

วิธีใหม่ในการวิเคราะห์ลักษณะแตกต่างกันเชิงเส้นสองมิติ
สำหรับใช้ในงานรู้จำใบหน้า

A NOVEL METHOD OF TWO-DIMENSIONAL LINEAR DISCRIMINANT
ANALYSIS FOR FACE RECOGNITION APPLICATION

วิญญู ธานี
WAIYAWUT SANAYHA

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2553

KMITL-2010-EN-D-018-004

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

วิธีใหม่ในการวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติ
สำหรับใช้ในงานรู้จำใบหน้า

A NOVEL METHOD OF TWO-DIMENSIONAL LINEAR DISCRIMINANT
ANALYSIS FOR FACE RECOGNITION APPLICATION



วิยวุฒิ เสน่หา

WAIYAWUT SANAYHA

เลขหมู่.....
เลขทะเบียน..... 107443
วัน,เดือน,ปี..... 29 ส.ค. 2553

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2553

KMIT-2010-EN-D-018-004

**A NOVEL METHOD OF TWO-DIMENSIONAL LINEAR DISCRIMINANT
ANALYSIS FOR FACE RECOGNITION APPLICATION**

WAIYAWUT SANAYHA

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
DOCTOR OF ENGINEERING IN ELECTRICAL ENGINEERING
FACULTY OF ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

2010

KMIT-2010-EN-D-018-004


COPYRIGHT 2010

FACULTY OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ใบรับรองวิทยานิพนธ์

หัวข้อวิทยานิพนธ์ วิธีใหม่ในการวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติสำหรับใช้ในงานรู้จำใบหน้า
Thesis Title A Novel Method of Two-Dimensional Linear Discriminant Analysis for Face
 Recognition Application
นักศึกษา นายวิษณุ เสน่หา
รหัสประจำตัว 45160315
ปริญญา วิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชา วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษาวิทยานิพนธ์ รศ.ดร.ยุทธพงษ์ รังสรรค์เสรี
หมายเลขวิทยานิพนธ์ KMITL-2010-EN-D-018-004

คณะกรรมการสอบวิทยานิพนธ์		ลายมือชื่อ
รศ.ดร.ฟูศักดิ์	ชีวิสุวิทย์	
รศ.ดร.สุรพันธุ์	เอื้อไพฑูลย์	
ผศ.ดร.สุภาวดี	อร่ามวิทย์	
รศ.ดร.สุวิพล	สิทธิชีวกาศ	
รศ.ดร.ยุทธพงษ์	รังสรรค์เสรี	

วัน / เดือน / ปี ที่สอบ วันพฤหัสบดีที่ 7 มกราคม พ.ศ. 2553 เวลา 09.00-11.00 น.

สถานที่สอบ ณ อาคาร A ชั้น 3 ห้องประชุม 2

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

KING MONKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG


คณะวิศวกรรมศาสตร์ รับรองแล้ว



(รองศาสตราจารย์ ดร.กอบชัย เดชหาญ)

คณบดี คณะวิศวกรรมศาสตร์

วันที่ 7 มกราคม พ.ศ. 2553

สำนักทะเบียนและประมวลผล สจส.
วันที่ส่งเล่มวิทยานิพนธ์ฉบับสมบูรณ์
วันที่ 2 เดือน ก.ม. พ.ศ. 53
ลงชื่อ..... 

หัวข้อวิทยานิพนธ์	วิธีใหม่ในการวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติสำหรับใช้ในงานรู้จำใบหน้า
นักศึกษา	นายวิษุฒิ เสน่หา
รหัสนักศึกษา	45160315
ปริญญา	วิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชา	วิศวกรรมไฟฟ้า
พ.ศ.	2553
อาจารย์ที่ปรึกษาวิทยานิพนธ์	รศ.ดร.ยุทธพงษ์ รังสรรค์เสรี

บทคัดย่อ

งานวิจัยนี้นำเสนอเทคนิคการฉายภาพสำหรับการประยุกต์ใช้ในงานรู้จำใบหน้าตามหลักการวิเคราะห์ลักษณะแตกต่างเชิงเส้น (Linear discriminant analysis) หรือ LDA ประยุกต์ร่วมกับฟังก์ชันการถ่วงน้ำหนัก (Weighting function) เทคนิคการฉายภาพกระทำในแบบสองมิติและสองทิศทาง (2-Dimensional 2-directional) หรือ (2D)²LDA นั่นคือ โดยปกติการทำงานของ LDA จะวิเคราะห์ข้อมูลด้านแถวของภาพเป็นหลัก แต่สำหรับวิธี (2D)²LDA จะเพิ่มการวิเคราะห์ข้อมูลด้านคอลัมน์ของภาพอีกแนวทางหนึ่งด้วย เมื่อรวมเทคนิคการวิเคราะห์ข้อมูลทั้งด้านแถวและคอลัมน์ของภาพทำให้สามารถแก้ปัญหาการกลุ่มตัวอย่างข้อมูลที่มีขนาดเล็กได้ ต่อจากนั้นมีการประยุกต์เพิ่มฟังก์ชันการถ่วงน้ำหนักเข้าไปในเมตริกซ์การกระจายระหว่างกลุ่ม (Between-class scatter matrix) และฟังก์ชันการถ่วงน้ำหนักสัมพัทธ์ในเมตริกซ์การกระจายภายในกลุ่ม (Within-class scatter matrix) เพื่อปรับปรุงข้อมูลให้สามารถแก้ปัญหาคงความสับสนของข้อมูลในแต่ละกลุ่มได้ เทคนิคนี้เรียกว่า การวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติสองทิศทางแบบถ่วงน้ำหนักสัมพัทธ์ เรียกย่อๆ ว่า RW(2D)²LDA สามารถเพิ่มความถูกต้องในการตัดสินใจวิเคราะห์ลักษณะแตกต่างได้ดีกว่าวิธี LDA หรือ 2DLDA โดยการทดสอบเทคนิคที่นำเสนอกับฐานข้อมูลภาพใบหน้าตามเงื่อนไขต่างๆ จำนวน 5 ฐานข้อมูลอย่างได้ผล ซึ่งผลการทดสอบสามารถสนับสนุนและยืนยันข้อสรุปได้ว่า กระบวนวิธี RW(2D)²LDA ที่นำเสนอมีประสิทธิภาพที่ดีกว่าวิธีแบบดั้งเดิม

Thesis Title	A Novel Method of Two-Dimensional Linear Discriminant Analysis for Face Recognition Application
Student	Mr.Waiyawut Sanayha
Student ID.	45160315
Degree	Doctor of Engineering
Program	Electrical Engineering
Year	2010
Thesis Advisor	Assoc. Prof. Dr. Yuttapong Rangsanseri

ABSTRACT

This thesis proposes a novel image projection technique for face recognition application based on Linear Discriminant Analysis (LDA) and combine with Weighting function. The projection is performed through 2-Directional 2-Dimensional LDA or $(2D)^2LDA$ that is essential working in the row-direction of images, and then proposes an alternative 2DLDA which works in the column-direction of images. By simultaneously combining row and column directions is overcome the “small sample size” problem. Moreover, a weighting discriminant hyperplanes are used in the between-class scatter matrix, and also relevance weighting in the within-class scatter matrix in order to weighted information for solving confusable data in these classes. This technique called the relevance weighted 2-dimensional 2-directional linear discriminant analysis or $RW(2D)^2LDA$ use for a more accurate discriminant decision than the one that produced by the conventional LDA, or 2DLDA. The proposed technique has been successfully tested on 5 face databases by several conditions. Experimental results can support and confirm the conclusion indicate that the proposed $RW(2D)^2LDA$ algorithm outperforms the conventional algorithms.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้อย่างดี ด้วยการให้คำแนะนำ ปรีกษา ตลอดจนการตรวจสอบเนื้อหาในวิทยานิพนธ์เล่มนี้อย่างละเอียดทุกขั้นตอนจาก รศ.ดร.ยุทธพงษ์ รังสรรค์เสรี ซึ่งเป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ ข้าพเจ้ารู้สึกทราบบ้างในความอนุเคราะห์จากท่านอาจารย์ และขอขอบพระคุณเป็นอย่างสูง

ข้าพเจ้าขอกราบขอบพระคุณทุกท่าน ที่ให้ความช่วยเหลือ และครอบครัวของข้าพเจ้าที่เป็นกำลังใจ และให้การสนับสนุนในทุกเรื่องๆ ทำให้ข้าพเจ้าสามารถทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดี

สุดท้ายนี้คุณค่าและประโยชน์อันพึงมีจากวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอบแต่ผู้มีพระคุณทุกท่าน

วิญฉฉฉ ฉฉฉฉ

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VII
สารบัญรูป.....	VIII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของวิทยานิพนธ์.....	1
1.2 ความมุ่งหมายและวัตถุประสงค์ของวิทยานิพนธ์.....	1
1.3 สมมติฐานของการศึกษา.....	2
1.4 ทฤษฎีหรือแนวความคิดที่ใช้ในการวิจัย.....	3
1.5 การเปรียบเทียบระหว่างวิธีการที่นำเสนอกับวิธีการแบบพื้นฐาน.....	4
1.6 ขอบเขตการวิจัย.....	5
1.7 ขั้นตอนของการศึกษา.....	5
บทที่ 2 ระบบการรู้จำใบหน้า.....	7
2.1 ปัจจัยที่มีผลต่อการรู้จำใบหน้า.....	7
2.1.1 ความสว่าง (Illumination)	7
2.1.2 มุมของใบหน้า (Pose)	8
2.1.3 การแสดงอารมณ์ทางใบหน้า (Facial expression).....	8
2.1.4 การหน่วงเวลา (Time delay)	9
2.1.5 การบังใบหน้า (Occlusions)	10
2.2 อัลกอริทึมการรู้จำใบหน้า (Face recognition algorithms).....	11
2.2.1 วิธีการฉายภาพแบบเชิงเส้นและไม่เป็นเชิงเส้น.....	11
2.2.2 โครงข่ายประสาทเทียม.....	13
2.2.3 กาบอร์ฟิลเตอร์และเวฟเล็ต.....	14
2.2.4 ระบบฟังก์ชันทำซ้ำและแฟลคทอล.....	15
2.2.5 ภาพอินฟราเรดและไฮเปอร์สเปกตรัม.....	16

สารบัญ (ต่อ)

	หน้า
2.3 ฐานข้อมูลภาพใบหน้า (Face databases)	17
2.3.1 ฐานข้อมูล LFW (Labeled Faces in the Wild).....	19
2.3.2 ฐานข้อมูล FERET (Facial Recognition Technology).	19
2.3.3 ฐานข้อมูล FRGC (Face Recognition Grand Challenge)	20
2.3.4 ฐานข้อมูล CAS-PEAL.....	21
2.3.5 ฐานข้อมูล PIE (CMU Pose, Illumination, and Expression).....	22
2.3.6 ฐานข้อมูล YALE.....	23
2.3.7 ฐานข้อมูล AR	23
2.3.8 ฐานข้อมูล ORL	24
2.3.9 ฐานข้อมูล Georgia Tech.....	24
บทที่ 3 การวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติ.....	26
3.1 แนวทางการวิเคราะห์เชิงเส้น (Linear analysis approach).....	28
3.1.1 การวิเคราะห์องค์ประกอบหลัก (Principal component analysis: PCA).....	29
3.1.2 การวิเคราะห์ลักษณะแตกต่างเชิงเส้น (Linear discriminant analysis: LDA).....	32
3.1.3 การวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติ (2-Dimensional LDA)	38
3.1.4 แนวทางการกระจายแบบคู่ (Pairwise scatter approach)	42
3.1.5 วิธี 2DLDA อีกแนวทางหนึ่ง (An alternative 2DLDA).....	44
3.1.6 การวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติสองทิศทาง ((2D) ² LDA)	45
3.2 วิธีการที่นำเสนอ.....	50
3.2.1 ฟังก์ชันการถ่วงน้ำหนัก (Weighting function)	50
3.2.2 วิธี Weighted (2D) ² LDA.....	51
3.2.3 วิธี Relevance Weighted (2D) ² LDA.....	51
3.3 หลักการคำนวณของวิธีที่นำเสนอ.....	52
3.4 การจำแนกข้อมูลด้วยวิธี k-NN (k-Nearest Neighbor Classification).....	55
3.4.1 ระยะยูคลิดีเนียน (Euclidean distance)	57
3.4.2 ระยะมาฮาลานอบิส (Mahalanobis distance).....	57

สารบัญ (ต่อ)

	หน้า
บทที่ 4 การทดสอบวิธีการวิเคราะห์หลักขณะแตกต่างเชิงเส้นสองมิติ.....	59
4.1 การทดสอบสมรรถนะของวิธีที่นำเสนอ.....	60
4.2 การทดสอบภายใต้เงื่อนไขมุมและความสว่างภาพใบหน้า.....	62
4.3 การทดสอบภายใต้เงื่อนไขการเปลี่ยนขนาดและการเอียงใบหน้า.....	65
4.4 การทดสอบภายใต้เงื่อนไขระยะเวลาและการบังใบหน้า.....	66
4.5 การทดสอบภายใต้เงื่อนไขการแสดงอารมณ์และความสว่างใบหน้า.....	68
4.6 สรุปผลการทดสอบ.....	69
บทที่ 5 สรุปผลการวิจัยและแนวทางในการพัฒนา.....	72
5.1 สรุปผลการวิจัย.....	72
5.2 แนวทางในการพัฒนา.....	73
บรรณานุกรม.....	74
ภาคผนวก.....	79
ภาคผนวก ก. ผลงานวิจัยที่ได้รับการตีพิมพ์เผยแพร่ในการประชุมวิชาการ.....	80
ภาคผนวก ข. ผลงานวิจัยที่ได้รับการตีพิมพ์เผยแพร่ในวารสารวิชาการระดับนานาชาติ.....	81
ประวัติผู้เขียน.....	103

สารบัญตาราง

ตารางที่	หน้า
2.1 การเปรียบเทียบฐานข้อมูลภาพใบหน้า.....	18
4.1 อัตราความถูกต้องในการรู้จำใบหน้าภายใต้เงื่อนไขของมุม และความสว่างใบหน้า.....	64
4.2 ผลการเปรียบเทียบสมรรถนะภายใต้เงื่อนไขการเปลี่ยนขนาดและการเอียงใบหน้า.....	66
4.3 ความถูกต้องการรู้จำใบหน้าภายใต้เงื่อนไขระยะเวลา และการบังใบหน้า.....	67
4.4 ผลการเปรียบเทียบสมรรถนะภายใต้เงื่อนไขระยะเวลาและความสว่างใบหน้า.....	69

สารบัญรูป

รูปที่	หน้า
2.1 ตัวอย่างภาพความสว่างใบหน้า.....	8
2.2 ตัวอย่างภาพใบหน้าที่มีมุมต่างๆ.....	8
2.3 ตัวอย่างภาพการแสดงอารมณ์ทางใบหน้า.....	9
2.4 ตัวอย่างภาพใบหน้าในเวลาต่างกัน.....	9
2.5 ตัวอย่างภาพการบังใบหน้าแบบต่างๆ.....	10
2.6 การลดมิติภาพตามวิธีการฉายภาพแบบเชิงเส้นและไม่เป็นเชิงเส้น.....	12
2.7 โครงสร้างตามแนวทางโครงข่ายประสาทเทียม.....	13
2.8 หลักการวิธีการบอร์และการลดมิติของผลตอบสนองฟิลเตอร์.....	15
2.9 ตัวอย่างภาพใบหน้าจากฐานข้อมูล LFW.....	19
2.10 ตัวอย่างภาพใบหน้าจากฐานข้อมูล FERET.....	20
2.11 ตัวอย่างภาพใบหน้าจากฐานข้อมูล FRGC.....	21
2.12 ตัวอย่างภาพใบหน้าจากฐานข้อมูล CAS-PEAL.....	22
2.13 ตัวอย่างภาพใบหน้าจากฐานข้อมูล PIE.....	23
2.14 ตัวอย่างภาพใบหน้าจากฐานข้อมูล YALE.....	23
2.15 ตัวอย่างภาพใบหน้าจากฐานข้อมูล AR.....	24
2.16 ตัวอย่างภาพใบหน้าจากฐานข้อมูล ORL.....	24
2.17 ตัวอย่างภาพใบหน้าจากฐานข้อมูล Georgia Tech.....	25
3.1 การพัฒนาอัลกอริทึมการรู้จำใบหน้าตามแนวทาง LDA.....	27
3.2 เทคนิคการฉายภาพแบบ PCA.....	30
3.3 การเปลี่ยนเมตริกซ์ภาพเป็นเวกเตอร์.....	31
3.4 การเลือกแกนการฉายที่เหมาะสม.....	34
3.5 กลุ่มข้อมูลตำแหน่งต่างๆ ที่มีความแปรปรวนเท่ากัน.....	43
3.6 การถ่วงน้ำหนักแบบ Weighted และ Relevance Weighted.....	51
3.7 การทำงานของวิธีการจำแนกด้วย k -NN.....	56
4.1 ผังการทดสอบภายใต้เงื่อนไขต่างๆ.....	59
4.2 ภาพใบหน้าที่ใช้ทดสอบจากฐานข้อมูล FERET.....	60
4.3 อัตราการรู้จำใบหน้าของวิธีใหม่ที่นำเสนอ.....	61
4.4 ภาพใบหน้าจากฐานข้อมูล PIE ภายใต้เงื่อนไขมุมและความสว่าง.....	63
4.5 อัตราการรู้จำใบหน้าภายใต้เงื่อนไขมุมและความสว่างใบหน้า.....	64

สารบัญรูป (ต่อ)

รูปที่	หน้า
4.6 ภาพใบหน้าจากฐานข้อมูล ORL ภายใต้เงื่อนไขการเปลี่ยนขนาดและการเอียงใบหน้า.....	65
4.7 อัตราการรู้จำใบหน้าภายใต้เงื่อนไขการเปลี่ยนขนาดและการเอียงใบหน้า.....	66
4.8 ภาพใบหน้าจากฐานข้อมูล AR ภายใต้เงื่อนไขระยะเวลาและการบังใบหน้า.....	67
4.9 อัตราการรู้จำใบหน้าภายใต้เงื่อนไขระยะเวลาและการบังใบหน้า.....	68
4.10 ภาพใบหน้าจากฐานข้อมูล YALE ภายใต้เงื่อนไขการแสดงอารมณ์และความสว่างใบหน้า.....	68
4.11 อัตราการรู้จำใบหน้าภายใต้เงื่อนไขการแสดงอารมณ์และความสว่างใบหน้า.....	69

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของวิทยานิพนธ์

การประมวลผลภาพดิจิทัล (Digital image processing) เป็นเทคโนโลยีสาขาหนึ่งที่ได้รับ ความสนใจและทำการวิจัยกันอย่างแพร่หลาย หลายปีที่ผ่านมาการพิสูจน์ตัวตนบุคคล (User Authentication) ที่อาศัยการประมวลผลภาพดิจิทัลได้เพิ่มความสำคัญขึ้นอย่างมาก เนื่องจากในปัจจุบัน การควบคุมรักษา ความปลอดภัยมีความจำเป็นอย่างยิ่งในสถานการณ์ที่มีการก่อการร้ายเกิดขึ้นในภูมิภาคต่างๆ ทั่วโลก โดยวิธีดั้งเดิมนิยมใช้บัตรประจำตัวและรหัสผ่านสำหรับการพิสูจน์ตัวตน อย่างไรก็ตามความน่าเชื่อถือ และความสะดวกในการใช้งานในลักษณะนี้ยังไม่เป็นที่น่าประทับใจเท่าใดนัก

เทคโนโลยีการพิสูจน์ตัวตนทางชีววิทยา หรือไบโอเมตริก (Biometric) เช่น เสียง ม่านตา ลายนิ้วมือ ลายฝ่ามือ และใบหน้า จึงถูกนำมาใช้ในการพิสูจน์ตัวตนบุคคลกันอย่างแพร่หลาย โดยเฉพาะ การวิเคราะห์ภาพใบหน้าเพื่อแยกแยะบุคคลที่ใช้กระบวนการรู้จำใบหน้า (Face recognition) ได้มีการ วิจัยและพัฒนาอย่างต่อเนื่อง เพราะนำไปประยุกต์ใช้งานได้หลากหลาย เช่น ระบบรักษาความปลอดภัย ในอาคารสถานที่ การติดตามข้อมูลบุคคลจากภาพถ่ายหรือกล้องโทรทัศน์ ตลอดจนใช้งานแทนระบบ บัตรหรือรหัสผ่าน

แม้ว่างานวิจัยด้านนี้กำลังได้รับความสนใจ และทำการศึกษาค้นคว้าวิธีการรู้จำใบหน้ากัน หลากหลายวิธี แต่การพัฒนาเพื่อประยุกต์ใช้งานยังมีเงื่อนไขที่เป็นข้อจำกัดต่อสมรรถนะของระบบอยู่ หลายประการ จึงต้องศึกษาอัลกอริทึมที่ละเอียดซับซ้อนเพื่อลดความผิดพลาดให้น้อยที่สุด รวมทั้งต้องใช้ อุปกรณ์คอมพิวเตอร์ที่มีประสิทธิภาพค่อนข้างสูง อย่างไรก็ตาม การรู้จำใบหน้าซึ่งเป็นการพิสูจน์ตัวตน ทางไบโอเมตริกก็เป็นที่นิยมใช้งานกันมากวิธีหนึ่ง และวิธีการทางสถิติเป็นแนวทางที่ให้ความถูกต้อง ในการรู้จำได้ดีในระดับหนึ่ง สามารถพัฒนาให้มีประสิทธิภาพในการรู้จำใบหน้าที่ดีขึ้นได้

1.2 ความมุ่งหมายและวัตถุประสงค์ของวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้มุ่งหวังเพื่อศึกษา เทคนิคการฉายภาพ (Image projection technique) สำหรับการประยุกต์ใช้ในงานรู้จำใบหน้า โดยใช้เทคนิคการฉายภาพตามหลักการวิเคราะห์ลักษณะแตกต่าง เชิงเส้นแบบสองมิติและสองทิศทาง (2-Dimensional and 2-directional linear discriminant analysis: (2D)²LDA) ซึ่งเป็นเทคนิคที่พัฒนามาจากการวิเคราะห์ลักษณะแตกต่างเชิงเส้น (Linear discriminant analysis: LDA) นั่นคือ โดยปกติการทำงานของ LDA จะวิเคราะห์ข้อมูลด้านแถวของภาพเป็นหลัก แต่สำหรับ วิธี (2D)²LDA จะเพิ่มการวิเคราะห์ข้อมูลด้านคอลัมน์ของภาพอีกแนวทางหนึ่งด้วย

นอกจากนี้ยังศึกษาวิธีการประยุกต์เทคนิคการฉายภาพด้วยฟังก์ชันการถ่วงน้ำหนัก (Weighting function) และฟังก์ชันการถ่วงน้ำหนักสัมพัทธ์ (Relevance Weighting function) เพื่อปรับปรุงข้อมูลระหว่างกลุ่มให้มีความแตกต่างกันมากขึ้น ขณะเดียวกันก็จะปรับปรุงข้อมูลภายในกลุ่มให้ใกล้เคียงกันมากขึ้น ทำให้ประสิทธิภาพโดยรวมในการตัดแยก (Classification) ข้อมูลได้ถูกต้องขึ้น

ดังนั้น วิทยานิพนธ์นี้จึงเป็นการนำเสนอวิธีใหม่โดยใช้เทคนิคการฉายภาพแบบสองมิติและสองทิศทาง หรือ $(2D)^2LDA$ มาประยุกต์เพิ่มฟังก์ชันการถ่วงน้ำหนักเข้าไปในเมตริกซ์การกระจายระหว่างกลุ่ม (Between-class scatter matrix) และเมตริกซ์การกระจายภายในกลุ่ม (Within-class scatter matrix) เพื่อปรับปรุงข้อมูลให้สามารถแก้ปัญหาความสับสนของข้อมูลในแต่ละกลุ่มได้ สุดท้ายจะดำเนินการทดสอบวิธีการรู้จำใบหน้าที่น่าสนใจกับฐานข้อมูลแบบต่างๆ ซึ่งผลการทดสอบต้องสามารถสนับสนุนและยืนยันข้อสรุปได้ว่า กระบวนการที่น่าสนใจมีประสิทธิภาพที่ดีกว่าวิธีการ LDA แบบดั้งเดิม

1.3 สมมุติฐานของการศึกษา

เทคนิคที่ใช้ในการรู้จำใบหน้าที่น่าสนใจงานกันวิธีหนึ่งคือ การวิเคราะห์องค์ประกอบหลัก (Principal component analysis: PCA) และการวิเคราะห์หลักขณะแตกต่างเชิงเส้น (LDA) บางครั้งเรียกว่า Fisher discriminant analysis (FDA) หรือ Fisher's linear discriminant (FLD) ซึ่งทั้งสองวิธีก็มีการศึกษาวิจัยและพัฒนากันอย่างต่อเนื่อง เช่น PCA ได้ถูกพัฒนาเป็นแบบสองมิติ (2-Dimensional PCA: 2DPCA) และแบบสองมิติสองทิศทาง (2-Dimensional and 2-directional PCA: $(2D)^2PCA$) ส่วน LDA ก็มีการพัฒนาเป็นแบบสองมิติ (2-Dimensional LDA: 2DLDA) และสองมิติสองทิศทาง (2-Dimensional 2-directional LDA: $(2D)^2LDA$) เช่นเดียวกัน ดังรายละเอียดต่อไปนี้

Turk และ Pentland [1] นำเสนอวิธี Eigenface ในปี ค.ศ. 1991 เป็นการฉายมิติภาพ (Image space) ไปสู่มิติคุณลักษณะ (Feature space) แบบเชิงเส้นที่มีขนาดเล็กกว่า โดยอาศัยหลักการ PCA ในการลดมิติ โดยคำนวณทิศทางการฉาย (Projection direction) จากค่ามากที่สุดของการกระจายรวม (Total scatter) ทุกกลุ่มข้อมูล

ต่อมา Belhumeur และคณะ [2] เสนอวิธี Fisherfaces ในปี ค.ศ. 1997 ที่เจาะจงกลุ่มข้อมูล โดยคำนวณการฉายที่เหมาะสม (Optimal projection) จากอัตราส่วนมากที่สุดระหว่างเมตริกซ์การกระจายระหว่างกลุ่มกับเมตริกซ์การกระจายภายในกลุ่มซึ่งมีประสิทธิภาพดีกว่าวิธี Eigenfaces

สำหรับวิธี LDA ก็มีการพัฒนาอย่างต่อเนื่องระหว่างปี ค.ศ. 2000-2007 งานวิจัย [3]-[9] มีการเผยแพร่งานวิจัย LDA ใหม่แบบต่างๆ เพื่อแก้ปัญหาตัวอย่างในกลุ่มข้อมูลมีจำนวนน้อย ที่เรียกว่า Small Sample Size (SSS) problem นั่นคือ กรณีที่จำนวนตัวอย่างฝึกฝน (Training samples) น้อยกว่าขนาดหรือมิติคุณลักษณะ (Feature dimension) ของเวกเตอร์ที่ใช้ทดสอบ ซึ่งมีผลให้เมตริกซ์การกระจายภายในกลุ่มของตัวอย่างอาจจะเป็นเมตริกซ์เอกฐาน (Singular matrix) ทำให้กระบวนการประมวลผลต้องมีการคำนวณที่ยุ่งยากและซับซ้อนขึ้น

งานวิจัย [10]-[12] เสนอวิธีการ 2DPCA ในปี ค.ศ. 2004 และ 2007 ซึ่งใช้หลักการเมตริกซ์สองมิติแทนเวกเตอร์ 1 มิติ โดยไม่ต้องมีการแปลงเมตริกซ์ภาพให้เป็นเวกเตอร์ภาพก่อน ทำการคำนวณเมตริกซ์ความแปรปรวนร่วม (Covariance matrix) ของภาพเริ่มต้นโดยตรง ขนาดของเมตริกซ์ความแปรปรวนร่วมของภาพด้วยวิธี 2DPCA นี้จึงเล็กกว่าวิธี PCA ทำให้ 2DPCA ง่ายต่อการคำนวณเมตริกซ์ความแปรปรวนร่วม และใช้เวลาในการคำนวณน้อยกว่าแบบ PCA ส่วนงานวิจัย [13]-[17] ในปี ค.ศ. 2005 และ 2007 นำเสนอรูปแบบ 2DLDA ที่เมตริกซ์การกระจายมีขนาดเล็กกว่าแบบ LDA และ PCA ยิ่งกว่านั้นการคำนวณ Eigenvector ก็ทำได้ง่ายกว่า

สำหรับการพัฒนา $(2D)^2$ PCA และ $(2D)^2$ LDA เริ่มในปี ค.ศ. 2005 โดย Zhang และ Zhou [18] เสนอวิธี $(2D)^2$ PCA กำหนดให้ 2DPCA ทำงานเฉพาะในทิศทางข้อมูลด้านแถวของภาพ และเสนอ 2DPCA อีกแนวทางหนึ่ง (Alternative 2DPCA) ให้ทำงานเฉพาะในทิศทางข้อมูลด้านคอลัมน์ของภาพ ผลการทดสอบสามารถเพิ่มความถูกต้องในการรู้จำได้โดยใช้มิติและเวลาน้อยกว่าวิธี PCA และ 2DPCA

ต่อมาในปี ค.ศ. 2006 มีงานวิจัยลักษณะเดียวกันเผยแพร่อีกสองผลงาน งานวิจัยแรกคือ Nagabhushan และคณะ [19] เสนอวิธีการ $(2D)^2$ FLD ส่วนงานวิจัยที่สอง Noushath และคณะ [20] เสนอวิธีการ $(2D)^2$ LDA ผลการทดสอบเป็นไปในแนวทางเดียวกันคือ มีอัตราความถูกต้องในการรู้จำที่ดีกว่าด้วยมิติของเวกเตอร์คุณลักษณะและใช้เวลาในการประมวลผลน้อยกว่าเมื่อเทียบกับวิธีแบบเดิม

งานวิจัยเพื่อเพิ่มประสิทธิภาพ LDA อีกแนวทางหนึ่งคือ การเพิ่มฟังก์ชันการถ่วงน้ำหนัก (Weighting function) ซึ่งเริ่มมีการเผยแพร่ผลงานวิจัยตั้งแต่ปี ค.ศ. 2000 การปรับปรุงประสิทธิภาพของ LDA ด้วยวิธีนี้ จะประยุกต์เพิ่มฟังก์ชันการถ่วงน้ำหนักที่เป็นระยะยูคลิเดียน (Euclidean distance) หรือระยะมาฮาลานอบิส (Mahalanobis distance) ในเมตริกซ์การกระจายระหว่างกลุ่ม [21]-[25] เพื่อปรับค่าตัวอย่างข้อมูลแต่ละกลุ่มให้แตกต่างกันมากขึ้น ต่อมาจึงพัฒนาฟังก์ชันการถ่วงน้ำหนักในเมตริกซ์การกระจายภายในกลุ่มเพิ่มเติม เพื่อปรับค่าตัวอย่างข้อมูลภายในกลุ่มให้มีการกระจายน้อยลงเรียกว่า ฟังก์ชันการถ่วงน้ำหนักสัมพัทธ์ (Relevance Weighting function) [26]-[29] และมีการประยุกต์ฟังก์ชันการถ่วงน้ำหนักสัมพัทธ์กับวิธี 2DLDA ในงานวิจัยล่าสุด [30]-[31]

จากการอ้างอิงงานวิจัยเกี่ยวกับการประยุกต์ใช้ฟังก์ชันการถ่วงน้ำหนัก [26]-[31] ผู้วิจัยจึงมีแนวคิดในการตั้งสมมุติฐานในการทำวิจัยได้ว่า วิธีการของ LDA เป็นเทคนิคการฉายภาพที่พัฒนามาจนถึงแบบสองมิติสองทิศทาง หรือ $(2D)^2$ LDA ตามงานวิจัย [19] และ [20] สามารถประยุกต์เพิ่มฟังก์ชันการถ่วงน้ำหนักเข้าไปในวิธีการดังกล่าวเป็นแนวทางใหม่ ที่สามารถเพิ่มประสิทธิภาพสมรรถนะการรู้จำใบหน้าให้ดีขึ้น จึงนำเสนอเป็นงานวิจัยต่อไป

1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย

1.4.1 การวิเคราะห์ลักษณะแตกต่างเชิงเส้น (Linear discriminant analysis: LDA)

1.4.2 การวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติ (2-Dimensional LDA: 2DLDA)

- 1.4.3 การวิเคราะห์หลักขณะแตกต่างเชิงเส้นสองมิติสองทิศทาง (2-Dimensional and 2-directional LDA: $(2D)^2LDA$)
- 1.4.4 การหาระยะยูคลิดีเนียน (Euclidean distance)
- 1.4.5 การหาระยะมาฮาลาโนบิส (Mahalanobis distance)
- 1.4.6 การวิเคราะห์หลักขณะแตกต่างเชิงเส้นสองมิติสองทิศทางแบบถ่วงน้ำหนัก (Weighted $(2D)^2LDA$)
- 1.4.7 การวิเคราะห์หลักขณะแตกต่างเชิงเส้นสองมิติสองทิศทางแบบถ่วงน้ำหนักสัมพัทธ์ (Relevance weighted $(2D)^2LDA$)

1.5 การเปรียบเทียบระหว่างวิธีการที่นำเสนอกับวิธีการแบบพื้นฐาน

วิธี PCA เป็นการฉายมิติภาพไปสู่มิติคุณลักษณะแบบเชิงเส้น ซึ่งลดมิติให้มีขนาดเล็กลง ทิศทางการฉายคำนวณจากค่ามากที่สุดของการกระจายรวมทุกกลุ่มข้อมูล วิธี PCA จึงมีข้อจำกัดที่ต้องมีการแปลงภาพใบหน้าสองมิติ (2D face images) ให้เป็นเวกเตอร์ภาพ 1 มิติ (1D image vectors) ก่อนเวกเตอร์ภาพจึงมีขนาดใหญ่ ทำให้ยุ่งยากในการคำนวณเมตริกซ์ความแปรปรวนร่วม จึงมีการพัฒนาวิธี 2DPCA เพื่อแก้ปัญหาความยุ่งยากในการคำนวณเนื่องจากเวกเตอร์ภาพของวิธี PCA มีขนาดใหญ่ เมตริกซ์ความแปรปรวนร่วมของภาพในวิธี 2DPCA เล็กกว่าวิธี PCA ทำให้วิธี 2DPCA คำนวณได้ถูกต้องและใช้เวลาในการคำนวณน้อยกว่า

ส่วนวิธี LDA คำนวณแกนการฉายที่เหมาะสมจากอัตราส่วนมากที่สุดระหว่างเมตริกซ์การกระจายระหว่างกลุ่มกับเมตริกซ์การกระจายภายในกลุ่ม ซึ่งมีประสิทธิภาพดีกว่าวิธี PCA และมีการพัฒนาไปสู่วิธี 2DLDA เช่นเดียวกัน โดยวิธี 2DLDA คำนวณ Eigenvector ของเมตริกซ์ความแปรปรวนร่วมโดยตรง ขนาดเมตริกซ์ความแปรปรวนร่วมเท่ากับความกว้างของภาพ ซึ่งเล็กมากเมื่อเทียบกับวิธี LDA ทำให้การคำนวณในวิธี 2DLDA ถูกต้องและมีประสิทธิภาพมากกว่าวิธี LDA ดังนั้น แนวทางการวิเคราะห์แบบสองมิติจึงเป็นแนวทางที่น่าสนใจ

วิธี 2DPCA และ 2DLDA คำนวณ Eigenvectors ของเมตริกซ์ความแปรปรวนร่วมของภาพโดยตรง ไม่มีการแปลงเมตริกซ์ภาพให้เป็นเวกเตอร์ ขนาดของเมตริกซ์ความแปรปรวนร่วมของภาพเท่ากับความกว้างของภาพ ซึ่งเล็กมากเมื่อเทียบกับวิธี PCA ทำให้คำนวณได้ถูกต้องมากกว่า แต่วิธี 2DPCA ก็มีข้อเสียที่ต้องการจำนวนสัมประสิทธิ์ในการแทนภาพมากกว่า เช่น ภาพขนาด 100×100 พิกเซล จำนวนสัมประสิทธิ์ของวิธี 2DPCA คือ $100 \times d$ พิกเซล ซึ่งค่า d ปกติจะกำหนดให้ไม่เกิน 5 สำหรับความถูกต้องที่พอใช้งานได้ แต่แม้ว่าปัญหานี้สามารถแก้ไขได้โดยใช้วิธี PCA ต่อจากวิธี 2DPCA สำหรับการลดมิติต่อไป แต่ก็ยังไม่ชัดเจนว่า มิติของวิธี 2DPCA จะลดลงอย่างที่ต้องการได้อย่างไร จึงมีการพัฒนาไปสู่วิธี $(2D)^2PCA$, $(2D)^2FLD$ และ $(2D)^2LDA$ ให้มีการทำงานกับข้อมูลทั้งด้านแถวและคอลัมน์ของภาพเป็นแบบสองมิติและสองทิศทาง (2-Dimensional and 2-directional) จะได้รับความถูกต้องในการ

รู้จำดีกว่าหรือใกล้เคียง แต่ใช้จำนวนสัมประสิทธิ์ในการแทนภาพน้อยกว่า ทำให้ประสิทธิภาพในการคำนวณดีกว่าวิธี 2DPCA และ 2DLDA อีกด้วย

งานวิจัยที่ผ่านมาแสดงให้เห็นว่า มีเพียงการประยุกต์เพิ่มฟังก์ชัน Weighting และ Relevance Weighting function กับวิธี LDA และ 2DLDA เท่านั้น ส่วนวิธี LDA ก็ถูกพัฒนาเป็น $(2D)^2LDA$ ที่ประกอบด้วยสองวิธีรวมกันคือ 2DLDA และ Alternative 2DLDA จากจุดนี้ เมื่อนำหลักการ Relevance Weighting function มาประยุกต์ใช้กับ $(2D)^2LDA$ จึงเป็นการนำเสนอหลักการใหม่ที่เรียกว่า การวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติสองทิศทางแบบถ่วงน้ำหนักสัมพัทธ์ (Relevance weighted $(2D)^2LDA$: RW $(2D)^2LDA$) ซึ่งจะแสดงรายละเอียดและผลการทดสอบในวิทยานิพนธ์ฉบับนี้ต่อไป

1.6 ขอบเขตการวิจัย

วิทยานิพนธ์ฉบับนี้นำเสนอ วิธีการสร้างระบบการรู้จำใบหน้า โดยใช้เทคนิคการฉายภาพตามหลักการวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติและสองทิศทาง ที่มีการวิเคราะห์ข้อมูลภาพทั้งด้านแนวและคมลึมน้ของภาพ โดยประยุกต์เพิ่มฟังก์ชันการถ่วงน้ำหนักในเมตริกซ์การกระจายระหว่างกลุ่มและฟังก์ชันการถ่วงน้ำหนักสัมพัทธ์ในเมตริกซ์การกระจายภายในกลุ่ม สำหรับข้อมูลด้านแนวและคมลึมน้ของภาพนั้น ทั้งนี้มีการเลือกใช้ตัวถ่วงน้ำหนักแบบระยะยูคลิเดียนเพื่อเปรียบเทียบประสิทธิภาพความถูกต้องในการแยกแยะใบหน้า ทดสอบโดยใช้คอมพิวเตอร์ในการจำลองระบบด้วยโปรแกรม MATLAB ผลที่ได้แสดงสมรรถนะของอัตราความถูกต้องในการรู้จำเมื่อทดสอบกับฐานข้อมูลที่เป็นที่ยอมรับ ภายใต้เงื่อนไขต่างๆ ได้แก่ ความสว่างภาพ มุมของใบหน้า การเปลี่ยนขนาดและการเอียงใบหน้า ภาพใบหน้าที่ระยะเวลาต่างกัน การบังใบหน้าและการแสดงอารมณ์ทางใบหน้า โดยเปรียบเทียบกับวิธีการแบบพื้นฐาน

1.7 ขั้นตอนของการศึกษา

วิทยานิพนธ์ฉบับนี้แบ่งเนื้อหาออกเป็น 5 บทด้วยกันคือ

บทที่ 1 กล่าวถึงความเป็นมาของงานวิจัย ความมุ่งหมายและวัตถุประสงค์ สมมติฐานของการศึกษา ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย การเปรียบเทียบระหว่างวิธีการที่นำเสนอกับวิธีการแบบพื้นฐาน ขอบเขตของการวิจัย และขั้นตอนของการศึกษา

บทที่ 2 กล่าวถึงปัจจัยต่างๆ ที่มีผลต่อระบบการรู้จำใบหน้า พื้นฐานการรู้จำใบหน้า วิธีที่ใช้ในการรู้จำใบหน้า และฐานข้อมูลแบบต่างๆ ที่นิยมใช้ในการทดสอบกับระบบการรู้จำใบหน้า

บทที่ 3 กล่าวถึงทฤษฎีพื้นฐานของการวิเคราะห์ลักษณะแตกต่างเชิงเส้น (LDA) ตลอดจนการวิเคราะห์ลักษณะแตกต่างเชิงเส้นแบบสองมิติ (2DLDA) สุดท้ายกล่าวถึง การวิเคราะห์ลักษณะแตกต่างเชิงเส้นแบบสองมิติและสองทิศทาง ($(2D)^2LDA$) และการประยุกต์ฟังก์ชันการถ่วงน้ำหนักในเทคนิคการฉายภาพแบบต่างๆ ได้แก่ วิธี LDA, 2DLDA และ $(2D)^2LDA$

บทที่ 4 กล่าวถึงการทดสอบหาค่าสมรรถนะของระบบการรู้จำใบหน้ากับฐานข้อมูลภาพใบหน้า และพารามิเตอร์ต่างๆ ที่มีผลต่อสมรรถนะของระบบ และผลที่ได้จากการจำลองระบบ เพื่อแสดงให้เห็นว่าวิธีการที่นำเสนอสามารถช่วยให้ระบบมีสมรรถนะที่ดีขึ้น

บทที่ 5 กล่าวถึงบทสรุปผลการวิจัย ข้อเสนอแนะ และแนวทางการวิจัยที่สามารถดำเนินการต่อไปในอนาคต

บทที่ 2

ระบบการรู้จำใบหน้า

เทคโนโลยีการพิสูจน์ตัวตนทางชีวมิติ (Biometrics) หรือไบโอเมตริก ที่จัดว่าธรรมดาที่สุดคือ ลายนิ้วมือ (Fingerprint) เพราะทำได้ง่ายในทางปฏิบัติ วิธีที่มีประสิทธิภาพดีที่สุดแต่ทำได้ยากคือ ม่านตา (Iris) เนื่องจากดวงตาเป็นอวัยวะที่สำคัญ บุคคลทั่วไปจึงไม่ยอมรับวิธีการสแกนม่านตา ฐานข้อมูลม่านตาเพื่อใช้ทดสอบมีน้อย ยังมีคุณลักษณะทางร่างกายมนุษย์อื่นๆ ที่มีการศึกษาวิจัย เช่น ลักษณะทางเรขาคณิตของลายนิ้วมือหรือฝ่ามือ (Finger/palm geometry) เสียง (Voice) ลายเซ็นต์ (Signature) หรือใบหน้า (Face) เป็นต้น ไบโอเมตริกที่นิยมใช้งานเชิงพาณิชย์ในช่วงหลายปีที่ผ่านมา แต่ละแบบมีข้อดีและข้อเสียแตกต่างกัน ได้แก่ การรู้จำม่านตามีความแม่นยำสูง แต่ราคาในการจัดทำระบบสูง และไม่เป็นที่ยอมรับของผู้ใช้งาน ลายนิ้วมือมีความน่าเชื่อถือและป้องกันการบุกรุกได้ดีในระดับหนึ่ง แต่ไม่เหมาะสำหรับผู้ที่ไม่ให้ความร่วมมือ (Non-collaborative individuals) ส่วนการรู้จำใบหน้าเป็นระบบที่มีความน่าเชื่อถือและการยอมรับของสังคม (Reliability and social acceptance) ที่เหมาะสม

2.1 ปัจจัยที่มีผลต่อการรู้จำใบหน้า

ระยะ 10 ปีที่ผ่านมา เทคโนโลยีการรู้จำใบหน้ามีความก้าวหน้าไปมาก ความสามารถของระบบให้อัตราการรู้จำที่ถูกต้องมากกว่า 90 เปอร์เซ็นต์ แต่ในทางปฏิบัติยังมีปัจจัยที่เป็นอุปสรรค และมีผลต่อความถูกต้องในการรู้จำหลายประการ เนื่องจากกระบวนการได้มาของภาพใบหน้าขึ้นอยู่กับเงื่อนไขการเปลี่ยนแปลงที่หลากหลาย ตัวแปรที่มีผลต่อความสามารถของระบบการรู้จำใบหน้าที่สำคัญมีอยู่ 5 ประการ ดังนี้

2.1.1 ความสว่าง (Illumination)

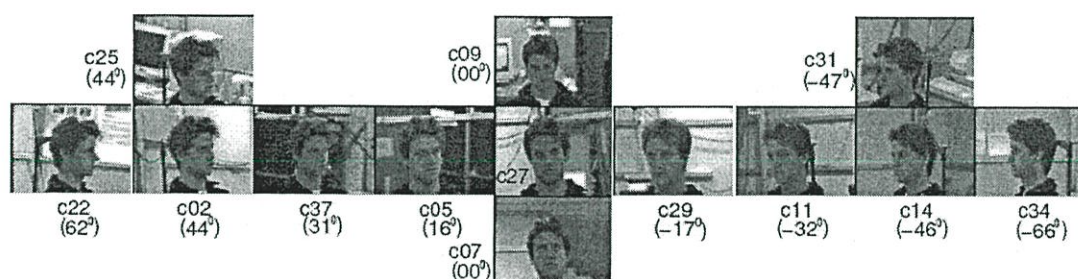
การเปลี่ยนแปลงความสว่างของภาพใบหน้าเป็นปัญหาที่ยุงยากที่สุดประการหนึ่ง เกิดจากการสะท้อนของผิวหนังจากสภาพแวดล้อมที่มีแสงสว่างแตกต่างกัน ตลอดจนระบบควบคุมภายในของกล้องที่ใช้ถ่ายภาพ ภาพที่แตกต่างกันเนื่องจากการเปลี่ยนแปลงของแสงมีผลต่อการรู้จำมากกว่าความแตกต่างเนื่องจากลักษณะของบุคคล ส่วนการเปลี่ยนทิศทางของแสงก็มีผลต่อการกระจายสเกลสีเทาของภาพใบหน้า การรู้จำใบหน้าสองมิติ (2-Dimensional method) ส่วนมาก มีอัตราการรู้จำที่ต่ำกว่าที่เงื่อนไขที่ไม่มีการเปลี่ยนแปลงความสว่างของภาพใบหน้ามากนัก ขณะที่ความสามารถในการรู้จำจะลดลงเมื่อเกิดการเปลี่ยนแปลงความสว่างและท่าทางของภาพใบหน้า รูปที่ 2.1 แสดงตัวอย่างภาพใบหน้าที่มีการเปลี่ยนแปลงความสว่างของภาพจากฐานข้อมูล CMU Pose Illumination Expression (PIE) โดยแถวบนเป็นภาพใบหน้าที่ถ่ายในห้องที่เปิดไฟสว่าง ส่วนแถวกลางถ่ายขณะปิดไฟ มีเพียงแสงจากแฟลชของกล้องถ่ายรูปเท่านั้น



รูปที่ 2.1 ตัวอย่างภาพความสว่างใบหน้า

2.1.2 มุมของใบหน้า (Pose)

ในทางปฏิบัติ มุมของใบหน้ามีผลต่อกระบวนการพิสูจน์ตัวตน เพราะภาพใบหน้าที่รวบรวมได้จากมุมกล้องต่างๆ กัน ทั้งมุมเอียงซ้าย-เอียงขวา และมุมก้ม-มุมเงย จะทำให้การฉายภาพ (Image projection) ผิดเพี้ยน เพราะถ้ามีการหมุนศีรษะไป 32 องศา ก็จะได้ภาพเอียงด้านข้าง ที่จะบังใบหน้าตนเองไปส่วนหนึ่ง และการใช้งานจริงกล้องที่ใช้ในระบบความปลอดภัยส่วนใหญ่จะถูกติดตั้งไว้คงที่เพื่อเก็บภาพใบหน้าบุคคลที่เดินผ่านกล้อง ภาพที่ได้จึงมีมุมแตกต่างกันหลากหลาย ซึ่งบางครั้งอยู่นอกขอบเขตที่กำหนดไว้ ทำให้ประสิทธิภาพการรู้จำลดลงไปมาก งานวิจัยที่ [34] - [35] นำเสนอผลการวิจัยว่า ภาพมุมด้านหน้าจะให้อัตราความถูกต้องในการรู้จำที่ดีกว่าภาพด้านอื่นๆ โดยที่อัตราความถูกต้องที่ยอมรับได้สูงกว่า 90% จะได้จากภาพที่มีมุมของใบหน้าเอียงไม่เกิน 32 องศา

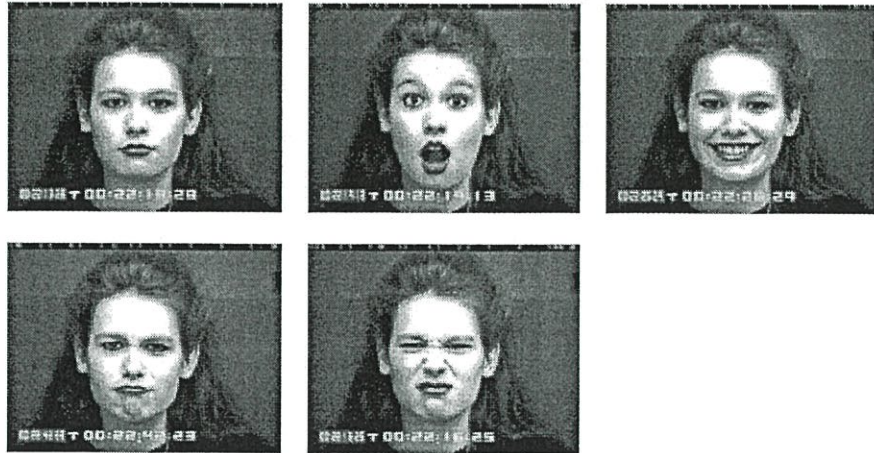


รูปที่ 2.2 ตัวอย่างภาพใบหน้าที่มุมต่างๆ

2.1.3 การแสดงอารมณ์ทางใบหน้า (Facial expression)

อัลกอริทึมที่ใช้กับการรู้จำใบหน้าสัมพันธ์กับการแสดงอารมณ์ทางใบหน้าโดยอติโนมัติ ยกเว้นโครงสร้างที่ผิดธรรมชาติอย่างมาก ไม่เหมือนกับมุมของใบหน้า ลักษณะท่าทางและสีหน้าที่เกิดจากอารมณ์ขณะการสื่อสารด้วยการพูดทำให้เกิดการเปลี่ยนแปลงที่ปรากฏบนใบหน้าอย่างมาก มีการรายงานว่าการเปลี่ยนแปลงการแสดงอารมณ์ทางใบหน้าที่เป็นไปได้มีจำนวนหลายพันรูปแบบ

สำหรับอัลกอริทึมที่ใช้หลักการทางโครงสร้าง (Structure-based algorithm) อาจจะได้ผลของการแสดงอารมณ์ทางใบหน้าในการรู้จำมากกว่าอัลกอริทึมแบบองค์รวม (Holistic-based algorithm)



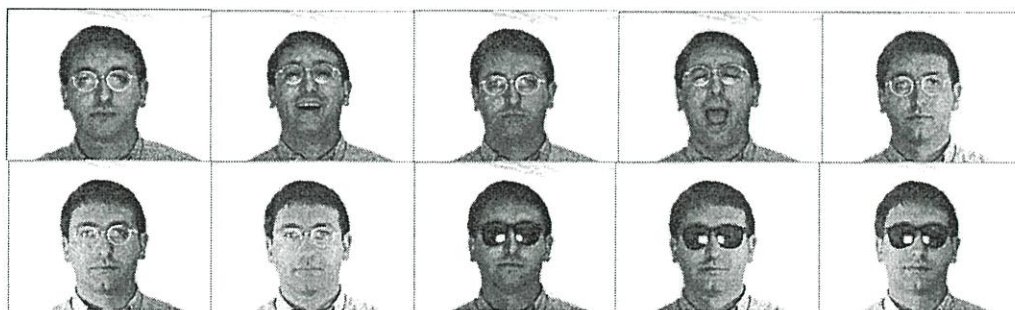
รูปที่ 2.3 ตัวอย่างภาพการแสดงอารมณ์ทางใบหน้า

2.1.4 การหน่วงเวลา (Time delay)

การหน่วงเวลาเป็นตัวแปรที่สำคัญอีกตัวแปรหนึ่ง เพราะเมื่อเวลาผ่านไป ใบหน้าและส่วนประกอบอื่นๆ จะเปลี่ยนไปแบบไม่เป็นเชิงเส้น (Non-linear)



(ก)



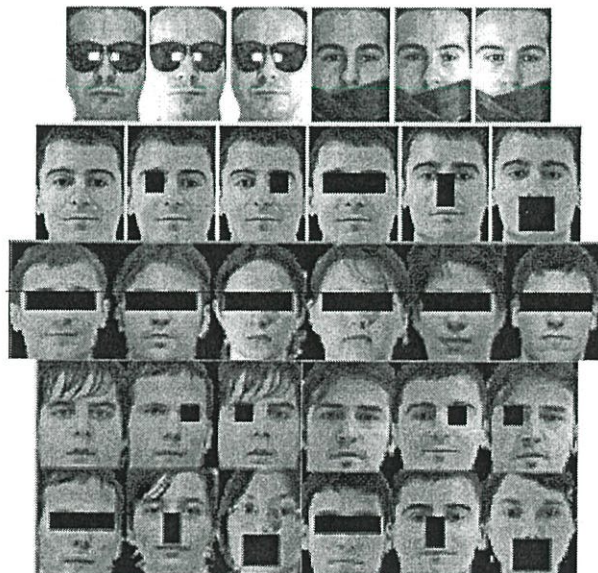
(ข)

รูปที่ 2.4 ตัวอย่างภาพใบหน้าทีเวลาต่างกัน (ก) ภาพสัปดาห์แรก (ข) ภาพสัปดาห์ที่สาม

ตัวอย่างสำหรับการเปลี่ยนแปลงที่ขึ้นกับเวลาได้แก่ ทรงผม การแต่งหน้า กล้ามเนื้อใบหน้า สิว หน้า ผมหัน ขมขยว ใส่แว่นหรือเครื่องประดับบนใบหน้า ดังนั้น จึงเป็นตัวแปรหนึ่งที่สำคัญ และ ต้องมีการทดสอบอัลกอริทึมที่ออกแบบเมื่อต้องการทราบประสิทธิภาพของการรู้จำใบหน้า โดยคำนึงถึง เวลาที่เปลี่ยนไปเป็นหลัก เมื่อเทียบกับตัวแปรอื่นๆ ปัญหานี้นับได้ว่า เป็นปัญหาที่แก้ไขได้ยากกว่า แต่ ปัญหาเรื่องเวลานี้ยังแก้ไขได้ง่ายกว่าปัญหาการเปลี่ยนแปลงของอายุ ตัวอย่างภาพใบหน้าที่รวบรวมจาก เวลาที่ต่างกันแสดงในรูปที่ 2.4 เป็นภาพใบหน้าจากฐานข้อมูล AR (Alex Martinez and Robert Benavente) ประกอบด้วยภาพบุคคลคนเดียวกัน 2 ชุด คือ ชุด ก. ซึ่งถ่ายในระยะเวลาต่างกับชุด ข. 2 สัปดาห์ และแต่ละชุดประกอบด้วยภาพที่มีความสว่าง การใส่แว่น และการแสดงอารมณ์ทางใบหน้าที่แตกต่างกัน

2.1.5 การบังใบหน้า (Occlusions)

การบังใบหน้าเป็นตัวแปรสุดท้ายที่มีผลต่อความสามารถในการรู้จำใบหน้าอย่างมาก โดยเฉพาะ ถ้ามีการบังด้านบนของใบหน้าโดยตรง ใบหน้าแต่ละบุคคลอาจจะสวมใส่แว่นต่างๆ ทำให้มีการบัง ใบหน้า เช่น แว่นตา หรือผ้าโพกหน้า รวมทั้งเครื่องประดับ หรือของใช้ส่วนตัว การบังใบหน้าอาจเกิด โดยความตั้งใจ หรือไม่ตั้งใจก็ได้ ภายใต้เงื่อนไขบางประการอาจถูกจูงใจเพื่อขัดขวางความพยายาม รู้จำด้วยการปิดบังขอบเขตของใบหน้า ในสถานการณ์หลากหลาย ขึ้นอยู่กับเป้าหมายการรู้จำว่าจะรวม บังจี้เหล่านี้หรือไม่ ถ้ารวมบังจี้การบังใบหน้าที่ก็มีความยุ่งยากมากขึ้นในกระบวนการรู้จำ การบัง ใบหน้าบุคคลทำให้ปริมาณและคุณภาพการรู้จำเปลี่ยนไป เท่าที่มีการทดสอบกับการรู้จำใบหน้าที่มีการ กำหนดเงื่อนไขการบังใบหน้าเป็นรูปแบบต่างๆ กัน เช่น ปิดตาข้างหนึ่ง และปิดส่วนล่างของใบหน้า ใส่แว่นตา ผ้าปิดหน้า หรือปิดจมูก ปิดปาก เป็นต้น ดังแสดงตัวอย่างภาพเหล่านี้ในรูปที่ 2.5



รูปที่ 2.5 ตัวอย่างภาพการบังใบหน้าแบบต่างๆ

ปัจจัยที่เป็นอุปสรรคต่อสมรรถนะของระบบการรู้จำใบหน้าเหล่านี้ ได้ถูกนำไปทดสอบกับวิธีการที่นำเสนอในงานวิจัยนี้ในภาพรวมทั้งหมด โดยคัดเลือกภาพใบหน้าที่มีเงื่อนไขแต่ละปัจจัยที่มีอยู่ในฐานข้อมูลต่างๆ มาทำการทดสอบหาสมรรถนะของระบบการรู้จำใบหน้าตามวิธีใหม่ที่น่าสนใจเปรียบเทียบกับวิธี (2D)²LDA แบบเดิม

2.2 อัลกอริทึมการรู้จำใบหน้า (Face recognition algorithms)

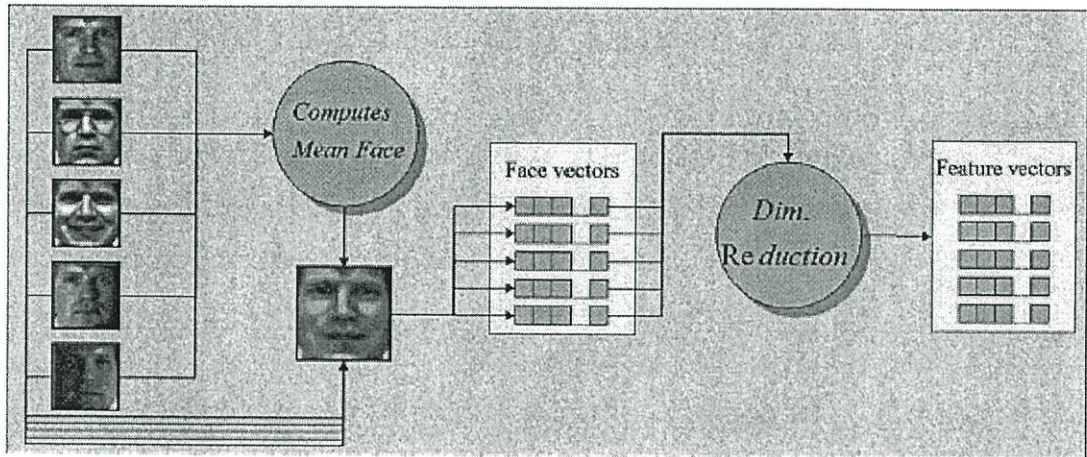
การรู้จำใบหน้าก็คือ การรู้จำรูปแบบ (Pattern recognition) ที่มีความยุ่งยากในการแก้ปัญหา เนื่องจากการวิเคราะห์ข้อมูลของภาพใบหน้าภายใต้เงื่อนไข หรือปัจจัยบางอย่างจะพบปัญหาที่ไม่เป็นเชิงเส้น (Nonlinearity problem) รวมอยู่ด้วย กล่าวในรายละเอียดคือ เป็นปัญหาในการเปรียบเทียบรูปแบบ (Template matching problem) ที่จะต้องมีการประมวลผลการรู้จำกับข้อมูลในปริภูมิที่มีมิติขนาดใหญ่ (High-dimensional space) นั่นเอง ทำให้ต้องใช้เวลามาก และมีขั้นตอนการคำนวณที่ซับซ้อน เพื่อให้ได้ผลการทดสอบเปรียบเทียบที่ตรงกัน ดังนั้น การวิจัยด้านนี้จึงมีอัลกอริทึมในการทดสอบ และการคัดแยกผลการรู้จำที่แตกต่างกัน หลักการสำคัญก็คือ การหาคุณลักษณะเด่น (Feature extraction) เป็นสำคัญ เพราะมีผลต่อประสิทธิภาพ และสมรรถนะของระบบการรู้จำใบหน้าโดยตรง อัลกอริทึมที่นิยมใช้ในงานวิจัยด้านนี้สามารถจำแนก และกล่าวโดยสังเขปได้ดังต่อไปนี้

2.2.1 วิธีการฉายภาพแบบเชิงเส้นและไม่เป็นเชิงเส้น (Linear/nonlinear projection methods)

เนื่องจากส่วนใหญ่การรู้จำใบหน้าอัตโนมัติเป็นปัญหาการรู้จำรูปแบบ แบบไม่เป็นเชิงเส้น และยากที่จะหาวิธีแก้ไข หรืออาจจำแนกได้ว่า เป็นปัญหาด้าน Template matching ซึ่งการรู้จำต้องทำในปริภูมิที่มีมิติขนาดใหญ่ การแมทซ์รูปแบบต้องใช้เวลาในการคำนวณมาก เทคนิคการลดมิติ (Dimensionality reduction technique) จึงถูกนำมาใช้เพื่อถ่ายถอดปัญหาไปสู่ปริภูมิที่มีมิติเล็กกว่า วิธีการใบหน้าไอเกน (Eigenfaces) จึงถูกพิจารณาเป็นแนวทางแรกๆ ด้วยหลักการนี้ ภาพ I ที่มีขนาด $N \times N$ ถูกแปลงให้เป็นเวกเตอร์ขนาด N^2 อย่างเป็นเชิงเส้น เพื่อแทนภาพนั้นด้วยจุดข้อมูลในปริภูมิที่มีมิติ N^2 อย่างไรก็ตาม การเปรียบเทียบไม่ได้กระทำในปริภูมิเหล่านี้ แต่จะกระทำในปริภูมิที่เล็กกว่าด้วยเทคนิคการลดมิติ ดังแสดงในรูปที่ 2.6

การฉายภาพแบบเชิงเส้นที่นิยมใช้มี 2 วิธี คือ การวิเคราะห์องค์ประกอบหลัก (Principal component analysis: PCA) และการวิเคราะห์ลักษณะแตกต่างเชิงเส้น (Linear discriminant analysis: LDA) วิธี PCA หลังจากการแปลงแบบเชิงเส้นให้เป็นเวกเตอร์เรียบร้อยแล้ว จะคำนวณเวกเตอร์เฉลี่ยของทุกภาพ และนำไปลบออกจากทุกเวกเตอร์ของภาพใบหน้าเริ่มต้น จากนั้นคำนวณหาเมตริกซ์ความแปรปรวนร่วม (Covariance matrix) เพื่อคัดแยกหาจำนวนจำกัดของไอเกนเวกเตอร์ (Eigenvectors) ของมัน ซึ่งตรงกับค่าไอเกนที่มากที่สุด (Greatest eigenvalues) ไอเกนเวกเตอร์นี้จะอ้างถึงใบหน้าไอเกน และเป็นฐานข้อมูลที่ใช้ในปริภูมิที่มีมิติต่ำกว่าด้วย เมื่อมีการทดสอบกับภาพใหม่ที่เข้ามา ก็จะคำนวณค่าใบหน้าไอเกนและ

เปรียบเทียบกับฐานข้อมูลที่มีอยู่ทั้งหมด ด้วยการวัดค่าระยะทาง (Distance measure) ซึ่งปกติจะใช้ระยะยูคลิเดียน (Euclidean distance)



รูปที่ 2.6 การลดมิติภาพตามวิธีการฉายภาพแบบเชิงเส้นและไม่เป็นเชิงเส้น

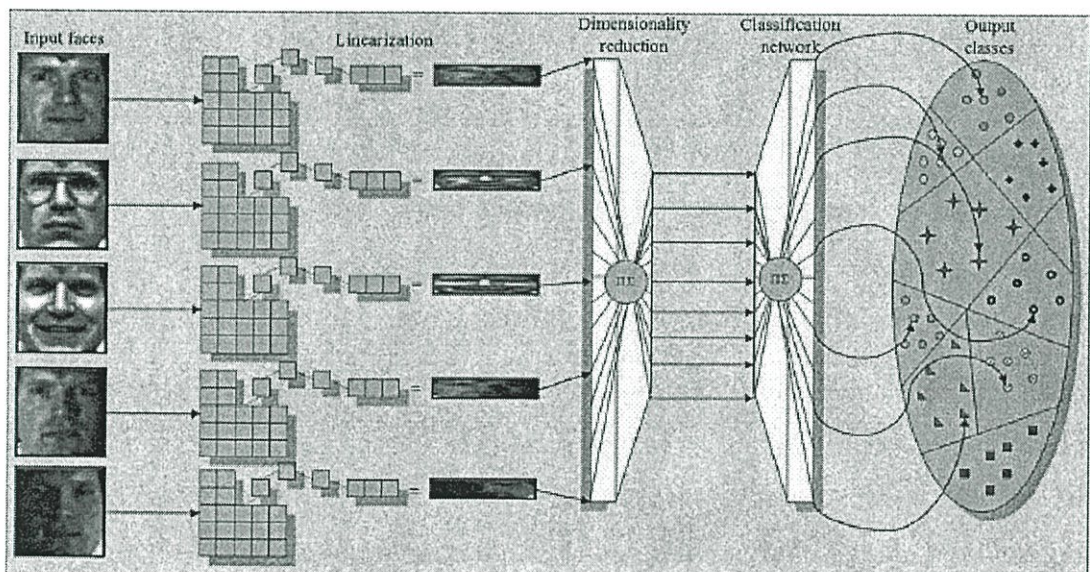
วิธี PCA ถูกนำมาประยุกต์ใช้อย่างมากกับการรู้จำใบหน้า แต่ก็มีวิธีการฉายภาพแบบเชิงเส้นอื่นๆ อีกมากที่มีการศึกษาวิจัยกันอยู่ ส่วน LDA [5] - [6] ดีกว่าวิธี PCA ซึ่งคำนวณหาความแตกต่างข้อมูลระหว่างกลุ่ม ขณะที่ PCA เกี่ยวกับข้อมูลอินพุทโดยตรง ไม่มีการวิเคราะห์โครงสร้างภายในวัตถุประสงค์หลักของ LDA คือ การหาเวกเตอร์พื้นฐานที่มีการคัดแยกกลุ่มข้อมูลที่ตีที่สุด โดยใช้หลักการความแตกต่างข้อมูลระหว่างกลุ่มมากที่สุด (Maximize between-classes difference) และความแตกต่างข้อมูลภายในกลุ่มเดียวกันน้อยที่สุด (Minimize within-classes difference)

สำหรับวิธีไม่เป็นเชิงเส้นมีงานวิจัยแรกที่น่าสนใจในปี ค.ศ. 2000 Sam และ Lawrence [36] แนะนำวิธี Locally Linear Embedding (LLE) ซึ่งเป็นอัลกอริทึมการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) ที่คำนวณมิติขนาดเล็ก (Low-dimensional) จากวิธีการฝังตัวเพื่อรักษาข้อมูลใกล้เคียงของอินพุทที่มีมิติขนาดใหญ่ โดย LLE จะทำการแมปอินพุทให้เป็นระบบพิกัดเดี่ยวโดยรวมกว้างๆ ที่มีมิติเล็กลง โดยค่าที่เหมาะสมของมันจะไม่เกี่ยวกับค่า Local ที่น้อยที่สุด การใช้ข้อดีของการสมมาตรภายในของการสร้างใหม่แบบเชิงเส้นทำให้ LLE สามารถเรียนรู้โครงสร้างภายนอกโดยรวมของความไม่เป็นเชิงเส้นได้หลากหลาย เช่น ใช้ในการสร้างภาพใบหน้า หรือสร้างเอกสารต้นฉบับ ต่อมาในปีเดียวกัน Tenenbaum และคณะ [37] เสนอแนวทางแก้ปัญหาการลดมิติโดยใช้การวัดข้อมูลเมตริกข้ออย่างง่าย เพื่อเรียนรู้ลักษณะสำคัญทางเรขาคณิตโดยรวมที่ซ่อนอยู่ของชุดข้อมูล ซึ่งไม่เหมือนกับวิธี PCA แนวทางนี้สามารถหาลำดับที่ไม่เป็นเชิงเส้นได้อย่างอิสระ ซึ่งโดยทั่วไปจะไม่สามารถแก้ปัญหาที่ซับซ้อนได้เช่นลายมือเขียนของบุคคล หรือภาพใบหน้าภายใต้เงื่อนไขมุมมองที่แตกต่างกัน วิธีนี้มีประสิทธิภาพในการคำนวณ สามารถแก้ปัญหาได้ทั้งหมด และใช้ได้กับกลุ่มข้อมูลสำคัญได้หลากหลาย งานวิจัยต่อมาในปี ค.ศ. 2003 Brand [38] เสนอแนวทางการลดมิติแบบไม่เป็นเชิงเส้น (Nonlinear dimensionality

reduction: NLDR) เพื่อสร้างกราฟด้วยข้อมูลข้างเคียงกับจุดสูงสุด และแสดงวิธีแก้ปัญหาทั้งในรูปแบบผสมหลักการเคอร์เนลของผังที่เกี่ยวข้องจากปริภูมิทั้งกลุ่ม ไปยังปริภูมิเป้าหมาย

2.2.2 โครงข่ายประสาทเทียม (Neural network)

การแก้ปัญหาการรู้จำใบหน้าที่ไม่เป็นเชิงเส้นอีกวิธีหนึ่งคือ โครงข่ายประสาทเทียม ซึ่งถูกนำมาใช้อย่างแพร่หลายกับปัญหาการรู้จำรูปแบบต่างๆ มากมาย และปรับใช้กับการพิสูจน์ตัวตนบุคคล ตัวคิดแยกประสาทเทียม (Neural classifiers) ดีกว่าวิธีที่เป็นเชิงเส้นคือ สามารถลดการคิดแยกที่ผิดพลาด (Misclassification) ระหว่างกลุ่มข้อมูลที่อยู่ชิดกัน หลักการโครงข่ายประสาทเทียมจะพิจารณาจุดภาพทุกจุดในภาพด้วยโครงข่ายประสาท อย่างไรก็ตาม ภาพอินพุตส่วนใหญ่จะมีปริมาณข้อมูลมาก จึงไม่ได้ป้อนภาพอินพุตเข้าสู่โครงข่ายประสาทโดยตรง แต่จะต้องผ่านการลดขนาดข้อมูลด้วยเทคนิคการลดมิติก่อน



รูปที่ 2.7 โครงสร้างตามแนวทางโครงข่ายประสาทเทียม

งานวิจัยแรกที่ใช้วิธีนี้แก้ปัญหา เสนอโดย Cottrell และ Fleming [39] ซึ่งแนะนำโครงข่ายประสาทสองชั้น (Second neural net) ทำงานในโหมดกึ่งอัตโนมัติ ดังแสดงในรูปที่ 2.8 ชั้นแรกภาพใบหน้าซึ่งแทนด้วยเวกเตอร์ x จะถูกลดมิติให้เป็นเวกเตอร์ตัวใหม่ h ด้วยโครงข่ายแรก จากนั้นเวกเตอร์ h จะเป็นอินพุตป้อนเข้าสู่โครงข่ายที่สอง ซึ่งเป็นโครงข่ายที่ทำหน้าที่คิดแยก (Classification net) Cottrell และ Fleming แสดงให้เห็นว่า ประสิทธิภาพโครงข่ายประสาทแบบนี้ไม่ได้ดีกว่าวิธีใบหน้าไอเกน แม้ว่าอยู่ในสถานะที่เหมาะสม ต่อมามีการทดสอบโครงข่ายประสาทเทียมแบบอื่นกับการรู้จำใบหน้า เพื่อหาคุณสมบัติเฉพาะที่ดีขึ้น ได้แก่ Self organizing map (SOM) ซึ่งไม่มีผลมากนักเมื่อเทียบกับการเปลี่ยนแปลงเล็กน้อยในภาพ ขณะที่โครงข่ายที่ซับซ้อนจะมีการเปลี่ยนแปลงบางส่วน

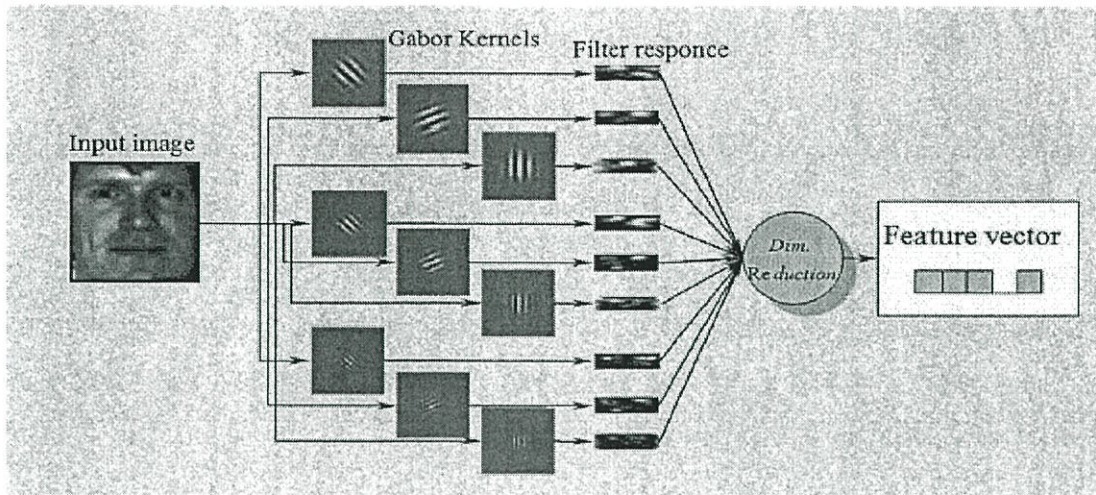
เมื่อเทียบกับการหมุน การแปลง และขนาดของภาพใบหน้า โดยทั่วไปโครงสร้างของโครงข่ายจะขึ้นอยู่กับประยุกต์ใช้งานในด้านต่างๆ อย่างมาก

ในปี ค.ศ. 1997 Lin และคณะ [40] ได้เสนอ โครงข่ายประสาทโดยใช้หลักความน่าจะเป็นในการตัดสินใจ (Probabilistic Decision Based Neural Network) ซึ่งนำไปประยุกต์ใช้งานต่างๆ ใน 3 รูปแบบได้แก่ ตัวตรวจจับใบหน้า (Face detection) ดวงตา (Eye detection) และการรู้จำใบหน้า (Face recognition) โครงข่ายเหล่านี้มีความยืดหยุ่นในการจัดลำดับชั้นกับฟังก์ชันที่ไม่เป็นเชิงเส้น ทำให้สามารถรู้จำได้มากถึง 200 คน และในปี ค.ศ. 2002 Meng และคณะ [41] นำเสนอแนวทางไฮบริด โดยขั้นแรกใช้วิธี PCA ในการคัดแยกคุณลักษณะเด่น จากนั้นใช้คุณลักษณะเด่นที่ได้เป็นอินพุตเข้าสู่โครงข่ายประสาทแบบ RBF (Radial Basis Function) อีกขั้นหนึ่ง ซึ่ง RBF สามารถจัดการกับปัญหาการรู้จำใบหน้าได้ดี เพราะมีอัลกอริทึมที่กระชับและเรียนรู้ระบบได้เร็ว

2.2.3 กาบอร์ฟิลเตอร์และเวฟเล็ต (Gabor filters and wavelets)

กาบอร์ฟิลเตอร์นับว่าเป็นเครื่องมือที่มีประสิทธิภาพเมื่อใช้ในการประมวลผลภาพ และการเข้ารหัสภาพ กรณีที่สำคัญที่สุด กาบอร์ฟิลเตอร์จะถูกใช้คัดแยกคุณลักษณะหลัก (Main features) ของภาพใบหน้า งานวิจัยของ Lades และคณะ [42] ที่เสนอในปี ค.ศ. 1993 ประยุกต์ใช้เฉพาะพื้นที่กรอบใบหน้า ซึ่งแทนด้วยโหนดของกริดตายตัว แต่ละโหนดของกริดนั้น ค่าสัมประสิทธิ์กาบอร์ (Gabor coefficients) จะถูกคัดแยกและจัดรวมเข้าเป็นกลุ่ม จากนั้นแต่ละโหนดจะถูกลิงค์ให้อยู่ในรูปไดนามิกลิงค์ (Dynamic link architecture: DLA) เพื่อทำการเปรียบเทียบค่าแตกต่างโดยใช้หลักการเทียบกราฟ (Graph matching strategy) ต่อมาปี ค.ศ. 1997 Wiskott และคณะ [43] ศึกษาวิธี DLA เพิ่มเติม และพัฒนา กาบอร์เวฟเล็ตด้วยการเทียบกราฟรวมแบบยืดหยุ่น (Elastic bunch graph matching method: EBGM) เพื่อการจำแนกและรู้จำใบหน้าบุคคล

แนวทางเวฟเล็ตที่เร็วกว่าถูกนำเสนอโดย Garcia และคณะ [44] ในปี ค.ศ. 2000 ซึ่งเสนอวิธีใหม่ในการรู้จำวิถีด้านหน้าของใบหน้าภายใต้เงื่อนไขความสว่างคงที่ วิธีนี้ใช้หลักการวิเคราะห์การแยกแยะเวฟเล็ตของภาพใบหน้าออกเป็นส่วนๆ ซึ่งกระบวนการนี้ทำได้รวดเร็วด้วยฮาร์ดแวร์ ขั้นแรกภาพอินพุตแต่ละภาพจะถูกแบ่งย่อยเป็นภาพที่ถูกฟิลเตอร์ที่มีค่าสัมประสิทธิ์เวฟเล็ต จากค่าสัมประสิทธิ์ที่มีคุณสมบัติโครงสร้างหรือความรู้สึกละเอียดของใบหน้าเหล่านี้ ทีมงานวิจัยได้สร้างเวกเตอร์คุณลักษณะ (Feature vector) โดยใช้การวัดทางสถิติอย่างง่าย จากนั้นแสดงให้เห็นว่า เมตริกซ์จากระยะ Bhattacharyya ความน่าเชื่อถือที่เป็นไปได้และประสิทธิภาพที่ได้สามารถใช้อัดแยกเวกเตอร์คุณลักษณะใบหน้าเข้ากลุ่มได้อย่างไร เพื่อแสดงให้เห็นว่า แม้แต่คุณลักษณะทางสถิติอย่างง่ายก็สามารถให้หลักการที่ดีสำหรับการคัดแยกใบหน้า



รูปที่ 2.8 หลักการวิธีกาบอร์และการลดมิติของผลตอบสนองฟิลเตอร์

ในปี ค.ศ. 2004 Liu [45] นำเสนอแนวทางที่ต่างออกไป (รูปที่ 2.8) โดยกำหนดเวฟเล็ตตัวหลัก (Mother wavelet) เพื่อให้ได้กาบอร์ฟิลเตอร์ 40 ตัว โดยเริ่มจากนำเวฟเล็ตตัวหลักมาปรับขนาดให้เป็น 5 สเกล และทำการหมุนแต่ละสเกลไปใน 8 ทิศทาง เพื่อให้ได้ฟิลเตอร์รวมทั้งหมด 40 ตัว เมื่อนำภาพอินพุตมาผ่านการกรองด้วยฟิลเตอร์เหล่านี้ จะได้ผลลัพธ์เป็นสำเนาของภาพ 40 ชุด เป็นการรวบรวมคุณลักษณะทุกตัวที่ได้จาก Gabor kernels ต่างๆ กัน คุณลักษณะของกาบอร์เวฟเล็ตที่ได้จะถูกนำมารวมกันเพื่อหาเวกเตอร์คุณลักษณะกาบอร์ที่ดีกว่า สำหรับการลดมิติของเวกเตอร์คุณลักษณะ จะใช้เทคนิค 2 ตัว ได้แก่ PCA และ EFM (Enhanced Fisher linear discriminant model) การใช้กาบอร์ฟิลเตอร์สามารถช่วยได้เป็นอย่างมากกับการทดสอบผลกระทบเกี่ยวกับลักษณะสีหน้าท่าทาง และความสว่างใบหน้า อย่างไรก็ตาม วิธีนี้ต้องมีการคำนวณคอร์เนลแต่ละภาพอินพุตเพิ่มมากขึ้น

2.2.4 ระบบฟังก์ชันทำซ้ำและแฟรคทอล (Fractals and iterated function systems: IFSs)

การเข้ารหัสข้อมูลภาพแบบแฟรคทอลเป็นการนำทฤษฎี Fractal มาใช้ในการบีบอัดข้อมูลภาพ ซึ่งผลที่ได้เป็นการบีบอัดแบบสูญเสียข้อมูลบางส่วน วิธีการเข้ารหัสข้อมูลภาพแบบแฟรคทอลนี้มีข้อดีคือ มีอัตราการบีบอัดข้อมูลสูง คุณภาพของภาพใกล้เคียงกับภาพต้นฉบับ อีกทั้งมีความเร็วในการคืนสภาพ (Reconstructed image) ที่เร็ว และมีความสามารถในการคืนสภาพผลลัพธ์ที่ไม่ขึ้นอยู่กับความละเอียด (Resolution independence) ของภาพ เพราะวิธีการเข้ารหัสข้อมูลภาพแบบแฟรคทอล นั้น ไม่ได้เก็บข้อมูลบางส่วนของภาพแบบวิธีการอื่นๆ แต่จะเป็นการเก็บข้อมูลในรูปแบบของรหัสแฟรคทอล (Fractal Code) ประกอบไปด้วยค่าคงที่ที่ใช้ในการปรับค่าความสว่าง ค่าความคมชัด และรูปแบบที่ใช้ในการเข้ารหัสข้อมูลภาพ จึงทำให้ข้อมูลที่ถูกเก็บมีค่าน้อยเมื่อเปรียบเทียบกับวิธีการอื่นๆ และจะใช้ข้อมูลจากรหัสแฟรคทอลนำมาสร้างภาพขึ้นใหม่ด้วยกระบวนการทำซ้ำ แต่วิธีการเข้ารหัสข้อมูลภาพแบบแฟรคทอลนี้ จะมีข้อเสีย คือใช้เวลาในการเข้ารหัสข้อมูลภาพหรือการบีบอัดข้อมูลภาพที่นานเมื่อเปรียบเทียบ

กับวิธีการอื่น เพราะต้องใช้เวลาในการหาค่าคงที่ ที่ใช้ในการปรับค่าความสว่าง ความคมชัด และรูปแบบที่ใช้ในการเข้ารหัสภาพของแต่ละส่วนในภาพต้นฉบับ

ข้อดีของระบบการรู้จำใบหน้าที่ใช้หลักการแฟรคทอลคือ ใช้การคำนวณน้อยกว่าเทคนิคแบบเชิงเส้น ไม่เป็นเชิงเส้น และโครงข่ายประสาทเทียม อีกทั้งเมื่อเพิ่มภาพใหม่เข้าไปในฐานข้อมูลก็ไม่จำเป็นต้องมีการสอนระบบใหม่ (Retraining) ในปี ค.ศ. 1999 Tan และ Yan [48] นำเสนอหลักการกลยุทธ์ IFS ที่แตกต่างออกไป โดยการนำ IFS ของภาพใบหน้า ให้เป็นเวกเตอร์คุณลักษณะและเก็บไว้ในฐานข้อมูล เมื่อมีภาพใหม่ที่ต้องระบุ การทำซ้ำเพียงครั้งเดียวเท่านั้นของรหัส IFS ทุกตัวในฐานข้อมูล จะใช้กับภาพที่ถูกทดสอบ และเลือกรหัส IFS ที่ให้ค่า PSNR (Peak Signal to Noise Ratio) สูงสุดซึ่งชี้ว่าเป็นการระบุที่ถูกต้อง มีผลให้เป็นการจำกัดในการใช้งานจริงของเทคนิคนี้ จริงๆ แล้วเมื่อพิจารณาฐานข้อมูลภาพใบหน้า ต้องพิจารณาเวลาที่ใช้เปรียบเทียบใบหน้าทั้งหมดด้วย วิธีแก้ไขข้อบกพร่องนี้คือ ต้องเก็บข้อมูลเวกเตอร์คุณลักษณะที่ถูกแยกเรียบร้อยแล้วขณะกำลังเข้ารหัส ข้อจำกัดที่สำคัญกว่าอีกข้อหนึ่งคือ การเข้ารหัสแฟรคทอลเป็นแบบไม่สมมาตร (Asymmetrical) ทำให้ขั้นตอนการเข้ารหัสยากกว่าการถอดรหัสมาก ดังนั้น บ่อยครั้งที่อัลกอริทึมการรู้จำใบหน้าแบบ IFS จึงประมวลผลช้า

2.2.5 ภาพอินฟราเรดและไฮเปอร์สเปกตรัม (Thermal and hyperspectral images)

ประสิทธิภาพที่ยอมรับได้ของเทคนิคที่ใช้หลักการภาพที่มีอยู่แล้วแบบเดิม ต้องมีการควบคุมเงื่อนไขต่างๆ ของภาพใบหน้าที่ได้มา แต่ส่วนมากวิธีเหล่านี้ไม่สามารถใช้ได้กับความผิดพลาดต่างๆ เช่น ไม่ได้ควบคุมเงื่อนไขความสว่างของภาพที่ได้มาเป็นสำคัญ จึงมีความพยายามแก้ปัญหาโดยใช้ภาพอินพุทที่มีความเข้มของภาพแตกต่างกัน ตัวอย่างเช่น ภาพอินฟราเรด ภาพชนิดนี้ทำงานด้วยคุณลักษณะพื้นผิวย่อย (Subsurface feature) ของใบหน้า ไม่นานมานี้งานวิจัยบางฉบับพิจารณาให้เป็นคุณลักษณะทางไบโอเมตริกได้จริงๆ หัวข้อที่น่าสนใจที่สุดคือ ความแตกต่างที่สำคัญระหว่างการรู้จำใบหน้าที่ใช้หลักการภาพที่มีความเข้มของภาพกับหลักการภาพอินฟราเรด เพื่อจะเน้นให้เห็นข้อดีของไบโอเมตริกวิธีใหม่เหล่านี้

งานวิจัยตามแนวทางภาพอินฟราเรดที่น่าสนใจนำเสนอโดย Chen และคณะ [49] ในปี ค.ศ. 2003 พวกเขาแนะนำ การทดสอบเปรียบเทียบประสิทธิภาพของเทคนิคอินฟราเรดกับความเข้มของภาพ แสดงให้เห็นว่า ดีกว่าวิธีก่อนหน้านี้เมื่อทำการทดสอบกับภาพภายใต้เงื่อนไขที่ไม่ได้ควบคุมเมื่อเปรียบเทียบกับ การเปลี่ยนแปลงความสว่าง อีกแนวทางหนึ่งถูกนำเสนอในปี ค.ศ. 2004 Socolinsky และ Selinger [50] - [51] เสนอวิธีการหาอัตราความถูกต้องในการรู้จำใบหน้าภายใต้เงื่อนไขเวลาที่ต่างกันของภาพฝึกฝน (Training images) กับภาพทดสอบ (Testing images) ข้อมูลที่ใช้ทดสอบได้จากเงื่อนไขต่างๆ และในสภาพแวดล้อมภายนอกอาคาร โดยที่ผลการทดสอบสนับสนุนวิธีที่ใช้ข้อมูลภาพอินฟราเรดและความเข้มของภาพร่วมกันว่า มีสมรรถนะการรู้จำที่ดีกว่า

ปี ค.ศ. 2004 Buddharaju และคณะ [52] ประยุกต์กระบวนการแบ่งด้วยฟัซซี่ (Fuzzy based segmentation process) เพื่อใช้แยกขอบเขตที่สนใจ ภาพออร์ฟิลเตอร์ถูกใช้แยกคุณลักษณะหลักแล้วใช้เป็นอินพุทป้อนเข้าสู่รูปแบบ Bessel สุดท้ายจะใช้ตัวคัดแยก Bayesian ในกระบวนการรู้จำ Buddharaju

สรุปว่า เทคนิคของเขาเป็นที่ยอมรับได้เมื่อเทียบกับวิธีใบหน้าไอเกน แต่ก็ยังมีข้อบกพร่องที่สำคัญคือ การทำงานกับภาพอินฟราเรดขึ้นอยู่กับอุณหภูมิผิวขณะกระบวนการได้มาของภาพ แต่ยังไม่เหมือนภาพ ความร้อนที่คาดไว้ ลักษณะการแยกสีและเงาของความร้อนจากใบหน้าขึ้นกับอุณหภูมิน้อยกว่าการแผ่รังสีความร้อน

นอกจากนี้ ในทางชีวการแพทย์ (Biomedicine) และการใช้งานตรวจจับระยะไกล (Remote sensing) มีการศึกษาวิจัยวิธีการตรวจจับสเปกตรัมของสีด้วยสเปกโตรสโคป (Spectroscope) กันอย่างกว้างขวาง เพราะบุคคลแต่ละคนจะแสดงคุณสมบัติสเปกตรัมการแยกสีเงาของส่วนใบหน้าได้กว้างมาก แต่บุคคลคนเดียวกันคุณลักษณะเหล่านี้จะคงที่ภายใต้เงื่อนไขเวลาและความสว่างแตกต่างกัน สมมุติฐานนี้ ทำให้มีการศึกษาความเป็นไปได้ในทางไบโอเมตริกจริงๆ

ปี ค.ศ. 2003 โดย Pan และคณะ [53] ได้เสนอเทคนิคที่ประยุกต์ใช้หลักการภาพไฮเปอร์สเปกตรัม (Hyperspectral images) บนฐานข้อมูลจำนวน 200 ออบเจ็กต์ ซึ่งได้มาจากใต้เงื่อนไขต่าง ๆ เช่น เงื่อนไขมุมของใบหน้า การแสดงอารมณ์ทางใบหน้า และความสว่างของใบหน้า แต่ละภาพใบหน้าถูกนำเสนอโดยใช้เวกเตอร์การสะท้อนสเปกตรัม ซึ่งถูกคัดแยกออกมาจากขอบเขตพื้นที่เล็กๆ บนใบหน้า สุดท้ายจะใช้ระยะมาฮาလာโนบิสวัดความเหมือนกันของภาพใบหน้า

2.3 ฐานข้อมูลภาพใบหน้า (Face databases)

ถ้าจะมีการประเมินว่า วิธีการที่นำเสนอมีข้อดีอย่างไรเมื่อมีตัวแปรที่เป็นเงื่อนไขหนึ่งตัว หรือหลายๆ เงื่อนไขเข้ามาเกี่ยวข้อง จะต้องสร้างฐานข้อมูลภาพใบหน้าขึ้นหลายฐานข้อมูลตามเงื่อนไขที่ต้องการแก้ปัญหา ขึ้นอยู่กับจำนวนและรูปแบบของปัญหาที่กำหนดไว้ รวมถึงพารามิเตอร์อื่นๆ เช่น จำนวนฐานข้อมูลที่ถูกต้องทดสอบ ขนาดของภาพใบหน้าทั้งหมดและชุดทดสอบ สิ่งเหล่านี้เป็นเหตุจูงใจให้นักวิจัยจำนวนมากสร้างฐานข้อมูลขึ้นหลายชุด ซึ่งให้ตัวแปรมากมายที่อาจเป็นไปได้บนภาพใบหน้าจำนวนมาก ฐานข้อมูลบางส่วน เช่น FERET [55], CMU-PIE [58], AR Faces [60] เป็นฐานข้อมูลภาพใบหน้า 2 มิติ ที่นิยมใช้กันมาก

ฐานข้อมูลแต่ละชุดถูกออกแบบให้ครอบคลุมเงื่อนไขการเปลี่ยนแปลงต่างๆ ที่หลากหลาย ตัวอย่างเช่น FERET เป็นฐานข้อมูลที่ดีสำหรับผู้ที่ต้องการจำนวนภาพฝึกฝนและภาพทดสอบระบบเป็นจำนวนมาก ขณะที่ CMU เน้นไปที่ปัญหาการทดสอบการเปลี่ยนแปลงของท่าทางและความสว่างของใบหน้าเป็นหลักมากกว่า ส่วน AR เป็นฐานข้อมูลที่เหมาะสมสำหรับการทดสอบเกี่ยวกับธรรมชาติการบังใบหน้าเท่านั้น ฐานข้อมูลภาพใบหน้าแบบ 2 มิติ ที่นิยมใช้กันมากที่สุดแสดงไว้ในตารางที่ 2.1 แต่ละแถวจัดแบ่งตามคุณสมบัติต่างๆ เริ่มจาก ชื่อฐานข้อมูล จำนวนบุคคลและจำนวนภาพต่อคน ชนิดภาพสีหรือสีเทา ความละเอียดของภาพ เงื่อนไขตัวแปรที่สามารถทำได้ ได้แก่ มุมของภาพ การแสดงออกทางอารมณ์ ความสว่างของภาพ เป็นต้น

ตารางที่ 2.1 การเปรียบเทียบฐานข้อมูลภาพใบหน้า

Database	LFW	FERET	FRGC	CAS-PEAL	PIE
Number of people	5749	1199	222 (large still training set) 466 (validation set)	1040 (595 males and 445 females) of Asians	68
Number of images	13233	14051	12,776 (large still training set) 943*8 (3D training set) 4007*8 (validation set)	30900	41368
Statics\Videos	Static	Static	Static	Static	Static
Gray\Color	Color	8-bit gray	Color	8-bit gray	Color
Resolution	250*250	256*384	1704*2272 or 1200*1600	360*480	640*486
Face pose	Various	7 categories: Frontal, quarter-left, quarter-right, half-left, half-right, full-left, full-right	Frontal view	21 pose angles vertical: up, middle, and down horizontal: left to right (67°, 45°, 22°, 0°, -22°, -45°, -67°)	13 pose angles in vertical
Facial expression	Various	Slight facial expression changes	Neutral and smiling	6 facial expressions: neutral, eye closing, frown, smile, surprise, and mouth open	4 facial expressions: neutral, eye closing, smiling, and talking
Illumination	Various	Controlled illumination	Controlled and uncontrolled illumination	15 lighting conditions	N / A
Accessories				3 kinds of glasses and 3 kinds of caps	Glasses
Ground truth	Identifications of subjects	Positions of eyes, nose, and mouth Identifications of subjects	Positions of eyes, nose, and mouth Identifications of subjects	Positions of eyes Identifications of subjects Face pose angles Facial expression labels Illumination positions	Some feature point data Identifications of subjects Measured locations of camera Head pose Facial expression labels Illumination positions
Number of people	15	10	126 (70 male and 56 female)	40	50
Number of images	165	5760	4000	400	750
Statics\Videos	Static	Static	Static images and image sequences	Static	Static
Gray\Color	8-bit gray	Color	Color	8-bit gray	Color
Resolution	320*243	640*480 (eye distance~90pixels)	768*576	92*112	640*480
Face pose	Frontal-view	9 poses	Nearly frontal-view or quarter-profile images	Moderate pose variation (up and down, quarter-profile to frontal-view)	Nearly frontal-view or quarter-profile
Facial expression	6 facial expressions: neutral, happiness, sadness, sleepiness, surprise, and wink	Neutral	4 facial expressions: neutral, smile, anger, and scream	3 facial expressions: neutral, smiling, closed eye	Various

ตารางที่ 2.1 (ต่อ)

Database	LFW	FERET	FRGC	CAS-PEAL	PIE
Illumination	3 lighting conditions: center-light, left-light, and right-light	64 lighting conditions and 1 ambient illumination	3 illumination conditions: left, right, and all side lights on	N / A	Various
Accessories	Glasses	N / A	Sun glasses, scarf	Glasses	Glasses
Ground truth	Identifications of subjects Facial expression labels Illumination positions	Identifications of subjects Illumination positions Coordinates of eyes and mouth (frontal view) Coordinates of face center (other views)	Identifications of subjects Facial expression labels	Cropped face region Identifications of subjects	Identifications of subjects Coordinates of left-upper corner and right corner of face rectangle

ต่อไปจะแสดงถึงรายละเอียดของฐานข้อมูลต่างๆ ที่สำคัญ และเป็นที่ยอมรับใช้งานกันอย่างแพร่หลายในงานวิจัยด้านการรู้จำใบหน้า รวมทั้งข้อสรุปและข้อกำหนดของแต่ละฐานข้อมูล ทั้งนี้เพื่อเป็นข้อมูลในการเปรียบเทียบ และเป็นประโยชน์ในการอ้างอิงต่อการวิจัยด้านการรู้จำใบหน้าต่อไป

2.3.1 ฐานข้อมูล LFW (Labeled Faces in the Wild) [54]

ฐานข้อมูล LFW เป็นฐานข้อมูลภาพใบหน้าที่ถูกสร้างขึ้นโดย มหาวิทยาลัยแมสซาชูเซตส์ (University of Massachusetts) ออกแบบสำหรับการศึกษาปัญหาการรู้จำใบหน้าที่ไม่มีการกำหนดขอบเขตแปรเปลี่ยนตามเงื่อนไขการแต่งกาย ท่าทาง ฉากหลัง และอื่นๆ อีกมาก ฐานข้อมูลนี้รวบรวมภาพใบหน้าที่เป็นภาพนิ่ง (Static) จากเว็บไซต์ต่างๆ มีจำนวนภาพมากถึง 13,233 ภาพ แต่ละภาพกำกับด้วยชื่อของบุคคล รวมทั้งสิ้น 5,749 คน บุคคลที่มีภาพตั้งแต่ 2 ภาพ หรือมากกว่านั้นมีจำนวน 1,680 คน ชนิดภาพสี ความละเอียด 250×250 พิกเซล มีการแสดงออกทางอารมณ์ และความสว่างของภาพใบหน้าแตกต่างกันหลากหลาย ดังตัวอย่างในรูปที่ 2.9



รูปที่ 2.9 ตัวอย่างภาพใบหน้าจากฐานข้อมูล LFW

2.3.2 ฐานข้อมูล FERET (Facial Recognition Technology) [55]

โปรแกรมการพัฒนาฐานข้อมูล FERET เริ่มจากปี ค.ศ. 1993 ถึงปี 1997 โดยได้รับการสนับสนุนจาก Department of Defense's Counterdrug Technology Development Program ผ่านทางหน่วยงาน Defense Advanced Research Products Agency (DARPA) ภารกิจแรกเริ่มเป็นการพัฒนา

ความสามารถในการรู้จำใบหน้าอัตโนมัติ ซึ่งใช้เป็นระบบเสริมความปลอดภัย ระบบอัจฉริยะ และการบังคับควบคุมทางกฎหมายในการทำหน้าที่ของบุคคล

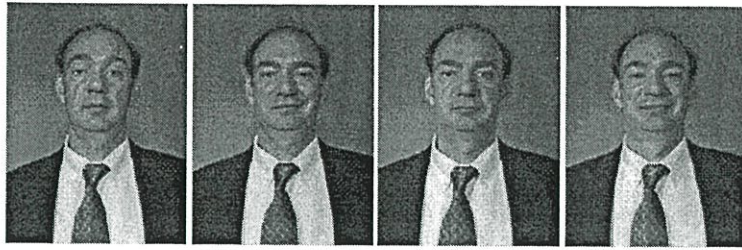


รูปที่ 2.10 ตัวอย่างภาพใบหน้าจากฐานข้อมูล FERET

ภาพใบหน้าของฐานข้อมูล FERET ถูกรวบรวมเพื่อสนับสนุนการทดสอบการเฝ้าระวังและประเมินกระบวนการรู้จำใบหน้า โดยใช้ขั้นตอนและการทดสอบที่เป็นมาตรฐาน ฐานข้อมูลนี้รวบรวมภาพใบหน้าจากบุคคลต่างๆ จำนวน 1,199 คน ประกอบด้วยภาพทั้งหมด 14,051 ภาพ ลักษณะภาพเป็นแบบสีเทา 8 บิต ด้วยมุมมองภาพจากด้านหน้า ด้านซีกซ้ายและซีกขวา เป็นภาพเดี่ยว แบบสแตติก ความละเอียด 256×384 พิกเซล มีการเปลี่ยนแปลงการแสดงอารมณ์ทางใบหน้าที่ละน้อย ควบคุมความสว่างของใบหน้า แต่ละใบหน้าที่มี 7 มุมมอง ได้แก่ มุมหน้าตรง เอียงซ้าย เอียงขวา หันครึ่งซ้าย หันครึ่งขวา เต็มด้านซ้าย และมุมเต็มด้านขวา ดังตัวอย่างในรูปที่ 2.10

2.3.3 ฐานข้อมูล FRGC (Face Recognition Grand Challenge) [56]

ฐานข้อมูล FRGC (Face Recognition Grand Challenge Database) ได้รับความร่วมมือจากหลายรัฐบาลที่สนใจในความสามารถของเทคโนโลยีการรู้จำใบหน้า เป้าหมายหลักเพื่อประเมินการรู้จำใบหน้าโดยเฉพาะ แต่อาจจะใช้กับการตรวจจับดวงตา (Eye detection) ด้วย และส่งเสริมความก้าวหน้าเพื่อสนับสนุนการใช้งานการรู้จำใบหน้าที่มีอยู่แล้วของรัฐบาลสหรัฐอเมริกา FRGC จะพัฒนาเทคนิคการรู้จำใบหน้าใหม่ๆ และพัฒนาระบบต้นแบบ อีกทั้งเพิ่มสมรรถนะพร้อมกันไปด้วย และจากการแข่งขันอัลกอริทึมในการเพิ่มประสิทธิภาพการรู้จำใบหน้าทั้ง 3 ด้าน คือ ภาพความละเอียดสูง (High resolution images) การรู้จำใบหน้า 3 มิติ (Three-dimensional face recognition) และเทคนิคใหม่ในขั้นตอนก่อนประมวลผล (New preprocessing techniques) FRGC กำลังติดตามเทคนิคทั้งสามและประเมินผลที่ได้รับไปพร้อมกัน ระบบการรู้จำใบหน้าปัจจุบันถูกออกแบบให้ทำงานกับภาพใบหน้าที่มีอยู่จำนวนน้อย วิธีดั้งเดิมสำหรับการวัดขนาดของใบหน้าที่จะบอกเป็นจำนวนพิกเซลระหว่างจุดกึ่งกลางของดวงตาทั้งสอง ปัจจุบันฐานข้อมูลภาพใบหน้าทั่วไปมีระยะห่างจุดกึ่งกลางดวงตาระหว่าง 40 ถึง 60 พิกเซล (ที่ความละเอียดภาพใบหน้า 10,000-20,000 พิกเซล) สำหรับฐานข้อมูล FRGC ประกอบด้วยภาพความละเอียดสูงที่มีระยะห่างจุดกึ่งกลางดวงตาเฉลี่ย 250 พิกเซล



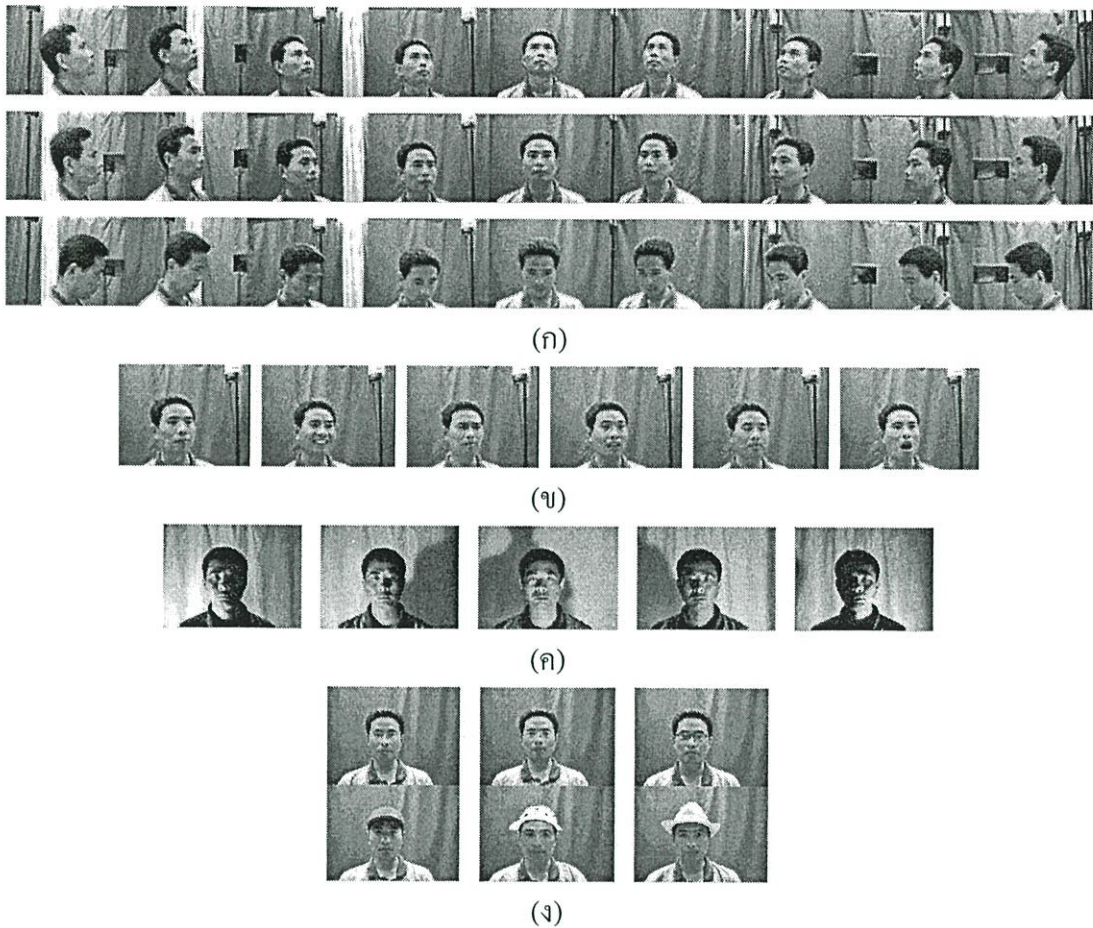
รูปที่ 2.11 ตัวอย่างภาพใบหน้าจากฐานข้อมูล FRGC

FRGC ถูกวางโครงสร้างให้ครอบคลุมปัญหาที่ทำหายสองปัญหา กำหนดเป็น Version 1 (ver1) และ Version 2 (ver2) สำหรับ Ver1 ถูกออกแบบเพื่อแนะนำผู้ใช้งานทั่วไปให้ร่วมแก้ปัญหการทำหายและสนับสนุนโครงสร้างพื้นฐานของ FRGC ส่วน Ver2 ถูกออกแบบเพื่อท้าทายให้นักวิจัยค้นหาเป้าหมายตามสมรรถนะของ FRGC

นอกจากนี้ FRGC มี 3 รูปแบบใหม่ในวงการการรู้จำใบหน้า รูปแบบแรกคือ ขนาดของ FRGC ในเทอมของข้อมูล ซึ่ง FRGC ver2 บรรจุ 50,000 เรคคอร์ด รูปแบบที่สองคือ ความซับซ้อนของ FRGC เนื่องจากชุดข้อมูลในการรู้จำใบหน้าก่อนหน้านี้อูกจำกัดด้วยภาพนิ่งในรูปแบบเดิมๆ แต่ FRGC ประกอบไปด้วย 3 โหมด ได้แก่ ภาพที่มีความละเอียดสูง ภาพ 3 มิติ และภาพหลายๆ ภาพต่อหนึ่งบุคคล รูปแบบที่สามคือ โครงสร้างพื้นฐาน (Infrastructure) สำหรับ FRGC จะจัดโครงสร้างพื้นฐานที่เรียกว่า สภาพแวดล้อมในการทดสอบทางไบโอเมตริก (Biometric Experimentation Environment: BEE) การอธิบายและเอกสารการทดสอบการคำนวณด้วยเฟรมเวิร์คพื้นฐานที่เป็น XML การทดสอบของ BEE จะยอมให้มีการอธิบายและกระจายการทดสอบ รวมทั้งการบันทึกผลข้อมูลดิบ เอกสารรูปแบบการทดสอบให้อยู่ในรูปแบบร่วมกัน รูปแบบนี้เป็นครั้งแรกที่มีสภาวะแวดล้อมการทดสอบการคำนวณที่สนับสนุนปัญหาท้าทายในการรู้จำใบหน้าและไบโอเมตริก

2.3.4 ฐานข้อมูล CAS-PEAL [57]

ฐานข้อมูล CAS-PEAL (Chinese Academy of Sciences-Pose, Expression, Accessories, and Lighting) ถูกสร้างขึ้นภายใต้การสนับสนุนของ National Hi-Tech Program และ ISVISION วัตถุประสงค์เพื่อสร้างฐานข้อมูลภาพใบหน้าของ PEAL รวมทั้งจัดเตรียมฐานข้อมูลภาพใบหน้าขนาดใหญ่ของชาวจีนสำหรับกลุ่มนักวิจัยการรู้จำใบหน้า ด้วยเงื่อนไขที่แตกต่างกัน เช่น การแสดงอารมณ์ทางใบหน้า ส่วนประกอบและความสว่างของใบหน้า ปัจจุบันฐานข้อมูล CAS-PEAL ประกอบด้วยภาพบุคคลต่างๆ จำนวน 1,040 คน รวมทั้งสิ้น 99,594 ภาพ เป็นชาย 595 คน หญิง 445 คน แต่ละเงื่อนไขใช้กล้อง 9 ตัว แนวนอนครั้งวงกลมติดตั้งจับภาพในท่าทางต่างๆ 1 ชุด และภาพมองขึ้นลงแต่ละชุด 18 ภาพ ดังตัวอย่างภาพในรูปที่ 2.12



รูปที่ 2.12 ตัวอย่างภาพใบหน้าจากฐานข้อมูล CAS-PEAL (ก) ภาพ 9 มุมมอง (ข) การแสดงอารมณ์ทางใบหน้า (ค) ความสว่าง (ง) การสวมอุปกรณ์

ภาพแต่ละภาพในฐานข้อมูลนี้เป็นภาพสี ใบหน้าเดี่ยว แบบสแตติก ความละเอียด 360×480 พิกเซล มุมท่าทาง 21 มุม การแสดงอารมณ์ทางใบหน้า 5 แบบ เงื่อนไขของแสง 15 เงื่อนไข ส่วนประกอบใบหน้าอีก 6 ชนิด (สวมแว่น 3 แบบ และสวมหมวก 3 แบบ)

2.3.5 ฐานข้อมูล PIE (CMU Pose, Illumination, and Expression) [58]

ฐานข้อมูลนี้เรียกย่อๆ ว่า PIE ซึ่งมาจากคำว่า Pose, Illumination and Expression เป็นฐานข้อมูลที่ได้รับจากสถาบัน Robotics Institute of Carnegie Mellon University เหมาะสำหรับงานวิจัยการรู้จำใบหน้าที่ต้องการทดสอบภายใต้เงื่อนไข ท่าทาง ความสว่าง และการแสดงอารมณ์ทางใบหน้า ฐานข้อมูลนี้ประกอบด้วยภาพใบหน้า 41,368 ภาพ จากบุคคลจำนวน 68 คน ภาพใบหน้าของแต่ละคนจะมีท่าทางแตกต่างกัน 13 แบบ เงื่อนไขความสว่างต่างกัน 43 แบบ และการแสดงอารมณ์ทางใบหน้าต่างกัน 4 แบบ ตัวอย่างภาพใบหน้าของฐานข้อมูล PIE แสดงในรูปที่ 2.13 เป็นภาพสี ใบหน้าเดี่ยว แบบสแตติก ความละเอียด 640×486 พิกเซล การแสดงอารมณ์ทางใบหน้าประกอบด้วยแบบปกติ ปิดตา ยิ้ม และพุดคุย



รูปที่ 2.13 ตัวอย่างภาพใบหน้าจากฐานข้อมูล PIE

2.3.6 ฐานข้อมูล YALE [59]

ฐานข้อมูล YALE เป็นฐานข้อมูลที่ถูกสร้างขึ้นโดย Yale University ประกอบด้วยภาพบุคคลจำนวน 15 คน แต่ละคนมี 11 ภาพ รวมภาพทั้งหมด 165 ภาพ ภาพเดี่ยว แบบสแตติก สีเทา 8 บิต ความละเอียด 320×243 พิกเซล มุมด้านหน้า การแสดงอารมณ์ทางใบหน้า 6 แบบ คือ ปกติ ยิ้ม เศร้า ง่วง ตกใจ และหลับตา เงื่อนไขความสว่าง 3 แบบ คือ สว่างตรงกลาง ด้านซ้าย และด้านขวา และอุปกรณ์ประกอบคือ สวมแว่นตา ตัวอย่างภาพใบหน้าแสดงในรูปที่ 2.14



รูปที่ 2.14 ตัวอย่างภาพใบหน้าจากฐานข้อมูล YALE

2.3.7 ฐานข้อมูล AR [60]

ฐานข้อมูล AR ถูกสร้างขึ้นโดย Aleix Martinez และ Robert Benavente ใน Computer Vision Center (CVC) ที่ UAB (Universitat Autònoma de Barcelona) ประกอบด้วยภาพบุคคลต่างๆ จำนวน 126 คน รวมทั้งสิ้น 4,000 ภาพ เป็นชาย 70 คน หญิง 56 คน ภาพสีเดี่ยว มีทั้งแบบสแตติกและภาพลำดับ ความละเอียด 768×576 พิกเซล มุมภาพเฉียงเกือบหน้าตรง เอียง 45 องศา การแสดงอารมณ์ทางใบหน้าที่มี 4 แบบ คือ หน้าปกติ ยิ้ม โกรธ และตกใจ เงื่อนไขความสว่าง 3 แบบ คือ สว่างตรงกลาง ด้านซ้าย และด้านขวา และอุปกรณ์ประกอบ 2 แบบ คือ สวมแว่นตาและผ้าพันคอ ดังตัวอย่างภาพแสดงในรูปที่ 2.15



รูปที่ 2.15 ตัวอย่างภาพใบหน้าจากฐานข้อมูล AR

2.3.8 ฐานข้อมูล ORL [61]

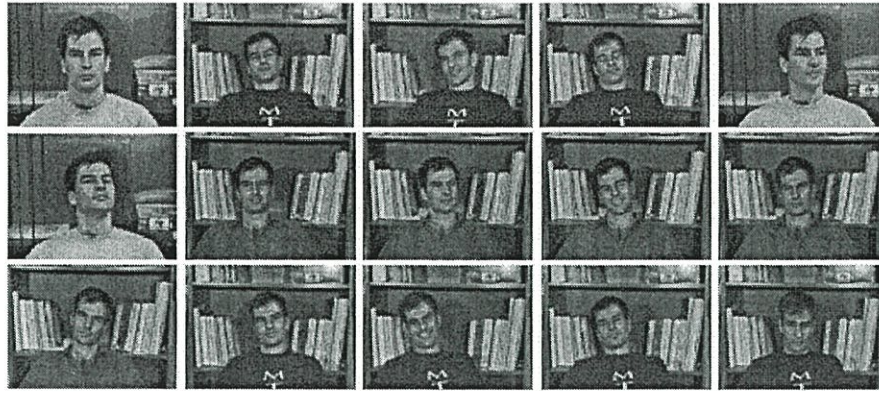
ฐานข้อมูล ORL ถูกสร้างขึ้นโดย AT&T Laboratories Cambridge ประกอบด้วยภาพบุคคลต่างๆ จำนวน 40 คน รวมทั้งสิ้น 400 ภาพ ภาพเดี่ยว แบบสแตติก สีเทา 8 บิต ความละเอียด 92×112 พิกเซล มุมภาพก้ม เงย จาก 45 องศา ถึงหน้าตรง การแสดงอารมณ์ทางใบหน้า 3 แบบ คือ หน้าปกติ ยิ้ม และปิดตา อุปกรณ์ประกอบ คือ สวมแว่นตา ตัวอย่างภาพแสดงในรูปที่ 2.16



รูปที่ 2.16 ตัวอย่างภาพใบหน้าจากฐานข้อมูล ORL

2.3.9 ฐานข้อมูล Georgia Tech [62]

ฐานข้อมูล Georgia Tech ถูกสร้างขึ้นโดย Georgia Institute of Technology ประกอบด้วยภาพจากบุคคล 50 คน รวมทั้งสิ้น 750 ภาพ เป็นแบบภาพสี ใบหน้าเดี่ยว แบบสแตติก ความละเอียด 640×480 พิกเซล มุมมองภาพเกือบจะเป็นด้านหน้ากับเอียง 45 องศา การเปลี่ยนการแสดงอารมณ์ทางใบหน้า และความสว่างหลากหลาย อุปกรณ์ประกอบ คือ สวมแว่นตา ดังตัวอย่างภาพในรูปที่ 2.17



รูปที่ 2.17 ตัวอย่างภาพใบหน้าจากฐานข้อมูล Georgia Tech

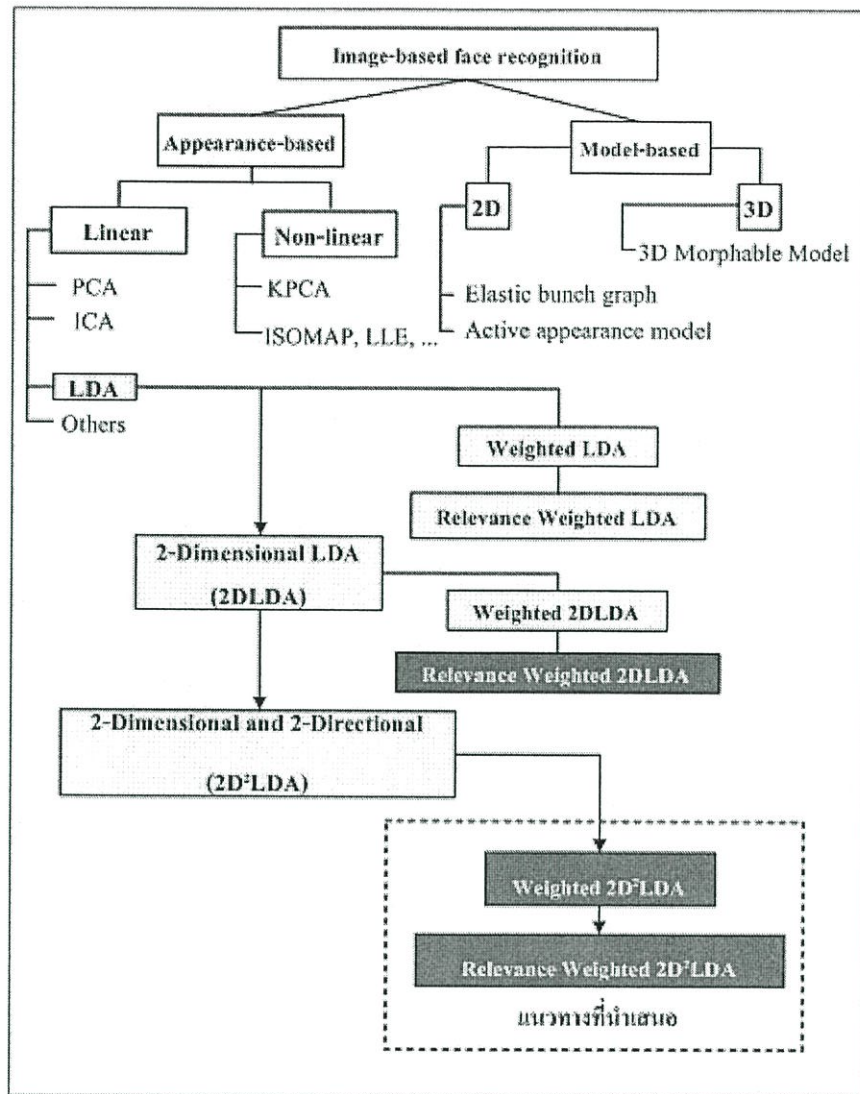
บทที่ 3

การวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติ

อัลกอริทึมการรู้จำใบหน้าที่นิยมใช้กันอย่างแพร่หลายคือ ทฤษฎีการวิเคราะห์องค์ประกอบหลัก (Principal component analysis: PCA) วิธีนี้ใช้หลักการของความสัมพันธ์เชิงเส้นระหว่างตัวแปรที่ใช้เป็นข้อมูลองค์ประกอบหลัก เป็นการวิเคราะห์ข้อมูลทางสถิติที่ใช้ลดขนาดข้อมูล โดยสร้างเมตริกซ์ความแปรปรวนร่วม (Covariance matrix) จากข้อมูลภาพ แล้วคำนวณไอเกนเวกเตอร์ (Eigenvectors) เพื่อหาแกนการฉาย (Projection axis) ที่เหมาะสม สำหรับฉายข้อมูลภาพให้มีขนาดเล็กลง อีกวิธีหนึ่งคือ ทฤษฎีการวิเคราะห์ลักษณะแตกต่างเชิงเส้น (Linear discriminant analysis: LDA) เป็นการหาปริภูมิเวกเตอร์ (Vector space) ใหม่ที่มีขนาดหรือมิติ (Dimension) ลดลงเช่นกัน โดยพิจารณาการกระจายข้อมูลภายในกลุ่มและข้อมูลระหว่างกลุ่มด้วย พบว่าเทคนิคทั้งสองต่างก็มีข้อเสีย คือ จะเกิดปัญหาที่เรียกว่า Small Sample Size problem (SSS) กรณีที่จำนวนตัวอย่างฝึกฝน (Training samples) น้อยกว่าขนาดหรือมิติของเวกเตอร์คุณลักษณะ (Feature dimension) ที่ใช้ทดสอบ เมตริกซ์การกระจายภายในกลุ่ม (Within-class scatter matrix) ของตัวอย่าง อาจจะเป็นเมตริกซ์เอกฐาน (Singular matrix) ทำให้กระบวนการประมวลผลต้องมีการคำนวณที่ยุ่ยากและซับซ้อนขึ้น

ต่อมาได้มีการนำเสนอเทคนิค การวิเคราะห์องค์ประกอบหลักสองมิติ (2-Dimensional PCA: 2DPCA) เพื่อแก้ปัญหา Small Sample Size ดังสรุปในรายงานการวิจัยที่ [10] และ [11] เทคนิคนี้ไม่มีการแปลงเมตริกซ์ภาพสองมิติให้เป็นเวกเตอร์หนึ่งมิติ แต่จะสร้างเมตริกซ์ความแปรปรวนร่วมที่มีมิติเท่ากับความกว้างของเมตริกซ์ของภาพใบหน้าเริ่มต้นโดยตรง ทำให้เมตริกซ์ความแปรปรวนร่วมของ 2DPCA มีขนาดเล็กกว่าวิธี PCA อีกทั้งการคำนวณไอเกนเวกเตอร์ที่สมนัย (Correspondence) กันใช้เวลาน้อยกว่า และแก้ปัญหา SSS ได้ดีกว่าวิธี PCA

วิธีการ PCA และ 2DPCA จัดว่าเป็นอัลกอริทึมการแทนภาพ (Face representation) มากกว่าการรู้จำภาพ (Face recognition) และการแปลงเชิงเส้นของวิธี 2DPCA จะลดจำนวนข้อมูลด้านคอสมินัสของเมตริกซ์ภาพเท่านั้น แต่จำนวนข้อมูลด้านแฉวยยังคงเท่ากับภาพเริ่มต้น ทำให้ไม่สามารถลดปัญหา SSS ได้ จึงมีการพัฒนาวิธี LDA ใหม่ให้เป็นแบบสองมิติเรียกว่า การวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติ (2-Dimensional LDA: 2DLDA) งานวิจัยที่มีการอ้างอิงวิธีนี้ได้แก่ [13], [14], [15] และ [17] ซึ่งนำเสนอว่า วิธี 2DLDA ไม่เพียงแก้ปัญหา SSS และลดมิติของเมตริกซ์ให้มีขนาดเล็กลงเท่านั้น แต่ 2DLDA ยังทำงานกับเมตริกซ์ภาพโดยตรง เพื่อหาเมตริกซ์การฉาย (Projection matrix) ที่ดีที่สุด ใช้หลักการสองพารามิเตอร์คือ เมตริกซ์การกระจายระหว่างกลุ่ม (Between-class scatter matrix) และเมตริกซ์การกระจายภายในกลุ่ม (Within-class scatter matrix) ที่มีขนาดเล็ก ทำให้สามารถแก้ปัญหากรณีเมตริกซ์การกระจายภายในกลุ่มอาจจะเป็นเมตริกซ์เอกฐานได้



รูปที่ 3.1 การพัฒนาอัลกอริทึมการรู้จำใบหน้าตามแนวทาง LDA

ขั้นตอนการพัฒนาตามแนวทางของวิธี LDA มีลำดับการพัฒนาดังแสดงในรูปที่ 3.1 ด้านบนแสดงวิธีการทั้งหมดของการรู้จำใบหน้าที่อาศัยหลักการภาพที่มีอยู่แล้วมาประมวลผล ซึ่งแบ่งเป็นแบบ Appearance-based และ Model-based

ด้านซ้ายแสดงหลักการของวิธี LDA ถูกพัฒนาเป็นวิธี 2DLDA และล่าสุดเป็น $(2D)^2LDA$ ซึ่งเป็นอัลกอริทึมที่มีการลดขนาดข้อมูลทั้งด้านแถวและคอลัมน์ของเมตริกซ์ภาพ อีกด้านหนึ่งมีความพยายามปรับปรุงประสิทธิภาพของ LDA โดยการเพิ่มฟังก์ชันการถ่วงน้ำหนักเข้าไปในเมตริกซ์การกระจายของข้อมูลในแต่ละกลุ่ม ซึ่งพบว่าการประยุกต์ฟังก์ชันการถ่วงน้ำหนักเป็น 2 แนวทางคือ ฟังก์ชันการถ่วงน้ำหนักและฟังก์ชันการถ่วงน้ำหนักสัมพัทธ์ (Weighting and Relevance Weighting function) การถ่วงน้ำหนักแบบ Weighting function ปกติใช้หลักการระยะยูคลิดีเนียน (Euclidean distance) หรือระยะมาฮาลานอบิส (Mahalanobis distance) ประยุกต์เป็นฟังก์ชันการถ่วงน้ำหนักในเมตริกซ์การกระจายระหว่างกลุ่ม ทำให้ข้อมูลระหว่างกลุ่มมีการแยกแยะกันได้ดีขึ้น ส่วนการเพิ่มฟังก์ชันการถ่วงน้ำหนักแบบ Relevance weighting function ใช้หลักการระยะยูคลิดีเนียนเป็นฟังก์ชันการถ่วงน้ำหนักในเมตริกซ์การ

กระจายภายในกลุ่ม เพื่อปรับปรุงข้อมูลภายในกลุ่มให้มีการกระจายของข้อมูลน้อยลง หรือจัดข้อมูลในกลุ่มเดียวกันให้ใกล้เคียงกันมากขึ้นซึ่งพบว่า มีการนำแนวคิดนี้มาประยุกต์ใช้กับงานรู้จำใบหน้าที่ใน 3 แนวทาง ได้แก่ Weighted LDA [21]-[26], Weighted 2DLDA [27]-[28] และ Relevance Weighted LDA [29]-[31] เป็นต้น

ไดอะแกรมด้านขวาในรูปที่ 3.1 แสดงการพัฒนาอัลกอริทึมของ LDA ที่มีการนำ Weighting function มาประยุกต์กับวิธี LDA ในงานการรู้จำใบหน้าที่เพื่อเพิ่มสมรรถนะการรู้จำให้ดีขึ้น งานวิจัยแรกที่พบคือ การประยุกต์ฟังก์ชันการถ่วงน้ำหนักเข้ากับวิธี LDA เรียกว่า Weighted LDA โดยใช้หลักการระยะยูคลิดีเนียน และระยะมาฮาลาโนบิสทั้งสองแบบประยุกต์เข้ากับเมตริกซ์การกระจายระหว่างกลุ่ม ต่อมา มีการนำ Relevance Weighting function ประยุกต์กับวิธี LDA ด้วยอีกแนวทางหนึ่ง ซึ่งเป็นการนำระยะยูคลิดีเนียนเป็นฟังก์ชันการถ่วงน้ำหนักในเมตริกซ์การกระจายภายในกลุ่ม เมื่ออัลกอริทึมของ LDA ถูกพัฒนาเป็น 2DLDA ก็พบงานวิจัยที่ประยุกต์ฟังก์ชันการถ่วงน้ำหนักที่เรียกว่า Weighted 2DLDA [27]-[28] ซึ่งใช้ระยะยูคลิดีเนียนเป็นฟังก์ชันการถ่วงน้ำหนักในเมตริกซ์การกระจายระหว่างกลุ่มกับวิธี 2DLDA เท่านั้น ยังไม่พบงานวิจัยที่เป็น Relevance Weighted 2DLDA เผยแพร่ออกมา ดังนั้น สามารถกำหนดเป็นแนวทางงานวิจัยได้ สำหรับอัลกอริทึมในกรอบเส้นประด้านล่างของรูปที่ 3.1 แสดงให้เห็นว่า ปัจจุบันยังไม่พบงานวิจัยที่ประยุกต์ฟังก์ชันการถ่วงน้ำหนักกับวิธี (2D)²LDA แต่อย่างใด จึงเป็นเหตุผลจูงใจในการนำเสนอเป็นงานวิจัยในวิทยานิพนธ์ฉบับนี้

เนื้อหาที่นำเสนอในบทที่ 3 นี้ กล่าวถึง ทฤษฎีที่นำมาประยุกต์ใช้ในงานรู้จำใบหน้าที่ โดยใช้แนวทางการวิเคราะห์เชิงเส้น (Linear analysis approach) ซึ่งประกอบไปด้วย เทคนิคการฉายภาพแบบต่างๆ ได้แก่ การวิเคราะห์องค์ประกอบหลัก การวิเคราะห์ลักษณะแตกต่างเชิงเส้น และการวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติและสองทิศทาง (2-Dimensional and 2-directional Linear Discriminant Analysis) หรือที่เรียกว่า (2D)²LDA นั่นเอง หลังจากนั้นจะอธิบาย ฟังก์ชันการถ่วงน้ำหนัก และการประยุกต์ฟังก์ชันนี้เข้ากับวิธีการวิเคราะห์ลักษณะแตกต่างเชิงเส้นแบบ (2D)²LDA เพื่อนำเสนอเป็นวิธีใหม่ในการรู้จำใบหน้าที่

3.1 แนวทางการวิเคราะห์เชิงเส้น (Linear analysis approach)

เทคนิคการรู้จำวัตถุตามแนวทาง Appearance-based ส่วนมากขึ้นอยู่กับ การแทนข้อมูลภาพ ด้วยวิธีการลดโครงสร้างปริภูมิเวกเตอร์ โดยพิจารณาภาพๆ หนึ่งให้เป็นจุดภาพที่อยู่ในปริภูมิเวกเตอร์ที่มีมิติขนาดใหญ่ (High-dimensional vector) และจะใช้เทคนิคทางสถิติเพื่อวิเคราะห์การกระจายของเวกเตอร์ภาพนั้นในปริภูมิเวกเตอร์ และหาประสิทธิภาพปริภูมิคุณลักษณะ (Feature space) ตามวิธีการประยุกต์แต่ละแบบ นั่นคือ การเปรียบเทียบหาความเหมือนกัน (Matching) ระหว่างภาพต้นแบบที่เก็บไว้กับภาพที่ทดสอบ ซึ่งจะต้องอยู่ในรูปปริภูมิคุณลักษณะเช่นกัน เทคนิคการฝึกสอนระบบใช้ข้อมูลภาพที่อยู่ในรูปเวกเตอร์เป็นตัววิเคราะห์และสังเคราะห์ภาพต่างๆ การรู้จำใบหน้าที่จัดเป็นปัญหาการค้นหาค้นหาปริภูมิ

(Space-searching problem) จึงขอกว่าถึง หลักการแทนภาพด้วยเวกเตอร์พอลังเจป เพื่อเป็นความรู้ความเข้าใจในการศึกษา เพื่อให้สามารถเข้าถึงการวิเคราะห์ที่ยุ่ยากซับซ้อนต่อไปได้

การเรียนรู้ปริภูมิย่อย (Subspace) เป็นแนวทางหลักที่ใช้ในการรู้จำใบหน้าและการวิเคราะห์โครงสร้างใบหน้า อัลกอริทึมปริภูมิย่อยในงานวิจัยทั้งหมดก่อนหน้านี้ เป็นการแปลงภาพใบหน้าใดๆ ที่อยู่ในรูปเมตริกซ์ให้เป็นรูปเวกเตอร์แถว (Row vector) ซึ่งปกติจะเกิดปัญหาในทางปฏิบัติยุ่งยากขึ้นคือ ข้อมูลโครงสร้างที่แท้จริงสูญหาย และมีมิติคุณลักษณะ (Feature dimension) มีขนาดใหญ่ ทำให้การประมวลผลช้า ที่เกิดขึ้นบ่อยคือ จำนวนตัวอย่างฝึกฝนของระบบมีจำนวนน้อยเมื่อเทียบกับมิติคุณลักษณะ ทำให้อัลกอริทึมได้รับผลจากปัญหา Small Sample Size เนื่องจากมิติของปริภูมิมีขนาดใหญ่ การเปรียบเทียบต้องใช้การคำนวณที่ซับซ้อน จึงมีการนำ เทคนิคการลดมิติ (Dimensionality reduction technique) มาใช้เพื่อแปลงปริภูมิที่มีขนาดใหญ่ไปสู่ปริภูมิที่เล็กกว่า

3.1.1 การวิเคราะห์องค์ประกอบหลัก (Principal component analysis: PCA)

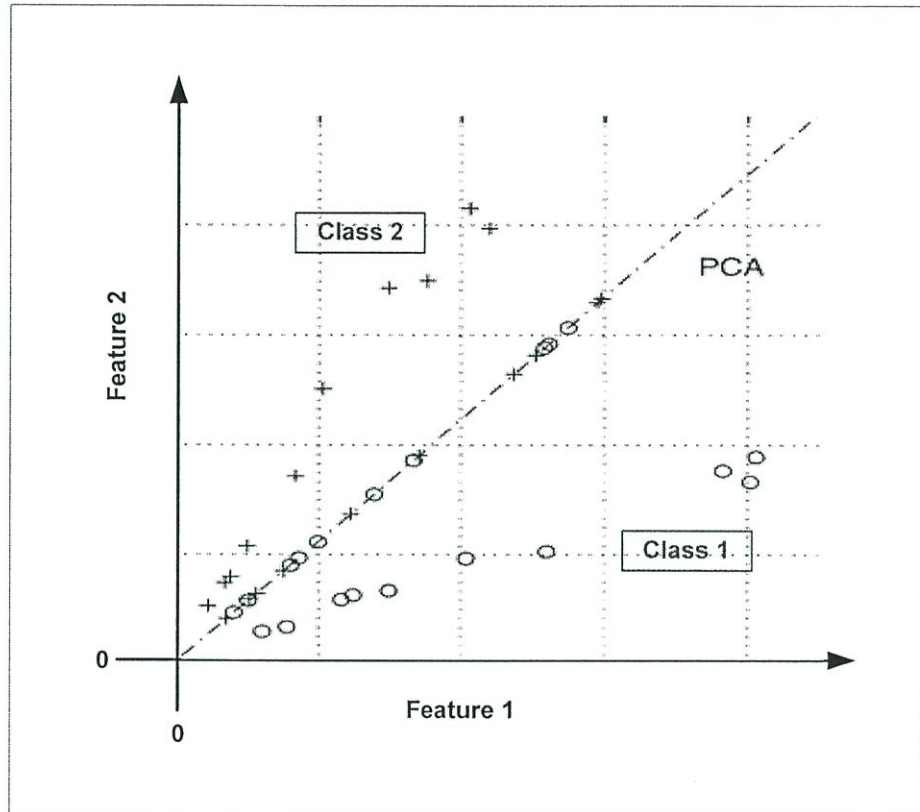
วิธีการวิเคราะห์องค์ประกอบหลักที่เรียกว่า PCA [1], [2] ถูกใช้ในการลดมิติ โดยแปลงเมตริกซ์ภาพให้เป็นเวกเตอร์ และคำนวณหาเวกเตอร์ที่ดีที่สุดสำหรับการกระจายข้อมูลภาพใบหน้าในปริภูมิภาพอินพุต เวกเตอร์เหล่านี้เป็นตัวกำหนดปริภูมิย่อยของภาพใบหน้า ซึ่งปริภูมิย่อยนี้เรียกว่า ปริภูมิใบหน้า (Face space) ระบบจะใช้เทคนิคการฉายภาพ (Image projection technique) ทำการฉายภาพใบหน้าในชุดฝึกฝนให้เข้าสู่ปริภูมิใบหน้าใหม่ จากนั้นคำนวณหาฟังก์ชันการถ่วงน้ำหนัก (Weighting function) ของการกระจายแต่ละเวกเตอร์ในปริภูมิใบหน้าเก็บไว้ในฐานข้อมูล ส่วนภาพใบหน้าที่ต้องการทดสอบก็จะถูกฉายไปบนปริภูมิใบหน้าเช่นเดียวกัน เมื่อต้องการทดสอบภาพใบหน้า ก็เปรียบเทียบกับปริภูมิใบหน้าที่ถูกฉายให้ตรงกับน้ำหนักชุดฝึกฝนที่เก็บในฐานข้อมูลในขั้นตอนสุดท้าย ก็สามารถระบุความถูกต้องของภาพที่ทดสอบได้

หลักการคำนวณของวิธี PCA

การวิเคราะห์องค์ประกอบหลักเป็นการแทนเซตของข้อมูลเดิมด้วยเซตข้อมูลใหม่ ที่มีขนาดเล็กกว่า ข้อมูลใหม่นี้เรียกว่า องค์ประกอบหลัก (Principal component) จะไม่มีความสัมพันธ์กันระหว่างข้อมูลภายในเซตเลย ถ้ามีเซตข้อมูลขนาด D มิติที่มีความสัมพันธ์ข้อมูลกันอยู่ภายใน สามารถใช้การรวมกันเชิงเส้นในการลดขนาดข้อมูลลงเป็น d มิติ โดยที่ d มีค่าน้อยกว่า D

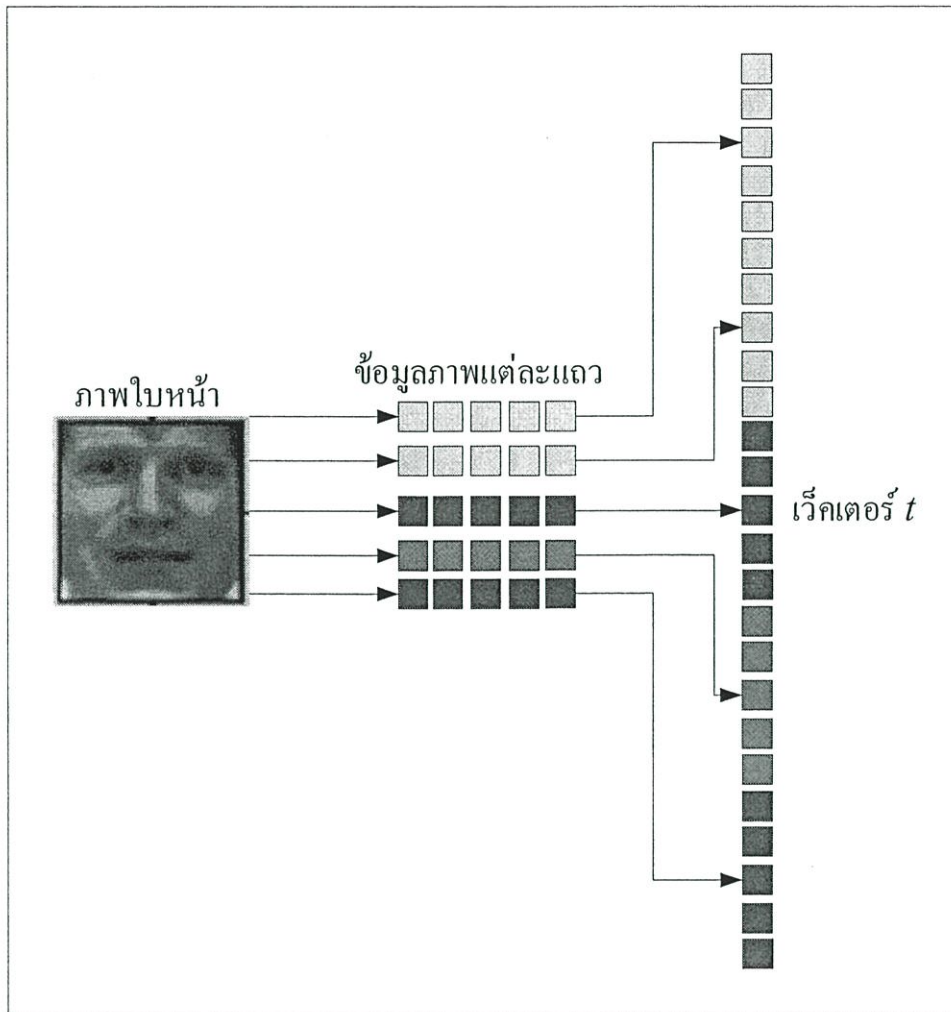
วิธี PCA เริ่มด้วยการนำภาพใบหน้าแต่ละภาพมาแทนด้วยเมตริกซ์ในปริภูมิใบหน้าที่มีมิติเท่ากับจำนวนพิกเซลของภาพ เช่น ภาพระดับสีเทามีขนาด $m \times n$ จะได้มิติของปริภูมิภาพคือ D โดยที่ $D = m \times n$ แล้วทำการลดขนาดมิติให้น้อยลง เพื่อหาคุณลักษณะของภาพ (Image feature) ซึ่งเรียกว่า ปริภูมิใบหน้าไอเกน (Eigenface Space) การหาปริภูมิใบหน้าไอเกนทำได้โดยการคำนวณหาไอเกนเวกเตอร์ (Eigenvector) ของเมตริกซ์ความแปรปรวนร่วมของภาพฝึกฝน (Training images) ที่สมนัยกับค่าไอเกน (Eigenvalue) ที่มีค่าความแปรปรวนจากมากไปหาน้อย เมื่อนำภาพไปฉายด้วยไอเกนเวกเตอร์ที่ได้ ก็จะได้ปริภูมิใบหน้าไอเกน การทดสอบ จะนำภาพฝึกฝนทั้งหมดฉายด้วยวิธีใบหน้าไอเกนเพื่อให้ได้ปริภูมิใบหน้าไอเกนฝึกฝน แล้วนำภาพทดสอบมาฉายด้วยไอเกนเวกเตอร์ของ

ภาพฝึกฝน เพื่อหาปริภูมิใบหน้าไอเกนทดสอบ แล้วจำแนกภาพใบหน้าโดยหาระยะ (Distance) ของ ปริภูมิฝึกฝนกับปริภูมิทดสอบ ตัวอย่างการทำงานของวิธี PCA กับปัญหาข้อมูลสองกลุ่มที่มีข้อมูลอยู่ ใกล้กัน ในปริภูมิเชิงเส้นแสดงในรูปที่ 3.2



รูปที่ 3.2 เทคนิคการฉายภาพแบบ PCA

วิธีการคำนวณ กำหนดให้มีภาพฝึกฝนจำนวน M ภาพ คือ A_1, A_2, \dots, A_M แต่ละภาพที่มี ขนาดมิติ $m \times n$ ถูกเปลี่ยนให้เป็นเวกเตอร์ t ตามรูปที่ 3.3 โดยนำข้อมูลแต่ละแถวของภาพมาเรียง ต่อกันจากบนลงล่างให้เป็นเวกเตอร์ แต่ละเวกเตอร์ที่ได้จึงมีขนาด $D \times 1$



รูปที่ 3.3 การเปลี่ยนเมตริกซ์ภาพเป็นเวกเตอร์

เมื่อภาพฝึกฝนทั้งหมดถูกเปลี่ยนให้เป็นเวกเตอร์แล้ว ก็นำเวกเตอร์ภาพฝึกฝนทั้งหมดมาเรียงตามคอลัมน์จะได้เป็นเมตริกซ์ของชุดภาพตัวอย่างฝึกฝนตามสมการที่ 3.1

$$T = [t_1, t_2, \dots, t_M] \quad (3.1)$$

ซึ่ง t_1, t_2, \dots, t_M คือ เวกเตอร์ข้อมูลค่าระดับสีเทา (Gray scale) ของจุดภาพในแต่ละภาพใบหน้า คำนวณหาค่าเฉลี่ยของภาพใบหน้า (Mean Face) เพื่อหาจุดศูนย์กลางของข้อมูลใหม่

$$m = \frac{1}{M} \sum_{i=1}^M t_i \quad (3.2)$$

คำนวณหาภาพที่ถูกลบด้วยค่าเฉลี่ยของภาพใบหน้า (Mean subtracted image)

$$T_i = t_i - m \quad (3.3)$$

สร้างเมตริกซ์ที่ใช้ในการคำนวณหาเมตริกซ์ความแปรปรวนร่วม

$$T = [T_1, T_2, \dots, T_M] \quad (3.4)$$

คำนวณเมตริกซ์ความแปรปรวนร่วมเพื่อวัดการกระจายของข้อมูล

$$S_T = \frac{1}{M} (T \cdot T^T) = \frac{1}{M} \sum_{i=1}^M T_i \cdot T_i^T \quad (3.5)$$

คำนวณไอเกนเวกเตอร์ที่สมนัยกับค่าไอเกน จากเมตริกซ์ความแปรปรวนร่วมด้วยสมการ

$$S_T u = \lambda u \quad (3.6)$$

เรียงลำดับไอเกนเวกเตอร์ที่สมนัยกับค่าไอเกนจากมากไปหาน้อย แล้วเลือกเฉพาะไอเกนเวกเตอร์ที่สัมพันธ์กับค่าไอเกนที่มากที่สุด ดังนั้นจะได้เมตริกซ์ใหม่ของไอเกนเวกเตอร์คือ

$$U_{PCA} = [u_1 \ u_2 \ u_3 \ \dots \ u_D] \quad (3.7)$$

การหาคุณลักษณะของภาพ (Image feature extraction)

เมื่อได้เวกเตอร์ U_{PCA} ที่เหมาะสมที่ใช้ในการฉายภาพของวิธี PCA โดยที่ d เป็นค่าที่เลือกระหว่าง 1 ถึง D และถ้ามีภาพฝึกฝนทั้งหมด M ภาพ จะสามารถหาคุณลักษณะของภาพฝึกฝนทั้งหมดจากสมการ

$$Y_k = U_d^T (t_k - m); \quad k = 1, 2, \dots, M \quad (3.8)$$

การสร้างภาพใหม่ (Reconstructed image)

วิธี PCA สามารถนำเอาคุณลักษณะของภาพและแกนการฉายที่เหมาะสมที่สุดของ PCA มาสร้างเป็นภาพใบหน้าใหม่ (Reconstructed face image) ได้ตามขั้นตอนดังนี้

โดยที่ $Y = [y_1, y_2, \dots, y_d]$ เป็นคุณลักษณะของภาพและ $U_d = [u_1, u_2, \dots, u_d]$ เป็นแกนการฉายที่เหมาะสมที่สุด ดังนั้นจาก $Y = U_d^T (t - m)$ และ U_d เป็นออร์โธนอร์มอล ($U^{-1} = U^T$) จะได้สมการการสร้างภาพใหม่คือ

$$\tilde{t}_k = U_d Y_k + m \quad (3.9)$$

3.1.2 การวิเคราะห์ลักษณะแตกต่างเชิงเส้น (Linear discriminant Analysis: LDA)

ในทางคณิตศาสตร์ ตัวอย่างของกลุ่มข้อมูลทั้งหมดอธิบายด้วยพารามิเตอร์สองตัวคือ เมตริกซ์การกระจายภายในกลุ่ม และเมตริกซ์การกระจายระหว่างกลุ่ม การใช้เซตของเวกเตอร์การฉายที่กำหนดโดยแกนการฉายตามวิธี LDA [2] ตัวอย่างที่ถูกฉายทั้งหมดจะอยู่ในรูปเมตริกซ์การกระจายระหว่างกลุ่มมากที่สุด และเมตริกซ์การกระจายภายในกลุ่มน้อยที่สุด ตามหลักการ Appearance-based ชั้นแรกภาพ

ใบหน้าที่อยู่ในรูปเมตริกซ์ 2 มิติต้องถูกแปลงให้เป็นเวกเตอร์ 1 มิติก่อน วิธี LDA จะคำนวณหาเวกเตอร์ภายใต้ปริภูมิที่ให้การแบ่งแยกกลุ่มต่างๆ ที่ดีที่สุด ยิ่งกว่านั้น ต้องได้จำนวนของคุณลักษณะที่มีความสัมพันธ์เป็นอิสระกับข้อมูลที่ถูกอธิบาย LDA จะสร้างการรวมกันแบบเชิงเส้น (Linear combination) ของกลุ่มข้อมูลเหล่านี้ ซึ่งจะให้ค่าเฉลี่ยความแตกต่างระหว่างกลุ่มที่กำหนดมากที่สุด

แนวคิดของวิธี LDA มาจากความต้องการที่จะฉายข้อมูลจาก D มิติลงบนแกนการฉายในปริภูมิที่มีมิติ d โดยที่ $d < D$ เพื่อให้ข้อมูลนั้นมีการแบ่งแยกของกลุ่มข้อมูลที่ดี ถ้าตัวอย่างข้อมูลมีขนาด D มิติ มีการรวมตัวกันแน่น เมื่อทำการฉายลงบนแกนใดๆ ทำให้เกิดการปนกันของตัวอย่างข้อมูลทุกๆ กลุ่ม จะทำให้กระบวนการรู้จำแย่ลงไปด้วย อย่างไรก็ตามการแก้ปัญหาสามารถทำได้โดยการทดลองเคลื่อนแกนการฉายไปรอบๆ เพื่อดูตำแหน่งของแกนการฉาย ที่ทำให้การฉายตัวอย่างข้อมูลมีการแบ่งแยกที่ดี

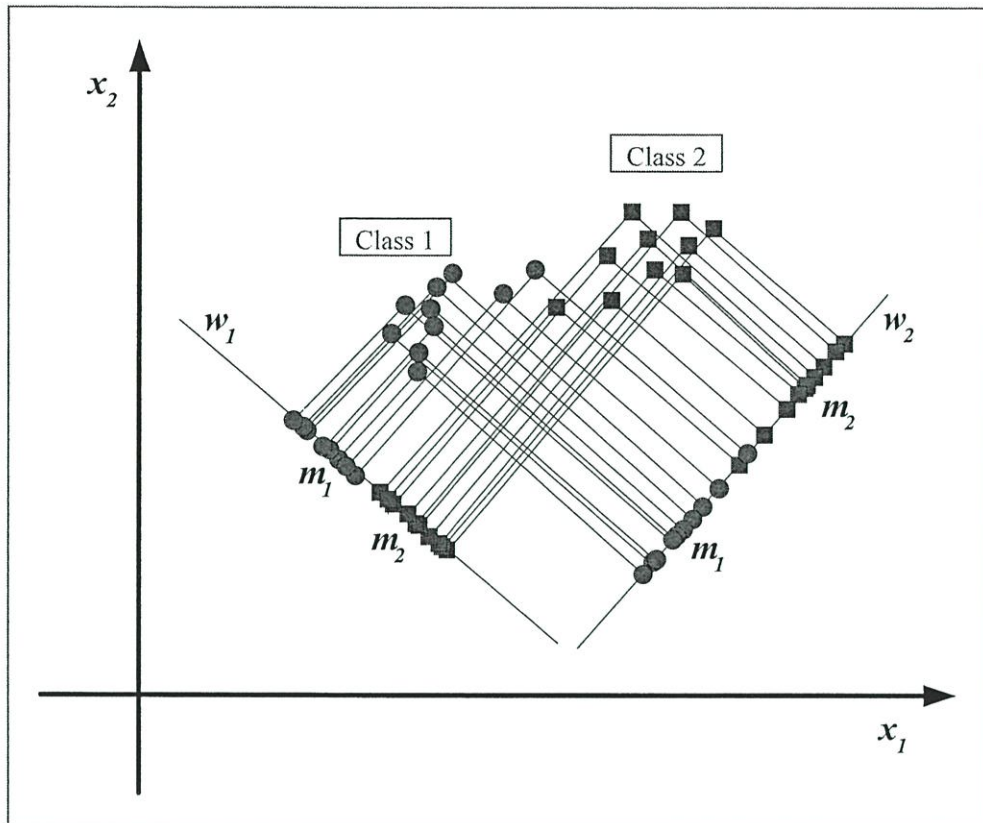
หลักการคำนวณของวิธี LDA

การหาแกนการฉายที่ทำให้ตัวอย่างข้อมูลแบ่งแยกได้ดี ขั้นแรกได้ทำการทดลองกับตัวอย่างเพียง 2 กลุ่มข้อมูล เพื่อดูความสัมพันธ์ หลังจากนั้นจึงขยายเป็นจำนวนกลุ่มที่มากขึ้น เมื่อกำหนดให้จำนวนตัวอย่างข้อมูลทั้งหมดคือ t_1, t_2, \dots, t_M แต่ละตัวอย่างมีขนาด D มิติ โดยถูกแบ่งออกเป็นสองกลุ่ม C_1 และ C_2 กลุ่มที่ 1 มีจำนวนตัวอย่าง N_1 และกลุ่มที่ 2 มีจำนวนตัวอย่าง N_2 จากหลักการรวมกันเชิงเส้นจะได้สมการ

$$y = w^T t \quad (3.10)$$

จะได้ y_1, y_2, \dots, y_M เป็นตัวอย่างที่ถูกฉายซึ่งถูกแบ่งเป็นสองกลุ่มคือ C_1 และ C_2 โดยที่แต่ละค่าของ y_i มาจากการฉายค่าของ t_i ลงบนเส้นในทิศทางของ w

จากแนวความคิดของวิธีการนี้คือ ต้องการให้ตัวอย่างในกลุ่ม C_1 ที่ถูกฉายตกลงบนเส้นให้อยู่ในกลุ่มเดียวกัน ส่วนตัวอย่างที่เหลือก็ให้ตกลงบนเส้นแล้วมาอยู่ในกลุ่มที่เหลือจึงจะทำให้การแบ่งแยกมีประสิทธิภาพ ดังนั้นสิ่งที่สำคัญที่สุดคือการเลือกแกนการฉาย w รูปที่ 3.4 แสดงผลของการเลือกค่าของ w ที่ใช้ในการฉายของตัวอย่างข้อมูลขนาด 2 มิติ จากรูปจะเห็นว่าเมื่อเลือกแกนการฉาย w_2 จะทำให้ตัวอย่างที่ถูกฉายของสองกลุ่มมีการปนกัน แต่ถ้าเลือกแกนการฉาย w_1 จะทำให้กลุ่มตัวอย่างมีการแบ่งแยกกันอย่างชัดเจน เมื่อเลือกแกนการฉายเป็น w_1 ตัวอย่างข้อมูลที่ถูกฉายลงบนแกนนี้จะมีการปนกันน้อยกว่าเมื่อเลือกแกนการฉายเป็น w_2 ทั้งที่ระยะห่างระหว่างค่าเฉลี่ยของตัวอย่างข้อมูลที่ถูกฉายของแกนการฉาย w_2 มีค่ามากกว่าแกนการฉาย w_1 ซึ่งแสดงให้เห็นว่าการใช้ระยะห่างระหว่างค่าเฉลี่ยของตัวอย่างที่ถูกฉายเพียงอย่างเดียวยังไม่เพียงพอที่จะทำให้การฉายตัวอย่างมีการแบ่งแยกที่ดี



รูปที่ 3.4 การเลือกแกนการฉายที่เหมาะสม

Fisher ได้นำเสนอการแก้ปัญหาี้ โดยการหาค่ามากที่สุดของฟังก์ชันอัตราส่วนของระยะทางระหว่างค่าเฉลี่ยของตัวอย่างข้อมูลที่ถูกฉายกับผลรวมการกระจายภายในกลุ่มของตัวอย่างข้อมูลที่ถูกฉาย เมื่อการกระจายภายในกลุ่มมีค่าเท่ากับความแปรปรวนของตัวอย่างภายในกลุ่ม โดยที่การกระจายของแต่ละกลุ่มตัวอย่างที่ถูกฉายคือ

$$\tilde{S}_i^2 = \sum_{y \in C_i} (y - \tilde{m}_i)(y - \tilde{m}_i)^T \quad (3.11)$$

ซึ่ง $\tilde{S}_1^2 + \tilde{S}_2^2$ เรียกว่า การกระจายภายในกลุ่มของตัวอย่างที่ถูกฉาย

ดังนั้นสมการการแบ่งแยกเชิงเส้นที่ได้จากการแปลงเชิงเส้น $Y = w^T t$ ซึ่งมีค่ามากที่สุดของข้อมูล 2 กลุ่มจะได้สมการวัตถุประสงค์เป็น

$$J(w) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{S}_1^2 + \tilde{S}_2^2} \quad (3.12)$$

สมการที่ 3.12 มีค่ามากที่สุดจะทำให้การแบ่งแยกระหว่างข้อมูล 2 กลุ่มทำได้ดี เมื่อได้แนวคิดในการแบ่งแยกข้อมูลที่ดีแล้ว ลำดับต่อมาคือหาวิธีในการหาแกนการฉายของตัวอย่างข้อมูล โดยสามารถ

หาจากเมตริกซ์การกระจายภายในกลุ่ม และเมตริกซ์การกระจายระหว่างกลุ่มของตัวอย่างข้อมูล กำหนดให้เมตริกซ์การกระจายภายในกลุ่มคือ

$$S_i = \sum_{t \in C_i} (t - m_i)(t - m_i)^T \quad (3.13)$$

กรณีพิจารณาแค่ 2 กลุ่มจะได้เมตริกซ์การกระจายภายในกลุ่มเป็น

$$S_W = S_1 + S_2 \quad (3.14)$$

ความสัมพันธ์ของเมตริกซ์การกระจายภายในกลุ่มของตัวอย่างข้อมูลที่ถูกฉายกับเมตริกซ์การกระจายภายในกลุ่มของตัวอย่างข้อมูลคือ

$$\begin{aligned} \tilde{S}_i &= \sum_{y \in C_i} (y - \tilde{m}_i)^2 \\ &= \sum_{t \in C_i} (w^T t - w^T m_i)^2 \\ &= \sum_{t_k \in C_i} w^T (t_k - m_i)(t_k - m_i)^T w \\ &= w^T S_i w \end{aligned} \quad (3.15)$$

ดังนั้นผลรวมของเมตริกซ์การกระจายของข้อมูลที่ถูกฉายจะเป็นดังนี้

$$\tilde{S}_1^2 + \tilde{S}_2^2 = w^T S_W w \quad (3.16)$$

ทำนองเดียวกัน ผลต่างของค่าเฉลี่ยที่ถูกฉายสามารถแสดงในรูปผลต่างของค่าเฉลี่ยตัวอย่างข้อมูลได้คือ

$$\begin{aligned} |\tilde{m}_1 - \tilde{m}_2|^2 &= |w^T m_1 - w^T m_2|^2 \\ &= w^T (m_1 - m_2)(m_1 - m_2)^T w \\ &= w^T S_B w \end{aligned} \quad (3.17)$$

S_W เรียกว่า เมตริกซ์การกระจายภายในกลุ่ม และ S_B เรียกว่า เมตริกซ์การกระจายระหว่างกลุ่ม ดังนั้นจากสมการที่ 3.12 สามารถแสดงสมการในตัวแปรของ S_W และ S_B เป็น

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \quad (3.18)$$

ในการแก้สมการ 3.18 เพื่อหาแกนการฉาย โดยการ Differentiate เทียบกับ w และกำหนดให้มีค่าเท่ากับศูนย์จะได้

$$S_B w = \lambda S_W w \quad (3.19)$$

เมื่อ S_W ไม่เป็นเอกฐาน (Non-singular) สามารถเปลี่ยนสมการใหม่เป็น

$$S_W^{-1} S_B w = \lambda w \quad (3.20)$$

แต่ในกรณีนี้พิจารณาจำนวนกลุ่มเพียง 2 กลุ่ม จึงไม่จำเป็นจะต้องแก้สมการที่ 3.20 โดยการหาค่าไอเกนเวคเตอร์และค่าไอเกนของ $S_W^{-1} S_B$ เพราะในความเป็นจริง $S_B w$ จะมีทิศทางเดียวกับ $(m_1 - m_2)$ ดังนั้นการหาค่า w ที่เหมาะสมที่สุดในกรณีนี้คือ

$$w = S_W^{-1}(m_1 - m_2) \quad (3.21)$$

จะได้แกนการฉาย w ที่มาจากการแบ่งแยกเชิงเส้นสามารถคำนวณได้โดยการหาค่าอัตราส่วนที่มากที่สุดของการกระจายภายในกลุ่มและการกระจายระหว่างกลุ่ม จากการศึกษาข้อมูล 2 กลุ่มมาทำการฉายจะพบว่าข้อมูลของตัวอย่าง D มิติจะสามารถลดมิติลงเป็น d มิติได้

หลังจากทดลองใช้ข้อมูล 2 กลุ่มมาทดสอบหาแกนการฉายที่เหมาะสมที่สุด ซึ่งจะแสดงให้เห็นความสัมพันธ์และวิธีในการหาแกนการฉาย ลำดับต่อมาจึงขยายจำนวนกลุ่มของตัวอย่างข้อมูลเป็นจำนวน L กลุ่ม การลดมิติจะมีเมตริกซ์ที่ใช้ในการหาแกนการฉายอยู่สองเมตริกซ์คือ เมตริกซ์การกระจายภายในกลุ่มและเมตริกซ์การกระจายระหว่างกลุ่ม ดังนั้นในกรณีของ L กลุ่ม ก็สามารถหาเมตริกซ์การกระจายภายในกลุ่มและเมตริกซ์การกระจายระหว่างกลุ่มได้ สำหรับสมการแสดงความสัมพันธ์เมตริกซ์การกระจายภายในกลุ่มคือ

$$S_W = \sum_{i=1}^L S_i \quad (3.22)$$

เมื่อการกระจายและค่าเฉลี่ยของแต่ละกลุ่มเป็น

$$S_i = \sum_{t \in C_i} (t - m_i)(t - m_i)^T \quad \text{และ} \quad m_i = \frac{1}{N_i} \sum_{t \in C_i} t \quad (3.23)$$

และเมตริกซ์การกระจายระหว่างกลุ่มคือ

$$S_B = \sum_{i=1}^L N_i (m_i - m)(m_i - m)^T \quad (3.24)$$

โดยที่ m คือ ค่าเฉลี่ยของตัวอย่างทั้งหมด

$$m = \frac{1}{M} \sum_{i=1}^M t_i \quad (3.25)$$

การฉายข้อมูลจากปริภูมิ D มิติ ไปเป็นปริภูมิ $L-1$ มิติด้วยวิธีการแบ่งแยกเชิงเส้นจะประกอบไปด้วยฟังก์ชันการแบ่งแยกทั้งหมด $L-1$

$$y_i = w_i^T t; \quad i = 1, 2, \dots, L-1 \quad (3.26)$$

เปลี่ยนสมการให้อยู่ในรูปของเมตริกซ์จะได้ดังสมการที่ 3.10 เมื่อตัวอย่างในการฉายมีความสัมพันธ์กับตัวอย่างที่ถูกฉาย ดังนั้นจึงสามารถหาเวกเตอร์ค่าเฉลี่ยและเมตริกซ์การกระจายของตัวอย่างที่ถูกฉายได้ดังนี้

$$\begin{aligned} \tilde{m}_i &= \frac{1}{N_i} \sum_{y \in C_i} y \\ \tilde{m} &= \frac{1}{M} \sum_{i=1}^L N_i \tilde{m}_i \\ \tilde{S}_W &= \sum_{i=1}^L \sum_{y \in C_i} (y - \tilde{m}_i)(y - \tilde{m}_i)^T \\ \tilde{S}_B &= \sum_{i=1}^L N_i (\tilde{m}_i - \tilde{m})(\tilde{m}_i - \tilde{m})^T \end{aligned} \quad (3.27)$$

จากแนวความคิดของตัวอย่างที่ถูกแบ่งเป็น 2 กลุ่ม จากความสัมพันธ์ผลรวมของเมตริกซ์การกระจายของข้อมูลที่ถูกฉายดังสมการที่ 3.16 สามารถหาเมตริกซ์การกระจายภายในกลุ่มและเมตริกซ์การกระจายระหว่างกลุ่มของตัวอย่างที่ถูกฉายได้ คือ

$$\tilde{S}_W = w^T S_W w \quad (3.28)$$

$$\tilde{S}_B = w^T S_B w \quad (3.29)$$

ดังนั้นความสัมพันธ์ระหว่างเมตริกซ์การกระจายภายในกลุ่มและเมตริกซ์การกระจายระหว่างกลุ่มที่ใช้ในการฉายข้อมูลให้มีปริภูมิมีขนาดลดลงสามารถทำได้โดยการหาอัตราส่วนที่มากที่สุดของเมตริกซ์การกระจายภายในกลุ่มและเมตริกซ์การกระจายระหว่างกลุ่ม แต่เนื่องจากการวัดการกระจายต้องมีขนาดเป็นสเกลาร์ จึงใช้เทคนิค Determinant เข้ามาช่วย จะได้สมการวัตถุประสงค์ดังนี้

$$J(w) = \frac{|\tilde{S}_B|}{|\tilde{S}_W|} = \frac{|w^T S_B w|}{|w^T S_W w|} \quad (3.30)$$

แกนการฉาย w ที่เหมาะสมเพื่อที่จะทำให้สมการ 3.30 มีค่ามากที่สุดคือ ไอเจนเวกเตอร์ที่สมนัยกับค่าไอเจนที่มากที่สุดของโครงสร้างไอเจน (Eigenstructure) ดังสมการ 3.31

$$S_B w_i = \lambda_i S_W w_i \quad (3.31)$$

คำนวณไอเจนเวกเตอร์ และค่าไอเจนจากเมตริกซ์ $S_W^{-1} S_B$ เมื่อเรียงลำดับไอเจนเวกเตอร์ที่สมนัยกับค่าไอเจนจากมากไปหาน้อย เลือกเฉพาะไอเจนเวกเตอร์ที่สัมพันธ์กับค่าไอเจนที่ค่ามากที่สุด จะได้เมตริกซ์ใหม่ของไอเจนเวกเตอร์คือ

$$W_{LDA} = [w_1 \ w_2 \ w_3 \ \dots \ w_D] \quad (3.32)$$

ดังนั้นจะได้การฉายภาพจากปริภูมิ D มิติ ไปเป็นปริภูมิ d มิติ ที่ทำให้อัตราส่วนระหว่างเมตริกซ์การกระจายระหว่างกลุ่มและเมตริกซ์การกระจายภายในกลุ่มมีค่ามากที่สุด

การหาคุณลักษณะของภาพ

เมื่อได้เวกเตอร์ที่เหมาะสมคือ W_{LDA} เพื่อนำมาใช้ในการฉายภาพของ LDA ที่มีค่าดังสมการที่ 3.32 เป็น $[w_1 \ w_2 \ w_3 \ \dots \ w_d]$ โดยที่ d เป็นค่าที่เลือกระหว่าง 1 ถึง D และมีภาพฝึกฝนทั้งหมด M ภาพ จะสามารถหาคุณลักษณะของภาพฝึกฝนทั้งหมดจากสมการหลักการรวมเชิงเส้นดังนี้

$$y_k = W_d^T (t_k - m); \quad k = 1, 2, \dots, M \quad (3.33)$$

การสร้างภาพใหม่

วิธีของ LDA สามารถนำเอาคุณลักษณะของภาพและแกนการฉายที่เหมาะสมที่สุดของวิธีการ LDA มาสร้างเป็นภาพใบหน้าใหม่ได้ตามขั้นตอนดังนี้

โดยที่ $Y = [y_1, y_2, \dots, y_d]$ เป็นคุณลักษณะของภาพ และ $U_d = [u_1, u_2, \dots, u_d]$ เป็นแกนการฉายที่เหมาะสมที่สุด ดังนั้นจาก $Y = W_d^T (t - m)$ และ W_d เป็นออร์โธโนมอล ($W^{-1} = W^T$) จะได้สมการการสร้างภาพใหม่คือ

$$\tilde{t} = W_d y + m \quad (3.34)$$

3.1.3 การวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติ (2-Dimensional LDA)

วิธี LDA มีข้อเสียอยู่สองประการ คือ ประการแรก LDA ทำงานกับเมตริกซ์ภาพอินพุทโดยตรง เมื่อใช้ภาพอินพุทที่มีส่วนประกอบอื่นของภาพแตกต่างกัน เช่น ฉากหลังภาพไม่เหมือนกัน อาจทำให้การคัดแยกและการระบุความถูกต้องผิดพลาด ประการที่สอง เมื่อเกิดปัญหาจำนวนตัวอย่าง

น้อย เมตริกซ์การกระจายภายในกลุ่มอาจจะเป็นเมตริกซ์เอกฐาน เกิด Singularity problem แม้ว่าการฉายปริภูมิอนพหุมิติขนาดใหญ่ไปสู่ปริภูมิย่อยมิติขนาดเล็กด้วย PCA สามารถแก้ปัญหานี้ได้ แต่ยังคงขาดรายละเอียดโครงสร้างข้อมูลที่สำคัญ

เพื่อแก้ไขข้อบกพร่องเหล่านี้จึงมีผู้วิจัยเสนอผลงานปรับปรุงวิธี LDA บนหลักการ 2DPCA ขึ้นมา เรียกว่า การวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติ (2-Dimensional LDA: 2DLDA) ซึ่งไม่เพียงแก้ปัญหา SSS และลดความยุ่งยากเรื่องมิติขนาดใหญ่ได้เท่านั้น แต่ 2DLDA ยังสามารถทำงานกับภาพโดยตรงกับการฉายภาพทุกทิศทาง ทำให้สามารถรักษารายละเอียดโครงสร้างข้อมูลที่สำคัญไว้ได้ และขนาดของเมตริกซ์การกระจายทั้งหมดเท่ากับความกว้างของภาพใบหน้า จึงไม่ต้องขยายขนาดภาพ สามารถรักษาข้อมูลทั้งหมดไว้ได้ครบถ้วน สรุปวิธี 2DLDA มีข้อดีคือ ง่ายในการคำนวณความถูกต้องของเมตริกซ์ความแปรปรวนร่วม และใช้เวลาน้อยกว่า

วิธีการ 2DLDA เป็นวิธีการใหม่ที่ประยุกต์มาจากวิธี LDA และ 2DPCA โดยนำเมตริกซ์ภาพมาคำนวณหาเมตริกซ์การกระจายภายในกลุ่ม และเมตริกซ์การกระจายระหว่างกลุ่ม แล้วคำนวณหาไอเกนเวกเตอร์ที่สมนัยกับค่าไอเกนจากมากไปหาน้อยตามลำดับ จากนั้นนำภาพใบหน้ามาฉายลงบนไอเกนเวกเตอร์ จะได้เมตริกซ์คุณลักษณะที่มีจำนวนมิติของคอลัมน์ภาพน้อยลง

หลักการคำนวณของวิธี 2DLDA

กำหนดให้ X ประกอบด้วยเวกเตอร์ขนาด n มิติ และ A เป็นภาพที่มีขนาด $m \times n$ จากการฉายภาพ A ลงบนแกน X ตามหลักการแปลงเชิงเส้นจะได้

$$Y = AX \quad (3.35)$$

ดังนั้นจะได้ Y ประกอบด้วยเวกเตอร์ขนาด m มิติ โดยแต่ละเวกเตอร์ของ Y คือเวกเตอร์คุณลักษณะของภาพ A เมื่อกำหนดให้มีตัวอย่างภาพทั้งหมด M ภาพ แต่ละภาพคือ A_j ($j = 1, 2, \dots, M$) ซึ่งแบ่งออกเป็นกลุ่ม แต่ละกลุ่มแทนด้วย C_i ($i = 1, 2, \dots, L$) โดยที่ L คือ จำนวนกลุ่มข้อมูลภาพทั้งหมด \bar{A} คือ ค่าเฉลี่ยรวมที่คำนวณจากภาพทั้งหมด ส่วน \bar{A}_i คือ ค่าเฉลี่ยที่คำนวณเฉพาะภาพภายในกลุ่ม C_i ซึ่งแต่ละกลุ่มมีจำนวน N_i ภาพ หลังจากภาพที่ใช้ฝึกสอนถูกฉายตามสมการที่ (3.35) ลงบนแกน X จะได้เมตริกซ์คุณลักษณะเป็น

$$Y_j = A_j X; \quad j = 1, 2, \dots, M \quad (3.36)$$

วิธีในการหาแกนการฉาย X สามารถคำนวณได้โดยการหาผลรวมค่าไอเกน (Trace) ของเมตริกซ์ความแปรปรวนร่วมของภาพตัวอย่างที่ถูกฉาย จากแนวความคิดนี้สามารถสร้างเป็นสมการได้คือ

$$J(X) = \frac{P_B}{P_W} \quad (3.37)$$

ซึ่ง P_B และ P_W คำนวณจาก

$$P_B = \text{tr}(\tilde{S}_B) \quad (3.38)$$

$$P_W = \text{tr}(\tilde{S}_W) \quad (3.39)$$

โดยที่ \tilde{S}_B คือ เมตริกซ์การกระจายระหว่างกลุ่มของภาพตัวอย่างที่ถูกฉาย และ \tilde{S}_W คือ เมตริกซ์การกระจายภายในกลุ่มของภาพตัวอย่างที่ถูกฉาย

$$\begin{aligned} \tilde{S}_B &= \sum_{i=1}^L N_i (\bar{Y}_i - \bar{Y})(\bar{Y}_i - \bar{Y})^T \\ &= \sum_{i=1}^L N_i [(\bar{A}_i - \bar{A})X][(\bar{A}_i - \bar{A})X]^T \end{aligned} \quad (3.40)$$

$$\begin{aligned} \tilde{S}_W &= \sum_{i=1}^L \sum_{Y_k \in C_i} (Y_k - \bar{Y}_i)(Y_k - \bar{Y}_i)^T \\ &= \sum_{i=1}^L \sum_{A_k \in C_i} [(A_k - \bar{A}_i)X][(A_k - \bar{A}_i)X]^T \end{aligned} \quad (3.41)$$

ถ้า

$$\begin{aligned} \text{tr}(\tilde{S}_B) &= X^T \left[\sum_{i=1}^L N_i (\bar{A}_i - \bar{A})^T (\bar{A}_i - \bar{A}) \right] X \\ &= X^T S_B X \end{aligned} \quad (3.42)$$

$$\begin{aligned} \text{tr}(\tilde{S}_W) &= X^T \left[\sum_{i=1}^L \sum_{A_k \in C_i} (A_k - \bar{A}_i)^T (A_k - \bar{A}_i) \right] X \\ &= X^T S_W X \end{aligned} \quad (3.43)$$

ในการคำนวณหา S_B และ S_W สามารถนำภาพที่ใช้ฝึกสอนมาคำนวณหาได้โดยตรง และจากสมการที่ 3.37 สามารถเปลี่ยนเป็นสมการใหม่ได้คือ

$$J(X) = \frac{X^T S_B X}{X^T S_W X} \quad (3.44)$$

สมการที่ 3.44 มีชื่อว่า สมการหลักการฉายเชิงเส้นที่ยึดถือประเภทเป็นเกณฑ์ (Class specific linear projection criterion) โดยเวกเตอร์ X ที่ทำให้ $J(X)$ มีค่ามากที่สุดมีชื่อว่า แกนการฉายที่เหมาะสมที่สุด

$$X_{opt} = \arg_{\max_X} J(X) \quad (3.45)$$

ถ้า S_W ไม่เป็นเมตริกซ์เอกฐาน สามารถหาเมตริกซ์ X โดยการคำนวณไอเกนเวกเตอร์และค่าไอเกนจากเมตริกซ์ S_B และ S_W ได้จากสมการ

$$S_B X_{opt} = \lambda S_W X_{opt} \quad (3.46)$$

โดยปกติแกนการฉายที่เหมาะสมที่สุดที่ใช้ในการฉายมาจากไอเกนเวกเตอร์ของ $S_W^{-1}S_B$ ที่สัมพันธ์กับค่าไอเกนที่มากที่สุดตั้งแต่ 1 ถึง n ซึ่งแต่ละไอเกนเวกเตอร์เป็นออร์โธโนมอล (Orthonormal)

$$X_{2DLDA} = [x_1, x_2, \dots, x_n] \quad (3.47)$$

การหาคุณลักษณะของภาพ

เวกเตอร์ที่เหมาะสมที่ถูกนำมาใช้หาคุณลักษณะในการฉายภาพของ 2DLDA คือ x_1, x_2, \dots, x_n ดังนั้น จึงเลือกแกนการฉายเพียง 1 ถึง d ในการหาคุณลักษณะของภาพเท่านั้น

$$Y_k = AX_k; \quad k=1, 2, \dots, d \quad (3.48)$$

เวกเตอร์ใหม่ที่ได้จากการฉาย y_1, y_2, \dots, y_d มีชื่อว่า เวกเตอร์คุณลักษณะ เมื่อนำเวกเตอร์คุณลักษณะรวมเป็นเมตริกซ์ $Y = [y_1, y_2, \dots, y_d]$ ที่มีขนาดเป็น $m \times d$ จะได้เมตริกซ์คุณลักษณะ Y ของภาพ A

การสร้างภาพใหม่

วิธีของ 2DLDA สามารถนำเอาเมตริกซ์คุณลักษณะ และแกนการฉายที่เหมาะสมที่สุดมาสร้างเป็นภาพใบหน้าใหม่ได้โดย กำหนดให้ A เป็นภาพ โดยที่ $Y = [y_1, y_2, \dots, y_d]$ เป็นเมตริกซ์คุณลักษณะของภาพ A และ $X_{2DLDA} = [x_1, x_2, \dots, x_n]$ เป็นแกนการฉายที่เหมาะสมที่สุด จากสมการ $Y = AX$ และ X เป็นออร์โธโนมอล $X^{-1} = X^T$ ดังนั้นสมการในการสร้างภาพใหม่คือ

$$\tilde{A} = YX^T = \sum_{k=1}^d y_k X_k^T \quad (3.49)$$

ในการสร้างภาพใบหน้าใหม่ขึ้นถ้าเลือกจำนวนแกนการฉาย d เท่ากับ n จะได้ภาพที่สร้างใหม่เหมือนกับภาพ A แต่ถ้าเลือก d น้อยกว่า n ภาพ \tilde{A} จะมีความเหมือนกับภาพ A น้อยลงตามลำดับ

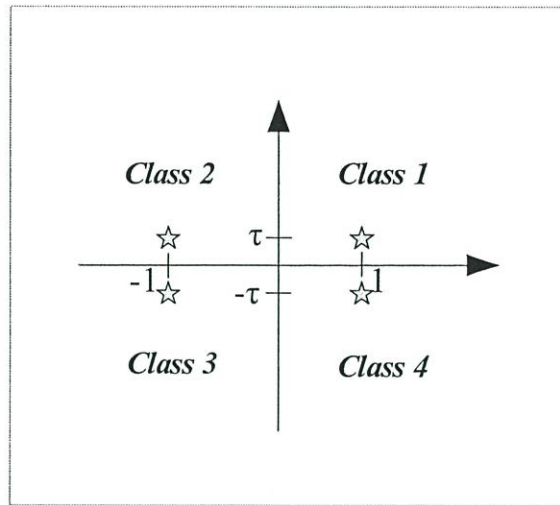
3.1.4 แนวทางการกระจายแบบคู่ (Pairwise scatter approach)

ในข้อมูลแต่ละชุดนั้น ระยะห่างหรือความสัมพันธ์ที่เหมือนกันระหว่างคู่ของกลุ่มข้อมูลเป็นสิ่งสำคัญสำหรับการคัดแยก ระยะหรือความสัมพันธ์ที่เหมือนกันนี้ โดยปกติจะได้มาด้วยการใช้แนวทางทางสถิติ เช่น ระยะยูคลิเดียน หรือระยะมาฮาလာโนบิส เป็นต้น ส่งผลให้การแยกกลุ่มข้อมูลสองกลุ่มได้ดีเพียงใดในปริภูมิคุณลักษณะ ยิ่งกว่านั้น ในระบบหลายๆ กลุ่ม (Multiclass) ตามวิธี LDA ความสัมพันธ์ระหว่างคู่ของกลุ่ม (Pairwise relationships) จะทำให้ต่างจากกลุ่มอื่นๆ ข้อมูลบางกลุ่มที่ใกล้ชิดกับกลุ่มอื่นๆ แต่ละกลุ่มจะทำให้ข้อมูลสับสนยุ่งยากมาก และมีผลต่อขั้นตอนการคัดแยกลักษณะเด่นด้วยงานวิจัย LDA ทั้งหมดจึงทำงานโดยใช้แนวทางเมตริกซ์การกระจายแบบคู่ (Pairwise scatter matrix) ซึ่งจะพิจารณาเฉพาะเมตริกซ์การกระจายระหว่างกลุ่มเท่านั้น คำตอบในการแก้ปัญหาข้อมูลบางกลุ่มที่ใกล้ชิดกับกลุ่มอื่นๆ นี้จะใช้ ไอเกนเวกเตอร์ที่สมนัยกับค่าไอเกนที่เรียงจากมากไปหาน้อยจากสมการ $S_B w_i = \lambda_i S_W w_i$ รูปแบบของเมตริกซ์การกระจายระหว่างกลุ่มสามารถหาได้จาก

$$S_B = \sum_{i=1}^L N_i (m_i - m)(m_i - m)^T \quad \text{โดยที่ } m \text{ คือ ค่าเฉลี่ยของตัวอย่างทั้งหมด}$$

สมการนี้เป็นตัววัดว่า ค่าเฉลี่ยของแต่ละกลุ่มห่างจากจุดศูนย์กลางข้อมูลถูกกระจายอย่างไร เป็นการดีที่จะได้ค่า S_B มากๆ เพราะแสดงว่ากลุ่มต่างๆ ถูกกระจายออกไปในปริภูมิที่ถูกแปลง (Transformed space) มากกว่าเดิม ทำให้แบ่งแยกง่ายขึ้น แต่ยังไม่ชัดเจนว่า ข้อมูลกลุ่มต่างๆ ถูกแยกออกจากกันได้อย่างไร ที่อยู่ติดกันได้อย่างไร เพื่อทำความเข้าใจกับประเด็นนี้ จึงเสนอตัวอย่างไดอะแกรมดังรูปที่ 3.5

พิจารณาข้อมูลในกลุ่มตามรูปที่ 3.5 กำหนดให้มีข้อมูลอยู่ 4 กลุ่ม แต่ละกลุ่มมีจำนวนของเวกเตอร์คุณลักษณะและความแปรปรวนเหมือนกัน ข้อมูลกลุ่มที่ 1 มีค่าเฉลี่ยเป็น $m_1 = (1, \tau)$ กลุ่มที่ 2 เป็น $m_2 = (-1, \tau)$ กลุ่มที่ 3 เป็น $m_3 = (-1, -\tau)$ และกลุ่มที่ 4 เป็น $m_4 = (1, -\tau)$ ตามลำดับ สำหรับกรณีนี้ จะได้เมตริกซ์การกระจายระหว่างกลุ่มดังนี้



รูปที่ 3.5 กลุ่มข้อมูลตำแหน่งต่างๆ ที่มีความแปรปรวนเท่ากัน

$$\frac{1}{4}S_B = \begin{pmatrix} 1 & 0 \\ 0 & \tau^2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad (3.50)$$

เมื่อ $\tau \rightarrow 0$ เมตริกซ์การกระจายระหว่างกลุ่มจะไม่มี การแบ่งแยกในทิศทางแนวตั้ง แต่จะมีการแบ่งแยกในทิศทางแนวนอนเท่านั้น กล่าวได้ว่า เมตริกซ์การกระจายระหว่างกลุ่มจะถูกครอบงำด้วยความแปรปรวนร่วมของคู่ข้อมูลกลุ่มอื่นๆ มากกว่าคู่ข้อมูลกลุ่ม (1,4) และ กลุ่ม (2,3) เห็นได้ชัดว่า ในเรื่องปัญหาการระบุ ความแปรปรวนร่วมของคู่ข้อมูลกลุ่ม (1,4) และ กลุ่ม (2,3) สำคัญกว่ากลุ่มอื่นๆ เนื่องจากมันติดกันมากกว่า ขณะนี้ การกระจายระหว่างกลุ่มแบบเดิมไม่ได้จับข้อมูลเหล่านี้อย่างเด่นชัดสรุปได้ว่า วิธีที่เป็นที่ยอมรับไม่นำเสนอการแบ่งแยกข้อมูลที่กำหนดอย่างถูกต้อง

จากตัวอย่างข้างบนนี้จะเห็นได้ว่า ถ้ามีกลุ่มข้อมูลบางกลุ่มอยู่ชิดมากกว่าเมื่อเทียบกับกลุ่มอื่น เมตริกซ์การกระจายระหว่างกลุ่มส่วนมากจะละเว้นการแบ่งแยกข้อมูลระหว่างกลุ่มที่ใกล้ชิดกับกลุ่มอื่นๆ แต่ละกลุ่ม เพื่อที่จะรักษาข้อมูลให้แยกห่างจากกันได้เพียงพอที่จะทำการแบ่งแยกข้อมูลแต่ละกลุ่ม ต้องปรับตัวถ่วงน้ำหนักให้เหมาะสม โดยธรรมชาติแล้ววิธีที่นำมาใช้คือ การนอร์มอลไลซ์ตัวถ่วงน้ำหนักด้วยกำลังสองของส่วนกลับของระยะยูคลิดีียนระหว่างกลุ่มข้อมูล ดังแสดงได้ในสมการที่ 3.51

$$\omega_B = \frac{1}{\|m_i - m\|} = \frac{1}{(m_i - m)^T (m_i - m)} \quad (3.51)$$

เราถ่วงน้ำหนักกลุ่มข้อมูลที่มีค่าเฉลี่ยของมันใกล้ชิดกับกลุ่มอื่นๆ มากกว่ากลุ่มที่มีค่าเฉลี่ยไกลออกไป ด้วยหลักนี้ กลุ่มข้อมูลที่มีความยุ่งยากมากกว่าจะถูกถ่วงน้ำหนักมากกว่า และกลุ่มข้อมูลที่ยุงยากน้อยกว่าก็จะถูกถ่วงน้ำหนักน้อยกว่า ดังนั้น ตามที่มีการนอร์มอลไลซ์ตัวถ่วงน้ำหนัก เราสามารถปรับเปลี่ยนสมการเมตริกซ์การกระจายระหว่างกลุ่ม ได้ดังสมการข้างล่างนี้

$$S_{B_{norm}} = \frac{1}{2N} \sum_{i=1}^{L-1} \sum_{j=i+1}^L \omega_B N_i N_j (m_i - m_j)(m_i - m_j)^T \quad (3.52)$$

สำหรับปัญหาที่กำหนดในรูปที่ 3.5 ความแปรปรวนร่วมระหว่างกลุ่ม สามารถคำนวณได้ดังนี้

$$\frac{1}{4} S_{B_{norm}} = \begin{pmatrix} 1 + \frac{1}{1 + \tau^2} & 0 \\ 0 & 1 + \frac{\tau^2}{1 + \tau^2} \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad (3.53)$$

ดังนั้น ด้วยการใช้เมตริกซ์ความแปรปรวนร่วมระหว่างกลุ่มแบบใหม่ดังสมการที่ 3.53 จะไม่มีผลอะไรกับคู่ข้อมูล (1,4) และคู่ (2,3) ที่ติดกัน การคำนวณของ (2D) LDA จะพิจารณาการกระจายข้อมูลเหล่านี้ ทำให้ได้ผลเป็นที่น่าพอใจครอบคลุมปัญหาการคัดแยกข้อมูล

3.1.5 วิธี 2DLDA อีกแนวทางหนึ่ง (An alternative 2DLDA)

2DLDA แบบเดิมกำหนดให้มีการทำงานในทิศทางด้านแถวของเมตริกซ์ข้อมูลภาพ จึงกำหนดสัญลักษณ์ของตัวแปรของวิธีสองมิติแบบเดิมนี้นี้ขึ้นใหม่ให้เป็นสัญลักษณ์เฉพาะสำหรับการทำงานข้อมูลด้านแถว และสอดคล้องกับการทำงานสำหรับข้อมูลด้านคอลัมน์ที่จะมีต่อไป อีกทั้งเพื่อไม่ให้สับสนกับตัวแปรในวิธี 2DLDA เดิม และแยกให้เห็นการทำงานด้านแถวของภาพให้ชัดเจน ดังนั้น จึงกำหนดให้ G_B คือ เมตริกซ์การกระจายระหว่างกลุ่ม และ G_W คือ เมตริกซ์การกระจายภายในกลุ่ม ซึ่งจะปฏิบัติตามสมการดังนี้

$$G_B = \frac{1}{2N} \sum_{i=1}^{L-1} \sum_{j=i+1}^L N_i N_j (\bar{A}_i - \bar{A}_j)^T (\bar{A}_i - \bar{A}_j) \quad (3.54)$$

$$G_W = \frac{1}{N} \sum_{i=1}^L \sum_{A_k \in C_i} (A_k - \bar{A}_i)^T (A_k - \bar{A}_i) \quad (3.55)$$

สมการที่ 3.55 แสดงเมตริกซ์การกระจาย G_W ที่ได้จาก Outer product ของเวกเตอร์แถวของภาพ สมมติให้ภาพฝึกฝนมีค่าเฉลี่ยเป็นศูนย์ เช่น $\bar{A} = (0)_{m \times n}$ ด้วยเหตุนี้ จึงกล่าวได้ว่า วิธี 2DLDA แบบเดิมเป็นการทำงานด้านแถวของภาพ ต่อไปจะพิจารณาใช้ Outer product ระหว่างเวกเตอร์คอลัมน์ของภาพเพื่อคำนวณหาเมตริกซ์ G_B และ G_W

กำหนดให้ $A_k = [(A_k^{(1)})(A_k^{(2)}) \dots (A_k^{(m)})]$ และ $\bar{A} = [(\bar{A}^{(1)})(\bar{A}^{(2)}) \dots (\bar{A}^{(m)})]$ ซึ่ง $A_k^{(j)}$ และ $\bar{A}^{(j)}$ คือ เวกเตอร์คอลัมน์ที่ j ของ A_k และ \bar{A} ตามลำดับ

เพื่อแนะนำวิธี 2DLDA อีกแนวทางหนึ่งที่เรียกว่า Alternative 2DLDA จะนำเสนอเมตริกซ์การกระจายระหว่างกลุ่ม H_B และเมตริกซ์การกระจายภายในกลุ่ม H_W ให้ถูกคำนวณดังต่อไปนี้

$$H_B = \frac{1}{2N} \sum_{i=1}^{L-1} \sum_{j=i+1}^L N_i N_j (\bar{A}_i - \bar{A}_j)(\bar{A}_i - \bar{A}_j)^T \quad (3.56)$$

$$H_W = \frac{1}{N} \sum_{i=1}^L \sum_{A_k \in C_i} (A_k - \bar{A}_i)(A_k - \bar{A}_i)^T \quad (3.57)$$

จะเห็นได้ว่า H_B และ H_W ในสมการที่ 3.56 และ 3.57 รูปแบบใหม่นี้เป็น Outer product ของคอลัมน์เวกเตอร์ ซึ่งไม่เหมือน G_B และ G_W ในสมการที่ 3.76 และ 3.77 ในกรณีของวิธี 2DLDA แบบเดิม การใช้เมตริกซ์การกระจายทั้งสองนี้เหมือนกับวิธี 2DLDA เริ่มแรก รูปแบบที่นำเสนอพบว่า เกณฑ์การฉายที่เหมาะสมคือ $W(m \times q)$ เพื่อให้การกระจายรวมของตัวอย่างที่ถูกฉายแล้วมีค่าสูงสุด โดยใช้สมการหลักเกณฑ์ของฟิชเชอร์ (Fisher's criterion) เช่นเดียวกัน แสดงได้โดย

$$J(W) = \frac{WH_B W^T}{WH_W W^T} \quad (3.58)$$

ดังนั้น ไอเกนเวกเตอร์ของเมตริกซ์ $H_W^{-1}H_B$ จะถูกคำนวณ และจำนวน q ของไอเกนเวกเตอร์ที่ตรงกับค่าไอเกนที่มากที่สุดค่าแรกของ $H_W^{-1}H_B$ จะถูกเลือก สุดท้ายการฉายของภาพฝึกฝนหนึ่งภาพบนเกณฑ์การฉายที่เหมาะสมทำให้ได้เมตริกซ์คุณลักษณะของภาพฝึกฝนตามลำดับ นั่นคือ ถ้า Z แทน เมตริกซ์คุณลักษณะของภาพ A ดังนั้น

$$Z = W^T A \quad (3.59)$$

สมการที่ 3.57 แสดงว่า เมตริกซ์ความแปรปรวนร่วม H_W ได้จาก Outer product ของคอลัมน์เวกเตอร์ของภาพฝึกฝน ถ้าสมมุติให้มีค่าเป็นศูนย์ ก็สามารถสรุปได้ว่า วิธี 2DLDA อีกแนวทางหนึ่งที่น่าสนใจในการทำงานในทิศทางด้านคอลัมน์ของภาพ

3.1.6 การวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติสองทิศทาง ((2D)²LDA)

วิธีการ (2D)²LDA เป็นวิธีการหาคุณลักษณะของภาพที่พัฒนามาจากวิธีการหาคุณลักษณะบนพื้นฐานของ LDA ต่อจากวิธี 2DLDA ซึ่งวิธีการนี้เป็นการพัฒนาคุณลักษณะให้สามารถลดมิติข้อมูลได้ทั้งด้านแถวและคอลัมน์ของภาพ

ถ้ากำหนดให้ X แทนเมตริกซ์การฉายที่เหมาะสมตามทิศทางด้านแถวของข้อมูลขนาด $m \times r$ ซึ่งได้อธิบายตามวิธี 2DLDA ในหัวข้อ 3.1.3 แล้ว และให้ W แทนเมตริกซ์การฉายที่เหมาะสมตามทิศทางด้านคอลัมน์ของข้อมูลขนาด $n \times c$ ที่ได้จากวิธี Alternative 2DLDA ในหัวข้อ 3.1.5 และขั้นตอนการฝึกฝนระบบ ภาพฝึกฝน A แต่ละภาพจะถูกฉายไปบนเมตริกซ์การฉาย X และ W ตามลำดับ เพื่อคำนวณหาเมตริกซ์คุณลักษณะ F ที่มีขนาด $r \times c$ จะใช้ความสัมพันธ์ตามสมการต่อไปนี้

$$F = W^T AX \quad (3.60)$$

เมื่อใช้ในกระบวนการแทนภาพ (Image representation) เมตริกซ์ F มักถูกเรียกอีกอย่างหนึ่งว่า เมตริกซ์สัมประสิทธิ์ (Coefficient matrix) สำหรับกระบวนการรู้จำใบหน้าจะเรียกเมตริกซ์ F ว่า เมตริกซ์คุณลักษณะ (Feature matrix)

หลักการคำนวณของวิธี (2D)²LDA

กำหนดให้เมตริกซ์ A แทนภาพที่มีขนาด $m \times n$ ซึ่งมีจำนวนภาพตัวอย่างทั้งหมด M ภาพ ดังนั้น ภาพตัวอย่างแต่ละภาพจะแทนด้วย $A_j = (j = 1, 2, \dots, M)$

ส่วนเมตริกซ์ $U \in R^{m \times r}$ และ $V \in R^{n \times c}$ เป็นเมตริกซ์การฉาย จะทำการฉายเมตริกซ์ภาพ A ลงบนเมตริกซ์การฉาย U และ V ตามหลักการแปลงเชิงเส้นจะได้สมการเมตริกซ์คุณลักษณะดังนี้

$$Y = U^T AV \quad (3.61)$$

ซึ่ง Y คือ เมตริกซ์คุณลักษณะของภาพ A ที่มีขนาด $r \times c$ โดยที่ $r < m$ และ $c < n$

เมื่อนำภาพตัวอย่างทั้งหมด M ภาพ แยกออกเป็นกลุ่มไปได้จำนวน L กลุ่ม ให้แต่ละกลุ่มมีจำนวนภาพในกลุ่ม N_i ภาพ ($i = 1, 2, \dots, L$) โดยมีค่าเฉลี่ยของภาพฝึกฝนทั้งหมดเป็น \bar{A} และกำหนดให้ \bar{A}_i คือ ค่าเฉลี่ยของภาพแต่ละกลุ่ม

หลังจากภาพตัวอย่างฝึกฝนกลุ่ม C_i ถูกฉายลงบนเมตริกซ์การฉาย U และ V ตามลำดับ จะได้เมตริกซ์คุณลักษณะเป็น

$$Y_j = U^T A_j V, \quad j = 1, 2, \dots, M \quad (3.62)$$

การหาเมตริกซ์การฉาย U และ V สามารถคำนวณได้โดยการหาผลรวมค่าไอเกนของเมตริกซ์ความแปรปรวนร่วมจากภาพตัวอย่างที่ถูกฉาย จากแนวความคิดสามารถสร้างเป็นสมการได้คือ

$$J(U, V) = \frac{P_B}{P_W} \quad (3.63)$$

ซึ่ง P_B และ P_W คำนวณได้จากสมการที่ 3.38 และ 3.39 ตามลำดับ ดังนั้น การคำนวณค่า \tilde{S}_B และ \tilde{S}_W แสดงได้ดังต่อไปนี้

$$\begin{aligned}\tilde{S}_B &= \sum_{i=1}^L N_i (\bar{Y}_i - \bar{Y})(\bar{Y}_i - \bar{Y})^T \\ &= \sum_{i=1}^L N_i [U(\bar{A}_i - \bar{A})V][U(\bar{A}_i - \bar{A})V]^T\end{aligned}\quad (3.64)$$

$$\begin{aligned}\tilde{S}_W &= \sum_{i=1}^L \sum_{Y_k \in C_i} (Y_k - \bar{Y}_i)(Y_k - \bar{Y}_i)^T \\ &= \sum_{i=1}^L \sum_{A_k \in C_i} [U(A_k - \bar{A}_i)V][U(A_k - \bar{A}_i)V]^T\end{aligned}\quad (3.65)$$

ถ้า

$$tr(\tilde{S}_B) = tr\left(\sum_{i=1}^L N_i U^T (\bar{A}_i - \bar{A}) V V^T (\bar{A}_i - \bar{A})^T U\right) \quad (3.66)$$

$$tr(\tilde{S}_W) = tr\left(\sum_{i=1}^L \sum_{A_k \in C_i} U^T (A_k - \bar{A}_i) V V^T (A_k - \bar{A}_i)^T U\right) \quad (3.67)$$

การคำนวณหา U และ V ไม่สามารถทำได้พร้อมกัน ดังนั้น จึงทำการคำนวณทีละค่าโดยกำหนดให้ $V = I$ (I คือ เมตริกซ์เอกลักษณ์) เพื่อคำนวณค่าของ U ลำดับต่อมากำหนดให้ $U = I$ เพื่อคำนวณค่าของ V

การคำนวณหา U

กำหนดให้ $V = I$

$$tr(\tilde{G}_B) = U^T G_B^U U, \quad tr(\tilde{G}_W) = U^T G_W^U U \quad (3.68)$$

โดยที่

$$G_B^U = \sum_{i=1}^L N_i (\bar{A}_i - \bar{A})(\bar{A}_i - \bar{A})^T \quad (3.69)$$

$$G_W^U = \sum_{i=1}^L \sum_{A_k \in C_i} N_i (A_k - \bar{A}_i)(A_k - \bar{A}_i)^T \quad (3.70)$$

สามารถแทนค่าสมการที่ 3.63 ใหม่ได้คือ

$$J(U) = \frac{U^T G_B^U U}{U^T G_W^U U} \quad (3.71)$$

การแก้สมการหา U ที่เหมาะสมที่สุดในการฉายสามารถทำได้โดย การหาค่ามากที่สุดของ $J(U)$

$$U_{opt} = \arg \max_x J(U) \quad (3.72)$$

ถ้า G_W^U ไม่เป็นเมตริกซ์เอกฐาน สามารถหาเมตริกซ์ U_{opt} โดยการคำนวณไอเกนเวกเตอร์ และค่าไอเกนจากเมตริกซ์ G_B^U และ G_W^U ได้จากสมการ

$$G_B^U U_{opt} = \lambda G_W^U U_{opt} \quad (3.73)$$

โดยปกติเมตริกซ์การฉายที่เหมาะสมที่สุดที่ใช้ในการฉายมาจากไอเกนเวกเตอร์ของ $G_W^{-1}G_B$ ที่สัมพันธ์กับค่าไอเกนที่มากที่สุดตั้งแต่ 1 ถึง m แต่ในความเป็นจริงเนื่องจากต้องการหาคุณลักษณะของภาพ ดังนั้นจึงเลือกเมตริกซ์ในการฉายเพียง 1 ถึง r เท่านั้น ดังนั้นเมตริกซ์ใหม่ที่ใช้ในการฉาย จึงมีขนาดเป็น $m \times r$

$$U = [u_1, u_2, \dots, u_r] \quad (3.74)$$

การคำนวณหา V

กำหนดให้ $U = I$

$$tr(\tilde{H}_B) = V^T H_B^V V, \quad tr(\tilde{H}_W) = V^T H_W^V V \quad (3.75)$$

โดยที่

$$H_B^V = \sum_{i=1}^L N_i (\bar{A}_i - \bar{A})^T (\bar{A}_i - \bar{A}) \quad (3.76)$$

$$H_W^V = \sum_{i=1}^L \sum_{A_k \in C_i} N_i (A_k - \bar{A}_i)^T (A_k - \bar{A}_i) \quad (3.77)$$

สามารถแทนค่าสมการที่ 3.52 ใหม่ได้คือ

$$J(V) = \frac{V^T H_B^V V}{V^T H_W^V V} \quad (3.78)$$

การแก้สมการหา V ที่เหมาะสมที่สุดในการฉายสามารถทำได้โดย การหาค่ามากที่สุดของ $J(V)$

$$V_{opt} = \arg \max_x J(V) \quad (3.79)$$

ถ้า H_W^V ไม่เป็นเมตริกซ์เอกฐาน สามารถหาเมตริกซ์ V_{opt} โดยการคำนวณไอเกนเวกเตอร์ และค่าไอเกนจากเมตริกซ์ H_B^V และ H_W^V ได้จากสมการ

$$H_B^V V_{opt} = \lambda H_W^V V_{opt} \quad (3.80)$$

โดยปกติเมตริกซ์การฉายที่เหมาะสมที่สุดที่ใช้ในการฉายมาจากไอเกนเวกเตอร์ของ $H_W^{V-1} H_B^V$ ที่สัมพันธ์กับค่าไอเกนที่มากที่สุดตั้งแต่ 1 ถึง n แต่ในความเป็นจริงเนื่องจากต้องการหาคุณลักษณะของภาพ ดังนั้นจึงเลือกเมตริกซ์ในการฉายเพียง 1 ถึง c เท่านั้น ดังนั้นเมตริกซ์ใหม่ที่ใช้ในการฉาย จึงมีขนาดเป็น $n \times c$

$$V = [v_1, v_2, \dots, v_c] \quad (3.81)$$

การหาคุณลักษณะของภาพ

หลังจากภาพฝึกฝนถูกฉายลงบนเมตริกซ์การฉาย U และ V โดย $A_j (j=1, 2, \dots, M)$ คือภาพฝึกฝนแต่ละภาพที่มีขนาดเป็น $m \times n$ และมีค่าเฉลี่ยของภาพฝึกฝนทั้งหมดเป็น \bar{A} และกำหนดให้ $\bar{A}_i (i=1, 2, \dots, L)$ คือ ค่าเฉลี่ยของภาพแต่ละกลุ่ม จะได้เมตริกซ์คุณลักษณะเช่นเดียวกับสมการที่ 3.51

เมตริกซ์ใหม่ที่ได้จากการฉาย Y_1, Y_2, \dots, Y_M มีชื่อว่า เมตริกซ์คุณลักษณะของภาพ A ซึ่งมีขนาด $r \times c$ โดยที่ $r \leq m$ และ $c \leq n$ ซึ่งแทนด้วยเมตริกซ์ Y นั่นเอง

การสร้างภาพใหม่

วิธี (2D)²LDA สามารถนำเอาเมตริกซ์คุณลักษณะและเมตริกซ์การฉายที่เหมาะสมที่สุด U และ V มาสร้างเป็นภาพใบหน้าใหม่ได้ตามขั้นตอนดังนี้

กำหนดให้ A เป็นภาพตัวอย่าง โดยที่ Y เป็นเมตริกซ์คุณลักษณะของภาพ A โดยที่ $U = [u_1, u_2, \dots, u_r]$ และ $V = [v_1, v_2, \dots, v_c]$ เป็นเมตริกซ์การฉายที่เหมาะสมที่สุด ดังนั้นสมการในการสร้างภาพใหม่คือ

$$\tilde{A} = UYV^T \quad (3.82)$$

3.2 วิธีการที่นำเสนอ

ตามแนวทางการวิจัยที่พัฒนาและมีการนำเสนอมาแล้วนั้น สรุปได้ว่า วิธี LDA พัฒนามาจนเป็นวิธีการวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติสองทิศทาง ((2D)²LDA) ที่มีการวิเคราะห์ข้อมูลทั้งในด้านคอลัมน์และแถวของภาพใบหน้า การประยุกต์ตัวถ่วงน้ำหนักก็เริ่มมีการนำมาใช้ในวิธีการเหล่านี้ทั้งในวิธี LDA และ 2DLDA ก็เริ่มมีงานวิจัยดังกล่าวออกมาให้เห็นแล้ว ดังนั้น วิทยานิพนธ์ฉบับนี้จึงขอเสนอแนวทางใหม่ที่น่าตัวถ่วงน้ำหนักมาประยุกต์ใช้ในวิธี (2D)²LDA เพื่อเพิ่มประสิทธิภาพการรู้จำให้ดีขึ้น

3.2.1 ฟังก์ชันการถ่วงน้ำหนัก (Weighting function)

จากสมการที่ 3.51 เมื่อนำมาประยุกต์กับการกระจายแบบคู่ สามารถแสดงสมการฟังก์ชันการถ่วงน้ำหนักในรูปการกระจายแบบคู่ได้ดังนี้

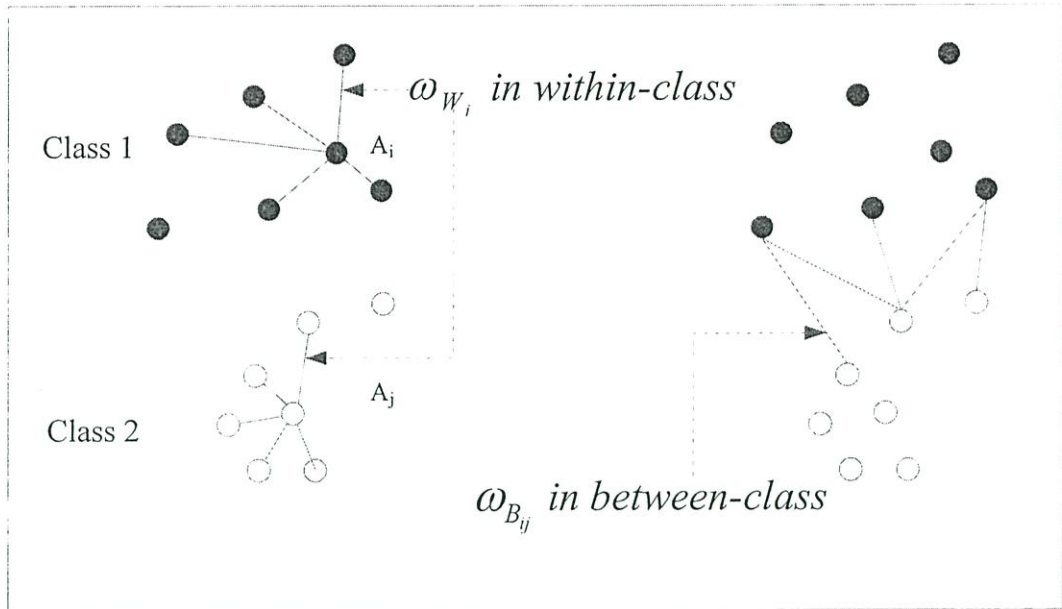
$$\omega_{B_{ij}} = \frac{1}{\|m_i - m_j\|} = \frac{1}{(m_i - m_j)^T (m_i - m_j)} \quad (3.83)$$

ซึ่ง $\omega_{B_{ij}}$ คือ Weighted pairwise scatter เป็นการคำนวณปกติ และกำหนดให้เป็นค่าระยะระหว่างค่าเฉลี่ยข้อมูลของกลุ่ม i และ j หรือมีการแยกข้อมูลระหว่างกลุ่ม i และ j ในมิติเริ่มต้นคืออย่างไร ในการทดสอบเริ่มแรกของงานวิจัยนี้พิจารณาฟังก์ชันการถ่วงน้ำหนักในรูปแบบของระยะยูคลิดีียน ซึ่งสมการของ $\omega_{B_{ij}}$ ก็คือ ค่าระยะความแตกต่างระหว่างค่าเฉลี่ยของภาพในกลุ่ม i และ j นั้นเอง

ส่วนฟังก์ชันการถ่วงน้ำหนักสัมพัทธ์กำหนดให้เป็น ω_{W_i} (Relevance weighted pairwise scatter) เป็นการคำนวณเพื่อแยกกลุ่มข้อมูลออกจากกลุ่มอื่นๆ ปกติจะกำหนดฟังก์ชันการถ่วงน้ำหนักนี้ให้เป็นไปตามสมการที่ 3.84

$$\omega_{W_i} = \sum_{i \neq j} \frac{1}{\omega_{B_{ij}}} \quad (3.84)$$

ω_{W_i} จะถูกนอร์มอลไลซ์เพื่อให้ได้ค่าที่มากที่สุดเป็น 1 แม้ว่าจะมีการนำเสนอตัวระยะต่างๆ ที่ผ่านมามากหลายแบบ ก็ควรเลือกตัวที่ดีที่สุดวิธีใดวิธีหนึ่งที่เป็นของชุดข้อมูล ตัวอย่างไดอะแกรมการถ่วงน้ำหนักที่แสดงความสัมพันธ์ของข้อมูลต่างๆ ในการถ่วงน้ำหนักข้อมูลระหว่างกลุ่ม และข้อมูลภายในกลุ่ม สามารถแสดงได้ดังรูปที่ 3.6



รูปที่ 3.6 การถ่วงน้ำหนักแบบ Weighted และ Relevance Weighted

3.2.2 วิธี Weighted (2D)²LDA

เนื่องจากสมการที่ (3.54) และ (3.56) ไม่สัมพันธ์โดยตรงกับความถูกต้องในการแยกแยะ และสนใจกลุ่มข้อมูลต่างๆ ที่อยู่นอกกลุ่มอาจจะมีผลไม่ดีต่อการประเมินเมตริกซ์ความแปรปรวนร่วมระหว่างกลุ่ม G_B และ H_B ทั้งหมด จึงมีการเพิ่มตัวถ่วงน้ำหนักเข้าไปที่เมตริกซ์ความแปรปรวนร่วมระหว่างกลุ่มในการคำนวณข้อมูลด้านแถวและคอลัมน์ทั้งสองด้านด้วยสมการที่ (3.85) และ (3.86) ดังต่อไปนี้

$$G_B = \frac{1}{2N} \sum_{i=1}^{L-1} \sum_{j=i+1}^L \omega_{B_{ij}} N_i N_j (\bar{A}_i - \bar{A}_j)^T (\bar{A}_i - \bar{A}_j) \quad (3.85)$$

$$H_B = \frac{1}{2N} \sum_{i=1}^{L-1} \sum_{j=i+1}^L \omega_{B_{ij}} N_i N_j (\bar{A}_i - \bar{A}_j) (\bar{A}_i - \bar{A}_j)^T \quad (3.86)$$

ซึ่ง การคำนวณฟังก์ชันการถ่วงน้ำหนัก $\omega_{B_{ij}}$ สอดคล้องเป็นไปตามสมการที่ 3.83

3.2.3 วิธี Relevance Weighted (2D)²LDA

จากวิธี Weighted (2D)²LDA หรือ W(2D)²LDA พบว่ามีการเพิ่มตัวถ่วงน้ำหนักเพื่อกำหนดให้เห็นความแตกต่างในการพิจารณาข้อมูลกลุ่มต่างๆ เมื่อประเมินเมตริกซ์ความแปรปรวนร่วมระหว่างกลุ่ม รูปแบบตัวถ่วงน้ำหนักนี้ก็สามารถนำมาใช้เมื่อมีการประเมินเมตริกซ์ความแปรปรวนร่วมภายในกลุ่มได้เช่นเดียวกัน เพื่อลดผลของข้อมูลจากกลุ่มภายนอก เรียกวิธีนี้ว่า RW(2D)²LDA ดังนั้น สามารถปรับปรุงสมการที่ (3.55) และ (3.57) ใหม่ได้ดังนี้

$$G_W = \frac{1}{N} \sum_{i=1}^L \sum_{A_k \in C_i} \omega_{W_i} (A_k - \bar{A}_i)^T (A_k - \bar{A}_i) \quad (3.87)$$

$$H_W = \frac{1}{N} \sum_{i=1}^L \sum_{A_k \in C_i} \omega_{W_i} (A_k - \bar{A}_i)(A_k - \bar{A}_i)^T \quad (3.88)$$

การรวมเอา ω_{W_i} เข้าไว้ในสมการที่ 3.55 และ 3.57 นั้น เพื่อถ่วงน้ำหนักข้อมูลภายในกลุ่มเดียวกันให้กระชับขึ้น ทำให้กลุ่มข้อมูลที่ i ที่เป็นกลุ่มภายนอกมีผลต่อ G_W และ H_W เพียงเล็กน้อยเท่านั้น เนื่องมาจาก ถ้าข้อมูลกลุ่มหนึ่งถูกแยกออกจากกลุ่มอื่นๆ ในชุดข้อมูลแล้ว เมตริกซ์ความแปรปรวนร่วมภายในกลุ่มสำหรับปริภูมิอันใหม่จะไม่มีผลต่อการตัดแยก ความถูกต้องในการตัดแยกก็จะถูกต้องมากขึ้น

3.3 หลักการคำนวณของวิธีที่นำเสนอ

จากการฉายภาพ A ลงบนเมตริกซ์ U และ V ตามหลักการแปลงเชิงเส้น ตามสมการ 3.62, 3.63 และ 3.64 ดังนั้น การคำนวณค่า \tilde{S}_B และ \tilde{S}_W แสดงได้ดังนี้

$$\begin{aligned} \tilde{S}_B &= \frac{1}{2N} \sum_{k,l=1}^L \omega_{B_{ij}} N_k N_l (\bar{Y}_k - \bar{Y}_l)(\bar{Y}_k - \bar{Y}_l)^T \\ &= \frac{1}{2N} \sum_{k,l=1}^L \omega_{B_{ij}} N_k N_l [U^T (\bar{A}_k - \bar{A}_l) V] [U^T (\bar{A}_k - \bar{A}_l) V]^T \end{aligned} \quad (3.89)$$

$$\begin{aligned} \tilde{S}_W &= \frac{1}{N} \sum_{i=1}^L \sum_{Y_k \in C_i} \omega_{W_i} (Y_k - \bar{Y}_i)(Y_k - \bar{Y}_i)^T \\ &= \frac{1}{N} \sum_{i=1}^L \sum_{A_k \in C_i} \omega_{W_i} [U^T (A_k - \bar{A}_i) V] [U^T (A_k - \bar{A}_i) V]^T \end{aligned} \quad (3.90)$$

ถ้า

$$tr(\tilde{S}_B) = tr \left(\frac{1}{2N} \sum_{k,l=1}^L \omega_{B_{ij}} N_k N_l U^T (\bar{A}_k - \bar{A}_l) V V^T (\bar{A}_k - \bar{A}_l)^T U \right) \quad (3.91)$$

$$tr(\tilde{S}_W) = tr \left(\frac{1}{N} \sum_{i=1}^L \sum_{A_k \in C_i} \omega_{W_i} U^T (A_k - \bar{A}_i) V V^T (A_k - \bar{A}_i)^T U \right) \quad (3.92)$$

ซึ่ง $\omega_{B_{ij}}$ คือ Weighted pairwise scatter และ ω_{W_i} คือ Relevance weighted pairwise scatter

การคำนวณหา U และ V ไม่สามารถทำได้พร้อมกัน ดังนั้น จึงทำการคำนวณทีละค่าโดยกำหนดให้ $V = I$ (I คือ เมตริกซ์เอกลักษณ์) เพื่อคำนวณค่าของ U ลำดับต่อมากำหนดให้ $U = I$ เพื่อคำนวณค่าของ V

การคำนวณหา U

กำหนดให้ $V = I$

$$\text{tr}(\tilde{G}_B) = U^T G_B^U U, \quad \text{tr}(\tilde{G}_W) = U^T G_W^U U \quad (3.93)$$

โดยที่

$$G_B^U = \frac{1}{2N} \sum_{k,l=1}^L \omega_{B_{kl}} N_k N_l (\bar{A}_k - \bar{A}_l)(\bar{A}_k - \bar{A}_l)^T \quad (3.94)$$

$$G_W^U = \frac{1}{N} \sum_{i=1}^L \sum_{A_k \in C_i} \omega_{W_i} (\bar{A}_k - \bar{A}_i)(\bar{A}_k - \bar{A}_i)^T \quad (3.95)$$

สามารถแทนค่าสมการที่ 3.52 ใหม่ได้คือ

$$J(U) = \frac{U^T G_B^U U}{U^T G_W^U U} \quad (3.96)$$

การแก้สมการหา U ที่เหมาะสมที่สุดในการฉายสามารถทำได้โดย การหาค่ามากที่สุดของ $J(U)$

$$U_{opt} = \arg \max_x J(U) \quad (3.97)$$

ถ้า G_W^U ไม่เป็นเมตริกซ์เอกฐาน สามารถหาเมตริกซ์ U_{opt} โดยการคำนวณไอเกนเวกเตอร์ และค่าไอเกนจากเมตริกซ์ G_B และ G_W ได้จากสมการ

$$G_B^U U_{opt} = \lambda G_W^U U_{opt} \quad (3.98)$$

โดยปกติเมตริกซ์การฉายที่เหมาะสมที่สุดที่ใช้ในการฉายมาจากไอเกนเวกเตอร์ของ $G_W^{U-1} G_B^U$ ที่สัมพันธ์กับค่าไอเกนที่มากที่สุดตั้งแต่ 1 ถึง m แต่ในความเป็นจริงเนื่องจากต้องการหาคุณลักษณะของภาพ ดังนั้นจึงเลือกเมตริกซ์ในการฉายเพียง 1 ถึง r เท่านั้น ดังนั้นเมตริกซ์ใหม่ที่ใช้ในการฉายจึงมีขนาดเป็น $m \times r$

$$U = [u_1, u_2, \dots, u_r] \quad (3.99)$$

การคำนวณหา V

กำหนดให้ $U = I$

$$\text{tr}(\tilde{H}_B) = V^T H_B^V V, \quad \text{tr}(\tilde{H}_W) = V^T H_W^V V \quad (3.100)$$

โดยที่

$$H_B^V = \frac{1}{2N} \sum_{k,l=1}^L \omega_{B_{kl}} N_k N_l (\bar{A}_k - \bar{A}_l)^T (\bar{A}_k - \bar{A}_l) \quad (3.101)$$

$$H_W^V = \frac{1}{N} \sum_{i=1}^L \sum_{A_k \in C_i} \omega_{W_i} (\bar{A}_k - \bar{A}_i)^T (\bar{A}_k - \bar{A}_i) \quad (3.102)$$

สามารถแทนค่าสมการที่ 3.52 ใหม่ได้คือ

$$J(V) = \frac{V^T H_B^V V}{V^T H_W^V V} \quad (3.103)$$

การแก้สมการหา V ที่เหมาะสมที่สุดในการฉายสามารถทำได้โดยการหาค่ามากที่สุดของ $J(V)$

$$V_{opt} = \arg \max_x J(V) \quad (3.104)$$

ถ้า H_W^V ไม่เป็นเมตริกซ์เอกฐาน สามารถหาเมตริกซ์ V_{opt} โดยการคำนวณไอเกนเวกเตอร์และค่าไอเกนจากเมตริกซ์ H_B และ H_W ได้จากสมการ

$$H_B^V V_{opt} = \lambda H_W^V V_{opt} \quad (3.105)$$

โดยปกติเมตริกซ์การฉายที่เหมาะสมที่สุดที่ใช้ในการฉายมาจากไอเกนเวกเตอร์ของ $H_W^{V^{-1}} H_B^V$ ที่สัมพันธ์กับค่าไอเกนที่มากที่สุดตั้งแต่ 1 ถึง n แต่ในความเป็นจริงเนื่องจากการหาคุณลักษณะของภาพ ดังนั้นจึงเลือกเมตริกซ์ในการฉายเพียง 1 ถึง c เท่านั้น ดังนั้นเมตริกซ์ใหม่ที่ใช้ในการฉายจึงมีขนาดเป็น $n \times c$

$$V = [v_1, v_2, \dots, v_c] \quad (3.106)$$

การหาคุณลักษณะของภาพ

หลังจากภาพฝึกฝนถูกฉายลงบนเมตริกซ์ U และ V โดย $A_j (j = 1, 2, \dots, M)$ คือ ภาพแต่ละภาพฝึกฝนที่มีขนาดเป็น $m \times n$ และมีค่าเฉลี่ยของภาพฝึกฝนทั้งหมดเป็น \bar{A} และกำหนดให้ $\bar{A}_i (i = 1, 2, \dots, L)$ คือค่าเฉลี่ยของภาพแต่ละกลุ่ม จะได้เมตริกซ์คุณลักษณะเช่นเดียวกับสมการที่ 3.51

เมตริกซ์ใหม่ที่ได้จากการฉาย Y_1, Y_2, \dots, Y_M เรียกว่า ฟิเจอร์ของภาพ A ที่มีขนาด $r \times c$ โดยที่ $r \leq m$ และ $c \leq n$ เมตริกซ์ Y มีชื่อว่า เมตริกซ์คุณลักษณะของภาพ A

การสร้างภาพใหม่

วิธี (2D)²LDA สามารถนำเอาเมตริกซ์คุณลักษณะและเมตริกซ์การฉายที่เหมาะสมที่สุด U และ V มาสร้างเป็นภาพใบหน้าใหม่ได้ตามขั้นตอนดังนี้

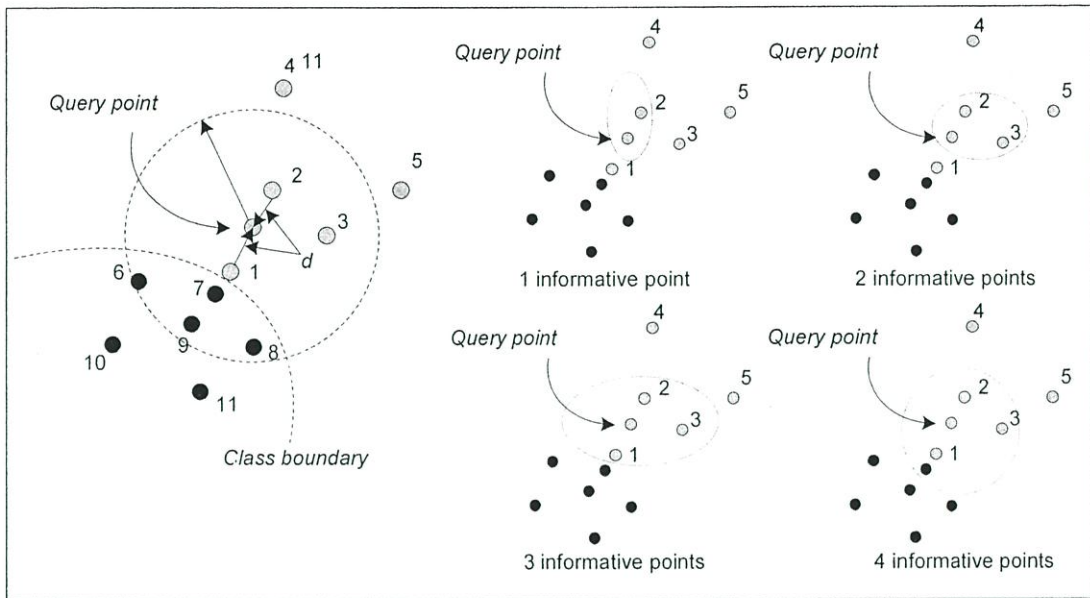
กำหนดให้ A เป็นภาพ โดยที่ Y เป็นเมตริกซ์คุณลักษณะของภาพ A โดยที่ $U = [u_1, u_2, \dots, u_r]$ และ $V = [v_1, v_2, \dots, v_c]$ เป็นเมตริกซ์การฉายที่เหมาะสมที่สุด ดังนั้นสมการในการสร้างภาพใหม่คือ

$$\tilde{A} = UYV^T \quad (3.107)$$

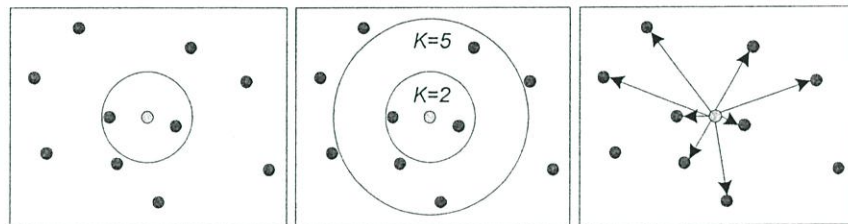
3.4 การจำแนกข้อมูลด้วยวิธี k -NN (k -Nearest Neighbor Classification)

เมื่อคำนวณหาเมตริกซ์คุณลักษณะของภาพด้วยวิธี (2D)²LDA แล้ว จะใช้วิธีการจำแนกข้อมูลด้วยวิธี 1-Nearest-Neighbor (1NN) ซึ่งเป็นการคำนวณหาระยะยูคลิดีเดียนเพื่อใช้กับเมตริกซ์ซึ่งจะกล่าวรายละเอียดต่อไป

วิธีการจำแนกโดยใช้ข้อมูลข้างเคียงใกล้สุด k ตัว เป็นอัลกอริทึมของการเรียนรู้แบบมีผู้สอน (Supervised learning) ซึ่งผลของการจัดกลุ่มของสมาชิกตัวใหม่จะขึ้นกับกลุ่มของสมาชิกตัวที่มีความใกล้เคียงที่สุด k ตัว เป้าหมายของอัลกอริทึมนี้คือ การจัดกลุ่มสมาชิกตัวใหม่โดยขึ้นกับแอมพริบิวต์ และข้อมูลตัวอย่างฝึกฝน ตัวจำแนกประเภทชนิดนี้จะเป็น โมเดลที่ไม่ใช้ความจำ เมื่อกำหนดจุดๆ หนึ่งขึ้นมา เราก็จะหาข้อมูลใกล้เคียงจำนวน k ตัว ที่ใกล้กับจุดที่กำหนดนั้น



(ก)



(ข)

รูปที่ 3.7 การทำงานของวิธีการจำแนกด้วย k -NN (ก) การเลือกพารามิเตอร์ k และ (ข) การคำนวณค่าระยะทาง

การทำงานของวิธีการจำแนกแสดงในรูปที่ 3.7 โดยใช้ข้อมูลข้างเคียงใกล้ที่สุด k ตัว เป็นอัลกอริทึมที่เรียบง่าย ทำงานโดยขึ้นกับระยะทางที่น้อยที่สุดจากสมาชิกใหม่ หรือข้อมูลที่ป้อนถาม (Input query instance) กับข้อมูลตัวอย่างฝึกฝน เพื่อจะคำนวณหาเพื่อนบ้านที่ใกล้ที่สุด k ตัว หลังจากนั้นจะรวบรวมสมาชิกที่ใกล้เคียงที่สุด k ตัวแล้วเลือกคลาสที่สมาชิกส่วนใหญ่อยู่ในกลุ่ม k ดังกล่าว สังกัดอยู่มากที่สุดให้กับสมาชิกใหม่

ขั้นตอนการทำงานของอัลกอริทึมการจำแนกโดยใช้ข้อมูลข้างเคียงใกล้ที่สุด k ตัว (k - Nearest Neighbors) เป็นวิธีการแยกประเภทข้อมูลโดยใช้หลักการที่พัฒนามาจากวิธี 1-NN โดยใช้ค่า k มาพิจารณา เป็นวิธีที่แยกประเภทของข้อมูลโดยพิจารณาข้อมูลที่มีความใกล้เคียงเป็นส่วนใหญ่ของกลุ่มที่อยู่ภายในหรือเท่ากับค่า k แล้วจัดข้อมูลนั้นเข้ากลุ่ม

วิธีการคำนวณ k -NN

กำหนดให้เซตของข้อมูลฝึกสอนที่ประกอบด้วยแอมพริบิวต์หลายตัวแปร $X_i | i = 1, 2, \dots, n$ ซึ่งจะนำมาใช้ในการแบ่งกลุ่มต่างๆ Y_i เนื่องจากเราให้ X_i เป็นตัวบอกปริมาณ จุดประสงค์เพื่อต้องการจัดตัวอย่างทดสอบ Y_i เข้ากับกลุ่มใดกลุ่มหนึ่ง สามารถคำนวณได้ตามขั้นตอนดังต่อไปนี้

- เลือก พารามิเตอร์ k ที่แสดงถึงจำนวนข้อมูลใกล้เคียงที่จะใช้ในการอ้างอิง
 - คำนวณค่าระยะทางระหว่างข้อมูลทดสอบ Y_i กับสมาชิกในข้อมูลตัวอย่างฝึกฝนทุกตัว ด้วยฟังก์ชันการหาระยะทาง $d_i = d(Y_i, X_i)$
 - ขั้นตอนต่อจากนั้นคือการนำระยะทางที่หาได้จากสมาชิกในข้อมูลตัวอย่างฝึกฝน มาเรียงลำดับจากน้อยไปหามากแล้วเลือกสมาชิกที่มีระยะทางใกล้เคียงที่สุดออกมา k ตัว แล้วเลือกคลาสตามคลาสส่วนใหญ่ของเพื่อนบ้านใกล้เคียงที่เลือกออกมา
 - รวบรวมกลุ่มข้อมูล หรือ Y จากข้อมูลใกล้เคียง k ตัวดังกล่าว
 - หากกลุ่ม (Class) จากค่าความเป็นส่วนใหญ่มหาของข้อมูลตัวอย่างกลุ่มใดๆ ที่อยู่ภายในค่าของ k ที่เกิดขึ้นในกลุ่มของข้อมูลใกล้เคียงที่ได้ในการตัดสินใจเพื่อทำนายคลาสให้กับข้อมูลที่สอบถาม (Query-instance) โดยคำนวณหาข้อมูล X_j ที่ใกล้กับ Y_i ที่สุด โดยการหา j จาก $d_j < d_i | i, j = 1, 2, \dots, n$
 - กำหนด d_j ไปเป็นกลุ่มของข้อมูลทดสอบ $C(Y) = C_j$
- ถ้าหากมีข้อมูลระยะทางใกล้เคียงข้อมูลที่เราสนใจหลายตัวที่มีกลุ่มตัวอย่างแตกต่างกัน ให้พิจารณาเลือกกลุ่มที่มีข้อมูลเป็นส่วนใหญ่มหาเป็นกลุ่มของข้อมูลที่เราสนใจนั้น
- วิธีการจำแนกข้อมูลด้วยวิธี k -NN สามารถเลือกใช้ได้หลายวิธี เช่น Euclidean distance, Mahalanobis distance, City Block metric, Cosine distance, Correlation distance, Hamming distance, Chebychev distance เป็นต้น โดยขอเสนอรายละเอียดวิธีที่นิยมใช้กันแพร่หลายไว้ 2 วิธี ดังนี้

3.4.1 ระยะยูคลิดีเนียน (Euclidean distance)

กำหนดให้ภาพใบหน้า A_1 และ A_2 ถูกนำมาหาเมตริกซ์คุณลักษณะเป็น Y_1 และ Y_2 ถ้าต้องการหาความเหมือนกันของภาพที่ถูกฉายทั้งสอง ทำได้โดยการหาระยะทาง

$$d(Y_1, Y_2) = \|Y_1 - Y_2\| = \sqrt{\sum_{k=1}^r \sum_{l=1}^c (y_1^{(k,l)} - y_2^{(k,l)})^2} \quad (3.108)$$

เมื่อ $\|Y_1 - Y_2\|$ คือ ระยะยูคลิดีเนียนระหว่างเมตริกซ์คุณลักษณะสองเมตริกซ์

ถ้าเมตริกซ์คุณลักษณะของภาพที่ใช้ฝึกสอนทั้งหมด คือ Y_1, Y_2, \dots, Y_M (M คือ จำนวนภาพที่ใช้ฝึกสอน) เมื่อแต่ละภาพเป็นส่วนหนึ่งของกลุ่ม C_i ในการทดสอบ จะนำภาพตัวอย่างหาระยะยูคลิดีเนียนเทียบกับภาพทุกภาพที่ใช้ในการฝึกสอน จากนั้นจะพิจารณาภาพตัวอย่างที่เทียบกับภาพฝึกสอนแล้วได้ระยะยูคลิดีเนียนที่สั้นที่สุดว่าอยู่ในกลุ่มใดของภาพฝึกสอน จากนั้นก็กำหนดให้คุณลักษณะของภาพตัวอย่าง Y กลุ่มนั้นด้วย

3.4.2 ระยะมาฮาลานอบิส (Mahalanobis distance)

เช่นเดียวกัน การกำหนดให้ภาพใบหน้า A_1 และ A_2 ถูกนำมาหาเมตริกซ์คุณลักษณะเป็น Y_1 และ Y_2 ถ้าต้องการหาความเหมือนกันของภาพที่ถูกฉายทั้งสองโดยการหาระยะทางมาฮาลานอบิสสามารถทำได้ดังสมการต่อไปนี้

$$d(Y_1, Y_2) = \|Y_1 - Y_2\| = \sqrt{\sum_{k=1}^r \sum_{l=1}^c (y_1^{(k,l)} - y_2^{(k,l)})^T S^{-1} (y_1^{(k,l)} - y_2^{(k,l)})} \quad (3.109)$$

เมื่อ $\|Y_1 - Y_2\|$ คือ ระยะมาฮาลาโนบิสระหว่างเมตริกซ์คุณลักษณะสองเมตริกซ์ และ S คือ เมตริกซ์ความแปรปรวนร่วม ซึ่งคำนวณได้จากความสัมพันธ์ตามสมการต่อไปนี้

$$S = \sqrt{\sum_{k=1}^r \sum_{l=1}^c (y_1^{(k,l)} - y_2^{(k,l)})^T S^{(k,l)-1} (y_1^{(k,l)} - y_2^{(k,l)})} \quad (3.110)$$

บทที่ 4

การทดสอบวิธีการวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติ

เนื้อหาภายในบทนี้กล่าวถึง การทดสอบระบบการรู้จำใบหน้าตามวิธีใหม่ที่น่าเสนอภายใต้เงื่อนไขพารามิเตอร์ต่างๆ ที่มีผลต่ออัตราความถูกต้องในการรู้จำ เป็นการนำเสนอผลที่ได้จากการทดสอบระบบรวมทั้งการสรุปและวิจารณ์ผลการทดสอบแต่ละแบบ โดยแสดงให้เห็นการทดสอบเปรียบเทียบสมรรถนะความถูกต้องในการรู้จำระหว่างวิธี (2D)²LDA แบบเดิม กับวิธีใหม่ที่น่าเสนอคือ W(2D)²LDA และ RW(2D)²LDA ภายใต้เงื่อนไขต่าง ๆ ที่มีผลต่อการรู้จำใบหน้า ซึ่งทำการทดสอบกับฐานข้อมูลภาพใบหน้าต่างๆ ได้แก่ ฐานข้อมูล FERET [55], PIE [58], ORL [61], AR [60] และฐานข้อมูล Yale [59] โดยที่ฐานข้อมูลแต่ละชุดใช้ทดสอบความถูกต้องในการรู้จำใบหน้าภายใต้เงื่อนไขต่างๆ ดังต่อไปนี้

หัวข้อแรกเป็นการทดสอบสมรรถนะความถูกต้องการรู้จำใบหน้าของวิธีใหม่ที่น่าเสนอโดยใช้ฐานข้อมูล FERET

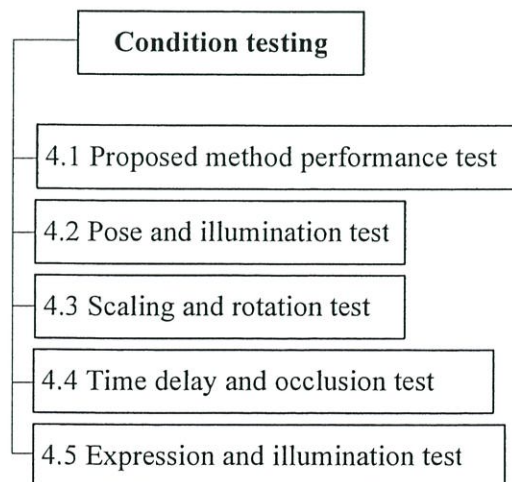
หัวข้อที่สองเป็นการทดสอบภายใต้เงื่อนไขของมุม และความสว่างภาพใบหน้า (Pose and illumination condition test) โดยใช้ฐานข้อมูล PIE

หัวข้อที่สามเป็นการทดสอบภายใต้เงื่อนไขการเปลี่ยนขนาดและการเอียงใบหน้า (Scaling and rotation condition test) โดยใช้ฐานข้อมูล ORL

หัวข้อที่สี่เป็นการทดสอบภายใต้เงื่อนไขของระยะเวลาในการถ่ายภาพ และการบังใบหน้า (Time delay and occlusion condition test) โดยใช้ฐานข้อมูล AR

หัวข้อสุดท้ายเป็นการทดสอบภายใต้เงื่อนไขการแสดงอารมณ์ทางใบหน้า และความสว่างบนใบหน้า (Expression and illumination condition test) โดยใช้ฐานข้อมูล Yale

ซึ่งแผนผังการทดสอบตามเงื่อนไขต่างๆ แสดงไว้ดังรูปที่ 4.1

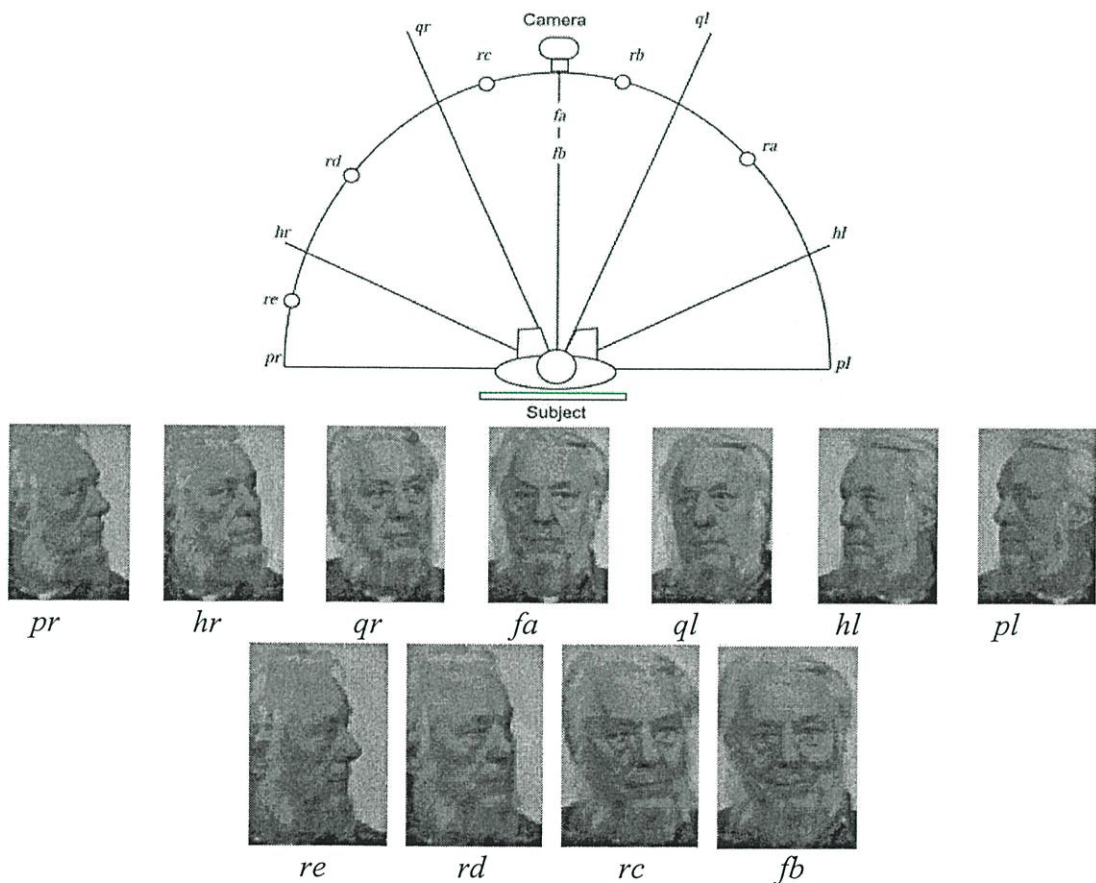


รูปที่ 4.1 แผนผังการทดสอบภายใต้เงื่อนไขต่างๆ

ดังนั้น จึงขอนำเสนอผลการทดสอบการรู้จำใบหน้ากับฐานข้อมูลแต่ละชุดภายใต้เงื่อนไขต่างๆ ตามลำดับดังรายละเอียดในหัวข้อต่อไปนี้

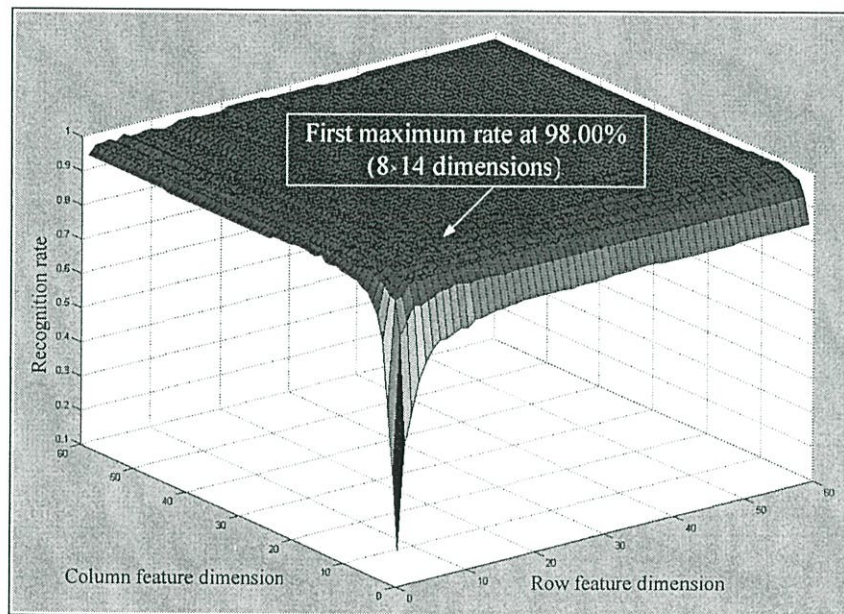
4.1 การทดสอบสมรรถนะของวิธีที่นำเสนอ

อันดับแรกเป็นการทดสอบการรู้จำใบหน้าด้วยวิธีใหม่ที่นำเสนอคือ RW(2D)¹LDA โดยเลือกใช้ฐานข้อมูล FERET สำหรับทำการทดสอบครั้งนี้ เพราะฐานข้อมูล FERET ถูกออกแบบมาเพื่อทดสอบความก้าวหน้าในการรู้จำใบหน้า มีภาพที่ถูกรวบรวมไว้เพื่อสนับสนุนทั้งการพัฒนาอัลกอริทึมและการทดสอบประเมิน ฐานข้อมูลแบ่งเป็นชุดพัฒนาสำหรับนักวิจัย และชุดที่แยกภาพสำหรับการทดสอบ เนื่องจากภาพแต่ละบุคคลในฐานข้อมูล FERET หนึ่งชุดประกอบด้วยภาพจำนวน 5 ถึง 11 ภาพ โดยสองภาพแรกเป็นวิถีด้านหน้ากำหนดเป็น *fa* และ *fb* โดยที่ *fa* เป็นภาพด้านหน้าปกติ ส่วน *fb* เป็นภาพแสดงอารมณ์ทางใบหน้า ภาพที่สาม (*fc*) มีมุมกล้องและแสงสว่างต่างออกไป ที่เหลือเป็นภาพด้านข้างซ้ายและขวาแตกต่างกัน ตัวอย่างตำแหน่งการติดตั้งกล้องเพื่อเก็บภาพ และภาพที่มุมต่างๆ ของฐานข้อมูล FERET แสดงในรูปที่ 4.2 ซึ่งนำมาใช้ทดสอบประสิทธิภาพทั้งสามวิธีในภาพรวม

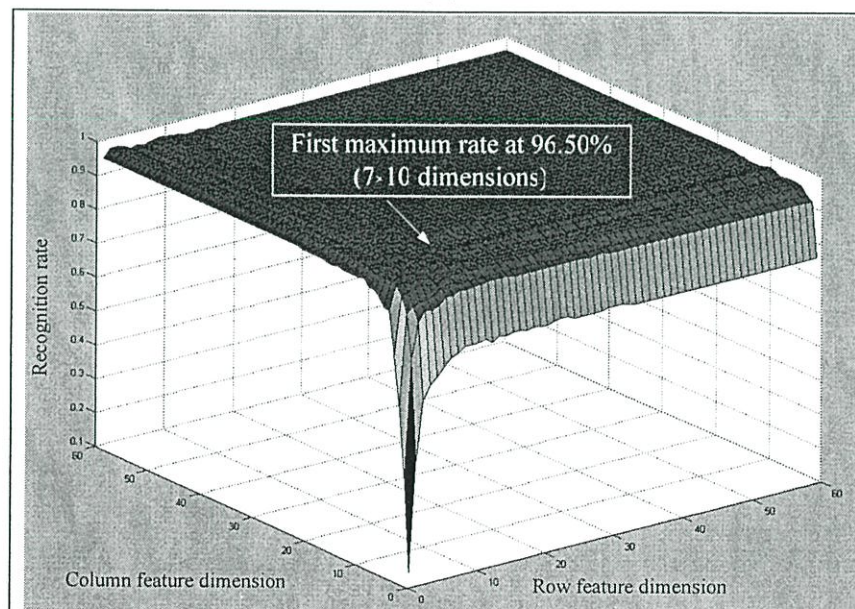


รูปที่ 4.2 ภาพใบหน้าที่ใช้ทดสอบจากฐานข้อมูล FERET

การทดสอบนี้เพื่อหาผลการเปรียบเทียบอัตราความถูกต้องในการรู้จำของวิธี $(2D)^2LDA$ แบบเดิม กับวิธีใหม่คือ $W(2D)^2LDA$ และ $RW(2D)^2LDA$ ภายใต้เงื่อนไขของจำนวนตัวอย่างภาพใบหน้าที่ใช้ในการฝึกฝนจากฐานข้อมูล FERET ภาพใบหน้าที่ใช้ในการทดสอบจากฐานข้อมูลที่มีอยู่ทั้งหมด 14,051 ภาพ (จำนวน 1,199 คน) การฝึกฝนระบบใช้ภาพใบหน้าที่คัดเลือกมาจำนวน 300 คน คนละ 5 ภาพ รวมภาพที่ใช้ฝึกฝน 1,500 ภาพ ส่วนการทดสอบระบบใช้ภาพใบหน้าจาก 500 คน คนละ 5 ภาพ รวมภาพที่ใช้ทดสอบ 2,500 ภาพ ทุกภาพมีขนาด 256×384 พิกเซล เป็นภาพครึ่งตัว ดังนั้น ทุกภาพถูกนอร์มอลไลซ์ให้มีขนาด 85×128 พิกเซล และตัดเฉพาะบริเวณใบหน้า มาทำการหาความถูกต้องในการรู้จำภาพใบหน้าที่ตั้งแต่ฟีเจอร์ที่ 1×1, 1×2, 1×3 ไปจนถึงฟีเจอร์ที่ 60×60

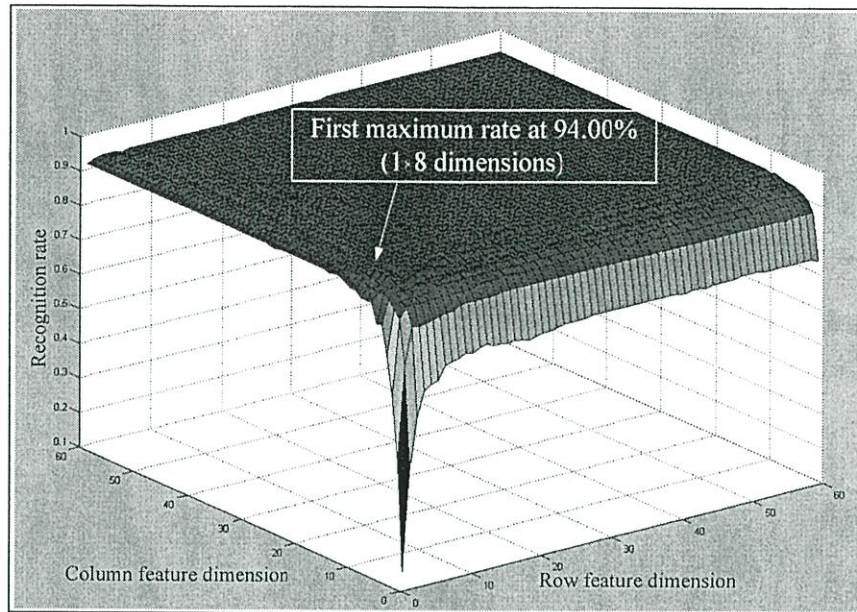


(ก)



(ข)

รูปที่ 4.3 อัตราการรู้จำใบหน้าของวิธีใหม่ที่น่าเสนอ (ก) $RW(2D)^2LDA$ และ (ข) $W(2D)^2LDA$ เปรียบเทียบกับวิธีแบบเดิม (ค) $(2D)^2LDA$



(ค)

รูปที่ 4.3 (ต่อ)

ผลการทดสอบความถูกต้องในการรู้จำที่นำเสนอในรูปกราฟ 3 มิติ แสดงในรูปที่ 4.3 ซึ่งมีแกนแนวนอน 2 แกน แสดงมิติคุณลักษณะด้านแถวและคอลัมน์ เริ่มจาก 0 ถึง 60 มิติ ส่วนแกนแนวตั้งเป็นแกนอัตราความถูกต้องในการรู้จำ กราฟทั้งสามรูป (รูปที่ 4.3 ก. ข. และ ค.) แสดงผลความถูกต้องในการรู้จำใบหน้าที่เป็นไปในลักษณะเดียวกัน ที่มีมิติคุณลักษณะ 3-5 มิติ ได้ความถูกต้องในการรู้จำมากกว่า 70% ขึ้นไป และเมื่อทดสอบไปถึง 6-10 มิติ อัตราการรู้จำที่ได้จะมากกว่า 97% หลังจากนั้นที่มีมิติคุณลักษณะมากกว่า 10 มิติ ขึ้นไป ความถูกต้องในการรู้จำไม่เปลี่ยนแปลงมากนัก ก่อนข้างเสถียร

จากกราฟแสดงผลการทดสอบจะเห็นว่า วิธีใหม่ที่นำเสนอคือ $RW(2D)^2LDA$ ได้ผลอัตราความถูกต้องในการรู้จำใบหน้าสูงสุด 98.00% ที่มีมิติคุณลักษณะ 8×14 ส่วนวิธี $W(2D)^2LDA$ มีอัตราการรู้จำสูงสุดที่ 96.50% ที่มีมิติคุณลักษณะ 7×10 และวิธี $(2D)^2LDA$ แบบเดิมมีอัตราการรู้จำสูงสุดที่ 94.00% ที่มีมิติคุณลักษณะ 1×8

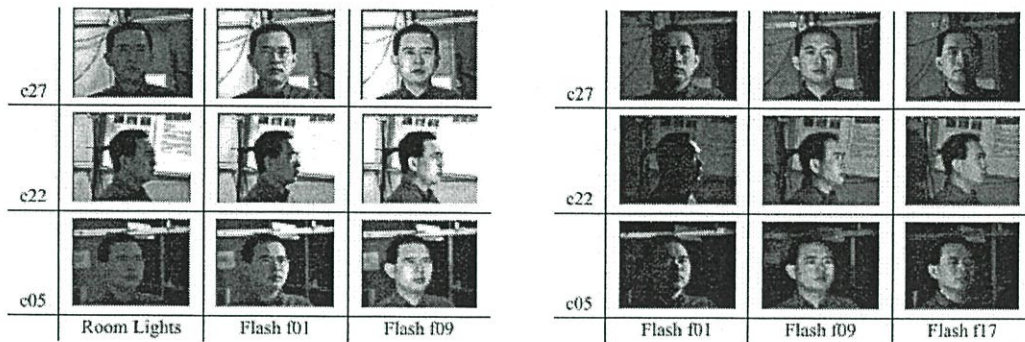
4.2 การทดสอบภายใต้เงื่อนไขมุมและความสว่างภาพใบหน้า (Pose and illumination)

การทดสอบนี้ใช้ฐานข้อมูลใบหน้าของ PIE สำหรับทดสอบเงื่อนไขมุมของภาพ และความสว่างภาพใบหน้า โดยใช้ภาพใบหน้าบุคคลจำนวน 68 คน แต่ละคนประกอบด้วยภาพที่มุมแตกต่างกัน 13 ภาพ และเงื่อนไขความสว่างต่างๆ 43 เงื่อนไข รวมทั้งหมดจำนวน 41,368 ภาพ ตัวอย่างภาพภายใต้เงื่อนไขมุมของภาพ และความสว่างภาพใบหน้าจากฐานข้อมูล PIE แสดงในรูปที่ 4.4 โดยรูป 4.4 ก. เป็นตัวอย่างภาพที่มีมุมแตกต่างกันทั้ง 13 มุม รูป 4.4 ข. เป็นภาพตามเงื่อนไขมุมของภาพและความสว่าง

ต่างๆ ภายใต้สภาวะแวดล้อมต่างกันคือ ด้านซ้ายมือเป็นสภาวะการเปิดไฟภายในห้อง ส่วนด้านขวามือเป็นสภาวะการปิดไฟภายในห้อง มีเพียงความสว่างจากแสงแฟลชของกล้องถ่ายรูปเท่านั้น



(ก)

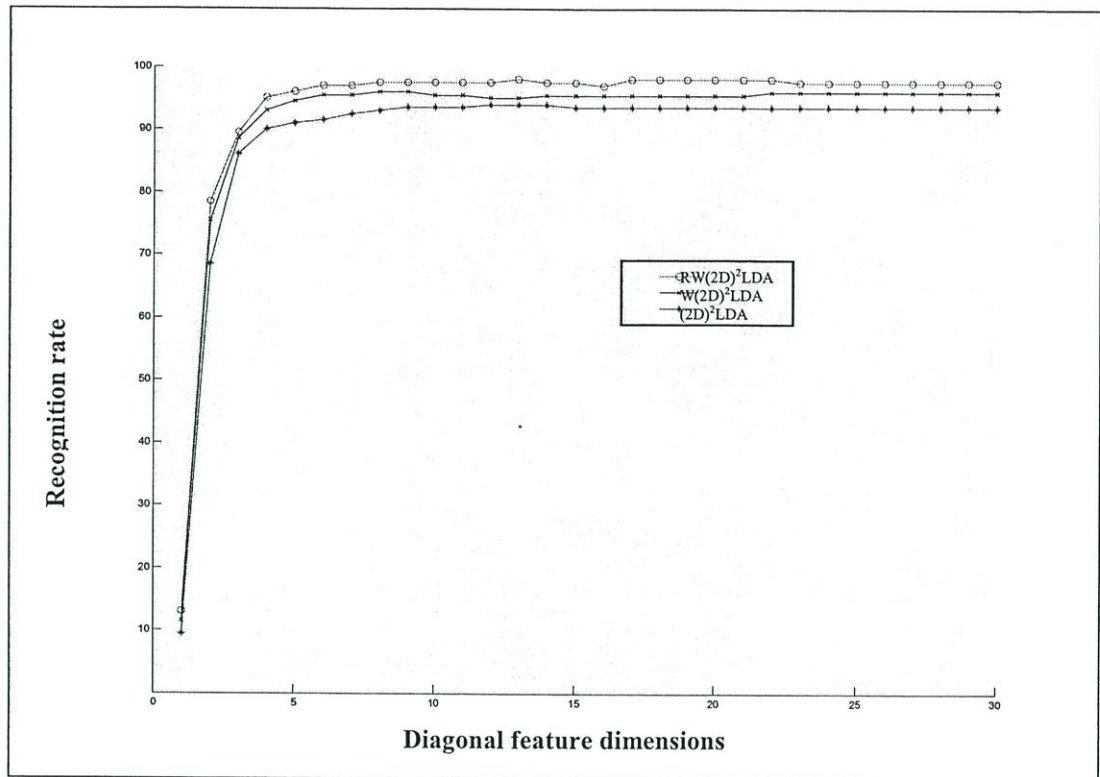


(ข)

รูปที่ 4.4 ภาพใบหน้าจากฐานข้อมูล PIE ภายใต้เงื่อนไขมุม (ก) และความสว่าง (ข)

การเปรียบเทียบสมรรถนะอัตราความถูกต้องในการรู้จำของทั้ง 3 วิธี ได้แก่ $RW(2D)^2LDA$, $W(2D)^2LDA$ และ $(2D)^2LDA$ ด้วยกราฟ 3 มิตินั้นทำได้ยากเนื่องจากเส้นกราฟ 3 มิติของแต่ละวิธีจะทับซ้อนกันเป็นส่วนใหญ่ จนไม่สามารถแยกแยะเส้นกราฟให้เห็นความแตกต่างของอัตราความถูกต้องในการรู้จำของแต่ละวิธีได้อย่างชัดเจน เพื่อให้ง่ายต่อการทำความเข้าใจ จึงเลือกใช้กราฟ 2 มิติ แสดงการเปรียบเทียบสมรรถนะดังกล่าว โดยเลือกข้อมูลอัตราความถูกต้องในการรู้จำในตามแนวแรงแมทริก (Diagonal) ที่มีค่ามีคุณลักษณะเท่ากันทั้งสองแกนมาใช้ในการเปรียบเทียบสมรรถนะ

ผลการเปรียบเทียบอัตราความถูกต้องในการรู้จำใบหน้า ภายใต้เงื่อนไขมุมของภาพ และความสว่างใบหน้าของตัวอย่างภาพใบหน้าที่ใช้ในการฝึกฝนของวิธีใหม่ $RW(2D)^2LDA$ และ $W(2D)^2LDA$ กับวิธีเดิม $(2D)^2LDA$ แสดงในรูปที่ 4.5 จะเห็นว่าวิธี $RW(2D)^2LDA$ และ $W(2D)^2LDA$ มีสมรรถนะการรู้จำที่ดีกว่าวิธี $(2D)^2LDA$ แบบเดิม



รูปที่ 4.5 อัตราการรู้จำใบหน้าภายใต้เงื่อนไขของมุมและความสว่างใบหน้า

ตารางที่ 4.1 อัตราการรู้จำใบหน้าภายใต้เงื่อนไขของมุม และความสว่างใบหน้า

มิติคุณลักษณะ	(2D) ² LDA	W(2D) ² LDA	RW(2D) ² LDA
1	14.50	13.50	19.50
2	72.50	75.00	81.00
3	88.50	89.50	91.50
4	90.00	94.00	95.50
5	91.50	95.00	95.00
6	91.50	95.00	96.50
7	93.00	95.50	96.50
8	93.00	95.50	97.00
9	93.50	96.00	97.50
10	93.50	96.50	97.00

ลำดับต่อไปเป็นการทดสอบความถูกต้องในการรู้จำ เพื่อหาประสิทธิภาพที่เพิ่มขึ้นของวิธีใหม่ ซึ่งเปรียบเทียบวิธีใหม่ที่นำเสนอ 2 วิธีคือ RW(2D)²LDA และ W(2D)²LDA กับวิธีแบบเดิมคือ (2D)²LDA ที่จำนวนมิติคุณลักษณะ 10 มิติ เพื่อให้เห็นสมรรถนะที่ดีขึ้นอย่างชัดเจนของวิธีใหม่ จากผลการทดสอบในตารางพบว่า วิธีใหม่ที่นำเสนอคือ RW(2D)²LDA และ W(2D)²LDA สามารถให้อัตรา

การรู้จำใบหน้าสูงกว่าวิธี (2D)²LDA 2.10% และ 8.11% ตามลำดับ ประเด็นที่น่าสนใจของวิธีใหม่ทั้งสองวิธีคือ มีประสิทธิภาพการคำนวณที่ดีกว่าเนื่องจากมีอัตราการเรียนรู้จำสูงกว่าที่มีคุณลักษณะต่ำกว่าวิธีเดิม

4.3 การทดสอบภายใต้เงื่อนไขการเปลี่ยนขนาดและการเอียงใบหน้า (Scaling and rotation)

เนื่องจากฐานข้อมูล ORL ถูกออกแบบมาเพื่อใช้สำหรับการทดสอบสมรรถนะของอัลกอริทึมในการรู้จำใบหน้าภายใต้เงื่อนไขการเปลี่ยนขนาดและการเอียงใบหน้าโดยเฉพาะ รวมทั้งการเก็บภาพก็ทำในเวลาต่างกัน และรวมเงื่อนไขการเปลี่ยนแปลงต่างๆ อีก เช่น การแสดงอารมณ์ทางใบหน้า (เปิดตา ปิดตา ยิ้ม หรือไม่ยิ้ม ใส่แว่น หรือไม่ใส่แว่น) มุมของภาพจะหมุนเอียงไปภาพละ 20 องศา ทุกภาพเป็นภาพสีเทา การทดสอบนี้แยกฐานข้อมูลออกเป็นสองชุดเท่าๆ กัน แบ่งเป็นกลุ่ม กลุ่มแรกใช้ฝึกฝนระบบภาพแต่ละคนมีชุดละ 5 ภาพ ส่วนภาพที่เหลือให้เป็นชุดทดสอบ นั่นคือ ภาพบุคคล 40 คน จำนวนทั้งหมด 400 ภาพ จะถูกแบ่งไปใช้เป็นชุดฝึกฝน 200 ภาพ และเป็นชุดทดสอบ 200 ภาพ ตัวอย่างภาพการเปลี่ยนขนาดและการเอียงใบหน้าแสดงในรูปที่ 4.6



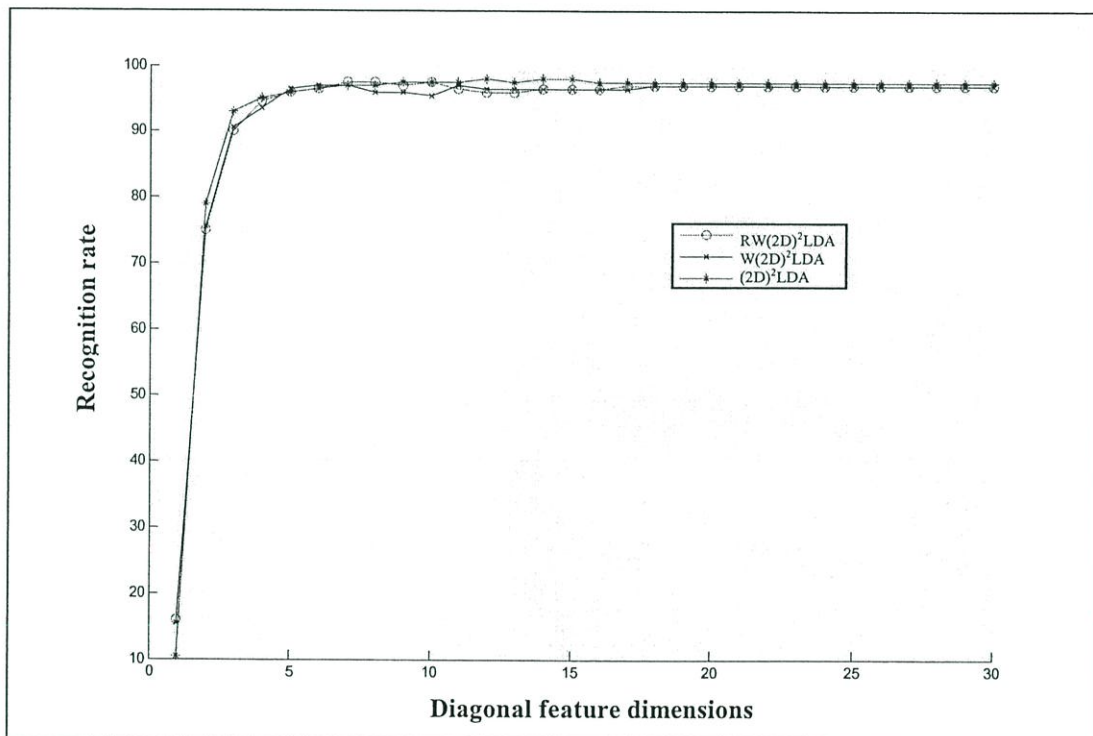
รูปที่ 4.6 ภาพใบหน้าจากฐานข้อมูล ORL ภายใต้เงื่อนไขการเปลี่ยนขนาดและการเอียงใบหน้า

ตารางที่ 4.2 แสดงผลการเปรียบเทียบสมรรถนะของอัลกอริทึมทั้ง 3 วิธี โดยแสดงเป็นค่าความถูกต้องในการรู้จำสูงสุด พร้อมกับมิติคุณลักษณะที่ให้ค่าความถูกต้องในการรู้จำสูงสุดนั้น จากตารางจะเห็นว่า วิธี RW(2D)²LDA และ W(2D)²LDA มีความถูกต้องในการรู้จำสูงกว่าวิธี (2D)²LDA และมิติคุณลักษณะที่ใช้ทดสอบเพื่อให้ได้การรู้จำใบหน้าสูงสุดลดลง แต่วิธี RW(2D)²LDA และ W(2D)²LDA ก็ใช้เวลาในการคำนวณมากกว่าวิธี (2D)²LDA เพราะอัลกอริทึมที่นำเสนอมีการคำนวณฟังก์ชันการถ่วงน้ำหนักเพิ่มขึ้น

ตารางที่ 4.2 ผลการเปรียบเทียบสมรรถนะภายใต้เงื่อนไขการเปลี่ยนขนาดและการเอียงใบหน้า

อัลกอริทึม	อัตราการเรียนรู้สูงสุด	มิติคุณลักษณะ	เวลาที่ใช้ในการคำนวณ (วินาที)
(2D) ² LDA	98.50	8×8	4.14
W(2D) ² LDA	98.75	7×8	5.26
RW(2D) ² LDA	99.15	6×7	7.85

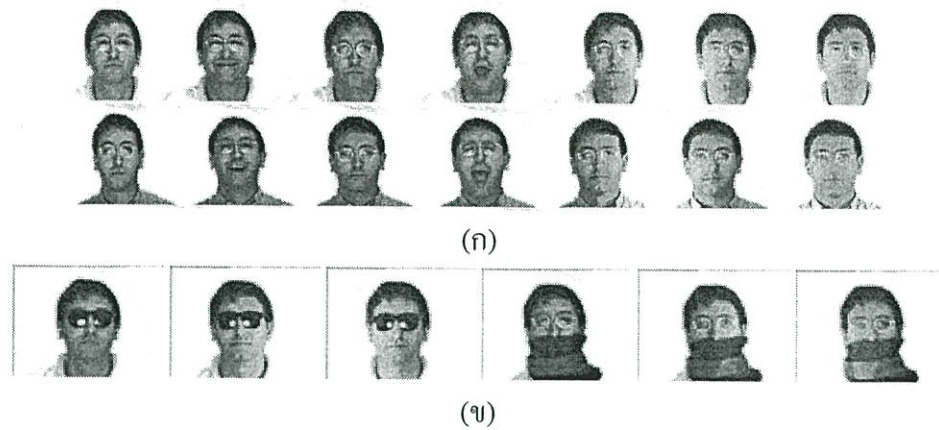
ผลการเปรียบเทียบอัตราความถูกต้องในการรู้จำภายใต้เงื่อนไขการเปลี่ยนขนาดและการเอียงใบหน้าแสดงดังรูปที่ 4.7



รูปที่ 4.7 อัตราการเรียนรู้ใบหน้าภายใต้เงื่อนไขการเปลี่ยนขนาดและการเอียงใบหน้า

4.4 การทดสอบภายใต้เงื่อนไขระยะเวลาและการบังใบหน้า (Time delay and occlusion)

ฐานข้อมูลภาพใบหน้าของ AR ที่นำมาทดสอบจำนวน 120 คน เป็นชาย 65 คน และหญิง 55 คน โดยแบ่งภาพใบหน้าแต่ละบุคคลออกเป็น 2 กลุ่ม ถ่ายภาพในเวลาที่แตกต่างกัน 2 สัปดาห์ ทำการคัดเลือกภาพบุคคลในแต่ละกลุ่มมาใช้ทดสอบบุคคลละ 7 ภาพ ประกอบด้วยภาพที่มีความสว่าง และการบังใบหน้า ดังตัวอย่างในรูปที่ 4.8 ภาพทั้งหมดถูกแปลงเป็นภาพระดับเทาขนาด 8 บิต ที่ความละเอียด 50×40 พิกเซลก่อนนำมาใช้งาน



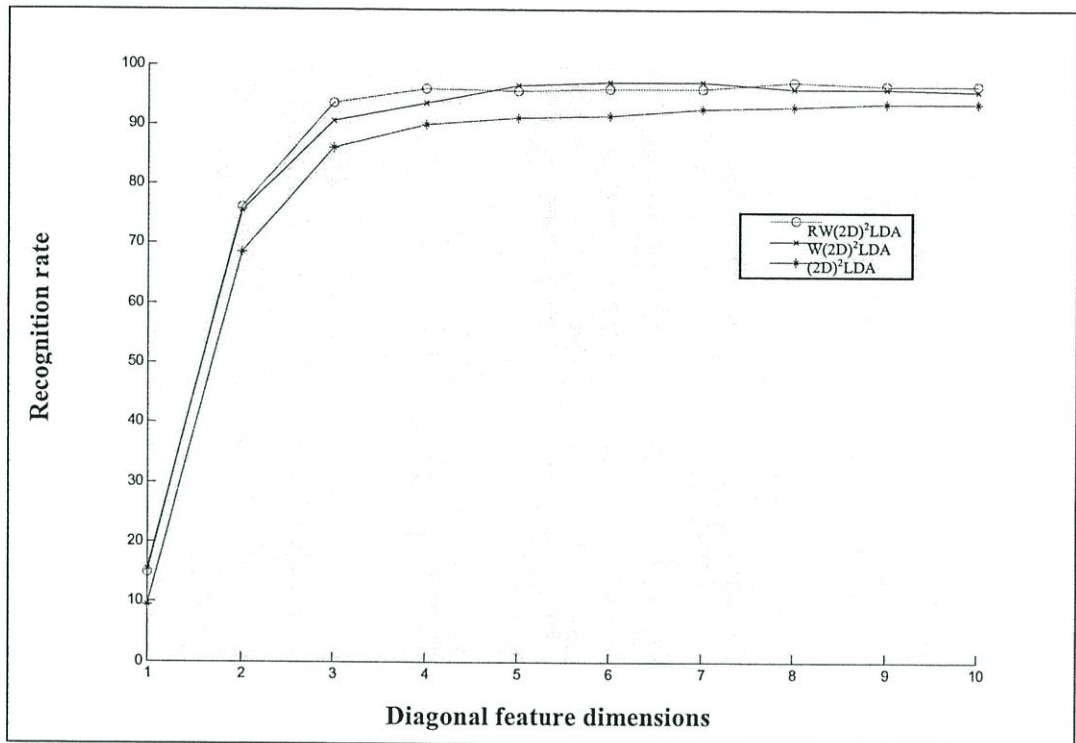
รูปที่ 4.8 ภาพใบหน้าจากฐานข้อมูล AR ภายใต้เงื่อนไขระยะเวลา (ก) และการบังใบหน้า (ข)

ต่อไปเป็นการทดสอบวิธี $(2D)^2LDA$ กับวิธี $W(2D)^2LDA$ และ $RW(2D)^2LDA$ ภายใต้เงื่อนไขการถ่ายภาพที่ระยะเวลาต่างกัน โดยใช้ภาพจากรูปที่ 4.8 (ก) กลุ่มที่ 1 (แถวบน) ทั้งหมดมาใช้ฝึกฝนระบบและภาพจากรูปที่ 4.8 (ก) กลุ่มที่ 2 (แถวล่าง) มาใช้ทดสอบ รวมภาพที่ใช้ในการฝึกฝนและทดสอบทั้งหมดจำนวน 840×2 ภาพ การหาคุณลักษณะของภาพด้วยวิธีทั้งสามภายใต้เงื่อนไขเวลานั้น ได้ทำการหาความถูกต้องในการรู้จำใบหน้าตั้งแต่มิติคุณลักษณะที่ $1 \times 1, 1 \times 2, 1 \times 3, \dots, 2 \times 1, 2 \times 2$ ไปจนถึง มิติที่ 10×10 จากนั้นจะใช้วิธี 1NN เป็นการจำแนกบุคคลโดยใช้ความใกล้เคียงกันมากที่สุดของภาพใบหน้าที่ถูกหาคุณลักษณะ ผลการเปรียบเทียบอัตราความถูกต้องในการรู้จำใบหน้าของทั้ง 3 วิธีนำเสนอในตารางที่ 4.3 แต่ละวิธีแสดงอัตราการรู้จำสูงสุดที่มีมิติคุณลักษณะต่างๆ

ตารางที่ 4.3 อัตราการรู้จำใบหน้าภายใต้เงื่อนไขระยะเวลาและการบังใบหน้า

อัลกอริทึม	อัตราการรู้จำสูงสุด	มิติคุณลักษณะ
$(2D)^2LDA$	93.50	7×9
$W(2D)^2LDA$	97.50	5×8
$RW(2D)^2LDA$	98.00	6×10

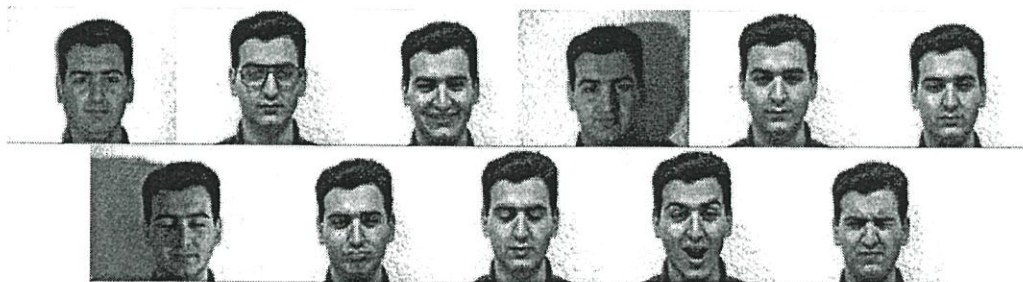
ส่วนการเปรียบเทียบด้วยกราฟกระทำในสองมิติ นำเสนอเพียง 10 มิติ ก็สามารถเห็นความแตกต่างได้ โดยผลการทดสอบแสดงในรูปที่ 4.9



รูปที่ 4.9 อัตราการเรียนรู้จำใบหน้าภายใต้เงื่อนไขระยะเวลาและการบังใบหน้า

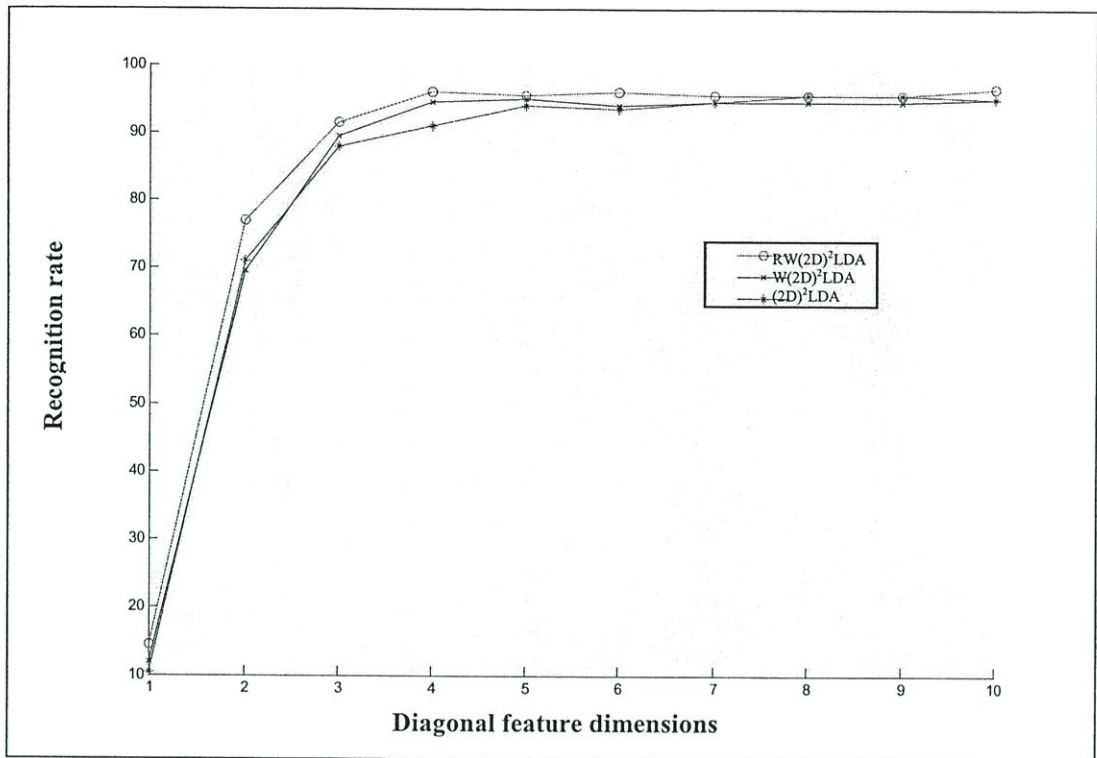
4.5 การทดสอบภายใต้เงื่อนไขการแสดงอารมณ์และความสว่างใบหน้า (Expression and illumination)

สุดท้ายเป็นการทดสอบสมรรถนะภายใต้เงื่อนไขการแสดงอารมณ์ทางใบหน้าและความสว่างภาพใบหน้าจากฐานข้อมูล YALE ภาพจากฐานข้อมูลนี้เป็นภาพสีเทามีฉากหลังเป็นสีขาว การทดสอบใช้ภาพทั้งหมด 165 ภาพ จากบุคคลจำนวน 15 คน แต่ละคนประกอบด้วย 11 ภาพ และการแสดงอารมณ์และรายละเอียดของแสงแตกต่างกัน ตัวอย่างภาพดังแสดงในรูปที่ 4.10



รูปที่ 4.10 ภาพใบหน้าจากฐานข้อมูล YALE ภายใต้เงื่อนไขการแสดงอารมณ์และความสว่างใบหน้า

รูปที่ 4.11 แสดงผลการทดสอบภายใต้เงื่อนไขการแสดงอารมณ์ทางใบหน้า และความสว่างใบหน้า ที่มีทัศนลักษณะตามแนวทแยง 10 มิติ



รูปที่ 4.11 อัตราการรู้จำใบหน้าภายใต้เงื่อนไขการแสดงอารมณ์และความสว่างของใบหน้า

ส่วนตารางที่ 4.3 แสดงความถูกต้องในการรู้จำใบหน้าสูงสุดภายใต้เงื่อนไขการเปลี่ยนแปลงการแสดงอารมณ์ทางใบหน้าและเงื่อนไขความสว่างของภาพ ซึ่งได้จากการประมวลผลของอัลกอริทึมทั้ง 3 วิธี เปรียบเทียบโดยการปรับเปลี่ยนเพิ่มจำนวนตัวอย่างฝึกฝนไปครั้งละ 2 ตัวอย่างต่อข้อมูลหนึ่งกลุ่ม

ตารางที่ 4.4 ผลการเปรียบเทียบสมรรถนะภายใต้เงื่อนไขระยะเวลาและความสว่างใบหน้า

อัลกอริทึม	จำนวนตัวอย่างภายในกลุ่ม			
	2	4	6	8
(2D) ² LDA	86.25	93.60	94.67	96.00
W(2D) ² LDA	87.36	95.00	96.50	98.45
RW(2D) ² LDA	89.57	96.48	97.20	99.40

4.6 สรุปผลการทดสอบ

จากผลการทดสอบอัตราการรู้จำใบหน้าภายใต้สภาวะและเงื่อนไขต่างๆ สามารถสรุปผลการทดสอบได้ดังต่อไปนี้

ฐานข้อมูลใบหน้า FERET

การทดสอบนี้ใช้ฐานข้อมูลภาพใบหน้า FERET เพื่อหาผลการทดสอบเปรียบเทียบอัตราความถูกต้องในการรู้จำของวิธี $(2D)^2LDA$ แบบเดิมกับวิธีใหม่ที่น่าเสนอคือ $W(2D)^2LDA$ และ $RW(2D)^2LDA$ ภายใต้เงื่อนไขของจำนวนตัวอย่างภาพใบหน้าที่ใช้ในการฝึกฝนระบบ โดยหาความถูกต้องในการรู้จำจากจำนวนมิติคุณลักษณะทั้งด้านแถวและคอลัมน์จำนวน 60×60 มิติ ในรูปแบบกราฟ 3 มิติ วิธี $RW(2D)^2LDA$ ได้รับความถูกต้องในการรู้จำใบหน้า 98.00% วิธี $W(2D)^2LDA$ มีอัตราความถูกต้องในการรู้จำใบหน้า 96.50% และวิธี $(2D)^2LDA$ มีอัตราความถูกต้องในการรู้จำใบหน้าต่ำสุดที่ 94.00%

ฐานข้อมูลใบหน้า PIE

การทดสอบนี้เป็นการทดสอบภายใต้เงื่อนไขมุมของภาพ และความสว่างภาพใบหน้า (Pose and Illumination) เนื่องจากฐานข้อมูล PIE มีภาพแต่ละบุคคลที่มุมแตกต่างกัน 13 ภาพ และเงื่อนไขความสว่างต่างๆ กัน 43 เงื่อนไข จึงเหมาะสำหรับการทดสอบตามเงื่อนไข และนำเสนอผลการทดสอบด้วยกราฟ 2 มิติ เลือกมิติคุณลักษณะในแนวทแยงมุมเพื่อให้การเปรียบเทียบชัดเจนขึ้น ผลการเปรียบเทียบอัตราความถูกต้องในการรู้จำใบหน้าภายใต้เงื่อนไขมุมของภาพ และความสว่างใบหน้าพบว่าวิธี $RW(2D)^2LDA$ และ $W(2D)^2LDA$ มีสมรรถนะการรู้จำที่ดีกว่าวิธี $(2D)^2LDA$ แบบเดิม และประสิทธิภาพที่เพิ่มขึ้นของวิธีใหม่ $RW(2D)^2LDA$ และ $W(2D)^2LDA$ เปรียบเทียบกับวิธีเดิม $(2D)^2LDA$ ด้วยมิติคุณลักษณะ 10 มิติ พบว่า วิธี $RW(2D)^2LDA$ และ $W(2D)^2LDA$ สามารถให้อัตราการรู้จำใบหน้าสูงกว่าวิธี $(2D)^2LDA$ 2.10% และ 8.11% ตามลำดับ

ฐานข้อมูลใบหน้า ORL

ฐานข้อมูล ORL ถูกออกแบบมาเพื่อใช้สำหรับการทดสอบสมรรถนะการรู้จำใบหน้าภายใต้เงื่อนไขการเปลี่ยนแปลงขนาดและการเอียงใบหน้า (Scaling and rotation) การทดสอบนี้ทำโดยแบ่งฐานข้อมูลออกเป็นสองชุดเท่าๆ กัน ใช้ฝึกฝนระบบและทดสอบ ผลการเปรียบเทียบสมรรถนะของทั้ง 3 วิธี พบว่าวิธี $RW(2D)^2LDA$ และ $W(2D)^2LDA$ มีความถูกต้องในการรู้จำสูงกว่าวิธี $(2D)^2LDA$ และมีมิติคุณลักษณะที่ให้ความถูกต้องในการรู้จำสูงสุดลดลง แต่ใช้เวลาในการคำนวณเพิ่มขึ้น

ฐานข้อมูลใบหน้า AR

ฐานข้อมูลภาพใบหน้าของ AR ใช้ทดสอบภายใต้เงื่อนไขระยะเวลาและการบังใบหน้า (Time delay and occlusion) เพราะประกอบด้วยภาพ 2 ชุด ที่ถ่ายในระยะเวลาต่างกัน 2 สัปดาห์ และภาพใบหน้าแต่ละชุดประกอบด้วยภาพการบังใบหน้า ผลการเปรียบเทียบอัตราความถูกต้องในการรู้จำใบหน้าของทั้ง 3 วิธี พบว่า วิธี $RW(2D)^2LDA$ และ $W(2D)^2LDA$ มีสมรรถนะเพิ่มสูงกว่าวิธี $(2D)^2LDA$ โดยที่มิติคุณลักษณะที่ให้ความถูกต้องในการรู้จำสูงสุดลดลงเช่นเดียวกัน

ฐานข้อมูลใบหน้า YALE

ฐานข้อมูล YALE ใช้ทดสอบสมรรถนะภายใต้เงื่อนไขการแสดงอารมณ์ทางใบหน้าและความสว่างของภาพ (Expression and illumination) เนื่องจากภาพจากฐานข้อมูลนี้ประกอบด้วยภาพที่มีการแสดงอารมณ์และรายละเอียดของแสงแตกต่างกัน ผลการทดสอบที่มีมิติคุณลักษณะตามแนวทแยง 10 มิติ เปรียบเทียบโดยการปรับเปลี่ยนเพิ่มจำนวนตัวอย่างฝึกฝนไปครั้งละ 2 ตัวอย่างต่อข้อมูลหนึ่งกลุ่ม วิธี RW(2D)²LDA แสดงความถูกต้องในการรู้จำใบหน้าสูงสุด 99.40% ที่มีมิติคุณลักษณะ 8 มิติ

บทที่ 5

สรุปผลการวิจัยและแนวทางในการพัฒนา

บทนี้เป็นการนำเสนอสรุปผลการวิจัย และแนวทางที่สามารถจะพัฒนา ศึกษาวิจัยต่อไปในอนาคต เนื่องจากงานวิจัยนี้เป็นการศึกษาการรู้จำใบหน้าด้วยหลักการวิเคราะห์ลักษณะแตกต่างเชิงเส้น (Linear Discriminant Analysis) หรือ LDA ในรูปแบบสองมิติและสองทิศทาง ประยุกต์กับฟังก์ชันการถ่วงน้ำหนัก ผู้วิจัยจึงขอเสนอสรุปผลการวิจัย และแนวทางที่อาจเป็นไปได้ในการพัฒนาวิธีใหม่ ต่อจากแนวทางที่นำเสนอในวิทยานิพนธ์ฉบับนี้ เพื่อเป็นประโยชน์สำหรับผู้วิจัยที่สนใจการวิจัยระบบการรู้จำใบหน้าตามแนวทางการวิเคราะห์ลักษณะแตกต่างเชิงเส้นต่อไป

5.1 สรุปผลการวิจัย

วิทยานิพนธ์นี้เป็นการศึกษาการรู้จำใบหน้าโดยการนำวิธี LDA แบบเดิมที่พัฒนาเป็นแบบสองมิติสองทิศทาง หรือ (2D²)LDA ในปัจจุบัน มาประยุกต์ด้วยการเพิ่มฟังก์ชันการถ่วงน้ำหนักให้เป็นวิธีใหม่สองวิธีคือ การวิเคราะห์ลักษณะแตกต่างเชิงเส้นสองมิติสองทิศทางแบบถ่วงน้ำหนัก (Weighted 2-dimensional and 2-directional LDA) และแบบถ่วงน้ำหนักสัมพัทธ์ (Relevance weighted 2-dimensional and 2-directional LDA) ซึ่งเรียกรวมกันว่า W(2D)²LDA และ RW(2D)²LDA ตามลำดับ เพื่อแก้ปัญหาการปนกันของข้อมูลระหว่างกลุ่ม โดยการคำนวณ Weighting function เพื่อถ่วงน้ำหนักข้อมูลให้มีการคัดแยกข้อมูลที่ปนกันระหว่างกลุ่ม และคำนวณ Relevance Weighting function เพื่อกระชับข้อมูลภายในกลุ่มให้จัดกลุ่มได้ดีขึ้น ซึ่งมีผล ต่อการแยกแยะ และคำนวณอัตราความถูกต้องในการรู้จำใบหน้าได้ดีขึ้น จากผลการทดสอบภายใต้เงื่อนไขต่างๆ กับฐานข้อมูลภาพใบหน้าจำนวน 5 ฐานข้อมูล สามารถสรุปผลการทดสอบได้ ดังนี้

- วิธีใหม่ที่นำเสนอทดสอบกับฐานข้อมูลภาพใบหน้า FERET ด้วย 60×60 Feature dimensions ได้อัตราความถูกต้องในการรู้จำสูงสุด 98.00%
- เงื่อนไขมุม และความสว่างภาพใบหน้าที่ทดสอบกับฐานข้อมูล PIE ด้วยกราฟ Diagonal feature dimension พบว่า วิธีใหม่ที่นำเสนอมีประสิทธิภาพการรู้จำใบหน้าเพิ่มขึ้น 5.1%
- เงื่อนไขการเปลี่ยนแปลงขนาดและการเอียงใบหน้าที่ทดสอบกับฐานข้อมูล ORL และเงื่อนไขระยะเวลาและการบังใบหน้าที่ทดสอบกับฐานข้อมูล AR พบว่า วิธีใหม่ที่นำเสนอมีประสิทธิภาพการรู้จำใบหน้าที่ดี ขณะที่มิติคุณลักษณะลดลง
- เงื่อนไขการแสดงอารมณ์ทางใบหน้าและความสว่างของภาพทดสอบกับฐานข้อมูล YALE ที่มีมิติคุณลักษณะตามแนวทแยงมุม 10 มิติ ด้วยการเพิ่มจำนวนตัวอย่างฝึกฝนไปครั้งละ 2 ตัวอย่างต่อข้อมูลหนึ่งกลุ่ม พบว่า วิธีใหม่ที่นำเสนอมีประสิทธิภาพการรู้จำใบหน้าสูงสุด 99.40%

- การทดสอบตามเงื่อนไขต่างๆ เป็นการทดสอบในภาพรวม เพื่อหาสมรรถนะการรู้จำใบหน้าของวิธีที่นำเสนอเปรียบเทียบกับวิธีเดิม โดยไม่ได้เน้นการทดสอบเฉพาะด้านในแต่ละเงื่อนไขให้เห็นความแตกต่างที่ละเอียดและชัดเจนขึ้นอย่างไร อีกทั้งวิธีที่นำเสนอมีข้อด้อยที่การใช้เวลาในการคำนวณมากกว่าวิธีเดิม ซึ่งไม่อาจหลีกเลี่ยงได้ เนื่องจากต้องเพิ่มฟังก์ชันการถ่วงน้ำหนัก ทำให้ต้องมีการคำนวณเพิ่มขึ้นในขั้นตอนการคำนวณหาเมตริกซ์ลักษณะ
- สุดท้ายการแสดงผลการทดสอบภายใต้เงื่อนไขการเปลี่ยนขนาดและการเอียงใบหน้าในรูปกราฟไม่สามารถแยกความแตกต่างของแต่ละวิธีได้ชัดเจน ทั้งนี้เนื่องจากภาพใบหน้าที่ใช้ทดสอบมีขนาดใบหน้าและมุมการเอียงไม่แตกต่างกันมากนัก ควรปรับเปลี่ยนฐานข้อมูลในการทดสอบ และการนำเสนอกราฟผลการทดสอบให้เหมาะสมยิ่งขึ้น

5.2 แนวทางในการพัฒนา

ไม่นานมานี้มีงานวิจัยที่นำเสนอวิธีใหม่เรียกว่า $(2D)^2PCALDA$ โดยคัดแยกคุณลักษณะเวกเตอร์การฉายที่เหมาะสมโดยตรงจากภาพ 2 มิติ โดยพิจารณาทิศทางด้านแถวของข้อมูลด้วยวิธี 2DPCA ส่วนด้านคอลัมน์ของข้อมูลพิจารณาด้วยวิธี 2DLDA สามารถหลีกเลี่ยงปัญหาเมตริกซ์คุณลักษณะที่มีขนาดใหญ่ของวิธี 2DPCA และ 2DLDA และแยกแยะและบ่งบอกข้อมูลภาพได้ดีขึ้นด้วย ผลการทดสอบกับฐานข้อมูลภาพใบหน้าของ YALE และ ORL แสดงถึงประสิทธิภาพที่เพิ่มขึ้นของวิธีที่นำเสนอจากผลงานวิจัยนี้ ผู้วิจัยจึงมีแนวคิดที่ว่า วิธี $(2D)^2PCALDA$ เป็นแนวทางในการศึกษาวิจัยต่อไป โดยสลับวิธีการพิจารณาทิศทางของข้อมูลด้านแถวและคอลัมน์ของวิธี $(2D)^2PCALDA$ เสียใหม่ ด้วยการพิจารณาด้านแถวของข้อมูลด้วยวิธี 2DLDA ส่วนทิศทางด้านคอลัมน์ของข้อมูลพิจารณาด้วยวิธี 2DPCA นอกจากนี้ยังประยุกต์ฟังก์ชันการถ่วงน้ำหนักเข้ากับวิธีทั้งสองได้อีก 5 แนวทาง ได้แก่ Weighted $(2D)^2PCALDA$, Relevance Weighted $(2D)^2PCALDA$, $(2D)^2LDAPCA$, Weighted $(2D)^2LDAPCA$, และ Relevance Weighted $(2D)^2LDAPCA$ เป็นต้น นอกจากนี้ ทั้ง 5 แนวทาง ยังสามารถเปลี่ยนตัวถ่วงน้ำหนักทั้งสองเป็นระยะมาฮาธาโนบิสได้อีกแนวทางหนึ่ง ซึ่งเป็นวิธีที่น่าสนใจถ้ามีการทดสอบเปรียบเทียบประสิทธิภาพของการรู้จำใบหน้าด้วยวิธีการประยุกต์ฟังก์ชันการถ่วงน้ำหนักด้วยระยะยูคลิเดียนและระยะมาฮาธาโนบิส จึงขอเสนอเป็นแนวทางในการทำวิจัยต่อไป

บรรณานุกรม

- [1] M. Turk and A. Pentland, “**Eigenfaces for Recognition,**” J. Cognitive Neuroscience, Vol. 3, No. 1, pp. 71-86, 1991.
- [2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “**Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,**” IEEE Trans. on PAMI, Vol. 19, No. 7, pp. 711-720, July 1997.
- [3] L. Chen, H. M. Liao, M. Ko, J. Lin, and G. Yu, “**A New LDA-Based Face Recognition System Which can Solve the Small Sample Size Problem,**” Pattern Recognition, Vol. 33, No. 10, pp. 1713-1726, 2000.
- [4] H. Yu and J. Yang, “**A Direct LDA Algorithm for High-Dimensional Data—with Application to Face Recognition,**” Pattern Recognition, Vol. 34, No. 10, pp. 2067-2070, 2001.
- [5] A.M. Martinez and A.C. Kak, “**PCA Versus LDA,**” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 23, No. 2, pp. 228-233, 2001.
- [6] H-C. Kim, D. Kim, and S.W. Bang, “**Face Recognition Using LDA Mixture Model,**” Pattern Recognition Letters, Vol. 24, No. 15, pp. 2815-2821, 2003.
- [7] M. Ordowski and G. Meyer, “**Geometric Linear Discriminant Analysis for Pattern Recognition,**” Pattern Recognition Letters, Vol. 37, No. 3, pp. 421-428, 2004.
- [8] X.-Y. Jing, D. Zhang, and Y.-F. Yao, “**Improvements on the Linear Discriminant Technique with Application to Face Recognition,**” Pattern Recognition Letters, Vol. 24, No. 15, pp. 2695-2701, 2003.
- [9] F. Song, D. Zhang, J. Wang, H. Liu, and Q. Tao, “**A Parameterized Direct LDA and Its Application to Face Recognition,**” Neurocomputing, Vol. 24, pp. 2695-2701, 2003.
- [10] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang, “**Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition,**” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 26, No. 1, pp. 131-137, 2004.
- [11] J. Meng and W. Zhang, “**Volume Measure in 2DPCA-Based Face Recognition,**” Pattern Recognition Letters, Vol. 28, No. 10, pp. 1203-1208, 2007.
- [12] Y. Wen and P. Shi, “**Image PCA: A New Approach for Face Recognition,**” Proc. IEEE Inter. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Vol. 1, pp. I-1241-1244, 2007.

- [13] H. Xiong, M.N.S. Swamy, and M.O. Ahmad, “**Two-Dimensional FLD for Face Recognition**,” *Pattern Recognition*, Vol. 38, No. 7, pp. 1121-1124, 2005.
- [14] J. Yang, D. Zhang, X. Yong, and J. Y. Yang, “**Two-Dimensional Discriminant Transform for Face Recognition**,” *Pattern Recognition*, Vol. 38, No. 7, pp. 1125-1129, 2005.
- [15] M. Li and B. Yuan, “**2D-LDA: A Statistical Linear Discriminant Analysis for Image Matrix**,” *Pattern Recognition Letters*, Vol. 26, No. 5, pp. 527-532, 2005.
- [16] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. J. Zhang, “**Multilinear Discriminant Analysis for Face Recognition**,” *IEEE Trans. Image Processing*, Vol. 16, pp. 212-220, 2007.
- [17] D.S. Guru and T.N. Vikram, “**2D Pairwise FLD: A Robust Methodology for Face Recognition**,” *IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 99-102, 2007.
- [18] D. Zhang and Z. Zhou, “**(2D)²PCA: Two-Directional Two-Dimensional PCA for Efficient Face Representation and Recognition**,” *Neurocomputing*, Vol. 69, pp. 224-231, 2005.
- [19] P. Nagabhushan, D.S. Guru, and B.H. Shekar, “**(2D)² FLD: An Efficient Approach for Appearance Based Object Recognition**,” *Neurocomputing*, Vol. 69, pp. 934-940, 2006.
- [20] S. Nousath, G. Hemantha, and P. Shivakumara, “**(2D)²LDA: An Efficient Approach for Face Recognition**,” *Pattern Recognition*, Vol. 39, No. 7, pp. 1396-1400, 2006.
- [21] Y. Li, Y. Gao, and H. Erdogan, “**Weighted Pairwise Scatter to Improve Linear Discriminant Analysis**,” *Proc. 6th Inter. Conf. Spoken Language Processing (ICSLP)*, Vol. 4, pp. 608-611, 2000.
- [22] R. Lotlikar and R. Kothari, “**Fractional-Step Dimensionality Reduction**,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 6, pp. 623-627, 2000.
- [23] M. Loog, R.P.W. Duin, and R. Haeb-Umbach, “**Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria**,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, No. 7, pp. 762-766, 2001.
- [24] B. Yu, L. Jin, and P. Chen, “**A New LDA-Based Method for Face Recognition**,” *Proc. IEEE Int. Conf. Pattern Recognition*, Vol. 1, pp. 168-171, 2002.
- [25] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, “**Face Recognition Using LDA-Based Algorithms**,” *IEEE Trans. Neural Networks*, Vol. 14, No. 1, pp. 195-200, 2003.
- [26] J.R. Price and T.F. Gee, “**Face Recognition Using Direct, Weighted Linear Discriminant Analysis and Modular Subspaces**,” *Pattern Recognition*, Vol. 38, No. 2, pp. 209-219, 2005.

- [27] H. Wang, S. Chen, and Z. Hu, “**Image Recognition Using Weighted Two-Dimensional Maximum Margin Criterion,**” Proc. 3rd Int. Conf. Natural Computation (ICNC), Vol. 1, pp. 582-58, 62007.
- [28] J. Wang, E. Sung, and W. Yau, “**Weighting Multi-Block Two-Dimensional Linear Discriminant Features for Face Verification,**” Proc. 6th Int. Conf. on Information, Communications and Signal Processing (ICICS), pp. 1-5, 2007.
- [29] E.K. Tang, P.N. Suganthan, X. Yao and A.K. Qin, “**Linear Dimensionality Reduction Using Relevance Weighted LDA,**” Pattern Recognition, Vol. 38, No. 4, pp. 485-493, 2005.
- [30] Y. X. Liang W.G. Gong, Y.J. Pan and W.H. Li, “**Generalizing Relevance Weighted LDA,**” Pattern Recognition, Vol. 38, No. 11, pp. 2217-2219, 2005.
- [31] K. Choug dali, M. Jedra, and N. Zahid, “**Face Recognition Using Relevance Weighted LDA with QR Decomposition,**” J. Graphics, Vision and Image Processing (GVIP), Vol. 6, No. 2, , pp. 27-312006.
- [32] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld, “**Face Recognition: A Literature Survey,**” ACM Computing Surveys, Vol. 35, No. 4, pp. 399-458, December 2003.
- [33] A. F. Abate, M. Nappi, D. Riccio and G. Sabatino, “**2D and 3D Face Recognition: A Survey,**” Pattern Recognition Letters, Vol. 28, pp. 1885 –1906, 2007.
- [34] Ralph Gross, Jianbo Shi, and Jeffrey Cohn, “**Quo Vadis Face Recognition?**” Third Workshop on Empirical Evaluation Methods in Computer Vision, December, 2001.
- [35] Ralph Gross, Simon Baker, Iain Matthews, and Takeo Kanade, “**Face Recognition Across Pose and Illumination,**” Handbook of Face Recognition, Stan Z. Li and Anil K. Jain, ed., Springer-Verlag, 2004.
- [36] Sam T. Roweis and Lawrence K. Saul, “**Nonlinear Dimensionality Reduction by Locally Linear Embedding,**” Science v.290 no.5500, pp.2323-2326, Dec.22, 2000.
- [37] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford, “**A Global Geometric Framework for Nonlinear Dimensionality Reduction,**” Science v.290 no.5500, pp.2319-2323, Dec.22, 2000.
- [38] M. Brand, “**Continuous Nonlinear Dimensionality Reduction by Kernel Eigenmaps,**” Technical Report 2003-21, Mitsubishi Electric Research Labs, Apr. 2003.
- [39] Cottrell, G.W., Fleming, M.K., “**Face Recognition Using Unsupervised Feature Extraction,**” Int. Proc. Intell. Neural Network Conf., pp. 322–325, 1990.

- [40] Lin, Shang-Hung, Kung, Sun-Yuan, Lin, Long-Ji, “**Face Recognition/Detection by Probabilistic Decision-Based Neural Network,**” IEEE Trans. Neural Networks, Vol. 8 No. 1, pp. 114–132, 1997.
- [41] Meng, J.E., Shiqian, W., Juwei, L., Hock, L.T., “**Face Recognition with Radial Basis Function (RBF) Neural Networks,**” IEEE Trans. Neural Networks, Vol. 13, No. 3, pp. 697–710, 2002.
- [42] Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., van der Malsburg, C., Wurtz, R.P., Konen, W., “**Distortion Invariant Object Recognition in The Dynamic Link Architecture,**” IEEE Trans. Comput. Vol. 42, pp. 300–311, March 1993.
- [43] Wiskott, L., Fellous, J.M., Kruger, N., von der Malsburg, C., “**Face Recognition by Elastic Bunch Graph Matching,**” IEEE Trans. Pattern Anal. Machine Intell. Vol. 19, pp. 75–779, July 1997.
- [44] Garcia, C., Zikos, G., Tziritas, G., “**Wavelet Packet Analysis for Face Recognition,**” Image Vision Comput. Vol. 18, pp. 289–297, 2000.
- [45] Liu, C., “**Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition,**” IEEE Trans. Pattern Anal. Machine Intell. Vol. 26, No. 5, pp. 572–581, 2004.
- [46] Riccio, D., Nappi, M., “**Deferring Range/Domain Comparisons in Fractal Image Compression,**” In: Proc. 12th Internat. Conf. on Image Analysis and Processing, Vol. 1 (September), pp. 412–417, 2003.
- [47] Kouzani, A.Z., He, F., Sammut, K., “**Fractal Face Representation and Recognition,**” In: 1997 IEEE Internat. Conf. on Systems, Man, and Cybernetics, Vol. 2 (October), pp. 1609–1613, 1997.
- [48] Tan, T., Yan, H., “**Face Recognition by Fractal Transformations,**” In: Proc. 1999 IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing, 1999. ICASSP’99, Vol. 6 (March), pp. 3537–3540, 1999.
- [49] Chen, X., Flynn, P.J., Bowyer, K.W., “**PCA-Based Face Recognition in Infrared Imagery: Baseline and Comparative Studies,**” In: Proc. IEEE Internat. Workshop on Analysis and Modeling of Faces and Gestures October, pp. 127–134, 2003.
- [50] Socolinsky, D.A., Selinger, A., “**Thermal Face Recognition in An Operational Scenario,**” In: Proc. 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR04), 2004a.

- [51] Socolinsky, D.A., Selinger, A., “**Thermal Face Recognition Over Time,**” In: Proc. 17th Internat. Conf. on Pattern Recognition (ICPR04), 2004b.
- [52] Buddharaju, P., Pavlidis, I., Kakadiaris, I., “**Face Recognition in The Thermal Infrared Spectrum,**” In: 2004 Conf. Computer Vision and Pattern Recognition Workshop June, pp. 133–138, 2004.
- [53] Pan, Z., Healey, G., Prasad, M., Tromberg, B., “**Face Recognition in Hyperspectral Images,**” IEEE Trans. Pattern Anal. Machine Intell. 25 (12), 1552–1560, 2003.
- [54] G. B. Huang, M. Ramesh, T. Berg, and Erik Learned-Miller, “**Labeled Faces in The Wild: A Database for Studying Face Recognition in Unconstrained Environments,**” University of Massachusetts, Amherst, Technical Report 07-49, October, 2007, and the online document “<http://vis-www.cs.umass.edu/lfw/>”.
- [55] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, “**The FERET Evaluation Methodology for Face Recognition Algorithm,**” IEEE Trans. on PAMI, Vol. 22, No. 10, pp. 1090-1104, October 2000, and the online document “http://www.itl.nist.gov/iad/humanid/feret/feret_master.html”.
- [56] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, “**Overview of The Face Recognition Grand Challenge,**” Proc. of CVPR05, No. 1, pp. 947–954, June 2005, and the original document under the directory “BEE_DIST\doc”.
- [57] The technical report JDL-TR-04-FR-001 “**The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations.**”
- [58] The CMU Technical Report CMU-RI-TR-01-02 “**The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces**”.
- [59] A. S. Georghiades and P. N. Belhumeur, “**From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose,**” IEEE Trans. on. PAMI, Vol.23, No.6, pp.643-660, June 2001.
- [60] A. M. Martinez and R. Benavente, “**The AR Face Database, CVC Technical Report #24,**” June 1998.
- [61] The ORL Face Database, <http://www.uk.research.att.com/facedatabase.html>
- [62] F. S. Samaria and A. C. Harter, “**Parameterisation of A Stochastic Model for Human Face Identification,**” Proc. of 2nd IEEE workshop on Applications of Computer Vision, pp. 138-142, 1994.
- [63] Y. Qi and J. Zhang, “**(2D)² PCALDA: An Efficient Approach for Face Recognition,**” Applied Mathematics and Computation, Vol. 213, Issue 1, pp. 1-7, July 2009.

ภาคผนวก

ภาคผนวก ก.

ผลงานวิจัยที่ได้รับการตีพิมพ์เผยแพร่ในการประชุมวิชาการ

1. **W. Sanayha and Y. Rangsanseri, "Hardware Implementation of the Data Encryption Standard (DES),"** 2nd WSEAS Int. Conf. on Signal, Speech and Image Processing (ICOSSIP 2002), pp. 1721-1724, Skiathos, Greece, September 25-28, 2002.
2. **วิญญูติ เสน่หา ปัญญา วิฑิมชฌิมา และยุทธพงษ์ รังสรรค์เสรี, "การรู้จำอักษรตัวพิมพ์ภาษาไทยโดยใช้คุณลักษณะความหนาแน่นและเส้นแถบของตัวอักษร,"** การประชุมวิชาการทางวิศวกรรมไฟฟ้า ครั้งที่ 28 (EECON 28), หน้า 921-924, ภูเก็ต, ประเทศไทย, 20-21 ตุลาคม 2548.
3. **วิญญูติ เสน่หา และยุทธพงษ์ รังสรรค์เสรี, "การหาลำดับข้อมูลการวาดเส้นจากภาพสแตติกตัวอักษรลายมือเขียนภาษาไทย,"** การประชุมวิชาการทางวิศวกรรมไฟฟ้า ครั้งที่ 29 (EECON-29), หน้า 1009-1012, พัทยา จ. ชลบุรี, ประเทศไทย, 9-10 พฤศจิกายน 2549.
4. **W. Sanayha and Y. Rangsanseri, "Recovery of Drawing Order from Static Thai Handwritten Images,"** Proceedings of International Workshop on Advance Image Technology (IWAIT 2007), pp. 365-369, Bangkok, Thailand, 8-9 January 2007.
5. **W. Leelatham, Y. Rangsanseri, and W. Sanayha, "Face Recognition Using Coupled Subspace Linear Discriminant Analysis,"** Proceedings of International Workshop on Advance Image Technology (ICEAST 2007), pp. 254-257, Bangkok, Thailand, 8-9 January 2007.
6. **W. Sanayha, and Y. Rangsanseri, "Off-line Thai Handwritten Character Recognition Using Tracing Order Information,"** Proceedings of International Conference on Engineering, Applied Sciences, and Technology (ICEAST 2007), pp. 258-261, Bangkok, Thailand, 8-9 January 2007.
7. **W. Sanayha, W. Leelatham, and Y. Rangsanseri, "WPS-CSLDA Application for Face Recognition,"** Proceedings of Joint International Conference on Information Communication Technology (JICT 2007), pp. 162-167, Vientiane, Lao PDR, 19-22 December 2007.
8. **W. Sanayha, and Y. Rangsanseri, "Relevance Weighted 2D²LDA Image Projection Technique for Face Recognition Application,"** Proceedings of the Annual International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON 2009), pp. 663-667, Chonburi, Thailand, 6-9 May 2009.

ภาคผนวก ข.

ผลงานวิจัยที่ได้รับการตีพิมพ์เผยแพร่ในวารสารวิชาการระดับนานาชาติ

1. **W. Sanayha, and Y. Rangsanseri, “Weighted LDA Image Projection Technique for Face Recognition,”** IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E92-A, No. 9, pp. 2257-2265, September 2009.
2. **W. Sanayha, and Y. Rangsanseri, “Relevance-Weighted (2D)²LDA Image Projection Technique for Face Recognition,”** ETRI Journal, Vol. 31, No. 4, pp. 438-447, August 2009.

IEICE **TRANSACTIONS**

**on Fundamentals of Electronics,
Communications
and Computer Sciences**

VOL.E92-A
NO.9
SEPTEMBER 2009

A PUBLICATION OF THE ENGINEERING SCIENCES SOCIETY



The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

PAPER

Weighted LDA Image Projection Technique for Face Recognition

Waiyawut SANAYHA^{†a)}, Member and Yuttapong RANGSANSERI[†], Nonmember

SUMMARY In this paper, we propose a novel image projection technique for face recognition applications based on Fisher Linear Discriminant Analysis (LDA). The projection is performed through a couple subspace analysis for overcoming the “small sample size” problem. Also, weighted pairwise discriminant hyperplanes are used in order to provide a more accurate discriminant decision than that produced by the conventional LDA. The proposed technique has been successfully tested on three face databases. Experimental results indicate that the proposed algorithm outperforms the conventional algorithms.

key words: face recognition, weighted pairwise scatter, coupled subspace, linear discriminant analysis (LDA)

1. Introduction

The linear discriminant technique is an important branch of pattern recognition. Owing to its simplicity and practicability, the technique has been widely applied in the field of pattern representation and classification, especially in the area of face recognition that has been a very hot research topic in recent years. Automatic face recognition can be seen as a pattern recognition problem, which is very hard to solve due to its nonlinearity. Particularly, we can think of it as a template matching problem, where recognition has to be performed in a high-dimensional space. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two classic linear discrimination methods.

The PCA is used to project images from the original image space into a face-subspace, where dimensionality is reduced. As a result, this method is very fast when testing new face images. The PCA has been intensively exploited in face recognition applications, but many other linear projection methods have been studied too. The PCA deals with the input data in their entirety, without paying any attention to the underlying structure.

The LDA is one of the most popular linear projection techniques for feature extraction. It expressly provides discrimination among the classes. It finds the set of the most discriminant projection vectors which can map high-dimensional samples onto a low-dimensional space. Using the set of projection vectors determined by LDA as the projection axes, all projected samples will form the maximum between-class scatter and the minimum within-class scatter simultaneously in the projective feature space. Even if the

LDA is often considered to outperform the PCA, an important qualification has to be made. Indeed the LDA provides better classification performances only when a wide training set is available. Besides recent studies also strengthen this argument expressly tackling this problem referred to as the Small Sample Size (SSS) problem. In some approaches, the PCA is considered as a preliminary step in order to reduce the dimensionality of the input space, and then the LDA is applied to the resulting space, in order to perform the real classification. However it has been demonstrated in recent works that, combining PCA and LDA in this way, discriminant as well as redundant information are both discarded. Thus, in some cases the LDA is applied directly on the input space.

The Fisherface method [1] is a well-known “PCA plus LDA” method, which has obtained a certain recognition effect. It first uses PCA to reduce the dimensionality of the feature space. Then, in the PCA transformed space, namely in the non-zero space of the total scatter matrix, S_T , the Fisher discrimination vectors are computed. In [2], an LDA-based technique attempting to solve the small sample size problem is proposed. It has been proved that the most expressive vectors derived in the null space of the within-class scatter matrix using PCA are equal to the optimal discriminant vector derived in the original space using LDA. The between-class scatter is projected into the null space of the within-class scatter and the eigenvectors corresponding to the largest eigenvalues of the transferred between-class scatter are chosen. In [3], a direct, exact LDA algorithm (D-LDA) for high-dimensional data set is presented. The null space of the between-class scatter is removed first, after that, the intrinsic dimensionality of the space is already very small, which significantly restricts further deriving discriminant basis vectors. The concept of fractional dimensionality is introduced in [4], an incremental procedure, called F-LDA algorithm, is more robust to the selection of weighting function and for any given weighting function. The authors in [5] derived a class of computationally inexpensive linear dimension reduction criteria by introducing a weighted variant of the well-known K -class Fisher criterion associated with LDA. They used the Mahalanobis distance between the class i and j in a weighting function of the original model. The method in [6] combines the strengths of the D-LDA and F-LDA approaches, called the Direct-Fractional step (DF-LDA). They introduced a new variant of D-LDA to lower the dimensionality of the original input space that results in a low-dimensional SSS-free subspace where the

Manuscript received May 21, 2008.

Manuscript revised February 1, 2009.

[†]The authors are with King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok 10520, Thailand.

a) E-mail: waiyawut.s@egat.co.th

DOI: 10.1587/transfun.E92.A.2257

most discriminatory features are preserved. A new LDA algorithm called the direct, weighted LDA (DW-LDA) is proposed in [7]. It simultaneously provides the advantages of both direct LDA [3] and weighted pairwise Fisher criteria [4]. A Generalized Null space Uncorrelated Fisher Discriminant Analysis (GNUFDA) technique is proposed in [8]. It integrates the uncorrelated discriminant analysis and weighted pairwise Fisher criterion, which can solve the problem in which some classes are far away and well separated from some other classes. In [9], two novel LDA-based methods, post-processed Fisherfaces (pFisherfaces) and bi-directional PCA plus LDA (BDPCA+LDA), are presented. The pFisherfaces utilizes 2D-Gaussian filter to smooth Fisherfaces, and the BDPCA+LDA is a LDA performed in the BDPCA subspace. In [10], a new feature extraction method-parameterized direct LDA (PD-LDA) presented for small sample size problems. PD-LDA inherits advantages of D-LDA such as “directness” and “efficiency.” Meanwhile, it greatly enhances the effectiveness and robustness of D-LDA. In [11], the weighted piecewise LDA (WPLDA) algorithm is used. It was applied to similarity, or matching error, data and provided a general solution for overcoming the SSS problem, where the lack of sufficient training samples causes improper estimation of a linear separation hyperplane between the classes. The processes can boost the performance of the verification algorithm significantly and drop of the Equal Error Rate (ERR) in the three experimental sets. The authors in [12] calculated the covariance matrix from 2D image straightforwardly and named Two-Dimension Principal Component Analysis (2DPCA). This method could especially reduce column dimension but it did not take on the row dimension. Similarly, the 2DLDA method [13] also computes the covariance matrix using two-dimensional image and considers the class information. This method reduces only the column dimension but does not consider the row dimension. The Eigenimages method [14] reduces both column and row dimensions by using row projected vectors. A novel algorithm, called Coupled Subspace Analysis (CSA), developed in [15] infers two low dimensional coupled subspaces that optimally reconstructed the original matrices from row and column directions collaboratively. The authors proved that PCA and the recently proposed 2DPCA were just simplified special cases of CSA and answered the unsolved theoretical problems in 2DPCA. The application of CSA to LDA is also recently developed in [16], named as Coupled Subspace Linear Discriminant Analysis (CSLDA). On the other hand, the Weighted Pairwise Scatter method is introduced in [17] for solving the confusable data based on LDA.

In this paper, the approach of weighted function in [17] is adapted into our recent work [16], in order to improve its discrimination performance. This algorithm will be subsequently referred as the Weighted Pairwise Scatter-Coupled Subspace LDA (WPS-CSLDA).

This paper is organized as follows: Sect. 2 presents the idea of Coupled Subspace LDA. Next, the weighted pairwise scatter approach and the proposed method are de-

scribed in Sect. 3. Experimental results and performance that compared with the other methods are presented in Sect. 4. A conclusion is finally given in Sect. 5.

2. Coupled Subspace LDA

Subspace learning is a fundamental approach for face recognition and facial expression analysis. In all previous work of subspace learning algorithms, a face image matrix is typically transformed to a vector by concatenating all row vectors, which usually results in some serious problems in practical applications. Firstly, the intrinsic spatial structure information is lost. Secondly, the feature dimension is extremely high even in moderate cases, which will result in the curse of the dimensionality dilemma. Finally, in many cases, the available number of training samples is relatively very small compared to the feature dimension, which will make the algorithms suffer from the small sample size problem.

Since the higher the dimension of the space, the more the computation we need to find a match, a dimensionality reduction technique is used to project the problem in a lower-dimensionality space. Thus, after the linearization the mean vector is calculated, among all images, and subtracted from all the vectors, corresponding to the original faces. The covariance matrix is then computed in order to extract a limited number of its eigenvectors corresponding to the greatest eigenvalues. These few eigenvectors, also referred to as Eigenfaces, represent a base in a low-dimensionality space. When a new image has to be tested, the corresponding Eigenface expansion is computed and compared against the entire database, according to a distance measure.

2.1 The Conventional LDA

The LDA problem is formulated as follows. Let us consider a set of N samples $\{x_1, x_2, \dots, x_N\}$ taking values in an n -dimensional space, and assume that each sample belongs to one of L classes $\{C_1, C_2, \dots, C_L\}$. Let us also consider a linear transformation mapping the original D -dimensional space into a d -dimensional feature space, where $d < D$. The new feature vectors $y_i \in \mathcal{R}^d$ are defined by the following linear transformation:

$$y_i = W^T x_i \quad i = 1, 2, \dots, N \quad (1)$$

where $W \in \mathcal{R}^{D \times d}$ is a matrix with orthonormal columns.

The LDA method selects W in (1) in the way that maximizes the ratio of the between-class scatter and the within-class scatter.

Let the between-class scatter matrix be defined as

$$S_B = \frac{1}{N} \sum_{i=1}^L n_i (\mu_i - \mu) (\mu_i - \mu)^T \quad (2)$$

and the within-class scatter matrix be defined as

$$S_W = \frac{1}{N} \sum_{i=1}^L \sum_{x_k \in C_i} (x_k - \mu_i) (x_k - \mu_i)^T \quad (3)$$

where μ_i is the mean of class C_i ,
 μ is the global sample mean,
and n_i is the number of samples in class C_i .

If S_W is nonsingular, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns which maximizes the ratio between the determinant of the between-class scatter matrix of the projected samples and the determinant of the within-class scatter matrix of the projected samples, i.e.,

$$W_{opt} = \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|} = [w_1 w_2 \dots w_m] \quad (4)$$

where $\{w_i | i = 1, 2, \dots, m\}$ is the set of generalized eigenvectors of S_B and S_W corresponding to the m largest generalized eigenvalues $\{\lambda_i | i = 1, 2, \dots, m\}$, i.e.,

$$S_B w_i = \lambda_i S_W w_i, \quad i = 1, 2, \dots, m \quad (5)$$

Note that there are at most $L - 1$ nonzero generalized eigenvalues, and so an upper bound on m is $L - 1$, where L is the number of classes.

2.2 Coupled Subspace Analysis [15]

An image is directly treated as a 2D matrix, and a new criterion is proposed to infer two low dimensional coupled subspaces that optimally reconstruct the original matrices from row and column directions collaboratively. An efficient approach, namely Coupled Subspace Analysis (CSA), is applied to learn these two subspaces. Then we reveal the essence of each step in CSA and propose an approach to select the dimension numbers for these two subspaces.

Denote matrix $Y_i \in \mathcal{R}^{r \times c}$ be the lower-dimensional matrix representation of sample $X_i \in \mathcal{R}^{m \times n}$ derived from two projection matrices $U \in \mathcal{R}^{m \times r}$ and $V \in \mathcal{R}^{n \times c}$, i.e. $Y_i = U^T X_i V$. Then, the optimal matrices U and V that best reconstruct the original matrices in the sense of least square error should satisfy the following objective function, i.e. Optimal Matrix Reconstruction Criterion.

$$U^*, V^* = \arg \min_{U, V} \sum_i \|U Y_i V^T - X_i\|_F^2 \quad (6)$$

The objective function on (6) is biquadratic and has no close-form solution. Therefore, we design a procedure to obtain the local optimal solution.

For given $U \in \mathcal{R}^{m \times r}$, the objective function of (6) can be rewritten as

$$V^* = \arg \min_V \sum_i \|X_i^U V V^T - X_i\|_F^2 \quad (7)$$

where $X_i^U = U U^T X_i$. The solution of (7) is the eigenvectors of the eigen-decomposition problem $F F^T x = \lambda x$ with

$$F = \begin{bmatrix} X_1^U(1, *)^T, \dots, X_1^U(m, *)^T, \\ X_2^U(1, *)^T, \dots, X_N^U(m, *)^T \end{bmatrix} \quad (8)$$

where $X_i^U(r, *)$ is the r -th row of the image matrix X_i^U and F is the concatenated matrix of all dimension-reduced samples X_i^U .

Similarly, for given $V \in \mathcal{R}^{n \times c}$, the optimization problem in (6) is changed to

$$U^* = \arg \min_U \sum_i \|U U^T X_i^V - X_i\|_G^2 \quad (9)$$

where $X_i^V = X_i V V^T$. The solution of (9) is the eigenvectors of the Eigen-decomposition problem $G G^T x = \lambda x$ with

$$G = \begin{bmatrix} X_1^V(*, 1)^T, \dots, X_1^V(*, n)^T, X_2^V(*, 1)^T, \dots, \\ \dots, X_N^V(*, n)^T \end{bmatrix} \quad (10)$$

where $X_i^V(*, c)$ is the c -th column of the image matrix X_i^V .

2.3 Connection with LDA [16]

From the projected vectors that we define $U \in \mathcal{R}^{m \times r}$ and $V \in \mathcal{R}^{n \times c}$. At this time, we want to project image X onto U and V projection matrices by the following linear transformation.

$$Y = U^T X V \quad (11)$$

Thus, we obtain Y , $r \times c$ feature matrix of image X , which $r < m$ and $c < n$. N denotes the total number of samples of all classes and there are L classes, each class has images N_i . Each training image is denoted by $m \times n$, X_j ($j = 1, 2, \dots, N$). And \bar{X}_i ($i = 1, 2, \dots, L$) denote the mean of class T_i , and the projected class is P_i . After the projection of training image onto U and V , we get the feature matrix

$$Y_j = U^T X_j V, \quad j = 1, 2, \dots, N \quad (12)$$

The U and V projection axis can be calculated by tracing of the covariance matrix in the projected image. From this point of view, we introduce the equation

$$J(U, V) = \frac{P_B}{P_W} \quad (13)$$

These are row parameters

$$P_B = \text{tr}(\bar{S}_B) \quad (14)$$

$$P_W = \text{tr}(\bar{S}_W) \quad (15)$$

where \bar{S}_B denotes the between-class scatter matrix of projected feature vectors of training images, and \bar{S}_W the within-class scatter matrix of projected feature vectors of training images. So,

$$\begin{aligned} \bar{S}_B &= \frac{1}{N} \sum_{i=1}^L n_i (\bar{Y}_i - \bar{Y})(\bar{Y}_i - \bar{Y})^T \\ &= \frac{1}{N} \sum_{i=1}^L n_i [U^T (\bar{X}_i - \bar{X}) V] [U^T (\bar{X}_i - \bar{X}) V]^T \end{aligned} \quad (16)$$

$$\bar{S}_W = \frac{1}{N} \sum_{i=1}^L \sum_{Y_k \in P_i} (Y_k - \bar{Y}_i)(Y_k - \bar{Y}_i)^T$$

$$= \frac{1}{N} \sum_{i=1}^L \sum_{X_k \in P_i} [U^T (X_k - \bar{X}_i) V] [U^T (X_k - \bar{X}_i) V]^T \quad (17)$$

And,

$$\text{tr}(\tilde{S}_B) = \text{tr} \left(\frac{1}{N} \sum_{i=1}^L n_i U^T (\bar{X}_i - \bar{X}) V V^T (\bar{X}_i - \bar{X})^T U \right) \quad (18)$$

$$\text{tr}(\tilde{S}_W) = \text{tr} \left(\frac{1}{N} \sum_{i=1}^L \sum_{X_k \in P_i} U^T (X_k - \bar{X}_i) V V^T (X_k - \bar{X}_i)^T U \right) \quad (19)$$

Two projection matrices (U and V) can not be calculated at the same time. So, we will firstly define the value of V equal to I (I is the identity matrix) to calculate U value. Then, we secondly define the value of U equal to I to calculate V value.

To Calculate U

Denote $V = I$

$$\text{tr}(\tilde{S}_B) = U^T S_B^U U, \quad \text{tr}(\tilde{S}_W) = U^T S_W^U U \quad (20)$$

where

$$S_B^U = \frac{1}{N} \sum_{i=1}^L n_i (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T$$

and

$$S_W^U = \frac{1}{N} \sum_{i=1}^L \sum_{X_k \in P_i} (X_k - \bar{X}_i)(X_k - \bar{X}_i)^T$$

From (13) we get

$$J(U) = \frac{U^T S_B^U U}{U^T S_W^U U} \quad (21)$$

The maximum $J(U)$ is used to evaluate the optimal U projection.

$$U_{opt} = \arg \max_U J(U) \quad (22)$$

S_B^U is non-singular, the solution of U_{opt} is calculated by the eigenvector and eigenvalue from equation.

$$S_B^U U_{opt} = \lambda S_W^U U_{opt} \quad (23)$$

In general, the optimal projection axis is from eigenvector $S_W^{-1} S_B^U$ which relates to the maximum eigenvalue from 1 to m . In fact, we need to select a set of projection axis, 1 to r , for feature extraction. The dimension of the projection matrix is reduced to $m \times r$.

$$U = [u_1, u_2, \dots, u_r] \quad (24)$$

To Calculate V

Denote $U = I$

$$\text{tr}(\tilde{S}_B) = V^T S_B^V V, \quad \text{tr}(\tilde{S}_W) = V^T S_W^V V \quad (25)$$

where

$$S_B^V = \frac{1}{N} \sum_{i=1}^L n_i (\bar{X}_i - \bar{X})^T (\bar{X}_i - \bar{X})$$

and

$$S_W^V = \frac{1}{N} \sum_{i=1}^L \sum_{X_k \in P_i} (X_k - \bar{X}_i)^T (X_k - \bar{X}_i)$$

Following the characteristic trace $\text{tr}(AB) = \text{tr}(BA)$, when A and B are matrix. From (13), we get

$$J(V) = \frac{V^T S_B^V V}{V^T S_W^V V} \quad (26)$$

The maximum $J(V)$ is used to evaluate the optimal V projection.

$$V_{opt} = \arg \max_V J(V) \quad (27)$$

If S_W^V is nonsingular, the solution of V_{opt} is calculated by the eigenvector and eigenvalue from equation

$$S_B^V V_{opt} = \lambda S_W^V V_{opt} \quad (28)$$

In general the optimal projection axis is from eigenvector $S_W^{-1} S_B^V$ which relates with the maximum eigenvalue from 1 to n . In fact, we need to select a set of projection axis, 1 to c , for feature extraction. The new projection matrix size becomes $n \times c$.

$$V = [v_1, v_2, \dots, v_c] \quad (29)$$

3. Weighted Pairwise Scatter LDA

The between-class covariance matrix is a measure of how distributed the means of each class is from the center. Intuitively, it is better to have a "bigger" value of S_B since it shows that the classes are more spread out in the transformed space, thus it is easier to discriminate them. From this expression, it is not clear how the classes are discriminated from each other pairwise. Next, the weighted pairwise scatter is proposed, and its application to the coupled subspace LDA is described.

3.1 Weighted Pairwise Scatter Approach [17]

To illustrate this point the classes are discriminated from each other pairwise, we consider an example in Fig. 1. Suppose there are four classes in \mathcal{R}^2 of them has the same

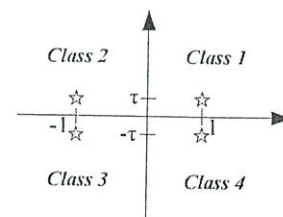


Fig. 1 Illustration of example.

number of feature vectors and equal variance. Let the mean of class 1 be $\mu_1 = (1, \tau)$, class 2 be $\mu_2 = (-1, \tau)$, class 3 be $\mu_3 = (-1, -\tau)$ and class 4 be $\mu_4 = (1, -\tau)$ respectively.

In this case, the between-class scatter matrix will be

$$\frac{1}{4}S_B = \begin{pmatrix} 1 & 0 \\ 0 & \tau^2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

when $\tau \rightarrow 0$, the between-class scatter matrix does not contain any discrimination in the vertical direction. Only discrimination is in the horizontal direction. We can say that the between-class scatter matrix is dominated by the covariance of the class pairs other than class pairs (1, 4) and (2, 3). Obviously, regarding the classification problem, the covariance of class pairs (1, 4) and class pairs (2, 3) are more important than others since they are more confusable, yet the conventional between-class scatter obviously does not capture this information.

From the above example, we can see that if there are some classes much closer relatively as compared to others, the between-class scatter matrix mostly ignores the discriminatory information between the classes that are close to each other.

This discussion leads us to define a general between-class scatter matrix that is equal to the sum of weighted "pairwise scatter" matrices.

$$S_B = \frac{1}{2N} \sum_{k,l=1}^L \omega_{kl} N_k N_l (\mu_k - \mu_l) (\mu_k - \mu_l)^T \quad (30)$$

where $\{\omega_{kl}\}$ is a set of weights. The ω_{kl} is a non-negative weight assigned to class pair $(k - l)$, and it represents how important it is to discriminate class k from class l .

In order to keep enough discriminant information, we need to adjust the weight. A natural candidate is a normalization weight equal to the square of the inverse of the Euclidean distance between class means.

$$\omega_{kl} = \frac{1}{\|\mu_k - \mu_l\|^2} = \frac{1}{(\mu_k - \mu_l)^T (\mu_k - \mu_l)} \quad (31)$$

We weight the classes which have their means closer to each other more than the ones which have means farther apart. In this sense, more confusable classes are weighted more and less confusable classes are weighted less. According to the normalization weight, the between-class scatter matrix is

$$S_{B_{norm}} = \frac{1}{2N} \sum_{k,l=1}^L N_k N_l \frac{(\mu_k - \mu_l) (\mu_k - \mu_l)^T}{(\mu_k - \mu_l)^T (\mu_k - \mu_l)} \quad (32)$$

For the problem defined in Fig. 1, the between-class covariance computed using (32) is now as follows:

$$\frac{1}{4}S_{B_{norm}} = \begin{pmatrix} 1 + \frac{1}{1+\tau^2} & 0 \\ 0 & 1 + \frac{\tau^2}{1+\tau^2} \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

So, using the new between covariance matrix (32), no matter how close the pairs (1, 4) and (2, 3) are, we take their spread into account in computing. This is very desirable

regarding the classification problem.

3.2 The Proposed Method

Then, we can use the new between-class covariance matrix (Weighted Pairwise Scatter) to apply (16) in coupled subspace LDA. The new between-class scatter matrix will be the same result in (33).

$$\begin{aligned} \tilde{S}_B &= \frac{1}{2N} \sum_{k,l=1}^L \omega_{kl} N_k N_l (\bar{Y}_i - \bar{Y}) (\bar{Y}_i - \bar{Y})^T \\ &= \frac{1}{2N} \sum_{k,l=1}^L \omega_{kl} N_k N_l [U^T (\bar{X}_i - \bar{X}) V] [U^T (\bar{X}_i - \bar{X}) V]^T \end{aligned} \quad (33)$$

So,

$$tr(\tilde{S}_B) = tr \left(\frac{1}{2N} \sum_{k,l=1}^L \omega_{kl} N_k N_l U^T (\bar{X}_i - \bar{X}) V V^T (\bar{X}_i - \bar{X})^T U \right) \quad (34)$$

In the WPS-CSLDA method, we can use the feature matrix and optimal projection axes, U and V respectively to reconstruct an image by following steps.

For a given image X , the feature matrix is $Y \in \mathcal{R}^{r \times c}$ and the optimal projection axes are $U \in \mathcal{R}^{m \times r}$ and $V \in \mathcal{R}^{n \times c}$, then it is easy to obtain the reconstructed image of X :

$$\tilde{X} = U Y V^T \quad (35)$$

\tilde{X} *reconstructed image* which has the same size as the image X . If we select $d = n$, we can completely reconstruct the images. But if $d < n$, the reconstructed image \tilde{X} is only an approximation for X .

When we have obtained the projection matrices U and V , projecting the m by n image X onto U and V simultaneously, yielding a r by c matrix Y in Eq. (12). The matrix Y is also called the coefficient matrix in image representation. After projecting each training image X_j ($j = 1, 2, \dots, N$) onto U and V , we obtain the feature matrices Y_j ($j = 1, 2, \dots, N$). Given a test face image X , first use (12) to get the feature matrix Y , then 1-Nearest Neighbor (1NN) classifier is used for classification. Here, the distance between Y and Y_j is defined by

$$d(Y, Y_j) = \|y - y_j\| = \sqrt{\sum_{k=1}^r \sum_{l=1}^c (C^{(k,l)} - C_j^{(k,l)})^2} \quad (36)$$

where $\|y - y_j\|$ denotes the Euclidean distance between the two feature vectors y and y_j .

4. Experimental Results

Our experiments operate with three face databases that are the ORL, AR and YALE database. The ORL face database (<http://www.cam-orl.co.uk>) contains all images from 40 individuals, each provides 10 different images. They are

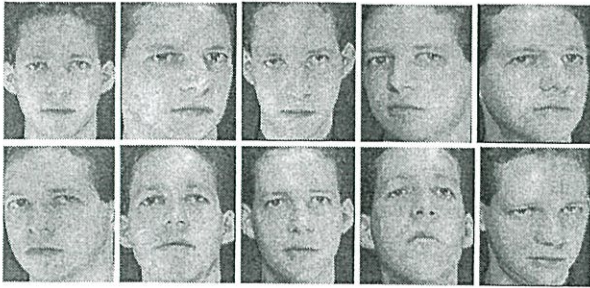


Fig. 2 Face image examples of the ORL face database.



Fig. 3 Face image examples of the AR face database.



Fig. 4 Face image examples of the YALE face database.

grayscale and normalized to a resolution of 112×92 pixels. Classifier in this experiment is 1NN. Example images of one person from the ORL face database are shown in Fig. 2.

The AR face database (Aleix Martinez and Robert Benavente) is composed of a resolution of 768×576 pixels with 24-bit color resolution. It has more than 4,000 images of 126 individuals, 70 males and 56 females. All images were especially taken in front view with different conditions such as facial emotion, facial brightness, wearing glasses or a scarf condition. Some examples of the AR face database are shown in Fig. 3.

In our work, we use 120 individuals, 65 males and 55 females, each divided into 2 groups. The first group is composed of 13 different images as second group was taken at different times (2 weeks). Both of the groups are composed of images which have difference facial aspect like ordinary, smile, angry, scare, right brightness, left brightness and both of brightness. So, we reject the glasses and scarf faces out of work. We then transform all of them to 8 bits grey scale of 120×100 pixels size, because face images of the AR face database have a large number of pixels size.

The Yale face database is grayscale with white background images. All images have a 320×243 pixels size. They were taken from 15 individuals; each provides 11 different images and consists of either male or female. For our work, all images from the Yale database were cropped

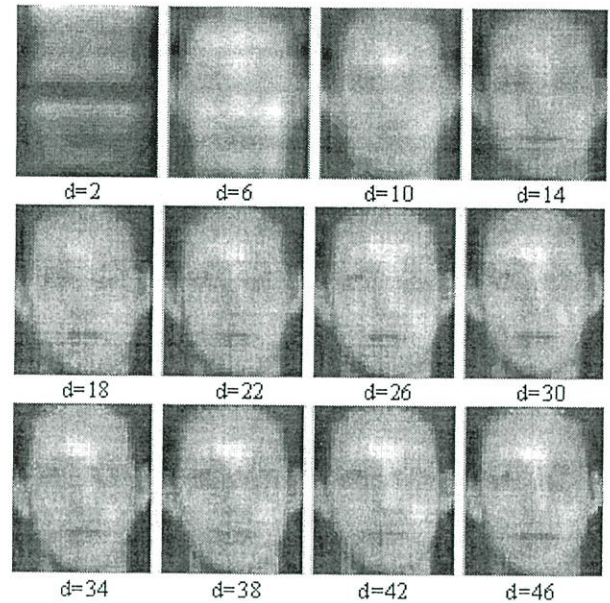


Fig. 5 The reconstructed images based on the WPS-CSLDA.

and resized into 100×80 pixels size, because the original images have a large size. So, they have to be decreased to a size that does not take over time for feature calculation. Figure 4 shows example images of one person of the Yale face database.

Our experiments follow by the databases. The first is the ORL face database testing. We tested to find the reconstruction images, performance of three algorithms that are the WPS-CSLDA (the proposed method), CSLDA [16] and Eigenimages [1], and maximum recognition rate. The second is the AR face database testing. It was tested in three conditions (different time of images, face emotion, and face darkness or brightness) with three algorithms. The third is the YALE face database testing. It is emotion and brightness faces conditions. All the experiments are described in the following sections.

The first experiment is the WPS-CSLDA reconstruction images using the ORL face database. We could project the first image onto the optimal project axis, and use the feature vector sets to reconstruct image. This process uses the WPS-CSLDA algorithm to take the original images. The result of reconstructed images were given out in Fig. 5, the number of dimension (d) used to map and reconstruct the face image. The reconstructed image is more and more similar to the original image when the d increases.

Then, the WPS-CSLDA accuracy is presented in 3D view, as shown in Fig. 6. The number of row and column are varied from 0 to 60.

Next, we compared the WPS-CSLDA, CSLDA and Eigenimages algorithms in one reference axis of recognition rate with only use the diagonal axis. Our experiment takes 5 face images of one person for the training group and 5 for the testing group. All of the images are 200 for training and 200 images for testing. Result of performance comparison is shown in Fig. 7.

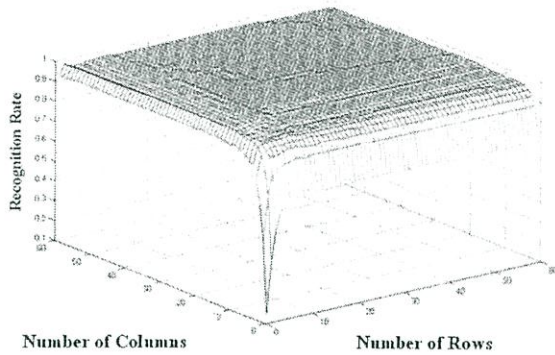


Fig. 6 Accuracy of the WPS-CSLDA algorithm in 3D view.

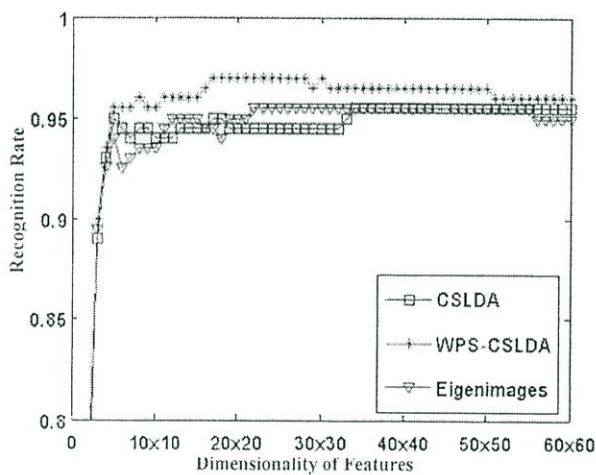


Fig. 7 The ORL database accuracy of three algorithms.

Table 1 Maximum accuracy comparison of the ORL face database.

Training Sample per class		2	3	4	5
WPS-CSLDA	Rate (%)	94.37	96.40	97.35	98.50
	Dimension	12x2	21x2	14x2	13x2
CSLDA	Rate (%)	93.44	95.71	96.25	97.50
	Dimension	14x2	22x3	10x2	13x2
Eigenimages	Rate (%)	91.01	93.33	94.17	95.50
	Dimension	13x2	13x3	21x2	18x2

We also tested the maximum accuracy of each training sample algorithm to compare their performance. For the ORL face database, we show the accuracy of the three algorithms under amount of training sample conditions at 2, 3, 4 and 5 images sample per class for training set and the remaining samples for testing set. Its result is shown in Table 1.

The second experiment takes by using the AR face database. We firstly provided an accuracy performance of the WPS-CSLDA, CSLDA and Eigenimages under the different time condition. Three algorithms were compared with all of 840 images of both training and testing group. The result of accuracy comparison is shown in Fig. 8.

Then, we secondly performed an accuracy performance under the face emotion condition. All of the images are 240 for training and 720 images for testing. The result of the facial emotion condition is shown in Fig. 9.

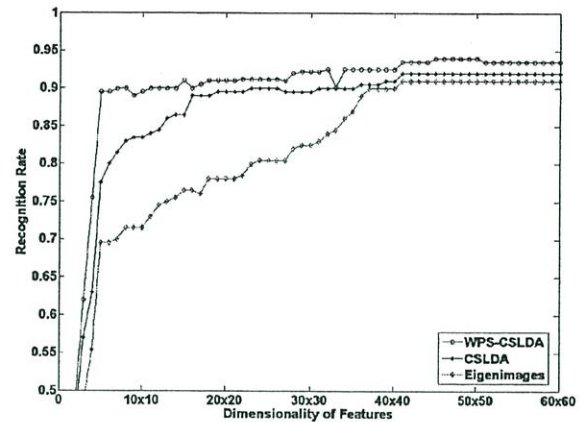


Fig. 8 The AR face database accuracy under different time condition.

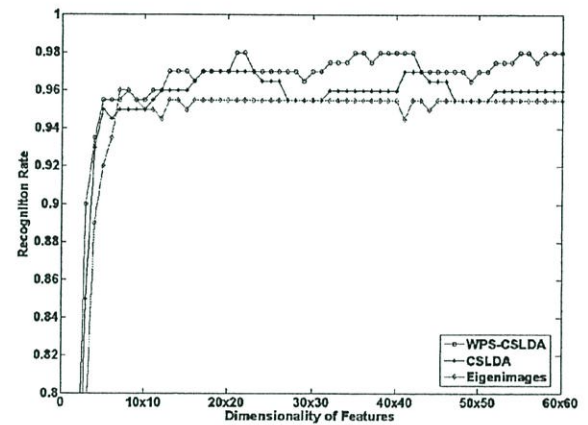


Fig. 9 The AR face database accuracy under facial emotion condition.

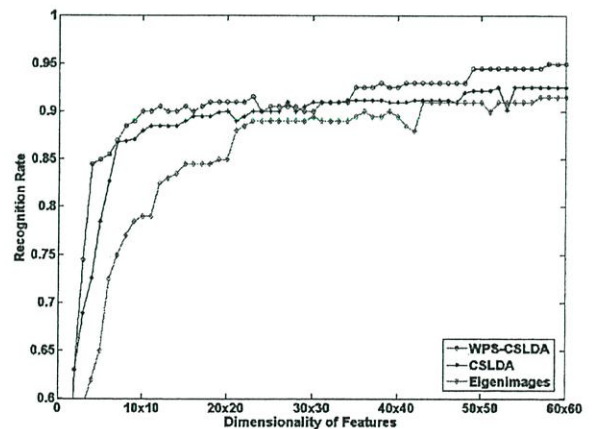


Fig. 10 The AR face database accuracy under facial brightness condition.

Next, the WPS-CSLDA, CSLDA and Eigenimages are compared using face images taken under different of facial brightness condition. All of the images are 240 for training and 720 images for testing. The result of the facial brightness condition is shown in Fig. 10.

We also tested the maximum accuracy of the AR face database under three different conditions by using different

Table 2 Maximum accuracy of the AR face database.

Condition	Methods	Feature Calculation Time (Second)	Maximum Accuracy (%)
Different Time	WPS-CSLDA	0.553	93.750
	CSLDA	0.501	91.400
	Eigenimages	0.482	90.275
Emotion	WPS-CSLDA	0.442	98.900
	CSLDA	0.410	98.120
	Eigenimages	0.378	97.470
Brightness	WPS-CSLDA	0.402	94.650
	CSLDA	0.384	92.630
	Eigenimages	0.327	91.020

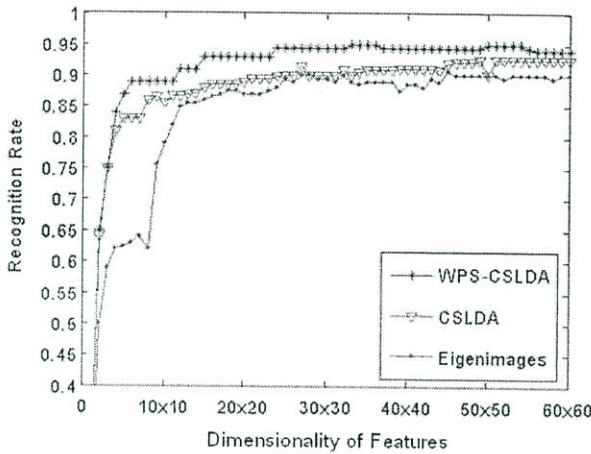


Fig. 11 The Yale face database accuracy of three algorithms.

Table 3 Maximum accuracy of the Yale face database.

Method	Maximum Accuracy (%)	Feature
WPS-CSLDA	95.70	12x2
CSLDA	93.85	13x2
Eigenimages	92.54	14x3

time, facial emotion and facial brightness condition to find feature calculation time and maximum accuracy. These are shown in Table 2.

The third experiment is the YALE face database testing. Our experiment takes 5 face images of one person for the training group and 6 for the testing group. Maximum accuracy comparison of the Yale database is shown in Fig. 11.

Finally, The YALE face database with maximum accuracy comparison is shown in Table 3.

5. Conclusion

In this paper, a new algorithm for face image feature extraction called the WPS-CSLDA has been proposed. The algorithm calculates directly based on original image matrices and uses a weighted pairwise linear discriminant analysis for image feature extraction. Besides, it can also solve confusion data of the inter-class. The recognition rate of the WPS-CSLDA is better than CSLDA and Eigenimages in all experiments, when testing with three face databases, but its computational cost is greater than the CSLDA and Eigenimages.

References

- [1] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherface; recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.19, no.7, pp.711-720, July 1997.
- [2] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit.*, vol.33, no.10, pp.1713-1726, Oct. 2000.
- [3] H. Yu and J. Yang, "A direct LDA algorithm for high dimensional data with application to face recognition," *Pattern Recognit.*, vol.34, no.10, pp.2067-2070, Oct. 2001.
- [4] R. Lotlikar and R. Kothari, "Fractional-step dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.22, no.6, pp.623-627, June 2000.
- [5] M. Loog, R.P.W. Duin, and R. Haeb-Umbach, "Multiclass linear dimension reduction by weighted pairwise Fisher criteria," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.23, no.7, pp.762-766, July 2001.
- [6] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face recognition using LDA-based algorithms," *IEEE Trans. Neural Netw.*, vol.14, no.1, pp.195-200, Jan. 2003.
- [7] J.R. Price and T.F. Gee, "Face recognition using direct, weighted linear discriminant analysis and modular subspaces," *Pattern Recognit.*, vol.38, no.2, pp.209-219, Feb. 2005.
- [8] A.K. Qin, P.N. Suganthan, and M. Loog, "Generalized null space uncorrelated Fisher discriminant analysis for linear dimensionality reduction," *Pattern Recognit.*, vol.39, no.9, pp.1805-1808, Sept. 2006.
- [9] W. Zuo, K. Wang, D. Zhang, and H. Zhang, "Combination of two novel LDA-based methods for face recognition," *Neurocomputing*, vol.70, no.4-6, pp.735-742, Jan. 2007.
- [10] F. Song, D. Zhang, J. Wang, H. Liu, and Q. Tao, "A parameterized direct LDA and its application to face recognition," *Neurocomputing*, vol.71, no.1-3, pp.191-196, Dec. 2007.
- [11] M. Kyperountas, A. Tefas, and I. Pitas, "Weighted piecewise LDA for solving the small sample size problem in face verification," *IEEE Trans. Neural Netw.*, vol.18, no.2, pp.506-519, March 2007.
- [12] J. Yang and J.Y. Yang, "From image vector to matrix: A straightforward image projection technique-IMPCA vs. PCA," *Pattern Recognit.*, vol.35, no.9, pp.1997-1999, Sept. 2002.
- [13] S. Kongsontana and Y. Rangsanseri, "Face recognition using 2DLDA algorithm," *Proc. 8th Inter. Sym. on Signal Processing and Its Applications (ISSPA 2005)*, vol.2, pp.675-678, Aug. 2005.
- [14] D. Zhang, S. Chen, and J. Liu, "Representing image matrices: Eigenimages vs. eigenvectors," *Proc. 2nd Inter. Sym. on Neural Networks (ISNN 2005)*, pp.659-664, May 2005.
- [15] D. Xu, S. Yan, L. Zhang, Z. Liu, and H. Zhang, "Coupled subspace analysis," *Microsoft Research Technical Report, MSR-TR-2004-106*, 2004.
- [16] W. Leelatham, Y. Rangsanseri, and W. Sanayha, "Face recognition using coupled subspace linear discriminant analysis," *Proc. Inter. Conf. on Engineering, Applied Sciences, and Technology (ICEAST 2007)*, pp.258-261, Nov. 2007.
- [17] Y. Li, Y. Gao, and H. Erdogan, "Weighted pairwise scatter to improve linear discriminant analysis," *Proc. 6th Inter. Conf. on Spoken Language Processing (ICSLP2000)*, vol.4, pp.608-611, Oct. 2000.



Waiyawut Sanayha received the B.Ind. degree in Instrumentation Technology from King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, in 1986, M.Eng. degree in Electrical Engineering from Prince of Songkla University, Thailand, in 1997 and M.Sc. degree in Information Technology from Rangsit University, Thailand, in 2003. He is currently working toward the D.Eng. in the Department of Telecommunications Engineering at KMITL. His research interests are in pattern recognition

that arise in character and face recognition.



Yuttapong Rangsanseri received the B.Eng. and the M.Eng. in electrical engineering, both from King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, in 1985 and 1987 respectively; and the Docteur in engineering from Institut National Polytechnique de Grenoble (France) in 1992. He is currently an associate professor of the Department of Telecommunications Engineering at KMITL. His Research interests include image processing, pattern recognition, and multimedia signal

processing.

ETRI⁹² Journal

VOLUME 31 · NUMBER 4 · AUGUST 2009 · ISSN 1225-6463

Relevance-Weighted $(2D)^2$ LDA Image Projection Technique for Face Recognition

Waiyawut Sanayha and Yuttapong Rangsanseri

PP. 438-447

Relevance-Weighted (2D)²LDA Image Projection Technique for Face Recognition

Waiyawut Sanayha and Yuttapong Rangsanseri

In this paper, a novel image projection technique for face recognition application is proposed which is based on linear discriminant analysis (LDA) combined with the relevance-weighted (RW) method. The projection is performed through 2-directional and 2-dimensional LDA, or (2D)²LDA, which simultaneously works in row and column directions to solve the small sample size problem. Moreover, a weighted discriminant hyperplane is used in the between-class scatter matrix, and an RW method is used in the within-class scatter matrix to weigh the information to resolve confusable data in these classes. This technique is called the relevance-weighted (2D)²LDA, or RW(2D)²LDA, which is used for a more accurate discriminant decision than that produced by the conventional LDA or 2DLDA. The proposed technique has been successfully tested on four face databases. Experimental results indicate that the proposed RW(2D)²LDA algorithm is more computationally efficient than the conventional algorithms because it has fewer features and faster times. It can also improve performance and has a maximum recognition rate of over 97%.

Keywords: Linear discriminant analysis (LDA), 2DLDA, 2-directional 2DLDA, (2D)²LDA, face recognition.

I. Introduction

Linear discriminant analysis (LDA) is one of the most popular linear projection techniques. It is a well-known feature extraction and data representation technique which is widely used in the areas of pattern recognition for feature extraction and dimension reduction. It finds the set of the largest discriminant projection vectors which can map high-dimensional samples onto a low-dimensional space. Principal component analysis (PCA) and LDA research (started in 1991 by Turk and Pentland [1]) presented the Eigenfaces method for the linear projection of face images onto a reduced dimension feature space. Belhumeur and others [2] presented a projection method based on Fisher's linear discriminant (FLD) in 1997. From 2000 to 2004, there was much work on the theory of PCA and LDA [3]-[9] which was motivated by the need to solve the small sample problem.

The objective of LDA is to find the optimal projection so that the ratio of the determinants of the between-class and within-class scatter matrices of the projected samples reaches its maximum. However, concatenating 2D matrices into a 1D vector leads to a very high-dimensional image vector, where it is difficult to evaluate the scatter matrices accurately due to its large size and the relatively small number of training samples. Furthermore, the within-class scatter matrix is always singular, making the direct implementation of the LDA algorithm an intractable task. To overcome these problems, a new technique called 2-dimensional LDA (2DLDA) was recently proposed. This method directly computes the eigenvectors of the scatter matrices without matrix-to-vector conversion. Thus, PCA and LDA were developed into the 2-dimensional space methods which are known as 2DPCA and 2DLDA, respectively [10]-[17].

The scatter matrices in 2DLDA are quite small compared to

Manuscript received Nov. 17, 2008; revised June 6, 2009; accepted June 24, 2009.

Waiyawut Sanayha (phone: +66 2326 4242, email: waiyawut.s@egat.co.th) and Yuttapong Rangsanseri (email: rangsanseri@yahoo.com) are with the Department of Telecommunications Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand.
doi:10.4218/etrij.09.0108.0667

the scatter matrices in LDA. The size of the 2DLDA matrix is proportional to the width of the image. 2DLDA evaluates the scatter matrices more accurately and computes the corresponding eigenvectors more efficiently than LDA or PCA. However, the main drawback of 2DLDA is that it needs more coefficients for image representation than the conventional PCA and LDA-based schemes.

In addition to the basic 2DLDA method, a two-directional LDA has been proposed. It works by simultaneously combining 2DLDA applied to the row direction of face images with alternative 2DLDA applied to the column direction of face images. This is called $(2D)^2$ LDA. Similarly, the 2-dimensional version of PCA is known as $(2D)^2$ PCA. These interesting algorithms were first developed in 2005. Zhang and Zhou [18] developed the 2-directional 2DPCA or $(2D)^2$ PCA, achieving the same or even higher recognition accuracy than 2DPCA. The main difference between $(2D)^2$ PCA and existing 2DPCA is that the latter only works in the row direction of face images, while the former works simultaneously in the row and column directions of face images. The main advantage of $(2D)^2$ PCA over 2DPCA is that far fewer coefficients are needed by $(2D)^2$ PCA for face representation and recognition than are needed by 2DPCA. Nagabhushan and others [19] introduced the $(2D)^2$ FLD method which has the advantage of higher recognition rates, a smaller memory requirement, and better computing performance than the standard PCA/2D-PCA/2D-FLD method. The major advantage of the proposed method is that it requires fewer coefficients for object/face image representation than the standard PCA/2D-FLD/2D-PCA because it works simultaneously in both row and column directions. Noushath and others [20] showed that $(2D)^2$ LDA requires fewer coefficients and less computing time for face image representation and recognition than standard PCA, 2DPCA, and 2DLDA methods.

In recent years, LDA research has developed several refinements to improve its performance, such as using the weighting function with the between-class scatter matrix. The relevance-weighted (RW) method is also combined with the within-class scatter matrix. In addition, both weights are used to perform the best recognition process for the LDA and 2DLDA. The research on weighted conventional LDA started by Li and others [21] introduced a weighting factor for each pairwise scatter that enables integration of confusable information into the between-class covariance matrix. There are many possibilities in choosing weighting factors. Li and others considered a few of them depending on the Euclidean and Kullback-Leibler distances between classes when a single Gaussian approximation is used for each class. This is called weighted pairwise scatter linear discriminant analysis (WPS-LDA) transform. Lotlikar and Kothari [22] introduced the

concept of fractional dimensionality and developed an incremental procedure called the fractional-step LDA (F-LDA) as the weighting function. Loog and others [23] introduced a weighted variant of the well-known K-class Fisher criterion associated with the LDA. It can be seen that the LDA weighs contributions of individual class pairs according to the Euclidian distance of the respective class means. An interesting subclass of these criteria is the approximate pairwise accuracy criteria (aPAC). Yu and others [24] redefined the between-class scatter by adding a weighted function according to the between-class distance, which helps to separate the classes as much as possible. At the same time, it projects the between-class scatter into the null space of the within-class scatter that contains the most discriminant information. Lu and others [25] combined the strengths of the direct LDA (D-LDA) [4] and F-LDA approaches in the proposed framework, which will hereafter be referred to as DF-LDA. Also, a weighting function is introduced into the proposed variant of D-LDA, so that a subsequent F-LDA step can be applied to carefully reorient the small sample size (SSS)-free subspace, which results in a set of optimal discriminant features for face representation. Price and Gee [26] proposed a new algorithm as direct weighted LDA or DW-LDA. It combines direct LDA (D-LDA) with weighted pairwise Fisher criteria. Wang and others [27] proposed the weighted two-dimensional maximum margin criterion (W2DMMC), which has an additional weighted parameter β that further broadens the margin. Wang and others [28] developed a multi-block 2DLDA (MB2DLDA) to apply to the sub-images instead of the whole image by weighting the 2DLDA feature of a block; thus, the verification performance is improved.

Another approach in the development of relevance-weighted LDA was introduced in 2005 by Tang and others [29] who incorporated the inter-class relationships as relevance weights into the estimation of the overall within-class scatter matrix. Liang and others [30] proposed a generalizing relevance-weighted LDA or GRW-LDA. When compared with the LDA-GSVD and Fisherfaces, the GRW-LDA can extract more powerful discriminatory features, thereby achieving the best performance using the least features. Chougali and others [31] presented a relevance-weighted LDA and QR decomposition matrix analysis. However, all algorithms presented in this paper are linear methods. Since facial variations are mostly non-linear, LDA, LDA/QR and RW-LDA/QR projections only provide suboptimal solutions. Jarchi and Boostani [32] proposed a variant LDA method for the multi-class problem which redefined the between-class and within-class scatter matrices by incorporating a weight function into each of them.

Our research indicates that LDA has in fact been improved in the $(2D)^2$ LDA method. In other words, the RW method can

be used to achieve better face recognition performance than that of LDA or 2DLDA. Therefore, it is best to combine the $(2D)^2$ LDA method with RW method in order to improve its discrimination performance. This algorithm will be hereafter referred to as the relevance-weighted $(2D)^2$ LDA (RW $(2D)^2$ LDA).

This paper is organized as follows. Section II presents background of the LDA and its development starting from the conventional LDA, the 2-dimensional LDA (2DLDA), an alternative 2DLDA, the 2-dimensional and 2-directional LDA, and application with the pairwise scatter approaches. Section III describes the proposed algorithm, which uses weighted and RW methods with the $(2D)^2$ LDA method. Experimental results and discussion are presented in section IV. The conclusion is given in section V. Finally, future work is proposed in section VI.

II. LDA Background

By using the set of projection vectors determined by the LDA as the projection axes, all projected samples form the maximum between-class scatter and the minimum within-class scatter simultaneously in the projective feature space.

1. Conventional LDA

When using appearance-based methods, the 2D face image matrices must be first transformed into 1D image vectors. The LDA maps high dimensional samples of the projection vectors onto a low-dimensional space and computes eigenvectors in the underlying space that give the best discrimination among classes. More formally, given a number of independent features relative to which the data is described, LDA creates a linear combination of these which yields the largest mean differences between the desired classes. Mathematically speaking, for all the samples of all classes, two measures are defined. One measure is called the within-class scatter matrix, as given by

$$S_w = \sum_{j=1}^C \sum_{i=1}^{N_j} (A_i^j - \mu_j)(A_i^j - \mu_j)^T, \quad (1)$$

where A_i^j is the i -th sample of class j , μ_j is the mean of class j , C is the number of classes, and N_j is the number of samples in class j . The other measure is called the between-class scatter matrix, as given by

$$S_b = \sum_{j=1}^C (\mu_j - \mu)(\mu_j - \mu)^T, \quad (2)$$

where μ represents the mean of all classes.

Once S_b and S_w are computed, we compute the optimal projection axes, denoted by X , so that the total scatter of the projected samples of the training images is maximized. To

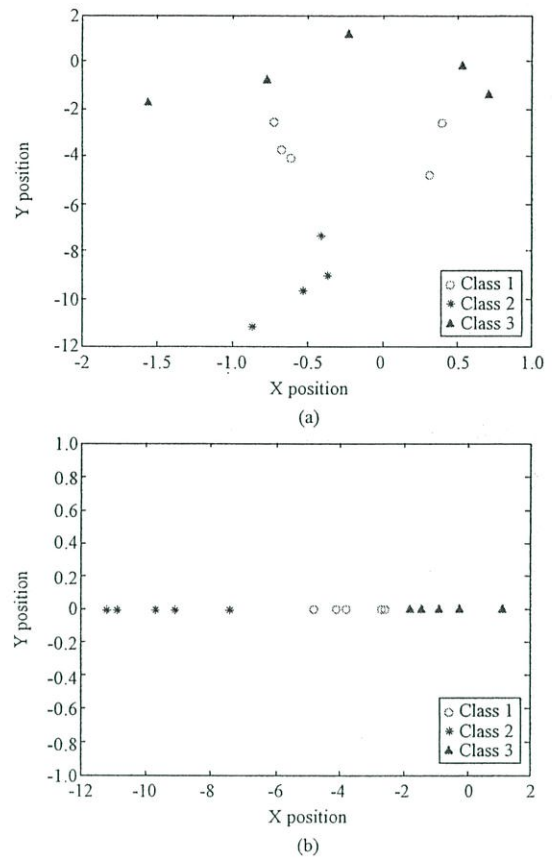


Fig. 1. LDA technique results: (a) data samples and (b) result of 1D projection.

maximize the total scatter of the projected samples we used Fisher's criterion:

$$J(X) = \frac{X^T S_b X}{X^T S_w X}. \quad (3)$$

The maximized $J(X)$ can be obtained by applying an eigenvector corresponding to the maximum eigenvalues of $S_w^{-1}S_b$. This is the optimal projection axis. It is usually not enough to have only one optimal projection axis, and we compute q projection axes, that is, X_1, X_2, \dots, X_q , which are eigenvectors corresponding to the first q largest eigenvalues of $S_w^{-1}S_b$.

To show the principle of the LDA algorithm, example results of the LDA projection technique are shown in Fig. 1. Example data of three classes is shown in Fig. 1(a), and 1D projection results using the LDA method are shown in Fig. 1(b).

2. Two-Dimensional LDA (2DLDA)

The LDA has two drawbacks when directly applied to the original input space. First, some non-face information such as image background data is misclassified when the face of the

same subject is presented on different backgrounds. Secondly, when the SSS problem occurs, the within-class scatter matrix is singular. This is the so-called singularity problem. Projecting the high-dimensional input space into low-dimensional subspace via PCA can solve these LDA problems. Nevertheless, the spatial structure information is still lost. To overcome this drawback, the 2-dimensional LDA is based on 2D matrices rather than 1D vectors. This means that the image matrix does not need to be converted into a vector. As a result, the 2DLDA has two advantages: it is easier to evaluate the covariance matrix accurately, and it has lower time-consumption.

Let A_i^j be an image of size $a \times b$ representing the i -th sample in the j -th class. The between-class scatter matrix G_b and within-class scatter matrix G_w are computed as

$$G_b = \sum_{j=1}^C (\mu_j - \mu)(\mu_j - \mu)^T, \quad (4)$$

$$G_w = \sum_{j=1}^C \sum_{i=1}^{N_j} (A_i^j - \mu_j)(A_i^j - \mu_j)^T. \quad (5)$$

Once G_b and G_w are computed, we compute the optimal projection axes, denoted by X , so that the total scatter of the projected samples of the training images is maximized. To maximize the total scatter of the projected images we used Fisher's criterion as follows:

$$J(X) = \frac{X^T G_b X}{X^T G_w X}. \quad (6)$$

Thus, the eigenvectors of the final covariance matrix $G_w^{-1}G_b$ are computed, and then q eigenvectors corresponding to the first q largest eigenvalues of $G_w^{-1}G_b$ are chosen. Thus, the dimension of the optimal projection axes X is $b \times q$.

Projection of a training image onto these optimal projection axes results in a feature matrix of the respective training image. That is, we define the feature matrix of A_i^j as

$$Z_i^j = A_i^j X, \quad (7)$$

where Z_i^j is the feature matrix of dimension $a \times q$.

The proposed 2DLDA works in the row-wise direction as the image covariance matrix G_w is obtained by the outer products of the row vectors of the training images.

Training images are subsequently projected onto the optimal projection axes, and their dimension is reduced. During the query phase, query images are also projected onto the optimal projection axes to reduce their dimensionality, and they are subjected to a Euclidean nearest neighbors classifier to contrast them with the projected training images. The class label of the training images which is nearest to the query is retrieved as their class label. In this case, the class of the nearest training

face to that of the query face image is identified as its class.

3. Alternative 2DLDA

Equation (5) reveals that the scatter matrix G_w can be obtained from the outer products of row vectors of images, assuming that the training images have a zero mean. For this reason, we claim that the original 2DLDA works in the row direction of images. Apparently, a natural extension is to use the outer product between column vectors of images to construct G_b and G_w . To devise an alternative 2DLDA, we propose that the between-class scatter matrix H_b and the within-class scatter matrix H_w be computed as

$$H_b = \sum_{j=1}^C (\mu_j - \mu)(\mu_j - \mu)^T, \quad (8)$$

$$H_w = \sum_{j=1}^C \sum_{i=1}^{N_j} (A_i^j - \mu_j)(A_i^j - \mu_j)^T. \quad (9)$$

It can be observed that H_b and H_w in (8) and (9) are obtained in this new formulation as outer products of column vectors, unlike G_b and G_w in (4) and (5) in the case of the original 2DLDA. Using these two scatter matrices, which are similar to the original 2DLDA, in this proposed model we also find the optimal projection axes W ($m \times q$) so that the total scatter of the projected samples is maximized using the same Fisher's criterion given by

$$J(W) = \frac{WH_b W^T}{WH_w W^T}. \quad (10)$$

Thus, the eigenvectors of $H_w^{-1}H_b$ are computed, and then q eigenvectors corresponding to the first q largest eigenvalues of $H_w^{-1}H_b$ are chosen. Finally, projection of a training image onto these optimal projection axes results in a feature matrix of the respective training image. That is, if Z_i^j represents the feature matrix of A_i^j , then

$$Z_i^j = W^T A_i^j. \quad (11)$$

Equation (9) reveals that the image covariance matrix H_w can be obtained from the outer products of the column vectors of the training images, assuming that they have a zero mean. Therefore, the proposed alternative 2DLDA works in the column direction of images. An illustration of the 2DLDA projection technique is shown in Fig. 2.

4. (2D)²LDA

Let X denote the $n \times d$ optimal projection matrix obtained in the original 2DLDA method as explained in section II.2, and let W denote the $m \times q$ matrix obtained by an alternative

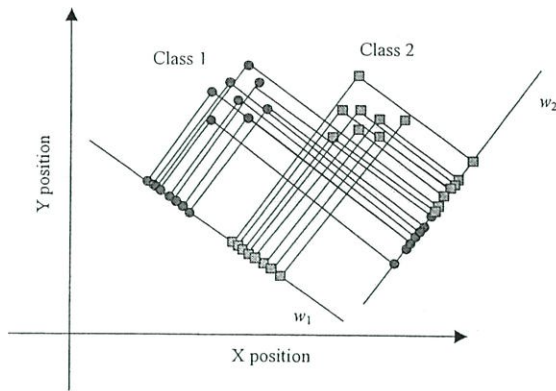


Fig. 2. 2DLDA projection technique.

2DLDA method as explained in section II.3. For $(2D)^2$ LDA method, each training image A_i^j is projected onto both X and W simultaneously to obtain the respective feature matrix F_i^j , which is of dimensions $q \times d$ as follows:

$$F_i^j = W^T A_i^j X. \quad (12)$$

The matrix F is also called the coefficient matrix in image representation. When used for face recognition, the matrix F is also called the feature matrix. After projecting each training image A_i^j onto X and W , the feature matrices F_i^j can be obtained.

In a data set, distance or similarity relationships between pairs of classes is the important information for classification. Distance or similarity relationships, usually acquired through statistical approaches such as Euclidean distance, Mahalanobis distance, and so on, reflect how well two classes are separated in the feature space. In multi-class LDA, the relationships between pairs of classes are likely to be different from one pair to another. The classes that are closer to each other are potentially more confusing, and they should be given more attention during the feature extraction stage. The weighted LDA works by using the pairwise scatter matrix approach which only points at the between-class scatter matrix. Generally, the S_b of this pairwise scatter is computed by

$$S_b = \sum_{j=1}^{C-1} \sum_{k=j+1}^C (\mu_j - \mu_k)(\mu_j - \mu_k)^T. \quad (13)$$

Therefore, the 2DLDA and an alternative 2DLDA method can rewrite the between-class scatter matrix (4) and (8) as follows:

$$G_b = \frac{1}{N} \sum_{j=1}^{C-1} \sum_{k=j+1}^C N_j N_k (\mu_j - \mu_k)(\mu_j - \mu_k)^T, \quad (14)$$

$$H_b = \frac{1}{N} \sum_{j=1}^{C-1} \sum_{k=j+1}^C N_j N_k (\mu_j - \mu_k)(\mu_j - \mu_k)^T. \quad (15)$$

Equation (14) is the between-class scatter matrix of 2DLDA which essentially works in the row-direction of images, and (15) is the between-class scatter matrix of an alternative 2DLDA which works in the column direction of images.

The within-class scatter matrix is still not changed because the pairwise scatter approach analyzes how the classes are discriminated from each other pairwise. Thus, it does not measure discrimination inside the within-class scatter matrix. However, we can rewrite the within-class scatter matrices (5) and (9) as

$$G_w = \frac{1}{N} \sum_{j=1}^C \sum_{i=1}^{N_j} N_j (A_i^j - \mu_j)^T (A_i^j - \mu_j), \quad (16)$$

$$H_w = \frac{1}{N} \sum_{j=1}^C \sum_{i=1}^{N_j} N_j (A_i^j - \mu_j)(A_i^j - \mu_j)^T. \quad (17)$$

Similar to (14) and (15), (16) is the within-class scatter matrix of 2DLDA which essentially works in the row-direction of images, and (17) is the within-class scatter matrix of an alternative 2DLDA which works in the column direction of images.

III. Proposed Algorithm

1. Weighted $(2D)^2$ LDA

Because (14) and (15) are not directly related to classification accuracy and focus equally on every pair of classes, the outlier classes may negatively influence the estimation of the overall between-class covariance matrices G_b and H_b . Therefore, Loog and others [23] proposed an extended criterion named the *approximate pairwise accuracy criterion* (aPAC) which replaces (14) and (15) with

$$G_b = \frac{1}{N} \sum_{j=1}^{C-1} \sum_{k=j+1}^C L_{jk} N_j N_k (\mu_j - \mu_k)(\mu_j - \mu_k)^T, \quad (18)$$

$$H_b = \frac{1}{N} \sum_{j=1}^{C-1} \sum_{k=j+1}^C L_{jk} N_j N_k (\mu_j - \mu_k)(\mu_j - \mu_k)^T, \quad (19)$$

where the weights L_{jk} are usually estimated based on relationships between classes j and k .

In order to keep enough discriminant information, we need to adjust the weights. A natural candidate is a normalization weight equal to the square of the inverse of the Euclidean distance between class means:

$$L_{jk} = \frac{1}{\|\mu_j - \mu_k\|^2} = \frac{1}{(\mu_j - \mu_k)^T (\mu_j - \mu_k)}. \quad (20)$$

Here, L_{jk} is defined as the dissimilarity between class j and k or how well classes j and k are separated in the original space.

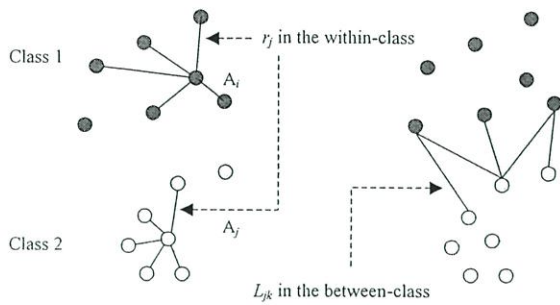


Fig. 3. Adjacency relationships in weighted and relevance-weighted calculation.

2. Relevance-Weighted (2D)²LDA

We can see that in addition to assigning different considerations to classes when estimating the between-class covariance matrix, a weighting scheme should also be employed when estimating the within-class covariance matrix. To reduce the influence of outlier classes, (16) and (17) are modified to give

$$G_w = \frac{1}{N} \sum_{j=1}^c \sum_{i=1}^{N_j} r_j N_j (A_i^j - \mu_j)^T (A_i^j - \mu_j), \quad (21)$$

$$H_w = \frac{1}{N} \sum_{j=1}^c \sum_{i=1}^{N_j} r_j N_j (A_i^j - \mu_j)(A_i^j - \mu_j)^T, \quad (22)$$

where r_j is the relevance-based weight. By integrating r_j in (16) and (17), we intend to ensure that if class i is an outlier class, it only has a slight influence on the estimated G_w and H_w . This is reasonable because if one class is well separated from the other classes in the data set, the within-class covariance matrix of this class in the new space is compact, and it will not have much influence on the classification.

To calculate a class's separation from other classes, a straightforward weighting function is defined:

$$r_j = \sum_{j \neq k} \frac{1}{L_{jk}}. \quad (23)$$

We normalize r_j so that the largest weight is 1. Although several dissimilarity measures have been proposed in the past, it is impossible to choose one of them as the best measure independent of the data set.

We weight the classes which have their means closer to each other more heavily than those that have means further apart. In this sense, more confusable classes are weighted more heavily, and less confusable classes are weighted more lightly. The weighted between-class relationships and the relevance-weighted within-class relationships are shown in Fig. 3.

3. Classification

Given a test face image A , we first use (12) to get the feature

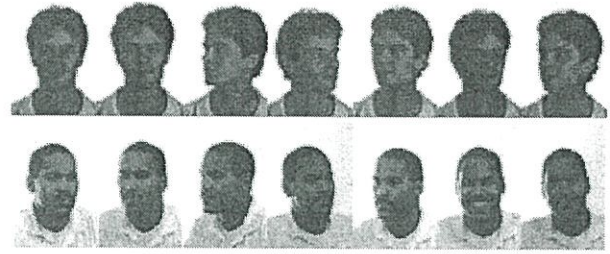


Fig. 4. Sample images from the FERET database.

matrix F_i^j . Then, we use a nearest neighbor classifier for classification. Here, the distance between $F1_i^j$ and $F2_i^j$ is defined by

$$\|F1_i^j - F2_i^j\| = \sqrt{\sum_{i=1}^q \sum_{j=1}^d (F1_i^j - F2_i^j)^2}. \quad (24)$$

IV. Experimental Results and Discussion

In our work, we have used four face databases, namely, the Facial Recognition Technology Database (FERET) [33]; the CMU Pose, Illumination, and Expression (PIE) Database [34]; the ORL Face Database [35]; and the YALE Face Database [36]. All of our experiments were carried out on a PC with an Intel Core2Duo CPU E6750 at 2.66 GHz with 1.96 GB of RAM memory under a MATLAB R2007b platform.

1. RW(2D)²LDA Performance Test

We first used the FERET face database, which is designed to advance the state of the art in face recognition, with the collected images directly supporting both algorithm development and the FERET evaluation tests. The database is divided into a development set which is provided to researchers and a set of sequestered images for testing. The dataset tested included 2,413 still facial images of 856 individuals. Some sample FERET face images are shown in Fig. 4.

The results of our first accuracy test of the RW(2D)²LDA algorithm are presented in a 3D view. We tested the recognition rate in row and column feature dimensions. The results of this performance test are shown in Fig. 5.

The number of feature dimensions was varied from 0 to 60. The RW(2D)²LDA accuracy is over 70% when tested with 2-3 dimensions and over 97% at 5-8 dimensions. It becomes less steady when there are over 10 feature dimensions.

2. Pose, Illumination, and Expression Condition Test

It is not easy to compare all three algorithms within a 3D view graph because the lines of the graph would be

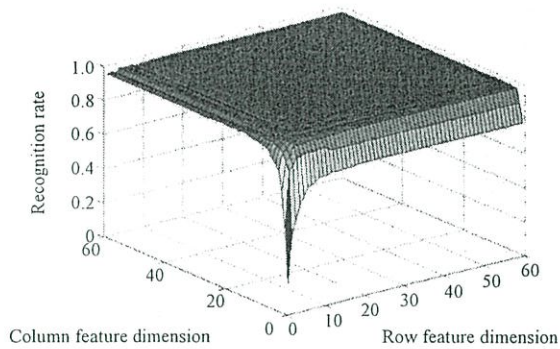


Fig. 5. RW(2D)²LDA accuracy in 3D view.



Fig. 6. Sample images from the PIE database.

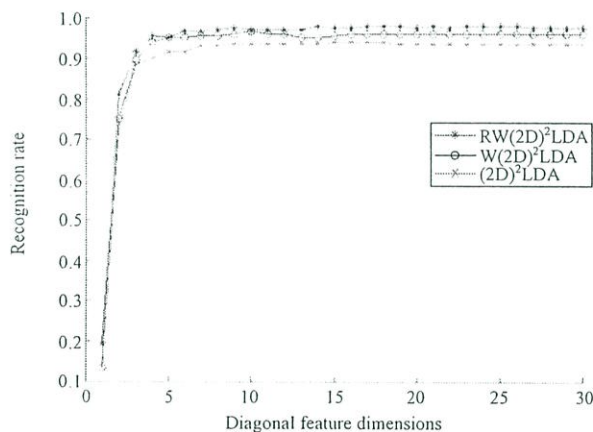


Fig. 7. Result of pose, illumination, and expression test in 2D.

complicated. Therefore, it is more suitable to use a 2D view graph for accurate comparisons between these algorithms. For this reason, we conducted a second experiment to show the performance improvement of the proposed method (RW(2D)²LDA and W(2D)²LDA) compared with the previous method ((2D)²LDA).

We used the PIE face database for pose, illumination, and expression condition tests. This database consists of 41,368 images of 68 individuals. The PIE database was able to image each person under 13 different poses, 43 different illumination conditions, and with 4 different expressions. Some sample PIE face images are shown in Fig. 6.

As seen in Fig. 7, the results of this test demonstrate that, for

Table 1. Result of pose, illumination, and expression test.

Feature dimensions	Methods		
	(2D) ² LDA	W(2D) ² LDA	RW(2D) ² LDA
1	14.50	13.50	19.50
2	72.50	75.00	81.00
3	88.50	89.50	91.50
4	90.00	94.00	95.50
5	91.50	95.00	95.00
6	91.50	95.00	96.50
7	93.00	95.50	96.50
8	93.00	95.50	97.00
9	93.50	96.00	97.50
10	93.50	96.50	97.00

the accuracy of training groups, the RW(2D)²LDA and W(2D)²LDA methods can improve the performance of the previous (2D)²LDA method.

The interesting aspect of the RW(2D)²LDA and W(2D)²LDA methodologies is that they are more computationally efficient than the original formulation of (2D)²LDA because they have fewer features and are more accurate.

Table 1 reveals the recognition accuracy obtained by three methods when the number of diagonal feature dimensions varies from 1 to 10. The W(2D)²LDA and RW(2D)²LDA methods can increase the performance by up to 2.10% and 8.11%, respectively, when compare with the (2D)²LDA.

3. Facial Scaling and Rotation Condition Test

Next, we experimentally evaluated our proposed RW(2D)²LDA, W(2D)²LDA, and (2D)²LDA methods on well-known face databases. The first of these is the ORL database. It is used to test the performance of face recognition algorithms under the condition of minor variation of scaling and rotation.

The ORL database contains 400 images of 40 individuals. These images were captured at different times and have different variations including expression (open or closed eyes, smiling or not smiling) and facial details (glasses or no glasses). The images were taken with a tolerance for some tilting and rotation of the face up to 20 degrees. All images are in grayscale and normalized to the resolution of pixels and histogram equilibrium in the preprocessing step. Some sample images from the ORL database are shown in Fig. 8. In our experiments, we split the whole database into two equal parts. The first five images of each class were used for training, and



Fig. 8. Sample images from the ORL database.

Table 2. Result of scaling and rotation.

Methods	Top recognition	Dimensions	Running time
$(2D)^2LDA$	98.50	8×8	4.14
$W(2D)^2LDA$	98.75	7×8	5.26
$RW(2D)^2LDA$	99.15	6×7	7.85



Fig. 9. Sample images from the YALE database.

the rest of the images were used for testing.

Table 2 compares the three methods in term of their top recognition accuracy and gives the corresponding dimensions of the feature matrices and running times. It can be found that the top recognition accuracy of the original $(2D)^2LDA$ method is comparable to that of the other methods. This table also reveals that the top recognition accuracy of the proposed $RW(2D)^2LDA$ and $W(2D)^2LDA$ methods is significantly higher than that of the existing $(2D)^2LDA$ method despite the $RW(2D)^2LDA$ having a reduced feature matrix. Conversely, the $RW(2D)^2LDA$ and $W(2D)^2LDA$ methods consume more running time than the $(2D)^2LDA$.

4. Facial Expression and Lighting Condition Test

Finally, we compared the proposed $RW(2D)^2LDA$ and $W(2D)^2LDA$ methods with the $(2D)^2LDA$ using the YALE face database. We tested the performance under the condition of minor variations of facial expression and lighting conditions. The images are in grayscale with white background and are 320×243-pixels in size. We used a total of 165 images of 15 individuals in GIF format. There are 11 images per subject, and each image shows a different facial expression or lighting configuration. Some sample images from the YALE database

Table 3. Result of facial expression and lighting conditions.

Methods	Number of training samples per class			
	2	4	6	8
$(2D)^2LDA$	86.25	93.60	94.67	96.00
$W(2D)^2LDA$	87.36	95.00	96.50	98.45
$RW(2D)^2LDA$	89.57	96.48	97.20	99.40

are shown in Fig. 9.

Table 3 shows the top recognition accuracy under the condition of minor variation of facial expression and lighting conditions obtained by three methods for varying numbers of training samples. The proposed $RW(2D)^2LDA$ and $W(2D)^2LDA$ methods are comparable to the $(2D)^2LDA$ method in terms of recognition accuracy.

V. Conclusion

This paper examined two efficient face recognition methods called the weighted $(2D)^2LDA$ or $W(2D)^2LDA$, and the relevance-weighted $(2D)^2LDA$ or $RW(2D)^2LDA$ incorporating weighted outlier class relationships into the estimation of the overall between-class scatter matrix. In the same way, relevance-weighted inner class relationships are incorporated into the overall within-class scatter matrix to improve the performance of the $(2D)^2LDA$ method. The experimental results have shown that our algorithms can be used under various conditions and can achieve better performance accuracy than existing methods.

VI. Future Work

The proposed $RW(2D)^2LDA$ method applies the weighting function in the within-class scatter matrix, but the $W(2D)^2LDA$ method applies the weighting function in the between-class scatter matrix. It would be good to combine the proposed methods with other methods to improve the performance of the $(2D)^2LDA$. However, the effectiveness of kernel-based methods has been reported recently. It may be interesting to focus on the problem of kernelizing an existing supervised Mahalanobis distance, such as the *kernelizing Mahalanobis distance learning algorithm*. Therefore, we plan to apply neighborhood component analysis, large margin nearest neighbors, and discriminant neighborhood embedding to this algorithm. An alternative kernelization framework will also be used in the between and within-class scatter matrix in our future work.

References

- [1] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, 1991, pp. 71-86.
- [2] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, 1997, pp. 711-720.
- [3] L. Chen et al., "A New LDA-Based Face Recognition System Which Can Solve the Small Sample Size Problem," *Pattern Recognition*, vol. 33, no. 10, 2000, pp. 1713-1726.
- [4] H. Yu and J. Yang, "A Direct LDA Algorithm for High-Dimensional Data—with Application to Face Recognition," *Pattern Recognition*, vol. 34, no. 10, 2001, pp. 2067-2070.
- [5] A.M. Martinez and A.C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, 2001, pp. 228-233.
- [6] H-C. Kim, D. Kim, and S.W. Bang, "Face Recognition Using LDA Mixture Model," *Pattern Recognition Letters*, vol. 24, no. 15, 2003, pp. 2815-2821.
- [7] M. Ordowski and G. Meyer, "Geometric Linear Discriminant Analysis for Pattern Recognition," *Pattern Recognition Letters*, vol. 37, no. 3, 2004, pp. 421-428.
- [8] X.-Y. Jing, D. Zhang, and Y.-F. Yao, "Improvements on the Linear Discriminant Technique with Application to Face Recognition," *Pattern Recognition Letters*, vol. 24, no. 15, 2003, pp. 2695-2701.
- [9] F. Song et al., "A Parameterized Direct LDA and Its Application to Face Recognition," *Neurocomputing*, vol. 24, 2003, pp. 2695-2701.
- [10] J. Yang et al., "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, 2004, pp. 131-137.
- [11] J. Meng and W. Zhang, "Volume Measure in 2DPCA-Based Face Recognition," *Pattern Recognition Letters*, vol. 28, no. 10, 2007, pp. 1203-1208.
- [12] Y. Wen and P. Shi, "Image PCA: A New Approach for Face Recognition," *Proc. IEEE Inter. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, 2007, pp. I-1241-1244.
- [13] H. Xiong, M.N.S. Swamy, and M.O. Ahmad, "Two-Dimensional FLD for Face Recognition," *Pattern Recognition*, vol. 38, no. 7, 2005, pp. 1121-1124.
- [14] J. Yang et al., "Two-Dimensional Discriminant Transform for Face Recognition," *Pattern Recognition*, vol. 38, no. 7, 2005, pp. 1125-1129.
- [15] M. Li and B. Yuan, "2D-LDA: A Statistical Linear Discriminant Analysis for Image Matrix," *Pattern Recognition Letters*, vol. 26, no. 5, 2005, pp. 527-532.
- [16] S. Yan et al., "Multilinear Discriminant Analysis for Face Recognition," *IEEE Trans. Image Processing*, vol. 16, 2007, pp. 212-220.
- [17] D.S. Guru and T.N. Vikram, "2D Pairwise FLD: A Robust Methodology for Face Recognition," *IEEE Workshop on Automatic Identification Advanced Technologies*, 2007, pp. 99-102.
- [18] D. Zhang and Z. Zhou, "(2D)²PCA: Two-Directional Two-Dimensional PCA for Efficient Face Representation and Recognition," *Neurocomputing*, vol. 69, 2005, pp. 224-231.
- [19] P. Nagabhushan, D.S. Guru, and B.H. Shekar, "(2D)²FLD: An Efficient Approach for Appearance Based Object Recognition," *Neurocomputing*, vol. 69, 2006, pp. 934-940.
- [20] S. Noushath, G. Hemantha, and P. Shivakumara, "(2D)²LDA: An Efficient Approach for Face Recognition," *Pattern Recognition*, vol. 39, no. 7, 2006, pp. 1396-1400.
- [21] Y. Li, Y. Gao, and H. Erdogan, "Weighted Pairwise Scatter to Improve Linear Discriminant Analysis," *Proc. 6th Inter. Conf. Spoken Language Processing (ICSLP)*, vol. 4, 2000, pp. 608-611.
- [22] R. Lotlikar and R. Kothari, "Fractional-Step Dimensionality Reduction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, 2000, pp. 623-627.
- [23] M. Loog, R.P.W. Duin, and R. Haeb-Umbach, "Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 7, 2001, pp. 762-766.
- [24] B. Yu, L. Jin, and P. Chen, "A New LDA-Based Method for Face Recognition," *Proc. IEEE Int. Conf. Pattern Recognition*, vol. 1, 2002, pp. 168-171.
- [25] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face Recognition Using LDA-Based Algorithms," *IEEE Trans. Neural Networks*, vol. 14, no. 1, 2003, pp. 195-200.
- [26] J.R. Price and T.F. Gee, "Face Recognition Using Direct, Weighted Linear Discriminant Analysis and Modular Subspaces," *Pattern Recognition*, vol. 38, no. 2, 2005, pp. 209-219.
- [27] H. Wang, S. Chen, and Z. Hu, "Image Recognition Using Weighted Two-Dimensional Maximum Margin Criterion," *Proc. 3rd Int. Conf. Natural Computation (ICNC)*, vol. 1, 2007, pp. 582-586.
- [28] J. Wang, E. Sung, and W. Yau, "Weighting Multi-Block Two-Dimensional Linear Discriminant Features for Face Verification," *Proc. 6th Int. Conf. on Information, Communications and Signal Processing (ICICS)*, 2007, pp. 1-5.
- [29] E.K. Tang et al., "Linear Dimensionality Reduction Using Relevance Weighted LDA," *Pattern Recognition*, vol. 38, no. 4, 2005, pp. 485-493.
- [30] Y. Liang et al., "Generalizing Relevance Weighted LDA," *Pattern Recognition*, vol. 38, no. 11, 2005, pp. 2217-2219.
- [31] K. Chougali, M. Jedra, and N. Zahid, "Face Recognition Using Relevance Weighted LDA with QR Decomposition," *J.*

Graphics, Vision and Image Processing (GVIP), vol. 6, no. 2, 2006, pp. 27-31.

- [32] D. Jarchi and R. Boostani, "A New Weighted LDA Method in Comparison to Some Versions of LDA," *Proc. World Academy of Science, Engineering and Technology (WASET)*, vol. 18, 2006, pp. 233-238.
- [33] P.J. Phillips et al., "The FERET Evaluation Methodology for Face Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, 2000, pp. 1090-1104.
- [34] The CMU Technical Report CMU-RI-TR-01-02, "The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces."
- [35] F.S. Samaria and A.C. Harter, "Parameterisation of a Stochastic Model for Human Face Identification," *Proc. 2nd IEEE Workshop on Applications of Computer Vision*, 1994, pp. 138-142.
- [36] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, 2001, pp. 643-660.



Waiyawut Sanayha received the BEng degree in instrumentation technology from King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, in 1986, the MEng degree in electrical engineering from Prince of Songkla University, Thailand, in 1997, and the MSc degree in information technology from Rangsit University, Thailand, in 2003. He is currently working toward the DEng with the Department of Telecommunications Engineering at KMITL. His research interests include pattern recognition in the areas of character and face recognition.



Yuttapong Rangsanseri received the BEng and the MEng in electrical engineering from King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, in 1985 and 1987, respectively, and the PhD in engineering from the Institute National Polytechnique de Grenoble, France, in 1992. He is currently an associate professor with the Department of Telecommunications Engineering at KMITL. His research interests include image processing, pattern recognition, and multimedia signal processing.

ประวัติผู้เขียน

ชื่อ-นามสกุล	นายวิฑูติ เสน่หา
วัน เดือน ปีเกิด	13 กุมภาพันธ์ พ.ศ.2505
ที่อยู่	101/2 หมู่ที่ 5 ซอย บางกรวย-ไทรน้อย 9 ถนน บางกรวย-ไทรน้อย ตำบล บางกรวย อำเภอบางกรวย จังหวัด นนทบุรี รหัสไปรษณีย์ 11130 โทร. 02-447-0303
ประวัติการศึกษา	2526 ประกาศนียบัตรวิชาชีพชั้นสูง ช่างอิเล็กทรอนิกส์ วิทยาลัยเอเชียอาคเนย์ 2530 ปริญญาตรี อดสาหกรรมศาสตรบัณฑิต สาขาเทคโนโลยีการวัดคุมทางอุตสาหกรรม สถาบันเทคโนโลยีพระจอมเกล้า เจ้าคุณทหารลาดกระบัง 2540 ปริญญาโท วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า มหาวิทยาลัยสงขลานครินทร์ 2546 ปริญญาโท วิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ มหาวิทยาลัยรังสิต
ประวัติการทำงาน	2526 เริ่มเข้าทำงานบรรจุเป็นพนักงาน ดำรงตำแหน่งช่างอันดับสี่ แผนกโทรศัพท์และอินเตอร์คอม กองระบบสื่อสารตามสาย ส่วนระบบ สื่อสาร การไฟฟ้าฝ่ายผลิตแห่งประเทศไทย (กฟผ.) 2538 ดำรงตำแหน่งหัวหน้าแผนกบำรุงรักษาระบบสื่อสาร 1 กองบำรุงรักษาสื่อสาร ฝ่ายปฏิบัติการภาคใต้ กฟผ. 2542 ดำรงตำแหน่งหัวหน้าแผนกพัฒนาระบบอินเทอร์เน็ต กองศูนย์อินเทอร์เน็ต สำนักงานวิจัยและพัฒนา กฟผ. 2545 ดำรงตำแหน่งหัวหน้าแผนกบริการผู้ใช้งานระบบอินเทอร์เน็ต กองศูนย์อินเทอร์เน็ต ฝ่ายเทคโนโลยีสารสนเทศ กฟผ. 2547 ดำรงตำแหน่งวิศวกรระดับ 9 ทำหน้าที่หัวหน้าแผนกปฏิบัติการ และบำรุงรักษาอินเทอร์เน็ต กองศูนย์อินเทอร์เน็ต ฝ่ายปฏิบัติการเทคโนโลยีสารสนเทศ กฟผ. 2553 ดำรงตำแหน่งวิศวกรระดับ 9 ทำหน้าที่หัวหน้าแผนกวางแผนและออกแบบอินเทอร์เน็ต กองศูนย์อินเทอร์เน็ต ฝ่ายปฏิบัติการเทคโนโลยีสารสนเทศ กฟผ.