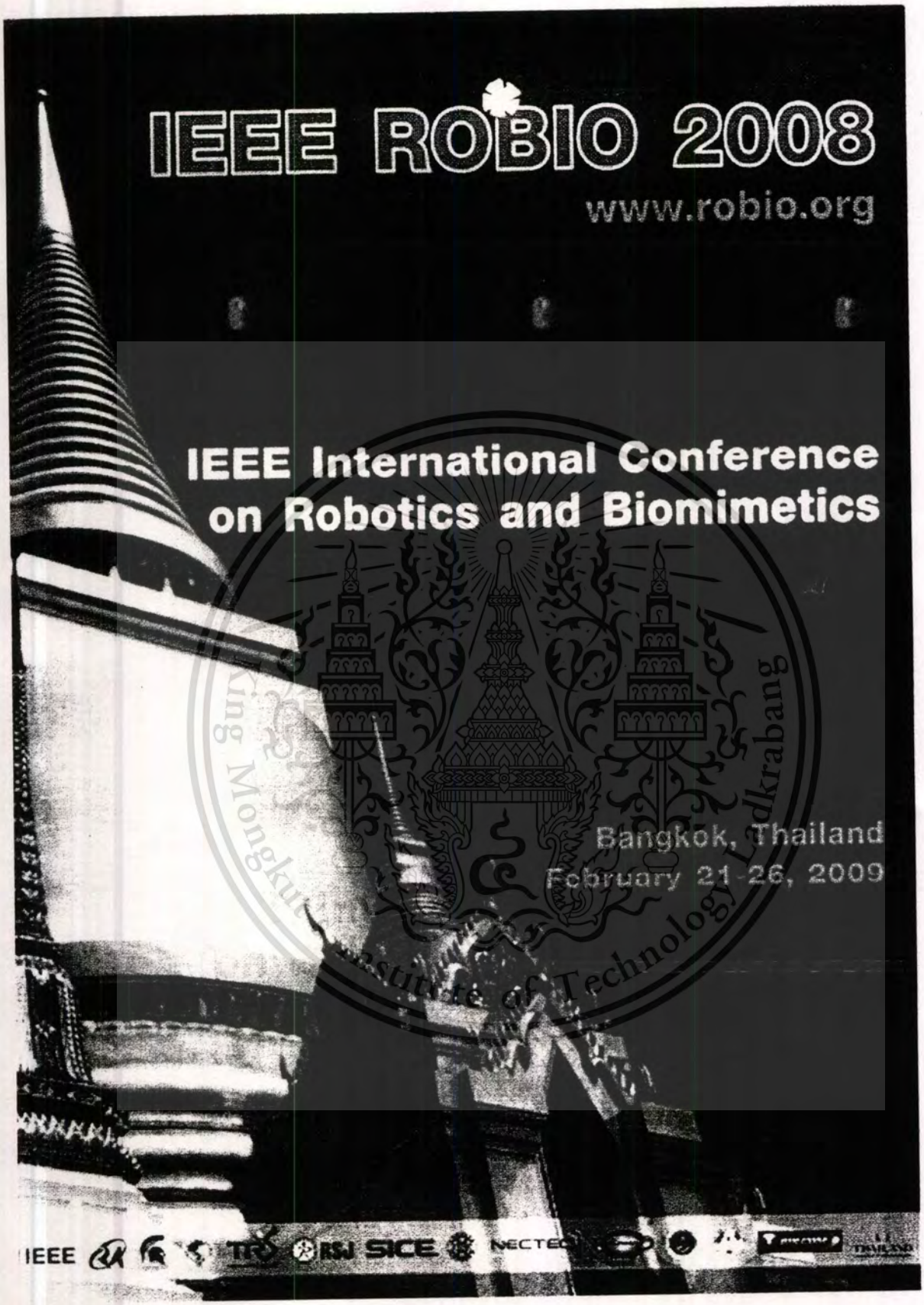
[8] M.S. Drew, T.K. Lee, and A. Rova, *Shape Retrieval with Eigen-CSS Search*, Technical Report, School of Computing Science, Simon Fraser University, SFU-CMPT-TR-2005-07, Feb.24, 2005.

[9] B. Wang and J.A. Bangham, "PCA Based Shape Descriptors for Shape Retrieval and the Evaluations", *International Conference on Computational Intelligence and Security*, 2006, pp. 1401-1406.

[10] G. Tzimiropoulos, N. Mitianoudis, and T. Stathaki, "Robust Recognition of Planar Shapes Under Affine Transforms Using Principal Component Analysis", *IEEE Signal Processing Letters*, vol. 14, no. 10, October 2007, pp. 723-726.

[11] L.da.F. Costa and R.M. Cesar Jr, *Shape Analysis and Classification: Theory and Practice*, CRC Press LLC, 2001.

[12] P. Thumwarin and T. Matsuura, "On-line writer recognition for Thai based on velocity of barycenter of the pen-point movement", *International Conference on Image Processing*, 2004, pp. 889-892.



Proceedings of the 2008 IEEE
International Conference on Robotics and Biomimetics
Bangkok, Thailand, February 21 -26, 2009

Robust Multi-TAR of Barycenter Contour for Multiple Views Shape Matching and Retrieval

Kosorl Thoum and Yuttana Kitjaidure

Department of Electronics, Faculty of Engineering
King Monkut's Institute of Technology, Ladkrabang
Bangkok 10520, Thailand

kosal_th@yahoo.com, kkyuttan@kmitl.ac.th

Abstract **– In this paper, we present a one dimensional descriptor for the two dimensional object silhouettes associated with each level of barycenter contour for multiple views shape matching and retrieval. Firstly, the barycenter contour is applied onto the shape contour. Then the averaging multi-triangle area representation (AMTAR) at each level of barycenter contour is computed as the shape descriptor. Finally, to classify among the shapes, the normalized cross-correlation function is used. Our algorithm is implemented on database consisting of 560 affine distorted shapes, chosen from MPEG-7 contour shape database CE-1. The experimental results indicate that the proposed method is invariant to affine transformation, robustness to the noise, and achieves high retrieval performance.**

Index Terms **–** ***Affine transformation, Barycenter Contour, Averaging Multi-Triangle Area Representation (AMTAR).***

I. Introduction

In real world situation when we capture the planar object in multiple viewpoints, we normally obtain differently deformed 2D shapes. These deformations can be modeled by affine transform, if the viewpoints are sufficiently far away. The important problem is how to increase the recognition performance of robot vision as well as human do. This research presents an approach to solve the problem. Typically, shape representation plays a major role in shape analysis and matching system. Many approaches since the past decade have been reported for shape representation and matching. Curvature scale space (CSS) [1], dynamic programming [2], shape context [3], B-spline moment [4], Fourier descriptor [5], [6], and wavelet descriptor [7] are as the example of these approaches.

These approaches can be classified into contour-based and region-based. In our work, the contour-based method has been chosen because region-based method [8] uses more data involved with more computation and storage. There are many shape representations introduced in the literature [5], [9].

Recently, growing research interest has been focused on the multi-scale technique for shape representation and matching in order to improve the accuracy and robustness to the moderate amount of noise. F. Mohkarian and F.K. Mackworth [1] have presented a multi-scale curvature-based descriptor technique for classification of planar curve. As known early, Fourier descriptor cannot provide the multiple resolutions for shape representation. However, I. Kunttu et al. [6] have applied the Fourier transform to the coefficients of the multi-scale wavelet transform in order to create a new Fourier descriptor that can be presented in multiple resolutions. In the literature [7], I.El Rube, M. Ahmed, and M. Kamel have examined on affine invariant shape classification based on dyadic wavelet transform. Both, Fourier-based and Wavelet-based are used to transform shape's boundary from spatial domain into spectral domain. In this work, the new multi-scale has been proposed in spatial domain that is called as the multi-level of barycenter contour. This approach can reduce the noise caused by deformation and distortion of boundary. Then, the Averaging Multi-Triangle Area Representation (AMTAR) is used as the shape representation by averaging the triangle area representation (TAR) of all Triangle Side Lengths (TSL). Finally, the matching strategy based on normalized cross correlation function is used to measure the similarity among the shapes. The results show that our algorithm is robust to affine transformation and moderate amount of noise, and achieves high retrieval accuracy.

The remainder of paper is organized as follows: Affine transformation is introduced in section II, then shape matching algorithm including preprocessing, feature extraction, and matching stage in section III. Next, the experiment results and discussion illustrate in section IV. Finally, conclusion and future works are given in section V.

II. Affine Transformation

The general affine transformation can be represented mathematically with the following equations:

$$\left.\begin{aligned} x' &= ax + by + e \\ y' &= cx + dy + f \end{aligned}\right\} \quad (1)$$

The equation (1) can be written in matrix form:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} = A \begin{bmatrix} x \\ y \end{bmatrix} + B \quad (2)$$

Where x' and y' are the distorted pixel coordinates by affine transformation, x and y are the pixel coordinates of original 2D shape. The translation is represented by matrix

978-1-4244-2679-9/08/$25.00 ©2008 IEEE

B while scaling, rotation and shear are reflected in matrix A. They are represented by the following matrices:

$$A_{Scaling} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix},\ A_{Rotation} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix},\ \text{and}$$

$$A_{Shear} = \begin{bmatrix} 1 & sh_x \\ sh_y & 1 \end{bmatrix}$$

Where s_x and s_y are the scaling parameters, sh_x and sh_y are the shear parameters, and θ is the rotation angle parameter.

If $s_x = s_y = sc$, $A_{Scaling}$ represents a uniform scaling. A shape is not deformed under rotation, uniform scaling and translation. However, non-uniform scaling and shear contribute to the shape deformation under general affine transforms.

Matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ in equation (2) is given by:

$$A = A_{Scaling} \cdot A_{Rotation} \cdot A_{Shear} \qquad (3)$$

Matrix $B = \begin{bmatrix} e \\ f \end{bmatrix}$ in equation (2) is removed by shifting the shape contour to its centre of gravity.

III. Shape Matching Algorithm

A. Preprocessing

In this stage, the shape is segmented from the image. Then, applying technique in [10] to the shape, the boundary of shape is extracted by starting from the specific selected point on the shape's boundary and moving counterclockwise, and the coordinates of each point are obtained denoted as the vector $x(k)$ and $y(k)$. Next, these vectors are sampled to have the same length N, denoted as $P_1, P_2, P_3, ..., P_N$ where $P_i = (x_i, y_i)$ and $i = (1,2,3,..., N)$. Finally, the shape's boundary is shifted to the origin by its center point $P_c = (x_c, y_c)$. It is given as following:

$$\left.\begin{aligned} \hat{x}_i &= x_i - x_c \\ \hat{y}_i &= y_i - y_c \end{aligned}\right\} \qquad (4)$$

Where $x_c = \frac{1}{N}\sum_{i=1}^{N} x_i$ and $y_c = \frac{1}{N}\sum_{i=1}^{N} y_i$ is the coordinates of center point, $\hat{x}_i$ and $\hat{y}_i$ is the coordinates of shifted shape's boundary.

B. Feature Extraction

In this section, we introduce how to get multi-level of barycenter contour. In the geometry, the barycentre of the set of coordinate points is the average of this set of coordinate points.

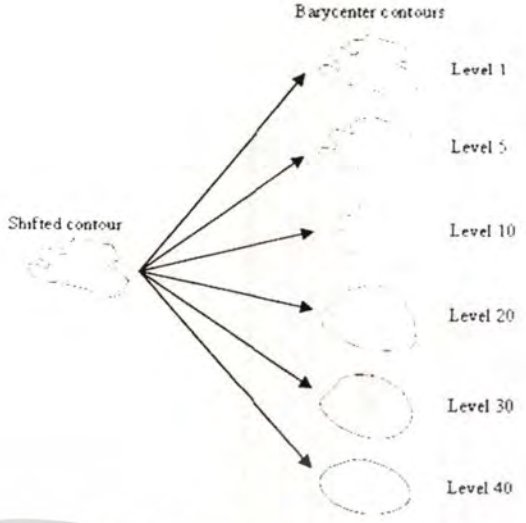


Figure 1. Sample of barycenter contour at level 1, 5, 10, 20, 30, and 40 for class 'teddy'

Figure 2. Example of any two sampled points used to calculate TAR-2p

In this paper, the first level of barycentre contour was adopted from [11] and it is determined from the coordinates of barycenter of triangle composed of $(0,0)$, (x_i, y_i), and (x_{i+1}, y_{i+1}). The second level is determined from the coordinates of barycenter of points $(0,0)$, (x_i, y_i), (x_{i+1}, y_{i+1}), and (x_{i+2}, y_{i+2}). So the next level is created by adding one adjacent point more to the past level. This approach can reduce the noise caused by the distortion and deformation of shape's boundary. If $(\tilde{x}_i, \tilde{y}_i)$ is the coordinates of barycenter contour, then it is determined by:

$$\left.\begin{aligned} \tilde{x}_i^m &= \frac{\hat{x}_i + \hat{x}_{i-1} + ... + \hat{x}_{i+m}}{m+2} \\ \tilde{y}_i^m &= \frac{\hat{y}_i + \hat{y}_{i+1} + ... + \hat{y}_{i+m}}{m+2} \end{aligned}\right\} \qquad (5)$$

Where $\hat{x}_i$ and $\hat{y}_i$ are the affine distorted boundary sequences, $\tilde{x}_i$ and $\tilde{y}_i$ are the barycenter boundary sequences, and m is the level of barycenter contour. Figure 1 shows the example of multi-level of barycenter contour for class 'teddy' at level 1, 5, 10, 20, 30, and 40.

Shape signature is mathematically represented by one dimensional function to describe shape. Many shape signatures have been presented in the literature [9].

1) AMTAR Signature

The AMTAR signature is proposed in this work, improving the TAR signature [12]. The TAR at any TSL is computed as the non-absolute triangle area of the three points on the contour $\tilde{P}_i = (\tilde{x}_i, \tilde{y}_i)$ and the other two points $\tilde{P}_{i-l} = (\tilde{x}_{i-l}, \tilde{y}_{i-l})$ and $\tilde{P}_{i+l} = (\tilde{x}_{i+l}, \tilde{y}_{i+l})$ further from point P_i by l sampled points (l is the index of triangle side length) as shown in Figure 2. It is calculated by:

$$TAR(m,i,l) = \frac{1}{2}\begin{vmatrix} \tilde{x}_{i-1}^m & \tilde{y}_{i-1}^m & 1 \\ \tilde{x}_i^m & \tilde{y}_i^m & 1 \\ \tilde{x}_{i+1}^m & \tilde{y}_{i+1}^m & 1 \end{vmatrix} \quad (6)$$

Then, the AMTAR signature is computed as the average of TAR of all TSL. It is defined by:

$$AMTAR(m,i) = \frac{1}{L}\sum_{l=1}^{L} TAR(m,i,l) \quad (7)$$

Where L is the number of triangle side length.

Under affine transformation, the relation between both areas (the distorted triangle and the original one) can be expressed by:

$$A' = |J|A \quad (8)$$

Where A' is the distorted triangle area by affine transformation, A is the original triangle area, and J is the Jacobin between $\{x', y'\}$ and $\{x, y\}$. By equation (1) this Jacobin can be calculated by:

$$J = \begin{vmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = (ab - cd) \quad (9)$$

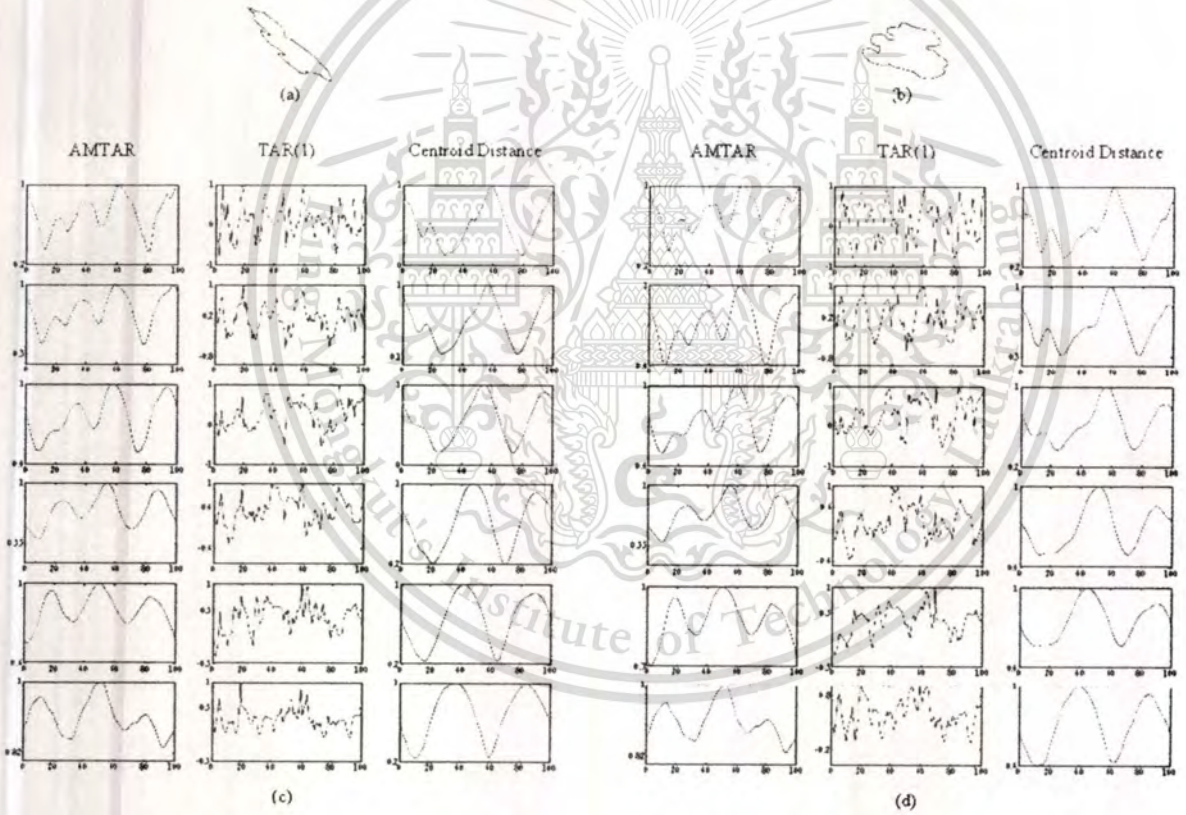


Figure 3 (a)-(b) show the examples of the shape 'teddy'. (c)-(d) is the examples of AMTAR, TAR at TSL = 1, and centroid distance of shape (a) and (b) respectively at level 1, 5, 10, 20, 30, and 40

Then $$A' = |ab - cd| A \quad (10)$$

For this reason, the TAR is invariant to affine transformation (translation, scale, rotation, and shear), so the AMTAR is invariant to affine transformation as well.

2) Centroid Distance Signature

The centroid distance is presented in many literatures [5] .[9]. It is invariant to translation, circularly shifted by rotation, and linearly changed by scaling. It is computed as the distance between the shape's boundary point and its center point. It can be calculated by:

$$r(m,t) = \sqrt{\left[\tilde{x}_t^m\right]^2 + \left[\tilde{y}_t^m\right]^2} \quad (11)$$

The AMTAR and the centroid distance representations are normalized by dividing equation (7), and (11) by its corresponding maximum value at each level of barycenter contour. Figure 3(a)-(b) illustrate the example of the shape in the same class 'teddy' and Figure 3(c)-(d) illustrates AMTAR, TAR at TSL = 1, and centroid distance of 'teddy' at level 1, 5, 10, 20, 30, and 40. The horizontal axis represents the location of barycenter contour points and the vertical axis represents the AMTAR, TAR at TSL = 1, and centroid distance by column respectively.

C. Matching Stage

In this stage, the normalized cross-correlation function is computed to measure the similarity between the query shapes denoted as $Q = [q_1, q_2, q_3, \ldots, q_N]$ and the database shapes denoted as $D = [d_1, d_2, d_3, \ldots, d_N]$. To be invariant to the starting point, the normalized cross-correlation value at each matching pair (the query shape and one of database shapes) is found from the maximum value calculated at j^{th} time step. It is calculated by:

$$R(Q,D) = \max_{1 \le j \le N} \frac{\sum_{i=1}^{N} q_i d_{i+j}}{\sqrt{\sum_{i=1}^{N} q_i^2} \cdot \sqrt{\sum_{i=1}^{N} d_i^2}} \quad (12)$$

Figure 4 Sample of affine distorted shape for class 'teddy'

Then the query shape will be assigned to one class in the database shapes corresponding to the maximum value of the normalized cross-correlation.

IV. Experimental Results and Discussions

Our algorithm has been performed on the dataset which consists of 40 classes of shape chosen from MPEG-7 contour shape database CE-1. Each class has 14 difference distorted shapes. So there are in total 560 affine distorted shapes in this dataset. The sample of affine distorted shape for class 'teddy' is shown in figure 4. These distorted shapes are obtained by transforming the original 2D shape using equation (1). The parameters of affine transformation are shown in table I.

In the preprocessing stage, after extracting the shape's contour as the sequences $x(k)$ and $y(k)$. Then these sequences are normalized to have the same length.

All the shape's boundary in our experiment are normalized to have 100 points, then the each level of barycenter contour is applied by equation (5) where m is the level of barycenter contour and it takes value from 1 until 40 ($m = 1, 2, \ldots, 40$). Then shape representation AMTAR is computed from average of TAR of all TSL by using equation (7) and the TSL in this experiment takes value from 1 until 49.

The retrieval results have been performed using the precision-recall curve which is commonly used by many researchers [5].

Precision is the ability to retrieve top-ranked shapes that are mostly relevant and recall is the ability of the search to find all of the relevant shapes in the same class. The precision and recall can be defined by;

$$\text{Precision} = \frac{\text{Number of retrieved relevant shape}}{\text{Total number of retrieved shape}} \quad (13)$$

$$\text{Recall} = \frac{\text{Number of retrieved relevant shape}}{\text{Total number of relevant shape}} \quad (14)$$

TABLE I: The Parameters of Affine Transformation

Affine Parameters	Shape's Number													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
sh_x	0	1	1.5	0.5	1.5	0.5	1	0	0	0	0	0	0	0
sh_y	0	0	0	0	0	0	0	0.5	0.5	1.4	0.5	1.5	0.5	1
sc	1	1.5	0.6	1.6	1	0.75	2.5	1.35	0.5	1.4	1.8	0.8	1.7	2.1
θ	0	25°	50°	75°	100°	125°	150°	175°	200°	225°	250°	275°	300°	325°

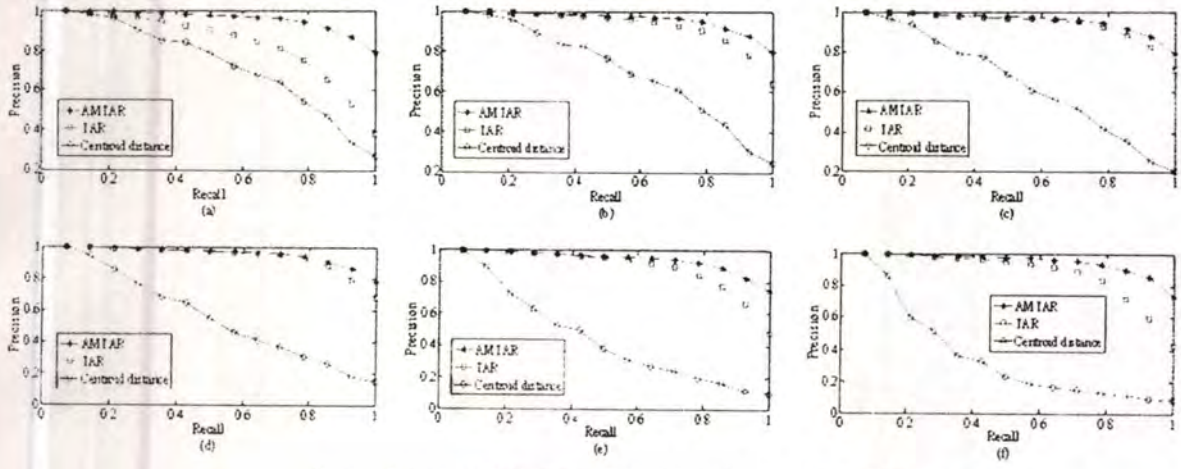


Figure 5 Precision-recall curve for AMTAR, TAR at TSL = 1, and Centroid distance at (a) level 1, (b) level 5, (c) level 10, (d) level 20, (e) level 30, and (f) level 40

The precision value at each recall is the average of the precision values over all the query shape Q.

We evaluate the effect of affine transformation on AMTAR feature and comparing the results to TAR at TSL = 1 and the centroid distance feature.

From the experimental results, we observe that when the level of barycenter contour is increased, it makes the shape's boundary less sensitive to noise and becoming smoother (ellipse and circle form). Typically, the centroid distance signature has lost the important data while the contour is smoothed. However, TAR still has much detail (noise) and the proposed feature, AMTAR, is in-between of the information of these two features as shown in Figure 3. That is why the performance of AMTAR outperforms as shown in Figure 5 by precision-recall curve.

V. Conclusion and Future Works

We have proposed a new multi-scale shape representation for multiple view shape matching and retrieval. The representation uses the AMTAR computing as the average of TAR of all triangle side length and associating with the barycentre contour. The TAR representation is computed as the area of the triangle formed by each three equal apart points on the shape boundary. The experimental results illustrate that the proposed multi-scale shape representation can reduce the small boundary distortion and deformation.

Additionally, our approach is invariant to affine transformation and achieves high retrieval efficiency. Research on this work is still going on, improving the algorithm that will further enhance its performance and capable of matching partially occluded shape.

References

[1] F. Mokhtarian and A.K. Mackworth, "A Theory of Multiscale Curvature-Based Shape Representation for Planar Curves," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Aug. 1992, vol. 14, no. 8, pp. 789-805.

[2] E.G.M. Petrakis, A. Diplaros, and E. Milios, "Matching and retrieval of distorted and occluded shape using dynamic programming," *IEEE Trans.on PAMI 24*, 2002, no. 11, pp. 1501-1516.

[3] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 2002, no. 24, pp. 509-522.

[4] Z. Huang and F.S. Cohen, "Affine-Invariant B-Spline Moment for Curve Matching," *Computer Vision and Pattern Recognition*, June 1994, pp. 490-495.

[5] D. Zhang and G. Lu, "Study and evaluation of different Fourier methods for image retrieval," *Image and Vision Computing 23*, 2005, pp. 33-49.

[6] I. Kunttu, L. Lepisto, J. Rauhamma, and A. Visa, "Multiscale Fourier Descriptor for Shape-Based Image Retrieval," *Proceeding of the 17th International Conference on Pattern Recognition* 2004, pp. 765-768.

[7] I.El Rube, M. Ahmed and M. Kamel, "Affine Invariant Multiscale Wavelet-base Shape Matching Algorithm," *Proceedings of the first Canadian Conference of Computer and Robot Vision*, May 2004, pp. 217-224.

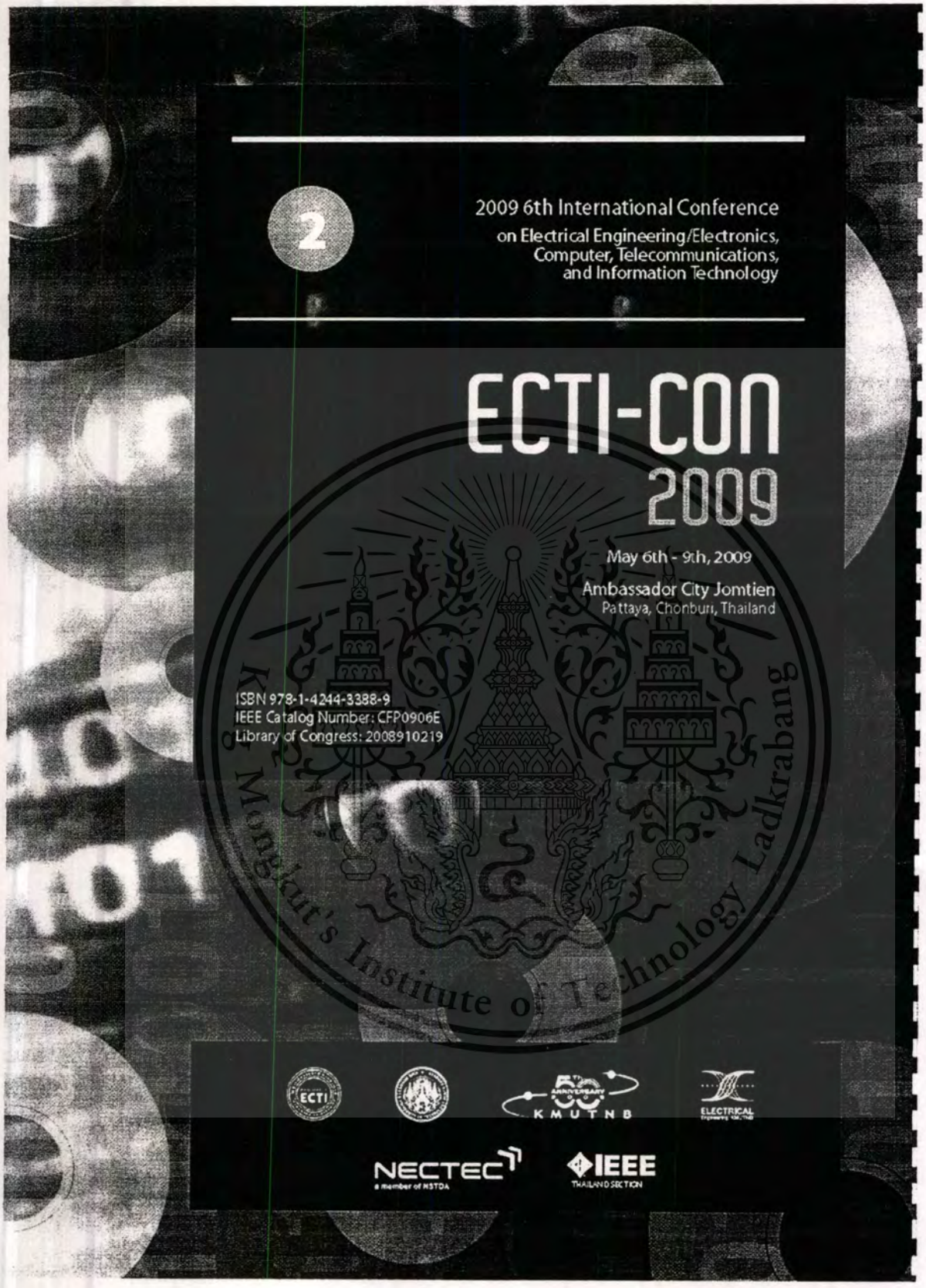
[8] H.K. Kim and J.D. Kim, "Region-based shape descriptor invariant to rotation, scale, and translation," *Signal Processing: Image Communication 16*, 2000, pp. 87-93.

[9] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition 37*, 2004, pp. 1-19.

[10] L.da F. Costa and R.M. Cesar Jr, *Shape Analysis and Classification: Theory and Pratice*, CRC Press LLC, 2001.

[11] P. Thumwarin and T. Matsuura, "On-line writer recognition for Thai based on velocity of barycenter of pen-point movement," *International Conference on Image Processing*, 2004, pp. 889-892.

[12] Kritsana Yawichai and Yuttana Kitjaidure, "A Simple and Efficient Algorithm for Affine Invariant Shape Matching," *International Conference on Engineering Applied Science and Technology*, Bangkok, Thailand, 2007, pp. 246-249.



Robust 2D Shape Matching and Retrieval Using Fisher Barycenter Contour

Kosorl Thourn[1], Yuttana Kitjaidure[1], and Shozo Kondo[2]
[1]Department of Electronics, Faculty of Engineering
King Mongkut's Institute of Technology, Ladkrabang, Bangkok 10520, Thailand
E-mail: kosal_th@yahoo.com, kkyuttan@kmitl.ac.th
[2]Department of Information and Telecommunication Engineering, Tokai University
E-mail: kondo@keyaki.cc.u-tokai.ac.jp

***Abstract*-In this paper, the algorithm for 2D shape matching and retrieval is developed by using Fisher Barycenter Contour (FBcC). First, the shape is represented into 3D format using the signed enclosed area at each scale level of Barycenter Contour (BcC). Because of high dimension of the feature representation, the Eigen Barycenter Contour (EBcC) is applied for dimensionality reduction. Then, the Fisher Barycenter Contour (FBcC) is used for making discrimination. Finally, the similarity is measured using the normalized cross correlation. The experimentation is tested on MPEG-7 contour shape database CE-1 part B of 1400 image shapes. The experimental results illustrate that our approach gives very high retrieval efficiency (or Bulls-eye test) of 89.60% and 98.62% using two parameters, shape signature and its power spectrum respectively, when comparing with all the existing methods.**

I. INTRODUCTION

Several properties of the objects have been used for recognition and categorization such as shape, color, texture, brightness, and others. However, the shape is considered as the most promising in a system for object recognition, matching, registration, and analysis. In addition, the users are interested in retrieval by shape rather than by the others as indicated in [1].

There are many applications relating to the shape recognition such as security systems, medical image analysis, computer vision, robot vision and others.

Many shape analysis techniques have been proposed over the past decade. The Accurate Retrieval based on Phase (WARP) [2], shape similarity measure based on correspondence of Visual Parts (VP) [3], Curve Edit Distance (CED) [4], Curvature Scale Space (CSS) [5], Beam Angle Statistics (BAS) [6], Multiscale Convexity Concavity representation (MCC) [7], Inner Distance Shape Context (IDSC+DP) [8], symbolic representation [9], and Triangle area representation with dynamic programming (DSW+Global) [10] are the examples of the shape recognition and classification methods. However, the current techniques have not given the satisfactory solution yet. This reason has motivated the researchers to develop the algorithm for reaching the efficient and effective results.

The Principal Component Analysis (PCA) and Fisher Linear Discriminant Analysis (FLDA) [11], [12] are known as the suitable data representation that can dramatically reduce the dimensionality of the original image and retain most of the information. It is frequently applied to face recognition. Some literatures applied PCA onto 2D closed boundary shape for classification and dimensionality reduction.

M.S. Drew et al. [13] have introduced the Eigen-CSS search for shape retrieval. A new feature vector for shape representation has been created and called the marginal sum feature vector which is composed of row-sum and column-sum of the raw CSS image. Then, the PCA (called in this literature as Eigen-CSS) is used in the matching stage.

B. Wang and J.A.Bangham [14] have presented an enhanced principal component descriptor (EPCD) for shape based image retrieval. The authors have stated that this descriptor outperforms the other descriptors including Fourier and Wavelet, and CSS descriptor.

In our previous work [15], the BcC is applied to decompose the shape boundary into multi-scale levels associated with the signed enclosed area signature to extract the feature from shape. Then, the shape feature at each scale level is transformed by Discrete Fourier Transform (DFT) to be invariant to starting point selection. The matching based on the PCA is tested on the affine invariant shape database. In this paper, we propose the algorithm for 2D closed boundary shape matching and retrieval (The diagram of the proposed technique is depicted in Fig. 1), using PCA and FLDA called Eigen Barycenter contour (EBcC) and Fisher Barycenter contour (FBcC) respectively. They are employed for dimensionality reduction of our data and furthermore the Fisher method will provide the discrimination by maximizing the discrimination between-class and minimizing the discrimination within-class.

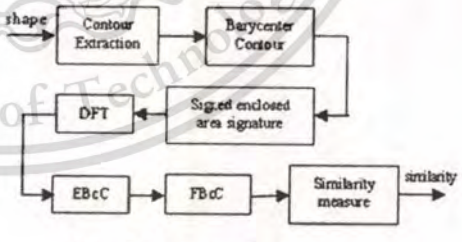


Figure 1 The diagram of our algorithm.

978-1-4244-3388-9/09/$25.00 ©2009 IEEE

The normalized cross correlation is taken into account for similarity measure. This technique gives higher retrieval performance than all the published algorithms tested on the MPEG-7 CE-shape-1 part B.

II. Feature Extraction

In order to compute the multiscale shape representation, the features are extracted from the segmented shapes by performing three steps as follows:

Boundary Extraction: The shape boundary is extracted from the image by using one of the conventional techniques [16] and then presented into 1-D sequence of $x(k)$ and $y(k)$. The contour sequences are re-sampled into N points and shifted about its center point. The appropriate number of sampled points for the experimentation is 128 points.

Barycenter Contour (BcC) Decomposition: The BcC [17], defined in (1), is applied onto the shifted shape contour in order to obtain the different scale level of shape boundary.

$$\left.\begin{aligned} \bar{x}_i^m &= \frac{x_i + x_{i+1} + \ldots + x_{i+m}}{m+2} \\ \bar{y}_i^m &= \frac{y_i + y_{i+1} + \ldots + y_{i+m}}{m+2} \end{aligned}\right\} \quad (1)$$

Where (x_i, y_i) is the coordinate of the shifted shape boundary, $(\bar{x}_i, \bar{y}_i)$ is the Barycenter coordinate of the shape boundary, and m is a level of the Barycenter Contour starting from 1 to 63. Fig. 2 shows the example of Barycenter contours of a bird and a device 9 in (a) and (b) respectively at level 1, 5, 10, 20, 30, 40 and 60.

Shape representation: The computation of the signed enclosed area signature [18] at each scale level of BcC is taken into account as given in (2) and these features are presented into 3D format as the shape representation as shown in Fig. 3.

$$A(m,i) = \frac{1}{2}\left(\bar{x}_i^m \bar{y}_{i+1}^m - \bar{x}_{i+1}^m \bar{y}_i^m\right) \quad (2)$$

The shape signatures are further transformed by the Discrete Fourier Transforms (DFT) in order to be invariant to

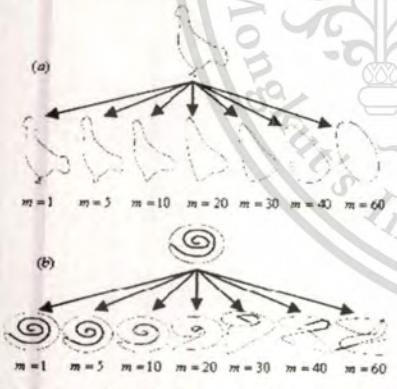


Figure 2 The example of Barycenter contours of a bird (a) and a device 9 (b) at scale level 1, 5, 10, 20, 30, 40, and 60.

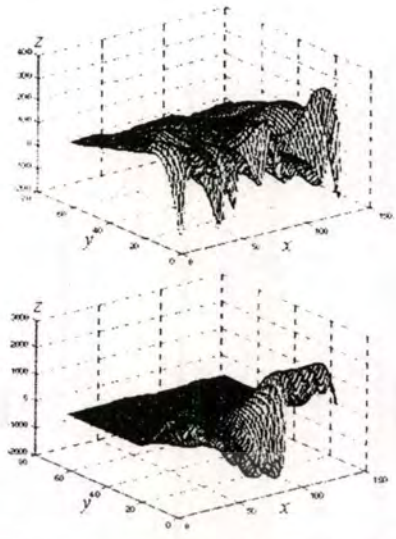


Figure 3 The signed enclosed area signature of Figure 2(a) and 2(b) respectively presented in 3D format. x is the sampling point index, y is the level of Barycenter contours, and z is the value of the signed enclosed area signature.

starting point selection. Then, all the DFT coefficients at each scale level are normalized by dividing with their corresponding DC component coefficients. At the final, they are concatenated into single feature vector as a new shape representation.

III. Shape Matching Using FBCC

Let $\{x_1, x_2, x_3, \ldots, x_M\}$ be a set of column vector of the shape representation taking value in n-dimensional space and suppose that each shape representation belongs to one of c classes $\{X_1, X_2, X_3, \ldots, X_c\}$.

The total covariance matrix Cov can be formed by:

$$Cov = \frac{1}{M}\sum_{i=1}^{M}(x_i - \mu).(x_i - \mu)^T \quad (3)$$

where $\mu = \frac{1}{M}\sum_{i=1}^{M} x_i$ is the mean of all training sample sets.

Let $Q = [x_1 - \mu, x_2 - \mu, x_3 - \mu, \ldots, x_M - \mu]$, so the total covariance matrix can be expressed as follows:

$$Cov = \frac{1}{M} QQ^T \quad (4)$$

The objective is to find the eigenvectors of Cov. However, directly calculating the eigenvectors from Cov is quite expensive computation. To avoid the time-consuming, the singular value decomposition (SVD) technique is applied.

The results from SVD are the eigenvectors $U = [u_1, u_2, \ldots, u_M]$ and the eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \lambda_M$ of Cov.

The optimum eigenvectors U_{opt} are chosen corresponding to the r largest eigenvalues, where the index r is calculated as the minimum index value of the ratio below:

$$\frac{\lambda_1+\lambda_2+\lambda_3+\ldots+\lambda_r}{\lambda_1+\lambda_2+\ldots+\lambda_r+\ldots+\lambda_M} \geq 0.95 \quad (5)$$

Then we can obtain the i[th] projected feature y_i from the sample x_i onto the eigenvector U_{opt} :

$$y_i = U_{opt}^T (x_i - \mu) \quad (6)$$

After dimensionality reduction by PCA, Fisher's LDA is applied to the projected vectors y_i for the discriminative measurement. The Fisher's idea is based on seeking projection direction W that maximizes the ratio of the between-class scatter matrix S_B and the within-class scatter matrix S_W .

Let the between-class scatter matrix be defined as:

$$S_B = \sum_{k=1}^{c} (\mu_k - \mu_c)(\mu_k - \mu_c)^T \quad (7)$$

and the within-class scatter matrix be defined as:

$$S_W = \sum_{k=1}^{c} \sum_{y_i \in Y_k} (y_i - \mu_k)(y_i - \mu_k)^T \quad (8)$$

where μ_k is the mean shape representation of input class Y_k .

μ_c is the mean of all the input y_i .

The Fisher's ratio is mathematically expressed as:

$$J(W) = \frac{W^T S_B W}{W^T S_W W} \quad (9)$$

The solution of maximizing this ratio is the eigenvectors W of the matrix $S_W^{-1} S_B$.

The top eigenvectors are selected corresponding to the $(c-1)$ largest eigenvalues of that matrix. A new projected feature z_i of the given feature input y_i is computed as follows:

$$z_i = W_{opt}^T y_i = W_{opt}^T U_{opt}^T (x_i - \mu) \quad (10)$$

The similarity between the unknown shape and all shapes in the database is measured by the normalized cross correlation value computed as follows:

$$\frac{z_{unknown} \cdot z_i}{\|z_{unknown}\| \cdot \|z_i\|} \quad (11)$$

Then, these values are ranked in decreasing order in which the most similar shape corresponds to the highest value of the normalized cross correlation.

IV. Experimental Results

The performance of FBcC is evaluated onto the MPEG-7 CE-shape-1 database. This database has been widely used for shape matching and retrieval. It consists of 1400 images classified into 70 classes and contains a mixture of natural and artificial objects under various rigid and non-rigid deformations. Some samples from the database are shown in Fig. 4.

The retrieval performances of our method are assessed using the precision-recall curves, where the precision value at a certain recall is the average of the precision value of all database shape at that recall.

Figure 4 Some sample shapes in the MPEG-7 CE-shape-1 database

In the experiment, each shape in the database is used as a test query. Two kinds of parameters are experimented. First, we test with the signed enclosed area signature and the other, with its power spectrum. In each test, the retrieval accuracy of FBcC and EBcC are compared. Fig. 5 shows that power spectrums of the shape signature (represented by solid symbol) achieve higher precision than the shape signature itself (represented by hole symbol) at a certain recall. Furthermore, we found that the FBcC method outperforms the EBcC method.

Additionally, we use the standard test called Bullseye test for the evaluation. The number of correct retrievals in the top of 40 ranks is counted, including the self-match. We have compared our results with the published results by the MPEG-7 database, in which the best result is reported in [19] using hierarchical deformable-shape tree with the rate of 87.7%, as listed in Table I. We found that our FBcC method achieves high retrieval performance of 89.60%. However, when the shape feature is changed to its power spectrum, our method outperforms all the published methods. A score can reach at 96.48% if we count matches in the top 20 rank onto the MPEG-7 database. Fig. 6 shows the retrieval rate of each class for the MPEG-7 Core Experiment CE-shape-1 database.

V. Conclusion

We propose the technique for 2D closed boundary shape matching and retrieval using Fisher Barycenter Contour (FBcC). In this method, the concept of the Barycenter has been employed for decomposing the shape into multiscale levels in order to reduce the noise and small boundary distortion.

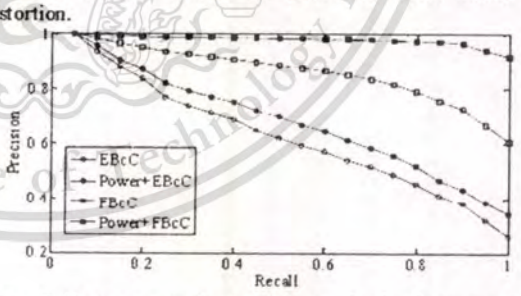


Figure 5 The precision-recall curves of the signed enclosed area signature and its power spectrum associated with EBcC and FBcC

TABLE I
COMPARISON OF THE BULLSEYE TEST FOR DIFFERENT ALGORITHMS

Method	Retrieval accuracy (%)
WARP [2]	58.50
VP [3]	76.45
Curve Edit Distance [4]	78.17
CSS [5]	81.12
BAS [6]	82.37
MCC [7]	84.93
IDSC + DP [8]	85.40
Symbolic Representation [9]	85.92
DSW+Global [10]	87.23
Hierarchical deformable-shape tree [19]	87.70
EBcC	**66.55**
Power spectrum + EBcC	**72.42**
FBcC	**89.60**
Power spectrum + FBcC	**98.62**

The enclosed area signature is calculated at each scale level of BcC, which is invariant to affine transforms, to form the shape representation in 3D format.

Because of the high dimensionality of data, EBcC is used for dimensionality reduction. Moreover, the FBcC technique is employed for making discrimination by maximizing the Fisher's ratio as stated in section III.

The experimental results illustrate that the proposed method achieved high retrieval accuracy. This technique outperforms all the reported technique, based on the Bulls-eye test, which is evaluated onto the MPEG-7 CE-shape-1 part B.

REFERENCES

[1] S. Lambert, E. de Leau, and L. Vuurpijl, "Using pen-based outlines for object-based annotation and image-based queries," *International Conference on Visual Information and Information Systems*, June 1999, pp. 585-592.

[2] I. Bartolini, P. Ciaccia, and M. Patella, "WARP: Accurate Retrieval of Shapes Using Phase of Fourier Descriptors and Time Warping Distance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, No. 1, January 2005, pp. 142-147.

[3] L.J. Latecki and R. Lakamper, "Shape Similarity Measure Based on Correspondence of Visual Parts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, No. 10, October 2000, pp. 1185-1190.

[4] T. Sebastian, P. Klein, and B. Kimia, "On Aligning Curves," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, No. 1, January 2003, pp. 116-125.

[5] F. Mokhtarian and M. Bober, *Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization*, Kluwer Academic Publishers, 2003.

[6] N. Arica and F. Vural, "BAS: A Perceptual Shape Descriptor Based on the Beam Angle Statistics," *Pattern Recognition Letter*, 2003, pp. 1627-1639.

[7] T. Adamek and N.E. O'Connor, "A Multiscale Representation Method for Nonrigid Shapes with a Single Closed Contour," *IEEE Trans. on Circuits System and Video Technology*, 2004, pp. 742-753.

[8] H. Ling, and D. Jacobs, "Shape Classification Using Inner-Distance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, No. 2, February 2007, pp. 286-299.

[9] M.P. Daniel and V. Torre, "Robust Symbolic Representation for Shape Recognition and Retrieval," *Pattern Recognition* 41, 2008, pp. 1782-1798.

[10] N. Alajlan, I. El Rube, M.S. Kamel, and G. Freeman, "Shape retrieval using triangle-area representation and dynamic space warping," *Pattern Recognition* 40, 2007, pp. 1911-1920.

[11] Marios Savvides, B.V.K. Vijaya Kumar and P.K. Khosla, "Eigenphases vs Eigenfaces," *Proceedings of the 17th International Conference on Pattern Recognition*, 2004, pp. 810-813.

[12] P.M. Belhumeur, J.P. Haspanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, July 1997, pp. 711-720.

[13] M.S. Drew, T.K. Lee, and A. Rova, "Shape Retrieval with Eigen-CSS Search," *Image and Vision Computing*, 2008, in press.

[14] B. Wang and J.A. Bangham, "PCA Based Shape Descriptors for Shape Retrieval and the Evaluations," *International Conference on Computational Intelligence and Security*, 2006, pp. 1401-1406.

[15] K. Thourn and Y. Kitjaidure, "Multi-View Shape Recognition Based on Principal Component Analysis," *International Conference on Advanced Computer Control*, Jan. 2009, in press.

[16] L.da F. Costa and R.M. Cesar Jr., *Shape Analysis and Classification: Theory and Practice*, CRC Press LLC, 2001.

[17] K. Thourn and Y. Kitjaidure, and S. Kondo, "Affine Invariant Shape Recognition Based on Multi-Level of Barycenter contour," *International Symposium on Communication and Information Technologies*, October 2008, pp. 145-149.

[18] D. Zhang and G. Lu, "Study and evaluation of different Fourier methods for image retrieval," *Image and Vision Computing* 23, 2005, pp. 33-49.

[19] P.F. Felzenszwalb and J.D. Schwartz, "Hierarchical Matching of Deformable Shapes," *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.

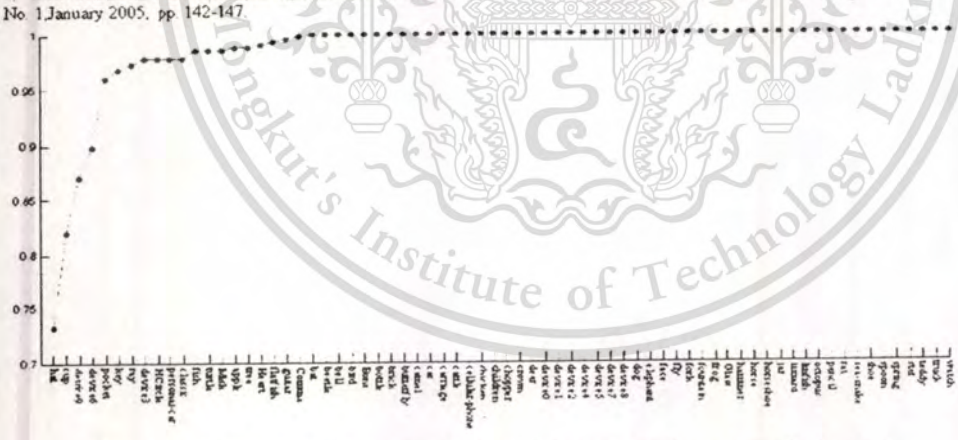


Figure 6 The retrieval accuracy for each class of the MPEG-7 CE-shape-1 part B using Power+FBcC method

2009 IEEE-RIVF International Conference on Computing & Communication Technologies

IEEE-RIVF 2009

Danang University of Technology, July 13-17, 2009

Tru Cao, Ralf-Detlef Kutsche, Akim Demaille Editors

Proceedings

IEEE

Celebrating 125 Years

Eigen and Fisher Barycenter Contour for 2D Shape Classification

Kosorl Thoum, Yuttana Kitjaidure
Department of Electronics, Faculty of Engineering
King Mongkut's Institute of Technology, Ladkrabang,
Bangkok 10520, Thailand
E-mail: kosal_th@yahoo.com, kkyuttan@kmitl.ac.th

Shozo Kondo
Department of Information and Telecommunication
Engineering, Tokai University, Japan
E-mail: kondo@keyaki.cc.u-tokai.ac.jp

***Abstract*—To achieve a good performance for shape classification, it requires both shape representation and classifier. In this paper, the so-called Eigen Barycenter Contour (EBcC) and Fisher Barycenter Contour (FBcC) techniques are presented for 2D shape classification. The representation utilizes the area of triangles at different scale level of Barycenter Contour (BcC). However, it is not invariant to starting point selection, so the phase normalization is applied. After that, we linearly project the shape feature in 3D format onto a subspace based on EBcC technique into low dimensional subspace. The FBcC, another similar method, also produces well separated classes in low dimensional subspace. Finally, the normalized cross correlation is used to measure the similarity among shapes. The experimental results demonstrate that the FBcC method outperforms the EBcC method and achieves high retrieval efficiency over other recent methods in the literature for tests on three different databases, the affine shape database, the MPEG-7 database CE-1 part B and the Kimia's database.**

Keywords-Barycenter Contour (BcC); Phase Normalization; Eigen Barycenter Contour(EBcC); Fisher Barycenter Contour (FBcC); Normalized Cross Correlation.

I. INTRODUCTION

Several properties of the objects have been used for recognition and categorization such as shape, color, texture, brightness, and others. However, the shape is considered as the most promising in a system for object recognition, matching, registration, and analysis. In addition, the users are interested in retrieval by shape rather than by the others as indicated in [1].

Many shape analysis techniques have been proposed over the past decade. The Accurate Retrieval based on Phase (WARP) [2], shape similarity measure based on correspondence of Visual Parts (VP) [3], Curve Edit Distance (CED) [4], Curvature Scale Space (CSS) [5], Beam Angle Statistics (BAS) [6], Multiscale Convexity Concavity representation (MCC) [7], Shape Context [8], Inner Distance Shape Context (IDSC+DP) [9], symbolic representation [10], and Triangle area representation with dynamic programming (DSW+Global) [11] are the examples of the shape recognition and classification methods. However, the current techniques have not given the satisfactory solution yet. This reason has motivated the researchers to develop the algorithm for reaching the efficient and effective results.

The Eigenface and Fisherface [12, 13] are the well-known techniques in computer vision that can dramatically reduce the dimension of the original image and retain most of the information. It is frequently applied to face recognition. Some literatures applied Eigenface onto 2D closed boundary shape for classification and dimensionality reduction.

M.S. Drew et al. [14] have introduced the Eigen-CSS search for shape retrieval. A new feature vector for shape representation has been created and called the marginal sum feature vector which is composed of row-sum and column-sum of the raw CSS image. Then, the Eigenface (called in this literature Eigen-CSS) is used in the matching stage.

B. Wang and J.A.Bangham [15] have presented an Enhanced Principal Component Descriptor (EPCD) for shape based image retrieval. The authors have stated that this descriptor outperforms the other descriptors including Fourier and Wavelet, and CSS descriptor.

In our previous work [16], the algorithm has been done in the spectrum domain to be invariant to starting point selection and the matching based on PCA is tested on the affine invariant shape database.

In this paper, we develop the algorithm for 2D shape classification in spatial domain using DFT onto the feature to automatically fix the starting point [2], then, transform back into spatial domain.

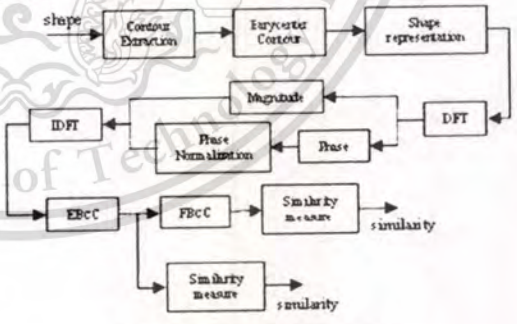


Figure 1: The diagram of our algorithm.

978-1-4244-4568-4/09/$25.00 ©2009 IEEE

Then, the representation at each level is concatenated into a single column vector producing the high dimension of shape feature vector. So, this feature is projected onto the subspace based on the Eigenface technique into low dimensional subspace. Similarly, for Fisher method, it does not only project data onto low dimensional subspace but also provides the discrimination by maximizing the discrimination between-class and minimizing the discrimination within-class. Finally, the normalized cross correlation is taken into account for similarity measure. This process is depicted in Fig. 1.

This technique gives higher retrieval performance than all the published algorithms tested on the MPEG-7 database CE-1 and the Kimia's database.

The paper is organized as follows: feature extraction is introduced in section II, followed by phase normalization in section III. Then, shape matching algorithm is given in section IV. Next, the experiment results are illustrated in section V. Finally, conclusion is given in section VI.

II. FEATURE EXTRACTION

In order to compute the shape representation, the features are extracted from the segmented shapes by performing three steps as follows:

Boundary Extraction: The shape boundary is extracted from the image by using one of the conventional techniques [17] and then presented into 1-D sequence of $x(k)$ and $y(k)$. The contour sequences are re-sampled into N points and shifted about its center point. The appropriate number of sampled points for the experimentation is 128 points.

Barycenter Contour (BcC) Decomposition: The BcC [18], defined in (1), is applied onto the shifted shape contour in order to obtain the different scale levels of shape boundary.

$$\left.\begin{aligned} \bar{x}_i^m &= \frac{x_i + x_{i+1} + \ldots + x_{i+m}}{m+2} \\ \bar{y}_i^m &= \frac{y_i + y_{i+1} + \ldots + y_{i+m}}{m+2} \end{aligned}\right\} \quad (1)$$

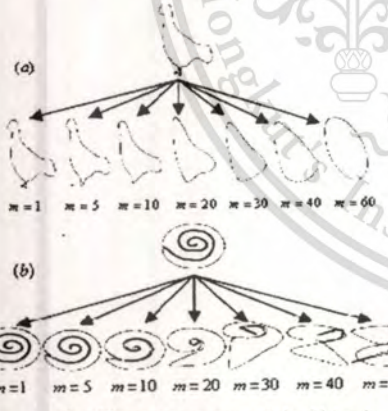


Figure 2: The example of Barycenter contours of a bird (a) and a device 9 (b) at scale level 1, 5, 10, 20, 30, 40, and 60.

Figure 3: The triangle area representation (TAR) signature of Figure 2(a) and 2(b) respectively. x is the sampled point index, y is the level of Barycenter contours, and z is the value of the TAR signature.

where (x_i, y_i) is the coordinate of the shifted shape boundary, $(\bar{x}_i, \bar{y}_i)$ is the Barycenter coordinate of the shape boundary, and m is a level of the Barycenter Contour starting from 1 to 63. Fig. 2 shows the example of Barycenter contours of a bird and a device 9 in (a) and (b) respectively at level 1, 5, 10, 20, 30, 40 and 60.

Shape representation: The computation of the triangle area representation (TAR) signature [11, 18, 19] at each scale level of BcC is taken into account as given in (2) and these features are presented into 3D format as the shape representation as shown in Fig. 3.

$$TAR(m,i) = \frac{1}{2}\begin{vmatrix} \bar{x}_{i-1}^m & \bar{y}_{i-1}^m & 1 \\ \bar{x}_i^m & \bar{y}_i^m & 1 \\ \bar{x}_{i+1}^m & \bar{y}_{i+1}^m & 1 \end{vmatrix} \quad (2)$$

III. PHASE NORMALIZATION

A main problem in getting a good result in shape classification is how to make the shape representation invariant to translation, scale, rotation, shear, reflection and starting point selection.

In our algorithm, the TAR is selected as the shape representation. As stated in [18, 19], this feature is invariant to affine transforms, its parameters are translation, scale, rotation and shear except reflection and starting point selection. The two exceptions will affect to the performance of classification.

Our solution to this problem is to carry out the phase normalization on TAR at each scale level of BcC. This process is in the same way as the Fourier phase normalization [2] which has been used to eliminate the starting point dependency.

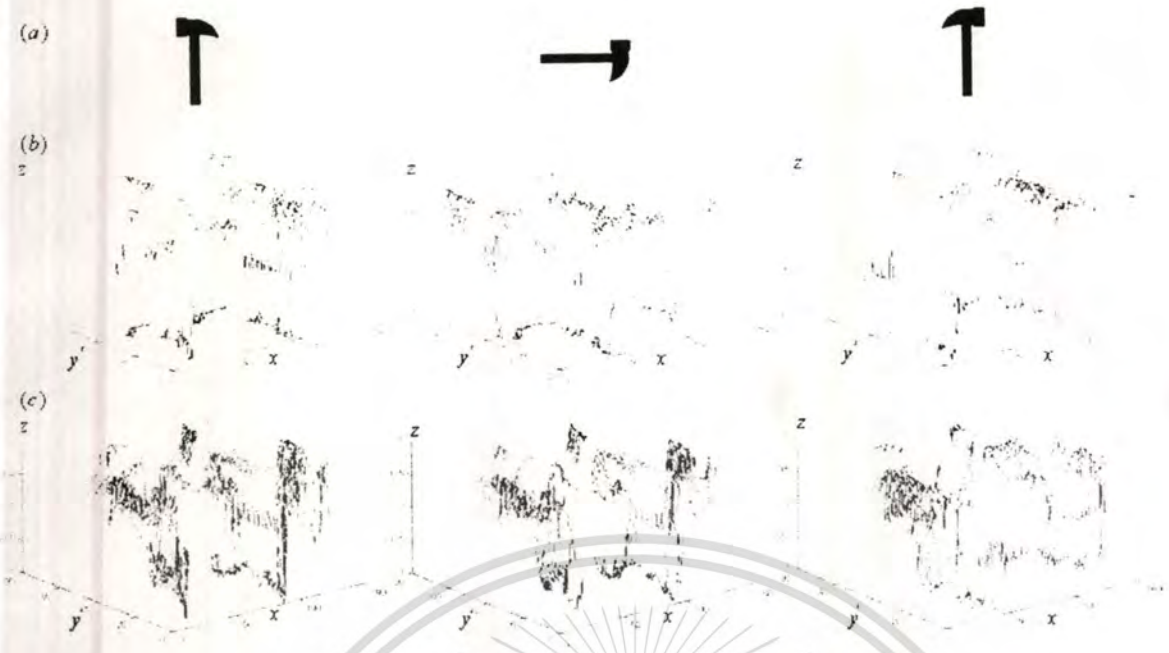


Figure 4: (a) The shape silhouette, its 90° rotation and its reflection transformation, and the corresponding of TAR at different scale of BcC (b) before phase normalization and (c) after phase normalization. x is the sampling point index, y is the level of Barycenter contours, and z is the value of the TAR signature.

and to be invariant to reflection transform. This can be accomplished by transforming TAR at each scale level of BcC into spectrum domain, calculating the magnitude and phase. Then, phase of Fourier is normalized to be invariant to starting point and reflection. We will get TARs by transforming them back into spatial domain. This process is depicted in Fig. 4.

IV. SHAPE MATCHING ALGORITHM

In this section, we examine two classification techniques, the Eigen Barycenter Contour (EBcC) and the Fisher Barycenter Contour, for 2D shape classification.

Let $\{x_1, x_2, x_3, \ldots x_N\}$ be a set of column vector of the shape representations taking value in n-dimensional space and suppose that each shape representation belongs to one of c classes $\{X_1, X_2, X_3, \ldots\ldots, X_c\}$.

A. *Eigen Barycenter Contour (EBcC)*

The total covariance matrix Cov can be formed by

$$Cov = \frac{1}{N}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T \qquad (3)$$

where $\mu = \frac{1}{N}\sum_{i=1}^{n} x_i$ is the mean of all training sample sets.

Let $Q = [x_1 - \mu, x_2 - \mu, \ldots\ldots, x_n - \mu]$, so the total covariance can be expressed as follows:

$$Cov = \frac{1}{N} QQ^T \qquad (4)$$

The objective is to find the projection axes which are the eigenvectors of the total covariance Cov. The finding of the eigenvectors of Cov is quite expensive computation and consumes much time. To avoid the time-consuming calculation, the singular value decomposition (SVD) technique is applied. The results from SVD are the required eigenvector $U = [u_1, u_2, u_3, \ldots\ldots, u_n]$ of Cov corresponding to the eigenvalue $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots\ldots \lambda_n$.

The optimum eigenvector U_{opt} is chosen corresponding to the r largest eigenvalue where the index value r is calculated as the minimum index value of the ratio below:

$$\frac{\lambda_1 + \lambda_2 + \lambda_3 + \ldots + \lambda_r}{\lambda_1 + \lambda_2 + \ldots + \lambda_r + \ldots + \lambda_M} \geq 0.95 \qquad (5)$$

Then we can obtain the i^{th} projected feature y_i from the sample x_i onto the optimum eigenvector U_{opt}:

$$y_i = U_{opt}^T (x_i - \mu) \qquad (6)$$

B. *Fisher Barycenter Contour (FBcC)*

This method is similar to the EBcC above for dimensionality reduction by finding the optimum projection axes. The difference between the algorithms is that the eigenvectors are found by the separated matrix derived from Fisher's linear discriminant function instead of the covariance

matrix. The separated matrices are the between-class scatter matrix S_B and the within-class scatter matrix S_W.

Let the between-class scatter matrix be defined as

$$S_B = \sum_{k=1}^{c} (\mu_k - \mu)(\mu_k - \mu)^T \quad (7)$$

and the within-class scatter matrix be defined as

$$S_W = \sum_{k=1}^{c} \sum_{x_i \in X_k} (x_i - \mu_k)(x_i - \mu_k)^T \quad (8)$$

where μ_k is the mean shape representation of class X_k.

The objective is to find the optimum projection axe W_{opt} which maximizes the Fisher's ratio below

$$J(W) = \frac{W^T S_B W}{W^T S_W W} \quad (9)$$

The solution of finding the maximum value of the Fisher's ratio is the eigenvectors of the matrix $S_W^{-1} S_B$.

Due to very expansive computation of directly finding the eigenvector from $S_W^{-1} S_B$ matrix with too large dimensional space, the resulted projected vectors from EBcC are applied to FBcC. So the eigenvectors from $S_{W(EBcC)}^{-1} S_{B(EBcC)}$ are resulted.

The top eigenvectors are selected corresponding to the $(c\text{-}1)$ largest eigenvalues of that matrix. A new projected feature z_i of the given feature x_i is computed as follows.

$$z_i = W_{opt}^T U_{opt}^T (x_i - \mu) \quad (10)$$

C *Similarity measure*

The similarity between the unknown shape $s_{unknown}$ and all shapes in the database s_i is measured by the normalized cross correlation value computed as follows:

$$\frac{s_{unknown} s_i}{\|s_{unknown}\| \|s_i\|} \quad (11)$$

Then, these values are ranked in decreasing order in which the most similar shape corresponds to the highest value of the normalized cross correlation.

Figure 5: The samples of affine distorted shape for class "butterfly".

V EXPERIMENTAL RESULTS

The performance of our algorithm is demonstrated using three different databases, the affine shape database, the MPEG-7 CE-shape-1 part B and the Kimia's database. In the experiment, all shapes in the database are used as a test query.

The retrieval performances of our method are assessed using the precision-recall curves, where the precision value at a certain recall is the average of the precision value of all database shapes at that recall. The precision and recall can be defined by

$$\text{Precision} = \frac{\text{Number of retrieved relevant shape}}{\text{Total number of retrieved shape}} \quad (12)$$

$$\text{Recall} = \frac{\text{Number of retrieved relevant shape}}{\text{Total number of relevant shape}} \quad (13)$$

A *Affine shape database*

The affine invariant shape database [16, 18] consists of 40 categories of shapes chosen from MPEG-7 contour shape database CE-1. Each category has 14 different distorted shapes. So there are in total 560 affine distorted shapes in this database. The sample of affine distorted shape for class 'butterfly' is shown in figure 5.

The retrieval performance is assessed by the precision-recall curve as depicted in Fig. 6. We found that the FBcC method outperforms the EBcC method. Furthermore, the correct retrievals of each category in the top 14 rank are counted in which the shapes in all class can retrieve 100 % except class of hat using the FBcC method as listed in Table I.

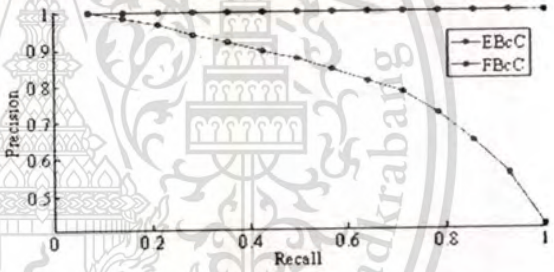


Figure 6: The precision-recall curves of the EBcC and FBcC method tested on the affine shape database.

TABLE I THE CORRECT RETRIEVALS OF EACH CLASS IN AFFINE SHAPE DATABASE USING EBCC AND FBCC ALGORITHM

Algorithm	apple	bat	beetle	bell	bird	Bone	brick	butterfly	camel	carriage
EBcC	93.82%	71.43%	86.22%	40.31%	66.33%	77.55%	50.51%	95.41%	96.43%	88.27%
FBcC	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Algorithm	cattle	phone	children	chopper	classic	Comma	crown	device1	device2	device5
EBcC	79.59%	81.63%	79.59%	56.12%	39.29%	76.02%	73.98%	70.41%	40.82%	62.76%
FBcC	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Algorithm	device8	elephant	face	fish	flatfish	fork	fountain	frog	guitar	hammer
EBcC	65.84%	80.61%	58.16%	68.88%	86.22%	99.49%	86.73%	84.69%	85.71%	66.33%
FBcC	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Algorithm	hat	Heart	horse	key	Misk	rat	spring	stef	teddy	tree
EBcC	44.39%	77.55%	93.88%	94.90%	77.55%	98.98%	100%	95.41%	93.37%	62.24%
FBcC	92.86%	100%	100%	100%	100%	100%	100%	100%	100%	100%

B *MPEG-7 database CE-shape-1*

The MPEG-7 CE-shape-1 database has been widely used for shape matching and retrieval. It consists of 1400 images classified into 70 classes and contains a mixture of natural and artificial objects under various rigid and non-rigid deformations. The sample shape of each class in MPEG-7 database is shown in Fig. 7.

The retrieval performance is assessed by the precision-recall curve as depicted in Fig. 8. We found that the FBcC method outperforms the EBcC method.

Additionally, we use the standard test called Bullseye test for the evaluation. The number of correct retrievals in the top of 40 ranks is counted, including the self-match. We have compared our results with the published results by the MPEG-7 database, in which the best result is reported in [20] using hierarchical deformable-shape tree with the rate of 87.7%, as listed in Table II. We found that our EBcC method give a very low accuracy. However, the FBcC method achieves high retrieval performance of 95.50% over all the published methods.

Figure 7: The sample of shape in each class of the MPEG-7 database.

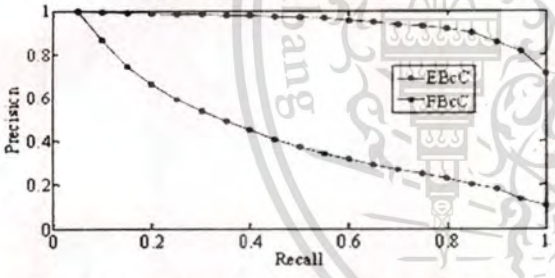


Figure 8: The precision-recall curves of the EBcC and FBcC method tested on the MPEG-7 database part B.

TABLE II
COMPARISION OF THE BULLSEYE TEST FOR DIFFERENT ALGORITHMS ON THE MPEG-7 CE-SHAPE-1 DATABASE

Method	Retrieval accuracy (%)
WARP [2]	58.50
VP [3]	76.45
Curve Edit Distance [4]	78.17
CSS [5]	81.12
BAS [6]	82.37
MCC [7]	84.93
IDSC + DP [9]	85.40
Symbolic Representation [10]	85.92
DSW+Global [11]	87.23
Hierarchical deformable-shape tree [20]	87.70
EBcC	**47.24**
FBcC	**95.50**

C *Kimia's database*

This database consists of 99 shapes from nine categories in which most of shapes are partially occluded. All the shapes in this database are shown in Fig. 9.

By the precision-recall curve in Fig. 10 shows that the retrieval performance using the FBcC technique is better than that using the EBcC technique.

A comparison of performances between different algorithms is summarized in Table III, showing the correct retrievals from the top 10 closet match. Notice that the maximum number of correctness for each category is 99.

As shown in Table III, we found that the EBcC method can retrieve in 10 cases of 591 shapes which is low performance comparing with other algorithms. However, the FBcC method achieves a good result of 100%.

Figure 9: All the shapes in the Kimia's database.

TABLE III
COMPARISION OF RETRIEVAL RATE FOR DIFFERENT ALGORITHMS ON THE KIMIA'S DATABASE

Method	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Total
Shape context [8]	97	91	88	85	84	77	75	66	56	37	756
MDS+SC+DP [9]	99	98	98	98	97	99	97	96	97	85	964
Symbolic representation [10]	99	99	99	98	99	98	98	95	96	94	975
DSW+Global [11]	99	99	99	98	98	97	98	95	93	80	956
EBcC	**99**	**84**	**74**	**64**	**59**	**43**	**37**	**50**	**36**	**45**	**591**
FBcC	**99**	**99**	**99**	**99**	**99**	**99**	**99**	**99**	**99**	**99**	**990**

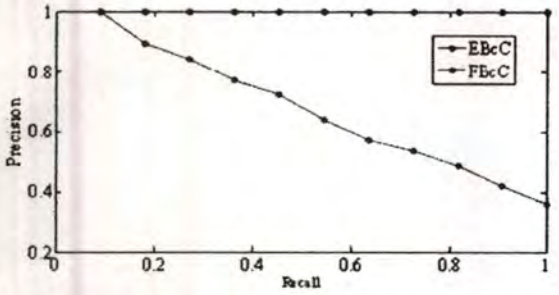


Figure 10. The precision-recall curves of the EBcC and FBcC method tested on the Kimia's database.

VI. CONCLUSION

We present the technique for 2D closed boundary shape matching and retrieval using Eigen Barycenter Contour (EBcC) and Fisher Barycenter Contour (FBcC). In our algorithm, the concept of the Barycenter has been employed for decomposing the shape into multiscale levels in order to reduce the noise and small boundary distortion.

The triangle area representation (TAR) signature is calculated at each scale level of BcC, which is invariant to affine transforms, to form the shape representation in 3D format.

Because of the high dimensionality of data, EBcC is used for dimensionality reduction. The retrieval accuracy experimented by using EBcC is not good over three databases. However, when using FBcC technique, it does not only outperform the EBcC method but also achieves high retrieval rate over the existing methods, based on the Bullseye test, which is tested on the two well-known database, the MPEG-7 database and the Kimia's database.

REFERENCES

[1] S. Lambert, E. de Leau, and L. Vuurpijl, "Using pen-based outlines for object-based annotation and image-based queries," *International Conference on Visual Information and Information Systems*, June 1999, pp. 585-592.

[2] I. Bartolini, P. Ciaccia, and M. Patella, "WARP: Accurate Retrieval of Shapes Using Phase of Fourier Descriptors and Time Warping Distance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol 27, No. 1, January 2005, pp. 142-147.

[3] L.J. Latecki and R. Lakamper, "Shape Similarity Measure Based on Correspondence of Visual Parts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, No. 10, October 2000, pp. 1185-1190.

[4] T. Sebastian, P. Klein, and B. Kimia, "On Aligning Curves," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, No. 1, January 2003, pp. 116-125.

[5] F. Mokhtarian and M. Bober, *Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization*, Kluwer Academic Publishers, 2003.

[6] N. Arica and F. Vural, "BAS: A Perceptual Shape Descriptor Based on the Beam Angle Statistics," *Pattern Recognition Letter*, 2003, pp. 1627-1639.

[7] T. Adamek and N.E. O'Connor, "A Multiscale Representation Method for Nonrigid Shapes with a Single Closed Contour," *IEEE Trans. on Circuits System and Video Technology*, 2004, pp. 742-753.

[8] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Context," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, No. 4, April 2002, pp. 509-522.

[9] H. Ling, and D. Jacobs, "Shape Classification Using Inner-Distance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, No. 2, February 2007, pp. 286-299.

[10] M.R. Daili and V. Torre, "Robust Symbolic Representation for Shape Recognition and Retrieval," *Pattern Recognition 41*, 2008, pp. 1782-1798.

[11] N. Alajlan, I. El Rube, M.S. Kamel, and G. Freeman, "Shape retrieval using triangle-area representation and dynamic space warping," *Pattern Recognition 40*, 2007, pp. 1911-1920.

[12] Marios Savvides, B.V.K. Vijaya Kumar and P.K. Khosla, "Eigenphases vs. Eigenfaces," *Proceedings of the 17th International Conference on Pattern Recognition*, 2004, pp. 810-813.

[13] P.M. Belhumeur, J.P. Haspanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, July 1997, pp. 711-720.

[14] M.S. Drew, T.K. Lee, and A. Rova, "Shape Retrieval with Eigen-CSS Search," *Image and Vision Computing*, 2008, in press.

[15] B. Wang and J.A. Bangham, "PCA Based Shape Descriptors for Shape Retrieval and the Evaluations," *International Conference on Computational Intelligence and Security*, 2006, pp. 1401-1406.

[16] K. Thoum and Y. Kitjaidure, "Multi-View Shape Recognition Based on Principal Component Analysis," *International Conference on Advanced Computer Control*, Jan. 2009, in press.

[17] L.da F. Costa and R.M. Cesar Jr, *Shape Analysis and Classification: Theory and Practice*, CRC Press LLC, 2001.

[18] K. Thoum and Y. Kitjaidure, and S. Kondo, "Affine Invariant Shape Recognition Based on Multi-Level of Barycenter contour," *International Symposium on Communication and Information Technologies*, October 2008, pp. 145-149.

[19] I.El Rube, M. Ahmed and M. Kamel, "Affine Invariant Multiscale Wavelet-base Shape Matching Algorithm," *Proceedings of the first Canadian Conference of Computer and Robot Vision*, May 2004, pp. 217-224.

[20] P.F. Felzenszwalb and J.D. Schwartz, "Hierarchical Matching of Deformable Shapes," *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.