

การวิเคราะห์ปริภูมีย่อยสองมิติสำหรับการรู้จำรูปแบบ



นายปริญญา สงวนสัตย์

สถาบันวิทยบริการ

จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต

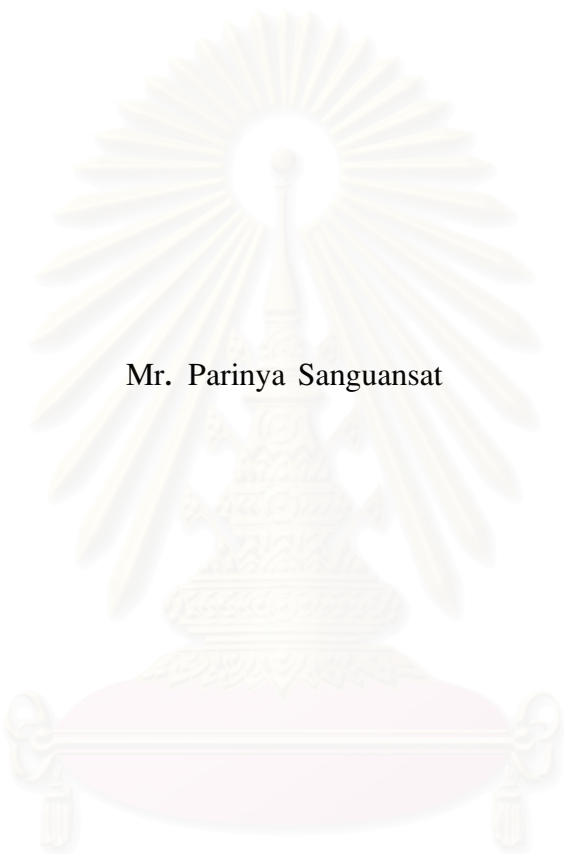
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2550

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

TWO-DIMENSIONAL SUBSPACE ANALYSIS FOR PATTERN RECOGNITION



Mr. Parinya Sanguansat

A Dissertation Submitted in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy Program in Electrical Engineering

Department of Electrical Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2007

Copyright of Chulalongkorn University

Thesis Title TWO-DIMENSIONAL SUBSPACE ANALYSIS FOR PATTERN RECOGNITION

By Mr. Parinya Sanguansat

Field of Study Electrical Engineering

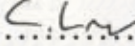
Thesis Advisor Associate Professor Somchai Jitapunkul, Dr.Ing.


Thesis Co-advisor Widhyakorn Asdornwised, Ph.D., Sanparith Marukatat, Ph.D.

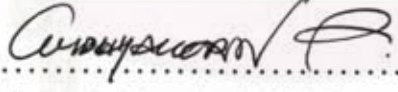
Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

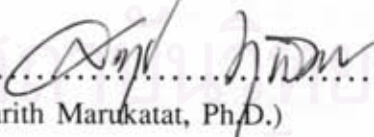
 ..... Dean of the Faculty of Engineering  
(Professor Direk Lavansiri, Ph.D.)

THESIS COMMITTEE

 ..... Chairman  
(Professor Chidchanok Lursinsap, Ph.D.)


 ..... Thesis Advisor  
(Associate Professor Somchai Jitapunkul, Dr.Ing.)

 ..... Thesis Co-advisor  
(Widhyakorn Asdornwised, Ph.D.)

 ..... Thesis Co-advisor  
(Sanparith Marukatat, Ph.D.)

 ..... Member  
(Assistant Professor Chedsada Chinrungrueng, Ph.D.)

 ..... Member  
(Sanun Srisuk, D.Eng.)

 ..... Member  
(Charnchai Pluempitiwiriyawej, Ph.D.)

ปริญา สงวนสัตย์: การวิเคราะห์ปริภูมิย่อยสองมิติสำหรับการรู้จำรูปแบบ (TWO-DIMENSIONAL SUBSPACE ANALYSIS FOR PATTERN RECOGNITION), อ. ที่ปรึกษา: รศ. ดร.สมชาย จิตะพันธ์กุล, อ. ที่ปรึกษาร่วม: อ.ดร.วิทยากร อัครวิเศษ, ดร.สรพฤทธิ์ มฤคทัต, 115 หน้า

วิทยานิพนธ์ฉบับนี้เสนอการรู้จำรูปแบบสำหรับภาพจำนวน 4 วิธี วิธีแรกนั้นอาศัยหลักการวิเคราะห์ทีสคริมิแนนด์ของส่วนประกอบสำคัญหรือหน้าพีชเซอร์ เนื่องจาก 2DPCA เหมาะสำหรับการแทนภาพมากกว่าการรู้จำดังนั้น 2DLDA จึงถูกเสนอขึ้นเพื่อเพิ่มประสิทธิภาพในการรู้จำ วิธีที่สองอยู่บนพื้นฐานของปริภูมิย่อยแบบเจาะจงคลาส (CSS) การนำ CSS มาประยุกต์ใช้กับ 2DPCA เป็นการนำข้อมูลของคลาสมาสู่วิธีการวิเคราะห์ปริภูมิแบบ Unsupervised โดยแต่ละปริภูมิย่อยของ CSS นั้นฝึกฝนด้วยตัวอย่างจากในคลาสของตัวเองเท่านั้น ด้วยวิธีนี้ทำให้วิธี CSS สามารถให้ค่าความผิดพลาดของการสร้างคินน้อยที่สุด ซึ่งเราสามารถให้ค่าความผิดพลาดดังกล่าวเป็นตัวแยกแยะคลาสของข้อมูลป้อนเข้าได้ วิธีที่สามนั้นอาศัยรูปแบบทั่วไปของเมตริกซ์ความแปรปรวนร่วมเชิงภาพที่เราเสนอขึ้น เรียกว่า เมตริกซ์ความแปรปรวนร่วมแบบข้ามเชิงภาพ โดยหากเปรียบเทียบเมตริกซ์ความแปรปรวนร่วมเชิงภาพกับเมตริกซ์ความแปรปรวนร่วมของ PCA จะพบว่าข้อมูลบางส่วนถูกละทิ้งไป ซึ่งข้อมูลส่วนนี้บางทีอาจมีประโยชน์ในการแยกแยะคลาส ในขณะที่เมตริกซ์ความแปรปรวนร่วมแบบข้ามเชิงภาพนั้นสร้างขึ้นจากตัวแปรสองตัวระหว่างภาพต้นฉบับ และภาพที่เกิดจากการเลื่อนด้วยวิธีการเลื่อนที่เราเสนอขึ้นเองของภาพต้นฉบับนั้น ด้วยวิธีเลื่อนที่เราแนะนำเสนอนั้นทำให้เมตริกซ์ความแปรปรวนร่วมแบบข้ามเชิงภาพสามารถเลือกใช้ข้อมูลที่เคยถูกละทิ้งไปในเมตริกซ์ความแปรปรวนร่วมเชิงภาพเดิม วิธีที่สี่วิธีปริภูมิย่อยแบบสุ่มถูกนำมาประยุกต์ใช้กับ 2DPCA ซึ่งโดยปกติแล้วคุณลักษณะของ 2DPCA อยู่ในรูปเมตริกซ์ 2 มิติ โดยจำนวนหลักของเมตริกซ์นี้ขึ้นกับจำนวนค่าเจาะจงของเมตริกซ์ความแปรปรวนร่วมเชิงภาพในขณะที่จำนวนแถวยังคงเท่ากับความสูงเดิมของภาพ ดังนั้นวิธีปริภูมิย่อยแบบสุ่มเหมาะที่จะประยุกต์ใช้ในแนวหลักเท่านั้น เพราะจำนวนค่าเจาะจงนี้ไม่มีผลกระทบต่อจำนวนแถวของเมตริกซ์คุณลักษณะ ปริภูมิย่อยแบบสุ่มจำนวนมากสามารถสร้างขึ้นโดยการสุ่มเลือกแถวของเมตริกซ์คุณลักษณะจำนวนหนึ่ง โดยมีตัวแยกประเภทหนึ่งตัวสำหรับแต่ละปริภูมิย่อยนี้ จากนั้นผลการตัดสินใจสุดท้ายได้จากการออกเสียงข้างมากของแต่ละตัวแยกประเภทเหล่านี้ โดยวิธีปริภูมิย่อยแบบสุ่มนี้ยังถูกนำมาใช้กับวิธี DiaPCA รวมทั้งใช้เพื่อเลือกใช้ปริภูมิย่อยที่เกิดขึ้นจากเมตริกซ์ความแปรปรวนร่วมแบบข้ามเชิงภาพในวิธีที่สาม

ผลการทดลองบนฐานข้อมูลมาตรฐาน ทั้งฐานข้อมูลภาพใบหน้ามนุษย์และแบบไม่ใช่ใบหน้ามนุษย์ทำให้เห็นว่าวิธีการที่เราเสนอขึ้นมาได้อัตราการรู้จำเพิ่มขึ้นจากวิธีดั้งเดิม

ภาควิชา ..... วิศวกรรมไฟฟ้า .....  
สาขาวิชา ..... วิศวกรรมไฟฟ้า .....  
ปีการศึกษา ..... 2550 .....

ลายมือชื่อนิสิต ..... ปริญา .....  
ลายมือชื่ออาจารย์ที่ปรึกษา .....  
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....  
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....



##4771860821: MAJOR ELECTRICAL ENGINEERING

KEY WORD: 2DPCA / 2DCSS / 2DRSA / ICCA / FACE RECOGNITION / ATR

PARINYA SANGUANSAT: TWO-DIMENSIONAL SUBSPACE ANALYSIS FOR PATTERN RECOGNITION. THESIS ADVISOR: ASSOC. PROF. SOMCHAI JI-TAPUNKUL, Dr.Ing., THESIS COADVISOR: WIDHYAKORN ASDORNWISED, Ph. D., SANPARITH MARUKATAT, Ph.D., 115 pp.

This dissertation proposes four novel frameworks for image pattern recognition. The first framework is based on discriminant analysis of principal components framework or Fisherface. Since 2DPCA is more suitable for face representation than face recognition, thus 2DLDA is proposed for improving performance in recognition task. The second framework is based on Class-Specific Subspace (CSS) method. By applying CSS over 2DPCA, the class information is introduced to an unsupervised method. Each subspace of CSS learned from only the training samples within their own class. In this way, the CSS representation can provide a minimum reconstruction error, which it can be used to classify the input data. The third framework is based on our proposed method by generalizing form of image covariance matrix, called image cross-covariance matrix. Comparing to the covariance matrix of PCA, the image covariance matrix discards some information. In practice, this disregard information may possibly be useful for discrimination. The image cross-covariance matrices are formulated by two variables, the original image and its shifted version. By our shifting algorithm, many image cross-covariance matrices are formulated to use another information in which discarded by the image covariance matrix. The fourth framework is to apply the random subspace method to 2DPCA. Normally, the feature of 2DPCA is a matrix. In the row direction, the number of the columns of this matrix is affected by the number of selected eigenvalues of the image covariance matrix while the number of selected eigenvalues is not influenced in the column direction. Thus, the number of the rows is still equal to the height of original image and the random subspace method can be apply in the column direction. The random subspaces are constructed by randomly selecting a number of rows of the original feature matrix. The multiple classifiers are constructed in these random subspaces of the data feature space. These classifiers are usually combined by simple majority voting in the final decision rule. Moreover, this framework is also applied to DiaPCA that used to select the subspaces which constructed by the third framework.

Experimental results on well-known image databases, both of face and non-face databases, show that all of our proposed techniques clearly gives a higher recognition accuracy than the conventional algorithm.

Department .. Electrical Engineering  
Field of study .. Electrical Engineering  
Academic year ..... 2007 .....

Student's signature ..... ปริญญา สังกุอานสาร  
Advisor's signature .....  
Co-advisor's signature .....  
Co-advisor's signature .....  
.....

## Acknowledgements

The work contained in this dissertation represents the accumulation of three years of my work made possible only by the collectively support of family, friends, colleagues and mentors. First and foremost, I would like to express my deep gratitude to my advisor Assoc. Dr. Somchai Jitapunkul, and my co-advisors, Dr. Widhyakorn Asdornwised and Dr. Sanparith Marukatat. It was a great privilege and honor to work and study under their guidance. I would like to thank them for their supports, friendship, empathy and great vision.

My thanks also to the members of my dissertation committees, Prof. Dr. Chidchanok Lursinsap, Assist. Dr. Chedsada Chinrungrueng, Dr. Sanun Srisuk, and Dr. Charnchai Pluempitiwiriyawej for their suggestions and invaluable comments.

I would also like to acknowledge Thailand Graduate Institute of Science and Technology (TGIST) and the Rajadapisek Sompoj endowment and Cooperation Project between Department of Electrical Engineering and Private Sector for Research and Development.

Many thanks to my all colleagues at DSPRL (Digital Signal Processing Research Laboratory) for academic documentary help, technical/theory information and programming/data information. Finally, my life has been constantly fulfilled by love and support of my family. I am extremely grateful to my parents, Nitti Sanguansat and Suchada Sanguansat, for their love, caring and sacrifices, my father and mother for educating and preparing me for my future life since I was a kid. This work is dedicated to them.

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

# Table of Contents

	Page
Abstract in Thai.....	iv
Abstract in English.....	v
Acknowledgements.....	vi
Table of Contents.....	vii
List of Tables.....	x
List of Figures.....	xi
Chapter	
<b>I Introduction.....</b>	<b>1</b>
1.1 Background and Signification of the Research Problems.....	1
1.2 Literature Review.....	1
1.3 Objectives.....	7
1.4 Scope.....	8
1.5 Expected Prospects.....	8
1.6 Research Procedure.....	9
<b>II Basic Background and Related Topics.....</b>	<b>10</b>
2.1 One-Dimensional Subspace Analysis and Its Variants.....	10
2.1.1 Principal Component Analysis (PCA).....	10
2.1.2 Linear Discriminant Analysis (LDA).....	12
2.1.3 Linear Discriminant Analysis of Principal Components.....	13
2.1.4 Class-Specific Subspace-Based PCA.....	14
2.1.5 Problems in Subspace Analysis.....	14
2.1.5.1 The Curse of Dimensionality.....	14
2.1.5.2 The Small Sample Size (SSS) Problem.....	14
2.1.5.3 Over-fitting Problem.....	16
2.1.6 Ensemble Methods for Subspace Analysis.....	16
2.1.6.1 Random Subspace Method (RSM).....	16
2.2 Two-Dimensional Subspace Analysis and Its Variants.....	18
2.2.1 Image Covariance Matrix and Two-Dimensional Principal Component Analysis (2DPCA).....	18
2.2.2 Column-Based 2DPCA.....	20
2.2.3 Bilateral-Projection-Based 2DPCA (B2DPCA).....	21
2.2.4 Diagonal-Based 2DPCA (DiaPCA).....	23
2.3 PCA Versus 2DPCA.....	27
2.3.1 The Relationship between PCA and 2DPCA.....	27
2.3.2 Evaluation of Over-fitting Problem in PCA and 2DPCA.....	27

Chapter	Page
2.3.3 The Distance Measurements . . . . .	30
2.3.3.1 1D Feature Distance Measurements . . . . .	30
2.3.3.2 2D Feature Distance Measurements . . . . .	30
<b>III The Proposed Frameworks . . . . .</b>	<b>33</b>
3.1 Two-Dimensional Linear Discriminant Analysis of Principal Component Vectors . . . . .	33
3.1.1 Two-Dimensional Linear Discriminant Analysis (2DLDA) . . . . .	35
3.1.2 2DPCA+2DLDA . . . . .	37
3.1.3 2D Fisherface . . . . .	38
3.1.4 2DPCA+2DLDA Versus 2D Fisherface . . . . .	38
3.2 Class-Specific Subspace-Based Two-Dimensional Principal Component Anal- ysis . . . . .	40
3.3 Image Cross-Covariance Analysis . . . . .	42
3.3.1 The Image Covariance Matrix Revisited . . . . .	42
3.3.1.1 Relation to Covariance Matrix . . . . .	43
3.3.2 From Image Covariance Matrix to Image Cross-Covariance Matrix . . . . .	44
3.3.3 Image Cross-Covariance Analysis (ICCA) . . . . .	46
3.4 Random Subspace Method-Based Two-Dimensional Subspace Analysis . . . . .	48
3.4.1 Two-Dimensional Random Subspace Analysis (2DRSA) . . . . .	49
3.4.2 Two-Dimensional Diagonal Random Subspace Analysis (2D <sup>2</sup> RSA) . . . . .	49
3.4.3 Random Subspace Method-based Image Cross-Covariance Analysis . . . . .	50
<b>IV The Experimental Results . . . . .</b>	<b>52</b>
4.1 Image Databases . . . . .	52
4.1.1 Yale Database . . . . .	52
4.1.2 ORL Database . . . . .	52
4.1.3 AR Database . . . . .	53
4.1.4 MSTAR Database . . . . .	53
4.2 Preprocessing . . . . .	56
4.2.1 Converting Color to Grayscale . . . . .	56
4.2.2 Cropping and Resizing . . . . .	56
4.2.3 Normalization . . . . .	57
4.3 Experiments and Analysis on Two-Dimensional Linear Discriminant Analysis of Principal Component Vectors . . . . .	58
4.3.1 Effect of Number of Features . . . . .	58
4.3.2 Effect of Number of Training Samples . . . . .	66
4.4 Experiments and Analysis on Class-Specific Subspace-Based Two-Dimensional Principal Component Analysis . . . . .	68
4.5 Experiments and Analysis on Image Cross-Covariance Analysis . . . . .	71
4.6 Experiments and Analysis on Two-Dimensional Random Subspace Analysis . . . . .	81
4.6.1 Effect of Feature Dimension of Random Subspace and Number of Classifiers . . . . .	81



Chapter	Page
4.6.2 The experiment results of Two-Dimensional Diagonal Random Subspace Analysis . . . . .	96
4.6.3 The experiment results of RSM-based ICCA . . . . .	98
4.7 Summary . . . . .	100
<b>V The Conclusions . . . . .</b>	<b>103</b>
5.1 Conclusions of The Dissertation . . . . .	103
5.2 Future Directions . . . . .	104
<b>References . . . . .</b>	<b>105</b>
<b>Appendices . . . . .</b>	<b>111</b>
Appendix A . . . . .	112
Appendix B . . . . .	113
<b>Vitae . . . . .</b>	<b>115</b>



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## List of Tables

Table	Page
2.1 The Bilateral Projection Scheme of 2DPCA with Iterative Algorithm. . . . .	22
2.2 Coupled Subspaces Analysis Algorithm. . . . .	23
2.3 Deleting The Columns Algorithm for Volume Measure . . . . .	32
3.1 The Image Shifting Algorithm for ICCA . . . . .	45
3.2 Two-Dimensional Random Subspace Analysis Algorithm . . . . .	50
3.3 Two-Dimensional Diagonal Random Subspace Analysis Algorithm. . . . .	50
4.1 MSTAR images comprising training set . . . . .	55
4.2 MSTAR images comprising testing set . . . . .	55
4.3 The Highest Recognition Accuracy Comparisons of 2DPCA, 2DLDA and 2DPCA+2DLDA on Yale, ORL, AR, and MSTAR Databases . . . . .	59
4.4 The Highest Recognition Accuracy Comparisons of 2DPCA and CSS-based 2DPCA on Yale, ORL, AR, and MSTAR databases. . . . .	68
4.5 The Data Statistics of Results comparisons of 2DRSA Method on Yale, ORL, AR, and MSTAR Databases . . . . .	95
4.6 The Highest Recognition Accuracy Comparisons of 2DPCA and 2DRSA Method on Yale, ORL, AR, and MSTAR Databases . . . . .	95
4.7 The Highest Recognition Accuracy Comparisons of 2DPCA and 2D <sup>2</sup> RSA on Yale databases . . . . .	96
4.8 The Data Statistics of Results of RSM-based ICCA Method Comparisons on Yale, ORL, AR, and MSTAR Databases . . . . .	99
4.9 The Highest Recognition Accuracy Comparisons of 2DPCA and ICCA Method on Yale, ORL, AR, and MSTAR Databases . . . . .	99
4.10 The Summary Comparison of The Highest Recognition Accuracy Results on Yale, ORL, AR, and MSTAR Databases . . . . .	101

## List of Figures

Figure	Page
2.1 Projection of the data points $(x,y)$ onto the principal component directions $(v_1,v_2)$ . . . . .	11
2.2 Illustration of within-class and between-class . . . . .	13
2.3 Illustration of the Linear Discriminant Analysis of principal components framework . . . . .	13
2.4 An illustration of the curse of dimensionality. . . . .	15
2.5 The flowchart of Coupled Subspaces Analysis: (a) The original image. (b) All the rows of the images are treated as the objects. (c) The low-dimensional representation after the projection with the right multiplying projection. (d) The reconstructed images with the right multiplying projection. (e) All the columns of the images are treated as the objects. (f) The low dimensional representation after the projection with left and right multiplying projection. .	24
2.6 Illustration of the ways for deriving the diagonal face images: If the number of columns is more than the number of rows . . . . .	26
2.7 Illustration of the ways for deriving the diagonal face images: If the number of columns is less than the number of rows . . . . .	26
2.8 The sample diagonal face images on Yale database. . . . .	26
2.9 The PCA's MSE on the training set and the testing set as the function of feature dimension. . . . .	29
2.10 The 2DPCA's MSE on the training set and the testing set as the function of feature dimension. . . . .	29
3.1 CSS-based 2DPCA diagram. . . . .	41
3.2 The relationship of covariance and image covariance matrix. . . . .	43
3.3 The samples of shifted images on the ORL database. . . . .	46
4.1 The sample images of one subject in the Yale database. . . . .	52
4.2 Five sample images of one subject in the ORL face database. . . . .	53
4.3 The sample images of one subject in the AR database. . . . .	53
4.4 Sample SAR images of MSTAR database: the upper row is BMP2 APCs, the middle row is BTR70 APCs, and the lower row is T72 tank. . . . .	54
4.5 Aspect and Depression Angles. . . . .	55
4.6 Preprocessing diagram . . . . .	56
4.7 Cropping image . . . . .	57
4.8 Recognition accuracy of 2DPCA+2DLDA on Yale database. . . . .	60
4.9 Performance map of 2DPCA+2DLDA on Yale database. . . . .	60
4.10 Recognition accuracy of 2DPCA+2DLDA on ORL database. . . . .	61
4.11 Performance map of 2DPCA+2DLDA on ORL database. . . . .	61
4.12 Recognition accuracy of 2DPCA+2DLDA on AR database. . . . .	62
4.13 Performance map of 2DPCA+2DLDA on AR database. . . . .	62
4.14 Recognition accuracy of 2DPCA+2DLDA on MSTAR database. . . . .	63
4.15 Performance map of 2DPCA+2DLDA on MSTAR database. . . . .	63
4.16 Recognition accuracy of 2DPCA, 2DLDA and 2DPCA+2DLDA on the Yale database. . . . .	64

Figure	Page
4.17 Recognition accuracy of 2DPCA, 2DLDA and 2DPCA+2DLDA on the ORL database. . . . .	64
4.18 Recognition accuracy of 2DPCA, 2DLDA and 2DPCA+2DLDA on the AR database. . . . .	65
4.19 Recognition accuracy of 2DPCA, 2DLDA and 2DPCA+2DLDA on the M-STAR database. . . . .	65
4.20 Performance of 2DPCA, 2DLDA and 2DPCA+2DLDA on the Yale database with different training numbers for each subject. . . . .	67
4.21 Performance of 2DPCA, 2DLDA and 2DPCA+2DLDA on the ORL database with different training numbers for each subject. . . . .	67
4.22 Recognition accuracy of CSS-based 2DPCA and 2DPCA on the Yale database.	69
4.23 Recognition accuracy of CSS-based 2DPCA and 2DPCA on the ORL database.	69
4.24 Recognition accuracy of CSS-based 2DPCA and 2DPCA on the AR database.	70
4.25 Recognition accuracy of CSS-based 2DPCA and 2DPCA on the MSTAR database. . . . .	70
4.26 Recognition accuracy of ICCA on the Yale database. . . . .	72
4.27 Performance map of ICCA on the Yale database. . . . .	72
4.28 Recognition accuracy of ICCA on the ORL database. . . . .	73
4.29 Performance map of ICCA on the ORL database. . . . .	73
4.30 Recognition accuracy of ICCA on the AR database. . . . .	74
4.31 Performance map of ICCA on the AR database. . . . .	74
4.32 Recognition accuracy of ICCA on the MSTAR database. . . . .	75
4.33 Performance map of ICCA on the MSTAR database. . . . .	75
4.34 Recognition accuracy of varying the number of image shifting of ICCA on the Yale database. . . . .	76
4.35 A period of varying the number of image shifting of ICCA on the Yale database. . . . .	76
4.36 Recognition accuracy of varying the number of image shifting of ICCA on the ORL database. . . . .	77
4.37 A period of varying the number of image shifting of ICCA on the ORL database. . . . .	77
4.38 Recognition accuracy of varying the number of image shifting of ICCA on the AR database. . . . .	78
4.39 A period of varying the number of image shifting of ICCA on the AR database.	78
4.40 Recognition accuracy of varying the number of image shifting of ICCA on the MSTAR database. . . . .	79
4.41 A period of varying the number of image shifting of ICCA on the MSTAR database. . . . .	79
4.42 The samples of the shifted images which obtain the best recognition accuracy on the Yale database. . . . .	80
4.43 The samples of the shifted images which obtain the best recognition accuracy on the ORL database. . . . .	80



Figure	Page
4.44 The samples of the shifted images which obtain the best recognition accuracy on the AR database. . . . .	80
4.45 The samples of the shifted images which obtain the best recognition accuracy on the MSTAR database. . . . .	80
4.46 Recognition accuracy of 2DRSA on the Yale database. . . . .	83
4.47 Performance map of 2DRSA on the Yale database. . . . .	83
4.48 Recognition accuracy of 2DRSA on the ORL database. . . . .	84
4.49 Performance map of 2DRSA on the ORL database. . . . .	84
4.50 Recognition accuracy of 2DRSA on the AR database. . . . .	85
4.51 Performance map of 2DRSA on the AR database. . . . .	85
4.52 Recognition accuracy of 2DRSA on the MSTAR database. . . . .	86
4.53 Performance map of 2DRSA on the MSTAR database. . . . .	86
4.54 The priori performance map of 2DRSA on Yale database. . . . .	87
4.55 The priori performance map of 2DRSA on ORL database. . . . .	87
4.56 The priori performance map of 2DRSA on AR database. . . . .	88
4.57 The priori performance map of 2DRSA on MSTAR database. . . . .	88
4.58 Averaging results of 2DRSA by varying the feature dimensions on Yale database. . . . .	89
4.59 Averaging results of 2DRSA by varying the number of classifiers on Yale database. . . . .	89
4.60 Averaging results of 2DRSA by varying the feature dimensions on ORL database. . . . .	90
4.61 Averaging results of 2DRSA by varying the number of classifiers on ORL database. . . . .	90
4.62 Averaging results of 2DRSA by varying the feature dimensions on AR database. . . . .	91
4.63 Averaging results of 2DRSA by varying the number of classifiers on AR database. . . . .	91
4.64 Averaging results of 2DRSA by varying the feature dimensions on MSTAR database. . . . .	92
4.65 Averaging results of 2DRSA by varying the number of classifiers on MSTAR database. . . . .	92
4.66 Boxplot of 2DRSA in term of the number of classifiers. . . . .	93
4.67 Boxplot of 2DRSA in term of the feature dimensions. . . . .	93
4.68 Boxplot of 2DRSA in term of the number of classifiers and the feature dimensions in prospected ranges. . . . .	94
4.69 Recognition accuracy of DiaPCA and 2DPCA on Yale database. . . . .	97
4.70 Recognition accuracy of 2D <sup>2</sup> RSA on Yale database. . . . .	97
4.71 Boxplot of RSM-based ICCA on Yale, ORL, AR, and MSTAR databases. . . . .	98
4.72 The Summary Comparison of The Highest Recognition Accuracy Results on Yale, ORL, AR, and MSTAR Databases. . . . .	102

# CHAPTER I

## INTRODUCTION

### 1.1 Background and Signification of the Research Problems

Pattern recognition is still a very active field of research. It can be applied to almost every branch of the engineering sciences, such as physics, chemistry and computer science. The aim of pattern recognition can be defined as the categorization of input data into identifiable classes via the extraction of significant features or attributes of the data from a background of irrelevant detail. The input and output of pattern recognition system are pattern (or feature) and pattern class, respectively. The pattern is the description of any member of a category representing a pattern class. For convenience, patterns are usually represented by a vector. It is often useful to think of a pattern vector as a point in an  $n$ -dimensional Euclidean space. Determination and then the discrimination of these pattern vectors form the two major problems in pattern recognition system design. If we can extract feature vector, then we do have a priori knowledge about the pattern to be recognized. Under these circumstances pattern recognizing machines are best designed using a training or learning procedure. Arbitrary decision functions are initially assumed, and through a sequence of iterative training steps these decision functions are made to approach optimum or satisfactory forms. The pattern class is given by user. The features of a pattern class are the characterizing attributes common to all patterns belonging to that class. Such features are often referred to as *intrasets* features. The features which represent the differences between pattern classes may be referred to as the *intersets* features.

After the image can be stored in digital format and the computer is powerful enough for processing these data, many pattern recognition systems for image analysis were arisen such as the automatic recognition of handwritten, the automatic recognition of images of human faces, the automatic target recognition for Synthetic Aperture Radar (SAR) images, etc. The number of real world applications (e.g. surveillance, secure access, human computer interface) and the availability of cheap and powerful hardware also lead to the development of commercial pattern recognition systems.

### 1.2 Literature Review

In pattern recognition, subspace analysis is the successful feature extraction method in many applications such as face recognition [1–20], Automatic Target Recognition (ATR) [21–23], speech recognition [24,25], character recognition [26,27], musical in-

strument sound recognition [28], etc. In this dissertation, only image pattern applications are discussed, especially for face recognition and ATR.

Generally, the original data space are high-dimensional that consume large amount of memory and time for computation. Furthermore, the non useful information, which contain in this space, will be the cause of the misclassification. In subspace analysis, the high-dimensional space is compacted into a considerably low dimensional subspace, and then we design classifiers in this subspace. Therefore, the low-dimensional subspace must be good enough to represent the original space. On the one hand, by reducing the dimensionality of data space, it alleviates the risk of bad estimation but improves the generalization capability [29]. However, it may cause the loss of discriminative information. Balancing these two sides is significant.

The subspace analysis techniques can be categorized into two main categories: unsupervised and supervised techniques. The supervised technique is a technique for creating a function from training data. The training data consist of pairs of input objects, and desired outputs. The task of the supervised technique is to predict the value of the function for any valid input object after having seen a number of training examples. Supervised technique can generate models of two types. Most commonly, it generates a global model that maps input objects to desired outputs. In some cases, the map is implemented as a set of local models. The unsupervised technique differs from the supervised technique by the fact that they do not use a priori output for each input object. Indeed, the supervised technique is usually part of the pattern recognition problem where we dispose class label for each object. This additional information allows choosing the subspace which suitable for classification problem. In the following we will review both subspace analysis techniques: unsupervised and supervised.

There are many works that unsupervised techniques are applied [1–3, 6, 8, 14, 15, 30]. Among these works, *Principal Component Analysis* (PCA) is the most commonly used technique. PCA is constructed around the criteria of preserving the data distribution. One of the widely known works on PCA is due to Turk and Pentland [3], called *Eigenface*. Eigenface is a successful technique in face recognition. The image is projected in a space in which the correlation among the components is zero. This space transformation is also known as *Karhunen-Loeve transform*. Every image in the database can be represented as a vector of weights which are obtained by projecting the image into Eigenface components by a simple inner product operation. When a new test image whose identification is required is given, the new image is also represented by its vector of weights. The identification of the test image is done by locating the image in the database whose weights are the closest, in Euclidean distance, to the weights of the test image.

In supervised techniques, *Linear Discriminant Analysis* (LDA) or Fisher Discriminant Analysis (FDA) is cited in a great number of successful works in recognition systems [16–18, 31–34]. In LDA's criterion, the image is projected in the subspace,

in which the variability among the image vectors of the same class (within class) is minimized and the variability among the face-vectors of different classes (between class) is maximized. However, LDA has two drawbacks when directly applied to the original input space [18]. First of all, some non useful information such as image background may be regarded by LDA as the discriminant information. This causes misclassification when the face of the same subject is presented on different background. Secondly, when the available number of training samples is relatively small compared to the feature dimension, the within-class scatter matrix be singular and the LDA cannot be applied.

The combination of unsupervised and supervised techniques can overcome some of the above problems. In Fisherface [16] and the discriminant analysis of principal components framework [17,18] demonstrate a significant improvement, they consist of LDA over the data projected onto the principal subspace since both techniques are complementary. Indeed, as described previously, PCA is a unsupervised technique that has no class information. Therefore, PCA does not provide any discriminative information. However, PCA can be used to discard non discriminative information which allows the improvement of the normal LDA. While LDA provides the discriminative information via the criterion which considers the relationships of both within and between-class scatter matrices. On the one hand, the dimension of the feature space is reduced by PCA. Therefore, the estimation of scatter matrices in LDA can be improved. Especially, when there are prominent variation in lighting condition and expression [16–18].

Another combination of unsupervised and supervised techniques framework is called Class-Specific Subspace (CSS), Individual Eigenface Subspace [35,36] or Face-Specific Subspace [37]. In traditional PCA [3], the samples are analyzed on the features extracted in a low-dimensional space learned from all training samples from all classes, called a universal Eigenface subspace. This subspace is optimal for representing training samples from all classes in the minimum mean squared error sense, it may not adequately capture or describe the detailed information that discriminates the samples in each class from another. While each subspace of CSS learned from only the training samples in same class. In this way, the CSS representation can provide a minimum reconstruction error. The reconstruction error is used to classify the input data via the Distance From CSS (DFCSS). Less DFCSS means more probability that the input data belongs to the corresponding class. There are many advantages of using CSS. First of all, the transformation matrices are trained from samples in their class, thus it is more suitable for representation the sample in their own class than a transformation matrix which is trained by samples in all classes. Secondly, the DFCSS is the distance between original image and its reconstruction image from CSS, therefore it requires only the memory for storing the transformation matrices as the number of classes. There is no more need to keep the feature vectors of all training samples when the nearest neighbor classifier are applied. Thirdly, the number of classes is usually less than the number of training samples, so the number of distance calculation of CSS-based method is also less than the number



of distance calculation in conventional methods. However, there are many drawbacks in this approach. First of all, since the number of training samples is limited thus the available number of training samples for separated estimating each individual subspace may be too small. To solve this problem, the authors [37] suggest a simple technique to derive multiple samples from a single example image. The technique is based on the following two intuitive propositions: proper geometric transforms (translation, rotation in image plane, scale changes, etc.) and proper gray-level transforms (simulative directional lighting, man-made noise, etc.). Secondly, this approach requires more computation time than the original PCA to calculate all of these CSSs.

However, all of above methods, the 2D image matrix must be previously transformed into vector. Therefore, the dimension of the feature space is very large. This leads to many problems in real application. First of all, working in the large dimensional feature space means that the dimension of the scatter matrix will be very large thus requires an enormous amount of training samples to obtain a good estimator. This is known as the curse of dimensionality [38]. Secondly, the spatial structure information could be lost in transformation process. Finally, if the available number of training samples is relatively very small compared to the feature dimension, the covariance matrix which estimated by these features will be singular, called singularity problem or Small Sample Size (SSS) problem [31].

Recently, Yang et al. [39] proposed an original technique called *Two-Dimensional Principal Component Analysis* (2DPCA), in which the image covariance matrix is computed directly on image matrices so the spatial structure information can be preserved. This yields a covariance matrix whose dimension just equals to the width of the face image. This is far smaller than the size of covariance matrix in PCA. Therefore, the image covariance matrix can be better estimated and will usually be full rank. That means the curse of dimensionality and the SSS problem can be avoided. Evidently, the experimental results in [39–42] have shown the improvement of 2DPCA over PCA on several face databases. As previously described, the directions that maximize the scatter of the data from all training samples might not be as adequate to discriminate between classes. In recognition task, a projection is required to emphasize the discrimination between classes.

A supervised subspace analysis technique in 2D subspace called *Two-Dimensional Linear Discriminant Analysis* (2DLDA) was proposed in [43]. This approach overcomes the SSS problem in classical LDA by working with images in matrix representation, like in 2DPCA. In particular, bilateral projection scheme was applied here same as in B2DPCA [40,41]. In this way, two eigenvalue problems were solved per iteration. One corresponds to the column direction and another one corresponds to the row direction of image, respectively. However, 2DLDA still gathers some non useful information such as image background as the discriminant information, like the classical LDA.

All of above techniques attempt to build only one strong classifier to manipulate

data space. However, when data are highly dimensional and the training sample size is small compared to the data dimensionality, it may be difficult to construct a good single classification rule. Usually, a classifier constructed on small training sets is biased and has a large variance as the classifier parameters (coefficients) are poorly estimated. Consequently, such a classifier may be weak, having a poor performance. Moreover, often it will be unstable: small changes in the training set cause large changes in the classifier. In general, bad performance of a classifier can be caused by different factors: incorrect assumptions about the model when constructing a classifier; too low a complexity of the classification rule to solve the problem; incorrect settings for classifier parameters; instability of the classifier, etc. Consequently, in the literature the term *weak classifier* can refer to different things: badly performing classifiers, unstable classifiers, classifiers of a low complexity, or classifiers depending upon certain assumed models that are not always true. But the performance of the weak classifier should satisfy slightly better than random guessing. However, in all cases when intending to improve a weak classifier, one actually aims to improve its performance. Therefore, describing a weak classifier as one that has a poor performance seems to be the most general definition.

To improve a weak classifier, one approach is to construct many weak classifiers instead of a single one, and to combine them into a powerful decision rule. Recently, a number of combining techniques has been developed. The most popular are bagging [44], boosting [45] and the Random Subspace Method (RSM) [46]. In bagging, one samples the training set, generating random independent bootstrap replicates, constructs the classifier on each of these, and aggregates them by a simple majority vote in the final decision rule. In boosting, classifiers are constructed on weighted versions of the training set, which is dependent on previous classification results. Initially, all objects have equal weights, and the first classifier is constructed on this data set. Then, weights are changed according to the performance of the classifier. Erroneously classified objects get larger weights, and the next classifier is boosted on the reweighted training set. In this way, a sequence of training sets and classifiers is obtained, these classifiers are then combined by simple majority voting or by weighted majority voting in the final decision. In the random subspace method, classifiers are constructed in random subspaces of the data feature space. These classifiers are usually combined by simple majority voting in the final decision rule. The ensemble methods were applied to face and automatic recognitions in [47–49].

The RSM is useful for weak linear classifiers obtained on small and critical training sample sizes. It improves the performance of linear classifiers which often suffer from the curse of dimensionality and having decreasing learning curves [50]. In the 2D subspace methods such as 2DPCA or 2DLDA, the performance will be decreasing when the dimension of feature increases. Therefore, the RSM is appropriate for the 2D subspace methods because the RSM constructs multiple classifiers in random feature subspaces, that means the dimension of feature is reduced. While Bagging and Boosting build the

multiple classifiers from training samples, the dimension of feature remain as ever.

In this dissertation, four frameworks are applied to improve the recognition performance of 2D subspace analysis. The first and the second frameworks are the combination of unsupervised and supervised techniques frameworks. The generalized version of 2D-PCA is presented in the third framework. And The random subspace method is applied to 2D subspace analysis in the fourth framework.

The first framework is based on discriminant analysis of principal components framework or Fisherface. Because of 2DPCA and PCA is more suitable for face representation than face recognition. Therefore, LDA is necessary for better performance in recognition task. Unfortunately, the linear transformation of 2DPCA reduces only the size of rows. The number of rows still equal to the height of original image. Thus, the SSS problem will be appeared when LDA is performed after 2DPCA directly. To overcome this problem, we proposed a simplified version of the bilateral 2DLDA by using only unilateral projection scheme, based on the 2DPCA concept. According to the previous discussion on 2DLDA, our proposed 2DLDA method is different from [43] on the fact that our method aims to find the optimal discriminated transformation for projecting the principal component vectors, which obtained from 2DPCA step, In this way, the projection direction of our 2DLDA is also corresponded to the direction which is projected in 2DPCA step, that is the row direction of images. Applying our proposed 2DLDA to 2DPCA can solve not only the SSS problem and the curse of dimensionality dilemma but also allows us to work directly on the image matrix in all projections. Hence, spatial structure information is employed and the size of all scatter matrices cannot be greater than the width of face image. Furthermore, computing with this dimension, the face image do not need to be resized, since all information still be preserved.

The second framework is based on Class-Specific Subspace (CSS). By applying CSS over 2DPCA, the class information is introduced to unsupervised method. Only one subspace is considered in the conventional 2DPCA. All training samples are used to construct this subspace. Unlike 2DPCA, many subspaces are constructed in CSS framework. Each subspace of CSS learned from only the training samples in own class. In this way, the CSS representation can provide a minimum reconstruction error. Which it can be used to classify the input data. However, because of the number of the training samples is limited in many real-world applications, thus the training samples which provides for each CSS will be very rare. It means the CSS in 1D subspace analysis will suffer from the SSS problem. Fortunately, in 2DPCA, the non-singular image covariance matrix can be estimated via only a few of training samples. Indeed, the exact number of training samples, in 2DPCA, is not just equal to the number of images but it equals to the number of images multiply by the number of rows of each image [51]. Therefore, the singularity or SSS problem is avertible even though in the CSS framework.

The third framework is base on the generalize form of image covariance matrix, called *image cross-covariance matrix*. Compare to the covariance matrix of PCA, the

image covariance matrix discard some of the information. This disregard information may possibly be useful for discrimination. The image cross-covariance matrices are formulated by two variables, the original image and its shifted version. By our shifting algorithm, many image cross-covariance matrices are formulated to cover all of the information in which discarded by the image covariance matrix.

The fourth framework applies the random subspace method [46] to 2DPCA. Normally, the feature of 2DPCA is a matrix. In the row direction, the number of the columns of these matrix is affected by the number of selected eigenvalues of the image covariance matrix while the number of selected eigenvalues is not influenced in the column direction. Thus, the number of the rows is still equal to the height of original image and the random subspace method can be apply in the column direction. The random subspaces are constructed by randomly selecting a number of rows of the original feature matrix. The multiple classifiers are constructed in these random subspaces of the data feature space. These classifiers are usually combined by simple majority voting in the final decision rule. Moreover, we applied this framework to the Diagonal Principal Component Analysis (DiaPCA) [52], a 2DPCA with diagonal face, and used for selecting the subspaces which constructed by the third framework.

This dissertation has a main objective to improve the recognition performance in two-dimensional subspace based technique, via including class information to conventional unsupervised techniques, taking the benefit of the Random Subspace Method and generalizing the 2DPCA. The results in the dissertation would be beneficial for face recognition and other image understanding systems, especially in the automatic target recognition.

### 1.3 Objectives

Propose the novel two-dimensional subspace analysis techniques with application to image pattern recognition. The two-dimensional subspace analysis directly performs with the images in original matrix form. In this way, it can overcome the Singularity or Small Sample Size (SSS) problem in traditional subspace analysis, and preserved the spatial structure information of images. The frameworks of *Two-Dimensional Linear Discriminant Analysis of Principal Component Vectors*, *Class-Specific Subspace of Two-Dimensional Principal Component Analysis*, *Image Cross-Covariance Analysis* and *Two-Dimensional Random Subspace Analysis* are proposed to increase the recognition performance.

The subspace analysis techniques aim to reduce noise and extract the discriminative information for classification. Therefore, the two case studies, face and automatic target recognitions are used to evaluate the performance of the proposed techniques. The face images inherently affected by pose, expression and illumination, while the Synthetic aperture radar (SAR) images are inherently affected by multiplicative speckle noise,



which is due to the coherent nature of the scattering phenomenon. These applications are totally affected by difference noises.

## 1.4 Scope

1. Study the performance and the limitation of the classical subspace analysis used in the pattern recognition.
2. Develop the novel two-dimensional subspace analysis techniques that can apply to the pattern recognition systems on standard databases.
  - (a) Face image recognition
    - i. Only the 2D image of full frontal view faces will be presented to the face recognition system.
    - ii. The face areas are detected manually.
    - iii. The frontal view face recognition system should be extended to a expression invariant face recognition system.
  - (b) The non-face image recognition such as Automatic Target Recognition (ATR).
3. Study the performance of the proposed techniques and provide a comparative study of proposed techniques with other traditional techniques proposed by many authors in the past.

## 1.5 Expected Prospects

1. Acquire a basic knowledge of subspace analysis for applying to pattern recognitions.
2. Obtain the new subspace analysis techniques.
3. Obtain new pattern recognition systems.
4. Publish the international journal or conference papers.
5. Know the advantages and disadvantages of using the proposed subspace analysis techniques in pattern recognitions.
6. Understand the necessity of the subspace analysis techniques for pattern recognitions.

## 1.6 Research Procedure

1. Study previous research papers relevant to the research works of the dissertation.
2. Develop the novel subspace analysis techniques.
3. Develop the high accuracy pattern recognition techniques.
4. Develop simulation programs.
5. Test the proposed algorithms by using standard face databases such as AR, ORL and Yale.
6. Perform the proposed algorithm on a non-face database, MSTAR, in the application of Automatic Target Recognition.
7. Collect and analyze computational results obtained from simulation programs.
8. Summarize the major findings as we found in step 7 and conclude the performance of the proposed framework in all concerned aspects.
9. Publish the international journal or conference papers.
10. Check whether the conclusions meet all the objectives of the research work of the dissertation.
11. Write the dissertation.

# CHAPTER II

## BASIC BACKGROUND AND RELATED TOPICS

In this chapter, the fundamental knowledge of the two-dimensional subspace analysis algorithm is described. First of all, the traditional 1D subspace, i.e. all points in this subspace are represented in vector form, analysis techniques and its variants. Including the serious problems of the subspace analysis and the ensemble method for subspace analysis, called Random Subspace Method (RSM), are reviewed. And then, the 2D subspace, i.e. all points in this subspace are represented in matrix form, analysis is introduced via a Two-Dimensional Principal Component Analysis (2DPCA) framework. Finally, the comparisons of PCA and 2DPCA are discussed, including the distance measurement, are described.

### 2.1 One-Dimensional Subspace Analysis and Its Variants

In this section, we reviewed the traditional 1D subspace analysis techniques, i.e. Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA). And then the frameworks of 1D subspace analysis, the linear discriminant of principal components and Class-Specific Subspace-based PCA are discussed. after that the one of ensemble method, the serious problems of the subspace analysis are criticized. Finally, the ensemble methods for subspace analysis are presented.

#### 2.1.1 Principal Component Analysis (PCA)

Principal components analysis (PCA) is the one of unsupervised subspace method, which used to reduce multidimensional data sets to lower dimensions for analysis.

Let  $\mathbf{A}$  is the  $m$  by  $n$  matrix of pixel's intensity of the image and the image vector,  $\gamma$ , is the vector of  $\mathbf{A}$  which was previously transformed by column-stack vectorization. Thus, the dimension of  $\gamma$  is  $mn \times 1$ . The average of  $\gamma$  can be found as

$$\psi = \frac{1}{M} \sum_{i=1}^M \gamma_i \quad (2.1)$$

where  $M$  is the number of training images. The zero-mean normalization is applied to all image vectors by

$$\phi_i = \gamma_i - \psi, i = 1, 2, 3, \dots, M, \quad (2.2)$$

where  $\phi_i$  is the  $i^{th}$  zero-mean normalized of  $\gamma_i$ . The covariance matrix,  $\mathbf{C}$ , of these

image vectors can be calculated as

$$\mathbf{C} = \Phi\Phi^T, \quad (2.3)$$

where  $\Phi = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_M \end{bmatrix}$ .

The PCA is defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the principal component directions, as shown in Fig. 2.1. This transformation is therefore equivalent to finding the eigenvalue decomposition of the matrix  $\mathbf{C}$ .

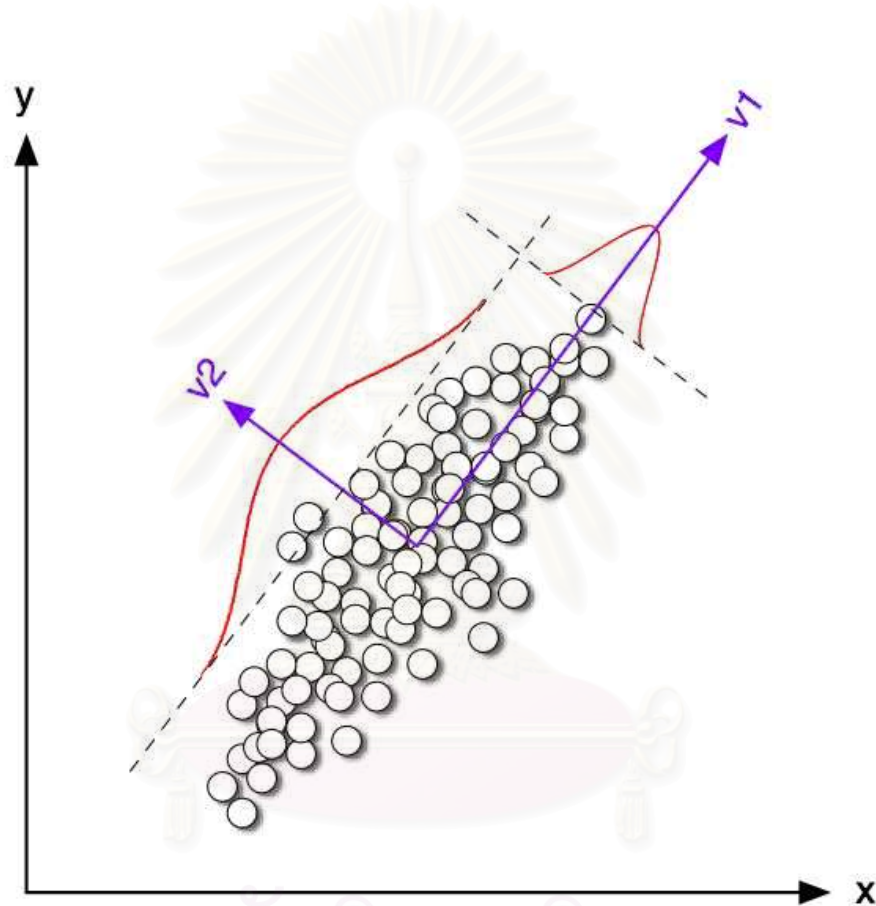


Figure 2.1: Projection of the data points  $(x,y)$  onto the principal component directions  $(v_1,v_2)$

According to the dimension of  $\phi$ , the dimension of  $\mathbf{C}$  will be  $mn \times mn$  which is normally quite large for calculating the eigenvalue decomposition. However, since the number training samples is normally smaller than the dimension of  $\phi$  therefore the non-zero eigenvalues of this covariance matrix can be found in another way via a new matrix,

$$\mathbf{L} = \Phi^T\Phi. \quad (2.4)$$

The dimension of  $\mathbf{L}$  is only  $M \times M$ , thus the eigenvalue decomposition of  $\mathbf{L}$  can be



done easier than  $\mathbf{C}$ . The eigenvalue decomposition of  $\mathbf{L}$ ,

$$\mathbf{L} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T, \quad (2.5)$$

where  $\mathbf{\Lambda}$  is the diagonal matrix which contains the eigenvalues of  $\mathbf{L}$  and  $\mathbf{V}$  contains a set of eigenvectors of  $\mathbf{L}$ . Finally, the eigenvectors of  $\mathbf{C}$  which correspond to the non-zero eigenvalues of  $\mathbf{C}$  can be determined by

$$\mathbf{U} = \mathbf{\Phi}\mathbf{V}, \quad (2.6)$$

where  $\mathbf{U}$  is the matrix that contains a set of eigenvectors of  $\mathbf{C}$ .

The eigenvector associated with the largest eigenvalue has the same direction as the first principal component, the eigenvector associated with the second largest eigenvalue determines the direction of the second principal component, and so on. Since the lower-order principal components often contain the most important aspects of the data, the dimension of projected space can be reduced by retaining those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones.

### 2.1.2 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis is a supervised subspace method in the sense that it represents data to that are useful for classification [31]. From Eq. (2.2), each normalized image vector  $\phi_i$  belongs to one of  $K$  classes  $\{\omega_1, \omega_2, \omega_3, \dots, \omega_K\}$ . The between-class scatter matrix  $\mathbf{S}_b$  and the within-class scatter matrix  $\mathbf{S}_w$  are defined as

$$\mathbf{S}_b = \sum_{k=1}^K \frac{n_k}{K} (\bar{\phi}_k - \bar{\phi})(\bar{\phi}_k - \bar{\phi})^T \quad (2.7)$$

$$\mathbf{S}_w = \sum_{k=1}^K \frac{n_k}{K} \sum_{\phi_i \in \omega_k} (\phi_i - \bar{\phi}_k)(\phi_i - \bar{\phi}_k)^T, \quad (2.8)$$

where  $n_k$  and  $\bar{\phi}_k$  are the number of samples and the expected value of class  $\omega_k$  respectively. And  $\bar{\phi}$  denotes the global mean of the entire samples. LDA finds a projection matrix,  $\mathbf{U}$ , that maximizes the Fisher's criterion, the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix as

$$\mathbf{U}_{opt} = \arg \max_{\mathbf{U}} \frac{|\mathbf{U}^T \mathbf{S}_b \mathbf{U}|}{|\mathbf{U}^T \mathbf{S}_w \mathbf{U}|}, \quad (2.9)$$

where  $\mathbf{U}_{opt}$  is the optimal solution of the projection matrix which is a set of generalized eigenvectors of  $\mathbf{S}_w^{-1} \mathbf{S}_b$ . This criterion attempts to minimize the variability among the image vectors of the same class (with-in class) while the variability among the image vectors of different classes (between class) is maximized, as shown in Fig. 2.2 Usually, the singularity of the within-class scatter matrix commonly encountered when the number of the training samples is too small.

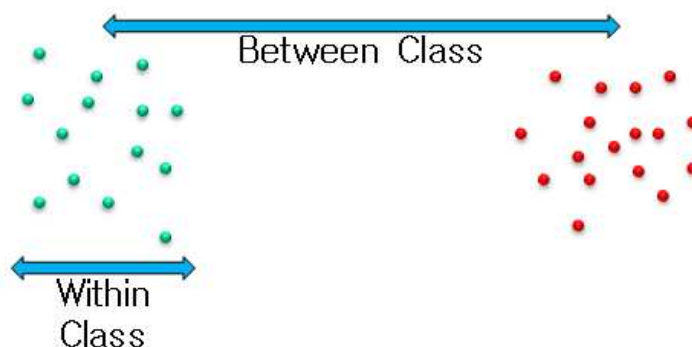


Figure 2.2 Illustration of within-class and between-class

### 2.1.3 Linear Discriminant Analysis of Principal Components

In linear discriminant analysis of principal components framework [17, 18] and Fisherface [16], The combination of unsupervised and supervised techniques had been used for face recognition.

The unsupervised technique, like PCA, has no class information. Thus, the pure PCA algorithm does not provide any discriminative information. While LDA does not perform very well when LDA directly applied to the original input space. Because of the first reason, some non useful information such as image background are regarded by LDA as the discriminant information. This causes misclassification when the face of the same subject is presented on different background. And the second reason, the within-class scatter matrix be singular when the available number of training samples is relatively small compared to the feature dimension, singularity problem. Therefore the LDA cannot be applied.

All above problems can be solved by firstly projecting the original images into PCA subspace, approximate the original data with lower dimensional feature vectors, in which the non discriminative information can be discarded. And then LDA are performed in this subspace for discrimination these features. The illustration of this framework is shown in Fig. 2.3



Figure 2.3: Illustration of the Linear Discriminant Analysis of principal components framework

### 2.1.4 Class-Specific Subspace-Based PCA

In [37], the Class-Specific Subspace (CSS) was proposed and applied to PCA, named Face-Specific Subspace (FSS). The difference from traditional method is the covariance matrix of the  $k^{th}$  class is evaluated from training samples of the  $k^{th}$  class, individually. The  $k^{th}$  CSS was represented as a 4-tuple: the projection matrix, the mean of the  $k^{th}$  class, the eigenvalues of covariance matrix and the dimension of the  $k^{th}$  FSS. For identification, the input sample is projected to all CSSs and then reconstruct by that CSS. If reconstruction error which obtained from the  $k^{th}$  CSS is minimum then the input sample is belong to the  $k^{th}$  class, so called Distance From CSS (DFCSS). In this dissertation, the CSS is adapted to 2DPCA framework.

### 2.1.5 Problems in Subspace Analysis

Three serious problems, i.e. the curse of dimensionality, the Small Sample Size (SSS) problem and the over-fitting problem, are reviewed here.

#### 2.1.5.1 The Curse of Dimensionality

The curse of dimensionality is a term coined by Richard Bellman [38] to describe the problem caused by the rapid increase in volume associated with adding extra dimensions to a space.

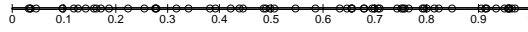
It is a significant obstacle in high dimension data analysis, which refers to the fact that a local neighborhood in higher dimensions is no longer local, or to put it another way, the sparsity increases exponentially given a fixed amount of data points.

For illustration, 64 data points are simulated from a uniform (0, 1) distribution. In one dimensional space, all the data points are clustered together, as shown in Fig. 2.4(a) However, in two dimensional space, the data become much more sparse, as shown in Fig. 2.4(b). And this is even obvious in three dimensional space, as shown in Fig. 2.4(c). Thus to achieve the same accuracy, much larger data sets are needed even when dimension is moderate and such large data sets are not available in practical situation.

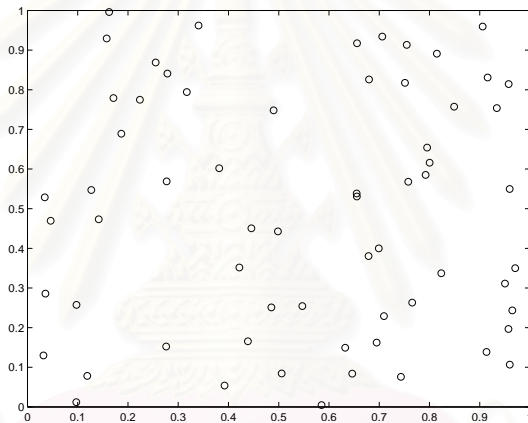
The curse of dimensionality is a significant obstacle in machine learning problems that involve learning a *state-of-nature* from a finite number of data samples in a high-dimensional feature space.

#### 2.1.5.2 The Small Sample Size (SSS) Problem

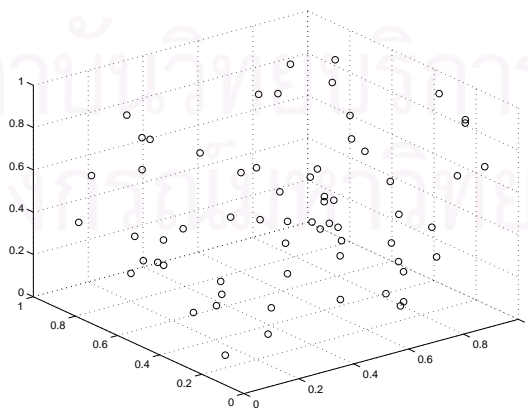
When the dimension of the feature space is larger than the number of training examples per class, not all parameters can be estimated accurately. Consequently, pattern



(a) One dimensional space



(b) Two dimensional space



(c) Three dimensional space

Figure 2.4 An illustration of the curse of dimensionality.



recognition techniques cannot perform well when they are used to analyze this space. This problem is called a *Small Sample Size* (SSS) problem [31].

For classification which is based on similarity measures that involve the inverse of the sample group covariance matrices, such as LDA. Unfortunately, these matrices are singular in SSS problems, therefore sometimes the SSS problem is called a *singularity* problem.

### 2.1.5.3 Over-fitting Problem

Usually a learning algorithm is trained using some set of training examples, i.e. exemplary situations for which the desired output is known. The learner is assumed to reach a state where it will also be able to predict the correct output for other examples, thus generalizing to situations not presented during training (based on its inductive bias). However, in cases where learning was performed too long or where training examples are rare, the learner may adjust to very specific random features of the training data, that have no causal relation to the target function. In this process of over-fitting, the performance on the training examples still increases while the performance on unseen data becomes worse.

Normally, in real-world application such as face recognition, the available number of training samples is very small. Thus when applied to image recognition, classical PCA is apt to be over-fitted to the training set due to the SSS problem, especially when the nearest neighbor classifier is applied.

### 2.1.6 Ensemble Methods for Subspace Analysis

In our focus application, the available number of training samples is inadequate. And the ensemble methods, like bagging or boosting, build the multiple classifiers from training samples. Therefore, in this dissertation, only Random Subspace Method (RSM) is applied to subspace analysis framework because the RSM constructs multiple classifiers in random feature subspaces, that means the dimension of feature is reduced. From this reason, the lower dimensional subspaces that were constructed by RSM could be far away from the problems in previous section.

#### 2.1.6.1 Random Subspace Method (RSM)

The Random Subspace Method (RSM) can be used to create multiple independent tree-classifiers that can be combined to improve accuracy [46]. And then apply the procedure to k-nearest-neighbor classifiers and show that it can achieve similar results [53]. In the RSM, one also modifies the training data like other ensemble methods. However, this modification is performed in the feature space. It randomly selects different feature dimensions and constructs multiple smaller subsets. When the feature dimension is smaller than the original one, the classifier, which is trained by this features in each

subset, will be weak classifier. However, the weak classifier will be more accurate when the combination rule is applied to a group of this weak classifiers in the end for prediction on the testing set. The RSM may benefit from using random subspaces for both constructing and aggregating the classifiers. When the number of training objects is relatively small compared with the data dimensionality, by constructing classifiers in random subspaces one may solve the SSS problem. The subspace dimensionality is smaller than in the original feature space, while the number of training objects remains the same. Therefore, the relative training sample size increases. When data have many redundant features, one may obtain better classifiers in random subspaces than in the original feature space. The combined decision of such classifiers may be superior to a single classifier constructed on the original training set in the complete feature space.



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## 2.2 Two-Dimensional Subspace Analysis and Its Variants

Traditionally, the image feature is storing in vector form, that leads to many problems in real world application, as describe in Chapter I.

The first 2D subspace based approach was introduced in [39], called Two-Dimensional Principal Component Analysis (2DPCA). The scatter matrix is computed directly on image matrices so the spatial structure information can be preserved. This yields a scatter matrix whose dimension equals to the width of the face image. This is far smaller than the size of covariance matrix in traditional one. Therefore, the scatter matrix can be better estimated and full rank in 2D subspace based method.

### 2.2.1 Image Covariance Matrix and Two-Dimensional Principal Component Analysis (2DPCA)

Normally in 1D subspace analysis, the 2D image matrices are first transformed to 1D image vectors by vectorization. The vectorization of a matrix is the column vector obtain by stacking the columns of the matrix on top of one another. The covariance or scatter matrix are formulated from the these image vectors. The covariance matrix will be well estimated if and only if the number of available training samples is not far smaller than the dimension of this matrix. In fact, it is too hard to collect this the number of samples. Then, normally in 1D subspace analysis, the estimated covariance matrix is not well estimated and not full rank.

Let each image is represented by a  $m$  by  $n$  matrix  $\mathbf{A}$  of its pixels' intensity. We consider linear projection of the form

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (2.10)$$

where  $\mathbf{x}$  is an  $n$  dimensional projection axis and  $\mathbf{y}$  is the projected feature of this image on  $\mathbf{x}$ , called *principal component vector*.

In original algorithm of 2DPCA [39], like PCA, 2DPCA search for the optimal projection by maximize the total scatter of projected data. Instead of using the criterion as in PCA, the total scatter of the projected samples can be characterized by the trace of the covariance matrix of the projected feature vectors. From this point of view, the following criterion was adopt as

$$J(\mathbf{x}) = \text{tr}(\mathbf{S}_\mathbf{x}), \quad (2.11)$$

where

$$\mathbf{S}_\mathbf{x} = E[(\mathbf{y} - E\mathbf{y})(\mathbf{y} - E\mathbf{y})^T]. \quad (2.12)$$

The total power equals to the sum of the diagonal elements or trace of the covariance matrix, the trace of  $\mathbf{S}$  can be rewritten as

$$\begin{aligned}
tr(\mathbf{S}_x) &= tr\{E[(\mathbf{y} - E\mathbf{y})(\mathbf{y} - E\mathbf{y})^T]\} \\
&= tr\{E[(\mathbf{A} - E\mathbf{A})\mathbf{x}\mathbf{x}^T(\mathbf{A} - E\mathbf{A})^T]\} \\
&= tr\{E[\mathbf{x}^T(\mathbf{A} - E\mathbf{A})^T(\mathbf{A} - E\mathbf{A})\mathbf{x}]\} \\
&= tr\{\mathbf{x}^T E[(\mathbf{A} - E\mathbf{A})^T(\mathbf{A} - E\mathbf{A})]\mathbf{x}\} \\
&= tr\{\mathbf{x}^T \mathbf{G}\mathbf{x}\}.
\end{aligned} \tag{2.13}$$

Giving that

$$\mathbf{G} = E[(\mathbf{A} - E\mathbf{A})^T(\mathbf{A} - E\mathbf{A})]. \tag{2.14}$$

This matrix  $\mathbf{G}$  is called *image covariance matrix*. Therefore, the alternative criterion can be expressed by

$$J(\mathbf{x}) = tr(\mathbf{x}^T \mathbf{G}\mathbf{x}), \tag{2.15}$$

where the image inner-scatter matrix  $\mathbf{G}\mathbf{x}$  is computed in a straightforward manner by

$$\mathbf{G} = \frac{1}{M} \sum_{k=1}^M (\mathbf{A}_k - \bar{\mathbf{A}})^T (\mathbf{A}_k - \bar{\mathbf{A}}), \tag{2.16}$$

where  $\bar{\mathbf{A}}$  denotes the average image,

$$\bar{\mathbf{A}} = \frac{1}{M} \sum_{k=1}^M \mathbf{A}_k. \tag{2.17}$$

It can be shown that the vector  $\mathbf{x}$  maximizing Eq. (2.13) correspond to the largest eigenvalue of  $\mathbf{G}$  [42]. This can be done, for example, by using the Eigenvalue decomposition or Singular Value Decomposition (SVD) algorithm. However, one projection axis is usually not enough to accurately represent the data, thus several eigenvectors of  $\mathbf{G}$  are needed. The number of eigenvectors  $d$  can be chosen according to a predefined threshold  $\theta$ .

Let  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  be eigenvalues of  $\mathbf{G}$  sorted in non-increasing order. We select  $d$  first eigenvectors such that their corresponding eigenvalues satisfy

$$\theta \leq \frac{\sum_{i=1}^d \lambda_i}{\sum_{i=1}^n \lambda_i}. \tag{2.18}$$

For feature extraction, Let  $\mathbf{x}_1, \dots, \mathbf{x}_d$  be  $d$  selected largest eigenvectors of  $\mathbf{G}$ . Each image  $\mathbf{A}$  is projected onto these  $d$  dimensional subspace according to Eq. (2.10). The projected image  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_d]$  is then an  $m$  by  $d$  matrix given by:

$$\mathbf{Y} = \mathbf{A}\mathbf{X}, \tag{2.19}$$

where  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_d]$  is a  $n$  by  $d$  projection matrix.



### 2.2.2 Column-Based 2DPCA

The original 2DPCA can be called the row-based 2DPCA. The alternative way of 2DPCA can be using the column instead of row, column-based 2DPCA [41].

Let  $\mathbf{A}_k = [(\mathbf{A}_k^{(1)})^T (\mathbf{A}_k^{(2)})^T \dots (\mathbf{A}_k^{(m)})^T]^T$  and  $\bar{\mathbf{A}} = [(\bar{\mathbf{A}}^{(1)})^T (\bar{\mathbf{A}}^{(2)})^T \dots (\bar{\mathbf{A}}^{(m)})^T]^T$ , where  $\mathbf{A}_k^{(i)}$  and  $\bar{\mathbf{A}}^{(i)}$  denote the  $i^{\text{th}}$  row vectors of  $\mathbf{A}_k$  and  $\bar{\mathbf{A}}$ , respectively. Then Eq. (2.16) can be rewritten as

$$\mathbf{G} = \frac{1}{M} \sum_{k=1}^M \sum_{i=1}^m (\mathbf{A}_k^{(i)} - \bar{\mathbf{A}}^{(i)})^T (\mathbf{A}_k^{(i)} - \bar{\mathbf{A}}^{(i)}). \quad (2.20)$$

This reveals that the image covariance matrix can be obtained from the outer product of row vectors of images, assuming the training images have zero mean, i.e.  $\bar{\mathbf{A}} = \mathbf{0}$ . For that reason, we claim that original 2DPCA is working in the row direction of images.

Illuminated by Eq. (2.20), a natural extension is to use the outer product between column vectors of images to construct the image covariance matrix. Let  $\mathbf{A}_k = [(\mathbf{A}_k^{(1)}) (\mathbf{A}_k^{(2)}) \dots (\mathbf{A}_k^{(n)})]$  and  $\bar{\mathbf{A}} = [(\bar{\mathbf{A}}^{(1)}) (\bar{\mathbf{A}}^{(2)}) \dots (\bar{\mathbf{A}}^{(n)})]$ , where  $\mathbf{A}_k^{(j)}$  and  $\bar{\mathbf{A}}^{(j)}$  denote the  $j^{\text{th}}$  column vectors of  $\mathbf{A}_k$  and  $\bar{\mathbf{A}}$ , respectively. Then the column-based definition for image covariance matrix  $\mathbf{H}$  can be defined as

$$\mathbf{H} = \frac{1}{M} \sum_{k=1}^M \sum_{j=1}^n (\mathbf{A}_k^{(j)} - \bar{\mathbf{A}}^{(j)}) (\mathbf{A}_k^{(j)} - \bar{\mathbf{A}}^{(j)})^T. \quad (2.21)$$

Eq. (2.21) can be derived at a similar way as in 2DPCA. Let  $\mathbf{Z} \in \mathbb{R}^m \times \mathbb{R}^q$  be a matrix with orthonormal columns. Projecting the random matrix  $\mathbf{A}$  onto  $\mathbf{Z}$  yields a  $q$  by  $n$  matrix  $\tilde{\mathbf{Y}} = \mathbf{Z}^T \mathbf{A}$ . Similar as in Eq. (2.13), the following criterion is adopted to find the optimal projection matrix  $\mathbf{Z}$ :

$$\begin{aligned} \text{tr}(\mathbf{S}_Z) &= \text{tr}\{E[(\tilde{\mathbf{Y}} - E\tilde{\mathbf{Y}})(\tilde{\mathbf{Y}} - E\tilde{\mathbf{Y}})^T]\} \\ &= \text{tr}\{E[(\mathbf{Z}^T \mathbf{A} - E[\mathbf{Z}^T \mathbf{A}])(\mathbf{Z}^T \mathbf{A} - E[\mathbf{Z}^T \mathbf{A}])^T]\} \\ &= \text{tr}\{\mathbf{Z}^T E[(\mathbf{A} - E\mathbf{A})(\mathbf{A} - E\mathbf{A})^T] \mathbf{Z}\} \\ &= \text{tr}\{\mathbf{Z}^T \mathbf{H} \mathbf{Z}\}. \end{aligned} \quad (2.22)$$

From Eq. (2.22), the column-based definition of image covariance matrix  $\mathbf{H}$  is:

$$\begin{aligned} \mathbf{H} &= E[(\mathbf{A} - E\mathbf{A})(\mathbf{A} - E\mathbf{A})^T] \\ &= \frac{1}{M} \sum_{k=1}^M (\mathbf{A}_k - \bar{\mathbf{A}})(\mathbf{A}_k - \bar{\mathbf{A}})^T \\ &= \frac{1}{M} \sum_{k=1}^M \sum_{j=1}^n (\mathbf{A}_k^{(j)} - \bar{\mathbf{A}}^{(j)}) (\mathbf{A}_k^{(j)} - \bar{\mathbf{A}}^{(j)})^T. \end{aligned} \quad (2.23)$$

Similarly, the optimal projection matrix can be obtained by computing the eigenvectors of Eq. (2.23) corresponding to the  $q$  largest eigenvalues. The value of  $q$  can also be controlled by setting a threshold as in Eq. (2.18). Because the eigenvectors of Eq. (2.23) only reflect the information between columns of images, than the column-based 2DPCA is working in the column direction of images.

### 2.2.3 Bilateral-Projection-Based 2DPCA (B2DPCA)

There are three major difference techniques in the framework of the bilateral projection scheme.

Firstly, the non-iterative bilateral projection scheme was applied to 2DPCA via left and right multiplying projection matrices [41, 54, 55] as follows

$$\mathbf{B} = \mathbf{L}^T \mathbf{A} \mathbf{R}, \quad (2.24)$$

where  $\mathbf{B}$  is a feature matrix which extracted from image  $\mathbf{A}$  and  $\mathbf{L}$  is a left multiplying projection matrix. Similar to the right multiplying projection matrix  $\mathbf{R}$  in Section 2.2.1, matrix  $\mathbf{L}$  is a  $m$  by  $l$  projection matrix that obtained by choosing the eigenvectors of image covariance matrix  $\mathbf{H}$  corresponding to the  $l$  largest eigenvalues. The matrix  $\mathbf{H}$ , which corresponds to the column direction of images, can be evaluated by

$$\mathbf{H} = \frac{1}{M} \sum_{k=1}^M (\mathbf{A}_k - \bar{\mathbf{A}})(\mathbf{A}_k - \bar{\mathbf{A}})^T. \quad (2.25)$$

Therefore, the dimension of feature matrix is decreasing from  $m \times r$  to  $l \times r$ . In this way, the computation time also be reducing. Moreover, the recognition accuracy of B2DPCA is better than 2DPCA.

Secondly, the bilateral projection scheme of 2DPCA with the iterative algorithm was proposed in [40, 56]. Let  $\mathbf{L} \in \mathbb{R}^m \times \mathbb{R}^l$  and  $\mathbf{R} \in \mathbb{R}^n \times \mathbb{R}^r$  be the left and right multiplying projection matrix respectively. For an  $m \times n$  image  $\mathbf{A}_k$  and  $l \times r$  projected image  $\mathbf{B}_k$ , the bilateral projection is formulated as follows:

$$\mathbf{B}_k = \mathbf{L}^T \mathbf{A}_k \mathbf{R} \quad (2.26)$$

where  $\mathbf{B}_k$  is the extracted feature matrix for image  $\mathbf{A}_k$ .

The optimal projection matrices,  $\mathbf{L}$  and  $\mathbf{R}$  in Eq. (2.26) can be computed by solving the following minimization criterion that the reconstructed image,  $\mathbf{L} \mathbf{B}_k \mathbf{R}^T$ , gives the best approximation of  $\mathbf{A}_k$ :

$$J(\mathbf{L}, \mathbf{R}) = \min \sum_{k=1}^M \|\mathbf{A}_k - \mathbf{L} \mathbf{B}_k \mathbf{R}^T\|_F^2, \quad (2.27)$$

where  $M$  is the number of data samples and  $\|\bullet\|_F$  is the Frobenius norm of a matrix.

The detailed iterative scheme designed to compute the optimal projection matrices,  $\mathbf{L}$  and  $\mathbf{R}$ , is listed in Table 2.1. The obtained solutions are locally optimal because

Table 2.1 The Bilateral Projection Scheme of 2DPCA with Iterative Algorithm.

---

$S_1$ :	Initialize $\mathbf{L}$ , $\mathbf{L} = \mathbf{L}_0$ and $i = 0$
$S_2$ :	While not convergent
$S_3$ :	Compute $\mathbf{G} = \frac{1}{M} \sum_{k=1}^M (\mathbf{A}_k - \bar{\mathbf{A}})^T \mathbf{L}_{i-1} \mathbf{L}_{i-1}^T (\mathbf{A}_k - \bar{\mathbf{A}})$
$S_4$ :	Compute the $r$ eigenvectors $\{\mathbf{e}_j^{\mathbf{R}}\}_{j=1}^r$ of $\mathbf{G}$ corresponding to the largest $r$ eigenvalues
$S_5$ :	$\mathbf{R}_i = [\mathbf{e}_1^{\mathbf{R}}, \dots, \mathbf{e}_r^{\mathbf{R}}]$
$S_6$ :	Compute $\mathbf{H} = \frac{1}{M} \sum_{k=1}^M (\mathbf{A}_k - \bar{\mathbf{A}}) \mathbf{R}_i \mathbf{R}_i^T (\mathbf{A}_k - \bar{\mathbf{A}})^T$
$S_7$ :	Compute the $l$ eigenvectors $\{\mathbf{e}_j^{\mathbf{L}}\}_{j=1}^l$ of $\mathbf{H}$ corresponding to the largest $l$ eigenvalues
$S_8$ :	$\mathbf{L}_i = [\mathbf{e}_1^{\mathbf{L}}, \dots, \mathbf{e}_l^{\mathbf{L}}]$
$S_9$ :	$i = i + 1$
$S_{10}$ :	End While
$S_{11}$ :	$\mathbf{L} = \mathbf{L}_{i-1}$
$S_{12}$ :	$\mathbf{R} = \mathbf{R}_{i-1}$
$S_{13}$ :	Feature extraction: $\mathbf{B}_k = \mathbf{L}^T \mathbf{A}_k \mathbf{R}$

---

the solutions are dependent on the initialized  $\mathbf{L}_0$ . In [40], the initialized  $\mathbf{L}_0$  sets to the  $m \times m$  identity matrix  $\mathbf{I}_m$ , while this value is set to  $\begin{bmatrix} \mathbf{I}_r \\ 0 \end{bmatrix}$  in [56], where  $\mathbf{I}_r$  is the  $l \times l$  identity matrix.

Finally, The criterion in Eq. (2.27) is biquadratic and has no closed-form solution. Therefore, an iterative procedure to obtain the local optimal solution was proposed in [57]. For  $\mathbf{R} \in \mathbb{R}^m \times \mathbb{R}^r$ , the criterion in Eq. (2.27) can be rewritten as

$$J(\mathbf{R}) = \min \sum_{k=1}^M \|\mathbf{A}_k - \mathbf{A}_k^{\mathbf{L}} \mathbf{R} \mathbf{R}^T\|_F^2, \quad (2.28)$$

where  $\mathbf{A}_k^{\mathbf{L}} = \mathbf{L} \mathbf{L}^T \mathbf{A}_k$ . The solution of Eq. (2.28) is the eigenvectors of the eigenvalue decomposition of image covariance matrix:

$$\mathbf{G} = \frac{1}{M} \sum_{k=1}^M (\mathbf{A}_k^{\mathbf{L}} - \bar{\mathbf{A}}^{\mathbf{L}})^T (\mathbf{A}_k^{\mathbf{L}} - \bar{\mathbf{A}}^{\mathbf{L}}). \quad (2.29)$$

Similarly, for  $\mathbf{L} \in \mathbb{R}^n \times \mathbb{R}^l$ , the criterion in Eq. (2.27) is changed to

$$J(\mathbf{L}) = \min \sum_{k=1}^M \|\mathbf{A}_k - \mathbf{L} \mathbf{L}^T \mathbf{A}_k^{\mathbf{R}}\|_F^2, \quad (2.30)$$

Table 2.2 Coupled Subspaces Analysis Algorithm.

---

$S_1$ :	Initialize $\mathbf{L}$ , $\mathbf{L} = \mathbf{I}_m$
$S_2$ :	For $i = 1, 2, \dots, T_{\max}$
$S_3$ :	Compute $\mathbf{A}_k^{\mathbf{L}} = \mathbf{L}_{i-1} \mathbf{L}_{i-1}^T \mathbf{A}_k$
$S_4$ :	Compute $\mathbf{G} = \frac{1}{M} \sum_{k=1}^M (\mathbf{A}_k^{\mathbf{L}} - \bar{\mathbf{A}}^{\mathbf{L}})^T (\mathbf{A}_k^{\mathbf{L}} - \bar{\mathbf{A}}^{\mathbf{L}})$
$S_5$ :	Compute the $r$ eigenvectors $\{\mathbf{e}_j^{\mathbf{R}}\}_{j=1}^r$ of $\mathbf{G}$ corresponding to the largest $r$ eigenvalues
$S_6$ :	$\mathbf{R}_i = [\mathbf{e}_1^{\mathbf{R}}, \dots, \mathbf{e}_r^{\mathbf{R}}]$
$S_7$ :	Compute $\mathbf{A}_k^{\mathbf{R}} = \mathbf{A}_k \mathbf{R}_i \mathbf{R}_i^T$
$S_8$ :	Compute $\mathbf{H} = \frac{1}{M} \sum_{k=1}^M (\mathbf{A}_k^{\mathbf{L}} - \bar{\mathbf{A}}^{\mathbf{L}}) (\mathbf{A}_k^{\mathbf{L}} - \bar{\mathbf{A}}^{\mathbf{L}})^T$
$S_9$ :	Compute the $l$ eigenvectors $\{\mathbf{e}_j^{\mathbf{L}}\}_{j=1}^l$ of $\mathbf{H}$ corresponding to the largest $l$ eigenvalues
$S_{10}$ :	$\mathbf{L}_i = [\mathbf{e}_1^{\mathbf{L}}, \dots, \mathbf{e}_l^{\mathbf{L}}]$
$S_{11}$ :	If $t > 2$ and $\ \mathbf{L}_i - \mathbf{L}_{i-1}\ _F < m\varepsilon$ and $\ \mathbf{R}_i - \mathbf{R}_{i-1}\ _F < n\varepsilon$
$S_{12}$ :	Then Go to $S_3$
$S_{13}$ :	Else Go to $S_{15}$
$S_{14}$ :	End For
$S_{15}$ :	$\mathbf{L} = \mathbf{L}_i$
$S_{16}$ :	$\mathbf{R} = \mathbf{R}_i$
$S_{17}$ :	Feature extraction: $\mathbf{B}_k = \mathbf{L}^T \mathbf{A}_k \mathbf{R}$

---

where  $\mathbf{A}_k^{\mathbf{R}} = \mathbf{A}_k \mathbf{R} \mathbf{R}^T$ . Again, the solution of Eq. (2.30) is the eigenvectors of the eigenvalue decomposition of image covariance matrix:

$$\mathbf{H} = \frac{1}{M} \sum_{k=1}^M (\mathbf{A}_k^{\mathbf{L}} - \bar{\mathbf{A}}^{\mathbf{L}}) (\mathbf{A}_k^{\mathbf{L}} - \bar{\mathbf{A}}^{\mathbf{L}})^T. \quad (2.31)$$

By iteratively optimizing the objective function with respect to  $\mathbf{L}$  and  $\mathbf{R}$ , respectively, we can obtain a local optimum of the solution. The whole procedure, namely Coupled Subspace Analysis (CSA) [57], is shown in Fig. 2.5.

#### 2.2.4 Diagonal-Based 2DPCA (DiaPCA)

The motivation for developing the DiaPCA method originates from an essential observation on the recently proposed 2DPCA [39]. That is, 2DPCA can be seen as the row-based PCA, which has been pointed out in [58,59]. Therefore, the projective vectors of 2DPCA only reflect variations between rows of images, while the omitted variations between columns of images are usually also useful for recognition, which implies some



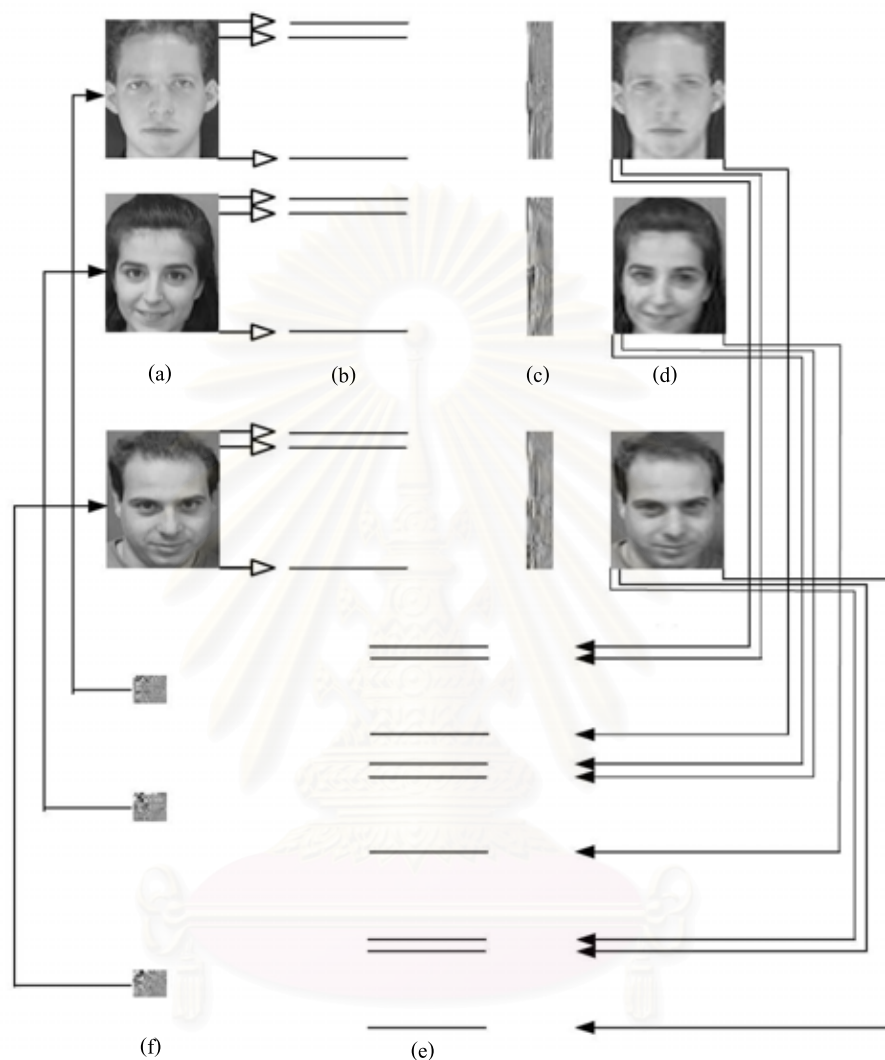


Figure 2.5: The flowchart of Coupled Subspaces Analysis: (a) The original image. (b) All the rows of the images are treated as the objects. (c) The low-dimensional representation after the projection with the right multiplying projection. (d) The reconstructed images with the right multiplying projection. (e) All the columns of the images are treated as the objects. (f) The low dimensional representation after the projection with left and right multiplying projection.

structure information (e.g. regions of a face like eyes, nose, etc.) cannot be uncovered by it. In that case, 2DPCA can hardly obtain improved accuracy.

In [52], a novel method called diagonal principal component analysis (DiaPCA) is proposed. In contrast to 2DPCA, DiaPCA seeks the optimal projective vectors from diagonal face images and therefore the correlations between variations of rows and those of columns of images can be kept. Therefore, this problem can solve by transforming the original face images into corresponding diagonal face images, as shown in Fig. 2.6 and Fig. 2.7. Because the rows (columns) in the transformed diagonal face images simultaneously integrate the information of rows and columns in original images, it can reflect both information between rows and those between columns. Through the entanglement of row and column information, it is expected that DiaPCA may find some useful block or structure information for recognition in original images. The sample diagonal face images on Yale database are displayed in Fig. 2.8.

Experimental results on a subset of FERET database [52] show that DiaPCA is more accurate than both PCA and 2DPCA. Furthermore, it is shown that the accuracy can be further improved by combining DiaPCA and 2DPCA together.

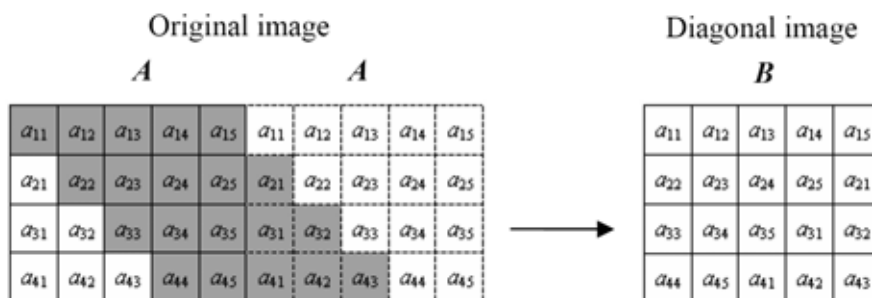


Figure 2.6: Illustration of the ways for deriving the diagonal face images: If the number of columns is more than the number of rows

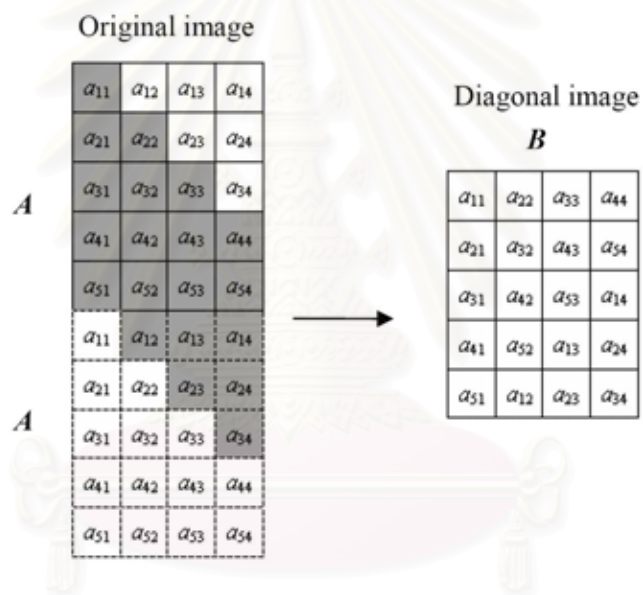


Figure 2.7: Illustration of the ways for deriving the diagonal face images: If the number of columns is less than the number of rows



Figure 2.8 The sample diagonal face images on Yale database.

## 2.3 PCA Versus 2DPCA

In this section, the relationship of PCA and 2DPCA and the distance measurements are reviewed.

### 2.3.1 The Relationship between PCA and 2DPCA

The sketch for the reason that why 2DPCA is better than PCA should lie on the answer of the following question:

*What is happen when the inputs of PCA are the rows of each image instead of entire images?*

Let  $\{\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_n\}$  be the set of rows of an image  $\mathbf{A}$  then the average of this training set will be  $\Psi = \frac{1}{n} \sum_{i=1}^n \Gamma_i$ . The  $i$ -th row differ from the average by the vector  $\Phi_i = \Gamma_i - \Psi$ . Therefore, the covariance matrix  $\mathbf{C}$  of PCA is given by

$$\mathbf{C} = \frac{1}{n} \sum_{i=1}^n \Phi_i \Phi_i^T = \frac{1}{n} \mathbf{Y} \mathbf{Y}^T, \quad (2.32)$$

where the matrix  $\mathbf{Y} = [\Phi_1 \ \Phi_2 \ \dots \ \Phi_n]$ . Following Eigenface's algorithm [3], the optimal projection vectors can be determined as the eigenvectors of matrix  $\mathbf{Y}^T \mathbf{Y}$ . If zero mean,  $\Psi = 0$ , the image covariance matrix  $\mathbf{G}$  in Eq.(2.14) can be rewritten as

$$\begin{aligned} \mathbf{G} &= E[(\mathbf{Y} - E\mathbf{Y})^T (\mathbf{Y} - E\mathbf{Y})] \\ &= E[\mathbf{Y}^T \mathbf{Y}] - E\mathbf{Y}^T E\mathbf{Y} \\ &= E[\mathbf{Y}^T \mathbf{Y}] - \beta, \end{aligned} \quad (2.33)$$

where  $\beta = E\mathbf{Y}^T E\mathbf{Y}$ .

From this point, 2DPCA can be explained in a novel perspective as a collection of biased PCAs. Indeed, the number of training samples of conventional PCA is only  $M$  while the number of training samples of 2DPCA is  $M \times m$ . Since the dimension of  $\mathbf{G}$  is  $n \times n$  with  $n < M \times m$ , thus 2DPCA can provide the full rank image covariance matrix. This is the reason of the improvement of 2DPCA over the original PCA [58, 59].

### 2.3.2 Evaluation of Over-fitting Problem in PCA and 2DPCA

From Section 2.1.5.3, the over-fitting problem is the one serious problem in traditional PCA. Therefore, it is interesting to investigate the over-fitting problem in 2DPCA.

The normalized Mean-Square Error (MSE) can be used to evaluate the over-fitting problem [54]. One statistical characteristic of PCA is that the MSE between random vector and its subspace projection is minimal. Thus the difference of MSE on the training



set and the validation set can be used to investigate the over-fitting problem. The MSE on the training set  $MSE^{train}$  and validation set  $MSE^{val}$  are defined as

$$MSE^{train} = \frac{\sum_{i=1}^{N_{train}} \left\| \mathbf{A}_i^{train} - \tilde{\mathbf{A}}_i^{train} \right\|^2}{\sum_{i=1}^{N_{train}} \left\| \mathbf{A}_i^{train} - \bar{\mathbf{A}}_i^{train} \right\|^2} \quad (2.34)$$

and

$$MSE^{val} = \frac{\sum_{i=1}^{N_{val}} \left\| \mathbf{A}_i^{val} - \tilde{\mathbf{A}}_i^{val} \right\|^2}{\sum_{i=1}^{N_{val}} \left\| \mathbf{A}_i^{val} - \bar{\mathbf{A}}_i^{val} \right\|^2}, \quad (2.35)$$

respectively. Where  $\mathbf{A}_i^{train}$  is the  $i^{th}$  training samples and  $\tilde{\mathbf{A}}_i^{train}$  is the reconstructed version of  $\mathbf{A}_i^{train}$ .  $\bar{\mathbf{A}}_i^{train}$  is the mean of all training samples.  $\mathbf{A}_i^{val}$  is the  $i^{th}$  validation samples and  $\tilde{\mathbf{A}}_i^{val}$  is the reconstructed version of  $\mathbf{A}_i^{val}$ .  $\bar{\mathbf{A}}_i^{val}$  is the mean of all validation samples.  $N_{train}$  and  $N_{val}$  are the number of samples in training and validation set, respectively.

The experiments on ORL face database was used to verify this perspective. The first 5 images per individual for training and calculating  $MSE^{train}$  and  $MSE^{val}$ .

The experimental result for PCA was shown in Fig. 2.9. When the value of the dimension of feature is small, the difference between  $MSE^{train}$  and  $MSE^{val}$  is small. But the difference is becoming great rapidly with the increase of the feature dimension. Thus classical PCA is over-fitted to the training set. Then MSE was used again to evaluate 2DPCA's capability in solving the over-fitting problem, as shown in Fig. 2.10. The difference of  $MSE^{train}$  and  $MSE^{val}$  is very small. Since the dimension of image covariance matrix is far smaller than the covariance matrix in PCA. Therefore, 2DPCA is not suffer from the over-fitting problem.

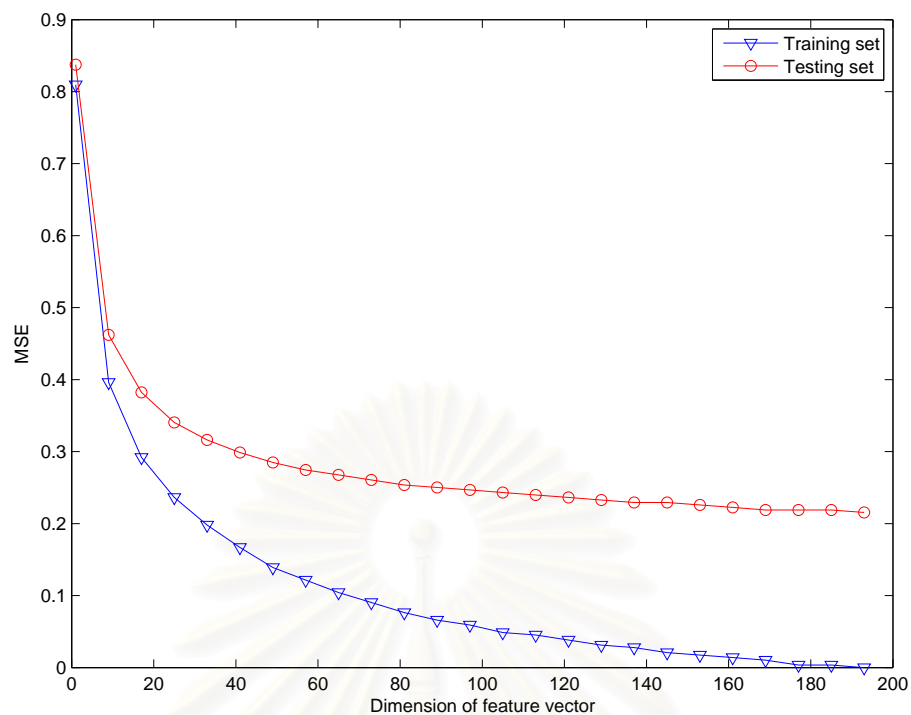


Figure 2.9: The PCA's MSE on the training set and the testing set as the function of feature dimension.

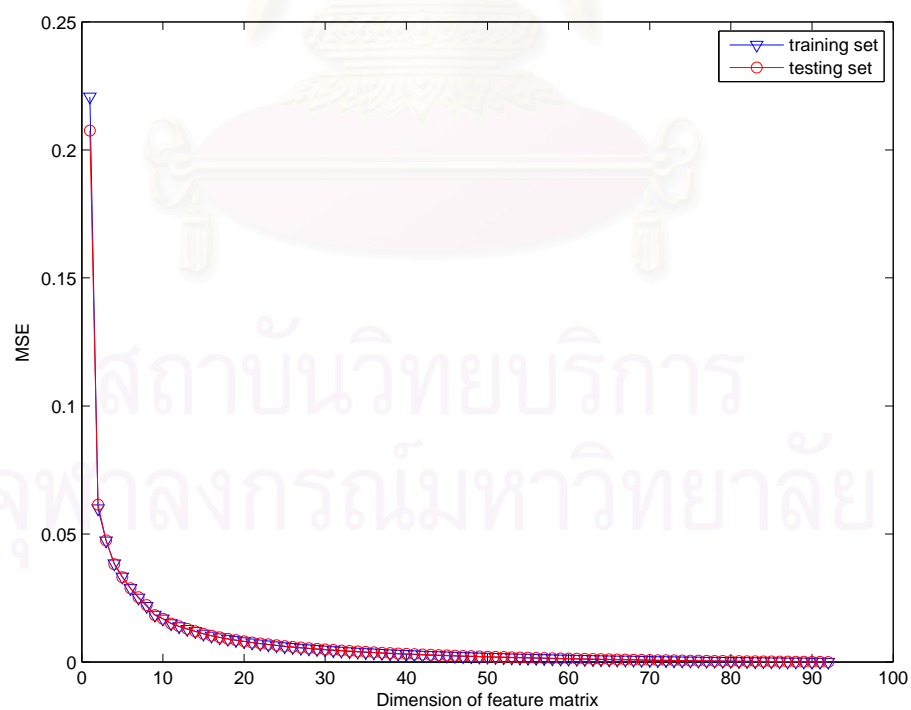


Figure 2.10: The 2DPCA's MSE on the training set and the testing set as the function of feature dimension.

### 2.3.3 The Distance Measurements

Normally, a nearest neighbor classifier is used for classification. The distance measurement is necessary for this classifier. In 1D subspace analysis, such as PCA, the feature is produced in vector form. While in the feature in 2D subspace analysis, such as 2DPCA, is presented in matrix form. For this reason the novel distance measurements were proposed to measure the distance between feature matrices.

#### 2.3.3.1 1D Feature Distance Measurements

In Euclidean space  $\mathbb{R}^n$ , the distance between two features is usually given by the Euclidean distance (L2-norm distance). Other distances, based on other norms, are sometimes used instead. For vector  $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$  and vector  $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_n]^T$ , the various distances [60] are defined as:

- *L1-norm distance*

$$d_{L1}(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^n |x_k - y_k| \quad (2.36)$$

- *L2-norm distance*

$$d_{L2}(\mathbf{x}, \mathbf{y}) = \left( \sum_{k=1}^n |x_k - y_k|^2 \right)^{1/2} \quad (2.37)$$

- *Lp-norm distance*

$$d_{Lp}(\mathbf{x}, \mathbf{y}) = \left( \sum_{k=1}^n |x_k - y_k|^p \right)^{1/p} \quad (2.38)$$

- *L $\infty$ -norm distance*

$$\begin{aligned} d_{L\infty}(\mathbf{x}, \mathbf{y}) &= \lim_{p \rightarrow \infty} \left( \sum_{k=1}^n |x_k - y_k|^p \right)^{1/p} \\ &= \max(|x_1 - y_1|, |x_2 - y_2|, \dots, |x_n - y_n|) \end{aligned} \quad (2.39)$$

- *Cosine distance*

$$d_{\cos}(\mathbf{x}, \mathbf{y}) = 1 - \cos(\mathbf{x}, \mathbf{y}) = 1 - \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \quad (2.40)$$

#### 2.3.3.2 2D Feature Distance Measurements

In the 2D subspace  $\mathbb{R}^m \times \mathbb{R}^n$ , The distance between two arbitrary feature matrices can be calculated by many methods such as Yang's distance [39], an assembled matrix distance [61], volume measure [62] and bilateral-projection-based distance [41]. In this dissertation, only Yang's distance was used as the distance measurement for 2D subspace analysis in all experiments.

- *Yang's distance*: The original distance between two arbitrary feature matrices, defined as  $\mathbf{B}_i = [\mathbf{b}_{i_1}, \mathbf{b}_{i_2}, \dots, \mathbf{b}_{i_q}]$  and  $\mathbf{B}_j = [\mathbf{b}_{j_1}, \mathbf{b}_{j_2}, \dots, \mathbf{b}_{j_q}]$  is defined by

$$d(\mathbf{B}_i, \mathbf{B}_j) = \sum_{k=1}^q \|\mathbf{b}_{i_k} - \mathbf{b}_{j_k}\|, \quad (2.41)$$

where  $\|\mathbf{b}_{i_k} - \mathbf{b}_{j_k}\|$  denotes the Euclidean distance between the two principal component vectors  $\mathbf{b}_{i_k}$  and  $\mathbf{b}_{j_k}$ . Suppose that the feature matrices of  $M$  training samples are  $\mathbf{B}_l$ ,  $l = 1, 2, 3, \dots, M$ , and that each of these samples is assigned a class  $\omega_k$ . Given a feature matrix of test sample  $\mathbf{B}_{test}$ , if

$$d(\mathbf{B}_{test}, \mathbf{B}_l) = \min_j d(\mathbf{B}_{test}, \mathbf{B}_j) \quad (2.42)$$

then the test sample  $\mathbf{B}_{test}$  is belong to class  $\omega_k$ .

- *Assembled Matrix Distance*: An assembled matrix distance (AMD) metric was proposed to calculate the distance between two feature matrices. Given two feature matrices  $\mathbf{B}_i$  and  $\mathbf{B}_j$ , the assembled matrix distance was defined as follows:

$$d_{AMD}(\mathbf{B}_i, \mathbf{B}_j) = \left( \sum_{v=1}^d \left( \sum_{u=1}^m (b_{i_{uv}} - b_{j_{uv}})^2 \right)^{p/2} \right)^{1/p}, \quad (p > 0), \quad (2.43)$$

where  $b_{i_{uv}}$  is the  $u^{th}$  row and the  $v^{th}$  column element of matrix  $\mathbf{B}_i$  and  $b_{j_{uv}}$  is the  $u^{th}$  row and the  $v^{th}$  column element of matrix  $\mathbf{B}_j$ .

- *Volume Measure*: The volume of an  $m \times n$  matrix of rank  $p$  was introduced in [63] as

$$Vol(\mathbf{A}) = \sqrt{\sum_{(i,j) \in N} \det^2 \mathbf{A}_{ij}}, \quad (2.44)$$

where  $\mathbf{A}_{ij}$  denotes the submatrix of  $\mathbf{A}$  with rows  $i$  and columns  $j$ ,  $N$  is the index set of  $p \times p$  nonsingular submatrix of  $\mathbf{A}$ , and if  $p = 0$ , then define  $Vol(\mathbf{A}) = 0$ . It is not easy to compute this volume directly by Eq. (2.44). Fortunately, from the Cauchy–Binet formula, we can get a very simple calculating formula for  $Vol(\mathbf{A})$  when  $\mathbf{A}$  is a matrix of full column rank:

$$Vol(\mathbf{A}) = \sqrt{\det \mathbf{A}^T \mathbf{A}}. \quad (2.45)$$

Then we can deduce the calculating formula to measure the similarities of two samples by Eq. (2.45). Given the feature matrices  $\mathbf{B}_i$  and  $\mathbf{B}_j$  after the transformation by 2DPCA, the distance measurement based on the matrix volume can be depicted as

$$d_{VM}(\mathbf{B}_i, \mathbf{B}_j) = \sqrt{\det((\mathbf{B}_i - \mathbf{B}_j)^T (\mathbf{B}_i - \mathbf{B}_j))}. \quad (2.46)$$

However, there is a very important point, i.e. to compute the matrix volume by Eq. (2.46), we must make sure that the matrix  $\mathbf{B}_i - \mathbf{B}_j$  is of full column rank. In fact, We can make it satisfy this condition conveniently by the following algorithm [62]. Which can be done by deleting the columns of  $\mathbf{B}_i - \mathbf{B}_j$ , as algorithm in Table 2.3.

Table 2.3 Deleting The Columns Algorithm for Volume Measure

---

$S_1$ : Initiate $k = 2$ .
$S_2$ : If the $k^{th}$ column is the last column of the matrix $\mathbf{B}_i - \mathbf{B}_j$ , then stop; otherwise decide the submatrix consisting of the first $k$ columns of the matrix $\mathbf{B}_i - \mathbf{B}_j$ is of full column rank.
$S_3$ : If it is of full column rank, then $k = k + 1$ , goto $S_2$ . Otherwise goto $S_4$ .
$S_4$ : Delete the $k^{th}$ column from the matrix $\mathbf{B}_i - \mathbf{B}_j$ , goto $S_2$ .

---

- **Bilateral-Projection-Based Distance:** For bilateral-projection-based scheme, the alternative distance measurement should be used. Since the feature matrices are extracted from both row and column directions of images, thus the distance will be sum of distances of each elements  $\mathbf{b}_{(m,n)}$  in feature matrix  $\mathbf{B}$ . The distance between two bilateral-projection-based scheme feature matrices,  $\mathbf{B}_i$  and  $\mathbf{B}_j$ , is defined by

$$d(\mathbf{B}_i, \mathbf{B}_j) = \sqrt{\sum_{m=1}^q \sum_{n=1}^d (\mathbf{b}_{(m,n)_i} - \mathbf{b}_{(m,n)_j})^2}. \quad (2.47)$$



# CHAPTER III

## THE PROPOSED FRAMEWORKS

In this chapter, the four frameworks were proposed for improving the performance of 2DPCA. We first proposed the two frameworks for image recognition that is designed to combine the unsupervised and supervised subspace analysis techniques. In Section 3.1, the first framework was called *Two-Dimensional Linear Discriminant Analysis of Principal Component Vectors*. This technique takes the advantage of the Fisher's criterion for ameliorating the discriminant power of classification. The second framework, *Class-Specific Subspace-Based Two-Dimensional Principal Component Analysis*, was introduced in Section 3.2. The class-specific technique is applied here to fulfill the information of class labels. We later proposed the generalized version of 2DPCA, called *Image Cross-Covariance Analysis (ICCA)*, in Section 3.3. Finally, the novel ensemble method for 2DPCA, named *Two-Dimensional Random Subspace Analysis (2DRSA)*, to increase the performance and robustness via the Random Subspace Method (RSM) was presented in Section 3.4.

### 3.1 Two-Dimensional Linear Discriminant Analysis of Principal Component Vectors

In face recognition, a great number of successful face recognition systems have been developed and reported in the literature [2, 3, 16–18]. Among these works, the linear subspace techniques, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are the most popular ones. The PCA's criterion chooses the subspace in the function of data distribution while LDA chooses the subspace which yields maximal inter-class distance, and at the same time, keeping the intra-class distance small. In general, LDA extracts features which are better suitable for classification task. Both techniques intend to project the vector representing face image onto lower dimensional subspace, in which each 2D face image matrix must be previously transformed into vector and then a collection of the transformed face vectors are concatenated into a matrix. This is the cause of three serious problems in particular approaches. First of all, the covariance matrix, which collects the feature vectors with high dimension, will leads to the curse of dimensionality, i.e. large memory required and high computation cost. Secondly, the spatial structure information could be lost when the column-stacking vectorization and image resizing are applied. Finally, normally in face recognition task, the available number of training samples is small compared to the feature dimension.

Therefore the covariance matrix estimated by these features will be singular and then cannot be inverted. This is called singularity problem or Small Sample Size (SSS) problem [31]. Especially, as a supervised technique, LDA has a tendency to over-fitting in SSS problems. And in the same situation, PCA faces with the over-fitting problem too, as discussed in Section 2.1.5.3.

Various solutions have been proposed for solving the SSS problem [16–18,32–34] within LDA framework. Among these LDA extensions, Fisherface [16] and the discriminant analysis of principal components framework [17, 18] demonstrates a significant improvement when applying Linear Discriminant Analysis (LDA) over principal components from the PCA-based subspace. Since both PCA and LDA can overcome the drawbacks of each other. PCA is constructed around the criteria of preserving the data distribution. Hence, it is suited for face representation and reconstruction from the projected face feature. However, in the classification tasks, PCA only normalize the input data according to their variance. This is not efficient since the between classes relationship is neglected. In general, the discriminatory power depends on both within and between classes relationship. LDA considers these relationships via the analysis of within and between-class scatter matrices. Taking this information into account, LDA allows further improvement. Especially, when there are prominent variation in lighting condition and expression. However, LDA has two drawbacks when directly applied to the original input space [18]. First of all, some non-face information such as image background are regarded by LDA as the discriminant information. This causes misclassification when the face of the same subject is presented on different background. Secondly, when SSS problem has occurred, the within-class scatter matrix is singular and LDA projection cannot be computed. Projecting the high dimensional input space into low dimensional subspace via PCA can solve these LDA problems. Nevertheless, all of above techniques, the spatial structure information still be not employed.

In Two-Dimensional Principal Component Analysis (2DPCA), the image covariance matrix is computed directly on image matrices so the spatial structure information can be employed. This yields a covariance matrix whose dimension just equals to the width of the face image. This is far smaller than the size of covariance matrix in PCA. Therefore, the image covariance matrix can be better estimated and full rank. Evidently, the experimental results in [39] have shown the improvement of 2DPCA over PCA on several face databases.

Subsequently, Two-Dimensional Linear Discriminant Analysis (2DLDA) was proposed in [43]. For overcoming the SSS problem in classical LDA by working with images in matrix representation, like in 2DPCA. In particular, bilateral projection scheme was applied there via left and right multiplying projection matrices. In this way, the eigenvalue problem was solved two times per iteration. One corresponds to the column direction and another one corresponds to the row direction of image, respectively

Because of 2DPCA is more suitable for face representation than face recognition,

like PCA. For better performance in recognition task, LDA is still necessary. Unfortunately, the linear transformation of 2DPCA reduces the input image to a vector with the same dimension as the number of rows or the height of the input image. Thus, the SSS problem may still occurred when LDA is performed after 2DPCA directly. To overcome this problem, we proposed a simplified version of the bilateral 2DLDA by using only unilateral projection scheme, based on the 2DPCA concept. Applying our proposed 2DLDA to 2DPCA not only can solve the SSS problem and the curse of dimensionality dilemma but also allows us to work directly on the image matrix in all projections. Hence, spatial structure information is maintained and the size of all scatter matrices cannot be greater than the width of face image. Furthermore, computing with this dimension, the face image do not need to be resized, since all information still be preserved.

In this section, we firstly show how to put the idea of 2DPCA under the LDA's criterion, namely 2DLDA. According to the previous discussion on 2DLDA, our proposed 2DLDA method is different from [43] on the fact that our method aims to find the optimal discriminated transformation for projecting the principal component vectors, which obtained from 2DPCA step, Our 2DLDA proposed subspace nested inside the subspace proposed by 2DPCA. Moreover, our 2DLDA formulation has a closed form solution, thus can be solved with non-iterative procedure. After that we apply it to the LDA of principal components [17,18] and Fisherface [16] frameworks, i.e. *Two-Dimensional Linear Discriminant Analysis of Principal Component Vectors* and *2D Fisherface*, respectively. The accuracy of this proposed framework is demonstrated on real world image databases. The experimental results can promise the performance of our proposed framework.

We would like to note that this framework was the detailed and expanded version of our published materials, which is based on [58,59].

### 3.1.1 Two-Dimensional Linear Discriminant Analysis (2DLDA)

Let  $\mathbf{z}$  be a  $q$  dimensional vector. A matrix  $\mathbf{A}$  is projected onto this vector via the similar transformation as Eq. (2.10):

$$\mathbf{v}=\mathbf{Az}. \quad (3.1)$$

This projection yields an  $m$  dimensional feature vector.

In 2DLDA, we search for the projection axis  $\mathbf{z}$  that maximizing the Fisher's discriminant criterion [16,31]:

$$J(\mathbf{z}) = \frac{tr(\mathbf{S}_b)}{tr(\mathbf{S}_w)}, \quad (3.2)$$

where  $\mathbf{S}_w$  is the *within-class scatter matrix* and  $\mathbf{S}_b$  is the *between-class scatter matrix*. In particular, the within-class scatter matrix describes how data are scattered around the

means of their respective class, and is given by

$$\mathbf{S}_w = \sum_{i=1}^K Pr(\omega_i) E [(\mathbf{H}\mathbf{z})(\mathbf{H}\mathbf{z})^T | \omega = \omega_i], \quad (3.3)$$

where  $K$  is the number of classes,  $Pr(\omega_i)$  is the prior probability of each class, and  $\mathbf{H} = \mathbf{A} - E\mathbf{A}$ . The between-class scatter matrix describes how different classes. Which represented by their expected value, are scattered around the mixture means by

$$\mathbf{S}_b = \sum_{i=1}^K Pr(\omega_i) E [(\mathbf{F}\mathbf{z})(\mathbf{F}\mathbf{z})^T], \quad (3.4)$$

where  $\mathbf{F} = E[\mathbf{A} | \omega = \omega_i] - E[\mathbf{A}]$ .

With the linearity properties of both the trace function and the expectation,  $J(\mathbf{z})$  may be rewritten as

$$\begin{aligned} J(\mathbf{z}) &= \frac{tr(\sum_{i=1}^K Pr(\omega_i) E [(\mathbf{F}\mathbf{z})(\mathbf{F}\mathbf{z})^T])}{tr(\sum_{i=1}^K Pr(\omega_i) E [(\mathbf{H}\mathbf{z})(\mathbf{H}\mathbf{z})^T | \omega = \omega_i])} \\ &= \frac{\sum_{i=1}^K Pr(\omega_i) E [tr((\mathbf{F}\mathbf{z})(\mathbf{F}\mathbf{z})^T)]}{\sum_{i=1}^K Pr(\omega_i) E (tr [(\mathbf{H}\mathbf{z})(\mathbf{H}\mathbf{z})^T | \omega = \omega_i])} \\ &= \frac{\sum_{i=1}^K Pr(\omega_i) E [tr((\mathbf{F}\mathbf{z})^T (\mathbf{F}\mathbf{z}))]}{\sum_{i=1}^K Pr(\omega_i) E (tr [(\mathbf{H}\mathbf{z})^T (\mathbf{H}\mathbf{z}) | \omega = \omega_i])} \\ &= \frac{tr(\mathbf{z}^T (\sum_{i=1}^K Pr(\omega_i) E [\mathbf{F}^T \mathbf{F}]) \mathbf{z})}{tr(\mathbf{z}^T (\sum_{i=1}^K Pr(\omega_i) E [\mathbf{H}^T \mathbf{H} | \omega = \omega_i]) \mathbf{z})} \\ &= \frac{tr(\mathbf{z}^T \tilde{\mathbf{S}}_b \mathbf{z})}{tr(\mathbf{z}^T \tilde{\mathbf{S}}_w \mathbf{z})}. \end{aligned} \quad (3.5)$$

Furthermore,  $\tilde{\mathbf{S}}_b$  and  $\tilde{\mathbf{S}}_w$  can be evaluated as follows:

$$\tilde{\mathbf{S}}_b = \sum_{i=1}^K \frac{n_i}{K} (\bar{\mathbf{A}}_i - \bar{\mathbf{A}})^T (\bar{\mathbf{A}}_i - \bar{\mathbf{A}}) \quad (3.6)$$

$$\tilde{\mathbf{S}}_w = \sum_{i=1}^K \frac{n_i}{K} \sum_{\mathbf{A}_k \in \omega_i} (\mathbf{A}_k - \bar{\mathbf{A}}_i)^T (\mathbf{A}_k - \bar{\mathbf{A}}_i), \quad (3.7)$$

where  $n_i$  and  $\bar{\mathbf{A}}_i$  are the number of elements and the expected value of class  $\omega_i$  respectively.  $\bar{\mathbf{A}}$  denotes the overall mean.

Then the optimal projection vector can be found by solving the following generalized eigenvalue problem:

$$\tilde{\mathbf{S}}_b \mathbf{z} = \lambda \tilde{\mathbf{S}}_w \mathbf{z}. \quad (3.8)$$

Again the SVD algorithm can be applied to solve this eigenvalue problem on the matrix  $\tilde{\mathbf{S}}_w^{-1} \tilde{\mathbf{S}}_b$ . Note that, in this size of scatter matrices involved in eigenvalue decomposition process is also become  $n$  by  $n$ . Thus, with the limited the training set, this decomposition



is more reliably than the eigenvalue decomposition based on the classical covariance matrix.

The number of projection vectors is then selected by the same procedure as in Eq.(2.18). Let  $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_q]$  be the projection matrix composed of  $q$  largest eigenvectors for 2DLDA. Given a  $m$  by  $n$  matrix  $\mathbf{A}$ , its projection onto the principal subspace spanned by  $\mathbf{z}_i$  is then given by

$$\mathbf{V} = \mathbf{AZ}. \quad (3.9)$$

The result of this projection  $\mathbf{V}$  is another matrix of size  $m$  by  $q$ . Like 2DPCA, this procedure takes a matrix as input and outputs another matrix. These 2 techniques can be further combined, their combination is explained in the next section.

### 3.1.2 2DPCA+2DLDA

In this section, we apply our proposed 2DLDA within the well-known frameworks for face recognition, the LDA of PCA-based feature. The discriminant analysis of principal components framework [17] is applied in this section. Our framework consists of 2DPCA step and 2DLDA step, namely 2DPCA+2DLDA. From Section 2.2.1, we obtain a linear transformation matrix  $\mathbf{X}$  on which each input face image  $\mathbf{A}$  is projected. At the 2DPCA step, a feature matrix  $\mathbf{Y}$  is obtained. The matrix  $\mathbf{Y}$  is then used as the input for the 2DLDA step. Thus, the evaluation of within and between-class scatter matrices in this step will be slightly changed. From Eqs.(3.6) and (3.7), the image matrix  $\mathbf{A}$  is substituted for the 2DPCA feature matrix  $\mathbf{Y}$  as follows

$$\tilde{\mathbf{S}}_b^Y = \sum_{i=1}^K \frac{n_i}{K} (\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}})^T (\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}}) \quad (3.10)$$

$$\tilde{\mathbf{S}}_w^Y = \sum_{i=1}^K \frac{n_i}{K} \sum_{\mathbf{Y}_k \in \omega_i} (\mathbf{Y}_k - \bar{\mathbf{Y}}_i)^T (\mathbf{Y}_k - \bar{\mathbf{Y}}_i) \quad (3.11)$$

where  $\mathbf{Y}_k$  is the feature matrix of the  $k$ -th image matrix  $\mathbf{A}_k$ ,  $\bar{\mathbf{Y}}_i$  be the average of  $\mathbf{Y}_k$  which belong to class  $\omega_i$  and  $\bar{\mathbf{Y}}$  denotes a overall mean of  $\mathbf{Y}$ ,

$$\bar{\mathbf{Y}} = \frac{1}{M} \sum_{k=1}^M \mathbf{Y}_k. \quad (3.12)$$

The 2DLDA optimal projection matrix  $\mathbf{Z}$  can be obtained by solving the eigenvalue problem in Eq.(3.8). Finally, the composite linear transformation matrix,  $\mathbf{L}=\mathbf{XZ}$ , is used to map the face image space into the classification space by,

$$\mathbf{D} = \mathbf{AL}. \quad (3.13)$$

The matrix  $\mathbf{D}$  is 2DPCA+2DLDA feature matrix of image  $\mathbf{A}$  with dimension  $m$  by  $q$ . However, the number of 2DLDA feature vectors  $q$  cannot exceed the number of principal component vectors  $d$ . In general case ( $q < d$ ), the dimension of  $\mathbf{D}$  is less than  $\mathbf{Y}$  in Section 2.2.1. Thus, 2DPCA+2DLDA can reduce the classification time compared to 2DPCA.



### 3.1.3 2D Fisherface

Fisherface [16] is applied in this section. By projecting the face image matrix  $\mathbf{A}$  onto  $\mathbf{U}$ , which is given by

$$\mathbf{U} = \mathbf{X}\mathbf{W}, \quad (3.14)$$

where  $\mathbf{X}$  is 2DPCA projection in Section 2.2.1 and  $\mathbf{W}$  is the optimal discrimination projection which can be obtained via optimizing an alternative criterion in [16],

$$J(\mathbf{W}) = \frac{\text{tr} \left( \mathbf{W}^T \mathbf{X}^T \tilde{\mathbf{S}}_b \mathbf{X} \mathbf{W} \right)}{\text{tr} \left( \mathbf{W}^T \mathbf{X}^T \tilde{\mathbf{S}}_w \mathbf{X} \mathbf{W} \right)}. \quad (3.15)$$

It should be note that the difference between this criterion and criterion in Eq.(3.5) is absence or presence of the projection of  $\tilde{\mathbf{S}}_b$  and  $\tilde{\mathbf{S}}_w$  by 2DPCA projection  $\mathbf{X}$ .

The new feature extraction can be obtained by the new projection matrix  $\mathbf{U}$  with dimension  $n$  by  $q$ ,  $q \leq p$ , as follow

$$\mathbf{B} = \mathbf{A}\mathbf{U}, \quad (3.16)$$

where  $\mathbf{B}$  is feature matrix of image  $\mathbf{A}$  with dimension  $m$  by  $q$ , which less than the dimension of  $\mathbf{Y}$  in Section 2.2.1, in case  $q < d$ . Thus, the classification time of 2D Fisherface can be decreasing.

### 3.1.4 2DPCA+2DLDA Versus 2D Fisherface

In this section, we will show that the 2DLDA of principal component vectors framework is same as 2D Fisherface expect computation time which were consumed. In 2DLDA of 2DPCA, the 2DPCA feature matrices  $\mathbf{Y}$  are used for calculating the within and between-class scatter matrices while image matrices  $\mathbf{A}$  are used in 2D Fisherface. The between-class scatter matrix in Eq.(3.10) can be rewritten as

$$\begin{aligned} \tilde{\mathbf{S}}_b^Y &= \sum_{i=1}^K \frac{n_i}{K} \mathbf{X}^T (\bar{\mathbf{A}}_i - \bar{\mathbf{A}})^T (\bar{\mathbf{A}}_i - \bar{\mathbf{A}}) \mathbf{X} \\ &= \mathbf{X}^T \sum_{i=1}^K \frac{n_i}{K} (\bar{\mathbf{A}}_i - \bar{\mathbf{A}})^T (\bar{\mathbf{A}}_i - \bar{\mathbf{A}}) \mathbf{X} \\ &= \mathbf{X}^T \tilde{\mathbf{S}}_b \mathbf{X} \end{aligned} \quad (3.17)$$

and the within-class scatter matrix in Eq.(3.11) can be rewritten in the same way as

$$\tilde{\mathbf{S}}_w^Y = \mathbf{X}^T \tilde{\mathbf{S}}_w \mathbf{X}. \quad (3.18)$$

From Eqs.(3.17), (3.18) and the criterion in Eq.(3.15)

$$J(\mathbf{W}) = \frac{\text{tr} \left( \mathbf{W}^T \tilde{\mathbf{S}}_b^Y \mathbf{W} \right)}{\text{tr} \left( \mathbf{W}^T \tilde{\mathbf{S}}_w^Y \mathbf{W} \right)} \quad (3.19)$$

Thus, Eqs.(3.13) and (3.16) are same the linear transformation,  $\mathbf{D}=\mathbf{B}$  and  $\mathbf{L}=\mathbf{U}$ . Computing these scatter matrices by using feature matrix  $\mathbf{Y}$  is easier than image matrix  $\mathbf{A}$ , hence 2DPCA+2DLDA consume the training time less than 2D Fisherface. However, they consumed the equivalent classification time.



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

### 3.2 Class-Specific Subspace-Based Two-Dimensional Principal Component Analysis

2DPCA is a unsupervised technique that is no information of class labels are considered. Therefore, the directions that maximize the scatter of the data from all training samples might not be as adequate to discriminate between classes. In recognition task, a projection that emphasize the discrimination between classes is more important. The extension of Eigenface, PCA-based, was proposed by using alternative way to represent by projecting to Class-Specific Subspace (CSS) [37]. In conventional PCA method, the images are analyzed on the features extracted in a low-dimensional space learned from all training samples from all classes. While each subspaces of CSS learned from training samples from one class. In this way, the CSS representation can provide a minimum reconstruction error. The reconstruction error is used to classify the input data via the Distance From CSS (DFCSS). Less DFCSS means more probability that the input data belongs to the corresponding class.

We would like to note that this framework was the detailed and expanded version of our published materials, which is based on [64].

Let  $\mathbf{G}_k$  be the image covariance matrix of the  $k^{th}$  CSS. Then  $\mathbf{G}_k$  can be evaluated by

$$\mathbf{G}_k = \frac{1}{M} \sum_{\mathbf{A}_c \in \omega_k} (\mathbf{A}_c - \bar{\mathbf{A}}_k)^T (\mathbf{A}_c - \bar{\mathbf{A}}_k), \quad (3.20)$$

where  $\bar{\mathbf{A}}_k$  is the average image of class  $\omega_k$ . The  $k^{th}$  projection matrix  $\mathbf{X}_k$  is a  $n$  by  $d_k$  projection matrix which composed by the eigenvectors of  $\mathbf{G}_k$  corresponding to the  $d_k$  largest eigenvalues. The  $k^{th}$  CSS of 2DPCA was represented as a 3-tuple:

$$\mathfrak{R}_k^{2DPCA} = \{\mathbf{X}_k, \bar{\mathbf{A}}_k, d_k\} \quad (3.21)$$

Let  $\mathbf{S}$  be a input sample and  $\mathbf{U}_k$  be a feature matrix which projected to the  $k^{th}$  CSS, by

$$\mathbf{U}_k = \mathbf{W}_k \mathbf{X}_k, \quad (3.22)$$

where  $\mathbf{W}_k = \mathbf{S} - \bar{\mathbf{A}}_k$ . Then the reconstruct image  $\mathbf{W}_k^r$  can be evaluates by

$$\mathbf{W}_k^r = \mathbf{U}_k \mathbf{X}_k^T. \quad (3.23)$$

Therefore, the DFCSS is defined by reconstruction error as follows

$$\varepsilon_k(\mathbf{W}_k^r, \mathbf{S}) = \sum_{m=1}^{n_{row}} \sum_{n=1}^{n_{col}} \left| \mathbf{w}_{(m,n)_k}^r - \mathbf{s}_{(m,n)} \right|. \quad (3.24)$$

If  $\varepsilon_t = \min_{1 \leq k \leq K} (\varepsilon_k)$  then the input sample  $\mathbf{S}$  is belong to class  $\omega_t$ .

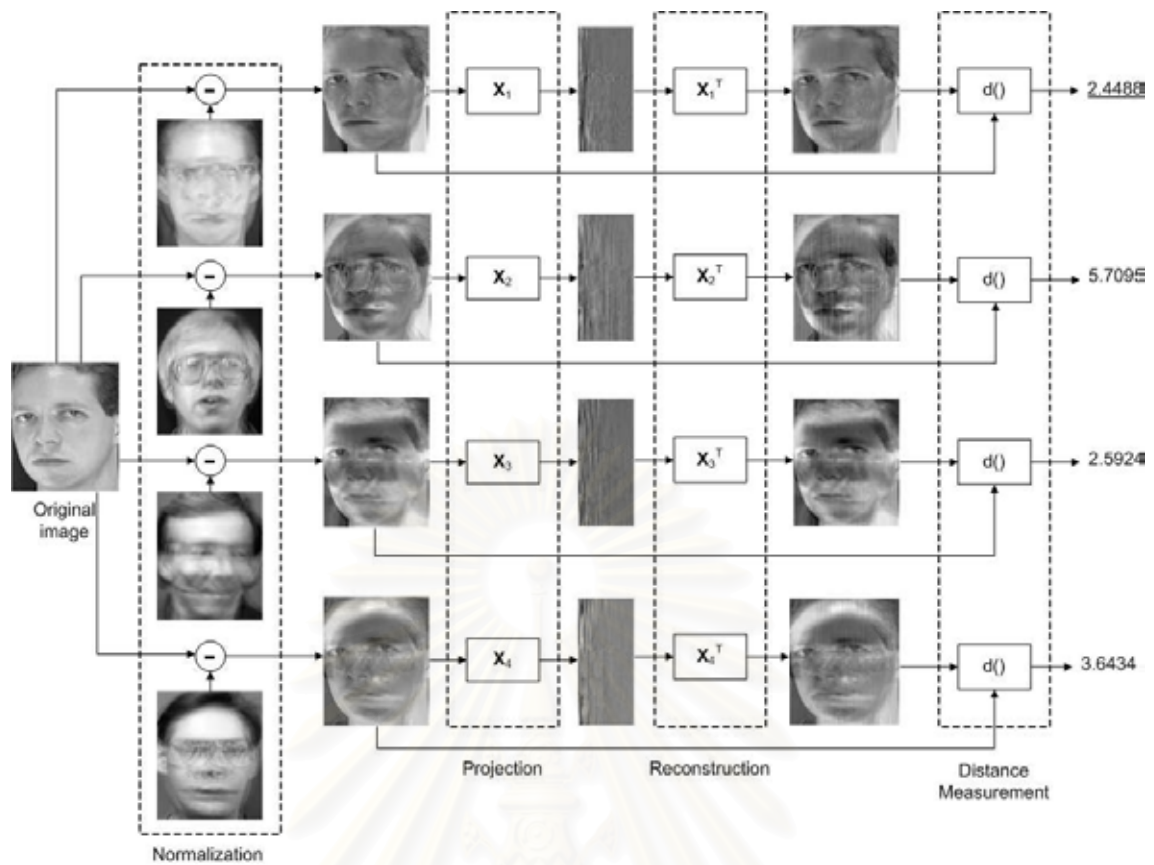


Figure 3.1 CSS-based 2DPCA diagram.

For illustration, we assume that there are 4 classes, as shown in Fig. 3.1. The input image must be normalized with the averaging images of all 4 classes. And then project to 2DPCA subspaces of each class. After that the image is reconstructed by the projection matrices ( $X$ ) in each class. The DFCSS is used now to measure the similarity between the reconstructed image and the normalized original image on each CSS. From Fig. 3.1, the DFCSS of the first class is minimum, thus we decide this input image is belong to the first class.

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

### 3.3 Image Cross-Covariance Analysis

In PCA, the covariance matrix provides a measure of the strength of the correlation of all pixel pairs. Because of the limit of the number of training samples, thus this covariance cannot be well estimated. While the performance of 2DPCA is better than PCA, although all of the correlation information of pixel pairs are not employed for estimating the image covariance matrix. Nevertheless, the disregard information may possibly include the useful information. In this dissertation, we proposed a framework for investigating the information which was neglected by original 2DPCA technique. To achieve this point, the *image cross-covariance matrix* is defined by two variables, the first variable is the original image and the second one is the shifted version of the former. By our shifting algorithm, many image cross-covariance matrices are formulated to cover all of the information. The Singular Value Decomposition (SVD) is applied to the image cross-covariance matrix for obtaining the optimal projection matrices. And we will show that these matrices can be considered as the orthogonally rotated projection matrices of traditional 2DPCA. According to the previous discussion, our proposed method, so called *Image Cross-Covariance Analysis (ICCA)*, is different from the original 2DPCA on the fact that the transformations of our method are generalized transformation of the original 2DPCA.

We would like to note that this framework was the detailed and expanded version of our published materials, which is based on [65].

#### 3.3.1 The Image Covariance Matrix Revisited

In 2D subspace analysis, the scatter matrix is straightforwardly estimated from the 2D image matrices. From Section 2.2.1, the *image covariance matrix*,  $\mathbf{G}$ , had given that

$$\mathbf{G} = E[(\mathbf{A} - E\mathbf{A})^T(\mathbf{A} - E\mathbf{A})]. \quad (3.25)$$

This is much smaller than the size of real covariance matrix needed in PCA, therefore it can be computed more accurately on small training set. Given a database of  $M$  training image matrices  $\mathbf{A}_k$ ,  $k = 1, \dots, M$  with same dimension  $m$  by  $n$ . The image inner-scatter matrix  $\mathbf{G}$  is computed in a straightforward manner by

$$\mathbf{G} = \frac{1}{M} \sum_{k=1}^M (\mathbf{A}_k - \bar{\mathbf{A}})^T (\mathbf{A}_k - \bar{\mathbf{A}}), \quad (3.26)$$

where  $\bar{\mathbf{A}}$  denotes the average image,

$$\bar{\mathbf{A}} = \frac{1}{M} \sum_{k=1}^M \mathbf{A}_k. \quad (3.27)$$



### 3.3.1.1 Relation to Covariance Matrix

First of all, the relationship between 2DPCA's image covariance matrix  $\mathbf{G}$ , in Eq.(2.14), and PCA's covariance matrix  $\mathbf{C}$  can be considered as

$$\mathbf{G}(i, j) = \sum_{k=1}^m \mathbf{C}(m(i-1) + k, m(j-1) + k) \quad (3.28)$$

where  $\mathbf{G}(i, j)$  and  $\mathbf{C}(i, j)$  are the  $i^{\text{th}}$  row,  $j^{\text{th}}$  column element of matrix  $\mathbf{G}$  and matrix  $\mathbf{C}$ , respectively. And  $m$  is the height of the image.

For illustration, let the dimension of all training images are 3 by 3. Thus, the covariance matrix of these images will be a 9 by 9 matrix and the dimension of image covariance matrix is only 3 by 3, as shown in Fig. 3.2.

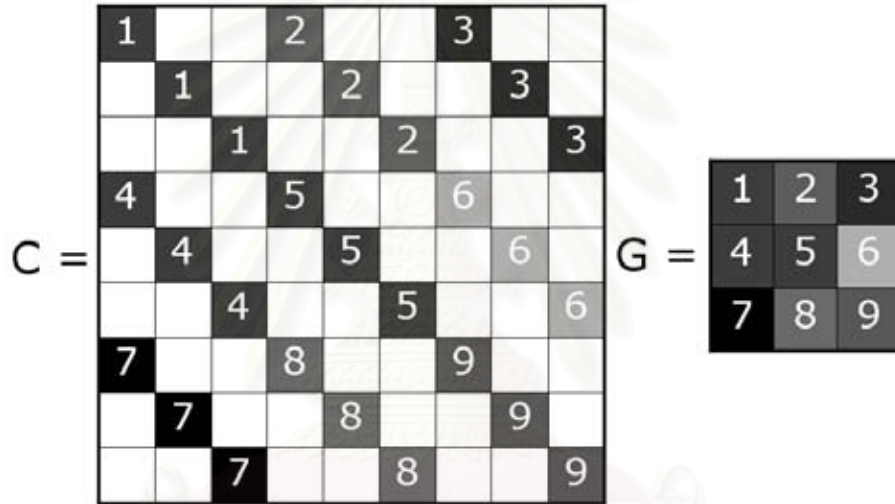


Figure 3.2 The relationship of covariance and image covariance matrix.

From Eq.(3.28), each elements of  $\mathbf{G}$  is the sum of all the same label elements in  $\mathbf{C}$ , for example:

$$\begin{aligned} \mathbf{G}(1, 1) &= \mathbf{C}(1, 1) + \mathbf{C}(2, 2) + \mathbf{C}(3, 3), \\ \mathbf{G}(1, 2) &= \mathbf{C}(1, 4) + \mathbf{C}(2, 5) + \mathbf{C}(3, 6), \\ \mathbf{G}(1, 3) &= \mathbf{C}(1, 7) + \mathbf{C}(2, 8) + \mathbf{C}(3, 9). \end{aligned} \quad (3.29)$$

It should be note that the total power of image covariance matrix equals and traditional covariance matrix  $\mathbf{C}$  are identical,

$$\text{tr}(\mathbf{G}) = \text{tr}(\mathbf{C}). \quad (3.30)$$

### 3.3.2 From Image Covariance Matrix to Image Cross-Covariance Matrix

From this point of view in Eq.(3.28), we can see that image covariance matrix is collecting the classification information only  $\frac{1}{m}$  of all information collected in traditional covariance matrix. Here there is a new raising question:

*Are the other  $\frac{m-1}{m}$  elements of the covariance matrix retaining unnecessary information?*

The answer of this question can be acquired by the experimental results in Section 4.5. However, since the performance of 2DPCA is still better than PCA, although only the limited information of the covariance matrix of PCA is used. Normally in face recognition task, the available number of training samples is relatively very small compared to the feature dimension, therefore the covariance matrix cannot be well estimated leading to the Small Sample Size (SSS) problem. For investigating how the retaining information in 2D subspace is rich for classification, the new  $\mathbf{G}$  is derived from the PCA's covariance matrix as

$$\mathbf{G}_L(i, j) = \sum_{k=1}^m \mathbf{C}(f(m(i-1) + k), m(j-1) + k), \quad (3.31)$$

$$f(x) = \begin{cases} x + L - 1, & 1 \leq x \leq mn - L + 1 \\ x - mn + L - 1, & mn - L + 2 \leq x \leq mn \end{cases} \quad (3.32)$$

where  $1 \leq L \leq mn$ .

We found that the  $\mathbf{G}_L$  can also be determined by applying the shifting to each images instead of averaging certain elements of covariance matrix. Therefore, the  $\mathbf{G}_L$  can alternatively be interpreted as the *image cross-covariance matrix* or

$$\mathbf{G}_L = E[(\mathbf{B}_L - E[\mathbf{B}_L])^T (\mathbf{A} - E[\mathbf{A}])] \quad (3.33)$$

where  $\mathbf{B}_L$  is the  $L^{\text{th}}$  shifted version of image  $\mathbf{A}$  that can be created via algorithm in Table 3.1. The samples of shifted images  $\mathbf{B}_L$  are presented in Fig. 3.3.

In 2DPCA, the columns of the projection matrix,  $\mathbf{X}$ , are obtained by selection the eigenvectors which corresponding to the  $d$  largest eigenvalues of image covariance matrix, in Eq.(2.14). While in ICCA, the eigenvalues of image cross-covariance matrix,  $\mathbf{G}_L$ , are complex number with non-zero imaginary part. The Singular Value Decomposition (SVD) is applied to this matrix instead of Eigenvalue decomposition. Thus, the ICCA projection matrix contains a set of orthogonal basis vectors which corresponding to the  $d$  largest singular values of image cross-covariance matrix.

For understanding the relationship between the ICCA projection matrix and the 2DPCA projection matrix, we will investigate in the simplest case, i.e. there are only one training image. Therefore, the image covariance matrix and image cross-covariance matrix are simplified to  $\mathbf{A}^T \mathbf{A}$  and  $\mathbf{B}_L^T \mathbf{A}$ , respectively.

Table 3.1 The Image Shifting Algorithm for ICCA

---

$S_1$ :	Input $m \times n$ original image $\mathbf{A}$ and the number of shifting $L$ ( $2 \leq L \leq mn$ ).
$S_2$ :	Initialize the row index, $irow = [2, \dots, n, 1]$ , and output image $\mathbf{B} = m \times n$ zero matrix.
$S_3$ :	For $i = 1, 2, \dots, L - 1$
$S_4$ :	Sort the first row of $\mathbf{A}$ by the row index, $irow$ .
$S_5$ :	Set the last row of $\mathbf{B} =$ the first row of $\mathbf{A}$ .
$S_6$ :	For $j = 1, 2, \dots, m - 1$
$S_7$ :	Set the $j^{th}$ row of $\mathbf{B} =$ the $(j + 1)^{th}$ row of $\mathbf{A}$ .
$S_8$ :	End For
$S_9$ :	Set $\mathbf{A} = \mathbf{B}$
$S_{10}$ :	End For

---

The image  $\mathbf{A}$  and  $\mathbf{B}_L$  can be decomposed by using Singular Value Decomposition (SVD) as

$$\mathbf{A} = \mathbf{U}_A \mathbf{D}_A \mathbf{V}_A^T, \quad (3.34)$$

$$\mathbf{B}_L = \mathbf{U}_{B_L} \mathbf{D}_{B_L} \mathbf{V}_{B_L}^T. \quad (3.35)$$

Where  $\mathbf{V}_A$  and  $\mathbf{V}_{B_L}$  contain a set of the eigenvectors of  $\mathbf{A}^T \mathbf{A}$  and  $\mathbf{B}_L^T \mathbf{B}_L$ , respectively. And  $\mathbf{U}_A$  and  $\mathbf{U}_{B_L}$  contain a set of the eigenvectors of  $\mathbf{A} \mathbf{A}^T$  and  $\mathbf{B}_L \mathbf{B}_L^T$ , respectively. And  $\mathbf{D}_A$  and  $\mathbf{D}_{B_L}$  contain the singular values of  $\mathbf{A}$  and  $\mathbf{B}_L$ , respectively. If all eigenvectors of  $\mathbf{A}^T \mathbf{A}$  are selected then the  $\mathbf{V}_A$  is the 2DPCA projection matrix, i.e.  $\mathbf{X} = \mathbf{V}_A$ .

Let  $\mathbf{Y} = \mathbf{A} \mathbf{V}_A$  and  $\mathbf{Z} = \mathbf{B}_L \mathbf{V}_{B_L}$  are the projected matrices of  $\mathbf{A}$  and  $\mathbf{B}$ , respectively. Thus,

$$\mathbf{B}_L^T \mathbf{A} = \mathbf{V}_{B_L} \mathbf{Z}^T \mathbf{Y} \mathbf{V}_A^T. \quad (3.36)$$

Denoting the SVD of  $\mathbf{Z}^T \mathbf{Y}$  by

$$\mathbf{Z}^T \mathbf{Y} = \mathbf{P} \mathbf{D} \mathbf{Q}^T, \quad (3.37)$$

and substituting into Eq.(3.36) gives

$$\begin{aligned} \mathbf{B}_L^T \mathbf{A} &= \mathbf{V}_{B_L} \mathbf{P} \mathbf{D} \mathbf{Q}^T \mathbf{V}_A^T \\ &= \mathbf{R} \mathbf{D} \mathbf{S}^T, \end{aligned} \quad (3.38)$$

where  $\mathbf{R} \mathbf{D} \mathbf{S}^T$  is the singular value decomposition of  $\mathbf{B}_L^T \mathbf{A}$  because of the unique properties of the SVD operation. It should be note that  $\mathbf{B}_L^T \mathbf{A}$  and  $\mathbf{Z}^T \mathbf{Y}$  have the same singular values. Therefore,

$$\mathbf{R} = \mathbf{V}_{B_L} \mathbf{P}, \quad (3.39)$$

$$\mathbf{S} = \mathbf{V}_A \mathbf{Q} = \mathbf{X} \mathbf{Q} \quad (3.40)$$

can be thought of as orthogonally rotated of projection matrices  $V_A$  and  $V_{B_L}$ , respectively.

As a result in Eq.(3.40), the ICCA projection matrix is the orthogonally rotated of original 2DPCA projection matrix.



Figure 3.3 The samples of shifted images on the ORL database.

### 3.3.3 Image Cross-Covariance Analysis (ICCA)

All training images are firstly transformed by the shifting algorithm in Table 3.1. And then using to estimate the image cross-covariance matrix in Eq.(3.33). The number of shifting  $L$  in Eq.(3.33) can be choosing between 1 to  $m \times n$ , where  $m$  and  $n$  are the number of the rows and columns of the image matrix, respectively.

The ICCA transformation matrix  $\mathbf{S}$  in Eq.(3.40) can be obtained by performing the SVD on the image cross-covariance matrix  $\mathbf{G}_L$ . Like in 2DPCA, the transformation matrix is consist of only the basis vectors which are corresponding to the first  $d$  singular values.

Therefore, there are two parameters here for adjusting the performance of ICCA, i.e. the number of shifting  $L$  and the number of basis vectors  $d$ .



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย



### 3.4 Random Subspace Method-Based Two-Dimensional Subspace Analysis

The main disadvantage of 2DPCA is that it needs many more coefficients for image representation than PCA. Many works try to solve this problem. In [39], PCA is used after 2DPCA for further dimensional reduction, but it is still unclear how the dimension of 2DPCA could be reduced directly. Many methods to overcome this problem were proposed by applied the bilateral-projection scheme to 2DPCA. In [41, 54], the right and left multiplying projection matrices are calculated independently while the iterative algorithm is applied to obtain the optimal solution of these projection matrices in [40, 56]. And the non-iterative algorithm for optimization was proposed in [66]. In [57], they proposed the iterative procedure which the right projection is calculated by the reconstructed images of the left projection and the left projection is calculated by the reconstructed images of the right projection. Nevertheless, all of above methods obtains only the local optimal solution.

Another method for dealing with high-dimensional space was proposed in [46], called Random Subspace Method (RSM). This method is the one of ensemble classification methods, like Bagging [44] and Boosting [45]. However, Bagging and Boosting are not reduce the high-dimensionality. Bagging randomly select a number of samples from the original training set to learn an individual classifier while Boosting specifically weight each training sample. The RSM can effectively exploit the high-dimensionality of the data. It constructs an ensemble of classifiers on independently selected feature subsets, and combines them using a heuristic such as majority voting, sum rule, etc.

There are many reasons the Random Subspace Method is suitable for face recognition task. Firstly, this method can take advantage of high dimensionality and far away from the curse of dimensionality [46]. Secondly, the random subspace method is useful for critical training sample sizes [50]. Normally in face recognition, the dimension of the feature is extremely large compared to the available number of training samples. Thus applying RSM can avoid both of the curse of dimensionality and the SSS problem. Thirdly, The nearest neighbor classifier, a popular choice in the 2D face-recognition domain [39–41, 54, 56, 66], can be very sensitive to the sparsity in the high-dimensional space. Their accuracy is often far from optimal because of the lack of enough samples in the high-dimensional space. The RSM brings significant performance improvements compared to a single classifier [50, 53]. Finally, since there is no hill climbing in RSM, there is no danger of being trapped in local optima [46].

The RSM was applied to PCA for face recognition in [47]. They apply the random selection directly to the feature vector of PCA for constructing the multiple subspaces. Nevertheless, the information which contained in each element of PCA feature vector is not equivalent. Normally, the element which corresponds to the larger eigenvalue,

contains more useful information. Therefore, applying RSM to PCA feature vector is seldom appropriate. Different from PCA, the 2DPCA feature is a matrix form. Thus, RSM is more suitable for 2DPCA, because the column direction does not depend on the eigenvalue.

In this dissertation, a framework of Two-Dimensional Random Subspace Analysis (2DRSA) is proposed to extend the original 2DPCA. The RSM is applied to feature space of 2DPCA for generating the vast number of feature subspaces, which be constructed by an autonomous, pseudorandom procedure to select a small number of dimensions from a original feature space. For a  $m$  by  $n$  feature matrix, there are  $2^m$  such selections that can be made, and with each selection a feature subspace can be constructed. And then individual classifiers are created only based on those attributes in the chosen feature subspace. The outputs from different individual classifiers are combined by the uniform majority voting to give the final prediction.

We would like to note that this framework was the detailed and expanded version of our published materials, which is based on [67,68].

#### 3.4.1 Two-Dimensional Random Subspace Analysis (2DRSA)

The Two-Dimensional Random Subspace Analysis consists of two parts, 2DPCA and RSM. After data samples was projected to 2D feature space via 2DPCA, the RSM are applied here by taking advantage of high dimensionality in these space to obtain the lower dimensional multiple subspaces. A classifier is then constructed on each of those subspaces, and a combination rule is applied in the end for prediction on the test sample. The 2DRSA algorithm is listed in Table 3.2, the image matrix,  $\mathbf{A}$ , is projected to feature space by 2DPCA projection in Eq.(2.19). In this feature space, it contains the data samples in matrix form, the  $m \times d$  feature matrix,  $\mathbf{Y}$  in Eq.(2.19). The dimensions of feature matrix  $\mathbf{Y}$  depend on the height of image ( $m$ ) and the number of selected eigenvectors of the image covariance matrix  $\mathbf{G}$  ( $d$ ). Therefore, only the information which embedded in each element on the row direction was sorted by the eigenvalue but not on the column direction. For this reason, we apply the random selection only on the the column direction or it means we randomly pick up some rows of feature matrix  $\mathbf{Y}$  to construct the new feature matrix  $\mathbf{Z}$ . The dimension of  $\mathbf{Z}$  is  $r \times d$ , normally  $r$  should be less than  $m$ . The results in [46] have shown that for a variety of data sets adopting half of the feature components usually yields good performance.

#### 3.4.2 Two-Dimensional Diagonal Random Subspace Analysis (2D<sup>2</sup>RSA)

The Two-Dimensional Diagonal Random Subspace Analysis consists of two parts, DiaPCA and RSM. Firstly, all image are transformed into the diagonal face images. After that the transformed image samples was projected to 2D feature space via DiaPCA, the RSM are applied here by taking advantage of high dimensionality in these space to obtain

Table 3.2 Two-Dimensional Random Subspace Analysis Algorithm

---

$S_1$ :	Project image, $\mathbf{A}$ , by Eq.(2.19).
$S_2$ :	For $i = 1$ to the number of classifiers
$S_3$ :	Randomly select a $r$ dimensional random subspace, $\mathbf{Z}_i^r$ , from $\mathbf{Y}$ ( $r < m$ ).
$S_4$ :	Construct the nearest neighbor classifier, $\mathbf{C}_i^r$ .
$S_5$ :	End For
$S_6$ :	Combine the output of each classifiers by using majority voting.

---

Table 3.3 Two-Dimensional Diagonal Random Subspace Analysis Algorithm.

---

$S_1$ :	Transforming images into diagonal images.
$S_2$ :	Project image, $\mathbf{A}$ , by Eq.(2.19).
$S_3$ :	For $i = 1$ to the number of classifiers
$S_4$ :	Randomly select a $r$ dimensional random subspace, $\mathbf{Z}_i^r$ , from $\mathbf{Y}$ ( $r < m$ ).
$S_5$ :	Construct the nearest neighbor classifier, $\mathbf{C}_i^r$ .
$S_6$ :	End For
$S_7$ :	Combine the output of each classifiers by using majority voting.

---

the lower dimensional multiple subspaces. A classifier is then constructed on each of those subspaces, and a combination rule is applied in the end for prediction on the test sample. Similar to 2DRSA, the 2D<sup>2</sup>RSA algorithm is listed in Table 3.3.

### 3.4.3 Random Subspace Method-based Image Cross-Covariance Analysis

As discussed in Section 3.3.2, not all elements of the covariance matrix is used in 2DPCA. Although, the image cross-covariance matrix can be switching these elements to formulate many versions of image cross-covariance matrix, the  $\frac{m-1}{m}$  elements of the covariance matrix are still not advertent in the same time. For integrating this information, the Random Subspace Method (RSM) can be using here via randomly select the number of shifting  $L$  to construct a set of multiple subspaces. That means each subspace is formulated from difference versions of image cross-covariance matrix. And then individual classifiers are created only based on those attributes in the chosen feature subspace. The outputs from different individual classifiers are combined by the uniform majority voting to give the final prediction. Moreover, the RSM can be used again for constructing the subspaces which are corresponding to the difference number

of basis vectors  $d$ . Consequently, the number of all random subspaces of ICCA reaches to  $d \times L$ . That means applying the RSM to ICCA can be constructed more subspaces than 2DRSA. As a result, the RSM-based ICCA can alternatively be apprehended as the generalized 2DRSA.



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## CHAPTER IV

### THE EXPERIMENTAL RESULTS

#### 4.1 Image Databases

In this dissertation, four databases were used to evaluate the performance of the proposed frameworks. Three from all databases are the human face image database, i.e. Yale, ORL and AR. The another database contains the Synthetic Aperture Radar (SAR) images of military vehicles.

##### 4.1.1 Yale Database

The Yale database [69] contains 165 images of 15 subjects. There are 11 images per subject, one for each of the following facial expressions or configurations: center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and wink. All sample images of one person from the Yale database are shown in Fig. 4.1. Each image was manually cropped and resized to  $100 \times 80$  pixels.



Figure 4.1 The sample images of one subject in the Yale database.

##### 4.1.2 ORL Database

The ORL database [70] contains images from 40 individuals, each providing 10 different images. For some subjects, the images were taken at different times. The facial expressions open or closed eyes, smiling or non smiling and facial details (glasses or no glasses) also vary. The images were taken with a tolerance for some tilting and rotation of the face of up to 20 degrees. Moreover, there is also some variation in the scale of up to about 10 percent. All images are gray scale and normalized to a resolution of  $112 \times 92$  pixels. The 5 sample images of one person from the ORL database are shown in Fig. 4.2.





Figure 4.2 Five sample images of one subject in the ORL face database.

#### 4.1.3 AR Database

The AR face database [71] was created by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) at the U.A.B. It contains over 1,000 color images corresponding to 134 people's faces (75 men and 59 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). The pictures were taken at the CVC under strictly controlled conditions. No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants. Each person participated in two sessions, separated by two weeks (14 days) time. The same pictures were taken in both sessions. All images were manually cropped and resized to  $100 \times 80$  pixels, and then convert to 256 level gray scale images. Only 14 images without occlusions (sun glasses and scarf) are used for each subject, the total number of a whole used image is 1,294. The sample cropped images of one person from the AR database are shown in Fig. 4.3.



Figure 4.3 The sample images of one subject in the AR database.

#### 4.1.4 MSTAR Database

The MSTAR public release data set [72] contains high resolution Synthetic Aperture Radar (SAR) data collected by the DARPA/Wright laboratory Moving and Stationary Target Acquisition and Recognition (MSTAR) program. The data set contains SAR images with size  $128 \times 128$  of three difference types of military vehicles—BMP2 armored personal carriers (APCs), BTR70 APCs and T72 tanks. The SAR image of a vehicle is

dependent on the pose of the target vehicle relative. For improving performance, it is desirable to ensure that the images for a given vehicle have the same pose. Thus, the preprocessor that can identify the aspect angle is required. The sample images from the MSTAR database are shown in Fig.4.4. All images were centrally cropped to  $80 \times 80$  pixels.

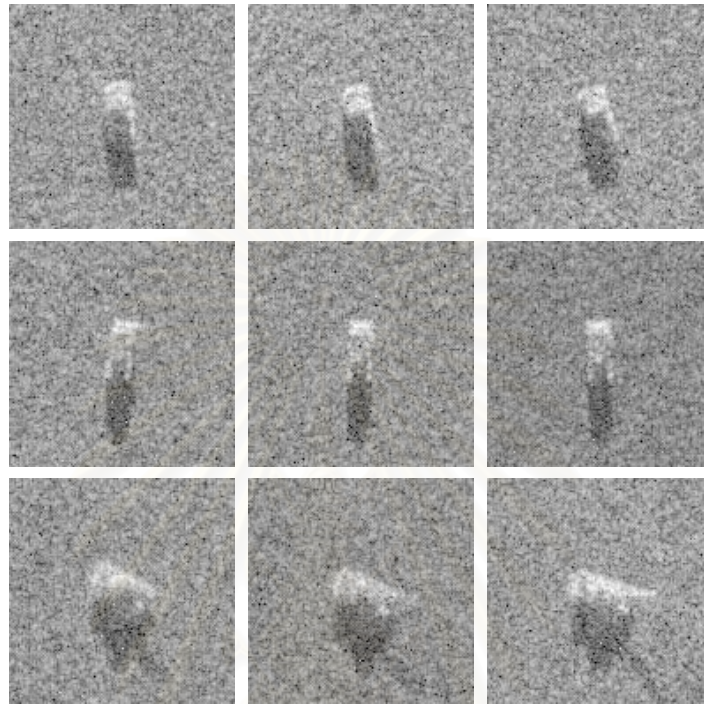


Figure 4.4: Sample SAR images of MSTAR database: the upper row is BMP2 APCs, the middle row is BTR70 APCs, and the lower row is T72 tank.

In all experiments in this dissertation, the training and testing sets of MSTAR database were referenced by Table 4.1 and Table 4.2. The aspect angle of the object was shown in Fig.4.5 (a). And the depression angle means the look angle pointed at the target by the antenna beam at the side of the aircraft, as shown in Fig.4.5 (b). Based on the different depression angles SAR images acquired at different times, the testing set can be used as a representative sample set of the SAR images of the targets for testing the recognition performance.

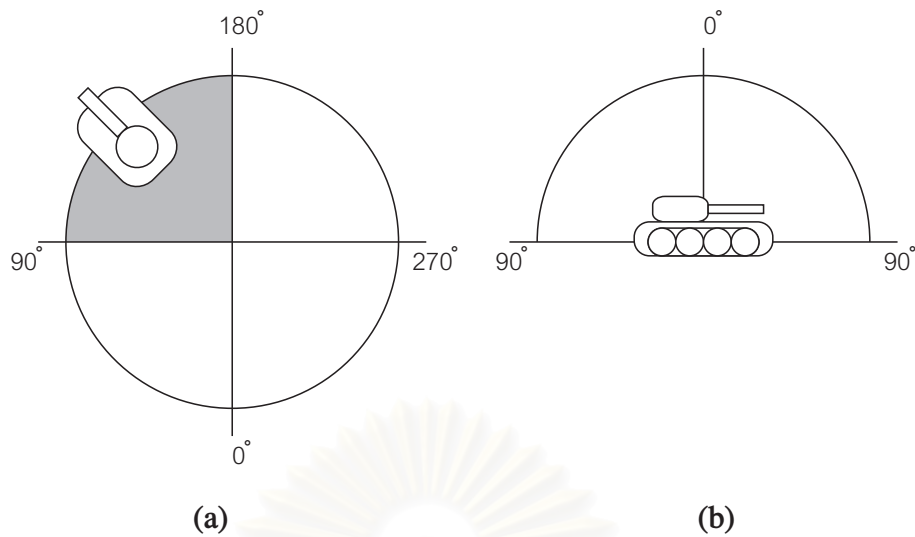


Figure 4.5 Aspect and Depression Angles.

Table 4.1 MSTAR images comprising training set

Class	Vehicle No.	Serial No.	Depression	#Images
BMP-2	1	9563		233
	2	9566	17°	231
	3	c21		233
BTR70	1	c71	17°	233
T-72	1	132		232
	2	812	17°	231
	3	s7		228
Total	-	-	17°	1,621

Table 4.2 MSTAR images comprising testing set

Class	Vehicle No.	Serial No.	Depression	#Images
BMP-2	1	9563		195
	2	9566	15°	196
	3	c21		196
BTR70	1	c71	15°	196
T-72	1	132		196
	2	812	15°	195
	3	s7		191
Total	-	-	15°	1,365

## 4.2 Preprocessing

In some databases, we notice that the background, some possible transformations of the object (scaling, rotation and translation) and sensor-dependent variations (for example, automatic gain control calibration and bad lens points) could undermine the recognition performance. This impact can be minimized by cropping and normalization. The preprocessing of this dissertation is following to the diagram in Fig. 4.6.

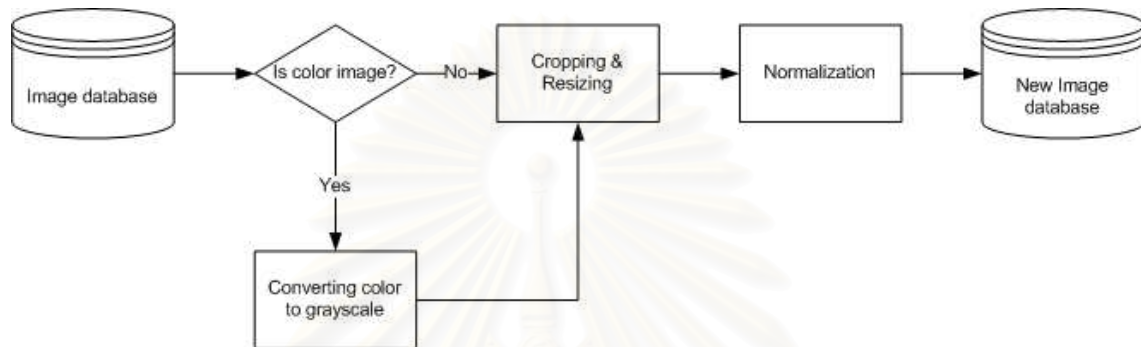


Figure 4.6 Preprocessing diagram

### 4.2.1 Converting Color to Grayscale

In some databases, they provide the data in color image which beyond the scope of this dissertation. Converting any color to its most approximate level of gray is required. First of all, we must obtain the values of its red, green and blue (RGB) primaries. And then it is sufficient to add about 30% of the red value plus about 59% of that of the green plus about 11% of that of the blue, no matter whatever scale is employed (0.0 to 1.0, 0 to 255, 0% to 100%, etc.). The resultant level is the desired gray value. These percentages are chosen due to the different relative sensitivity of the normal human eye to each of the primary colors (higher to the green, lower to the blue).

In this dissertation, the red ( $R$ ), green ( $G$ ) and blue ( $B$ ) primaries are converted to the grayscale by

$$Gray = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B, \quad (4.1)$$

where  $Gray$  is the intensity of each pixel in grayscale. Only AR database was converted to grayscale. The other databases are originally presented in grayscale.

### 4.2.2 Cropping and Resizing

In this dissertation, the cropping procedure was manually implemented by human. By attempting to align images such that the faces are the same size, in the same position

and at the same orientation. Specifically, the image is scaled and translated to make the eye coordinates coincident with pre-specified locations in the output. After cropping, all of the images were resized to same dimensions by linear interpolation. The cropping for a sample image on Yale database was shown in Fig. 4.7. The cropping was only applied to Yale, AR and MSTAR databases. For ORL database, the images were already cropped by originator.



Figure 4.7 Cropping image

### 4.2.3 Normalization

The normalization is to compensate for intensity variations. By

$$\mathbf{A}' = \frac{\mathbf{A}}{\|\text{vec}(\mathbf{A})\|}, \quad (4.2)$$

where  $\mathbf{A}$  is the original image matrix,  $\mathbf{A}'$  is the normalized image matrix and  $\|\text{vec}(\mathbf{A})\|$  represents to the norm of the vectorization of the image matrix.



### 4.3 Experiments and Analysis on Two-Dimensional Linear Discriminant Analysis of Principal Component Vectors

In this section, we experimentally evaluate our proposed technique by using 3 methods: 2DLDA, 2DPCA+2DLDA, and 2DPCA. 2DPCA is used as baseline method for comparison on face databases (Yale, ORL, and AR) and non-face database (MSTAR).

On Yale, AR, and MSTAR databases, the dimension of the image covariance matrix  $\mathbf{G}$ , the within-class scatter matrix  $\mathbf{S}_w$  and the between-class scatter matrix  $\mathbf{S}_b$  are equal to  $80 \times 80$ . For ORL, the dimension of these matrices is equal to  $92 \times 92$ . Thus, solving the eigenvalue problems with these sizes is uncomplicated.

#### 4.3.1 Effect of Number of Features

The first experiment is the performance comparison via varying the number of feature vectors.

On Yale database, the five image samples (centerlight, glasses, happy, leftlight, and noglasses) are used to train, and the six remaining images (normal, rightlight, sad, sleepy, surprised and wink) for test. For ORL and AR database, the first five images are used to train per subject, and the remaining images for test. For MSTAR database, the training and testing samples are shown in Table 4.1 and Table 4.2, respectively.

Firstly, we vary both of the number of principal component vectors ( $d$ ) and the number of 2DLDA feature vectors ( $q$ ). Since the input data of 2DLDA is the 2DPCA feature matrices, thus the number of 2DLDA feature vectors cannot exceed the number of principal component vectors that was used in the previous 2DPCA feature extraction ( $q \leq d$ ). It should be note that the performance of 2DPCA+2DLDA actually depends on both principal component vector and 2DLDA feature vector, as shown in Figs. 4.8, Fig. 4.10, Fig. 4.12, and Fig. 4.14. And the performance maps of these results are presented in Fig. 4.9, Fig. 4.11, Fig. 4.13, and Fig. 4.15. As this results on all databases, the numbers of 2DLDA feature vectors, which achieve the high recognition accuracies, are approximately ten percent of all 2DLDA feature vectors.

Secondly, for comparison reason, we plot the top recognition of Figs. 4.8, Fig. 4.10, Fig. 4.12, and Fig. 4.14, in the direction of the number of 2DLDA feature vectors axis, in Figs. 4.16, Figs. 4.17, Figs. 4.18, and 4.19, respectively. The results of pure 2DLDA are in agreement with pure LDA. That is the pure 2DLDA method includes the information which not useful for classification as its discriminant information, while the 2DPCA+2DLDA method can achieve higher recognition rate than other methods. Table 4.3 shows the comparisons of all three methods on the highest recognition accuracy, the number of principal component vectors ( $d$ ), the number of 2DLDA feature vectors ( $q$ ), and the dimension of feature matrices.

The almost recognition accuracies of our proposed 2DLDA and 2DPCA+2DLDA are superior to 2DPCA in all experiments. Moreover, since the number of feature vectors in case of 2DPCA+2DLDA must be less or equal to the number of principal component vectors of 2DPCA in the first step, so the computational time consumed for classification can be reduced.

Table 4.3: The Highest Recognition Accuracy Comparisons of 2DPCA, 2DLDA and 2DPCA+2DLDA on Yale, ORL, AR, and MSTAR Databases

Database	Method	Accuracy (%)	$d$	$q$	Dimension
Yale	2DPCA	95.56	11	-	$100 \times 11$
	2DLDA	97.78	-	14	$100 \times 14$
	2DLDA+2DPCA	98.89	18	17	$100 \times 17$
ORL	2DPCA	92.50	5	-	$112 \times 5$
	2DLDA	94.00	-	6	$112 \times 6$
	2DLDA+2DPCA	94.50	33	6	$112 \times 6$
AR	2DPCA	60.74	2	-	$100 \times 2$
	2DLDA	60.90	-	4	$100 \times 4$
	2DLDA+2DPCA	65.06	6	5	$100 \times 5$
MSTAR	2DPCA	99.49	11	-	$80 \times 11$
	2DLDA	99.56	-	78	$80 \times 78$
	2DLDA+2DPCA	99.63	37	35	$80 \times 35$

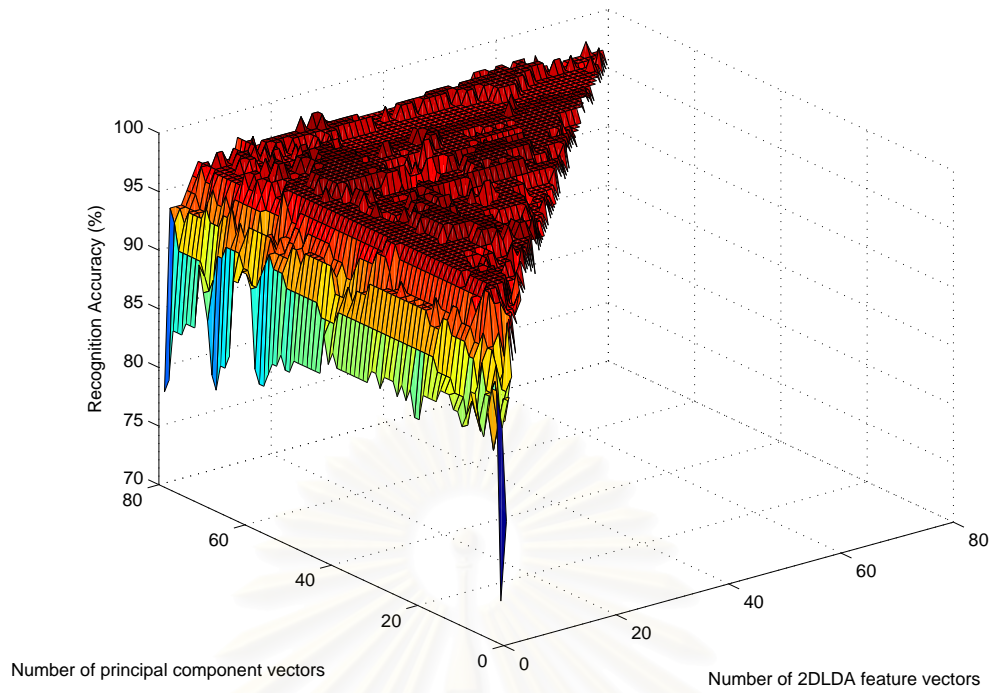


Figure 4.8 Recognition accuracy of 2DPCA+2DLDA on Yale database.

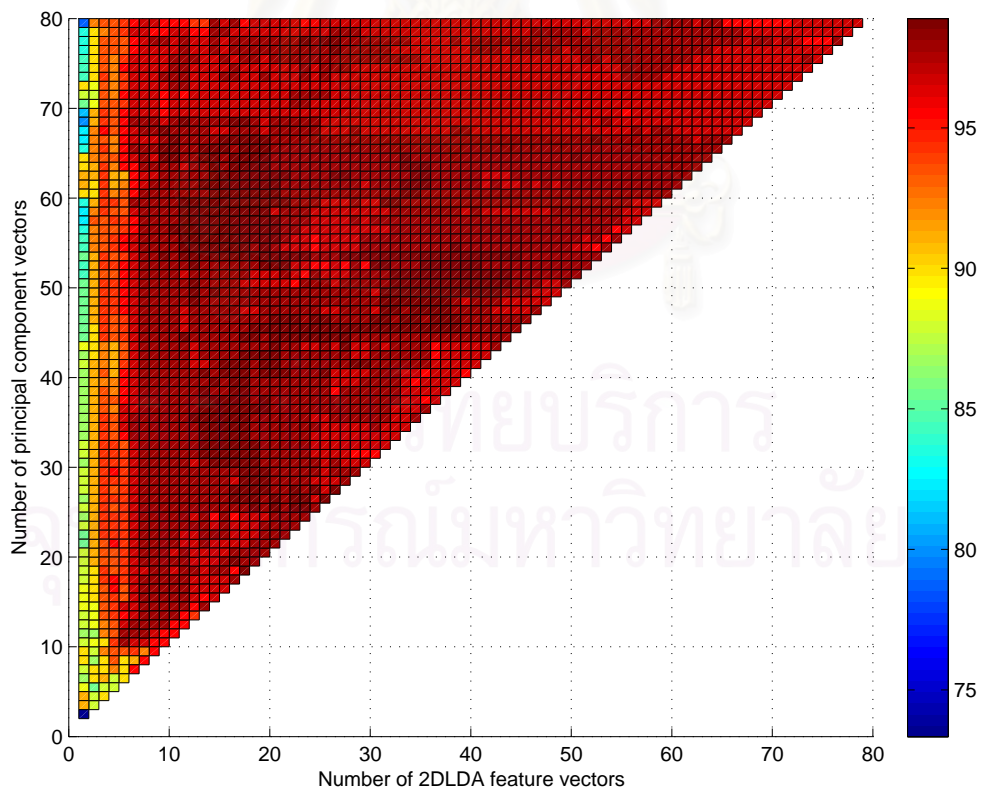


Figure 4.9 Performance map of 2DPCA+2DLDA on Yale database.

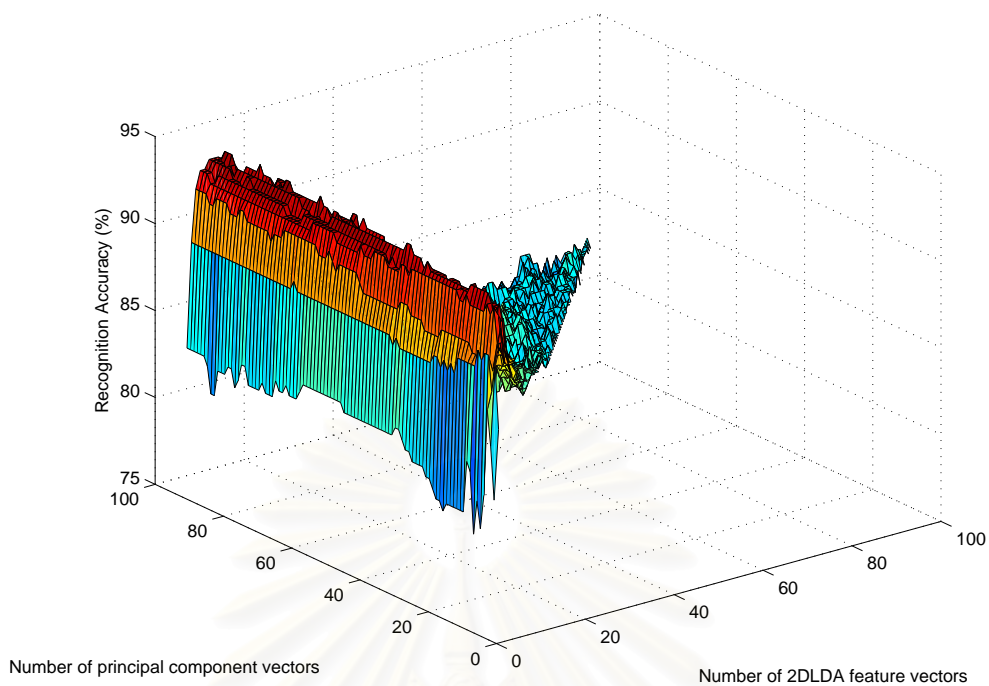


Figure 4.10 Recognition accuracy of 2DPCA+2DLDA on ORL database.

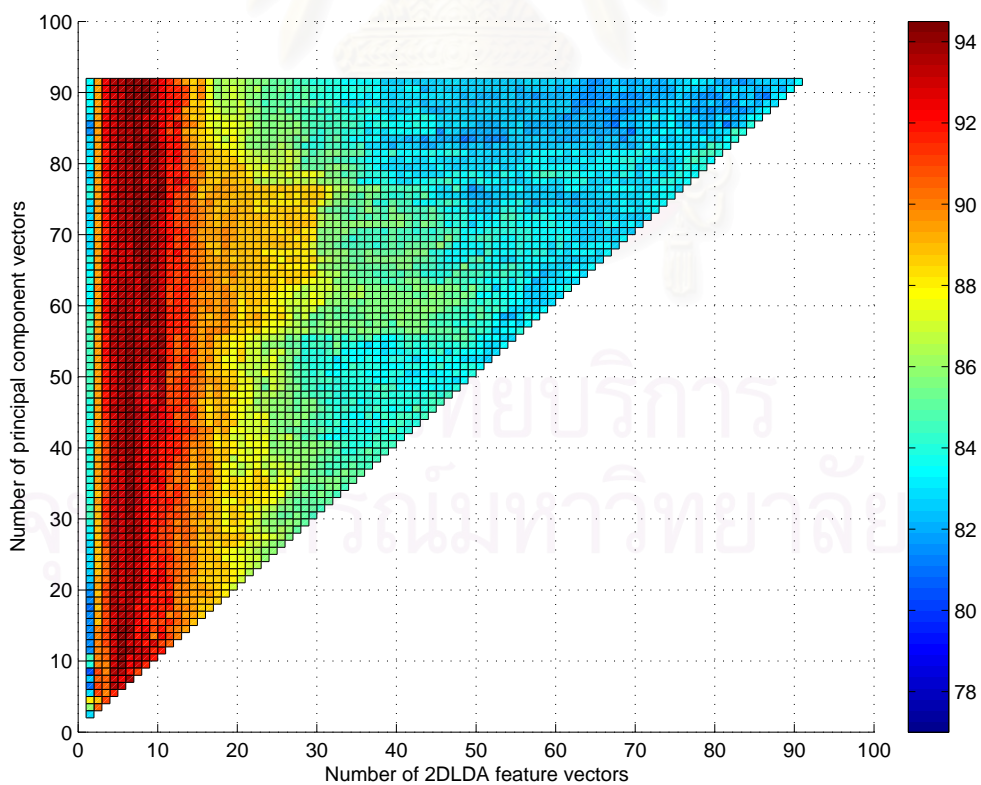


Figure 4.11 Performance map of 2DPCA+2DLDA on ORL database.

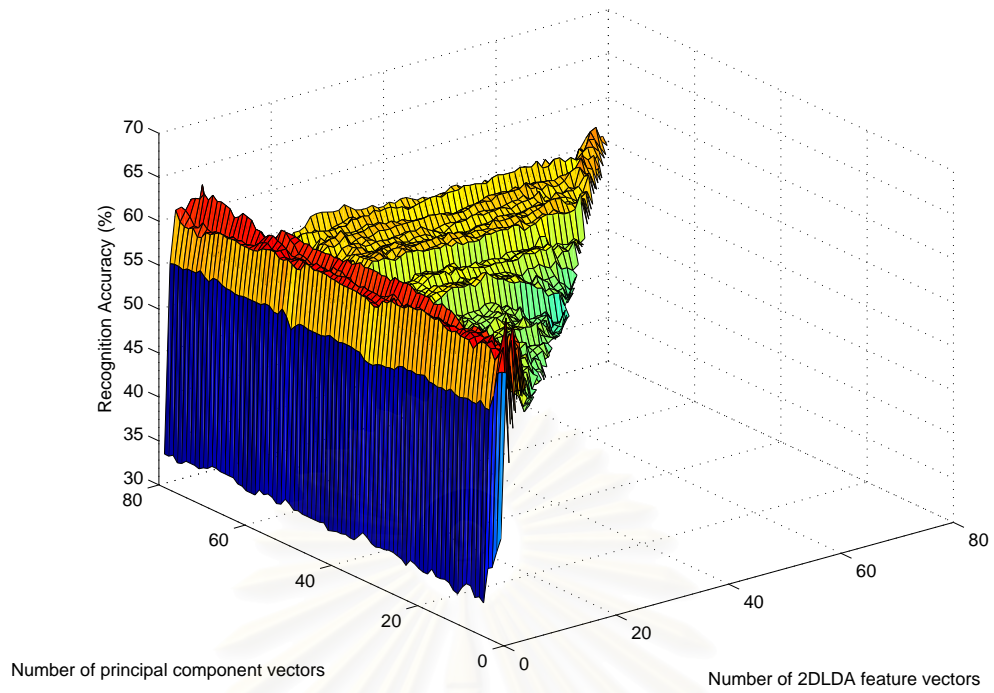


Figure 4.12 Recognition accuracy of 2DPCA+2DLDA on AR database.

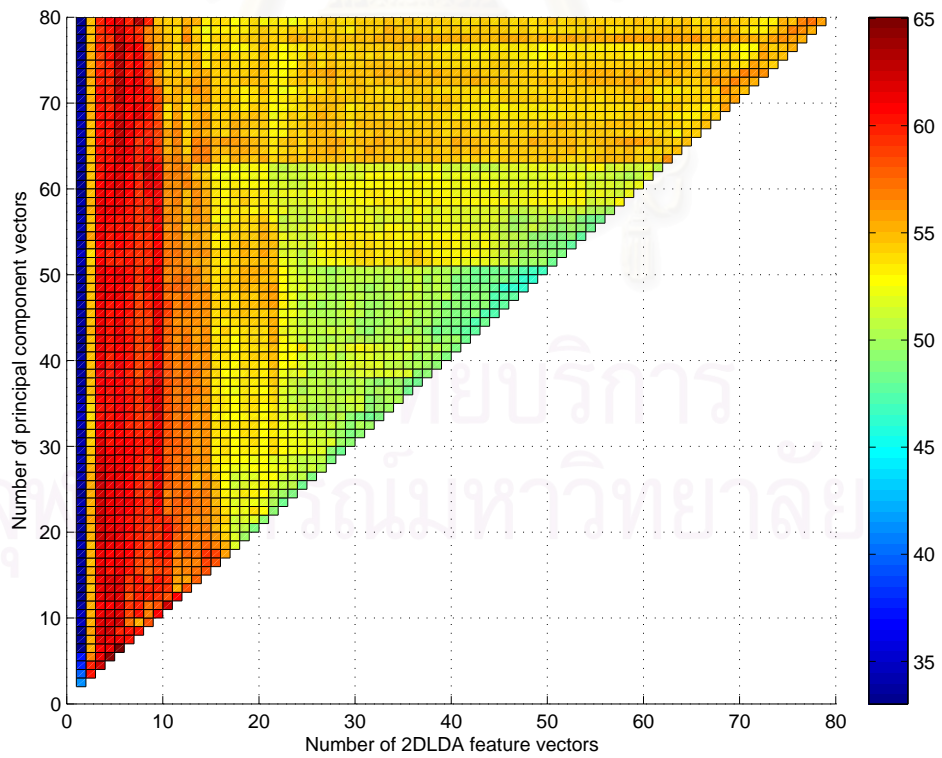


Figure 4.13 Performance map of 2DPCA+2DLDA on AR database.



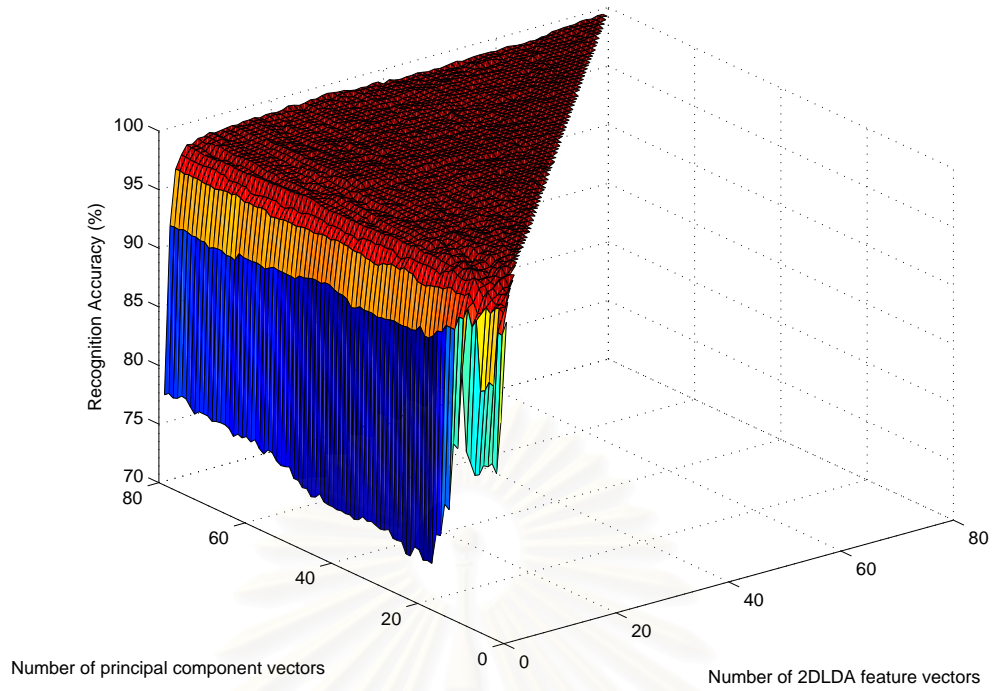


Figure 4.14 Recognition accuracy of 2DPCA+2DLDA on MSTAR database.

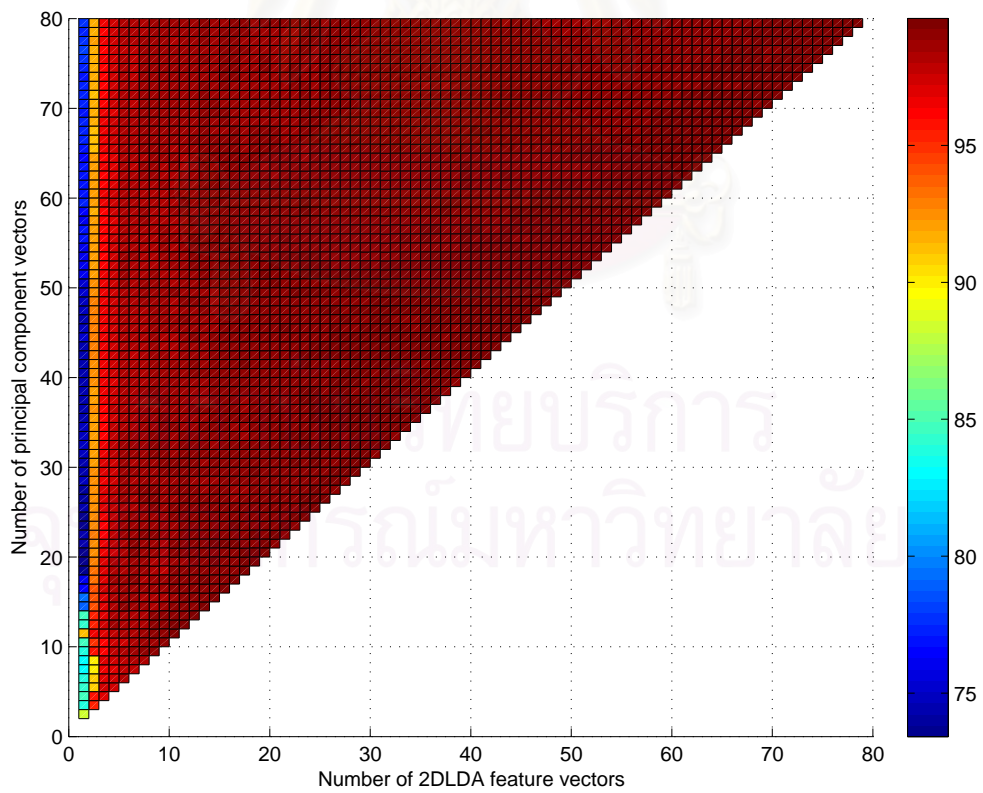


Figure 4.15 Performance map of 2DPCA+2DLDA on MSTAR database.

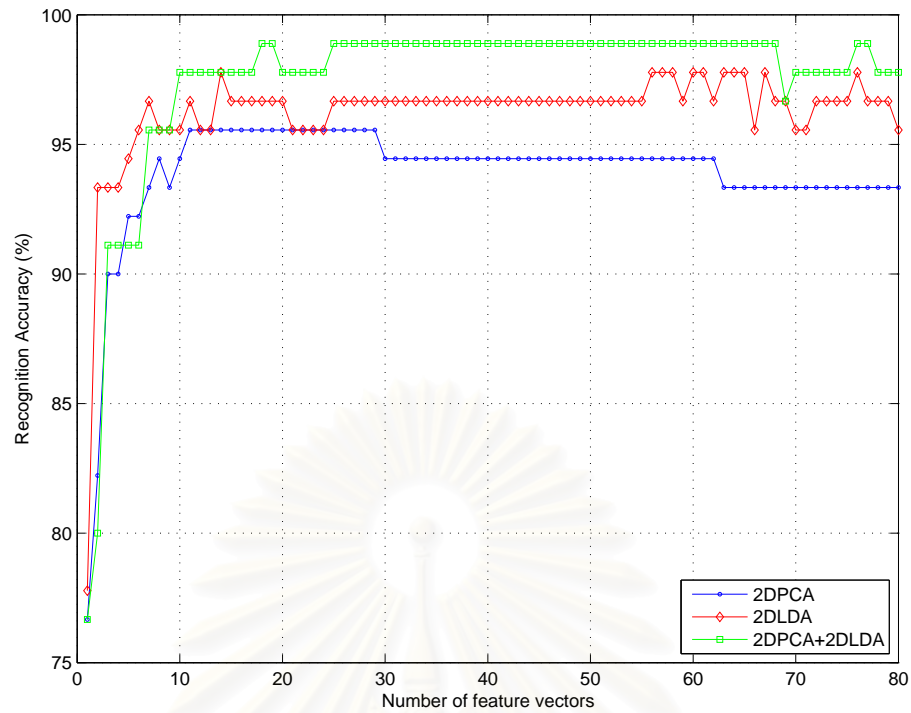


Figure 4.16: Recognition accuracy of 2DPCA, 2DLDA and 2DPCA+2DLDA on the Yale database.

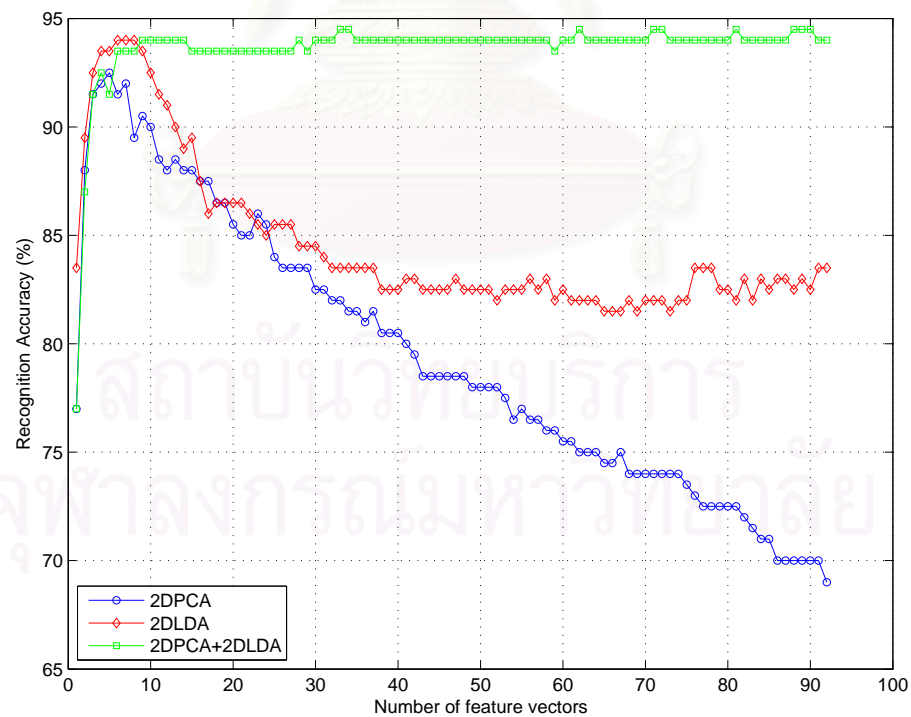


Figure 4.17: Recognition accuracy of 2DPCA, 2DLDA and 2DPCA+2DLDA on the ORL database.

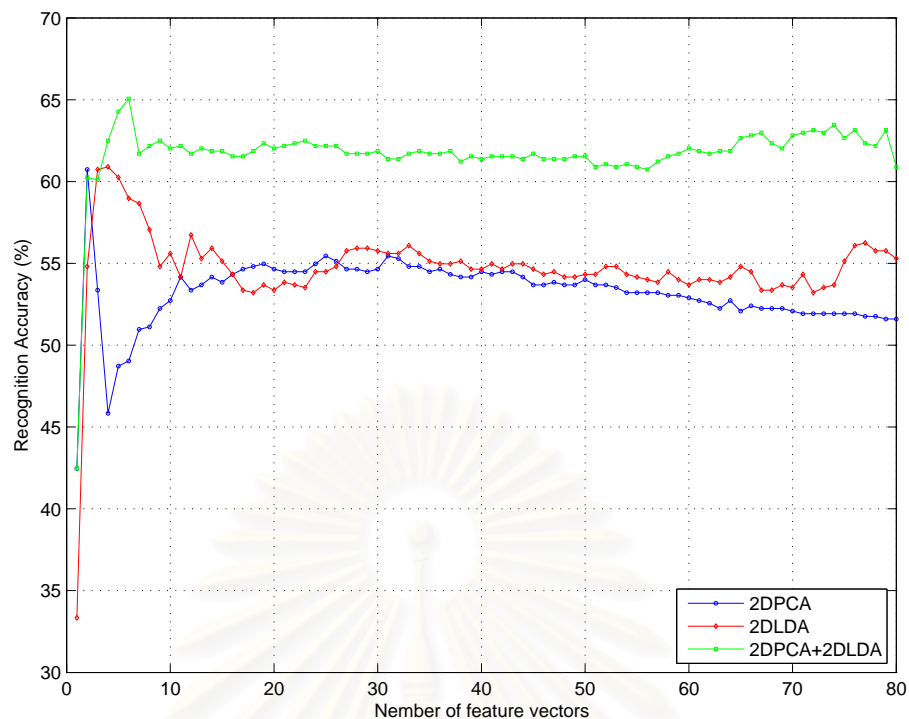


Figure 4.18: Recognition accuracy of 2DPCA, 2DLDA and 2DPCA+2DLDA on the AR database.

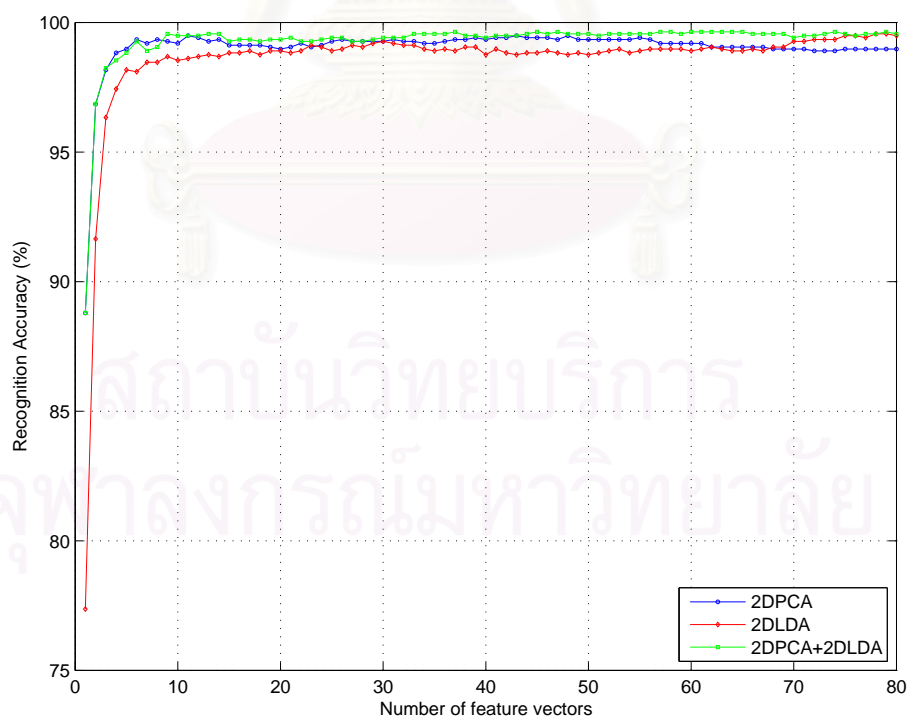


Figure 4.19: Recognition accuracy of 2DPCA, 2DLDA and 2DPCA+2DLDA on the MSTAR database.

### 4.3.2 Effect of Number of Training Samples

In this experiment, the effect of the number of training samples is considered by fixing the optimal number of feature vectors from Table 4.3, we use the different training numbers for each subject on Yale and ORL databases.

The number of training images were varied from 2 to 10 for Yale database and 2 to 9 for and ORL database and the remaining images are used to test. The experiment was repeated 100 times which the training and testing sets were randomly selected in each time. The averaging recognition accuracies of 2DPCA, 2DLDA, and 2DPCA+2DLDA on Yale and ORL databases were plotted against the number of training images for each subject as shown in Figs. 4.20 and 4.21, respectively.

Since the number of principal component vectors and number of 2DLDA feature vectors were set to the values that obtain the best accuracy when five images are used to train. Consequently, the performance of 2DPCA+2DLDA at the point of five training samples gives the best results on both databases, as shown in Figs. 4.20 and 4.21. However, the performance of 2DLDA and 2DPDA+2DLDA achieve the higher than 2DPCA when the number of the training samples is increasing while the most of results of both 2DLDA and 2DPDA+2DLDA obtain better the performance than 2DPCA on ORL database.

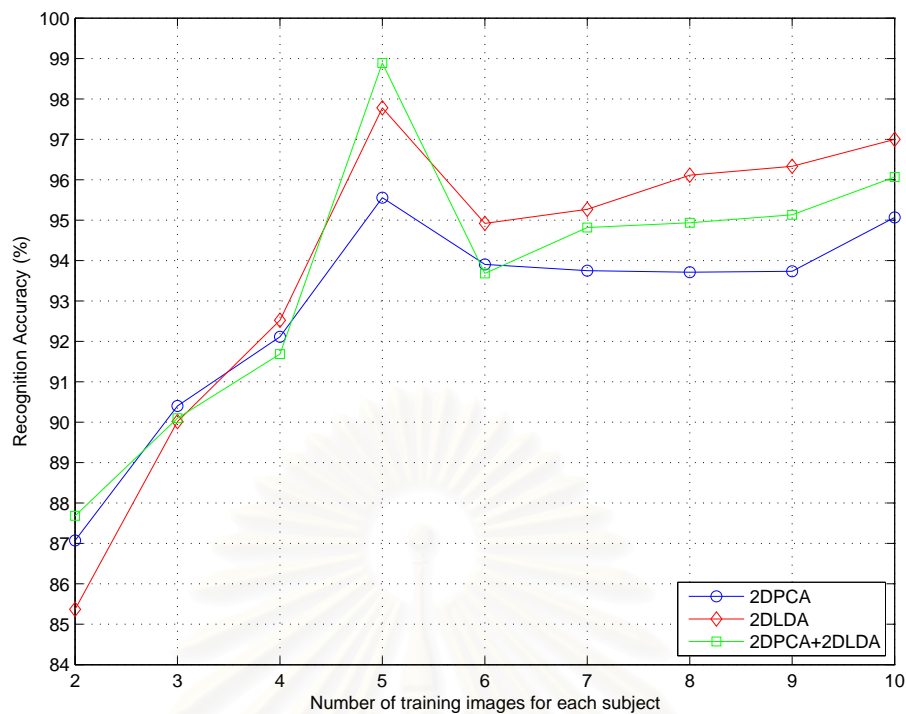


Figure 4.20: Performance of 2DPCA, 2DLDA and 2DPCA+2DLDA on the Yale database with different training numbers for each subject.

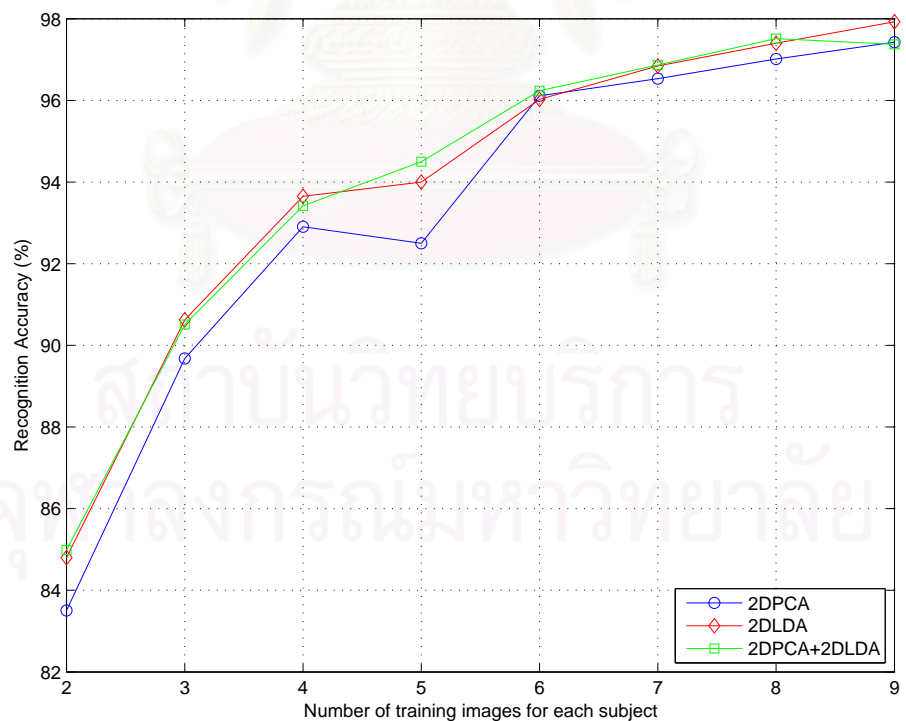


Figure 4.21: Performance of 2DPCA, 2DLDA and 2DPCA+2DLDA on the ORL database with different training numbers for each subject.



#### 4.4 Experiments and Analysis on Class-Specific Subspace-Based Two-Dimensional Principal Component Analysis

In this section, we experimentally evaluate our proposed framework, CSS-Based 2DPCA, by using 2DPCA as the baseline method for comparison.

In all experiments, the training set is divided to  $K$  sets, where  $K$  is the number of classes in each database. There are 15, 40, 134 and 3 classes on Yale, ORL, AR and MSTAR database, respectively.

The first experiment is the performance comparison via varying the number of feature vectors.

On Yale database, the five image samples (centerlight, glasses, happy, leftlight, and noglasses) are used to train, and the six remaining images (normal, rightlight, sad, sleepy, surprised and wink) for test. For ORL and AR database, the first five images are used to train per subject, and the remaining images for test. For MSTAR database, the training and testing samples are shown in Table 4.1 and Table 4.2, respectively.

We vary the number of principal component vectors ( $d$ ) from 1 to the number of pixels in the width direction. The compared results with 2DPCA are shown in Fig. 4.22, Fig. 4.23, Fig. 4.24, and Fig. 4.25. And the highest recognition accuracy comparisons of 2DPCA and CSS-based 2DPCA on Yale, ORL, AR, and MSTAR databases are presented in Table. 4.4.

From these results, the performance of CSS-based 2DPCA give the best results when a few of the number of principal component vectors are used on all face databases. Especially, it obtains the best recognition accuracy on Yale database. However, the performance of this method is lower than 2DPCA on the big database, such as MSTAR database.

Table 4.4: The Highest Recognition Accuracy Comparisons of 2DPCA and CSS-based 2DPCA on Yale, ORL, AR, and MSTAR databases

Database	Method	Accuracy (%)	$d$	Dimension
Yale	2DPCA	95.56	11	$100 \times 11$
	2DPCA+CSS	97.78	9	$100 \times 9$
ORL	2DPCA	92.50	5	$112 \times 5$
	2DPCA+CSS	86.00	4	$112 \times 4$
AR	2DPCA	60.74	2	$100 \times 2$
	2DPCA+CSS	51.28	1	$100 \times 1$
MSTAR	2DPCA	99.49	11	$80 \times 11$
	2DPCA+CSS	78.53	5	$80 \times 5$

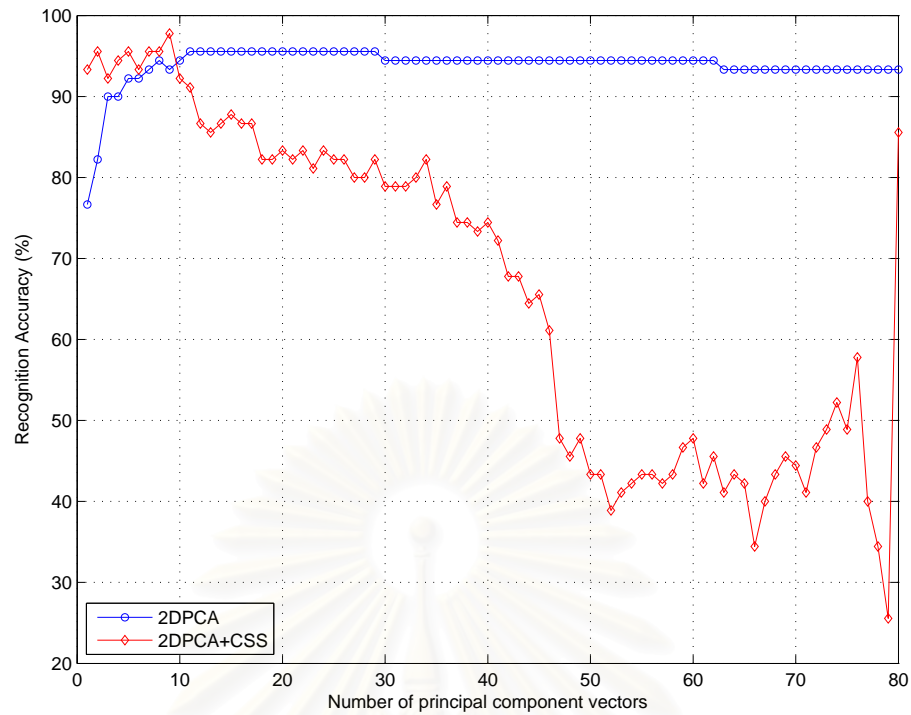


Figure 4.22: Recognition accuracy of CSS-based 2DPCA and 2DPCA on the Yale database.

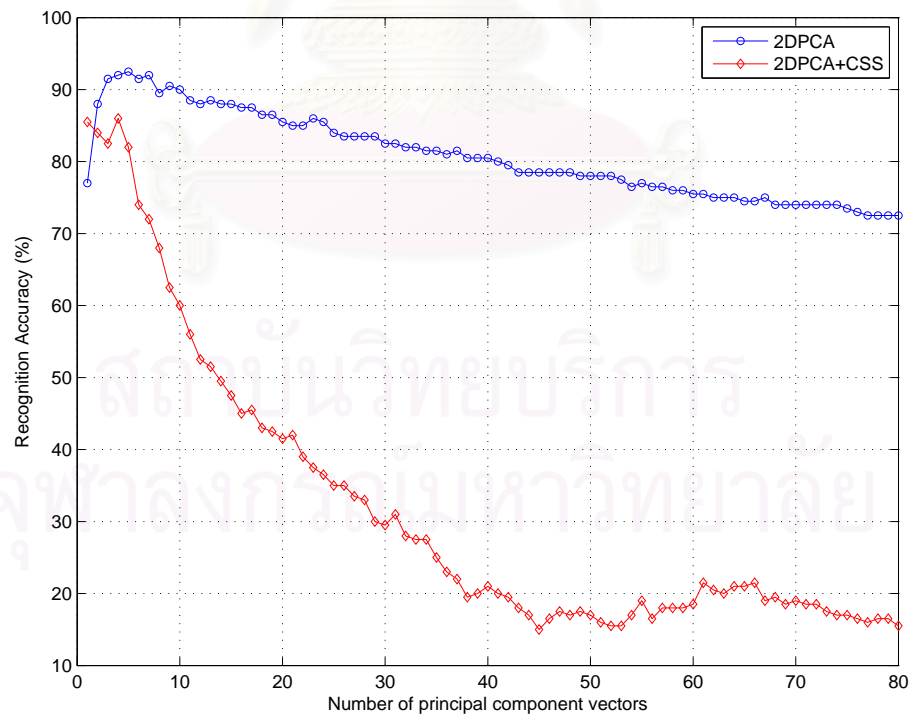


Figure 4.23: Recognition accuracy of CSS-based 2DPCA and 2DPCA on the ORL database.

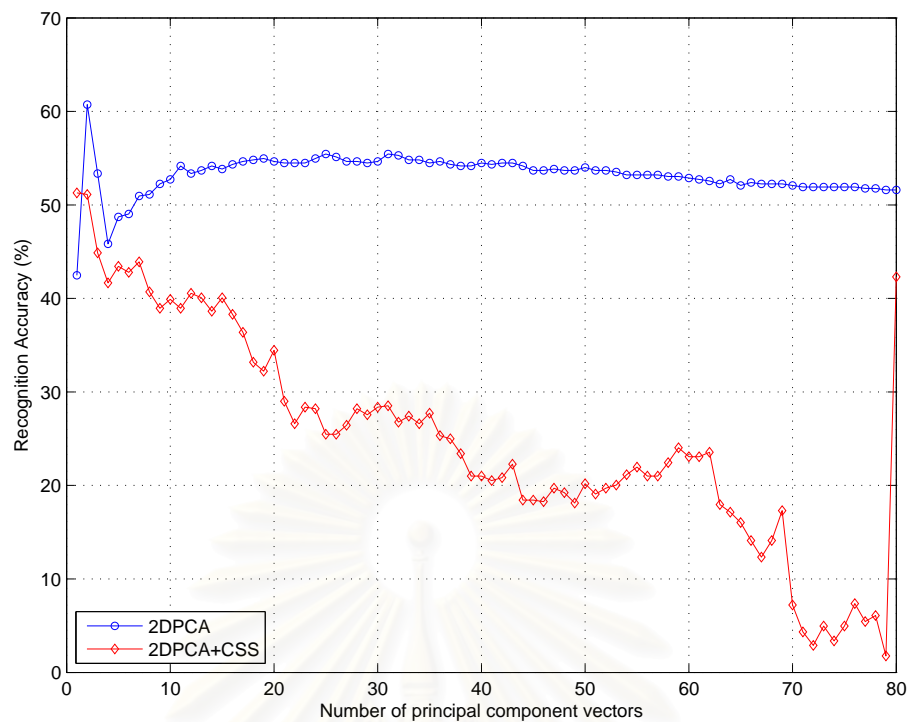


Figure 4.24: Recognition accuracy of CSS-based 2DPCA and 2DPCA on the AR database.

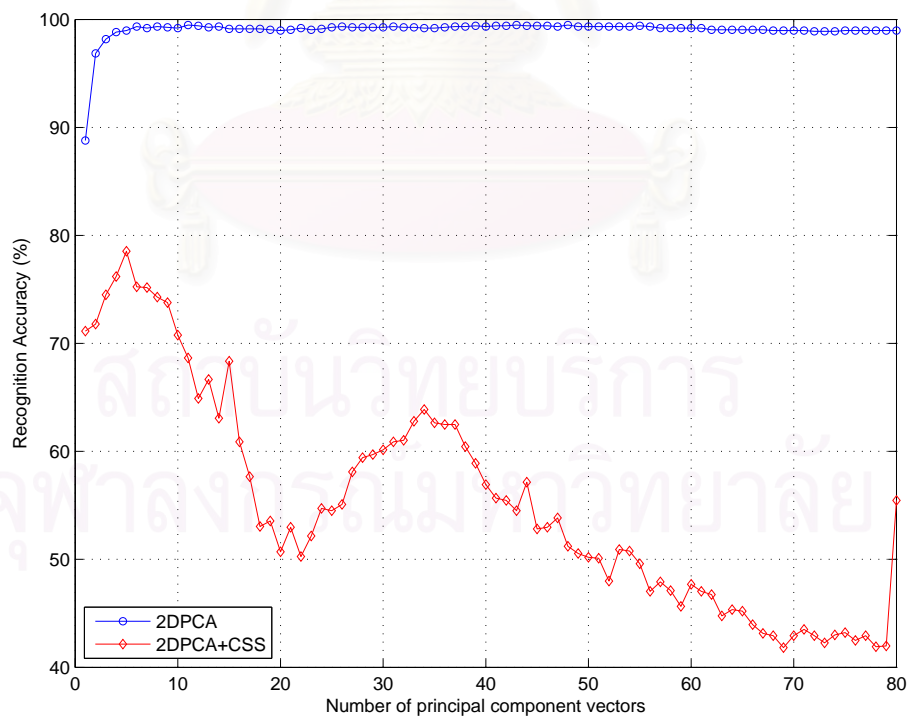


Figure 4.25: Recognition accuracy of CSS-based 2DPCA and 2DPCA on the MSTAR database.

## 4.5 Experiments and Analysis on Image Cross-Covariance Analysis

In this section, we experimentally evaluate our proposed framework, Image Cross-Covariance Analysis (ICCA), for the effect of the number of image shifting and principal component vectors.

In these experiments, the performance comparison via varying the number of image shifting and the number of the principal component vectors.

On Yale database, the five image samples (centerlight, glasses, happy, leftlight, and noglasses) are used to train, and the six remaining images (normal, rightlight, sad, sleepy, surprised and wink) for test. For ORL and AR database, the first five images are used to train per subject, and the remaining images for test. For MSTAR database, the training and testing samples are shown in Table 4.1 and Table 4.2, respectively.

Firstly, we vary both of the number of image shifting ( $L$ ) and the number of principal component vectors ( $d$ ). If  $L = 1$  then it is original 2DPCA. Actually, the maximum number of image shifting can be reached to the number of a whole pixels in the image ( $width \times height$ ), i.e. 8000 on Yale, 10304 on ORL, 8000 on AR, and 6400 on MSTAR. However, only a few number of shifting image can achieve the highest recognition accuracy. Therefore, the number of image shifting was varied only from 1 to 80. While the number of principal component vectors were varied from 1 to 80, 92, 64, and 30 on Yale, ORL, AR, and MSTAR databases, respectively. The results on Yale, ORL, AR, and MSTAR databases are shown in Fig. 4.26, Fig. 4.28, Fig. 4.30, and Fig. 4.32, respectively. And the performance maps of these results are presented in Fig. 4.26, Fig. 4.28, Fig. 4.30, and Fig. 4.32.

Secondly, we fixed the number of principal component vectors ( $d$ ) to the value that obtains the highest recognition accuracies on each database in the former experiments while we vary the number of image shifting ( $L$ ) from 1 to 1000 on Yale, ORL, AR, and MSTAR. The results are shown in Fig. 4.34, Fig. 4.36, Fig. 4.38, and Fig. 4.40. From these results, we found that they are the periodic curve which the number of image shifting in a period equals to the height of the image. The one period of these results are shown in Fig. 4.35, Fig. 4.37, Fig. 4.39, and Fig. 4.41.

The samples of shifted image which obtain the best recognition accuracy on the Yale, ORL and AR databases are illustrated with the its original images in Fig. 4.42, 4.43 and 4.44, respectively. The best of recognition accuracy of these methods are presented in Table 4.9.

Our proposed method achieves the higher recognition accuracy than 2DPCA, while the complexity of testing process is almost same as 2DPCA. Because only one test sample is shifted. However, the complexity of training process of ICCA depends on the number and resolution of training images, thus it usually be higher than 2DPCA.

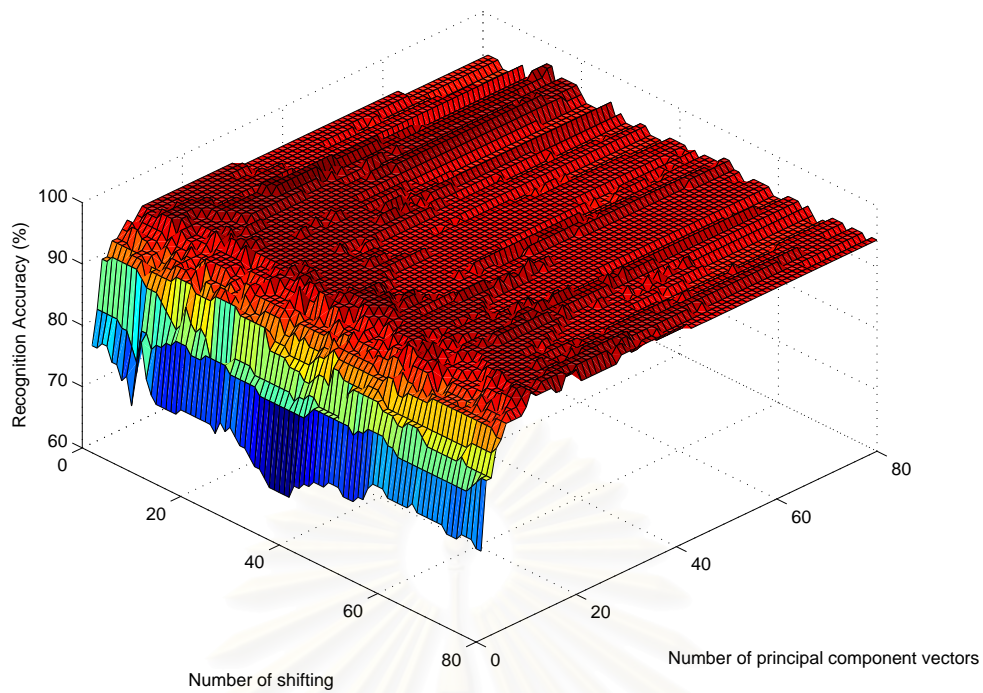


Figure 4.26 Recognition accuracy of ICCA on the Yale database.

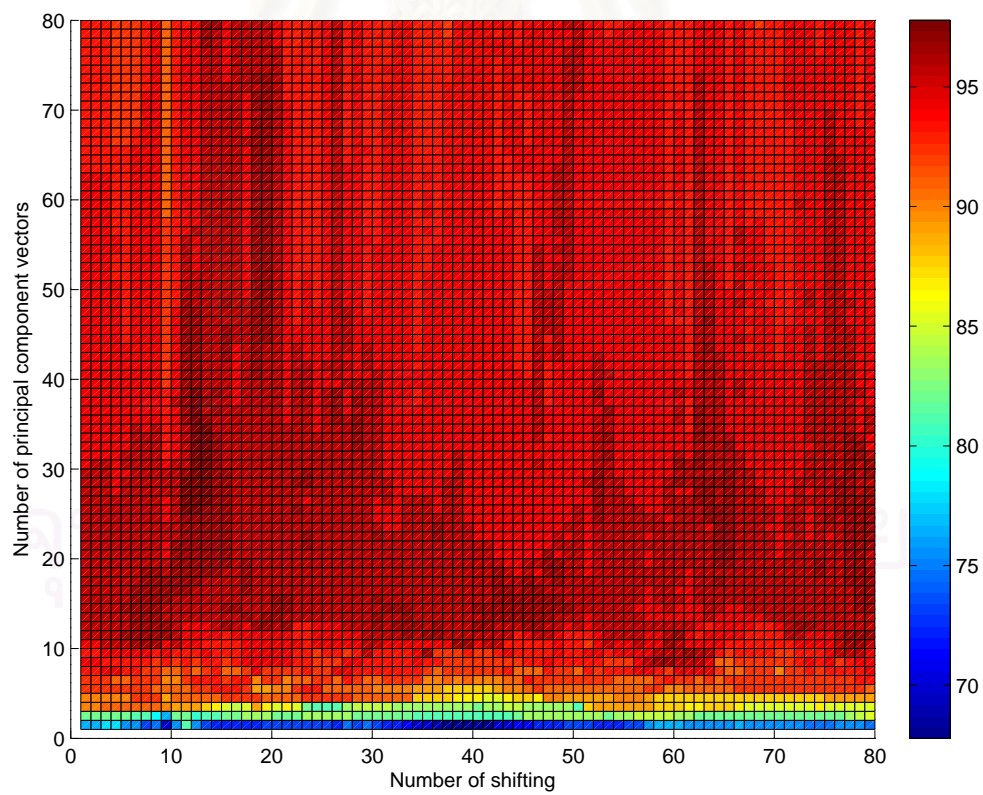


Figure 4.27 Performance map of ICCA on the Yale database.



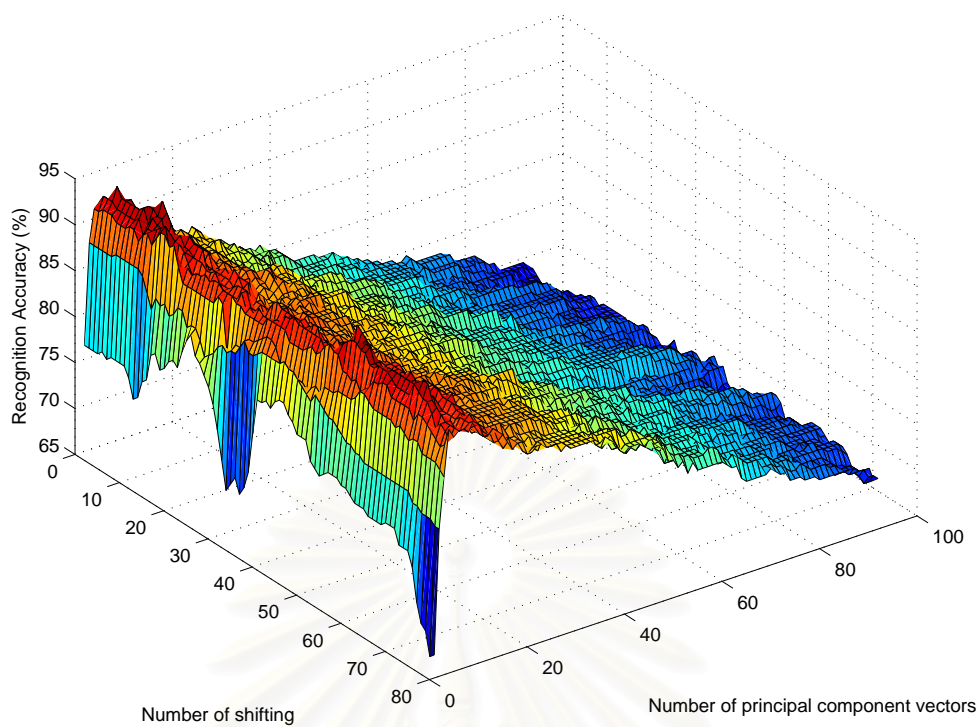


Figure 4.28 Recognition accuracy of ICCA on the ORL database.

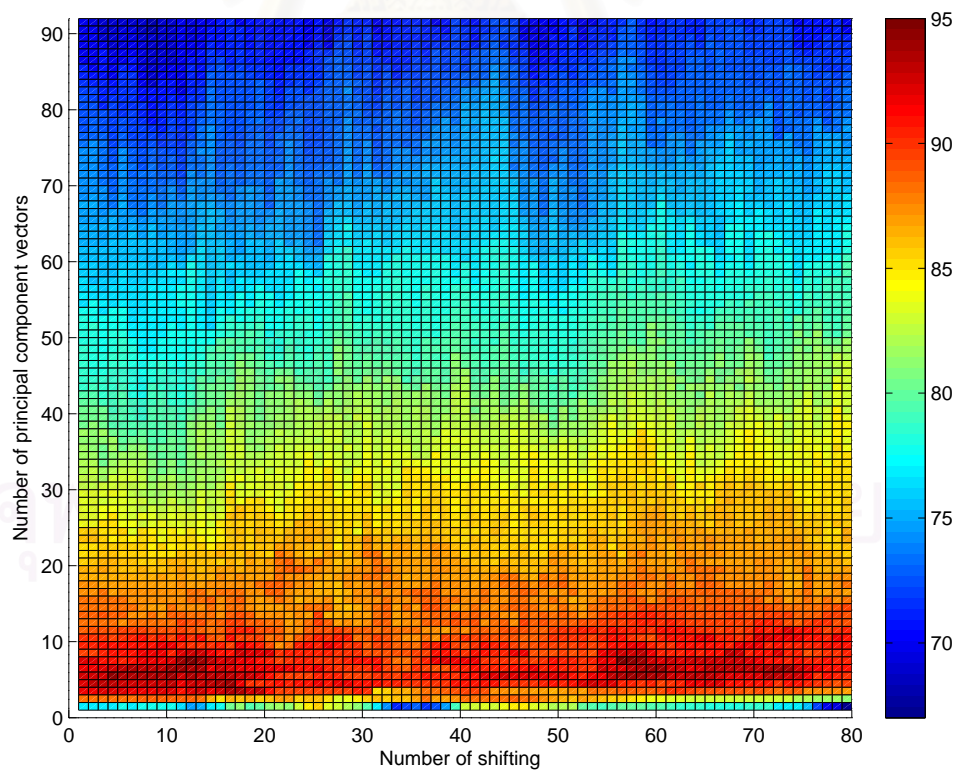


Figure 4.29 Performance map of ICCA on the ORL database.

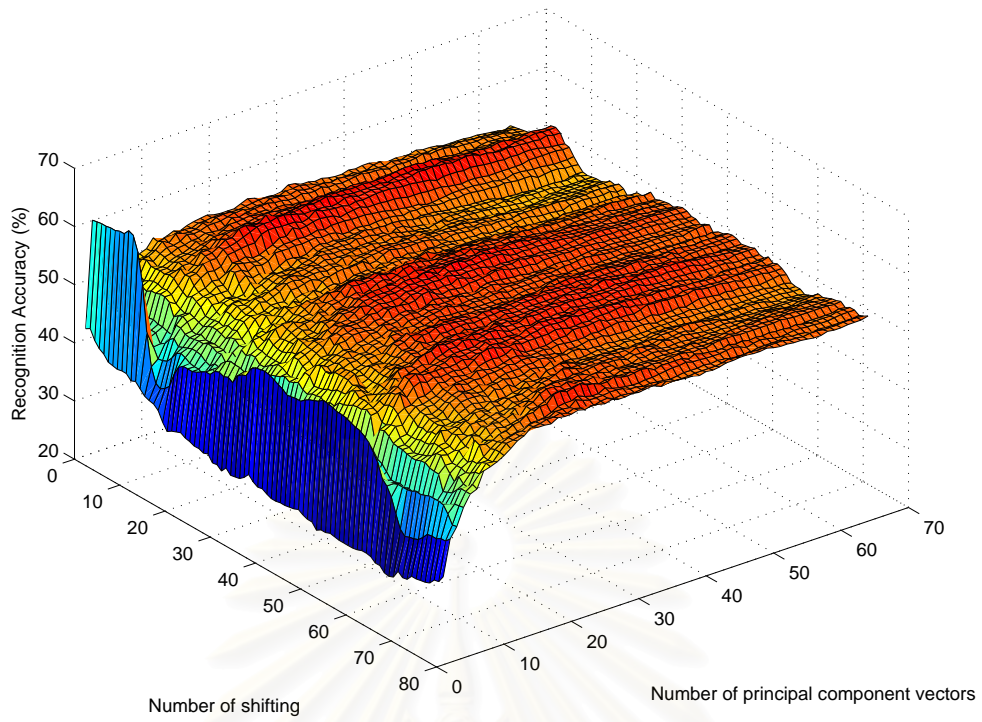


Figure 4.30 Recognition accuracy of ICCA on the AR database.

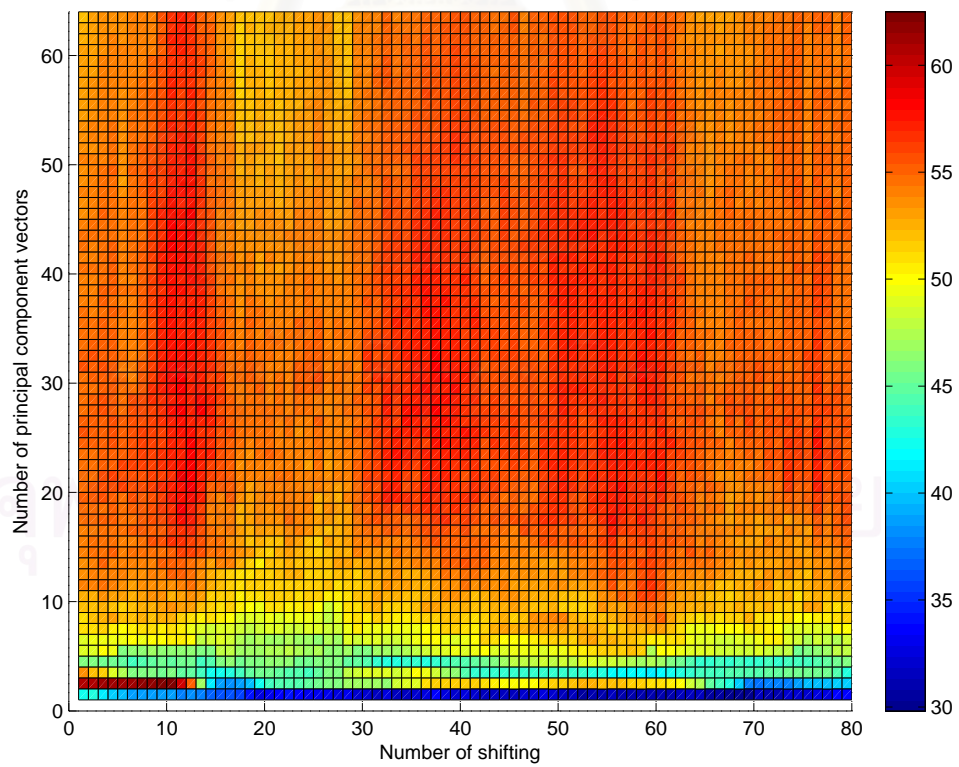


Figure 4.31 Performance map of ICCA on the AR database.

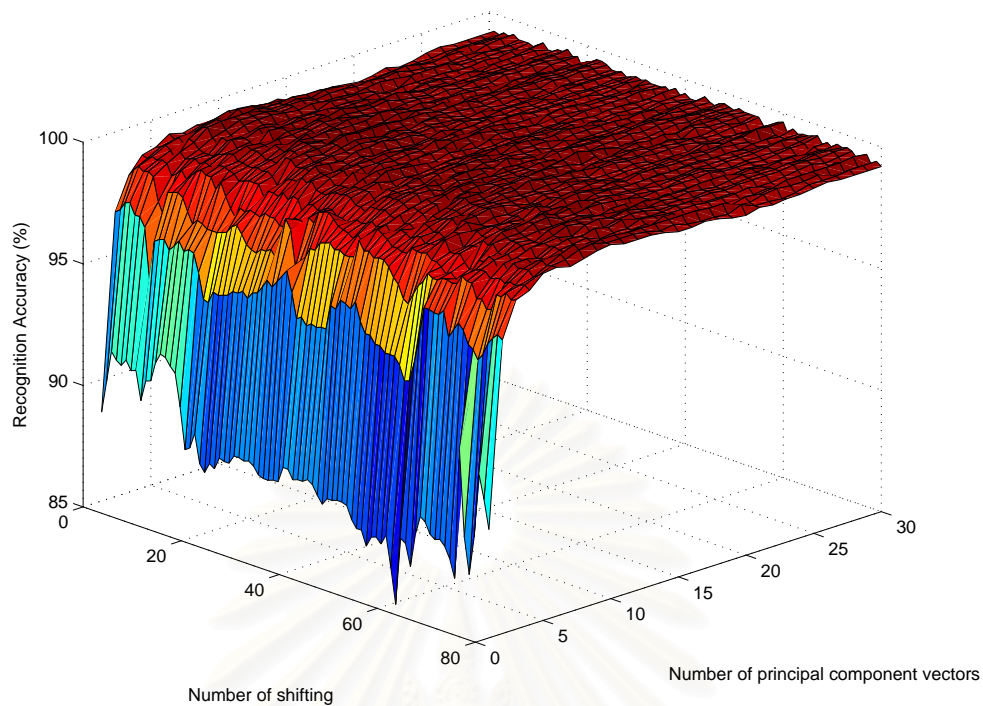


Figure 4.32 Recognition accuracy of ICCA on the MSTAR database.

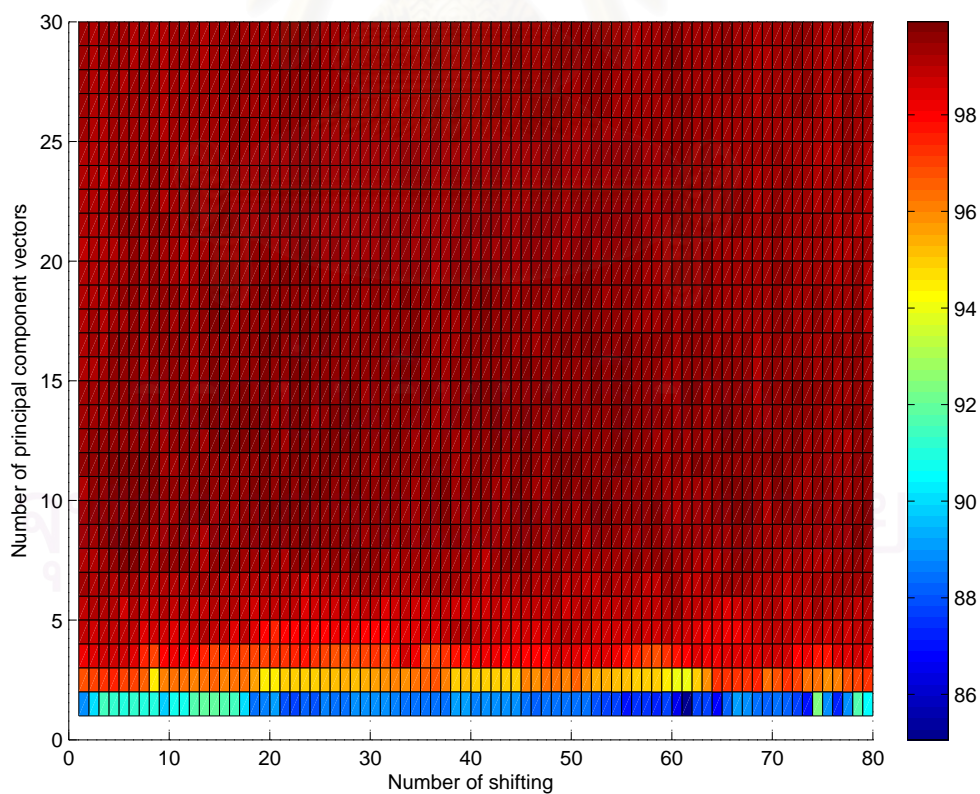


Figure 4.33 Performance map of ICCA on the MSTAR database.

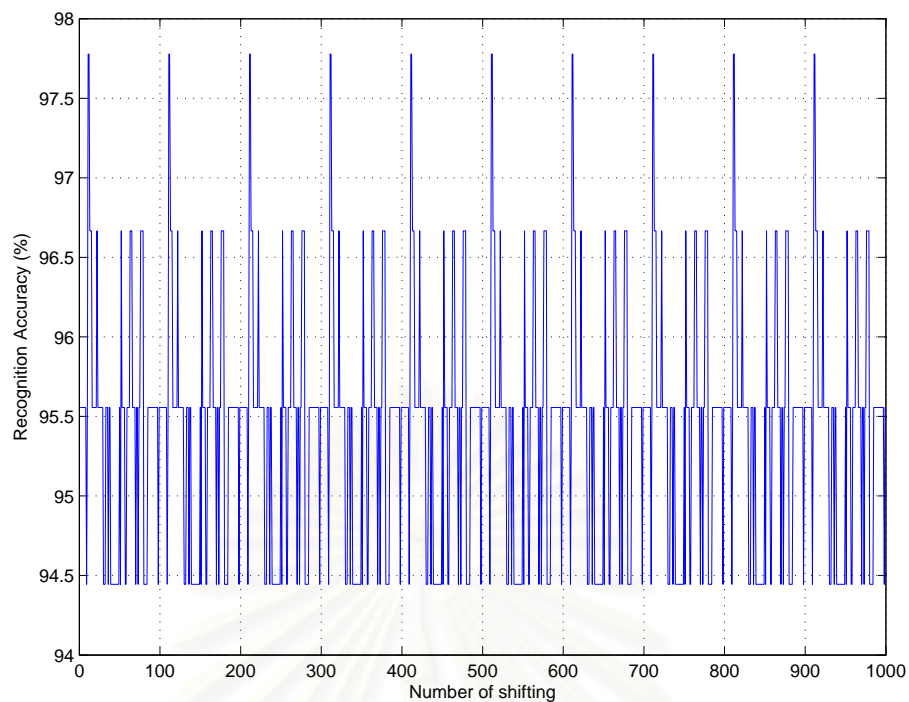


Figure 4.34: Recognition accuracy of varying the number of image shifting of ICCA on the Yale database.

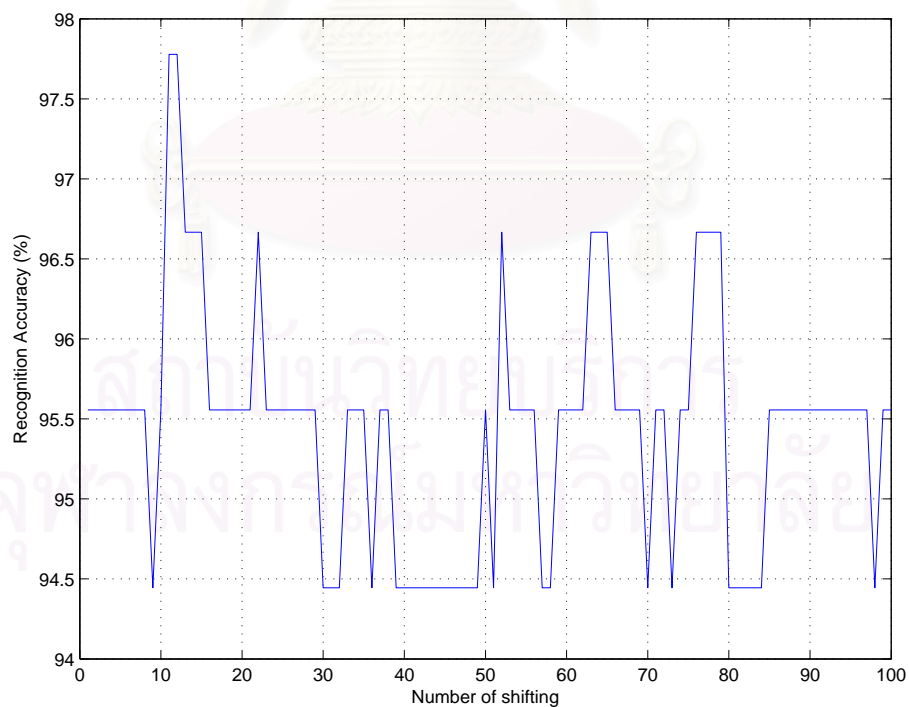


Figure 4.35: A period of varying the number of image shifting of ICCA on the Yale database.



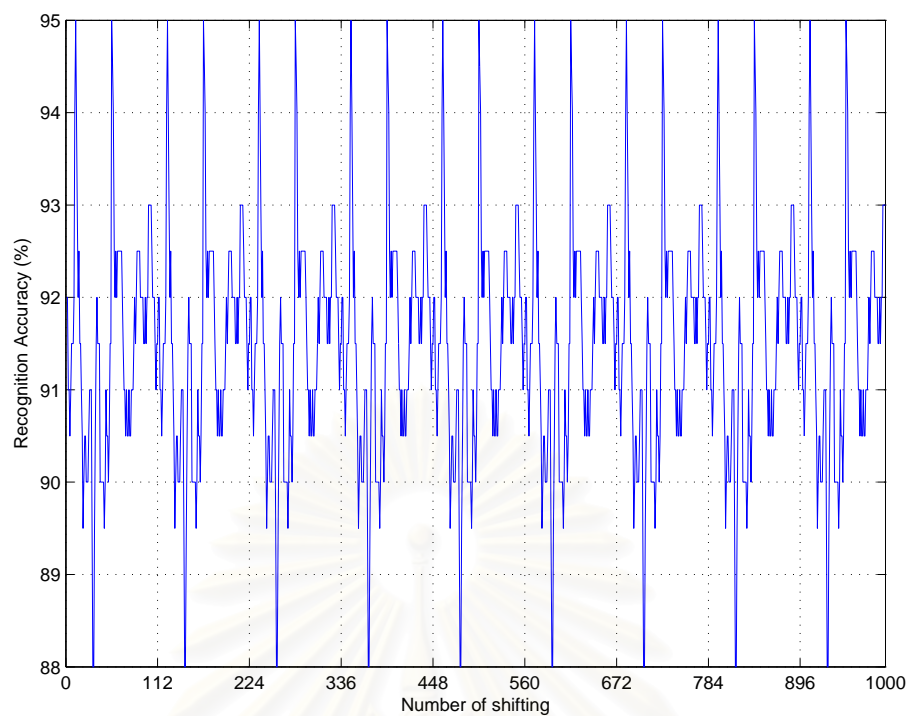


Figure 4.36: Recognition accuracy of varying the number of image shifting of ICCA on the ORL database.

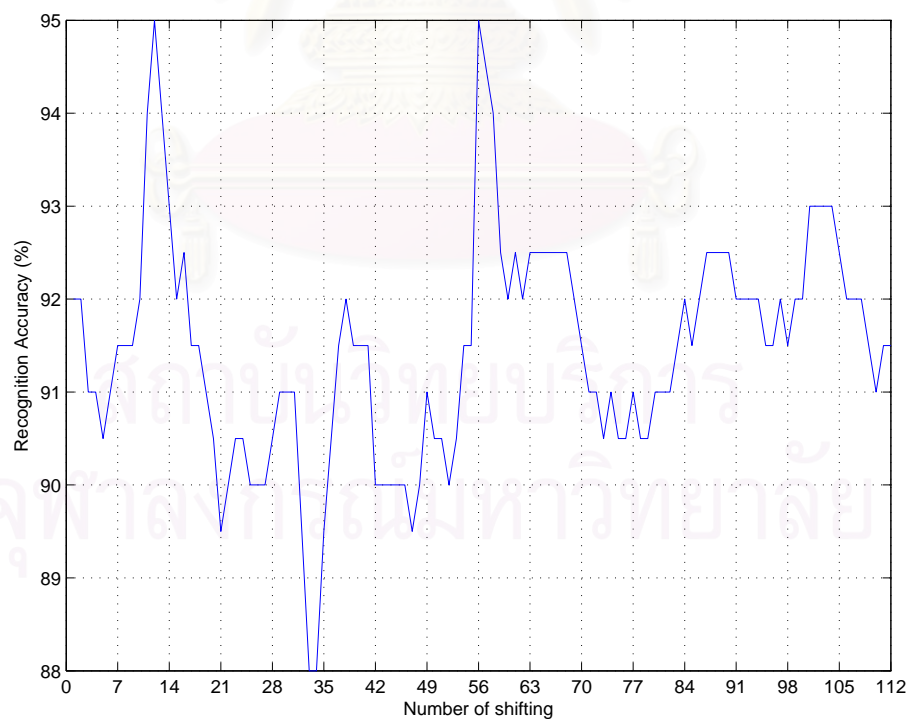


Figure 4.37: A period of varying the number of image shifting of ICCA on the ORL database.



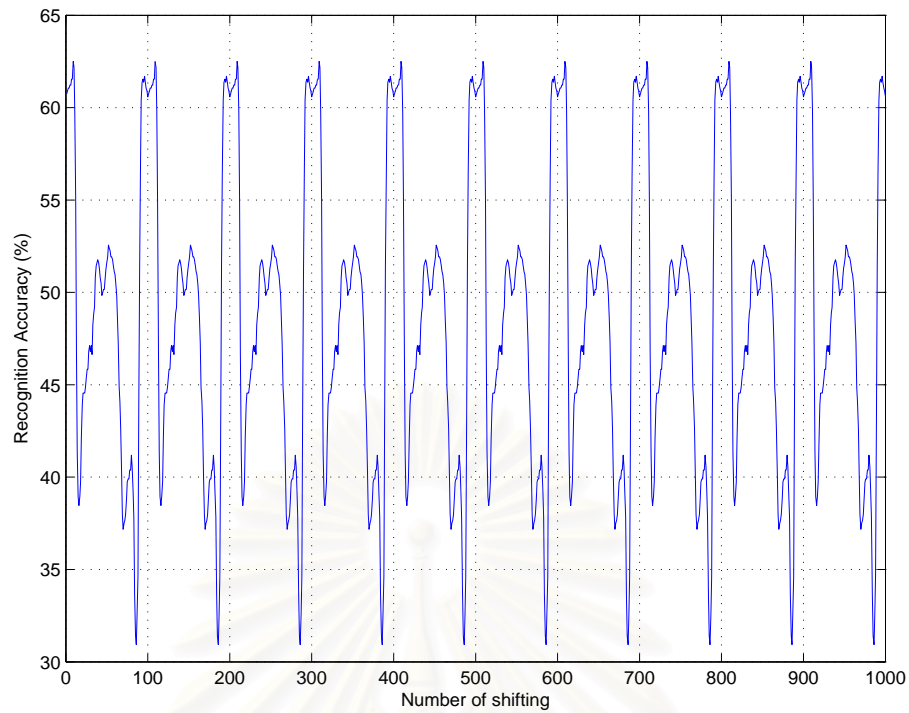


Figure 4.38: Recognition accuracy of varying the number of image shifting of ICCA on the AR database.

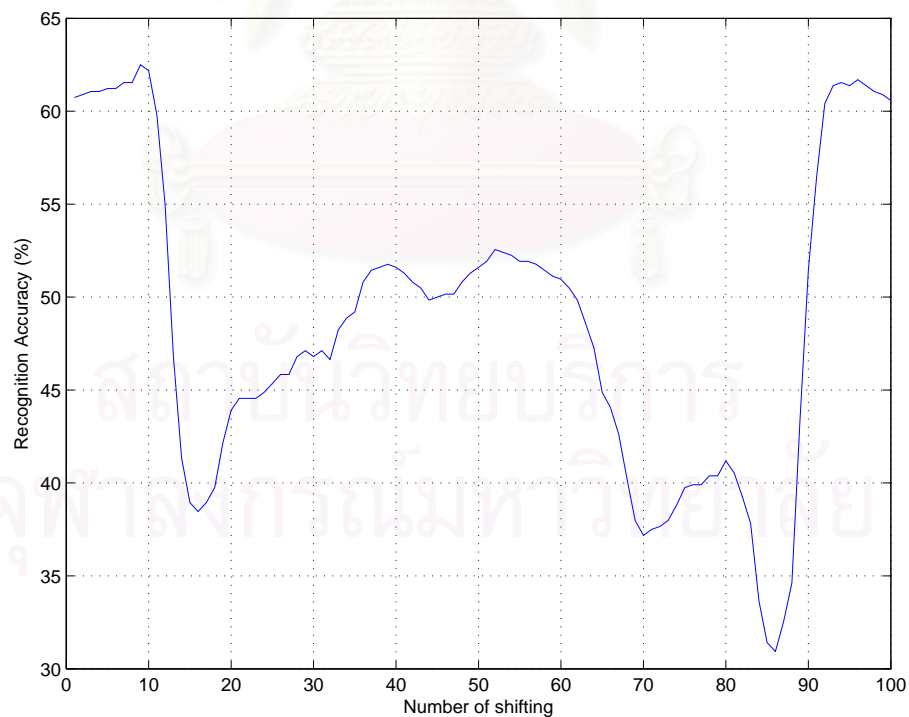


Figure 4.39: A period of varying the number of image shifting of ICCA on the AR database.

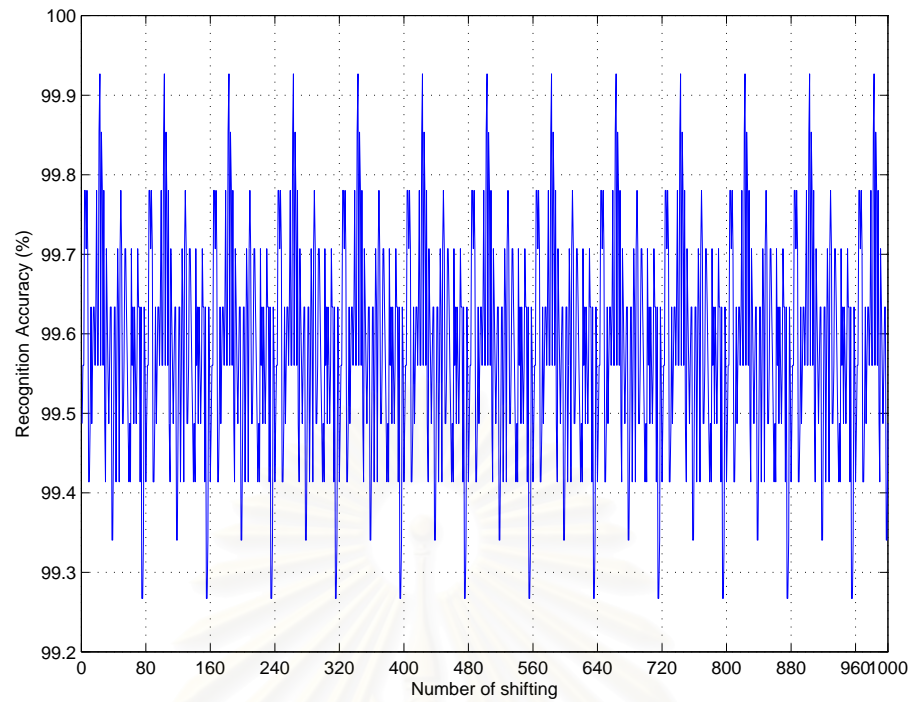


Figure 4.40: Recognition accuracy of varying the number of image shifting of ICCA on the MSTAR database.

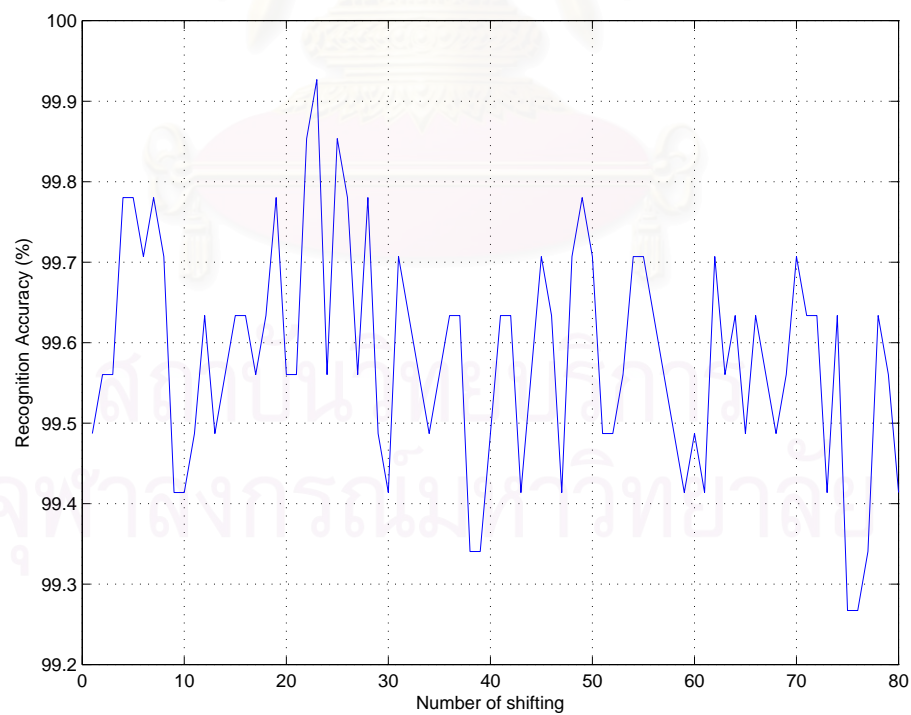


Figure 4.41: A period of varying the number of image shifting of ICCA on the MSTAR database.

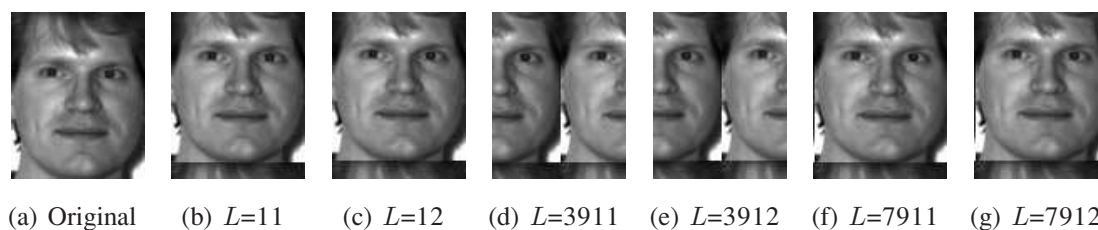


Figure 4.42: The samples of the shifted images which obtain the best recognition accuracy on the Yale database.

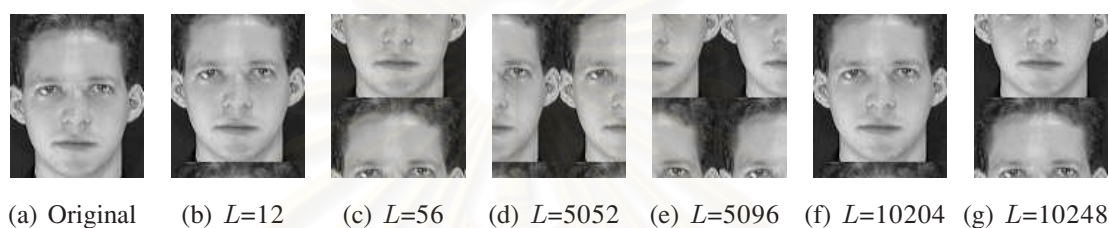


Figure 4.43: The samples of the shifted images which obtain the best recognition accuracy on the ORL database.

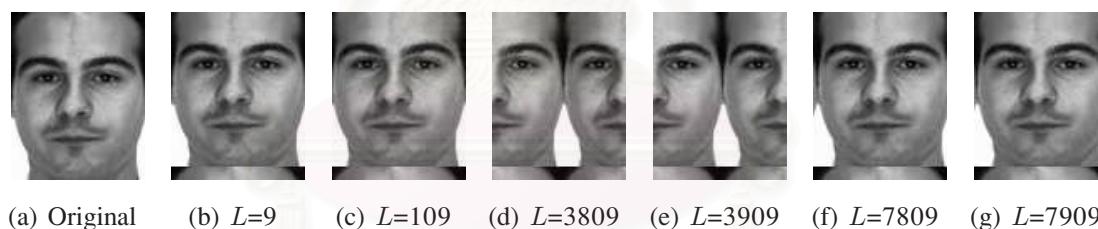


Figure 4.44: The samples of the shifted images which obtain the best recognition accuracy on the AR database.

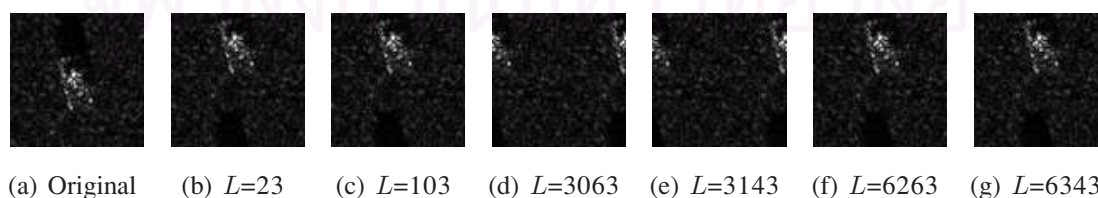


Figure 4.45: The samples of the shifted images which obtain the best recognition accuracy on the MSTAR database.

## 4.6 Experiments and Analysis on Two-Dimensional Random Subspace Analysis

In this section, we experimentally evaluate our proposed framework, 2DRSA and its extension, 2D<sup>2</sup>RSA and RSM-based ICCA. by using 2DPCA as the baseline method for comparison.

### 4.6.1 Effect of Feature Dimension of Random Subspace and Number of Classifiers

The first experiment was setup for investigating the effect of the feature dimension of random subspace and the number of classifiers.

On Yale database, the five image samples (centerlight, glasses, happy, leftlight, and noglasses) are used to train, and the six remaining images (normal, rightlight, sad, sleepy, surprised and wink) for test. For ORL and AR database, the first five images are used to train per subject, and the remaining images for test. For MSTAR database, the training and testing samples are shown in Table 4.1 and Table 4.2, respectively.

We vary the number of random selected rows of feature matrix ( $r$ ) from 1 to the height of image (100 on Yale and AR, 112 on ORL and 80 on MSTAR). The number of classifiers varies from 1 to 99 step by 2. We fix the number of principal component vectors ( $d$ ) to the value that obtains the highest recognition accuracy in 2DPCA, excepting MSTAR. The number of classifiers on MSTAR was changed only from 1 to 25, stepping by 2. If  $r$  equals to the number of pixels in the height direction of image, it is original 2DPCA. The results of these experiments on the Yale, ORL, AR, and MSTAR databases are presented in Fig. 4.46, Fig. 4.48, Fig. 4.50 and Fig. 4.52, respectively.

For investigating the optimal region, the priori performance map of these results on each database were plotted by selecting the points which obtain the high performance and shown in Fig. 4.54, Fig. 4.55, Fig. 4.56 and Fig. 4.57. The average of recognition accuracies across the number of random selected rows of feature matrix on the Yale, ORL, AR and MSTAR databases are presented in Fig. 4.58, Fig. 4.60, Fig. 4.62 and Fig. 4.64, respectively. From these figures, we found that only a little number of selected rows of feature matrix can be improved the performance. The average of recognition accuracies across the number of classifiers on the Yale, ORL, AR and MSTAR databases are presented in Fig. 4.59, Fig. 4.61, Fig. 4.63 and Fig. 4.65, respectively. From these figures, we found that the recognition accuracy will be improved when the number of classifiers are increased. And then the highest recognition accuracies of each databases is presented in Table 4.6. The average of recognition accuracies across the number of classifier and the number of random selected rows of feature matrix on the Yale, ORL, AR and MSTAR databases are presented in Fig.4.66 and Fig.4.67, respectively.

The second experiment was setup for investigation the salability of 2DRSA. A

thousand experiments were operated by randomly selecting the number of random selected rows of feature matrix and classifiers in small ranges. From the results in Fig. 4.54, Fig. 4.55, Fig. 4.56 and Fig. 4.57, we decided to vary the number of random selected rows of feature matrix ( $r$ ) from 10 to 15 and the number of classifiers from 29 to 99 step by 2 on all databases. The maximum, minimum, mean and the standard derivation are presented in Table 4.5. And the boxplot of there results were depicted in Fig. 4.68.

Our proposed method achieves the higher recognition accuracy than 2DPCA in Table 4.6. However, the complexity of 2DRSA is higher than 2DPCA because many classifiers are used to obtain the high performance.



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย



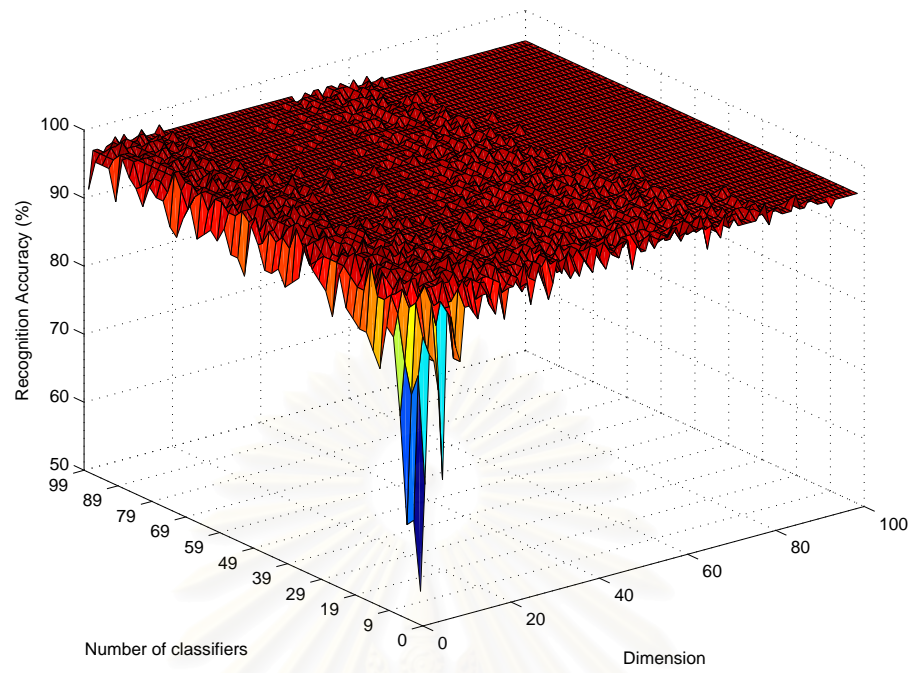


Figure 4.46 Recognition accuracy of 2DRSA on the Yale database.

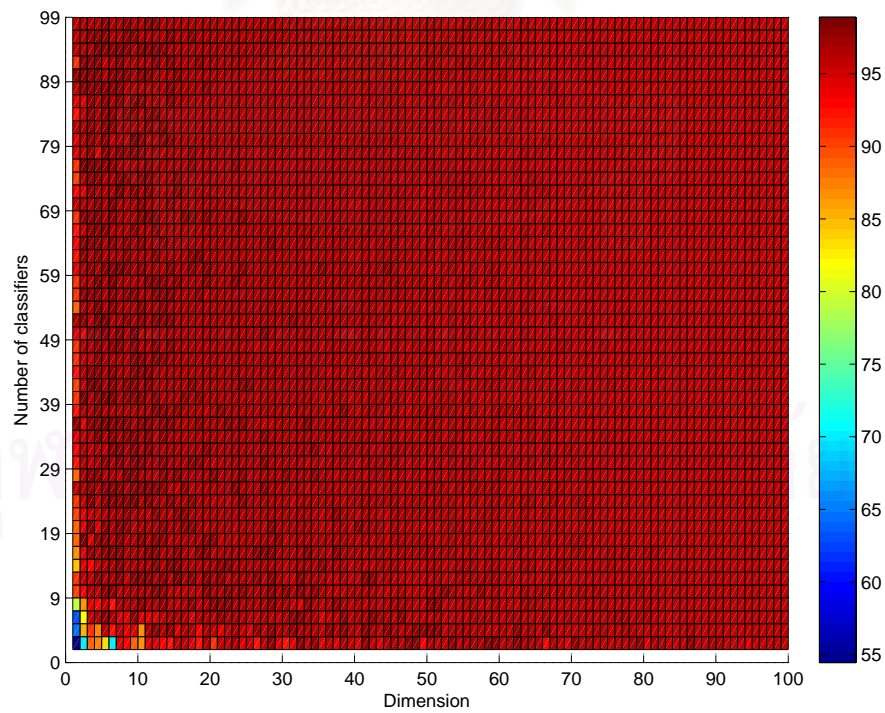


Figure 4.47 Performance map of 2DRSA on the Yale database.

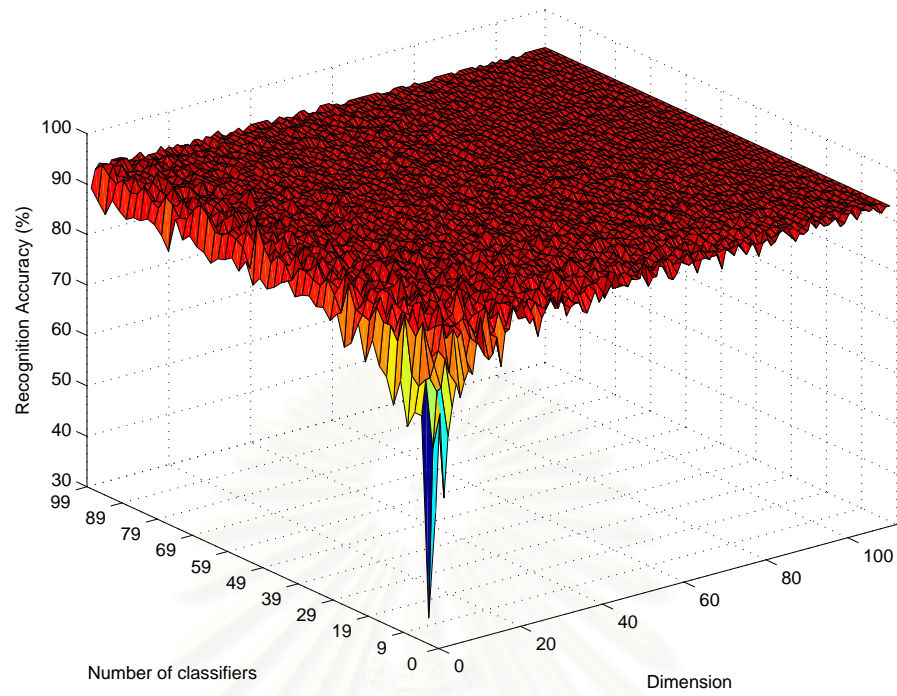


Figure 4.48 Recognition accuracy of 2DRSA on the ORL database.

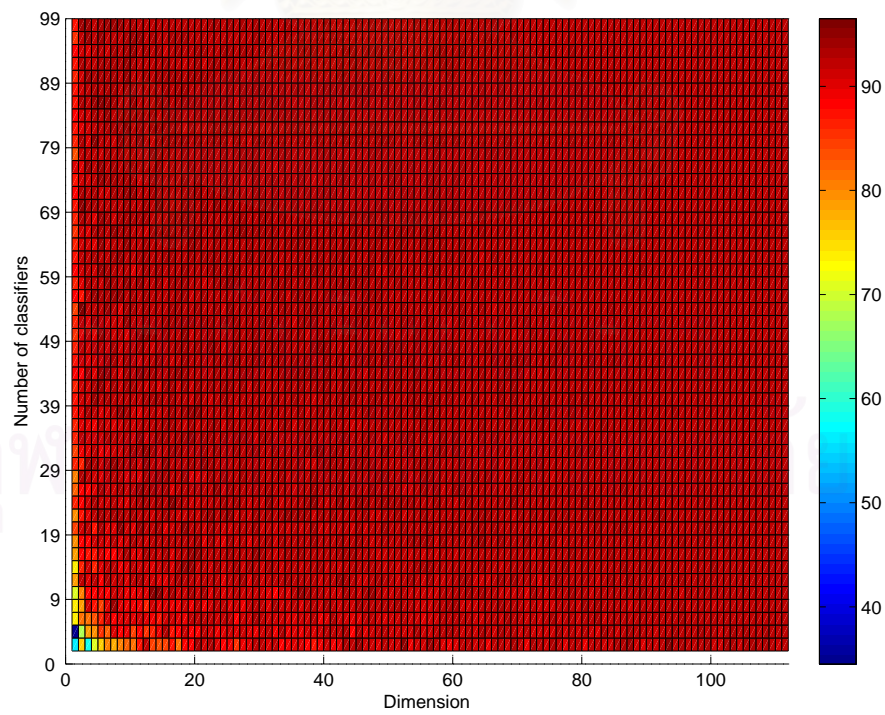


Figure 4.49 Performance map of 2DRSA on the ORL database.

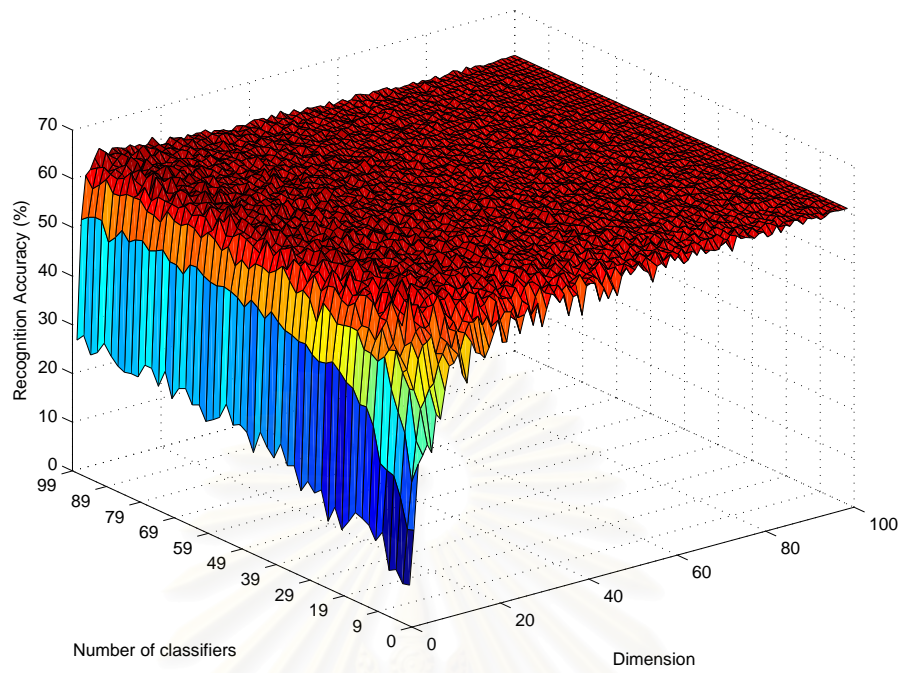


Figure 4.50 Recognition accuracy of 2DRSA on the AR database.

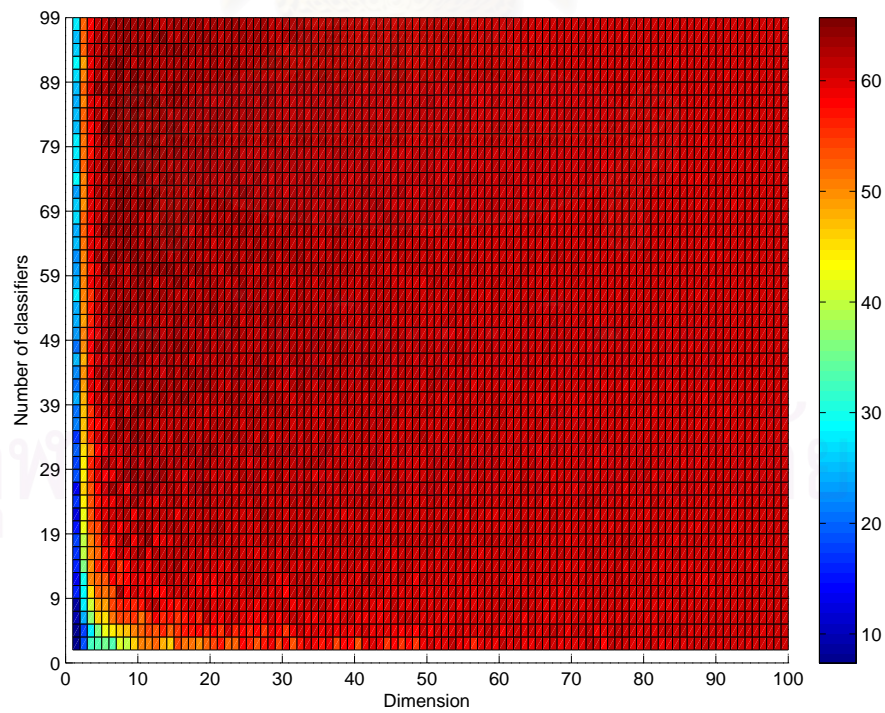


Figure 4.51 Performance map of 2DRSA on the AR database.

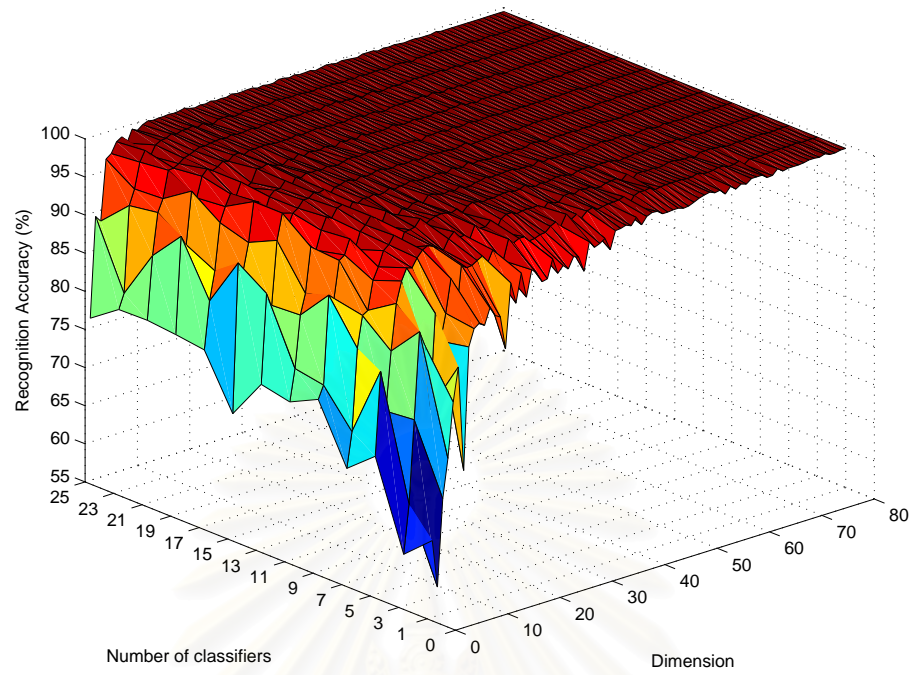


Figure 4.52 Recognition accuracy of 2DRSA on the MSTAR database.

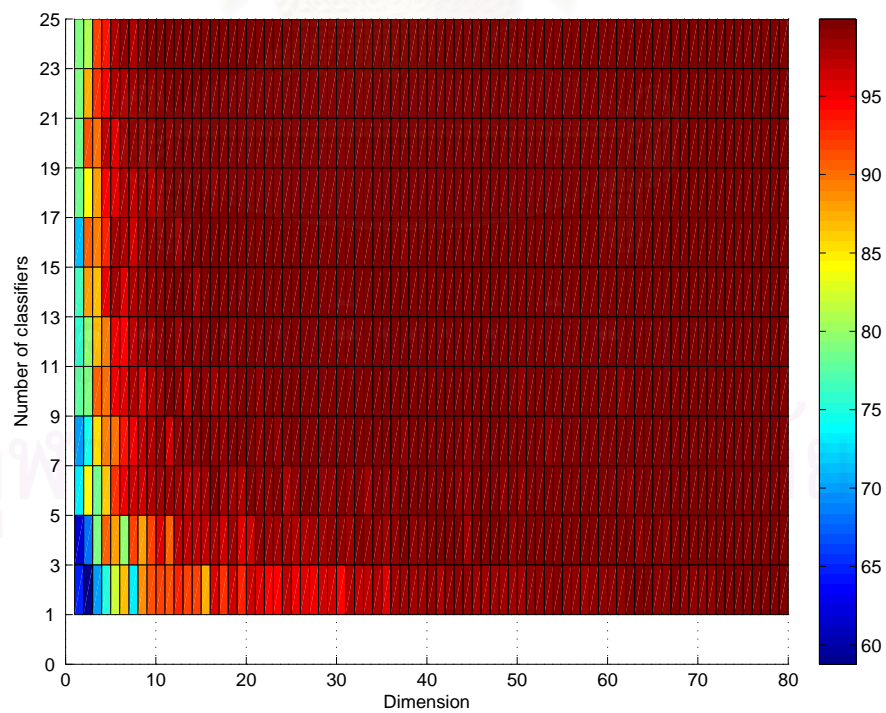


Figure 4.53 Performance map of 2DRSA on the MSTAR database.



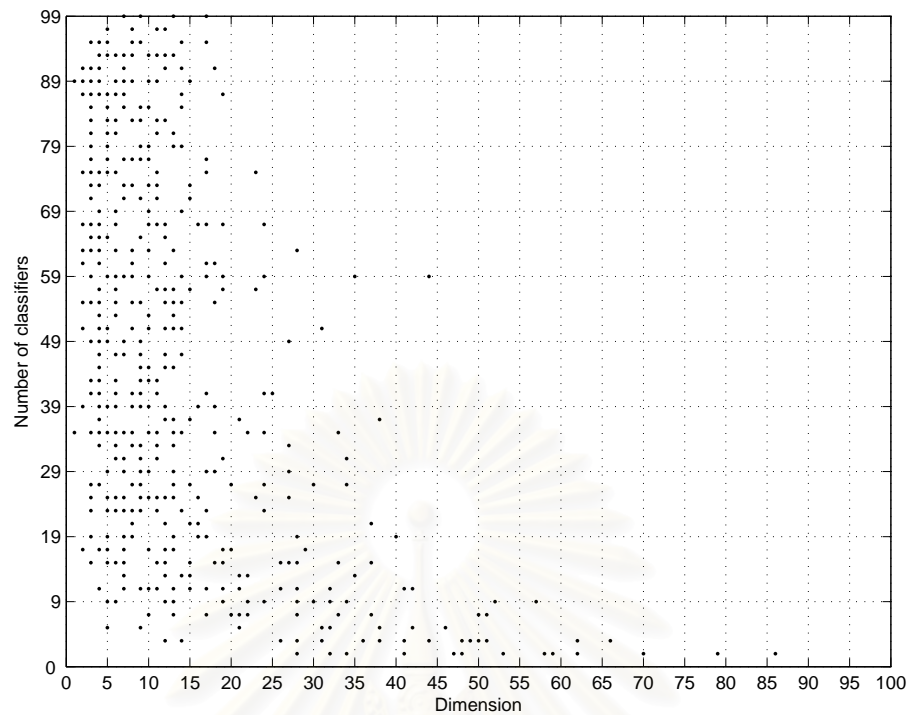


Figure 4.54 The priori performance map of 2DRSA on Yale database.

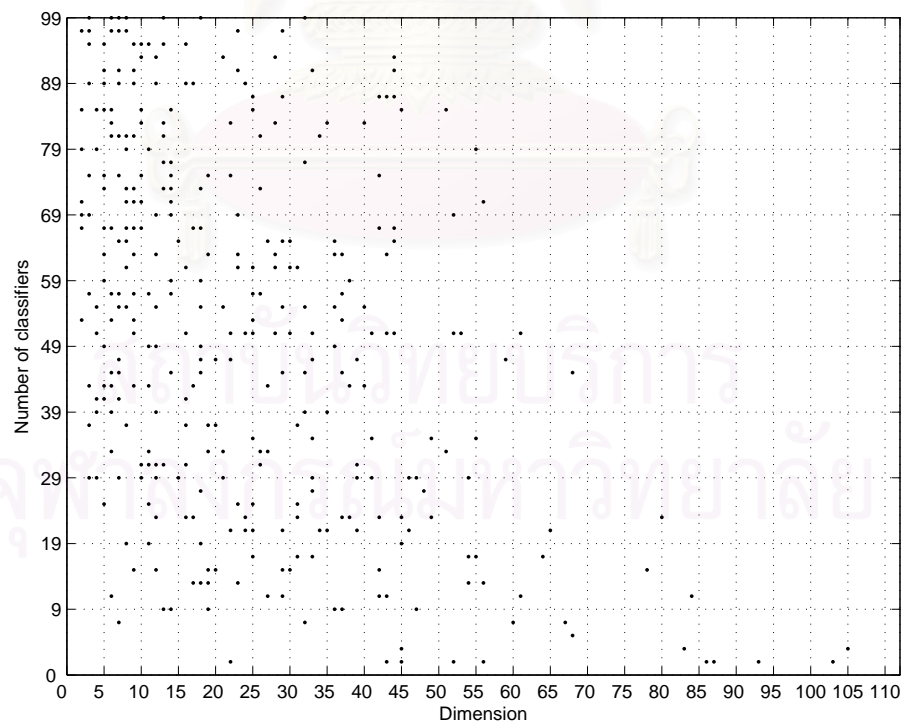


Figure 4.55 The priori performance map of 2DRSA on ORL database.



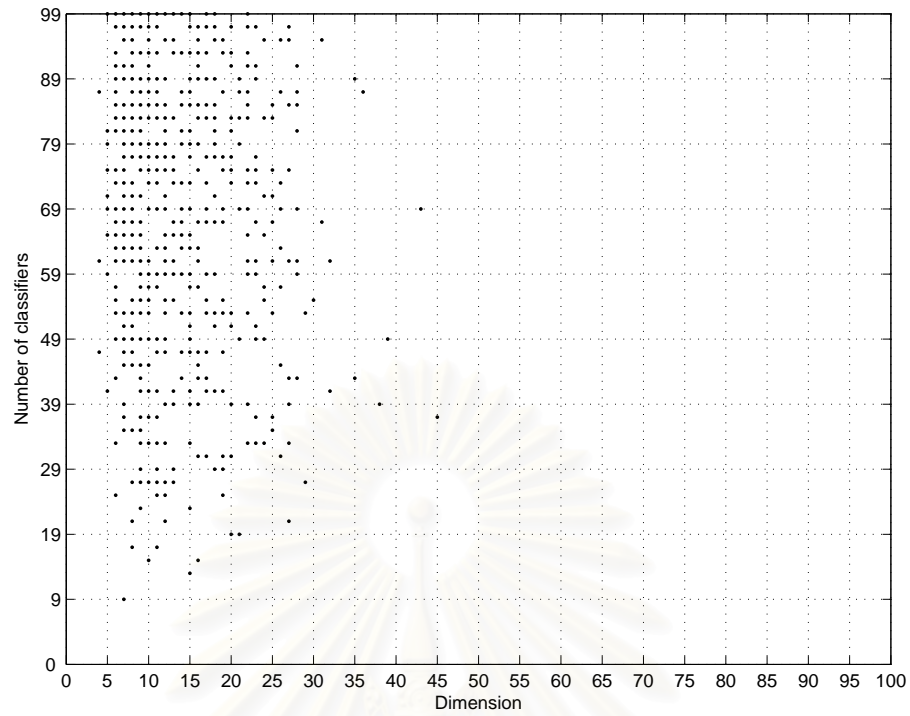


Figure 4.56 The priori performance map of 2DRSA on AR database.

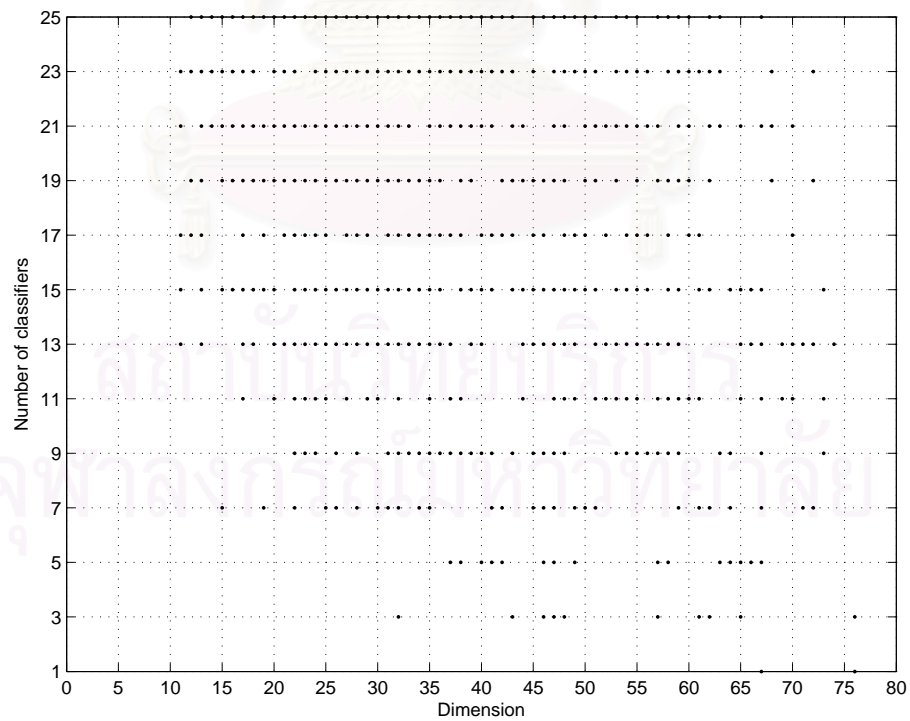


Figure 4.57 The priori performance map of 2DRSA on MSTAR database.

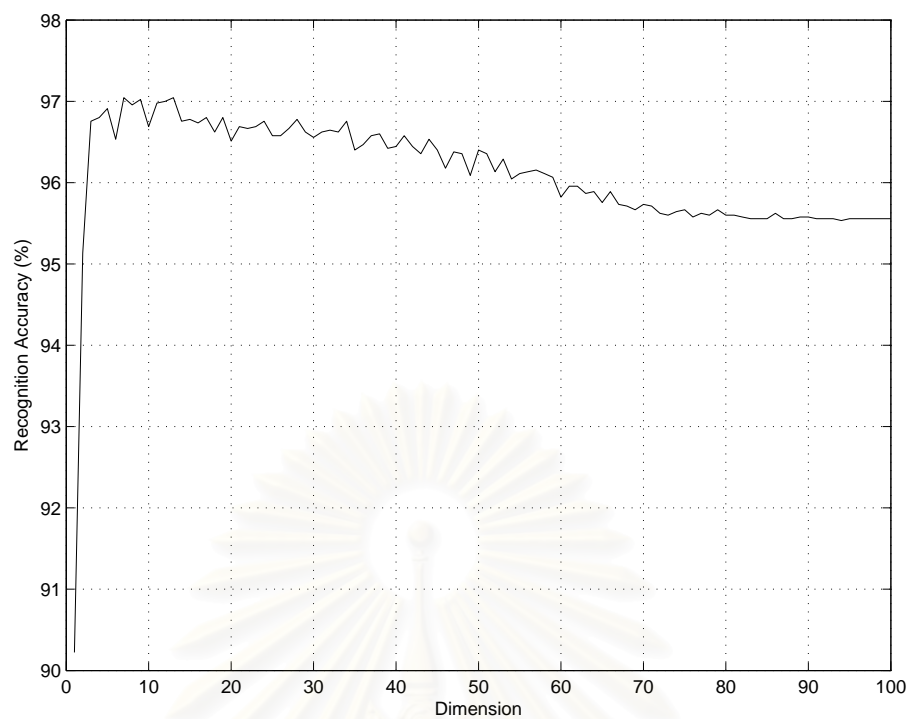


Figure 4.58: Averaging results of 2DRSA by varying the feature dimensions on Yale database.

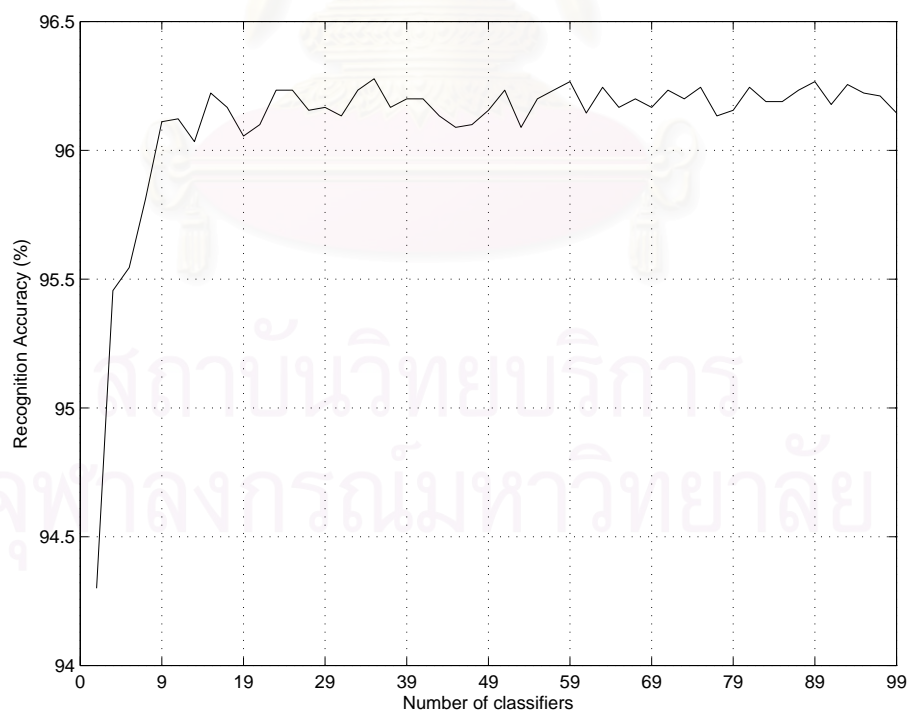


Figure 4.59: Averaging results of 2DRSA by varying the number of classifiers on Yale database.

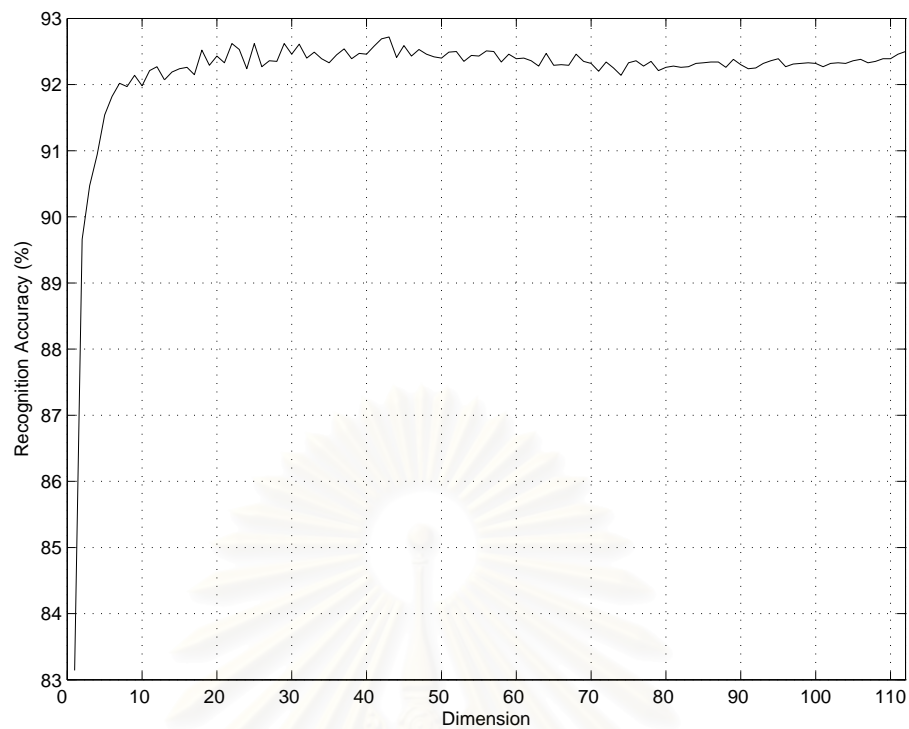


Figure 4.60: Averaging results of 2DRSA by varying the feature dimensions on ORL database.

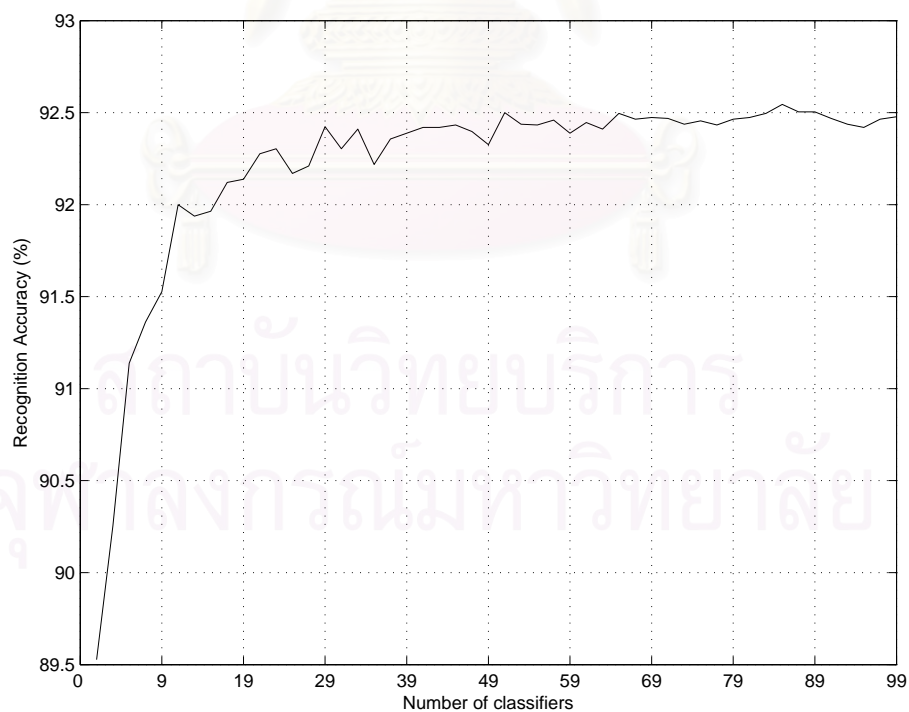


Figure 4.61: Averaging results of 2DRSA by varying the number of classifiers on ORL database.

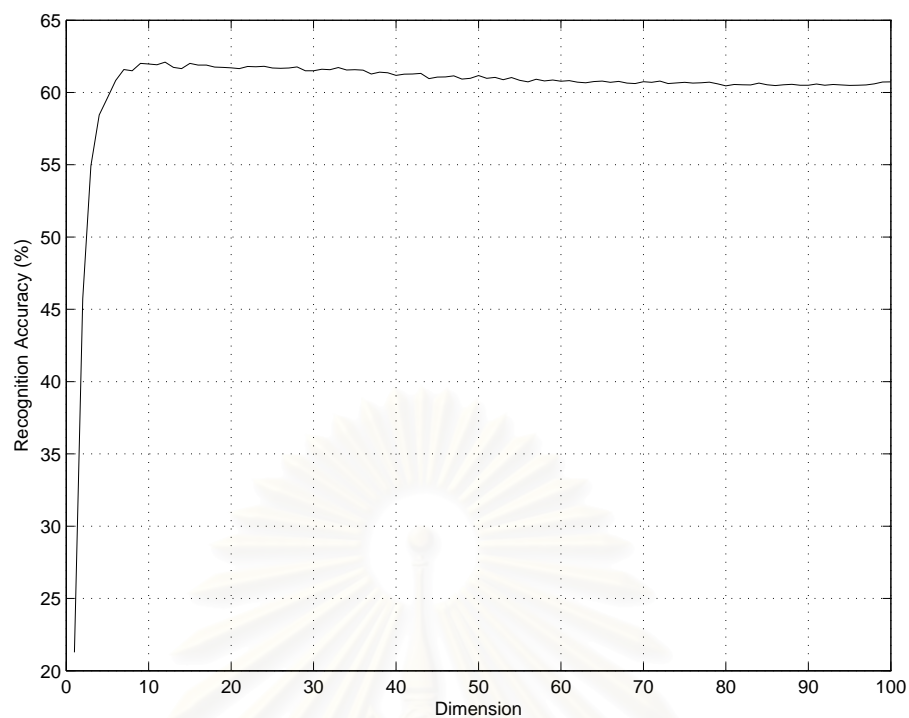


Figure 4.62: Averaging results of 2DRSA by varying the feature dimensions on AR database.

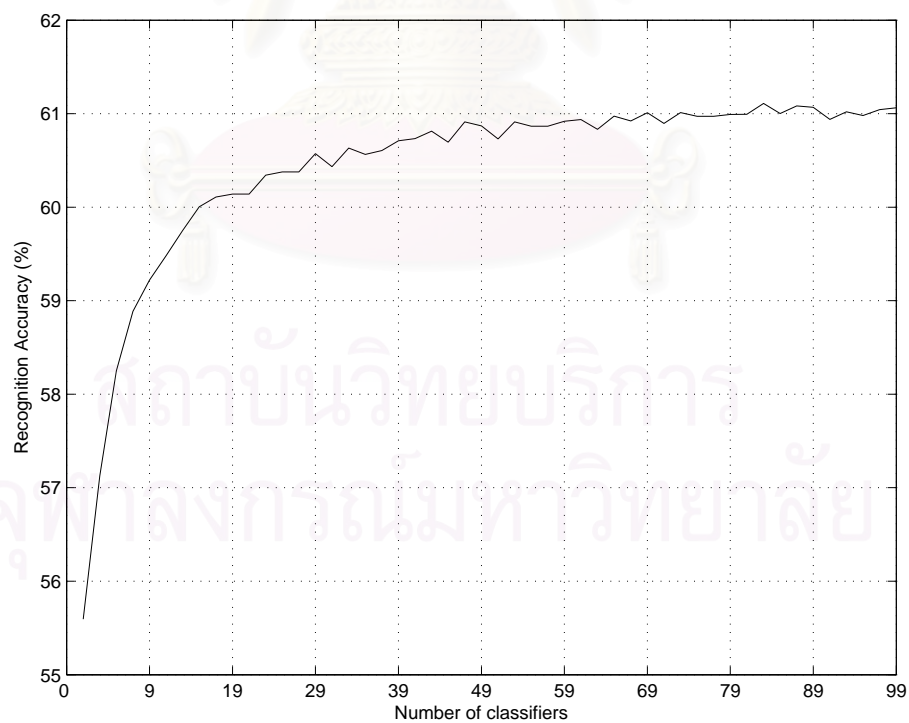


Figure 4.63: Averaging results of 2DRSA by varying the number of classifiers on AR database.

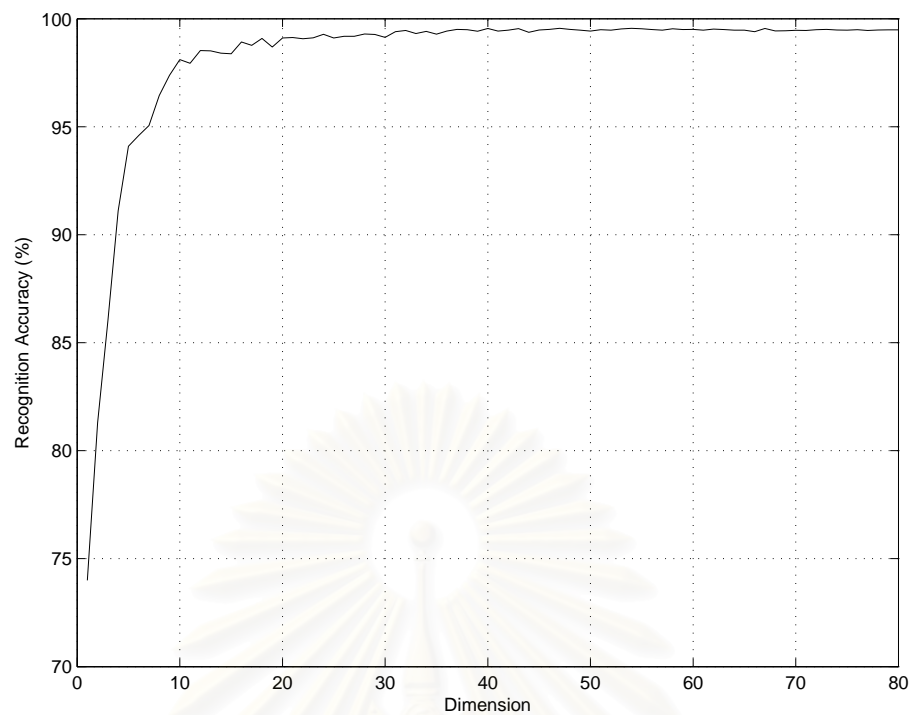


Figure 4.64: Averaging results of 2DRSA by varying the feature dimensions on MSTAR database.

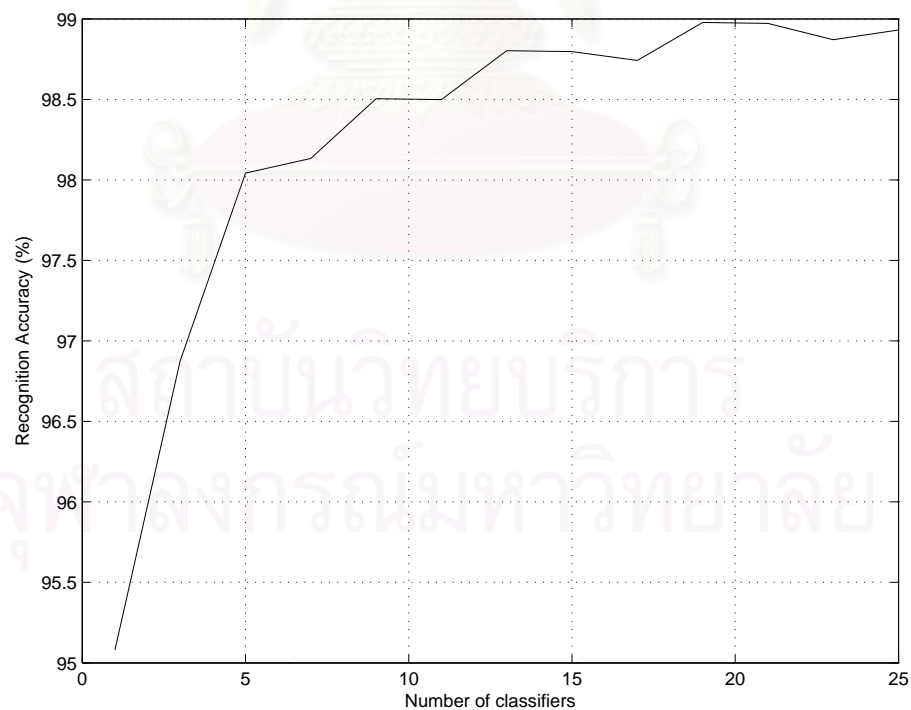


Figure 4.65: Averaging results of 2DRSA by varying the number of classifiers on MSTAR database.



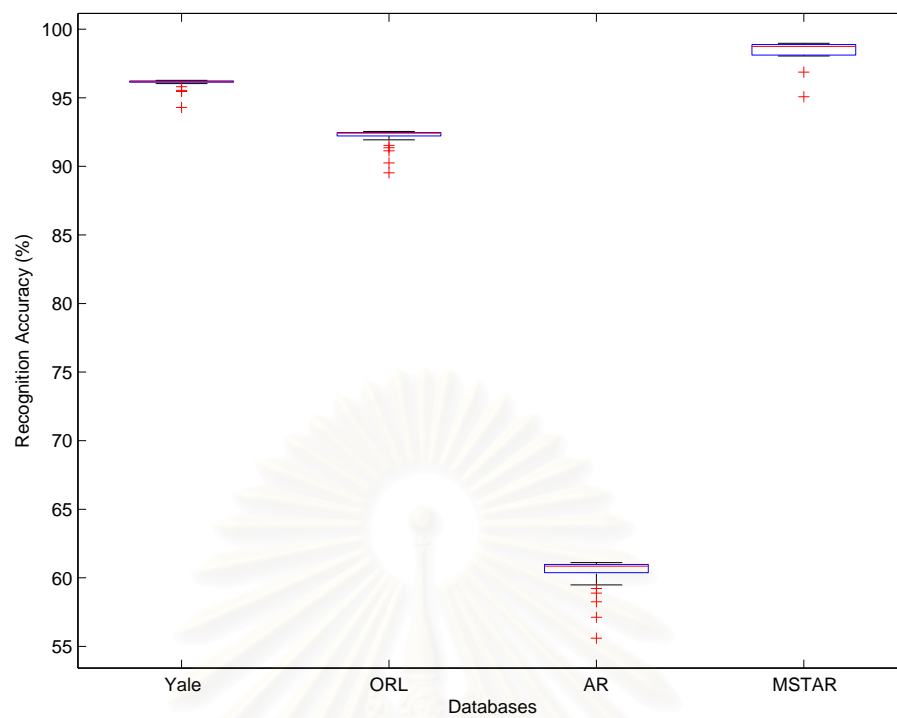


Figure 4.66 Boxplot of 2DRSA in term of the number of classifiers.

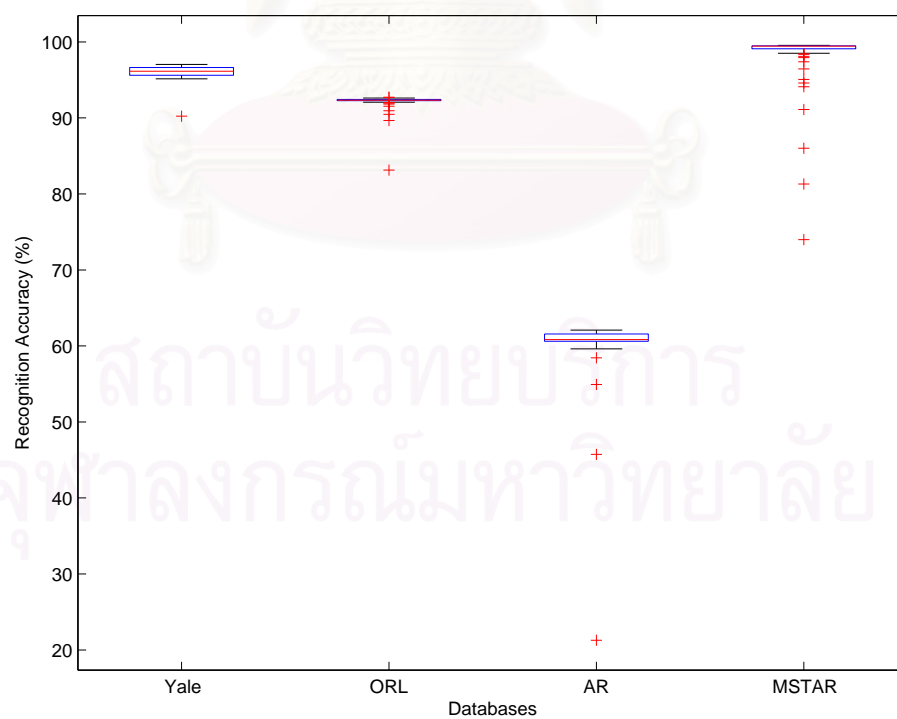


Figure 4.67 Boxplot of 2DRSA in term of the feature dimensions.

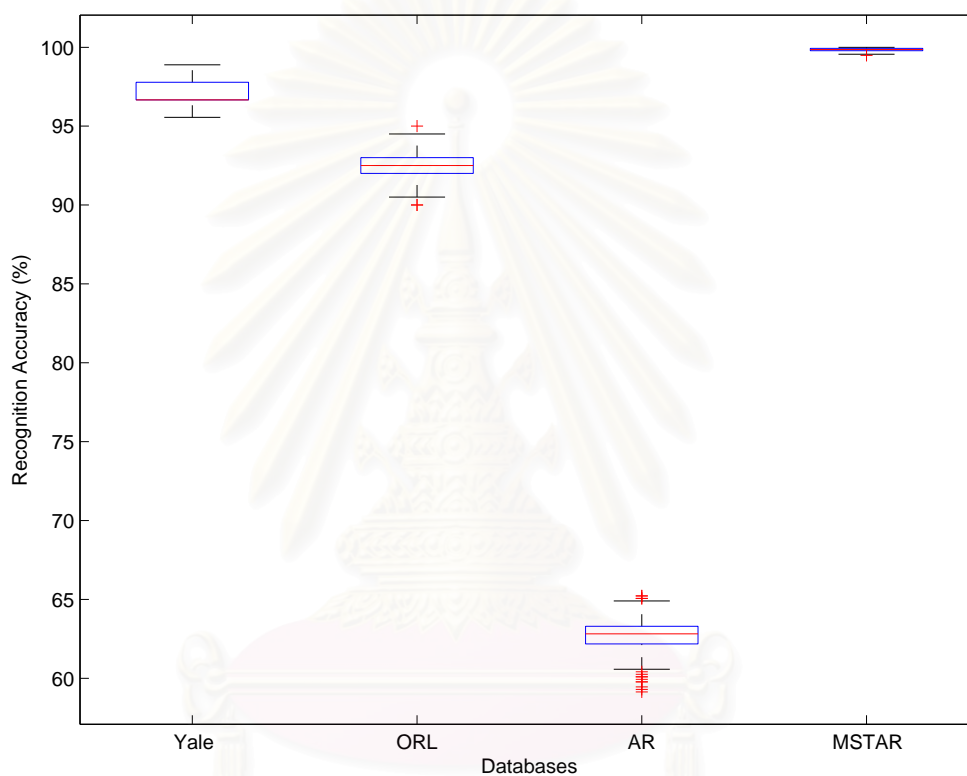


Figure 4.68: Boxplot of 2DRSA in term of the number of classifiers and the feature dimensions in prospected ranges.

Table 4.5: The Data Statistics of Results comparisons of 2DRSA Method on Yale, ORL, AR, and MSTAR Databases

Database	Data statistics		#Classifiers	Dimension
Yale	Max	98.8889	31	$12 \times 20$
	Mean	97.0244	-	-
	Min	95.5556	31	$10 \times 20$
	Std	0.5427	-	-
ORL	Max	95	31	$14 \times 5$
	Mean	92.7345	-	-
	Min	90	29	$15 \times 5$
	Std	0.787	-	-
AR	Max	65.2244	29	$14 \times 2$
	Mean	62.745	-	-
	Min	59.1346	45	$13 \times 2$
	Std	0.9326	-	-
MSTAR	Max	100	45	$12 \times 11$
	Mean	99.8228	-	-
	Min	99.4872	77	$11 \times 11$
	Std	0.1055	-	-

Table 4.6: The Highest Recognition Accuracy Comparisons of 2DPCA and 2DRSA Method on Yale, ORL, AR, and MSTAR Databases

Database	Method	Accuracy	$d$	$L$	Dimension
Yale	2DPCA	95.56 %	11	1	$100 \times 11$
	2DRSA	98.89 %	20	11	$12 \times 20$
ORL	2DPCA	92.50 %	5	1	$112 \times 5$
	2DRSA	95.00 %	5	12	$14 \times 5$
AR	2DPCA	60.74 %	2	1	$100 \times 2$
	2DRSA	65.22 %	2	9	$14 \times 2$
MSTAR	2DPCA	99.49 %	11	1	$80 \times 11$
	2DRSA	100.00 %	11	23	$12 \times 11$

#### 4.6.2 The experiment results of Two-Dimensional Diagonal Random Subspace Analysis

The first experiment, we compare the performance of 2DPCA and DiaPCA when the number of principal component vectors is varying on Yale database. The result is plotted in Fig. 4.69, DiaPCA can achieve the highest recognition accuracy by using the number of principal component vectors lower than 2DPCA. The second experiment, we fixed the number of principal component vectors as the value that obtain the best performance in the previous experiment. And the dimension of random subspace is varying from 1 to the height of the image (100). If this value equals to 1, it means normal DiaPCA. For the number of classifiers, we used only the odd number form 1 to 99. The results of 2D<sup>2</sup>RSA on Yale database are presented in Fig. 4.70. The best of recognition accuracy of these methods are compared in Table 4.7, where  $d$  is the number of principal component vectors,  $r$  is the dimension of random subspace and  $C$  is the number of classifiers. Our proposed method achieves the higher recognition accuracy than 2DPCA and DiaPCA. However, the complexity of our proposed methods depends on the number of classifiers and the dimension of random subspace, thus it usually be higher than 2DPCA.



Table 4.7: The Highest Recognition Accuracy Comparisons of 2DPCA and 2D<sup>2</sup>RSA on Yale databases

Database	Method	Accuracy	$d$	$r$	$C$
Yale	2DPCA	95.56 %	11	100	1
	DiaPCA	95.56 %	10	100	1
	2D <sup>2</sup> RSA	97.78 %	10	4	9

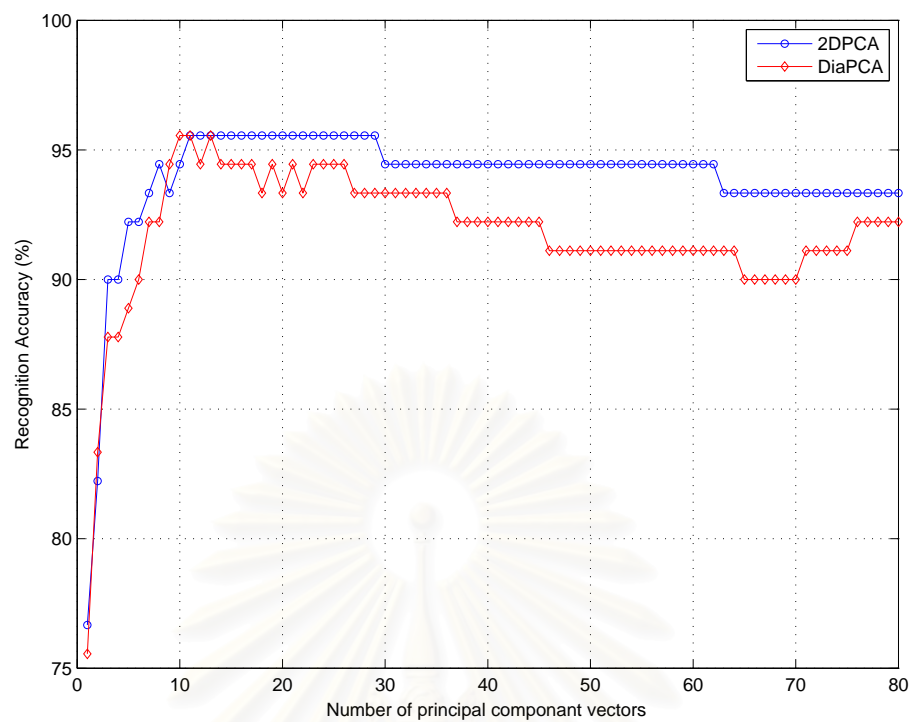


Figure 4.69 Recognition accuracy of DiaPCA and 2DPCA on Yale database.

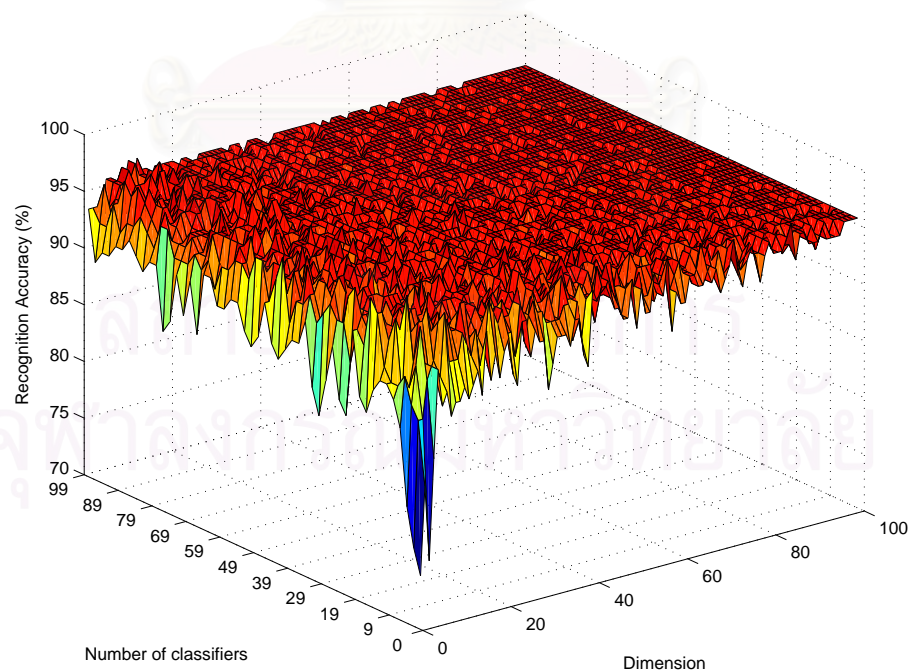


Figure 4.70 Recognition accuracy of 2D<sup>2</sup>RSA on Yale database.



### 4.6.3 The experiment results of RSM-based ICCA

In this experiment, the performance of RSM-based ICCA was investigated. For comparison with 2DRSA, we vary values in the same intervals as used in the experiments on 2DRSA, as following: The number of classifiers was randomly selected from the odd numbers between 29 to 99. The dimension of the random subspaces was randomly selected between 10 and 15. And the five value of the number of the image shifting ( $L$ ) was randomly selected from the first period in the previous experiments.

A thousand of experiments, the parameters were randomly selected in above intervals in each experiments, were performed on Yale, ORL, AR, and MSTAR databases. The maximum, mean, minimum, and standard derivation of the results are presented in Table 4.8 and Fig. 4.71. And the highest recognition accuracy of RSM-based ICCA are compared with 2DPCA and ICCA in Table 4.9.

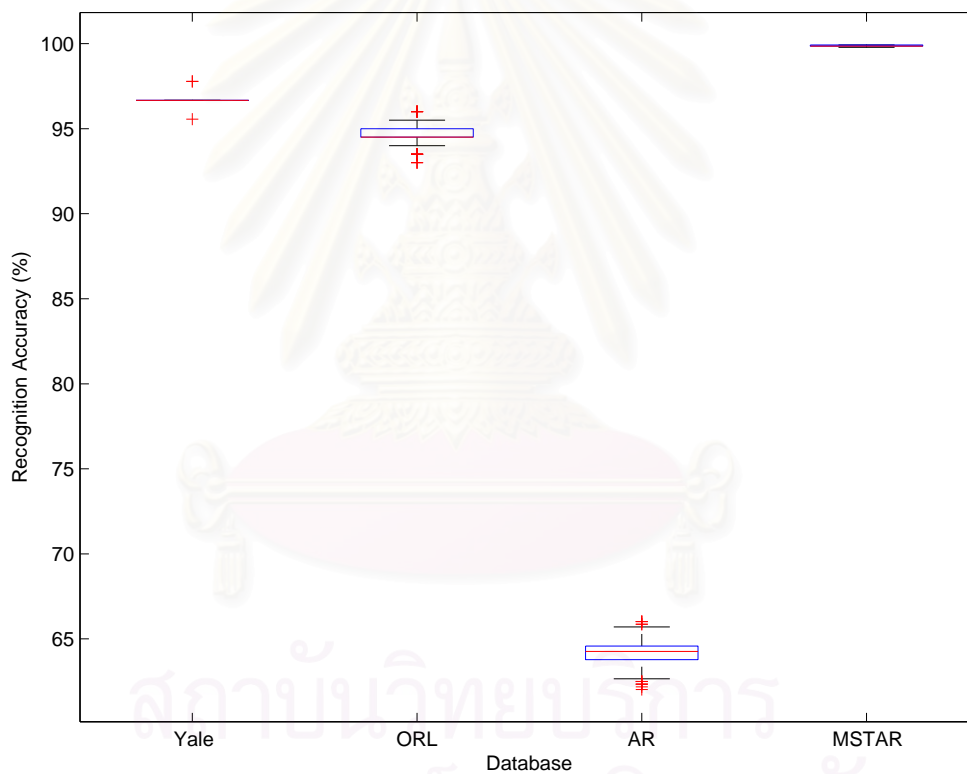


Figure 4.71 Boxplot of RSM-based ICCA on Yale, ORL, AR, and MSTAR databases.

Table 4.8: The Data Statistics of Results of RSM-based ICCA Method Comparisons on Yale, ORL, AR, and MSTAR Databases

Database	Data statistics		#Classifiers	Dimension	#L
Yale	Max	97.7778	29	11 × 20	5
	Mean	96.6833	-	-	5
	Min	95.5556	39	11 × 20	5
	Std	0.144	-	-	5
ORL	Max	96	29	11 × 5	5
	Mean	94.5865	-	-	5
	Min	93	29	12 × 5	5
	Std	0.5348	-	-	5
AR	Max	66.0256	41	10 × 2	5
	Mean	64.1951	-	-	5
	Min	62.0192	69	15 × 2	5
	Std	0.6302	-	-	5
MSTAR	Max	99.9267	59	13 × 11	5
	Mean	99.8557	-	-	5
	Min	99.7802	33	12 × 11	5
	Std	0.0514	-	-	5

Table 4.9: The Highest Recognition Accuracy Comparisons of 2DPCA and ICCA Method on Yale, ORL, AR, and MSTAR Databases

Database	Method	Accuracy	$d$	$L$	Dimension
Yale	2DPCA	95.56 %	11	1	100 × 11
	ICCA	97.78 %	25	11	100 × 25
	ICCA+RSM	97.78 %	20	-	11 × 20
ORL	2DPCA	92.50 %	5	1	112 × 5
	ICCA	95.00 %	7	12	112 × 7
	ICCA+RSM	96.00 %	5	-	11 × 5
AR	2DPCA	60.74 %	2	1	100 × 2
	ICCA	62.50 %	2	9	100 × 2
	ICCA+RSM	66.03 %	2	-	10 × 2
MSTAR	2DPCA	99.49 %	11	1	80 × 11
	ICCA	99.93 %	11	23	80 × 11
	ICCA+RSM	99.93 %	11	-	13 × 11

## 4.7 Summary

In this section, all our proposed methods are compared with the original 2DPCA here. The proposed methods are listed as following,

- Two-Dimensional Linear Discriminant Analysis (2DLDA):  
The non-iterative unilateral projection based 2DLDA.
- Two-Dimensional Linear Discriminant Analysis of Principal Component Vectors (2DPCA+2DLDA) or 2D Fisherface:  
The 2DPCA is performed as the feature extraction of 2DLDA.
- Class-Specific Subspace-Based Two-Dimensional Principal Component Analysis (2DPCA+CSS):  
The 2DPCA which each subspace is trained by only the samples in own class.
- Image Cross-Covariance Analysis (ICCA):  
The generalized 2DPCA with the generalized image covariance matrix, namely *image cross-covariance matrix*.
- Two-Dimensional Random Subspace Analysis (2DRSA):  
The 2DPCA with ensemble method, Random Subspace Method (RSM).
- Random Subspace Method-Based Image Cross-Covariance Analysis (ICCA+RSM):  
The ICCA with ensemble method, Random Subspace Method (RSM).

The highest recognition accuracy of above methods on three face databases (Yale, ORL, and AR) and a SAR image database (MSTAR) are presented in Table 4.10 and Fig. 4.72.

From these results, we found that the 2DLDA, 2DPCA+2DLDA, and ICCA can perform a improvement from 2DPCA in all databases with a little bit complexity. While the 2DRSA and RSM-based ICCA achieve the highest recognition accuracy on all databases but they take the highest complexity in the same time. For Class-Specific Subspace-Based 2DPCA, it achieves a good performance only in a small database, i.e. Yale. Unfortunately, the performance of 2DPCA+CSS will drop when it is applied to the big database.

Table 4.10: The Summary Comparison of The Highest Recognition Accuracy Results on Yale, ORL, AR, and MSTAR Databases

Database	Method	Accuracy (%)	Dimension
Yale	2DPCA	95.5556	100 × 11
	2DLDA	97.7778	100 × 14
	2DPCA+2DLDA	98.8889	100 × 17
	2DPCA+CSS	97.7778	100 × 9
	2DRSA	98.8889	12 × 20
	ICCA	97.7778	100 × 25
	ICCA+RSM	97.7778	11 × 20
ORL	2DPCA	92.5000	112 × 5
	2DLDA	94.0000	112 × 6
	2DPCA+2DLDA	94.5000	112 × 6
	2DPCA+CSS	86.0000	112 × 4
	2DRSA	95.0000	14 × 5
	ICCA	95.0000	112 × 7
	ICCA+RSM	96.0000	11 × 5
AR	2DPCA	60.7372	100 × 2
	2DLDA	60.8974	100 × 4
	2DPCA+2DLDA	65.0641	100 × 5
	2DPCA+CSS	51.2821	100 × 1
	2DRSA	65.2244	14 × 2
	ICCA	62.5000	100 × 2
	ICCA+RSM	66.0256	10 × 2
MSTAR	2DPCA	99.4872	80 × 11
	2DLDA	99.5604	80 × 78
	2DPCA+2DLDA	99.6337	80 × 35
	2DPCA+CSS	78.5348	80 × 25
	2DRSA	100.0000	12 × 11
	ICCA	99.9267	80 × 11
	ICCA+RSM	99.9267	13 × 11

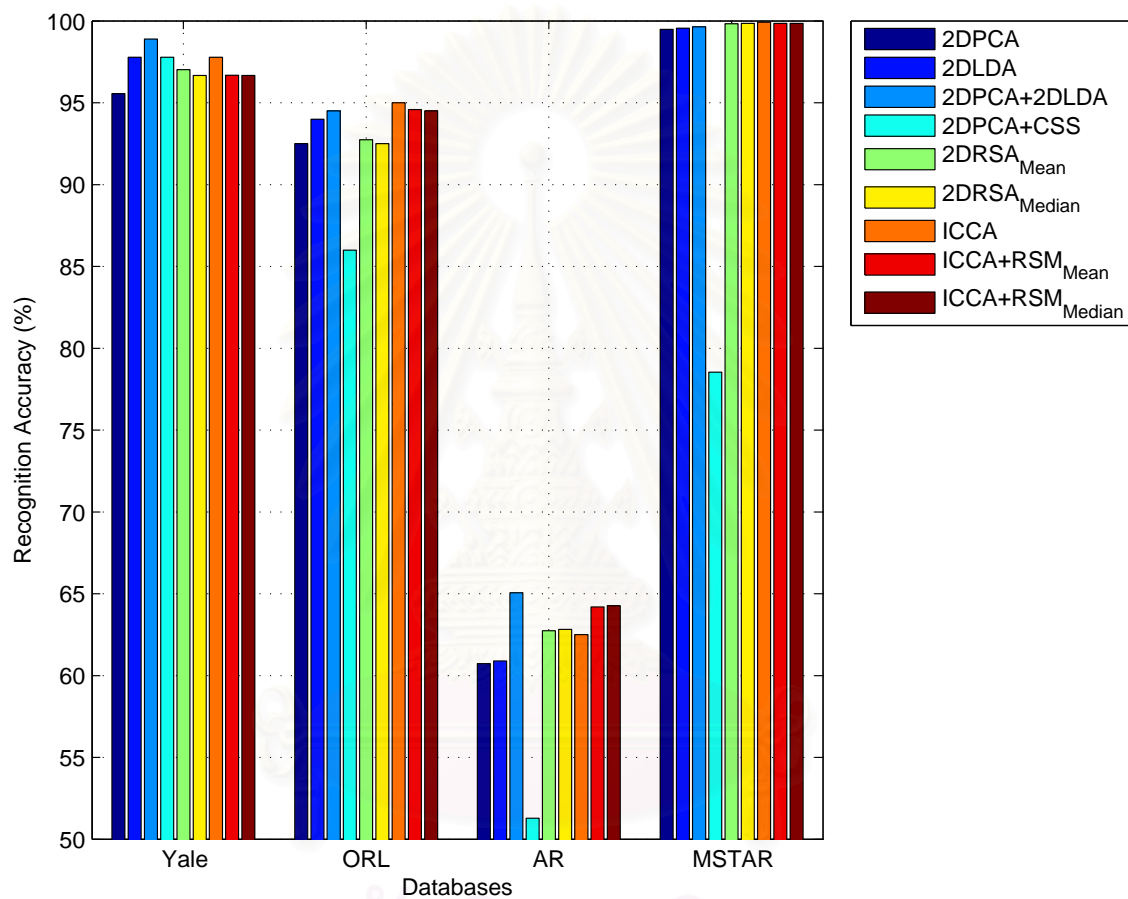


Figure 4.72: The Summary Comparison of The Highest Recognition Accuracy Results on Yale, ORL, AR, and MSTAR Databases.



# CHAPTER V

## THE CONCLUSIONS

This chapter summarizes the works presented in this dissertation including conclusions and future directions.

### 5.1 Conclusions of The Dissertation

In this dissertation, the improved frameworks of two-dimensional subspace analysis are proposed and applied to face and automatic target recognition (ATR). Firstly, this dissertation proposed the two-dimensional linear discriminant analysis of principal component vectors framework to improve the discriminant power of system. Since 2D-PCA is more suitable for face representation than face recognition, like PCA. For better performance in recognition task, linear discriminant still be necessary.

Secondly, Class-Specific Subspace (CSS) is applied to 2DPCA for providing the information of class labels. By applying CSS over 2DPCA, the class information is introduced to unsupervised method. Each subspace of CSS learned from only the training samples in own class. In this way, the CSS representation can provide a minimum reconstruction error. Which it can be used to classify the input data.

Thirdly, the random subspace method is applied to 2DPCA. Normally, the feature of 2DPCA is a matrix. In the row direction, the number of the columns of these matrix is affected by the number of selected eigenvalues of the image covariance matrix while the number of selected eigenvalues is not influenced in the column direction. Thus, the number of the rows is still equal to the height of original image and the random subspace method can be apply in the column direction. The random subspaces are constructed by randomly selecting a number of rows of the original feature matrix. The multiple classifiers are constructed in these random subspaces of the data feature space. These classifiers are usually combined by simple majority voting in the final decision rule.

Finally, this dissertation proposed the generalize form of image covariance matrix, called image cross-covariance matrix. And then combine with RSM for improving the accuracy and the robustness of system. Compare to the covariance matrix of PCA, the image covariance matrix discard some of the information. This disregard information may possibly be useful for discrimination. The image cross-covariance matrices are formulated by two variables, the original image and its shifted version. By our shifting algorithm, many image cross-covariance matrices are formulated to cover all of the information in which discarded by the image covariance matrix. And then the random subspace method is applied to this framework for taking the benefit of a huge number of these subspaces.

From the experimental results, we found that the main advantage of our proposed frameworks that is almost of them improve the recognition accuracy of face and automatic target recognitions. Moreover, they can be generally applied to other image recognition tasks. However, there are some disadvantages in our proposed frameworks. Since the several parameters cannot automatically specify to the optimal values. And the complexity of our proposed frameworks increase from the original 2DPCA, especially when RSM is applied.

## 5.2 Future Directions

- Several parameters (such as the number of feature vectors of 2DPCA and 2DLDA, the dimension of random subspace, the number of image shifting and the number of classifiers) are still manually specified. The optimal values are found by experiments for the best recognition accuracy. Nevertheless, automatic parameter specification is necessary for the practical applications in the future research.
- The two-dimensional subspace analysis is more appropriate for 2D data such as grayscale image than traditional one-dimensional subspace analysis. However, some databases furnish the color information which is neglected in this dissertation or some databases provide the image sequence (video). These information should be improving the performance of the recognition system. To combine all information in multi-dimensional data, It can be done by using *Multi-Linear Dimensional Subspace Analysis* which the dimension of the data is unlimit. The *Tensor* analysis and *High Order Singular Value Decomposition* (HOSVD) are the keywords of this framework.
- The distance measurement between two feature matrices is the one of the interesting point for improving the recognition accuracy. Moreover, the knowledge of matrix distance measurement probably be using to improve the distance measurement of feature tensors.
- The kernel method is the one technique which is opposite to the subspace analysis, because the dimension of the data are increased in kernel method while the subspace analysis produce the subspace with lower dimension data. Apply kernel method to two or multi-linear dimensional subspace analysis should be leading ultimately to improvements in pattern recognition.

## References

- [1] Kirby, M. and Sirovich, L. Application of the karhunen-lokve procedure for the characterization of human faces. IEEE Trans. Pattern Anal. and Mach. Intell. 12 (Jan. 1990): 103–108.
- [2] Sirovich, L. and Kirby, M. Low-dimensional procedure for characterization of human faces. J. Optical Soc. Am. 4 (1987): 519–524.
- [3] Turk, M. and Pentland, A. Eigenfaces for recognition. J. of Cognitive Neuroscience. 3, 1 (1991): 71–86.
- [4] Samal, A. and Iyengar, P. A. Automatic recognition and analysis of human faces and facial expressions: A survey. Pattern Recognition. 25, 1 (1992): 65–77.
- [5] O’Toole, A. J., Abdi, H., Deffenbacher, K. A., and Valentin, D. Low-dimensional representation of faces in higher dimensions of the face space. J. of the Optical Society of America A. 10, 3 (1993): 405–411.
- [6] Pentland, A., Moghaddam, B., and Starner, T., View-based and modular eigenspaces for face recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 1994.
- [7] Chellappa, R., Wilson, C. L., and Sirohey, S. Human and machine recognition of faces: A survey. Proceedings of the IEEE. 83 (May 1995): 705–740.
- [8] Hancock, P. J. B., Mike, A. B., and Bruce, V. Face processing: Human perception and principal components analysis. Memory and Cognition. 24, 1 (1996): 26–40.
- [9] Daugman, J. Face and gesture recognition: Overview. IEEE Transactions on Pattern Analysis and Machine Intelligence. 19 (July 1997): 675–676.
- [10] Belhumeur, P. N., P. Hespanha, J., and Kriegman, D. J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 19 (July 1997): 711–720.
- [11] Zhang, J., Yan, Y., and Lades, M. Face recognition: Eigenface, elastic matching, and neural nets. Proceedings of the IEEE. 85 (Sept. 1997): 1423–1435.
- [12] Phillips, P. J., Wechsler, H., Huang, J. S., and Rauss, P. J. The FERET database and evaluation procedure for face-recognition algorithms. Image and Vision Computing. 16, 5 (1998): 295–306.

- [13] Burton, A. M., Bruce, V., and Hancock, P. J. From pixels to people: A model of familiar face recognition. Cognitive Science. 23 (Jan. 1999): 1–31.
- [14] Martínez, A. M., Recognition of partially occluded and/or imprecisely localized faces using a probabilistic approach. Proceedings of Computer Vision and Pattern Recognition. 1 (June 2000): 712–717.
- [15] Moon, H. and Phillips, P. Computational and performance aspects of PCA-based face recognition algorithms. Perception. 30 (2001): 303–321.
- [16] Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. and Mach. Intell. 19 (July 1997): 711–720.
- [17] Zhao, W., Chellappa, R., and Krishnaswamy, A., Discriminant analysis of principle components for face recognition. IEEE 3rd Inter. Conf. on Automatic Face and Gesture Recognition, Japan, Apr. 1998.
- [18] Zhao, W., Chellappa, R., and Nandhakumar, N., Empirical performance analysis of linear discriminant classifiers. Computer Vision and Pattern Recognition, IEEE Computer Society. (1998): 164–171.
- [19] Marcialis, G. and Roli, F., Fusion of LDA and PCA for face recognition. Proceedings of the Workshop on Machine Vision and Perception, 8 th Workshop of the Italian Association for Artificial Intelligence (AIIA'02), Sept. 2002.
- [20] Lu, J., Plataniotis, K. N., and Venetsanopoulos, A. N. Face recognition using LDA-based algorithms. IEEE Trans. on Neural Networks. 14 (Jan. 2003): 195–200.
- [21] Zhao, Q., Principe, J., Brennan, V., Xu, D., and Wang, Z. Synthetic aperture radar automatic target recognition with three strategies of learning and representation. Optical Engineering. 39, 5 (2000): 1230–1244.
- [22] Mishra, A. and Mulgrew, B., Radar signal classification using pca-based features. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2006.
- [23] Brennan, V. L., Principal Component Analysis with Multiresolution. Ph.D. dissertation Florida University, 2001.
- [24] Hung, J. W., Wang, H. M., and Lee, L. S., Comparative analysis for data-driven temporal filters obtained via principal component analysis (PCA) and linear discriminant analysis (LDA) in speech recognition. 2001 Eurospeech Conference on Speech Communications and Technology, Sept. 2001.

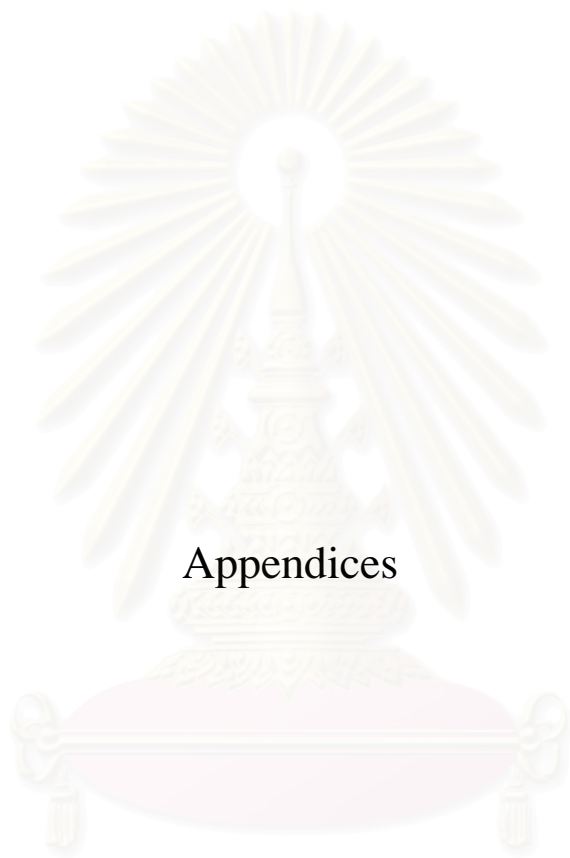
- [25] Kwon, O. W., Lee, T. W., and Chan, K., Application of variational bayesian pca for speech feature extraction. IEEE International Conference on Acoustics, Speech, and Signal Processing. 1 (2002): 825–828.
- [26] Deepu, V., Madhvanath, S., and A.G., R., Principal component analysis for online handwritten character recognition. International Conference on Pattern Recognition. 2 (2004): 327–330.
- [27] Sun, J., Hotta, Y., Katsuyama, Y., and Naoi, S., Low resolution character recognition by dual eigenspace and synthetic degraded patterns. HDP '04: Proceedings of the 1st ACM workshop on Hardcopy document processing, New York, NY, USA, ACM Press. (2004): 15–22.
- [28] Benetos, E., Kotti, M., Kotropoulos, C., Burred, J. J., Eisenberg, G., Haller, M., and Sikora, T., Comparison of subspace analysis-based and statistical model-based algorithms for musical instrument classification. Proceedings of 2nd Workshop on Immersive Communication and Broadcast Systems (ICOB '05), Berlin, Germany, Oct. 27-28 2005.
- [29] Yang, Q. and Tang, X. Recent advances in subspace analysis for face recognition. SINOBIOMETRICS. (2004): 275–287.
- [30] Zhao, W. Y., Discriminant component analysis for face recognition. 15th International Conference on Pattern Recognition (ICPR'00). 2 (2000): 818–821.
- [31] Fukunaga, K. Introduction to Statistical Pattern Recognition. Academic Press, second ed., 1990.
- [32] Chen, L., Liao, H., Ko, M., Lin, J., and Yu, G. A new LDA based face recognition system which can solve the small sample size problem. Pattern Recognition. 33, 10 (2000): 1713–1726.
- [33] Lu, J., Plataniotis, K. N., and Venetsanopoulos, A. N. Regularized discriminant analysis for the small sample size problem in face recognition. Pattern Recogn. Lett. 24, 16 (2003): 3079–3087.
- [34] Huang, R., Liu, Q., Lu, H., and Ma, S. Solving the small sample size problem of LDA. Pattern Recognition. 3 (2002): 29–32.
- [35] Liu, X., Chen, T., and Kumar, B. Face authentication for multiple subjects using eigenflow. Pattern Recognition. 36 (February 2003): 313–328.
- [36] Savvides, M., Kumar, B. V., and Khosla, P. K., Face verification using correlation filters. IEEE Workshop on Automatic Identification Advanced Technologies, March 2002.



- [37] Shan, S., Gao, W., and Zhao, D. Face recognition based on face-specific subspace. International Journal of Imaging Systems and Technology. 13, 1 (2003): 23–32.
- [38] Bellman, R. Adaptive Control Processes. Princeton, NJ. : Princeton University Press, 1961.
- [39] Yang, J., Zhang, D., Frangi, A. F., and Yang, J. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Anal. and Mach. Intell. 26 (Jan. 2004): 131–137.
- [40] Kong, H., Li, X., Wang, L., Teoh, E. K., Wang, J.-G., and Venkateswarlu, R. Generalized 2D principal component analysis. IEEE International Joint Conference on Neural Networks (IJCNN). 1 (2005): 108–113.
- [41] Zhang, D. and Zhou, Z. H. (2D)<sup>2</sup>PCA: 2-directional 2-dimensional PCA for efficient face representation and recognition. Neurocomputing. 69 (2005): 224–231.
- [42] Yang, J. and Yang, J. Y. From image vector to matrix: A straightforward image projection technique IMPCA vs. PCA. Pattern Recognition. 35, 9 (2002): 1997-1999.
- [43] Ye, J., Janardan, R., and Li, Q., Two-dimensional linear discriminant analysis. in *Advances in Neural Information Processing Systems 17* (Saul, L. K., Weiss, Y., and Bottou, L., eds.), pp. 1569–1576, Cambridge, MA : MIT Press, 2005.
- [44] Breiman, L. Bagging predictors. Machine Learning. 24, 2 (1996): 123–140.
- [45] Freund, Y. and Schapire, R. E., A decision-theoretic generalization of on-line learning and an application to boosting. European Conference on Computational Learning Theory. (1995): 23–37.
- [46] Ho, T. K. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. and Mach. Intell. 20 (Aug. 1998): 832–844.
- [47] Chawla, N. V. and Bowyer, K., Random subspaces and subsampling for 2D face recognition. Computer Vision and Pattern Recognition. 2 (2005): 582–589.
- [48] Lu, J., Plataniotis, K., Venetsanopoulos, A., and Li, S. Ensemble-based discriminant learning with boosting for face recognition. IEEE Trans. on Neural Networks. 17 (Jan. 2006): 166–178.
- [49] Asdornwised, W. and Jitapunkul, S. Multiple description pattern analysis: Robustness to misclassification using local discriminant frame expansions. IEICE Trans. Inf. & Syst. Special Section on Machine Vision Applications. E88-D, 10 (Oct. 2005): 2296–2307.

- [50] Skurichina, M. and Duin, R. P. W. Bagging, boosting and the random subspace method for linear classifiers. Pattern Anal. Appl. 5, 2 (2002): 121–135.
- [51] Xu, D., Yan, S., Zhang, L., Li, M., Ma, W., Liu, Z., and Zhang, H., Parallel image matrix compression for face recognition. the 11th International Multimedia Modelling Conference. 00 (2005): 232–238.
- [52] Zhang, D., Zhou, Z.-H., and Chen, S. Diagonal principal component analysis for face recognition. Pattern Recognition. 39, 1 (2006): 133–135.
- [53] Ho, T. K., Nearest neighbors in random subspaces. Proceedings of the 2nd Int’l Workshop on Statistical Techniques in Pattern Recognition, Sydney, Australia, Aug. 1998.
- [54] Zuo, W., Wang, K., and Zhang, D., Bi-directional PCA with assembled matrix distance metric. International Conference on Image Processing. 2 (2005): 958–961.
- [55] Xu, A., Jin, X., Jiang, Y., and Guo, P., Complete two-dimensional PCA for face recognition. International Conference on Pattern Recognition. 3 (2006): 481–484.
- [56] Ye, J., Generalized low rank approximations of matrices. International Conference on Machine Learning. (2004): 887–894.
- [57] Xu, D., Yan, S., Zhang, L., Liu, Z., and Zhang, H., Coupled subspaces analysis. tech. rep., Microsoft Research, Oct. 2004.
- [58] Sanguansat, P., Asdornwised, W., Jitapunkul, S., and Marukatat, S., Two-dimensional linear discriminant analysis of principle component vectors for face recognition. IEEE International Conference on Acoustics, Speech, and Signal Processing. 2 (May 2006): 345–348.
- [59] Sanguansat, P., Asdornwised, W., Jitapunkul, S., and Marukatat, S. Two-dimensional linear discriminant analysis of principle component vectors for face recognition. IEICE Trans. Inf. & Syst. Special Section on Machine Vision Applications. E89-D, 7 (2006): 2164–2170.
- [60] Abdi, H., Distance, pp. 280–284. Thousand Oaks (CA) : Sage, 2007.
- [61] Zuo, W., Zhang, D., and Wang, K. An assembled matrix distance metric for 2DPCA-based. Pattern Recognition Letters. 27 (2006): 210–216.
- [62] Meng, J. and Zhang, W. Volume measure in 2DPCA-based face recognition. Pattern Recognition Letters. 28 (2007): 1203–1208.

- [63] Zuo, W., Zhang, D., and Wang, K. An application of the matrix volume in probability. Linear Algebra Application. 321 (2000): 9–25.
- [64] Sanguansat, P., Asdornwised, W., Jitapunkul, S., and Marukatat, S., Class-specific subspace-based two-dimensional principal component analysis for face recognition. International Conference on Pattern Recognition 2 (Aug. 2006): 1246–1249.
- [65] Sanguansat, P., Asdornwised, W., Jitapunkul, S., and Marukatat, S., Image cross-covariance analysis for face recognition. IEEE Region 10 Conference on Convergent Technologies for the Asia-Pacific, 2007.
- [66] Liu, J. and Chen, S. Non-iterative generalized low rank approximation of matrices. Pattern Recognition Letters. 27 (2006): 1002–1008.
- [67] Sanguansat, P., Asdornwised, W., Jitapunkul, S., and Marukatat, S., Two-dimensional random subspace analysis for face recognition. 7th International Symposium on Communications and Information Technologies, 2007.
- [68] Sanguansat, P., Asdornwised, W., Jitapunkul, S., and Marukatat, S., Two-dimensional diagonal random subspace analysis for face recognition. International Conference on Telecommunications, Industry and Regulatory Development. 1 (Aug 2007): 66–69.
- [69] Yale University, The Yale face database., 1997. Available from <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.
- [70] The AT&T (Olivetti) Research Laboratory, The (Olivetti) research laboratory face database (ORL)., 1992. Available from <http://www.cl.cam.ac.uk/Research/DTG/attarchive/facedatabase.html>.
- [71] Martinez, A. and Benavente, R., The AR face database., 1998. Available from [http://rv11.ecn.purdue.edu/~aleix/aleix\\_face\\_DB.html](http://rv11.ecn.purdue.edu/~aleix/aleix_face_DB.html).
- [72] The Center for Imaging Science (CIS), DARPA Moving and Stationary Target Recognition Program (MSTAR)., 1997. Available from <http://cis.jhu.edu/data.sets/MSTAR>.



Appendices

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## Appendix A

### List of Abbreviations

2DFDA	Two-Dimensional Fisher Discriminant Analysis
2DLDA	Two-Dimensional Linear Discriminant Analysis
2DPCA	Two-Dimensional Principal Component Analysis
2DPCA+2DLDA	2DLDA of Principal Component Vectors
2DPCA+CSS	Class-Specific Subspace-Based 2DPCA
2DRSA	Two-Dimensional Random Subspace Analysis
2D <sup>2</sup> RSA	Two-Dimensional Diagonal Random Subspace Analysis
ATR	Automatic Target Recognition
B2DPCA	Bilateral-Projection-Based 2DPCA
CSA	Coupled Subspace Analysis
CSS	Class-Specific Subspace
DFCSS	Distance From Class-Specific Subspace
DFSS	Distance From Face-Specific Subspace
DiaPCA	Diagonal Principal Component Analysis
FDA	Fisher Discriminant Analysis
FSS	Face-Specific Subspace
HOSVD	High Order Singular Value Decomposition
ICCA	Image Cross-Covariance Analysis
ICCA+RSM	Random Subspace Method-Based ICCA
LDA	Linear Discriminant Analysis
NN	The nearest neighbor classifier
PCA	Principal Component Analysis
RSM	Random Subspace Method
SAR	Synthetic Aperture Radar
SSS	Small Sample Size Problem
SVD	Singular Value Decomposition



## Appendix B

### Publications and Presentations

Parinya Sanguansat, Widhyakorn Asdornwised, Sanparith Marukatat and Somchai Jitapunkul,

“Two-Dimensional Random Subspace Analysis,” In preparation to submit to IEICE Transaction on Information and Systems.

Parinya Sanguansat, Widhyakorn Asdornwised, Sanparith Marukatat and Somchai Jitapunkul,

“Generalized 2DPCA using Image Cross-Covariance Matrix,” In preparation to submit to IEICE Transaction on Information and Systems.

Parinya Sanguansat, Widhyakorn Asdornwised, Sanparith Marukatat and Somchai Jitapunkul,

“Two-Dimensional Linear Discriminant Analysis of Principle Component Vectors for Face Recognition,” IEICE Transaction on Information and Systems: Special Section on Machine Vision Applications, Vol. E89-D, No. 7, pp. 2164–2170, July. 2006.

Parinya Sanguansat, Widhyakorn Asdornwised, Sanparith Marukatat and Somchai Jitapunkul,

“Two-Dimensional Linear Discriminant Analysis of Principle Component Vectors for Face Recognition,” Proceeding on IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2006), Vol. 2, pp. 345–348, Toulouse, France, 15-19 May 2006.

Parinya Sanguansat, Widhyakorn Asdornwised, Sanparith Marukatat and Somchai Jitapunkul,

“Class-Specific Subspace-Based Two-Dimensional Principal Component Analysis for Face Recognition,” Proceeding on the 18<sup>th</sup> International Conference on Pattern Recognition (ICPR 2006), Vol. 2, pp. 1246–1249, Hong Kong, China, 20-24 August 2006.

Parinya Sanguansat, Widhyakorn Asdornwised, Sanparith Marukatat and Somchai Jitapunkul,

“Image Cross-Covariance Analysis for Face Recognition,” Proceeding on IEEE Region 10 Conference on Convergent Technologies For The Asia-Pacific (TEN-CON 2007), Taipei, Taiwan, 30 October-2 November 2007.

Parinya Sanguansat, Widhyakorn Asdornwised, Sanparith Marukatat and Somchai Jitapunkul,

“Two-Dimensional Random Subspace Analysis for Face Recognition,” Proceeding on the 7<sup>th</sup> International Symposium on Communications and Information Technologies (ISCIT 2007), Sydney, Australia, 16-19 October 2007.

Parinya Sanguansat, Widhyakorn Asdornwised, Sanparith Marukatat and Somchai Jitapunkul,

“Two-Dimensional Diagonal Random Subspace Analysis for Face Recognition,” Proceeding on the International Conference on Telecommunications, Industry and Regulatory Development, pp. 66–69, Bangkok, Thailand, 19-21 August 2007.



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## Vitae

Parinya Sanguansat was born in Bangkok, Thailand in 1980. He received the B.Eng. and M.Eng. degrees from the Department of Electrical Engineering at the Chulalongkorn University, Bangkok, Thailand, in 2001 and 2004 respectively. He is currently pursuing the Doctoral degree in electrical engineering at Chulalongkorn University, Bangkok, Thailand, since 2004. His research areas are digital signal processing in pattern recognition including on-line handwritten recognition, face and automatic target recognition.



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย