

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การระบุตัวบุคคลจากภาพวีดิทัศน์โดยใช้หลายเอกลักษณ์

PERSON IDENTIFICATION FROM VIDEO SEQUENCE USING
MULTIPLE CUES



จพ.
๑/๒๔ ๗
๒๐๕๐

เลขหมู่.....
เลขทะเบียน..... 70925
วัน,เดือน,ปี... 2.2 ... ๒๕๕๐

b. 115 90087
i.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศ
บัณฑิตวิทยาลัย
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. ๒๕๕๐

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**PERSON IDENTIFICATION FROM VIDEO SEQUENCE USING
MULTIPLE CUES**



**A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY
SCHOOL OF GRADUATE STUDIES
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

2007

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2007

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารลิขสิทธิ์สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อผู้ยืมได้เห็นว่าเอกสารนี้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การระบุตัวบุคคลจากภาพวีดิทัศน์โดยใช้หลายเอกลักษณ์
นักศึกษา	นายคนูวัฒน์ แสงผล
รหัสประจำตัว	44067410
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
พ.ศ.	2550
อาจารย์ผู้ควบคุมวิทยานิพนธ์	ผศ.ดร.ธนารัตน์ ชลิตาพงศ์

บทคัดย่อ

วิทยานิพนธ์นี้ได้เสนอแนวทางในการพัฒนาระบบระบุตัวบุคคลจากวีดิทัศน์โดยอัตโนมัติ โดยใช้หลักการประมวลผลภาพวีดิทัศน์และคอมพิวเตอร์วิชัน มาใช้ในการวิเคราะห์ คัดแยก เอกลักษณ์ต่างๆของบุคคล ได้แก่ ภาพใบหน้า สีผิว ท่าเดิน สีเครื่องแต่งกาย และ ข้อมูลเชิงเวลา เพื่อใช้ในการระบุตัวบุคคล ซึ่งจะทำให้การระบุตัวบุคคลมีความถูกต้องแม่นยำมากขึ้น ขั้นตอนของกระบวนการระบุตัวบุคคลประกอบไปด้วย ขั้นตอนแรก จะเป็นการสร้างแบบจำลองบุคคลระยะยาว โดยรับภาพวีดิทัศน์เข้ามา จากนั้นทำการลบฉากหลัง เมื่อบริเวณที่เป็นฉากหลังถูกลบออกไป จะทำการค้นหาใบหน้า สีผิว และลักษณะการเดิน นำข้อมูลที่ได้จากการวิเคราะห์ไปสร้างเป็นแบบจำลองบุคคล โดยในส่วนของใบหน้า และสีผิวนั้นจะใช้โครงข่ายประสาทเทียมแบบแพร่กลับ เข้ามาช่วยในการระบุตัวบุคคล ขั้นตอนที่สอง เป็นการเป็นส่วนของการระบุตัวบุคคล โดยรับภาพวีดิทัศน์เพื่อทำการตรวจสอบว่าเป็นบุคคลใด โดยภาพวีดิทัศน์ที่ส่งไปจะเป็นภาพที่ถูกลบฉากหลังออก จากนั้นจะถูกส่งไปยัง 3 กระบวนการ กระบวนการแรก คือ การรู้จำใบหน้า กระบวนการที่สองคือ การรู้จำสีผิว กระบวนการสุดท้ายคือ การรู้จำลักษณะการเดิน เมื่อสามารถระบุตัวบุคคลได้แล้ว ระบบจะทำการเก็บรายละเอียดอื่น ๆ ของบุคคลนั้นไว้ในแบบจำลองบุคคลระยะสั้น โดยจะมีสีเสื้อผ้าที่บุคคลนั้นสวมใส่ และเวลาที่บุคคลนั้นปรากฏขึ้นภายในบริเวณเฟรมะวัง เพื่อใช้เป็นเอกลักษณ์เสริม ในกรณีที่ใบหน้าไม่สามารถระบุตัวบุคคลได้อย่างชัดเจน แต่เมื่อนำเอกลักษณ์อื่น ๆ มาประกอบเข้าด้วยกัน เช่น เมื่อดูจาก สีผิว ลักษณะการเดิน สีเสื้อผ้าที่สวมใส่ ก็จะทำให้ระบบสามารถที่จะระบุตัวบุคคลได้ ความผิดพลาดของการระบุตัวบุคคลก็จะลดน้อยลง ถ้าระบบยังไม่สามารถที่จะระบุตัวบุคคลได้หรือระบุได้ว่าเป็นบุคคลแปลกหน้าระบบจะทำการส่งสัญญาณไปยังผู้ดูแลให้เข้ามาตรวจสอบต่อไป งานวิจัยชิ้นนี้ได้ทำการทดลองกับบุคคล 20 คน จากผลการทดลองพบว่างานวิจัยที่นำเสนอนี้ สามารถใช้ในการระบุตัวบุคคลได้แม่นยำกว่าการใช้เอกลักษณ์เพียงชนิดเดียว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis	Person Identification from Video Sequence using Multiple Cues
Student	Mr.Danuwat Swangpol
Student ID	44067410
Degree	Master of Science
Program	Information Technology
Year	2007
Thesis Advisor	Asst.Prof. Dr.Thanarat Chalidabhongse

ABSTRACT

This thesis proposes a development of a system for automatic person identification from videos. The system employs techniques of image/video processing and computer vision to analyze and extract the detected person's features such as face region, skin color, gait, clothes color, and temporal data of his appearance to identify the person. By using multiple cues of a person, the identification yields better accuracy. The system composes of two main steps. The first step is to construct the long-term person models. First, the input video is segmented using background subtraction. Then, face and skin regions are located from foreground pixels. Two backpropagation neural networks are used for face and skin training. Gait also extracted and registered in the long-term person model. The second step is to identify the person using trained neural networks and long-term person models. When person approach the watching area, the system subtract background from input image. From foreground image, face detection/recognition, skin detection/ recognition, and gait recognition are processed. If the person is known before, the system will keep log of his appearance and extract clothes color for the short-term person model. These short-term person models are used in the case of system cannot identify person with the long-term person models. If the person is unknown or identified as a stranger to the system, the system will send an alarm message to an authorized person. We performed some experiments on 20 persons in our lab using two fixed cameras. The experimental results show the system with multiple cues yields better results than using single cues.

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สำเร็จลุล่วงได้อย่างดี ด้วยความกรุณาจากอาจารย์ผู้ควบคุมวิทยานิพนธ์ ผศ.ดร.ธนารัตน์ ชลิตาพงศ์ ผู้ซึ่งให้คำแนะนำ ชี้แนะแนวทางที่เป็นประโยชน์ และให้ความช่วยเหลือมาโดยตลอดในทุกๆ เรื่อง ผู้จัดทำขอขอบพระคุณเป็นอย่างสูง

ขอขอบคุณเจ้าหน้าที่ฝ่ายต่างๆ ของคณะเทคโนโลยีสารสนเทศ ที่ให้ความช่วยเหลือและอำนวยความสะดวกในทุกๆ ด้าน

ขอขอบคุณบิดามารดา ที่ให้ความอนุเคราะห์ทางด้านทุนทรัพย์และคอยเป็นห่วงให้กำลังใจ อยู่ตลอด

สุดท้ายขอขอบคุณเพื่อนๆ ร่วมรุ่น IS12.1 วามม์ สน และ วิวิวัฒน์ ใน PIC Lab ที่คอยมาเป็นนายแบบให้ และอีกหลายๆ ท่านที่ไม่ได้เอ่ยนามที่ให้คำปรึกษาทำให้ผู้จัดทำเข้าใจแนวทาง รวมทั้งช่วยให้มีกำลังใจในการทำงานวิจัยเสมอมา

คุณวิวัฒน์ แสงผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VIII
สารบัญรูป.....	IX
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา.....	2
1.3 ทฤษฎีหรือแนวความคิดที่ใช้ในการวิจัย.....	2
1.4 ขอบเขตของการวิจัย.....	2
1.5 ขั้นตอนการศึกษา.....	2
1.6 ประโยชน์ที่คาดว่าจะได้รับ.....	3
บทที่ 2 งานวิจัยที่เกี่ยวข้อง.....	4
2.1 งานวิจัยที่เกี่ยวกับการรู้จำใบหน้า (Face Recognition).....	4
2.1.1 วิธีการวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis).....	6
2.1.2 วิธีการวิเคราะห์ด้วยโครงข่ายประสาทเทียมแบบแพร่กลับ.....	8
2.2 งานวิจัยที่เกี่ยวกับการรู้จำลักษณะการเดิน (Gait Recognition).....	11
2.2.1 วิธีการรู้จำโดยใช้แบบจำลอง.....	11
2.2.2 วิธีการรู้จำโดยวิธีการวิเคราะห์ทางสถิติ.....	11
2.2.3 วิธีการรู้จำโดยวิธีการวิเคราะห์ทางกายภาพ.....	12
2.2.4 วิธีการรู้จำโดยวิธีการวิเคราะห์การเคลื่อนไหวเชิงพื้นที่ และเชิงเวลา.....	12
2.3 การตรวจจับใบหน้า.....	14
2.4 การตรวจจับสีผิว.....	16
บทที่ 3 ทฤษฎีการประมวลผลภาพวิดิทัศน์และคอมพิวเตอร์ชัน.....	18
3.1 มอร์โฟโลยีทางคณิตศาสตร์ (Mathematical Morphology).....	18

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่บนสื่อออนไลน์ การคัดลอกโดยไม่ได้รับอนุญาตถือว่าผิดกฎหมาย และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
3.1.1 อีรอสชัน (Erosion).....	20
3.1.2 ไดเลชัน (Dilation).....	20
3.1.3 โอเพนนิ่ง (Opening).....	21
3.1.4 โคลสซิง (Closing).....	21
3.2 แบบจำลองสี (Color Model).....	22
3.2.1 แบบจำลองสี RGB.....	23
3.2.2 แบบจำลองสีอนอร์เมลไลซ์ RGB (rgb).....	24
3.2.3 แบบจำลองสีแบบ HIS.....	25
3.2.2 แบบจำลองสี YCbCr.....	27
3.3 การวิเคราะห์องค์ประกอบที่เชื่อมต่อกันของภาพ (Connected Component Analysis)...	27
3.4 การแยกวัตถุออกจากฉากหลังในภาพวิดีโอ.....	29
3.4.1 เฟรมดิฟเฟอเรนซ์ซิง (Frame Differencing).....	29
3.4.2 การลบฉากหลัง (Background Subtraction).....	28
3.4.3 อิมเมจโฟลว์ (Image Flow).....	36
3.5 การแยกส่วนประกอบต่าง ๆ ของบุคคล.....	39
3.6 โครงข่ายประสาทเทียมแบบแพร่กลับ (Backpropagation Neural Networks).....	40
3.7 สหสัมพันธ์.....	45
3.7.1 การแปลสหสัมพันธ์จากกราฟ.....	45
3.7.1.1 ข้อมูลสัมพันธ์กันอย่างสมบูรณ์.....	46
3.7.1.2 ข้อมูลสัมพันธ์กันแต่ไม่สมบูรณ์.....	47
3.7.1.3 ข้อมูลสัมพันธ์ต่ำ.....	47
3.7.1.4 ข้อมูลสัมพันธ์เป็นศูนย์.....	48
3.7.2 การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์.....	49
บทที่ 4 กรรมวิธีที่นำเสนอ.....	50
4.1 แบบจำลองบุคคล (Person Model).....	51
4.1.1 แบบจำลองบุคคลระยะยาว.....	51
4.1.2 แบบจำลองบุคคลระยะสั้น.....	52

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นับญาติให้มาใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
4.2 การสร้างแบบจำลองบุคคล (Person modeling).....	51
4.2.1 การสร้างแบบจำลองบุคคลระยะยาว.....	52
4.2.2 การสร้างแบบจำลองบุคคลระยะสั้น.....	53
4.3 การเก็บวีดิทัศน์ตัวอย่าง.....	54
4.4 เอกลักษณ์ที่ใช้ในการระบุตัวบุคคล.....	55
4.4.1 ภาพใบหน้า.....	55
4.4.2 ภาพสีผิว.....	57
4.4.3 ภาพลักษณะการเดิน.....	57
4.5 การรู้จำและการปรับปรุงแบบจำลองส่วนบุคคล.....	58
4.5.1 การระบุตัวบุคคลด้วยใบหน้า.....	59
4.5.2 การระบุตัวบุคคลด้วยสีผิว.....	60
4.5.3 การระบุตัวบุคคลด้วยลักษณะการเดิน.....	60
4.5.4 การระบุตัวบุคคลด้วยสีเครื่องแต่งกาย.....	62
4.5.5 การระบุตัวบุคคลด้วยเวลา.....	65
4.5.6 การปรับปรุงแบบจำลองระยะสั้น.....	66
4.6 การรวบรวมผลของการระบุตัวบุคคล.....	66
4.6.1 ค่าความมั่นใจของเอกลักษณ์ใบหน้าและสีผิว.....	67
4.6.2 ค่าความมั่นใจของลักษณะการเดิน.....	68
4.6.3 การรวมค่าความมั่นใจ.....	67
4.6.3.1 การปรับค่าความมั่นใจด้วยเวลา.....	68
4.6.3.2 การปรับค่าความมั่นใจด้วยสีเครื่องแต่งกาย.....	69
บทที่ 5 วิธีการทดลองและผลการทดลอง.....	70
5.1 การลบฉากหลังออกจากภาพวีดิทัศน์.....	70
5.2 การรู้จำตัวบุคคลด้วยใบหน้า.....	71
5.3 การรู้จำตัวบุคคลด้วยสีผิว.....	74
5.4 การรู้จำตัวบุคคลด้วยลักษณะการเดิน.....	76
5.5 การรู้จำตัวบุคคลด้วยใบ สีผิว เวลา และสีเสื้อ.....	77
5.6 การรู้จำตัวบุคคลด้วยสีผิว ลักษณะการเดิน เวลา และสีเสื้อ.....	79

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
บทที่ 6 สรุปผลการวิจัยและข้อเสนอแนะ.....	81
6.1 สรุปผลการวิจัย.....	81
6.1.1 การระบุตัวบุคคลด้วยใบหน้า.....	81
6.1.2 การระบุตัวบุคคลด้วยสีผิว.....	81
6.1.3 การระบุตัวบุคคลด้วยลักษณะการเดิน.....	82
6.1.4 การระบุตัวบุคคลด้วยเอกลักษณ์ใบหน้า สีผิว เวลา และ สีเสื้อ.....	83
6.1.5 การระบุตัวบุคคลด้วยเอกลักษณ์สีผิว เวลา ลักษณะการเดิน และ สีเสื้อ.....	83
6.2 ข้อเสนอแนะในการทำวิจัยต่อไป.....	83
บรรณานุกรม	84
ภาคผนวก.....	87
ภาคผนวก ก. ผลงานวิจัยที่ได้รับการตีพิมพ์เผยแพร่.....	88
ประวัติผู้เขียน	94

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
5.1 ผลของการรู้จำด้วยสีผิว.....	75
5.2 ผลการรู้จำด้วยเกท.....	76
5.3 ผลการรู้จำด้วยใบหน้า สีผิว เวลา และสีเสื้อ.....	77
5.4 แสดงผลการรู้จำด้วยสีผิว ลักษณะการเดิน เวลา และสีเสื้อ.....	79



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 ขั้นตอนการทำงานของระบบรู้จำใบหน้า.....	4
2.2 การจัดรูปแบบของภาพในระบบเวกเตอร์.....	6
2.3 ข้อมูลภาพที่เตรียมก่อนการคำนวณ PCA.....	6
2.4 ภาพใบหน้าไอเคน.....	8
2.5 ระบบโครงข่ายประสาทเทียม.....	9
2.6 PCA และ โครงข่ายประสาทเทียมแบบแพร่กลับ.....	9
2.7 ค่าไอเคน และ ไอเคนเวกเตอร์.....	10
2.8 ซิมิลาริตี พล็อต.....	13
2.9 โครงสร้างการทำงานของระบบค้นหาภาพใบหน้าโดยใช้โครงข่ายประสาทเทียม.....	14
2.10 แสดงเอกลักษณ์ฮาร์ (Haar-like Feature).....	15
2.11 การกระจายตัวของบริเวณที่เป็นสีผิวและไม่เป็นสีผิว.....	16
3.1 ตัวอย่างสตริกเจอร์อีลิเมนต์ขนาด 3×3 พิกเซล.....	19
3.2 การทำอีรอสชัน $A \ominus B$	20
3.3 การทำไดเลชัน $A \oplus B$	20
3.4 การทำโอเพนนิ่ง $A \circ B$	21
3.5 การทำโคลสซิง $A \bullet B$	22
3.6 โครมาติซิตี โคอะเกรม.....	23
3.7 แบบจำลองสี RGB.....	24
3.8 ความแตกต่างของแสงในปริภูมิสี rgb.....	25
3.9 แบบจำลองสี HIS.....	25
3.10 องค์ประกอบของภาพ.....	28
3.11 การระบุหมายเลขให้กับองค์ประกอบของภาพโดยพิจารณาใน 4 ทิศทาง.....	28
3.12 การทำเฟรมดิฟเฟรนซ์ซิง.....	30
3.13 แบบจำลองสีในรูปแบบ 3 มิติในระบบสี RGB.....	31
3.14 ความเปลี่ยนแปลงของค่าความสว่างและค่า Chromaticity Distortion.....	34
3.15 อีสโทแกรม ของค่า Normalized Brightness Distortion (a) และแสดงอีสโทแกรมของ ค่า Normalized Color Distortion(b).....	36
3.16 แสดงการลบฉากหลังออกจากภาพ.....	36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่บนสื่อออนไลน์
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
3.17 แสดงแบบจำลองบุคคล.....	39
3.18 แสดงอัตราส่วนในการแยกส่วนประกอบของร่างกาย.....	39
3.19 ขั้นตอนการค้นหาส่วนประกอบของร่างกาย.....	40
3.20 ขั้นตอนการปรับขนาด.....	40
3.21 โครงสร้างแบคพรอพพาเกชันนิวรอนเน็ตเวิร์ก.....	41
3.22 กราฟแสดงสหสัมพันธ์อย่างสมบูรณ์เป็นไปทางบวก.....	46
3.23 กราฟแสดงสหสัมพันธ์อย่างสมบูรณ์เป็นไปทางลบ.....	46
3.24 กราฟแสดงสหสัมพันธ์ค่อนข้างสูงเป็นไปทางบวก.....	47
3.25 การแสดงสหสัมพันธ์ค่อนข้างสูงเป็นไปทางลบ.....	47
3.26 กราฟแสดงสหสัมพันธ์ค่อนข้างต่ำเป็นไปทางบวก.....	48
3.27 กราฟแสดงสหสัมพันธ์ค่อนข้างต่ำเป็นไปทางลบ.....	48
3.28 กราฟแสดงสหสัมพันธ์เป็นศูนย์.....	48
4.1 แสดงขั้นตอนการสร้างแบบจำลองบุคคล.....	51
4.2 แสดงขั้นตอนการระบุตัวบุคคล.....	51
4.3 ภาพแบบจำลองบุคคลระยะยาว.....	53
4.4 แสดงการรู้จำและแบบจำลองระยะสั้น.....	53
4.5 แสดงวิถีทัศน์ของบุคคลที่เดินเข้าหากล้อง.....	54
4.6 แสดงวิถีทัศน์ของบุคคลเดินหันข้างกับกกล้อง.....	55
4.7 แสดงผลการลบฉากหลังในมุมมองทั้งสอง.....	55
4.8 แสดงการค้นหาใบหน้า.....	56
4.9 แสดงภาพตัวอย่างใบหน้า.....	56
4.10 แสดงภาพตัวอย่างสีผิวในระบบสี RGB.....	57
4.11 แสดงภาพตัวอย่างการเดินหันข้างกับกกล้อง.....	58
4.12 ผลทางสถิติของความกว้างในการก้าวขาของบุคคล 20 คน.....	58
4.13 แสดงภาพใบหน้าที่ผ่านการปรับสี.....	59
4.14 แสดงผลของการรู้จำด้วยระบบโครงข่ายประสาทเทียม.....	59
4.15 แสดงผลการรู้จำด้วยโครงข่ายประสาทเทียม.....	60
4.16 ตัวอย่างโครงร่างของโครงข่ายประสาทเทียมแบบแบคพรอพพาเกชันที่ใช้รู้จำสีผิว.....	60

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยมนำไปเผยแพร่โดยไม่ได้รับอนุญาตจากเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
4.17 ภาพความกว้างของการก้าวขาในการเดิน.....	61
4.18 แสดงการหาค่าความกว้างสูงสุด.....	62
4.19 ลักษณะสี่เหลี่ยม.....	63
4.20 แสดงการหาส่วนประกอบของร่างกาย.....	63
4.21 แสดงขั้นตอนการค้นหาส่วนประกอบของร่างกาย.....	64
4.22 แสดงขั้นตอนการปรับขนาด.....	64
4.23 แสดงส่วนของเครื่องแต่งกายที่ถูกคัดแยก.....	64
4.24 แสดงการกำจัดบริเวณที่ไม่ใช่สี่เหลี่ยม.....	65
4.25 แสดงค่าความมั่นใจที่ได้.....	66
4.26 ผลทางสถิติของความกว้างในการก้าวขาสูงสุดของบุคคล 20 คน.....	67
4.27 แสดงขั้นตอนการรวมค่าความมั่นใจ.....	68
5.1 ผลของกระบวนการลบฉากหลัง.....	70
5.2 ผลของการกำจัดสิ่งรบกวน.....	71
5.3 ข้อมูลใบหน้าที่ใช้ในการเรียนรู้.....	71
5.4 การเรียนรู้ของระบบโครงข่ายประสาทเทียม.....	72
5.5 ผลของการเรียนรู้ของระบบโครงข่ายประสาทเทียม.....	72
5.6 ผลการเรียนรู้ของโครงข่ายประสาทเทียมครั้งที่สอง.....	73
5.7 ผลของการระบุตัวบุคคลได้อย่างถูกต้อง.....	73
5.8 แสดงผลของการระบุตัวบุคคลที่ผิดพลาด.....	74
5.9 แสดงสีผิวในระบบสี RGB.....	74
5.10 ผลทางสถิติของความกว้างในการก้าวขาของบุคคล 20 คน.....	76
5.11 แสดงอัตราความถูกต้องกับจำนวนบุคคลที่เพิ่มขึ้น.....	78
5.12 แสดงอัตราความถูกต้องกับจำนวนบุคคลที่เพิ่มขึ้น.....	80
6.1 แสดงการเกาะกลุ่มของสีผิว.....	82
6.2 แสดงการเกาะกลุ่มของลักษณะการเดิน.....	82

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบัน อาคาร สถานที่สำคัญต่าง ๆ ได้ถูกสร้างเพิ่มขึ้นอย่างรวดเร็ว ไม่ว่าจะเป็นอาคารสำนักงานบริษัทต่าง ๆ ระบบรถไฟฟ้าใต้ดิน หรือแม้กระทั่งสนามบิน เป็นต้น ซึ่งเป็นสถานที่ที่ต้องให้ความสำคัญในเรื่องของการรักษาความปลอดภัย อันเนื่องมาจากสภาพสังคมในปัจจุบัน ซึ่งสถานที่สำคัญเหล่านี้ก็จะติดตั้งระบบรักษาความปลอดภัยต่าง ๆ ไม่ว่าจะเป็นการเข้าออกอาคารด้วยลายนิ้วมือ หรือมีการติดตั้งระบบวิดิทัศน์รักษาความปลอดภัยตามบริเวณสำคัญ ๆ ภายในอาคาร

ระบบวิดิทัศน์รักษาความปลอดภัยกำลังเป็นที่นิยมในปัจจุบัน โดยภาพวิดิทัศน์จากกล้องที่ติดตั้งอยู่ตามบริเวณต่าง ๆ จะถูกส่งไปยังศูนย์ควบคุม ซึ่งที่ศูนย์ควบคุมนี้ทางหนึ่งก็จะทำการบันทึกภาพเหตุการณ์ไว้ในสื่อบันทึกข้อมูล อีกทางหนึ่งก็จะส่งภาพไปยังจอรับภาพต่าง ๆ ซึ่งแต่ละจอรับภาพก็จะแสดงภาพบริเวณที่กล้องตัวนั้น ๆ ติดตั้งอยู่เพื่อแสดงภาพให้บุคลากรที่เฝ้าสังเกตการณ์ ซึ่งบุคลากรจะคอยเฝ้าสังเกตเหตุการณ์ที่ผิดปกติจากหลาย ๆ จอรับภาพและเป็นเวลานาน จึงอาจทำให้เกิดความเหนื่อยล้า ทำให้มีโอกาที่จะพลาดเหตุการณ์ที่ผิดปกติได้ จึงได้มีแนวคิดที่จะพัฒนาระบบที่จะสามารถตรวจสอบระบุตัวบุคคลได้โดยอัตโนมัติ โดยใช้ข้อมูลที่ได้จากทางระบบวิดิทัศน์ โดยใช้กรรมวิธีการทางคอมพิวเตอร์วิชัน

วิธีการระบุตัวบุคคลที่ใช้กันอยู่ในปัจจุบันนี้จะเป็นการค้นหาเอกลักษณ์สำคัญๆ ที่บุคคลนั้นมี เช่น ลายนิ้วมือ ลายมลายตา ใบหน้า สีผิว ลักษณะการเดิน เป็นต้น โดยในกรณีของลายนิ้วมือและลายมลายตานั้นจะต้องมีอุปกรณ์เฉพาะทาง ซึ่งผู้ที่ถูกตรวจสอบนั้นจะต้องมีการสัมผัสกับอุปกรณ์จึงจะสามารถค้นหาเอกลักษณ์ที่ต้องการได้ ในส่วนของการระบุตัวบุคคลจากภาพวิดิทัศน์นั้นจะใช้เอกลักษณ์ ใบหน้า สีผิว ลักษณะการเดิน หรือ อื่น ๆ ที่สามารถหาได้จากภาพระบบวิดิทัศน์ เนื่องด้วยการตรวจสอบบุคคลผ่านทางระบบวิดิทัศน์นั้นจะเป็นการตรวจสอบจากระยะไกล ผู้ที่ถูกตรวจสอบจะไม่มีสัมผัสกับอุปกรณ์ และผู้ที่ถูกตรวจสอบอยู่นั้นอาจจะรู้ตัวหรือไม่รู้ตัวก็ได้

ในวิทยานิพนธ์ฉบับนี้จึงได้นำเสนอวิธีการวิเคราะห์ในการค้นหาเอกลักษณ์ของตัวบุคคลผ่านทางระบบวิดิทัศน์โดยใช้ หลายเอกลักษณ์ ไม่ว่าจะเป็น ใบหน้า สีผิว ลักษณะการเดิน สีของเครื่องแต่งกาย จากนั้นจะนำค่าความมั่นใจ (Confidential) ของแต่ละเอกลักษณ์มารวมเข้าด้วยกันเพื่อใช้ในการรู้จำบุคคลโดยอัตโนมัติ โดยในส่วนของความรู้จำนั้น ได้นำระบบโครงข่ายประสาทเทียมแบบแพร่กลับเข้ามาใช้ โดยเป็นแบบที่ต้องมีผู้ฝึกสอนก่อนจึงจะรู้จำได้ (Supervised Learning)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 วัตถุประสงค์ของการศึกษา

1. เพื่อศึกษาและพัฒนาวิธีการตรวจจับบุคคลจากวิดีโอแบบอัตโนมัติ
2. เพื่อศึกษาและพัฒนาวิธีการค้นหาเอกลักษณ์ของตัวบุคคลจากภาพวิดีโอ
3. เพื่อศึกษาและพัฒนาวิธีการระบุตัวบุคคลโดยใช้เอกลักษณ์หลาย ๆ ชนิด
4. เพื่อสร้างระบบตรวจสอบบุคคลอัตโนมัติผ่านทางระบบวิดีโอโดยใช้กระบวนการทางคอมพิวเตอร์วิชัน

1.3 ทฤษฎีและแนวคิดที่ใช้ในการวิจัย

ในงานวิจัยนี้ได้นำแนวความคิดที่จะนำภาพวิดีโอมาใช้ในการระบุตัวบุคคลเพื่อรักษาความปลอดภัย โดยใช้วิธีการทางคอมพิวเตอร์วิชันมาวิเคราะห์ เพื่อให้ได้เอกลักษณ์ของตัวบุคคลที่สามารถค้นพบได้ในระบบวิดีโอ เช่น ใบหน้า สีผิว ลักษณะการเดิน สีเครื่องแต่งกาย เป็นต้น มาทำการระบุตัวบุคคล โดยใช้ระบบ โครงข่ายประสาทเทียมได้โดยอัตโนมัติ

1.4 ขอบเขตการวิจัย

1. ใช้ใบหน้า สีผิว ลักษณะการเดิน และ สีของเครื่องแต่งกาย เป็นเอกลักษณ์ในการระบุตัวบุคคล
2. ตำแหน่งของกล้องติดตั้งในตำแหน่งที่เหมาะสมต่อการจับภาพด้านหน้าหรือด้านข้างของบุคคล
3. ข้อมูลที่ใช้ในการทดสอบ เป็นข้อมูลวิดีโอวีดีทัศน์ของการเดินของบุคคล 20 คน

1.5 ขั้นตอนการศึกษา

1. ศึกษาวิธีการนำภาพวิดีโอเข้ามาประมวลผล
2. ศึกษาวิธีการแยกบุคคลออกจากฉากหลัง (Background Subtraction)
3. ศึกษาวิธีการแยกส่วนประกอบต่าง ๆ ของบุคคล (Body Parts Segmentation)
4. ศึกษาวิธีการค้นหาเอกลักษณ์ เช่น ใบหน้า สีผิว ลักษณะการเดิน สีเครื่องแต่งกาย
5. ศึกษาวิธีการรู้จำด้วยโครงข่ายประสาทเทียม
6. สร้างโครงสร้างและกระบวนการต่าง ๆ ตามที่ได้ศึกษา เพื่อให้ได้มาซึ่งลักษณะพิเศษ
7. ทดสอบระบบที่ใช้ในการวิเคราะห์ลักษณะพิเศษ และระบบรู้จำ
8. สรุปและอภิปรายผลการทดลอง

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์หรือการขงในเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ทราบถึงกระบวนการ วิธีการต่าง ๆ ที่เกี่ยวข้องกับงานวิจัยนี้
2. ได้วิธีการนำภาพวิดิทัศน์มาทำการประมวลผลเพื่อทำการตรวจจับและระบุตัวบุคคลอย่างอัตโนมัติ
3. ผลงานที่ได้สามารถนำไปพัฒนาต่อยอดเป็นระบบรักษาความปลอดภัยอัจฉริยะหรือตีกอัจฉริยะได้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

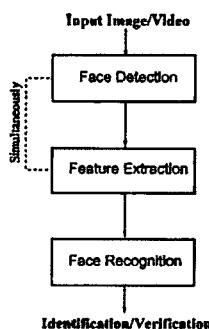
บทที่ 2

งานวิจัยที่เกี่ยวข้อง

การตรวจสอบหรือระบุตัวบุคคลสามารถกระทำได้โดยการรู้จำเอกลักษณ์ของบุคคลนั้น ๆ ไว้ในฐานข้อมูล เมื่อต้องการที่จะตรวจสอบบุคคล ระบบจะทำการวิเคราะห์ภาพที่ได้รับเข้ามาเพื่อค้นหาบริเวณที่เป็นเอกลักษณ์ (Feature Detection) จากนั้นนำเอกลักษณ์ที่ได้มาสกัดตามกระบวนการสกัดเอกลักษณ์ (Feature Extraction) เพื่อให้เหลือข้อมูลที่เป็นประโยชน์ต่อการรู้จำ ซึ่งกระบวนการรู้จำจะนำเอกลักษณ์ที่ได้มาตรวจสอบเปรียบเทียบตามวิธีการจากนั้นจึงระบุว่าเป็นบุคคลใด โดยเอกลักษณ์ที่ใช้ในระบบวิทัศน์นั้น ได้แก่ ใบหน้า ลักษณะการเดิน เป็นต้น ซึ่งแต่ละเอกลักษณ์ก็จะมีวิธีการตรวจจับเอกลักษณ์, สกัดเอกลักษณ์ และรู้จำเอกลักษณ์แตกต่างกันไป ในบทนี้จึงได้นำเสนอลงงานวิจัยที่ใช้ในการตรวจจับ สกัด และรู้จำเอกลักษณ์ที่ใช้ในระบบวิทัศน์ โดยมีรายละเอียดดังนี้

2.1 งานวิจัยเกี่ยวกับการรู้จำใบหน้า (Face Recognition)

การรู้จำตัวบุคคลโดยใช้ใบหน้านั้น ได้มีการนำมาใช้ตั้งแต่ปี 1954 โดย Bruner and Tagiuri ซึ่งเป็นการวิเคราะห์และรู้จำตัวบุคคลโดยใช้มนุษย์เป็นคนตัดสินใจ ซึ่งได้มีการพัฒนาวิธีวิเคราะห์เรื่อยมาจนกระทั่งปี 1970 จึงได้มีการค้นคว้าระบบรู้จำตัวบุคคลโดยอัตโนมัติขึ้นมาเป็นครั้งแรกโดย Kelly et al. ในช่วง 2-3 ปีที่ผ่านมา ระบบระบุตัวบุคคลจากใบหน้าได้มีการพัฒนาไปอย่างรวดเร็วทำให้ระบบมีความถูกต้องแม่นยำมากยิ่งขึ้น โดยขั้นตอนหลัก ๆ ของระบบระบุตัวบุคคลโดยใช้ภาพใบหน้าในแบบภาพนิ่งหรือวิทัศน์นั้นจะมีอยู่ 3 ส่วนด้วยกัน คือ การค้นหาตำแหน่งของใบหน้า (face detection), การค้นหาเอกลักษณ์สำคัญบนใบหน้า (face feature extraction) และระบบรู้จำใบหน้า (face recognition) โดยระบบรู้จำใบหน้าแบบอัตโนมัติในปัจจุบันจะมีขั้นตอนในการทำงานดังรูปที่ 2.1



รูปที่ 2.1 ขั้นตอนการทำงานของระบบรู้จำใบหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.1 ในขั้นตอนของการตรวจจับใบหน้า (Face Detection) จะเป็นขั้นตอนการวิเคราะห์ภาพว่าในภาพนั้นมีใบหน้าอยู่ตรงบริเวณใดในภาพ เพื่อที่จะนำเฉพาะบริเวณที่เป็นใบหน้าไปใช้ในการรู้จำต่อไป ซึ่งวิธีที่ใช้ในการวิเคราะห์ค้นหา ตำแหน่งของใบหน้าจะมีอยู่หลายวิธี ดังเช่น การค้นหาใบหน้าด้วยโครงข่ายประสาทเทียม (Neural Network Face Detection) , ฮาร์-ฟีเจอร์ (Haar Features) เป็นต้น ขั้นตอนต่อมาจะเป็นส่วนของการสกัดเอกลักษณ์ของใบหน้า (Feature Extraction) และ ส่วนของการรู้จำใบหน้า (Face Recognition) จะมีวิธีการวิเคราะห์ภาพใบหน้าแต่ละภาพว่ามีเอกลักษณ์ใดประกอบอยู่บ้าง และจะรู้จำได้อย่างไร โดยวิธีวิเคราะห์จะแบ่งไปตามกลุ่มของวิธีการ ดังเช่น การใช้ลักษณะที่ปรากฏ วิธีการวิเคราะห์จะใช้วิธีการวิเคราะห์องค์ประกอบหลัก (PCA) หรือ โครงข่ายประสาทเทียม เป็นต้น ถ้าเป็นกลุ่มของการใช้แบบจำลองใบหน้า วิธีการวิเคราะห์จะใช้ กราฟกลุ่มข้อมูลแบบยืดหยุ่น (Elastic Bunch Graph) เป็นต้น

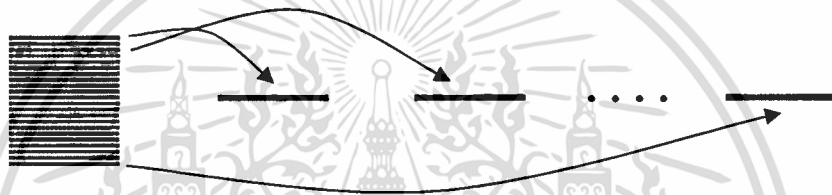
ในปัจจุบันเทคโนโลยีและวิธีการวิเคราะห์เพื่อการรู้จำตัวบุคคลจากใบหน้านั้นได้มีการพัฒนาไปมาก มีความรวดเร็วในการรู้จำ และความถูกต้องแม่นยำก็สูงขึ้น แต่ก็ยังมีข้อจำกัดอยู่บางประการ เช่น ต้องเป็นภาพหน้าตรง, เป็นภาพที่ไม่แสดงอารมณ์ออกทางสีหน้า เช่น ยิ้มหรือ หัวเราะ เป็นต้น วิธีการวิเคราะห์และการรู้จำตัวบุคคลโดยใช้ใบหน้านั้นสามารถแบ่งได้เป็น 2 กลุ่มตามลักษณะของข้อมูลที่ได้รับเข้ามา คือ ภาพใบหน้าที่เป็นภาพนิ่ง และอีกกลุ่มหนึ่งจะเป็นการรู้จำตัวบุคคลจากใบหน้าโดยใช้ภาพวิดีโอ โดยที่วิธีการวิเคราะห์ภาพใบหน้าที่จากวิดีโอ นั้นจะนำวิธีการวิเคราะห์ภาพใบหน้าที่เป็นภาพนิ่งมาประยุกต์ใช้เข้าด้วยกัน นอกจากจะแบ่งออกได้เป็น 2 กลุ่มข้างต้นแล้ว ระบบการรู้จำด้วยใบหน้าที่ยังสามารถแบ่งออกได้อีก 2 กลุ่มตามลักษณะของวิธีการวิเคราะห์ โดยกลุ่มแรกจะเป็นการใช้ลักษณะที่ปรากฏ (Appearance-Based) เป็นวิธีการวิเคราะห์ลักษณะที่ปรากฏอยู่ในภาพใบหน้า โดยนำภาพใบหน้าที่เป็นภาพขาวดำมาทำการวิเคราะห์ โดยนำรูปภาพแต่ละรูปมาจัดเรียงให้อยู่ในรูปแบบของเวกเตอร์ ซึ่งวิธีการวิเคราะห์แบบนี้จะใช้วิธีการวิเคราะห์ในเชิงสถิติ เพื่อมาช่วยวิเคราะห์การกระจายตัวของเวกเตอร์รูปภาพหลาย ๆ เวกเตอร์ ดังเช่นวิธีการวิเคราะห์องค์ประกอบหลัก เป็นต้น กลุ่มที่สองจะเป็นกลุ่มที่ใช้แบบจำลอง (Model-Based) วิธีการนี้จะเน้นไปที่การสร้างแบบจำลองใบหน้าของมนุษย์ที่ประกอบไปด้วยค่าตัวแปรต่างๆ ของใบหน้า เช่น ระยะห่างระหว่างตา, ระยะระหว่างตาทั้งสองข้างถึงจมูก และปาก เป็นต้น เพื่อใช้เป็นค่าเปรียบเทียบ ที่จะนำไปใช้ในการรู้จำใบหน้าต่อไป ดังเช่น วิธีการวิเคราะห์จะใช้กราฟกลุ่มข้อมูลแบบยืดหยุ่น (Elastic Bunch Graph) เป็นต้น

ซึ่งในวิทยานิพนธ์ฉบับนี้ จะใช้วิธีการทางลักษณะที่ปรากฏ (Appearance-Based) มาใช้ในการรู้จำภาพใบหน้า โดยมีรายละเอียดดังนี้

2.1.1 วิธีการวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis, PCA)

วิธีการวิเคราะห์องค์ประกอบหลักเป็นวิธีการทางสถิติ ซึ่งถูกนำไปประยุกต์ใช้ในงานต่าง ๆ เช่น การสร้างภาพใบหน้าไอเกน (EigenFaces) เพื่อใช้ในการรู้จำ เป็นต้น วิธีการนี้ได้ถูกนำมาใช้ในการสร้างภาพใบหน้าไอเกนครั้งแรกโดย Sirovich และ Kirby ในปี 1987 จนกระทั่งปี 1991 M.A. Turk และ Alex P. Pentland ได้สร้างระบบรู้จำภาพใบหน้า โดยใช้ภาพใบหน้าไอเกนเป็นข้อมูลที่ใช้ในการรู้จำ

การที่จะนำภาพใบหน้ามาวิเคราะห์ด้วยวิธี PCA นั้น จะต้องทำการจัดให้อยู่ในรูปแบบของเวกเตอร์เสียก่อน เช่น ถ้าภาพมีขนาดกว้าง (w) และสูง (h) ตามลำดับ จำนวนองค์ประกอบในเวกเตอร์ที่ได้จากภาพจะเท่ากับ $w \times h$ มิติ ในแต่ละจุดภาพ (pixel) จะแทนด้วย 1 องค์ประกอบในเวกเตอร์ นั่นคือ แต่ละแถวของภาพจะถูกเขียนเรียงต่อกันเป็นเวกเตอร์ ดังแสดงในรูปที่ 2.2



รูปที่ 2.2 การจัดรูปแบบของภาพในระบบเวกเตอร์

จากนั้นนำค่าเวกเตอร์ที่ได้มาจัดเรียงให้อยู่ในระบบเมทริก โดยให้ข้อมูลของแต่ละแถวในเมทริกเป็นข้อมูลของภาพเวกเตอร์แต่ละภาพ ดังนั้นจะได้เมทริก A ซึ่งหมายถึงข้อมูลของภาพที่ n และมีขนาด m ดังรูปที่ 2.3

$$\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nm} \end{bmatrix} \quad \begin{matrix} m = w * h \\ n = \text{number of image} \end{matrix}$$

รูปที่ 2.3 ข้อมูลภาพที่เตรียมก่อนการคำนวณ PCA

ขั้นตอนในการคำนวณ PCA มีดังต่อไปนี้

1. คำนวณค่าเฉลี่ยของภาพใบหน้า

$$M_j = \frac{1}{n} \sum_{i=1}^n A_{ij} \quad \text{โดยที่ } 1 \leq j \leq m \quad (2.1)$$

2. คำนวณค่าเบี่ยงเบนมาตรฐานของภาพใบหน้า

$$S_{ij} = A_{ij} - M_j \quad \text{โดยที่ } 1 \leq i \leq n \text{ และ } 1 \leq j \leq m \quad (2.2)$$

3. สร้างเมทริกซ์ของความแปรปรวนร่วม

$$C = \frac{1}{n} \sum_{i=1}^n S_i S_i^T \quad (2.3)$$

4. คำนวณค่าไอเกน

$$\lambda = CC^T \quad (2.4)$$

5. คำนวณไอเกนเวกเตอร์

$$e = \frac{SC}{\sqrt{\lambda}} \quad \text{เมื่อ } S = [S_{ij}] \quad (2.5)$$

ผลลัพธ์ที่ได้จากการคำนวณด้วยวิธีการวิเคราะห์องค์ประกอบหลักจะได้ค่าไอเกน (λ) และไอเกนเวกเตอร์ (e) ซึ่งข้อมูลทั้งสองมีความสมนัย (Correspondence) ซึ่งกันและกัน ซึ่งไอเกนเวกเตอร์นี้จะเปลี่ยนรูปไปแล้ว ไม่สามารถมองเห็นเป็นภาพใบหน้าได้ แต่หากต้องการนำกลับมาสร้างเป็นภาพใบหน้าใหม่อีกครั้งจะเรียกว่า ภาพใบหน้าไอเกน (Eigen Faces) ซึ่งจะต้องนำไอเกนเวกเตอร์ที่ได้มาคำนวณอีกครั้ง ดังสมการที่ 2.6

$$F_k = e_k^T (A - M) \quad (2.6)$$

โดยที่

$$A = [A_{ij}]$$

$$M = [M_j]$$

เมื่อทำการคำนวณภาพใบหน้าไอเกน (F) เป็นที่เรียบร้อยแล้ว ดังรูปที่ 2.4



รูปที่ 2.4 ภาพใบหน้าไอเกน

ภาพใบหน้าแต่ละภาพจะถูกแทนด้วยค่าน้ำหนัก (Weight) ที่ได้จากการคำนวณภาพใบหน้าไอเกน เมื่อต้องการตรวจสอบว่าภาพใบหน้าที่ได้รับเข้ามาใหม่นั้นเป็นภาพของบุคคลใด ให้นำภาพใบหน้าใหม่นั้นไปคำนวณหาค่าน้ำหนัก และทำการค้นหาภาพใบหน้าในระบบใบหน้า (Face Space) โดยใช้วิธีการวัดระยะทางแบบยูคลิดีเนียน (Euclidean Distance) ถ้าภาพบุคคลที่นำมาตรวจสอบมีระยะใกล้เคียงกับกลุ่มของข้อมูลภาพใบหน้าใดในระบบใบหน้ามากที่สุด ภาพที่นำมาตรวจสอบจะถูกระบุว่าเป็นบุคคลนั้น

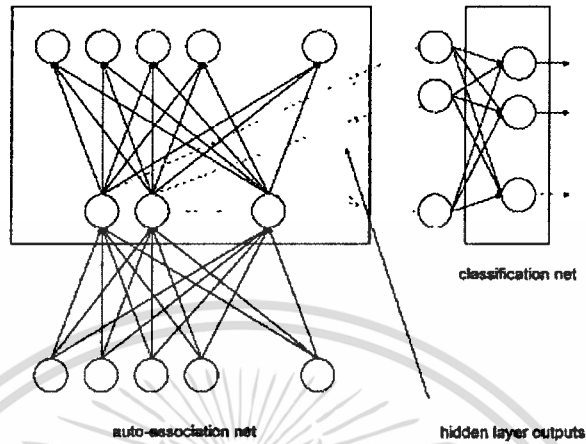
2.1.2 วิธีการวิเคราะห์ด้วยโครงข่ายประสาทเทียมแบบแพร่กลับ

โครงข่ายประสาทเทียมจะมีความทำงานโดยเลียนแบบการทำงานของระบบโครงข่ายประสาทของมนุษย์ แต่ก่อนที่ระบบโครงข่ายประสาทเทียมจะทำการรู้จำได้นั้น จะต้องมีการฝึกสอน (Training) หรือ เรียนรู้ (Learning) มาก่อน ระบบที่พัฒนาด้วยโครงข่ายประสาทเทียมจึงจะมีความสามารถในการแยกแยะ (Classification) หรือ จดจำ (Recognition) ข้อมูลที่นำมาทดสอบได้ แต่ระบบโครงข่ายประสาทเทียมจะมีความซับซ้อนถ้าภาพข้อมูลมีขนาดที่ใหญ่จะทำให้ระบบนั้นมีความยุ่งยากในการฝึกสอน หรือฝึกสอนได้ช้า หรือไม่กระทั่งไม่สามารถฝึกสอนได้ เช่น ถ้าภาพขนาด 128×128 จุดภาพ จำนวนข้อมูลขาเข้า (input) ของระบบโครงข่ายประสาทเทียมจะมีขนาดถึง 16384 ข้อมูล ซึ่งเป็นขนาดข้อมูลที่ใหญ่มาก เพื่อลดความซับซ้อนของข้อมูลลงจึงได้มีนักวิจัยที่ได้พัฒนาวิธีการต่าง ๆ ที่จะช่วยลดความซับซ้อนดังต่อไปนี้

Cottrell and Fleming (1990) ได้ใช้โครงข่ายประสาทเทียมแบบแพร่กลับ 2 โครงข่าย ดังแสดงในรูปที่ 2.6 ซึ่งจะประกอบไปด้วย โครงข่ายแรกคือ โครงข่ายที่ใช้เปรียบเทียบความเกี่ยวข้องกันแบบอัตโนมัติ (Auto association) และสกัดเอกลักษณ์ เพื่อใช้เป็นข้อมูลขาเข้าให้กับโครงข่ายส่วนที่สอง โดยโครงข่ายที่ 2 จะเป็นกระบวนการคัดแยก (Classification) ที่ใช้กันโดยทั่วไป ในส่วนของโครงข่ายที่ใช้เปรียบเทียบความเกี่ยวข้องกันแบบอัตโนมัตินั้นจะมีข้อมูลขาเข้า n ข้อมูล และข้อมูลขาออก n ข้อมูล และชั้นซ่อน p ข้อมูล ซึ่ง $p \ll n$ จากนั้นนำเวกเตอร์ใบหน้า x ไปทำการสอนให้กับโครงข่ายประสาทเทียมแบบแพร่กลับ เพื่อให้ได้ผลลัพธ์ y ซึ่งเป็นค่าที่ใกล้เคียงที่สุด (best approximation) ของ x ซึ่งทำให้ผลลัพธ์ของชั้นซ่อน h จะประกอบไปด้วยค่าของ x ที่ถูกบีบ

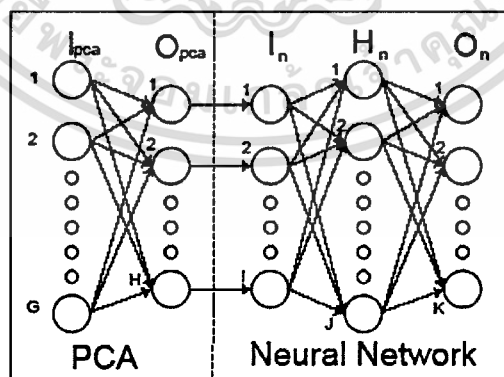
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อัด หรือเรียกว่า ฟีเจอร์เวกเตอร์ (Feature Vector) ซึ่งสามารถนำไปใช้เป็น ข้อมูลขาเข้าให้กับ โครงข่ายที่สองซึ่งเป็นส่วนของการคัดแยก (Classification) ต่อไป



รูปที่ 2.5 แสดงระบบ โครงข่ายประสาทเทียม

จากงานวิจัยของ สนั่น ศรีสุข (2541) ได้นำเสนอวิธีการลดความซับซ้อนของข้อมูลด้วยวิธีการวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis, PCA) เพื่อช่วยลดข้อมูลที่ไม่จำเป็นจากรูปภาพใบหน้าทิ้งไป ทำให้ใช้ข้อมูลสำหรับการรู้จำน้อยลง ซึ่งเป็นวิธีการที่สามารถลดขนาดของข้อมูลให้เหลือเท่าที่จำเป็น โดยข้อมูลภาพที่ใช้มีขนาด กว้าง 50 จุดภาพ สูง 55 จุดภาพ 255 สี เมื่อเปลี่ยนให้อยู่ในรูปเวกเตอร์แล้วจะมีขนาดเป็น $55 \times 50 = 2750$ มิติ การทำงานของระบบแสดงดังรูปที่ 2.6

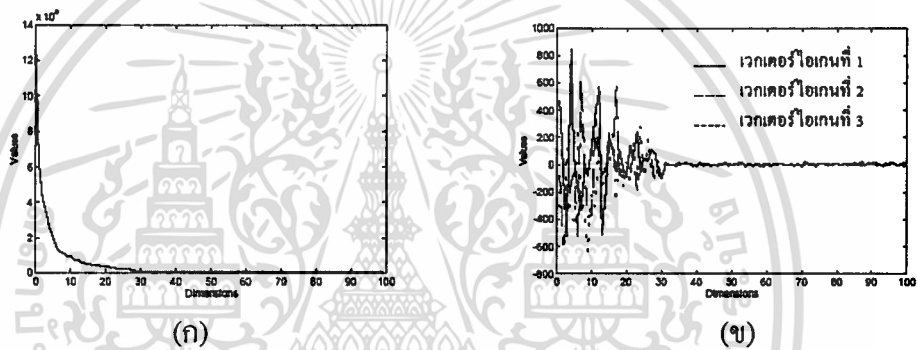


รูปที่ 2.6 PCA และ โครงข่ายประสาทเทียมแบบแพร์กัลบ

วิธีการวิเคราะห์องค์ประกอบหลักจะทำการสร้างเมทริกความแปรปรวนร่วม (Covariance Matrix) จากข้อมูลภาพ จากนั้นคำนวณหาค่าไอเกน (Eigen value) และ เวกเตอร์ไอเกน (Eigen vector) ดังสมการที่ 2.1 – 2.5

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การจะเลือกจำนวนมิติของข้อมูลจะพิจารณาจากค่าไอเกน และ ไอเกนเวกเตอร์ เมื่อพิจารณาค่าไอเกนจากรูป 2.7 (ก) ที่มีการเรียงข้อมูลแล้ว จะเห็นการลดลงของค่าไอเกน เมื่อเกินองค์ประกอบที่ 30 จะมีค่าประมาณศูนย์ทั้งหมด ซึ่งค่าไอเกนที่มีค่าต่ำ ๆ ที่สมนัยกับเวกเตอร์ไอเกนนั้น ๆ หมายถึงข้อมูลที่ไม่มีความสำคัญสามารถตัดทิ้งได้ และเมื่อพิจารณาตัวอย่างเวกเตอร์ไอเกนจากรูป ที่ 2.7(ข) จะเห็นว่าเมื่อเกินองค์ประกอบที่ 30 จะมีค่าประมาณศูนย์เมื่อนำไปให้โครงข่ายประสาทเทียมเรียนรู้ ข้อมูลส่วนนี้จะไม่มีส่วนช่วยในการแยกแยะข้อมูลของโครงข่ายประสาทเทียม ดังนั้นข้อมูลที่จะถูกนำไปรู้จำจึงเหลือ 30 องค์ประกอบแรกเท่านั้น ซึ่งคิดเป็น $(\frac{30}{2750} \times 100 = 1.09\%)$ จากข้อมูลเดิม



รูปที่ 2.7 แสดงค่าไอเกน และ ไอเกนเวกเตอร์

จากนั้นทำการปรับค่าผลลัพธ์ที่ได้จากการคำนวณ ให้เหมาะสมกับโครงข่ายประสาทเทียม โดยจะถูกปรับให้อยู่ในช่วง $[0,1]$ เพื่อป้อนค่าที่เหมาะสมให้กับโครงข่ายประสาทเทียม โดยการปรับค่า ดังสมการที่ 2.7

$$E_{i,j} = \frac{e_{i,j} - Min}{Max - Min} \quad 1 \leq i \leq n; 1 \leq j \leq 30 \quad (2.7)$$

เมื่อ e เป็นผลลัพธ์จากการคำนวณด้วย PCA

Min, Max เป็นค่าสุดและสูงสุดจากผลลัพธ์ของ PCA

2.2 งานวิจัยที่เกี่ยวกับการรู้จำลักษณะการเดิน (Gait Recognition)

วิธีการระบุตัวบุคคลด้วยลักษณะการเดินหรือเกท (gait) นั้นจะเป็นวิธีการที่พัฒนาขึ้นเพื่อให้สามารถระบุตัวบุคคลได้จากระยะไกล โดยใช้ระบบวิดิทัศน์ทำการรับภาพที่ต่อเนื่องเข้ามาทำการวิเคราะห์ ซึ่งลักษณะของเกท นั้นก็จะมียู่หลายลักษณะดังนี้

2.2.1 วิธีการสร้างแบบจำลอง (Model Based Methods)

วิธีการสร้างแบบจำลองเกท นั้นจะใช้บริเวณที่เป็นจุดหมุน, ความยาวของแขนขา ทั้ง 2 ข้าง, ความเร็วของการเปลี่ยนแปลงบริเวณที่เป็นเหลี่ยมมุม ซึ่งเป็นค่าที่สามารถวัดได้ จึงสามารถนำมาใช้แบบจำลองสร้างเกทได้

จากงานวิจัยของ Cunado et al. (1997) ได้มีการสร้างแบบจำลองของเกท โดยใช้การเคลื่อนไหวที่มีลักษณะเหมือนกับการแกว่งของลูกตุ้ม และใช้วิธีการแปลงฮอร์ทแบบพลวัต (Dynamic Hough Transform) ในการที่จะค้นหาเส้นที่แสดงถึงบริเวณที่เป็นต้นขา ในทุก ๆ ภาพ จากนั้นนำวิธีการหาค่ากำลังสองน้อยที่สุด (Least Square) มาใช้ในการทำให้ค่าความโน้มเอียงของข้อมูลที่เป็นบริเวณต้นขา และเป็นการกำจัดจุดผิดพลาดที่เกิดจากการที่ขาทั้งสองข้างบนบังคับกันเอง จากนั้นนำค่าแถบขนาดความถี่แบบถ่วงน้ำหนักเฟส (Phase-Weighted Magnitude Spectra) มาเป็นเอกลักษณ์ของเกท เพื่อใช้ในการระบุตัวบุคคล

จากงานวิจัยของ Yam et al. (2001) ได้นำเสนอวิธีการใหม่โดยนำลักษณะการเดินและการวิ่งมาใช้เป็นโครงสร้างในการสร้างแบบจำลองเกท โดยเรียกวิธีนี้ว่า การแกว่งต่อเนื่องแบบพลวัต (Dynamic Coupled Oscillator) เพื่อนำมาใช้สร้างแบบจำลองการเคลื่อนไหวของสะโพก ต้นขา และขาส่วนล่าง จากนั้นนำวิธีการจับคู่แม่แบบเชิงเวลา (Temporal Template Matching) มาใช้ในการค้นหามุมที่หมุนไประหว่างต้นขากับขาส่วนล่าง โดยเอกลักษณ์ของเกท นั้นจะได้จากวิธีการ แถบขนาดความถี่แบบถ่วงน้ำหนักเฟสลำดับต่ำ (Lower-Order Phase-Weighted Magnitude Spectra)

2.2.2 วิธีการวิเคราะห์ทางสถิติ (Statistical methods)

เป็นวิธีการที่จะแสดงถึงลักษณะพิเศษทางสถิติที่อธิบายการเคลื่อนที่ของข้อมูลภาพว่าเป็นไปในลักษณะใด

จากงานวิจัยของ Murase et al. (1996) ได้นำค่าตัวแปรต่าง ๆ ในระบบไอเกน (Eigen Space) มาช่วยลดการคำนวณลงและเพิ่มประสิทธิภาพในการประมาณเกท (Gait Estimation) หลังจากนั้น Hung et al. ได้นำมาพัฒนาต่อโดยเพิ่มวิธีการวิเคราะห์ระบบแบบบัญญัติ (Canonical space analysis) เพื่อให้ได้ผลที่มีการแบ่งแยกอย่างชัดเจนและมีประสิทธิภาพมากยิ่งขึ้น

จากงานวิจัยของ Shutler et al. (2000) ได้พัฒนาวิธีการวิเคราะห์โมเมนต์ความเร็ว (Velocity-moment-based) เพื่อใช้ในการอธิบายถึงการเคลื่อนที่ของวัตถุ จากข้อมูลชุดภาพ (image sequence)

วิธีการทางสถิตินั้นจะไม่พบปัญหาที่เกี่ยวกับสิ่งรบกวนและช่วงเวลาที่เปลี่ยนไปของชุดข้อมูลภาพ และเมื่อทำการเปลี่ยนเทียบกับวิธีการทางสร้างแบบจำลอง (Model-based) แล้ววิธีการคำนวณทางสถิตินั้นก็จะใช้การคำนวณที่น้อยกว่า

2.2.3 วิธีการวิเคราะห์ทางกายภาพ (Physical – Parameter – Based methods)

วิธีการนี้จะนำข้อมูลทางกายภาพของร่างกายมนุษย์มาใช้ในการสร้างรูปแบบของลักษณะการเดิน เช่น ส่วนสูง, จังหวะในการก้าวขา, ความกว้างในการก้าวขา เป็นต้น

จากงานวิจัยของ Bobick et al. (2001) ได้เสนอวิธีการที่จะค้นหาลักษณะการเดินโดยใช้ลักษณะการเคลื่อนไหวของการเดินมาเป็นต้นแบบ เพื่อที่จะให้ได้ค่าตัวแปรของร่างกายในการเดิน เช่น ระยะห่างในแนวตั้งระหว่างศีรษะถึงเท้า, ระยะห่างระหว่างศีรษะกับกระดูกเชิงกราน, ระยะห่างระหว่างเท้ากับกระดูกเชิงกราน และ ความกว้างของการก้าวขา เป็นต้น

วิธีการนี้จะมีค่าเป็นอิสระต่อมุมมอง เนื่องจากค่าตัวแปรที่ต้องการนั้นจะต้องทำในระบบสามมิติ อย่างไรก็ตามยังมีข้อจำกัดทางวิธีการประมวลผลภาพเพื่อให้ได้ค่าตัวแปรที่ต้องการ เช่น การทำแทบลิตามส่วนต่าง ๆ ของร่างกาย (Body-Parts Labeling), ความลึกชดเชย (Depth-Compensation), การเทียบวัดกล้อง (Camera Calibration), การลบเงาออกจากภาพ (Shadow Removal) เป็นต้น และค่าตัวแปรที่นำมาใช้ในการรู้จำนั้นอาจจะไม่เพียงพอ ถ้านำไปใช้กับสถานที่ที่มีคนจำนวนมาก

2.2.4 วิธีการวิเคราะห์การเคลื่อนไหวเชิงพื้นที่ และ เชิงเวลา (Spatio -Temporal Based Methods)

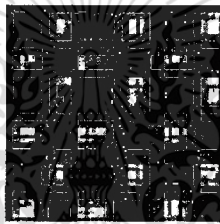
สำหรับระบบรู้จำบุคคลด้วยการเคลื่อนไหวโดยใช้วิธีการวิเคราะห์เชิงพื้นที่และเชิงเวลา (Spatio – Temporal analysis) จะทำการพิจารณาลักษณะการกระทำหรือการเคลื่อนไหวต่าง ๆ ที่เกิดขึ้น และจะถูกกำหนดโดยระบบ 3 มิติ ในทางพื้นที่และเวลา โดยใช้ข้อมูลการเคลื่อนที่ของบุคคลนั้น

จากงานวิจัยของ Niyogi et al. (1994) ได้นำเสนอ ระบบการเคลื่อนที่ (Translation) และ เวลา (Time) (XT Space) การเคลื่อนที่ของศีรษะและขาจะมีรูปแบบ (Pattern) ที่แตกต่างกัน ซึ่งรูปแบบของการเคลื่อนที่นี้จะหาได้ โดยเริ่มจาก การค้นหาเส้นกรอบ (Bounding Box) ของร่างกายที่กำลังเคลื่อนไหว จากนั้นนำเส้นกรอบที่ได้ไปปรับขนาดให้มีความเหมาะสมกับแบบจำลองห้าแท่ง (Five-stick model) โดยที่คุณลักษณะของเกทสามารถหาได้จากแบบจำลองชนิดปรับให้พอดีแบบ

เอกสแอมพลีไฟด์โมเดล (Velocity – Normalized fitted Model) หลังจากนั้น Niyogi et al. (1994) ได้
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำการขยายวิธีการวิเคราะห์ของตนเองโดยนำวิธีการวิเคราะห์เชิงพื้นที่และเชิงเวลามาใช้ในการค้นหาเขต หลังจากที่ทำกรตรวจจับการเคลื่อนไหวได้แล้วนั้น รูปแบบ (Pattern) XYT (2-D space and 1-D time) จะถูกปรับขนาดให้เหมาะสมกับ พื้นผิวเชิงพื้นที่และเชิงเวลา (Spatio-temporal surface) ซึ่ง พื้นผิวนี้อาจแสดงถึง การรวมกันของค่าตัวแปรพื้นฐานทางพื้นผิวและพื้นผิวต่าง ๆ ซึ่งสามารถนำมาใช้ในการรู้จำลักษณะการกระทำต่าง ๆ ที่เป็นขั้นพื้นฐานได้

ความคล้ายของภาพ (Image Self-Similarity) ในระบบ XYT ได้ถูกนำเสนอโดย Benaddelkader et al. (2001) โดยทำการสร้าง ซิมิลาริตีพล็อต (Similarity Plots, SPs) ของชุดข้อมูลภาพที่มีการเคลื่อนไหวของวัตถุ ซึ่ง ซิมิลาริตีพล็อต นี้จะประกอบไปด้วยข้อมูลต่าง ๆ ที่สำคัญของเขต ลักษณะของซิมิลาริตีพล็อต แสดงดังรูปที่ 2.8



รูปที่ 2.8 ซิมิลาริตีพล็อต

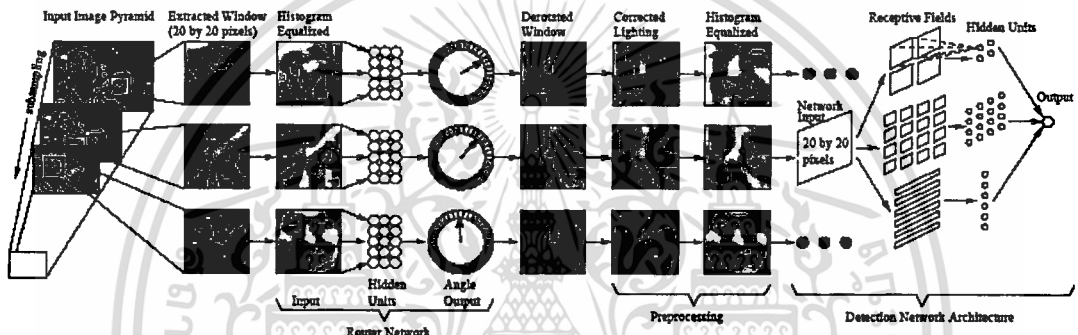
HMM-Based (Hidden Markov Model) เพื่อใช้ในการแสดงถึงคุณลักษณะของเขต และการรู้จำเขต ได้ถูกนำเสนอโดย Kale et al. J (2002) โดยจะทำการเลือกกลุ่มของเฟรมหลัก (key frames) ที่เกิดขึ้นระหว่างการเดิน จากนั้นทำการแปลงข้อมูลภาพให้อยู่ในรูปแบบของไบนารี (Binary) และนำค่าความกว้างในการก้าวขา มาเป็นข้อมูลขาเข้า (Input Feature) ขั้นตอนส่วนต่อไปจะทำการสร้างเวกเตอร์การวัดแบบน้อยมิติ (Low-Dimensional Measurement Vector) โดยใช้การวัดระยะทางแบบยูคลิดีส (Euclidean Distance) ระหว่างภาพที่จะเปรียบเทียบกับกลุ่มของเฟรมหลัก (key frames) ค่าของเวกเตอร์การวัดแบบน้อยมิติ (Low-Dimensional Measurement Vector) ที่ได้จะนำไปใช้ในกระบวนการเรียนรู้ของ HMM

การวิเคราะห์เชิงพื้นที่และเชิงเวลานั้น สามารถค้นหาคุณสมบัติของเขต ได้ทั้งเชิงพื้นที่และเชิงเวลา ซึ่งวิธีการวิเคราะห์นี้มีข้อได้เปรียบอยู่ตรงที่มีความซับซ้อนในการคำนวณต่ำ และสามารถนำไปพัฒนาได้ไม่ยากนัก แต่ข้อเสียก็คือ อ่อนไหวต่อสิ่งรบกวน และความคลาดเคลื่อนในการเคลื่อนไหว

2.3 การตรวจจับใบหน้า (Face Detection)

วิธีการตรวจจับและติดตามบุคคลในระบบวิทัศน์ในปัจจุบันนี้ ได้มีนักวิจัยได้พัฒนาวิธีต่าง ๆ ขึ้นมามากมาย โดยใช้องค์ประกอบต่าง ๆ ที่สามารถค้นหาได้จากชุดข้อมูลภาพ เช่น ใบหน้า, สีมืด, ร่างกาย เป็นต้น เพื่อใช้ในการกำจัดส่วนที่ไม่ต้องการทิ้งออกไป ให้เหลือแต่สิ่งที่สนใจ โดยวิธีการตรวจจับต่าง ๆ มีดังนี้

จากงานวิจัยของ Rowley et al. (1998) นำเสนอระบบค้นหาภาพใบหน้าโดยใช้โครงข่ายประสาทเทียม ซึ่งมีการใช้โครงข่ายประสาทเทียมหลายโครงข่ายช่วยในการทำงาน โครงสร้างการทำงานของระบบแสดงดังรูปที่ 2.9



รูปที่ 2.9 โครงสร้างการทำงานของระบบค้นหาภาพใบหน้าโดยใช้โครงข่ายประสาทเทียม

จากรูปที่ 2.9 ในส่วนแรกจะเป็นการแก้ปัญหาภาพเอียงในมุมต่าง ๆ โดยให้ชื่อว่าเครือข่ายเรเตอร์ (Router) การทำงานในส่วนนี้จะเป็นการนำภาพใบหน้าประมาณ 1,050 ภาพ มาปรับเปลี่ยนขนาด โดยมีอัตราการปรับเปลี่ยนขนาดจาก 1 ถึง 1.2 หมุนภาพ และเปลี่ยนตำแหน่งด้วยการซูม ทำให้เกิดภาพใบหน้าที่มีการเปลี่ยนแปลง 15,720 ภาพ นำภาพใบหน้าเหล่านี้ไปรู้จำโดยใช้โครงข่ายประสาทเทียมแบบแพร่กลับ ที่มีโครงสร้างชั้นรับข้อมูล 400 หน่วย ชั้นซ่อน 15 หน่วย และชั้นผลลัพธ์ 36 หน่วย จากนั้นหากมีภาพที่ต้องการจะค้นหาให้นำภาพนั้นมาทดสอบก่อนว่าภาพนั้นมีการเอียงเกิดขึ้นหรือเปล่าโดยใช้โครงข่ายเรเตอร์ หากมีการเอียงเกิดขึ้นให้ทำการปรับภาพนั้นให้เป็นภาพหน้าตรงเพื่อจะได้นำไปทดสอบในขั้นตอนต่อไป ในส่วนที่สองเป็นขั้นตอนของการค้นหาภาพใบหน้า โดยเรียกส่วนนี้ว่า โครงข่ายค้นหา (Detector networks) ข้อมูลที่ใช้ในการรู้จำของโครงข่ายนี้ประกอบไปด้วย ภาพที่มีภาพใบหน้า และภาพที่ไม่มีภาพใบหน้า โดยให้ผลลัพธ์เป็น +1.0 เมื่อภาพนั้นเป็นภาพใบหน้า และผลลัพธ์เป็น -1.0 เมื่อภาพนั้นไม่ใช่ภาพใบหน้า

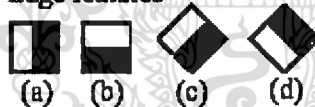
การทำงานดังรูปที่ 2.9 จะมีข้อเสียคือ ต้องมีภาพใบหน้าจำนวนมากเพื่อไปรู้จำ ทำให้ระบบทำงานได้ช้า และต้องมีหลายโครงข่ายเพื่อช่วงแบ่งการทำงาน และหากภาพที่จะนำมาค้นหาไม่ใช่ภาพที่มีการรู้จำมาก่อน อาจจะทำให้ไม่สามารถค้นหาภาพใบหน้านั้นได้

Viola et al. (2001) ได้นำเสนอวิธีการเชิงสถิติเพื่อใช้ในการการค้นหาวัดดู จากนั้นได้ถูกนำมาปรับปรุงเพิ่มเติมโดย Lienheart et al. (2002) โดยวิธีการนี้ได้ใช้เอกลักษณ์ฮาร์ (Haar-Feature) และ ตัวแยกประเภทแบบเรียงต่อกัน (Cascade Classifier) ซึ่งอยู่ในรูปของโมเดลทางสถิติ มาช่วยในการแยกเอกลักษณ์ของวัตถุที่ต้องการออกมาจากภาพ ซึ่งก่อนที่จะนำไปใช้ในการค้นหาวัดดูได้นั้นจะต้องมีการสอนให้ตัวแยกประเภททำการเรียนรู้ก่อนว่าต้องการค้นหาวัดดูชนิดใด เช่น ใบหน้ารถยนต์ เป็นต้น

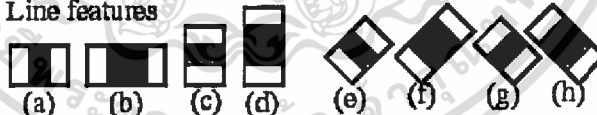
ตัวแยกประเภทจะทำการเรียนรู้ภาพใบหน้าจากรูปภาพที่ขนาดเท่ากันทั้งหมดเสียก่อนจึงจะใช้ตัวตัดแยกในการตรวจจับบริเวณที่เป็นใบหน้าได้ (Viola ใช้ขนาด 20×20 เพื่อใช้ในการตรวจจับใบหน้า) โดยกระบวนการตรวจจับนั้นจะทำการเลื่อนกรอบค้นหา (Search window) ว่าบริเวณใดในภาพที่มีลักษณะใกล้เคียงหรือเหมือนกับใบหน้าที่มากที่สุด สำหรับการค้นหาภาพใบหน้าที่มีขนาดแตกต่างกันนั้น สามารถกระทำได้เนื่องจากตัวแยกประเภทนั้นมีความสามารถในการหดหรือขยาย (scale) ได้เป็นอย่างดี

โครงสร้างของระบบนั้นจะใช้เอกลักษณ์ฮาร์ และ กลุ่มของตัวแยกประเภทที่อ่อนแอจำนวนมาก ซึ่งในแต่ละตัวแยกประเภทนั้นจะใช้เพียงเอกลักษณ์เดียวที่จะทำการตรวจสอบว่าบริเวณนี้เป็นใบหน้าที่หรือไม่เป็นใบหน้าที่ โดยลักษณะของเอกลักษณ์นั้นจะอยู่ในรูปของแม่แบบ (template) ดังแสดงในรูปที่ 2.10

1. Edge features



2. Line features



3. Center-surround features



รูปที่ 2.10 แสดงเอกลักษณ์ฮาร์ (Haar-like Feature)

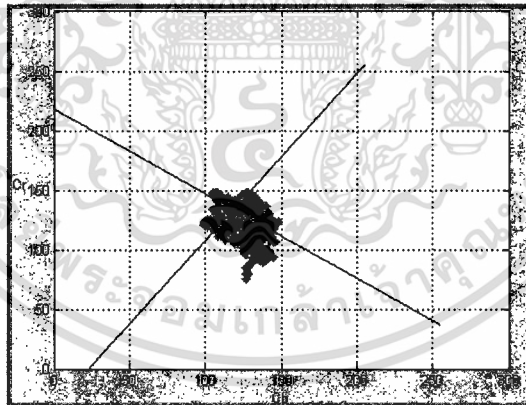
2.4 การตรวจจับสีผิว (Skin Color Detection)

เป้าหมายหลักของการตรวจจับสีผิว คือ การที่สามารถบอกได้ว่าบริเวณใดเป็นสีผิว และ บริเวณใด ไม่ใช่สีผิว ซึ่งวิธีการวิเคราะห์นั้นจะมีรายละเอียดดังนี้

2.4.1 การตรวจจับสีผิวโดยใช้แบบจำลองชนิดวงรี (Skin Detection With Ellipse Model) (Lee et al. 2002)

วิธีการนี้จะเป็นการสร้างโมเดลการกระจายตัวของผิวแบบมีตัวแปร (Parametric Skin Distribution Modeling) ทำให้โมเดลการกระจายของสีผิวที่สร้างขึ้นมีขนาดเล็ก และมีนัยสำคัญ (Generalize) และสามารถนำมาประมาณค่าในช่วง (Interpolate) ได้

การตรวจจับสีผิวจะใช้วิธีการตรวจจับสีผิวที่เรียกว่าการตรวจจับสีผิวโดยใช้แบบจำลองชนิดวงรี (Elliptic Boundary Model) โดยการสร้างโมเดลสีผิวเป็นลักษณะเป็นวงรีซึ่งการแยกแยะสีผิวโดยวิธีนี้จะนำปริภูมิสี YCbCr มาใช้ในการหาโดยปริภูมิสีนี้จะแก้ไขปัญหาในเรื่องความไม่แน่นอนของแสงที่ตกกระทบบนใบหน้าและมือซึ่งมีค่าเปลี่ยนแปลงอยู่ตลอดเวลา ซึ่งในการพิจารณาจะพิจารณาเพียงองค์ประกอบของสี (CbCr) โดยละทิ้งองค์ประกอบทางแสงสว่าง(Y) ดังแสดงในรูปที่ 2.11



รูปที่ 2.11 การกระจายตัวของบริเวณที่เป็นสีผิวและไม่เป็นสีผิว

จากรูปที่ 2.11 บริเวณที่เป็นสีจะเขียวจะเป็นบริเวณที่เป็นสีผิว ซึ่งเมื่อสังเกตดู จะพบว่ามีการกระจายตัวในลักษณะของวงรี ส่วนบริเวณที่เป็นสีแดงคือบริเวณที่ไม่ใช่สีผิว โดยสามารถสร้างแบบจำลองวงรีได้โดย

$$L_1 : AC_b + BC_r + C = 0 \tag{2.8}$$

$$L_2 : BC_b + AC_r + D = 0 \tag{2.9}$$

$$\frac{L_1^2}{a^2} + \frac{L_2^2}{b^2} = 1 \tag{2.10}$$

โดยที่ L_1 และ L_2 คือแกนเอกและแกนโทของวงรี ตามลำดับ
 A, B, C และ D สามารถหาได้จากค่าเฉลี่ย และค่าความแปรปรวนร่วมของ $C_b C_r$
 a และ b คือ ความยาวของแกนเอก และแกนโทของวงรีตามลำดับ

โดยการพิจารณาว่าจุดภาพใดในภาพเป็นสี่เหลี่ยมหรือไม่นั้นสามารถพิจารณาได้จากค่าของ

$C_b C_r$

$$MSkin(i, j) = \begin{cases} 1, & \text{if enclose with ellipse} \\ 0, & \text{otherwise} \end{cases} \tag{2.11}$$

โดยที่ $MSkin(i, j)$ คือ มาตรการของบริเวณที่เป็นสี่เหลี่ยม
 i, j คือ ตำแหน่งของจุดสี่ในภาพ

บทที่ 3

ทฤษฎีที่ใช้ในการวิจัย

ในบทนี้จะกล่าวถึงทฤษฎีพื้นฐานในเรื่องการประมวลผลภาพและหลักการทางคอมพิวเตอร์วิชัน รวมไปถึงโครงข่ายประสาทเทียมที่ต้องใช้ในวิทยานิพนธ์ฉบับนี้

3.1 มอร์โฟโลยีทางคณิตศาสตร์ (Mathematical Morphology) (Gonzalez and Woods, 1992)

ภาพของวัตถุที่ได้จากการสแกนด้วยเครื่องสแกน (Scanner) หรือได้จากการถ่ายภาพของกล้องดิจิทัล (Digital Camera) เพื่อนำมาประมวลผลภาพด้วยคอมพิวเตอร์นั้น ภาพที่ได้อาจจะมีสัญญาณรบกวน (Noise) เกิดขึ้น ซึ่งสัญญาณรบกวนที่เกิดขึ้นคือจุดภาพที่ไม่พึงประสงค์อาจเกิดได้จากหลายสาเหตุเช่น ความไม่สม่ำเสมอของแสงในการสแกน กระจัดที่วางภาพในการสแกนไม่สะอาด รอยนิ้วมือที่เกิดจากการสัมผัส ความผิดพลาดในการส่งผ่านข้อมูล เป็นต้น ซึ่งจะทำให้คุณภาพของภาพนั้นลดลงไป ดังนั้นก่อนที่จะทำการประมวลผลภาพต่อไปจึงต้องมีการวิเคราะห์เพื่อกำจัดสัญญาณรบกวนที่เกิดขึ้น

มอร์โฟโลยีทางคณิตศาสตร์เป็นวิธีที่นำหลักการทางคณิตศาสตร์คือทฤษฎีเซต มาใช้ในการนำเสนอรูปแบบหรือโครงสร้างของวัตถุในภาพ ซึ่งสามารถเป็นวิธีที่จะนำมาใช้ในการกำจัดสัญญาณรบกวนแบบเกลือและพริกไทย (Salt-and-pepper Noise) ที่เกิดขึ้นบนภาพไบนารีได้ โดยมีนิยามเบื้องต้นดังนี้

กำหนดให้ A และ B เป็นเซตของจุดภาพในเชิง 2 มิติ ซึ่งประกอบด้วยจุดภาพ $a = (a_1, a_2)$ และ $b = (b_1, b_2)$ ตามลำดับ

1. ทรานสเลชัน (Translation) การทำทรานสเลชันของ A โดย $x = (x_1, x_2)$ เขียนแทนด้วย $(A)_x$ เป็นการย้ายตำแหน่งจุดภาพที่อยู่ในเซต A ด้วยขนาด x มีนิยามดังนี้

$$(A)_x = \{c \mid c = a + x ; a \in A\} \quad (3.1)$$

2. รีเฟล็กชัน (Reflection) การทำรีเฟล็กชันของ A เขียนแทนด้วย \hat{A} เป็นการกลับตำแหน่งจุดภาพที่อยู่ในเซต A มีนิยามดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับ $\hat{A} = \{x \mid x = -a ; a \in A\}$ ไม่อนุญาตให้นำไปใช้ประโยชน์ (3.2) คำ
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

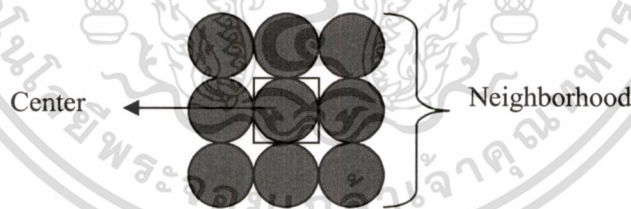
3. คอมพลิเมนต์ (Complement) การทำคอมพลิเมนต์ของเซต A เขียนแทนด้วย A^C เป็นการกำหนดตำแหน่งจุดภาพที่ไม่ได้อยู่ในเซต A มีนิยามดังนี้

$$A^C = \{x \mid x \notin A\} \quad (3.3)$$

4. ผลต่าง (Difference) การทำผลต่างของเซต A และเซต B เขียนแทนด้วย $A - B$ เป็นการกำจัดจุดภาพบางส่วนในเซต A ซึ่งเป็นจุดภาพที่อยู่ในเซต B มีนิยามดังนี้

$$A - B = \{x \mid x \in A ; x \notin B\} = A \cap B^C \quad (3.4)$$

หลักการงานของมอร์โฟโลยีทางคณิตศาสตร์จะพิจารณาเฉพาะกลุ่มสีที่เป็นวัตถุในภาพตามที่ได้กำหนดไว้ นั่นคืออาจจะพิจารณาเฉพาะกลุ่มจุดภาพดำซึ่งอยู่บนพื้นหลังสีขาว หรืออาจจะพิจารณาเฉพาะกลุ่มจุดภาพขาวซึ่งอยู่บนพื้นหลังสีดำ ในแต่ละกลุ่มจุดภาพอาจจะมีสัญญาณรบกวนเกิดขึ้น จึงควรจะมีการกำจัดสัญญาณรบกวนที่เกิดขึ้นนี้ออกไปก่อนที่จะนำภาพไปประมวลผลต่อไป ในการพิจารณาการกำจัดสัญญาณรบกวนนั้นจะต้องมีการกำหนดสตรักเจอร์อีลิเมนต์ (Structure Element) ขึ้นมาช่วยในการพิจารณา กำหนดให้เป็นสตรักเจอร์อีลิเมนต์ B มีตัวอย่างดังต่อไปนี้



รูปที่ 3.1 ตัวอย่างสตรักเจอร์อีลิเมนต์ขนาด 3x3 พิกเซล

จากตัวอย่างสตรักเจอร์อีลิเมนต์ B ดังแสดง เป็นโครงสร้างจุดภาพดำที่มีขนาด 3x3 พิกเซล โดยจะนำจุดกลาง (Center) และจุดรอบข้าง (Neighborhood) ของสตรักเจอร์อีลิเมนต์ B มาพิจารณาร่วมกับวัตถุที่เป็นเซตของจุดภาพดำ

การทำงานโดยพื้นฐานของมอร์โฟโลยีจะประกอบไปด้วย อีรอสัน (Erosion) และไดเลชัน (Dilation) ซึ่งมีการทำงานดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

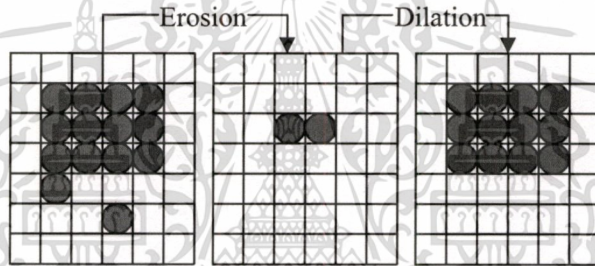
ในการเพิ่มประสิทธิภาพของมอร์โฟโลยีทางคณิตศาสตร์สามารถรวมการทำงานของทั้ง 2 พื้นฐานเข้าด้วยกัน ทำให้เกิดวิธีการโอเพนนิ่ง (Opening) และ โคลสซิง (Closing) ซึ่งมีการทำงานดังนี้

3.1.3 โอเพนนิ่ง (Opening)

การทำโอเพนนิ่งของเซต A กับสตรักเจอร์อีลิเมนต์ B เขียนแทนด้วย $A \circ B$ มีนิยามดังนี้

$$A \circ B = (A \ominus B) \oplus B \quad (3.7)$$

เป็นการทำงานร่วมกันระหว่างการทำอีรอสันและไคเลชันตามลำดับ คือจะเริ่มจากการทำอีรอสันโดยลดจำนวนจุดภาพดำ แล้วตามด้วยการทำไคเลชันโดยเพิ่มจำนวนจุดภาพดำลงไป ในภาพให้เท่ากับจำนวนรอบของการทำอีรอสัน ดังตัวอย่างต่อไปนี้



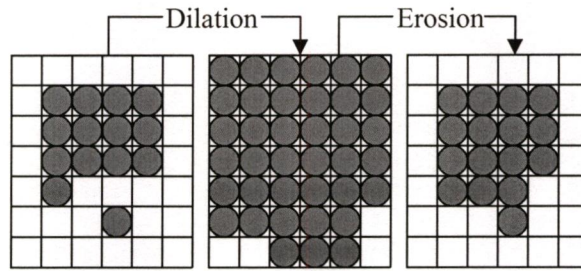
รูปที่ 3.4 การทำโอเพนนิ่ง $A \circ B$

3.1.4 โคลสซิง (Closing)

การทำโคลสซิงของเซต A กับสตรักเจอร์อีลิเมนต์ B เขียนแทนด้วย $A \bullet B$ มีนิยามดังนี้

$$A \bullet B = (A \oplus B) \ominus B \quad (3.8)$$

เป็นการทำงานที่มีลำดับตรงข้ามกับการทำโอเพนนิ่ง โดยจะทำไคเลชันและอีรอสันตามลำดับคือจะเริ่มจากการทำไคเลชันโดยเพิ่มจำนวนจุดภาพดำ แล้วตามด้วยการทำอีรอสันโดยลดจำนวนจุดภาพดำของภาพให้เท่ากับจำนวนรอบของการทำไคเลชัน ดังตัวอย่างต่อไปนี้

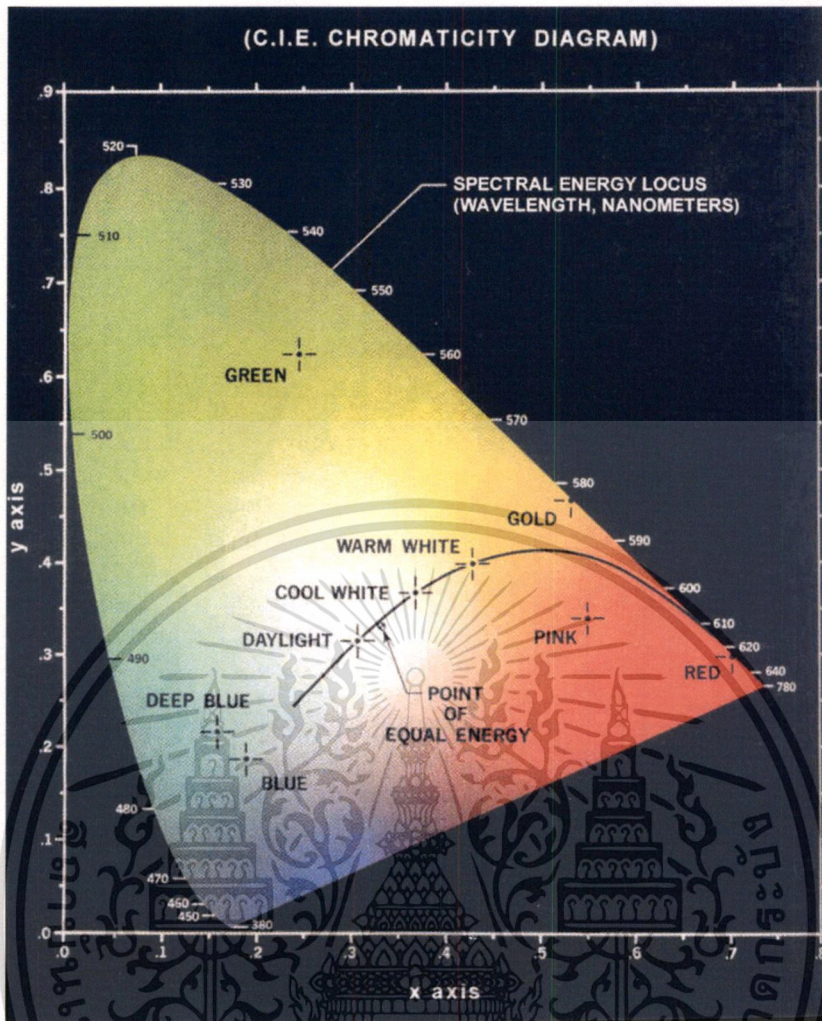


รูปที่ 3.5 การทำโคลสซิง $A \cdot B$

จากวิธีข้างต้นอาจยังไม่ครอบคลุมในการกำจัดสัญญาณรบกวนแบบเกลือและพริกไทย จึงมีการพัฒนาผลลัพธ์ที่ได้จากการทำโอเพนนิ่งและโคลสซิง โดยนำทั้งสองการทำงานนี้มาทำงานร่วมกันคือ อาจจะทำโอเพนนิ่งก่อนแล้วตามด้วยโคลสซิง หรือในทางตรงกันข้าม อาจจะทำโคลสซิงก่อนแล้วตามด้วยโอเพนนิ่ง ทำให้ผลลัพธ์ที่ได้มีประสิทธิภาพมากขึ้น

3.2 แบบจำลองสี (Color Model) (Gonzalez and Woods, 1992)

สีเป็นคุณสมบัติทางกายภาพอย่างหนึ่งของวัตถุ เกิดจากการรับรู้สัญญาณในรูปของคลื่นแสงที่ตกกระทบตามนุษย์และส่งสัญญาณผ่านประสาทตาไปยังสมอง จากนั้นจึงแปลงค่าเป็นค่าสีต่างๆ ในระบบสีจะมีแบบจำลองสีหลายรูปแบบให้เลือกใช้ตามความเหมาะสมเช่น แบบจำลองสี XYZ ซึ่งเป็นแบบจำลองสีมาตรฐานที่พัฒนาขึ้นโดย CIE (Commission Internationale de l'Eclairage) ถูกใช้เป็นตัวกลางในการอ้างอิงและแปลงค่าระหว่างแบบจำลองสีทั้งหมด ในงานทางด้านกราฟิกจะใช้แบบจำลองสี CMYK ในการแสดงภาพบนหน้าจอคอมพิวเตอร์จะใช้แบบจำลองสี RGB ในงานทางด้านคอมพิวเตอร์กราฟิกหรืองานที่ต้องคำนึงในด้านของศิลปะจะใช้แบบจำลองสี HSI หรือมาตรฐานในการวัดสีเช่นเครื่องมือในการวัดสีของผลิตภัณฑ์ทางการเกษตรจะใช้แบบจำลองสี $L^*a^*b^*$ เป็นต้น ในงานวิจัยนี้ได้มีการศึกษาแบบจำลองสีต่างๆ ดังต่อไปนี้



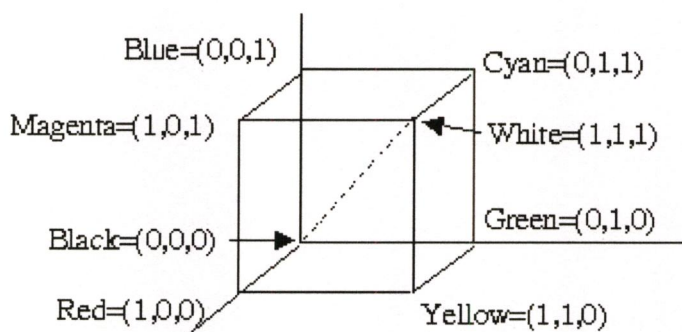
รูปที่ 3.6 โครมาติคิตี ไดอะแกรม

3.2.1 แบบจำลองสี RGB

แบบจำลองสี RGB เป็นระบบสีที่ใช้กันทั่วไปจะประกอบไปด้วยแม่สีหลัก 3 สีคือ แดง (R) เขียว (G) และน้ำเงิน (B) ซึ่งแต่ละแม่สีจะมีขนาดเป็น 8 บิต ดังนั้นในความลึกของแต่ละพิกเซลจะมีขนาดเป็น 24 บิต ซึ่งในแต่ละแม่สีจะมีค่าตั้งแต่ 0 ถึง 255 ดังนั้นสามารถที่จะนำมาสร้างสีได้ถึง $256 \times 256 \times 256$ เท่ากับ 16,777,216 สี

แบบจำลองสี RGB สามารถนำมาแสดงในระบบพิกัดปริภูมิสี (Color Space) ในแบบลูกบาศก์ได้ดังรูปที่ 3.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.7 แบบจำลองสี RGB

แต่อย่างไรก็ตามปริภูมิสี RGB เป็นข้อมูลของสีในแต่ละค่าของ R G หรือ B เป็นการผสมกันระหว่างค่าสีและค่าความเข้มของแสง ซึ่งสีในลักษณะนี้ไม่ค่อยเป็นที่นิยมใช้ในงานที่เป็นการวิเคราะห์และวิธีการรู้จำโดยใช้สี

3.2.2 ปริภูมิสีนอร์มัลไลซ์ RGB (rgb)

ปัญหาของปริภูมิสี RGB คือการรวมกันของความสว่างสี โดยปริภูมิสีนี้มีวัตถุประสงค์เพื่อตัดความสนใจขององค์ประกอบความสว่างทิ้งไป โดยหากค่าจุดภาพในปริภูมิสี RGB นั้นมีสีเดียวกันแต่ความสว่างต่างกัน จะถูกกำหนดเป็นค่าเดียวกันในปริภูมิสีนอร์มัลไลซ์ RGB โดยแต่ละค่าสามารถหาได้จาก

$$r = \frac{R}{(R+G+B)} \quad (3.9)$$

$$g = \frac{G}{(R+G+B)} \quad (3.10)$$

$$b = \frac{B}{(R+G+B)} \quad (3.11)$$

- ตัวอย่างเช่น
1. ค่าในปริภูมิสี RGB (0,0,255) ทำการแปลงไปสู่ rgb จะได้(0,0,1)
 2. ค่าในปริภูมิสี RGB (0,0,100) ทำการแปลงไปสู่ rgb จะได้(0,0,1)

ซึ่งจะสังเกตได้ว่าค่าในปริภูมิสีของตัวอย่างมีเฉพาะองค์ประกอบของสีน้ำเงินเพียงอย่างเดียวแต่ค่าความเข้มของแสงต่างกัน เมื่อลองมองดูปริภูมิสี rgb จะเห็นได้ว่าองค์ประกอบทางแสงจะถูกละทิ้งไปทำให้ค่าทั้ง 2 นี้เป็นค่าเดียวกันดังแสดงในรูปที่ 3.8



RGB(0,0,255) -> rgb(0,0,1)



RGB(0,0,100) -> rgb(0,0,1)

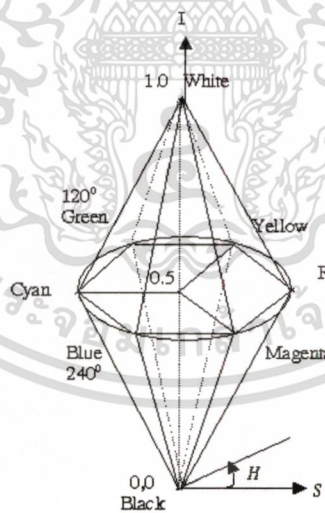
รูปที่ 3.8 ความแตกต่างของแสงในปริภูมิสี rgb จะถูกละทิ้ง

3.2.3 แบบจำลองสี HSI

ค่าสี (Hue) ความอิ่มสี (Saturation) และความเข้มสี (Intensity) เป็นคุณสมบัติหนึ่งที่จะนำมาใช้ในการแสดงสี ซึ่งเป็นแบบจำลองสีที่สื่อความหมายในการรับรู้ของมนุษย์ได้ดีกว่าแบบจำลองสี RGB

แบบจำลองสี HSI สามารถนำมาแสดงในระบบพิกัดปริภูมิสีด้วยพิกัดทรงกรวย ได้ดังรูปที่

3.9



รูปที่ 3.9 แบบจำลองสี HIS

H ใช้แสดงสีที่ต่างกัน ซึ่งสีต่างๆ ถูกนำมาจัดด้วยการสร้างเป็นวงล้อสี โดยกำหนดค่าสีด้วยค่ามุม ซึ่งมีค่าตั้งแต่ 0 ถึง 360 องศา โดยที่ H=0 คือค่าสีแดง H=120 คือค่าสีเขียว H=240 คือค่าสีน้ำเงิน เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

S ใช้กำหนดความอิ่มสี เป็นระดับความสดหรือความหมองที่เกิดขึ้นจากการผสมกันระหว่างสีที่อยู่ตรงข้ามในวงล้อสี โดยระดับสีที่สดที่สุดคือสีที่ไม่ถูกนำมาผสมเลข $S=1$ และระดับสีที่หมองที่สุดคือสีที่ถูกผสมกันในอัตราส่วนครึ่งต่อครึ่งระหว่างสีที่อยู่ตรงข้ามกัน ในวงล้อสี $S=0$

I ใช้กำหนดความมืดหรือความสว่างของสี โดยระดับที่สว่างที่สุดของทุกสีคือสีขาว $I=1$ และระดับที่มืดที่สุดของทุกสีคือสีดำ $I=0$

ในการแปลงค่าสีจากแบบจำลองสี RGB เป็นแบบจำลองสี HSI สามารถทำได้ดังสมการต่อไปนี้

$$I = \frac{1}{3}(R + G + B) \quad (3.12)$$

$$S = 1 - \frac{3 \min(R, G, B)}{R + G + B} \quad (3.13)$$

$$H = \cos^{-1} \left[\frac{\frac{1}{2}[(R - G) + (R - B)]}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right] \quad (3.14)$$

เมื่อค่า B มากกว่าค่า G จะกำหนดให้ $H = 360 - H$

และในการแปลงค่าสีจากแบบจำลองสี HSI เป็นแบบจำลองสี RGB สามารถทำได้โดยแยกกรณีตามค่า H ดังต่อไปนี้
กรณีที่ $0^\circ \leq H < 120^\circ$

$$R = \frac{I}{\sqrt{3}} \left[1 + \frac{S \cos(H)}{\cos(60^\circ - H)} \right] \quad (3.15)$$

$$G = \sqrt{3}I - R - B \quad (3.16)$$

$$B = \frac{I}{\sqrt{3}}(1 - S) \quad (3.17)$$

กรณีที่ $120^\circ \leq H < 240^\circ$

$$R = \frac{I}{\sqrt{3}}(1 - S) \quad (3.18)$$

$$G = \frac{I}{\sqrt{3}} \left[1 + \frac{S \cos(H - 120^\circ)}{\cos(180^\circ - H)} \right] \quad (3.19)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$B = \sqrt{3}I - R - G \quad (3.20)$$

กรณีนี้ $240^\circ \leq H < 360^\circ$

$$R = \sqrt{3}I - G - B \quad (3.21)$$

$$G = \frac{I}{\sqrt{3}}(1-S) \quad (3.22)$$

$$B = \frac{I}{\sqrt{3}} \left[1 + \frac{S \cos(H - 240^\circ)}{\cos(300^\circ - H)} \right] \quad (3.23)$$

3.2.4 ปริภูมิสี YCbCr

ปริภูมิสี YCbCr เป็นปริภูมิสีดิจิทัล และมีการกำหนดมาตรฐานโดย BT.601 ได้จากการแปลงองค์ประกอบ YUV ซึ่งเป็นสัญญาณแบบอนาล็อก ซึ่งจะมีการแบ่งข้อมูลของสีออกในลักษณะที่สอดคล้องกับการรับรู้ของมนุษย์เช่นเดียวกับปริภูมิ HIS โดยที่ Y เป็นองค์ประกอบความสว่างมีค่าตั้งแต่ 16-235 CbCr คือองค์ประกอบสีซึ่งเป็นค่าความแตกต่างของสีน้ำเงินกับค่าความสว่าง และ สีแดงกับค่าความสว่างตามลำดับโดยที่ Cb และ Cr มีค่าตั้งแต่ 16-240 โดยในการแปลงปริภูมิสี RGB ไปยัง YCbCr ทำได้โดย

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2568 & 0.5041 & 0.0980 \\ -0.1482 & -0.2910 & 0.4392 \\ 0.4392 & -0.3678 & -0.0714 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (3.24)$$

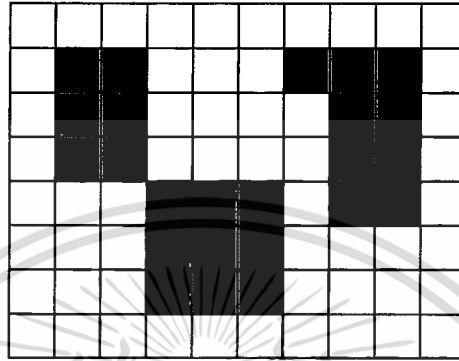
และการแปลงกลับจากปริภูมิสี YCrCb ไปสู่ RGB ทำได้โดย

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.1643 & -0.0018 & 1.5958 \\ 1.1643 & -0.3914 & -0.8135 \\ 1.1643 & 2.0178 & -0.0012 \end{bmatrix} \begin{bmatrix} Y-16 \\ Cb-128 \\ Cr-128 \end{bmatrix} \quad (3.25)$$

3.3 การวิเคราะห์องค์ประกอบที่เชื่อมต่อกันของภาพ (Connected Components) (Lohmann, 1998)

ในภาพๆ หนึ่งอาจจะมียอดองค์ประกอบหรือวัตถุอยู่วัตถุเดียวหรือมากกว่าหนึ่งวัตถุขึ้นไป โดยในแต่ละวัตถุอาจจะอยู่ในตำแหน่งที่ต่อเนื่องกัน ซ้อนทับกัน หรือแยกออกจากกัน ซึ่งในภาพที่เป็นแบบไบนารีจะมีการพิจารณาการเชื่อมต่อของวัตถุโดยพิจารณาจากพิกเซลที่เป็นส่วนของวัตถุ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนูญาติเห็นไปเซปรีเซชันด้านกรรค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กับพิกเซลรอบข้าง (Neighborhoods) ว่ามีการเชื่อมต่อกันหรือไม่ ซึ่งสามารถพิจารณาได้ทั้งใน 4 ทิศทาง (4-connected Neighborhoods) หรือใน 8 ทิศทาง (8-connected Neighborhoods) ซึ่งถ้าพิกเซลที่เป็นส่วนของวัตถุมีการเชื่อมต่อกันจะถือว่าเป็นวัตถุเดียวกัน แต่ถ้าพิกเซลที่เป็นส่วนของวัตถุแยกออกจากกันหรือไม่เชื่อมต่อกันจะกำหนดให้เป็นวัตถุที่ต่างกัน ดังแสดงในรูปที่ 3.10



รูปที่ 3.10 องค์ประกอบของภาพ

จากรูปที่ 3.10 นั้นกำหนดให้พิกเซลสีขาวเป็นส่วนของพื้นหลัง และพิกเซลสีดำเป็นส่วนของวัตถุ เมื่อนำภาพมาวิเคราะห์องค์ประกอบที่เชื่อมต่อกันของภาพทำให้สามารถระบุหมายเลข (Labeling of Connected Components) ให้กับองค์ประกอบของภาพได้ดังรูปที่ 3.11

0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	3	3	3	0
0	1	1	0	0	0	0	3	3	0
0	1	1	0	0	0	0	3	3	0
0	0	0	2	2	2	0	3	3	0
0	0	0	2	2	2	0	0	0	0
0	0	0	2	2	2	0	0	0	0
0	0	0	0	0	0	0	0	0	0

รูปที่ 3.11 การระบุหมายเลขให้กับองค์ประกอบของภาพโดยพิจารณาใน 4 ทิศทาง

จากรูปที่ 3.11 เป็นผลการระบุหมายเลขให้กับองค์ประกอบของภาพ ซึ่งภาพจะประกอบไปด้วย 4 องค์ประกอบด้วยกัน โดยหมายเลข 0 แสดงส่วนของพื้นหลัง หมายเลข 1 แสดงวัตถุที่หนึ่ง หมายเลข 2 แสดงวัตถุที่สอง และหมายเลข 3 แสดงวัตถุที่สามของภาพ ซึ่งจากผลที่ได้ข้างต้นเป็นการวิเคราะห์องค์ประกอบที่เชื่อมต่อกันของภาพโดยใช้ 4 ทิศทาง แต่ถ้าเป็นการวิเคราะห์โดยใช้ 8 ทิศทาง จะทำให้ภาพข้างต้นประกอบไปด้วย 2 วัตถุเท่านั้น นั่นคือวัตถุที่หนึ่งและวัตถุที่สองเป็นวัตถุที่เชื่อมต่อกันในแนวทแยงจึงถือว่าเป็นวัตถุเดียวกัน

3.4 การแยกวัตถุออกจากฉากหลังในภาพวีดิทัศน์

กล้องที่ใช้ในระบบภาพวีดิทัศน์นั้นจะทำการรับภาพเข้ามาเป็นภาพนิ่งที่ต่อเนื่องกันเป็นลำดับเรียงกันไป จึงทำให้ภาพที่ได้เป็นภาพที่มีการเคลื่อนไหว โดยแต่ละภาพจะมีลักษณะที่แตกต่างกันไป การที่ภาพมีความแตกต่างกันนั้นอาจจะเกิดได้จากการที่วัตถุเคลื่อนไหว หรือกล้องมีการเคลื่อนไหว การเปลี่ยนแปลงของแสง เป็นต้น โดยปกติแล้วภาพแต่ละภาพก็จะมีวัตถุประกอบอยู่ในภาพ การที่จะแยกวัตถุที่ปรากฏอยู่ในภาพออกจากฉากหลังนั้น สามารถกระทำได้หลายวิธี โดยขึ้นอยู่กับสภาพแวดล้อมที่ใช้ และขึ้นอยู่กับลักษณะของกล้องที่ติดตั้งว่าเป็นอย่างไร รวมไปถึงลักษณะของฉากหลังด้วยว่าเป็นฉากหลังที่มีการเคลื่อนไหวหรือไม่ โดยสามารถแบ่งออกได้เป็น 4 ประเภทดังนี้

กล้องอยู่กับที่, วัตถุอยู่กับที่
 กล้องอยู่กับที่, วัตถุเคลื่อนไหว
 กล้องเคลื่อนไหว, วัตถุอยู่กับที่
 กล้องเคลื่อนไหว, วัตถุเคลื่อนไหว

การวิเคราะห์ภาพในแต่ละประเภทที่ได้กล่าวไปนั้นก็จะต้องใช้กระบวนการวิเคราะห์ที่หลากหลายแตกต่างกันไปตามความซับซ้อนของภาพที่ได้ การวิเคราะห์ภาพในแบบแรกนั้นจะมีความยุ่งยากน้อยที่สุด และในแบบสุดท้ายจะมีความยุ่งยากในการวิเคราะห์มากที่สุด

กระบวนการที่ใช้ในการวิเคราะห์ข้อมูลวีดิทัศน์เพื่อสกัดวัตถุที่สนใจออกจากฉากหลังที่นิยมใช้กันอย่างแพร่หลายมีดังนี้

3.4.1 เฟรมดิฟเฟอเรนซ์ซิง (Frame Differencing)

วิธีการนี้เป็นวิธีที่ใช้ในการตรวจจับการเปลี่ยนแปลงที่เกิดขึ้นในภาพ โดยใช้การเปรียบเทียบระหว่างภาพสองภาพว่ามีจุดภาพ (pixel) ใดในภาพทั้งสองที่มีความแตกต่างกันหรือเหมือนกัน โดยนำจุดที่ตำแหน่งเดียวกันของทั้งสองภาพมาทำการลบกัน ผลลัพธ์ที่ได้จะแสดงส่วนที่มีความแตกต่างกันในรูปแบบภาพไบนารี คือมีแค่สองสีคือ ขาว กับ ดำ ถ้าผลลัพธ์ที่ได้มีค่ามากกว่าค่า ชิดแบ่ง (threshold) จะมีค่าเป็น 1 แทนด้วยสีขาว ซึ่งเกิดจากการที่วัตถุมีการเคลื่อนที่ หรือมีการเปลี่ยนแปลงของแสง เป็นต้น ส่วนผลลัพธ์ที่มีค่าน้อยกว่าค่าชิดแบ่ง ที่ทำให้เกิดการตอบสนอง จะมีค่าเป็น 0 แทนด้วยสีดำ เนื่องจากบริเวณที่เป็นฉากหลังจะไม่ค่อยมีความเปลี่ยนแปลง เมื่อนำมาลบกันก็จะมีค่าน้อยกว่าค่าชิดแบ่ง จึงทำให้ฉากหลังของภาพถูกลบออกไป ดังแสดงในสมการที่ 3.26 แต่จะมีข้อเสียคือ ถ้าวัตถุนั้นเคลื่อนที่อย่างช้า ๆ หรือแสงค่อย ๆ เปลี่ยนไป วิธีการนี้จะไม่สามารถตรวจจับการเคลื่อนไหวได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

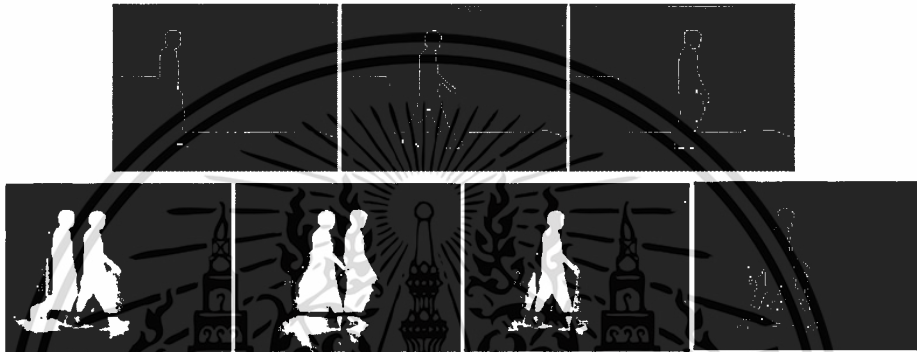
$$I_n = \begin{cases} 1 & \text{if } |I_{t+1} - I_t| > \tau \\ 0 & \text{otherwise} \end{cases} \quad (3.26)$$

โดยที่ I_n = รูปใหม่ที่ได

I_t = รูป ณ เวลา t

I_{t+1} = รูป ณ เวลา t+1

τ = ค่าขีดแบ่งที่ทำให้เกิดการตอบสนอง



รูปที่ 3.12 แสดงการทำเฟรมดิฟเฟรนซ์ซึ่ง

จะเห็นได้ว่าภาพที่ได้จะเกิดการเหลื่อมล้ำระหว่างภาพทั้งสอง โดยเกิดจากการที่ภาพมีการเคลื่อนไหวหรือเกิดความเปลี่ยนแปลง การที่จะทำให้ภาพที่เกิดการเหลื่อมล้ำนั้นถูกลบออกไปสามารถกระทำได้โดยการนำภาพผลลัพธ์ภาพที่หนึ่งมาทำกระบวนการทางตรรกะ ในที่นี้คือ AND โดย AND กับภาพผลลัพธ์ภาพที่สอง ก็จะได้ภาพผลลัพธ์สุดท้ายที่ไม่มีการเหลื่อมล้ำของภาพเกิดขึ้น ซึ่งภาพผลลัพธ์ที่ได้ก็คือภาพที่สองที่ภาพของฉากหลังได้ถูกลบออกไป โดยที่อาจจะมีส่วนที่เป็นภาพฉากหลังเหลืออยู่บางพอสมควร จากนั้นจึงนำภาพที่ 2 มาแทนลงในส่วนที่เป็นสีขาว ก็จะได้ภาพผลลัพธ์สุดท้ายขึ้นมา ดังแสดงในรูปที่ 3.12

3.4.2 การลบฉากหลัง (Background Subtraction) (Horprasert. 1999)

วิธีการลบฉากหลังเป็นวิธีการที่ได้รับความนิยมวิธีหนึ่ง โดยมีขั้นตอนการทำงานหลัก ๆ ดังนี้ การสร้างแบบจำลองฉากหลัง หรือ Background Modeling เพื่อใช้เป็นภาพอ้างอิง , การกำหนดค่าขีดแบ่ง (Threshold) ที่เหมาะสม เพื่อใช้ในการลบฉากหลัง ในส่วนสุดท้ายจะเป็นการลบฉากหลังออกจากภาพ โดยใช้ค่าขีดแบ่งที่ได้เป็นค่าประกอบในการตัดสินใจ

เนื่องจากสภาพแวดล้อมมีการเปลี่ยนแปลงอยู่ตลอดเวลา เช่น แสงสว่าง ดังนั้นถ้าในสภาพแวดล้อมที่เป็นบริเวณฉากหลังนั้นมีการเปลี่ยนแปลงอาจทำให้กระบวนการลบฉากหลังมี

เอกสารความผิดพลาดได้ ซึ่งสามารถแบ่งออกได้เป็น 2 ประเภทคือ การเปลี่ยนแปลงของแสงสว่างโดยรวมราคา ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

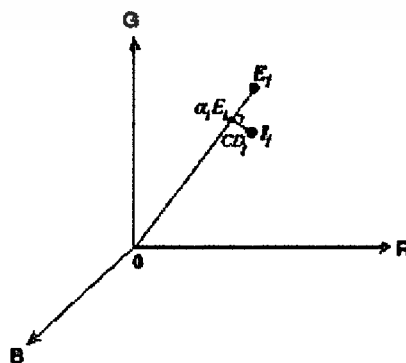
(Global Illumination Change) และ การเปลี่ยนแปลงของแสงสว่างเฉพาะที่ (Local Illumination Change)

- การเปลี่ยนแปลงของแสงสว่างโดยรวม คือ การที่แสงภายในพื้นที่ที่เฝ้าสังเกต เกิดการเปลี่ยนแปลงโดยมีผลกระทบต่อส่วนใหญ่ของพื้นที่ที่เฝ้าสังเกต เช่นการเปิดปิดไฟ เป็นต้น ซึ่งจะมีผลต่อการลบฉากหลัง
- การเปลี่ยนแปลงของแสงสว่างเฉพาะที่ คือ การที่แสงภายในพื้นที่ที่เฝ้าสังเกตเกิดการเปลี่ยนแปลง โดยเปลี่ยนเพียงเล็กน้อยจากพื้นที่ทั้งหมด หรืออีกนัยหนึ่งคือการเปลี่ยนแปลงของแสงที่ตกกระทบพื้นผิวของวัตถุที่เฝ้าสังเกต เกิดขึ้นเฉพาะบางส่วนหรือเฉพาะจุด เช่น เกิดมีเงามาบังเพียงบางจุด หรือมีแสงสะท้อนจากแหล่งแสงอื่นมาตกกระทบ เป็นต้น

งานวิจัยนี้ได้นำเสนอกระบวนการลบฉากหลัง โดยได้นำเสนอแบบจำลองสี ซึ่งเป็นการแยกองค์ประกอบการบิดเบือนของสีที่มองเห็นออกเป็น 2 องค์ประกอบได้แก่ การบิดเบือนของสีของวัตถุ (Chromaticity or Color Distortion) และการบิดเบือนของความสว่าง (Brightness Distortion) ซึ่งสามารถแก้ปัญหาที่ได้กล่าวไปแล้วได้เป็นอย่างดี

3.4.2.1 แบบจำลองสี (Color Model)

แบบจำลองสี (Color Model) นี้จะอยู่ในระบบสี RGB โดยประกอบไปด้วย 2 ส่วนที่สำคัญด้วยกันคือ E_i และ I_i ดังแสดงในรูปที่ 3.13 กำหนดให้ i แทน ลำดับจุดภาพ โดย $E_i = [E_R(i), E_G(i), E_B(i)]$ แทนค่าของสี RGB ของจุดภาพที่ i ในรูปภาพอ้างอิงหรือภาพที่เป็นฉากหลัง โดยเส้น OE_i คือเส้นที่ลากจากจุดโอริจิน (Origin) ผ่านจุด E_i เรียกว่าเส้นสีที่คาดการณ (Expected Chromaticity Line) ส่วน $I_i = [I_R(i), I_G(i), I_B(i)]$ แทนค่าของสี RGB ของจุดภาพที่ i ในรูปภาพรับเข้าที่กำลังสนใจอยู่ และที่ต้องการลบฉากหลังออก



รูปที่ 3.13 แสดงแบบจำลองสีในรูปแบบ 3 มิติในระบบสี RGB

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการพิจารณาความแตกต่างระหว่าง I_i และ E_i งานวิจัยนี้ได้คำนวณค่าความบิดเบือนทั้งสององค์ประกอบ ดังต่อไปนี้ คือ ค่าความบิดเบือนของความสว่าง (Brightness Distortion) และค่าความบิดเบือนของสี (Color Distortion) โดยมีรายละเอียดในการคำนวณดังนี้

Brightness Distortion (α)

ค่าความบิดเบือนของความสว่าง คือ ค่าปริมาณเชิงสเกลาร์ (scalar value) ซึ่งเป็นค่าที่ทำให้ค่าสีที่ผู้วัดสังเกตเห็นใกล้เคียงกับค่าสีจริงของวัตถุในแบบจำลองที่สุด ดังสมการที่ 3.27

$$\phi(\alpha_i) = (I_i - \alpha_i E_i)^2 \quad (3.27)$$

α_i คือ ค่าความเข้มของความสว่าง (Brightness) กล่าวคือ $\alpha_i = 1$ จะหมายถึงค่าความสว่าง (Brightness) ของจุดภาพที่นำมาเปรียบเทียบกับจุดภาพอ้างอิงมีค่าเท่ากัน ถ้า $\alpha_i < 1$ แสดงว่าจุดภาพที่นำมาเปรียบเทียบกับจุดภาพอ้างอิงจะมีค่าต่ำกว่าจุดภาพอ้างอิง และถ้า $\alpha_i > 1$ จุดภาพที่นำมาเปรียบเทียบกับจุดภาพอ้างอิงจะสว่างกว่าจุดภาพอ้างอิง

Color Distortion (CD)

ค่าความบิดเบือนของสี คือ ค่าระยะขจัดระหว่างค่าสีที่สังเกตอยู่ (observed color) กับเส้นสีที่คาดการณ์ (Expected Chromaticity Line) ดังสมการที่ 3.28

$$CD_i = \|I_i - \alpha_i E_i\| \quad (3.28)$$

3.4.2.2 การสร้างภาพแบบจำลองฉากหลัง (Background Modeling)

งานวิจัยนี้มีสมมติฐานที่ว่าฉากหลังเป็นชนิดที่ไม่มีเปลี่ยนแปลง (static background) เช่น ภาพของทางเดิน ในขณะที่ไม่มีบุคคลเดินไปมา โดยขั้นตอนในการสร้างแบบจำลองฉากหลังนั้นจะเริ่มจากนำค่าสีของจุดภาพต่าง ๆ มาทำการคำนวณโดยใช้จำนวนของภาพฉากหลัง N ภาพ ในการทำแบบจำลองฉากหลัง แบบจำลองฉากหลังที่ได้จะถูกเก็บให้อยู่ในรูปแบบทางสถิติ โดยจะกระทำทีละจุดภาพ โดยกำหนดให้แต่ละจุดภาพมี 4-tuple คือ $\langle E_i, s_i, a_i, b_i \rangle$ โดย E_i คือ ค่า Expected Color, s_i คือ ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation (SD)) ของสี a_i คือ ค่าแปรผันความสว่าง และ b_i คือ ค่าแปรผันของสีของจุดภาพ ที่ i โดย E_i สามารถหาได้ดังนี้

$$E_i = [\mu_R(i), \mu_G(i), \mu_B(i)] \quad (3.29)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่ $\mu_R(i)$, $\mu_G(i)$ และ $\mu_B(i)$ จะเป็นค่าเฉลี่ยของค่าสี R,G,B ที่จุดภาพที่ i โดยคำนวณจากภาพฉากหลังจำนวน N ภาพ แต่ในความเป็นจริงในการที่จะหาค่าของจุดภาพในช่วงระยะเวลาหนึ่งจะพบปัญหาที่เรียกว่าสิ่งรบกวนหรือ Noise เกิดขึ้น และการผันแปรของความสว่างจากแหล่งที่ให้แสงสว่าง แต่ค่าที่ผันแปรเหล่านี้สามารถจำลองได้โดยใช้ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation(SD)) ในแต่ละ Band (s_i) จะได้

$$s_i = [\sigma_R(i), \sigma_G(i), \sigma_B(i)] \quad (3.30)$$

โดยที่ $\sigma_R(i)$, $\sigma_G(i)$ และ $\sigma_B(i)$ คือ ค่า SD ของค่าสี R,G,B ที่จุดภาพที่ i โดยคำนวณจากภาพฉากหลังจำนวน N ภาพ

จากนั้นทำการนอร์มอลไลซ์ หรือ ทำการสมดุลช่วงค่าสี (Balance Color Band) จะได้สมการค่าความบิดเบือนของความสว่าง (Brightness Distortion) และค่าความบิดเบือนของสี (Color Distortion) จากสมการที่ 3.27 และ 3.28 จะเป็นดังนี้

$$\alpha_i = \min \left[\sum_{C=R,G,B} \left(\frac{I_c(i) - \alpha_i \mu_c(i)}{\sigma_c(i)} \right)^2 \right] \\ = \frac{\left(\frac{I_R(i)\mu_R(i)}{\sigma_R^2(i)} + \frac{I_G(i)\mu_G(i)}{\sigma_G^2(i)} + \frac{I_B(i)\mu_B(i)}{\sigma_B^2(i)} \right)}{\left(\left[\frac{\mu_R(i)}{\sigma_R(i)} \right]^2 + \left[\frac{\mu_G(i)}{\sigma_G(i)} \right]^2 + \left[\frac{\mu_B(i)}{\sigma_B(i)} \right]^2 \right)} \quad (3.31)$$

$$CD_i = \sqrt{\sum_{C=R,G,B} \left(\frac{I_c(i) - \alpha_i \mu_c(i)}{\sigma_c(i)} \right)^2} \quad (3.32)$$

เมื่อพิจารณาถึงค่าแปรผันความสว่างและค่าแปรผันของสี ที่เกิดขึ้นระหว่างการสร้างภาพแบบจำลองฉากหลัง จะพบว่าจุดภาพที่แตกต่างกันจะส่งผลให้เกิดผลการกระจายที่แตกต่างกันทั้งค่าของ α และ CD ดังแสดงในรูปที่ 3.14 (3.14a, 3.14b) ดังนั้นจึงมีความจำเป็นที่จะเปรียบเทียบแต่ละจุดภาพที่มาตรฐานเดียวกันโดยการหาค่าแปรผันความสว่าง a_i และค่าแปรผันของสี b_i ซึ่งเป็นค่าที่อยู่ในแบบ 4-Tuple $\langle E_i, s_i, a_i, b_i \rangle$ โดยค่าของความแปรผันนี้จะถูกฝังลงไปในการแบบจำลองฉากหลังทุก ๆ จุดภาพ



รูปที่ 3.14 แสดงค่าแปรผันความสว่าง(a) และ ค่าแปรผันของสี(b)

โดย α_i คือค่าแปรผันความสว่าง

$$\alpha_i = RMS(\alpha_i) = \sqrt{\frac{\sum_{i=0}^N (\alpha_i - 1)^2}{N}} \quad (3.33)$$

โดย b_i คือค่าแปรผันของสี

$$b_i = RMS(CD_i) = \sqrt{\frac{\sum_{i=0}^N (CD_i)^2}{N}} \quad (3.34)$$

3.4.2.2.1 การแบ่งแยกชนิดของจุดภาพและการลบจุดภาพ (Pixel Classification and Subtraction Operation)

ในการตอนนี้จะเป็นการคำนวณหาค่าความแตกต่างของภาพที่เป็นแบบจำลองจากหลังกับภาพรับเข้าปัจจุบัน ซึ่งความแตกต่างนั้นจะถูกแยกออกเป็นค่าความสว่าง และค่าสี โดยสามารถคำนวณได้โดยนำค่าขีดแบ่ง (Threshold) ที่เหมาะสม มาทำการเปรียบเทียบกับความบิดเบือนของค่าความสว่าง (α) และความบิดเบือนของค่าสี (CD) ที่จุดภาพ i จะทำให้ได้ หน้ากากจำแนก $M(i)$ ซึ่งจะเป็นตัวบ่งบอกชนิดของจุดภาพว่าเป็นแบบใด โดยสามารถจำแนกชนิดของจุดภาพ ของภาพปัจจุบัน ได้ดังนี้

1. ฉากหลัง (Original Background (B)) : ถ้าภาพทั้งสองมีค่าความสว่างและค่าสีที่จุดภาพเดียวกันมีความคล้ายคลึงกันจะถือว่าจุดภาพนั้นเป็นฉากหลัง
2. เงา (Shades Background or Shadow (S)) : ถ้าที่จุดภาพในภาพทั้งสองมีค่าสีคล้ายคลึงกัน แต่ค่าความสว่างของภาพปัจจุบันมีค่าน้อยกว่าภาพอ้างอิงที่จุดภาพเดียวกัน ซึ่งจะ

ถือว่าจุดภาพนั้นเป็นบริเวณที่เป็นฉากหลังที่มีแสงมาต่น้อยลงหรืออาจเป็นเงาจากวัตถุอื่น

3. ส่วนสว่าง (Highlighted Background (H)) : ถ้าภาพทั้งสองมีค่าสีที่คล้ายคลึงกันแต่ค่าความสว่างในภาพปัจจุบันสูงกว่าภาพอ้างอิง
4. วัตถุที่กำลังเคลื่อนที่ (Moving Foreground Object (F)) : ถ้าจุดภาพในภาพปัจจุบันนั้นมีค่าสีที่แตกต่างกับภาพอ้างอิง

จากที่ได้กล่าวมาแล้วว่าจุดภาพที่แตกต่างกันจะทำให้ผลของการกระจายของค่าความบิดเบือนของความสว่าง (α_i) และค่าความบิดเบือนของสี (CD_i) มีความแตกต่างกันขึ้นอยู่กับตำแหน่งจุดภาพ ดังนั้นเพื่อที่จะทำให้ใช้ค่าขีดแบ่ง (Threshold) เพียงค่าเดียวสำหรับทุก ๆ จุดภาพ จึงได้ทำการนอร์มอลไลซ์ค่า α_i และ CD_i จะได้

$$\hat{\alpha}_i = \frac{\alpha_i - 1}{a_i} \quad (3.35)$$

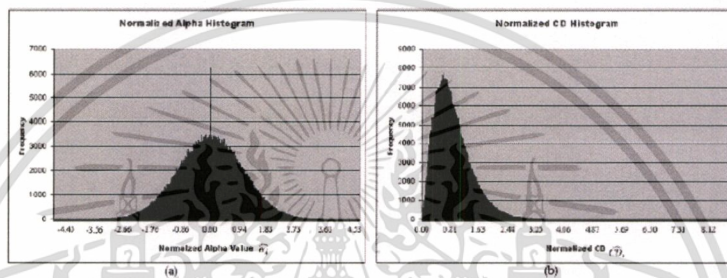
$$\hat{CD}_i = \frac{CD_i}{b_i} \quad (3.36)$$

3.4.2.2 การเลือกค่าขีดแบ่งโดยอัตโนมัติ (Automatic Threshold Selection)

โดยทั่วไปแล้วการกระจายของค่าความบิดเบือนจะถูกกำหนดให้อยู่ในรูปของการกระจายแบบเกาส์เซียน เพื่อการคำนวณหาค่าอัตราการตรวจจับที่ต้องการ (desired detection rate) r ซึ่งสามารถกำหนดค่าขีดแบ่งได้โดยใช้ค่า $K\sigma$ โดย K เป็นค่าคงที่เป็นจำนวนเท่า ซึ่งจะถูกกำหนดโดย r และ σ คือค่า เบี่ยงเบนมาตรฐาน (standard deviation) แต่จากผลการทดลองในงานวิจัยนี้จะพบว่าค่าความบิดเบือนของความสว่างแบบนอร์มอลไลซ์ (Normalized Brightness Distortion ($\hat{\alpha}_i$)) เท่านั้นที่มีการกระจายตัวแบบเกาส์ ดังแสดงในรูปที่ 3.15(a) แต่ค่าความบิดเบือนของสีแบบนอร์มอลไลซ์ (Normalized Color Distortion (\hat{CD}_i)) ไม่ได้มีการกระจายแบบเกาส์เซียน (Gaussian) ดังแสดงในรูปที่ 3.15(b) ดังนั้นจึงมีวิธีการคำนวณหาค่าขีดแบ่ง (threshold) โดยใช้หลักการทางสถิติ โดยเริ่มจากการทำฮิสโทแกรมของค่าความบิดเบือนของความสว่างแบบนอร์มอลไลซ์ (Normalized Brightness Distortion ($\hat{\alpha}_i$)) และ บิดเบือนของสีแบบนอร์มอลไลซ์ (Normalized Color Distortion (\hat{CD}_i)) ดังแสดงในรูปที่ 3.15 ซึ่งค่าของฮิสโทแกรมนั้นจะได้มาจากการรวบรวมชุด

เอกสาร
ข้อมูลระหว่างกระบวนการสร้างแบบจำลองฉากหลัง จำนวนชุดข้อมูลที่ได้จะมีจำนวนเป็น $N \times Y$ ค่า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพื่อใช้ในการสร้างฮิสโทแกรม โดยที่ขนาดของรูปภาพคือ $X \times Y$ และ N คือจำนวนภาพทั้งหมดที่ใช้ในการคำนวณ เมื่อสร้างฮิสโทแกรมได้สำเร็จแล้ว ค่าขีดแบ่งจะสามารถเลือกได้โดยอัตโนมัติ โดยขึ้นอยู่กับค่าอัตราการตรวจจับที่ต้องการ (desired detection rate) r โดยค่าขีดแบ่งสำหรับค่าความบิดเบือนของสี τ_{CD} คือค่าออร์มอลไลซ์ของค่าความบิดเบือนของสีที่ค่าอัตราการตรวจจับที่ต้องการ r ในกรณีของค่าความบิดเบือนของความสว่างจะมีค่าขีดแบ่งอยู่ 2 ค่า คือ $\tau_{\alpha 1}$ และ $\tau_{\alpha 2}$ โดยที่ $\tau_{\alpha 1}$ คืออัตราการตรวจจับที่ต้องการที่ (r) และ $\tau_{\alpha 2}$ คือค่าที่อัตราการตรวจจับที่ต้องการที่ $(1-r)$



รูปที่ 3.15 แสดงฮิสโทแกรม ของค่า Normalized Brightness Distortion (a) และแสดงฮิสโทแกรม ของค่า Normalized Color Distortion (b)

ผลที่ได้จากกระบวนการลบฉากหลังดังแสดงในรูปที่ 3.16



รูปที่ 3.16 แสดงการลบฉากหลังออกจากภาพ

3.4.3 การเคลื่อนที่ของภาพ (Image Flow)

การเคลื่อนที่ของรูปภาพ คือ การกระจายของอัตราความเร็วที่เกิดขึ้น โดยมีความสัมพันธ์กับผู้สังเกตการณ์บนจุดใด ๆ ในรูปภาพ โดย การเคลื่อนที่ของภาพ นั้นจะประกอบไปด้วยข้อมูลที่มีความสำคัญต่อการวิเคราะห์ฉากหลังที่มีการเคลื่อนไหว โดยที่การเคลื่อนที่ของภาพสามารถกำหนดได้จากอัตราความเร็วโดยแสดงอยู่ในรูปของเวกเตอร์ บนทุก ๆ จุดภาพในรูปภาพ ซึ่งวิธีการคำนวณหาค่า การเคลื่อนที่ของภาพบางกระบวนการจะยึดหลักการคำนวณโดยใช้รูปภาพที่เรียงกัน เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เป็นลำดับจำนวนสองภาพ หรือมากกว่านั้น ซึ่งวิธีการที่ใช้กระบวนการคำนวณแบบนี้สามารถแบ่งออกได้เป็น 2 วิธีคือ

3.4.3.1 Feature-Based Methods

วิธีการคำนวณหาการเคลื่อนที่ของภาพแบบการใช้เอกลักษณ์ (Feature-Based) ดังเช่น เส้นขอบ, มุม ที่ปรากฏในภาพ โดยในขั้นแรกจะทำการหาเอกลักษณ์ภายในภาพแต่ละภาพ จากนั้นจะทำการเปรียบเทียบเอกลักษณ์ทุกตัวที่ได้และทำการคำนวณหาค่าความไม่เหมือนกันต่างๆ ระหว่างภาพที่ได้ อย่างไรก็ตามในการที่จะเลือกเอกลักษณ์นั้นก็ไม่สามารถกระทำได้ง่าย เนื่องจากว่าวิธีการนี้จะให้ผลที่เป็นเวกเตอร์อัตราความเร็วอย่างกระจัดกระจาย

3.4.3.2 Gradient Based Methods

วิธีการแบบค่าความชัน (Gradient Based) นั้นจะใช้ประโยชน์จากความสัมพันธ์ระหว่าง เชิงเวลา (Temporal) และเชิงพื้นที่ (Spatial) ของ ค่าความชันของความเข้ม (Gradients of Intensity) ซึ่งความสัมพันธ์นี้สามารถนำมาใช้ในการแบ่งแยก (Segment) โดยใช้ Velocity of Points เป็นหลัก

กำหนดให้ค่าความเข้มของรูปภาพ (image intensity) ที่จุดใดๆ บนภาพให้มีค่าเป็น $E(x, y, t)$ และเมื่อมีการเคลื่อนที่เพียงเล็กน้อย ค่าของความเข้ม ณ จุดนั้นๆ จะเป็นค่าคงที่ ดังนั้น

$$\frac{dE}{dt} = 0 \quad (3.37)$$

เมื่อใช้กฎ Chain rule ในการทำ differentiation จะได้

$$\frac{\partial E}{\partial x} \frac{dx}{dt} + \frac{\partial E}{\partial y} \frac{dy}{dt} + \frac{\partial E}{\partial t} = 0 \quad (3.38)$$

กำหนดให้

$$u = \frac{dx}{dt} \quad (3.39)$$

และ

$$v = \frac{dy}{dt} \quad (3.40)$$

โดยค่าความชันทางเชิงพื้นที่และเชิงเวลาของค่าความชัน และอัตราความเร็วจะได้

$$E_x u + E_y v + E_t = 0 \quad (3.41)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากสมการที่ได้กล่าวมาค่า E_x , E_y , E_t สามารถคำนวณหาค่าได้โดยตรงจากรูปภาพ ดังเช่น ทุก ๆ จุดบนภาพ จะมีตัวแปรอยู่ 2 ตัวที่ไม่ทราบค่า คือ u และ v และมีสมการเพียงสมการเดียว การใช้ข้อมูลจากจุดบนรูปภาพเพียงอย่างเดียวนั้นไม่สามารถคำนวณหาค่าการเคลื่อนที่ของภาพ ได้ ซึ่งเป็นปัญหาที่เรียกว่า aperture problem กล่าวคือ อัตราความเร็วที่จุดใด ๆ ไม่สามารถกำหนดได้โดยใช้ข้อมูลที่จุดใด ๆ ในภาพเพียงจุดเดียว โดยไม่มีสมมุติฐานอื่นๆ เข้ามาช่วย

โดยเราสามารถกำหนดได้ว่าสนามของอัตราความเร็วเปลี่ยนแปลงอย่างราบรื่นบนรูปภาพ ภายในข้อสมมุติฐานนี้ กระบวนการคำนวณเคลื่อนที่ของภาพ โดยใช้รูปภาพ 2 ภาพ หรือมากกว่า นั้น สามารถกระทำได้

$$u = u_{average} - E_x \frac{P}{D} \quad (3.42)$$

$$v = v_{average} - E_y \frac{P}{D} \quad (3.43)$$

โดยที่

$$P = E_x u_{average} + E_y v_{average} + E_t \quad (3.44)$$

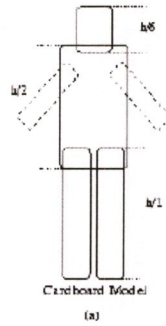
และ

$$D = \lambda^2 + E_x^2 + E_y^2 \quad (3.45)$$

จากสมการที่ 3.45 ค่า E_x , E_y , E_t และ λ แสดงให้เห็นถึงค่าความชันเชิงพื้นที่ (spatial gradient) ในทิศทางแกน x และแกน y ค่าความชันเชิงเวลา (temporal gradient) และค่าตัวคูณคงที่ตามลำดับ เมื่อใช้ภาพสองภาพในการคำนวณจะเป็นการคำนวณซ้ำ ๆ บนภาพเดิม ในกรณีที่ใช้หลายภาพการคำนวณจะใช้ภาพใหม่ในการคำนวณ

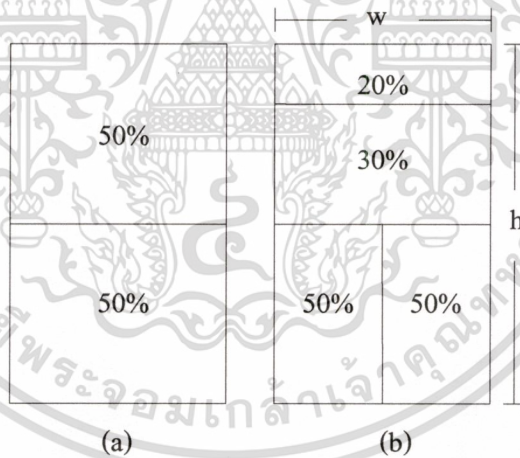
3.5 การแยกส่วนประกอบต่างๆ ของบุคคล (Ju et al. 1996)

การติดตามการเคลื่อนไหวของมนุษย์นั้นจะทำการหาขอบเขตรอบตัวบุคคลทั้งหมด จึงได้มีการเพิ่มวิธีการติดตาม โดยจะทำการติดตามการเคลื่อนไหวของส่วนประกอบต่าง ๆ ของบุคคล เช่น หัว แขน ลำตัว ขา และ เท้า เป็นต้น เพื่อให้ทราบว่าขณะนี้ท่าทางของบุคคลนั้น ๆ เป็นอย่างไร โดยการใช้แบบจำลองบุคคล (Cardboard Model) ซึ่งแสดงถึงตำแหน่งและขนาดของร่างกาย



รูปที่ 3.17 แสดงแบบจำลองบุคคล

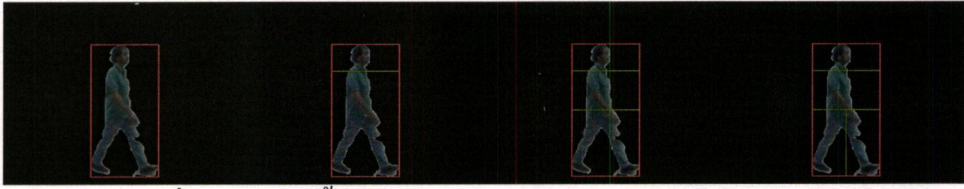
แบบจำลองบุคคล เป็นแบบจำลองร่างกายมนุษย์ในทำขึ้น ดังแสดงในรูปที่ 3.17 โดยถูกนำมาใช้ในการคำนวณตำแหน่งของส่วนต่าง ๆ ของร่างกาย โดยการที่จะแยกส่วนประกอบต่าง ๆ บนร่างกายของมนุษย์นั้น ระบบจะต้องทำการค้นหาตำแหน่งของมนุษย์ทั้งหมดเสียก่อน โดยทำเป็นกรอบสี่เหลี่ยมล้อมไว้ จากนั้นระบบจึงค่อยค้นหาและแยกส่วนประกอบต่าง ๆ ออกมาในภายหลัง โดยระบบจะนำค่าความสูงของกรอบและความกว้างของกรอบมาใช้ในการแยกส่วนประกอบต่าง ๆ ของร่างกาย ดังแสดงในรูปที่ 3.18



รูปที่ 3.18 แสดงอัตราส่วนในการแยกส่วนประกอบของร่างกาย

เมื่อได้ตำแหน่งของกรอบสี่เหลี่ยมล้อมตัวบุคคลตัวเป็นที่เรียบร้อยแล้ว ขั้นตอนต่อไปจะทำการแยกส่วนของร่างกายเป็นส่วน ๆ โดยทำการแบ่งกรอบออกเป็น 2 ส่วนตามความสูงของกรอบ ส่วนละ 50% ดังรูปที่ 3.18(a) จากนั้นจะทำการแบ่งกรอบส่วนบน โดยจะแบ่งออกเป็น 20% ซึ่งจะเป็นบริเวณศีรษะ ส่วน 30% ที่เหลือจะเป็นบริเวณลำตัว และกรอบส่วนล่าง 50% นั้นจะเป็นบริเวณขา จากทำการแบ่งกรอบส่วนล่างตามความกว้างส่วนละ 50% จะเป็นบริเวณขาทั้งสอง ขั้นตอนการแยกส่วนประกอบของร่างกายแสดงดังรูปที่ 3.19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.19 แสดงขั้นตอนการค้นหาส่วนประกอบของร่างกาย

ซึ่งเมื่อทำตามวิธีการข้างบนแล้วจะทำให้ได้บริเวณที่เป็นส่วนประกอบต่าง ๆ บนร่างกายอย่างคร่าว ๆ ซึ่งจะต้องทำการปรับปรุงให้มีขนาดที่พอดีกับขนาดที่แท้จริงของส่วนประกอบนั้น ๆ โดยจะเริ่มทำกับบริเวณที่เป็นส่วนหัวก่อน จากนั้นก็จะเป็นบริเวณลำตัว ส่วนบริเวณขา นั้นจะทำให้เป็นขั้นตอนสุดท้าย เพื่อจะได้มีความถูกต้องมากยิ่งขึ้นเมื่อนำไปใช้ในการรู้จำ



รูปที่ 3.20 แสดงขั้นตอนการปรับขนาด

3.6 การรู้จำโดยใช้ โครงข่ายประสาทเทียม (Richard O. Duda 2001)

โครงข่ายประสาทเทียม (Neural Networks) คือ รูปแบบการประมวลผลที่ได้แนวคิดมาจากการทำงานของระบบประสาทภายในสมองของมนุษย์ โดยที่การทำงานของโครงข่ายประสาทเทียมจะเลียนแบบการทำงานของสมอง เช่น การรู้จำและจำแนกรูปแบบต่างๆ การประมวลผลสัญญาณ การทำความเข้าใจในรูปภาพและเสียง การควบคุมหุ่นยนต์ ฯลฯ โดยที่การจะทำให้โครงข่ายประสาทเทียมสามารถทำงานเหล่านี้ได้จะต้องให้โครงข่ายประสาทเทียมกระทำการเรียนรู้ โดยสามารถแบ่งประเภทของการเรียนรู้ได้เป็น 2 ประเภทใหญ่ๆ คือ

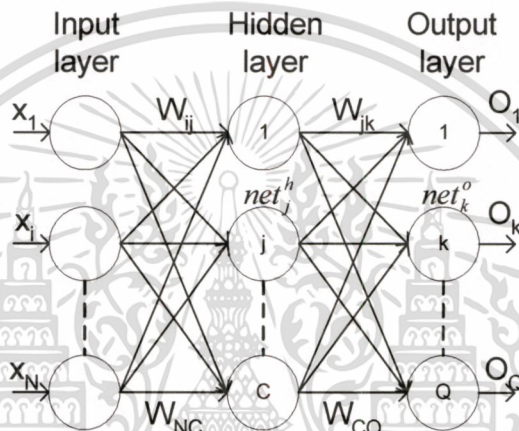
1. การเรียนรู้โดยมีผู้สอน (Supervised learning) โดยที่วิธีการนี้ จะต้องทราบผลลัพธ์ที่แท้จริงของข้อมูลที่นำมาเรียนรู้เสียก่อน จากนั้นจะนำผลลัพธ์ที่ได้จากระบบ มาเปรียบเทียบกับผลลัพธ์ที่แท้จริง ปรับค่าน้ำหนักจนได้ผลลัพธ์ที่ใกล้เคียงหรือเหมือนผลลัพธ์ที่แท้จริงมากที่สุด
2. การเรียนรู้ด้วยตนเอง (Unsupervised learning) ซึ่งการเรียนรู้วิธีนี้จะไม่ทราบผลลัพธ์ที่แท้จริง ดังนั้นจะต้องทำการฝึกสอนโครงข่ายประสาทเทียมด้วยข้อมูลหลากหลายรูปแบบ โดยที่โครงข่ายประสาทเทียมจะทำการแบ่งแยกข้อมูลออกเป็น Cluster โดยสังเกตที่ระยะทางจากข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ไปยังจุดศูนย์กลางของ Cluster หากมีข้อมูลที่ไม่สามารถจัดเข้ากับ Cluster ใดได้ ก็กระจายประสาทเทียมจะสร้าง Cluster ขึ้นมาใหม่

ซึ่งในที่นี้จะใช้โครงข่ายประสาทเทียมแบบแพร่กระจายย้อนกลับ (Back-propagation neural networks) ซึ่งมีรายละเอียดพอสังเขปดังนี้

โครงข่ายประสาทเทียมแบบแพร่กระจายย้อนกลับ (Back-propagation neural networks) ซึ่งจะใช้โครงสร้างแบบ Multilayer Feedforward Neural Networks เป็นโครงสร้าง ดังแสดงในรูปที่ 3.21



รูปที่ 3.21 โครงสร้าง Back-propagation neural networks

Back-propagation neural networks สามารถนำไปใช้ได้กับปัญหาชนิดต่างๆ ไม่ว่าจะเป็นการจำแนกกลุ่ม การทำนาย การประมวลผลสัญญาณเชิงเลข เป็นต้น

หลักการคร่าวๆ ของ Back-propagation คือ ในระหว่างทำการเรียนรู้ (Learning) จะต้องเป็นการเรียนรู้แบบมีผู้สอน เมื่อป้อน Data เข้าสู่ระบบ ระบบจะทำการประมวลผลไปที่ละชั้น (layer) จนกระทั่งได้ Output ของระบบที่ Output layer จากนั้นทำการเปรียบเทียบ Output ที่ได้กับ Target ผลต่างที่เกิดขึ้นจะถูกป้อนกลับไปปรับปรุงค่า Weight โดยจะทำการปรับปรุงถอยหลังไปเรื่อยๆ จากชั้น Output Layer จนถึงชั้น Input Layer

รูปที่ 3.21 แสดงโครงสร้าง Back-propagation neural networks ซึ่งประกอบไปด้วย 3 ชั้น และค่า Weight ในชั้น Hidden Layer และ Output Layer โดยที่ในชั้น Hidden Layer นั้นในแต่ละ Node จะประกอบไปด้วยสมการที่ 3.46

$$sgm(x) = \frac{1}{1 + e^{-x}} \quad (3.46)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากสมการที่ 3.46 สามารถเขียนเป็นสมการของ Hidden Node ได้ดังนี้

$$net_j^h = \sum_{i=1}^N X_i W_{ij} \quad ; j = \{1, 2, \dots, C\} \quad (3.47)$$

$$H_j = f(net_j^h) = sgm(net_j^h) \quad (3.48)$$

โดยที่ H_j คือ Output ของ Hidden Node
 C คือจำนวนของ Hidden Node
 N คือขนาดของ Data หรือจำนวนของ Input Node
 X คือ Input Data
 W_{ij} คือค่า Weight ระหว่าง Hidden Layer กับ Input Layer

และในส่วนของ Output Layer จะใช้สมการดังนี้

$$net_k^o = \sum_{j=1}^C H_j W_{jk} \quad ; k = \{1, 2, \dots, Q\} \quad (3.49)$$

$$O_k = f(net_k^o) = sgm(net_k^o) \quad (3.50)$$

โดยที่ O_k คือ Output ของ Output Node
 Q คือจำนวนของ Output Node
 C คือจำนวนของ Hidden Node
 W_{jk} คือค่า Weight ระหว่าง Hidden Layer กับ Output Layer

เมื่อได้ผลลัพธ์จาก Output แล้ว นำมาหาค่า Error Rate เพื่อจะนำไปปรับปรุงค่า Weight ในแต่ละ Layer ซึ่งการปรับปรุงค่า Weight ของ Back-propagation นี้จะเปรียบเทียบความแตกต่างระหว่าง Output ที่ได้กับ Target โดยจะใช้ LMS algorithm เพื่อหาความแตกต่างนี้

$$E(W) = \frac{1}{2} \sum_{k=1}^Q (T_k - O_k)^2 \quad (3.51)$$

โดยที่ O_k คือ Output ของ Output Node
 เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Q คือจำนวนของ Output Node
 T_k คือ Output ที่แท้จริง (Target)

การปรับปรุ้ค่าน้ำหนัก (Update weight) นั้นจะกระทำได้โดยใช้สมการที่ 3.52

$$W(new) = W + \Delta W \quad (3.52)$$

โดยที่ W คือค่า Weight

$$\text{เมื่อ } \Delta W = -\eta \frac{\partial E}{\partial W} \quad (3.53)$$

โดยที่ η คืออัตราการเรียนรู้ (Learning rate)

จากรูปที่ 3.21 เป็นรูปโครงสร้าง Back-propagation neural networks ซึ่งมีจำนวน Layer เท่ากับ 3 Layer และมี Weight อยู่ 2 ช่วงคือ จาก Hidden to Output และจาก Input to Hidden ซึ่งในขั้นตอนแรกจะต้องพิจารณาจากส่วน Hidden to Output เสียก่อน

$$\Delta W_{jk} = -\eta \frac{\partial E}{\partial W_{jk}} = \eta \delta_k \frac{\partial net_k^o}{\partial W_{jk}} \quad (3.54)$$

โดยที่ W_{jk} คือค่า Weight ระหว่าง Hidden Layer กับ Output Layer

$$\text{เนื่องจาก } net_k^o = \sum_{j=1}^C H_j W_{jk}$$

$$\text{ดังนั้น } \frac{\partial net_k^o}{\partial W_{jk}} = \frac{\partial \sum_{j=1}^C H_j W_{jk}}{\partial W_{jk}} = H_j \quad (3.55)$$

โดยที่ H_j คือ Output ของ Hidden Node
 C คือจำนวนของ Hidden Node

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และ
$$\delta_K = -\frac{\partial E}{\partial net_k^O} = (T_k - O_k) f'(net_k^O) \quad (3.56)$$

โดยที่ O_k คือ Output ของ Output Node

T_k คือ Output ที่แท้จริง (Target)

เมื่อนำสมการที่ 3.54 , 3.55 และ 3.56 แทนลงในสมการที่ 3.50 จะได้สมการที่ใช้ในการปรับปรุงค่า Weight จาก Hidden Node ไปยัง Output Node ดังสมการที่ 3.57

$$W_{jk} (new) = W_{jk} + \eta [(T_k - O_k) f'(net_k^O) H_j] \quad (3.57)$$

โดยที่ H_j คือ Output ของ Hidden Node

O_k คือ Output ของ Output Node

T_k คือ Output ที่แท้จริง (Target)

W_{jk} คือค่า Weight ระหว่าง Hidden Layer กับ Output Layer

η คืออัตราการเรียนรู้ (Learning rate)

ในขั้นตอนถัดมาเป็นการปรับปรุง Weight ในส่วน Input to Hidden ซึ่งจะใช้หลักการเหมือนกับการปรับปรุง Weight ในส่วน Hidden to Output โดยจะใช้สมการที่ 3.52 และ 3.53 เป็นหลัก

$$\Delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}} = \eta \delta_j \frac{\partial net_j^h}{\partial W_{ij}} \quad (3.58)$$

โดยที่ W_{ij} คือค่า Weight ระหว่าง Input Layer กับ Hidden Layer

เนื่องจาก
$$net_j^h = \sum_{i=1}^N X_i W_{ij}$$

ดังนั้น
$$\frac{\partial net_j^h}{\partial W_{ij}} = \frac{\partial \sum_{i=1}^N X_i W_{ij}}{\partial w_{ij}} = X_i \quad (3.59)$$

โดยที่ N คือขนาดของ Data หรือจำนวนของ Input Node

X คือ Input Data

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และ
$$\delta_j = -\frac{\partial E}{\partial net_j^h} = \left(\sum_{k=1}^Q \delta_k W_{jk} \right) f'(net_j^h) \quad (3.60)$$

โดยที่ W_{jk} คือค่า Weight ระหว่าง Hidden Layer กับ Output Layer
 δ_k หาได้จากสมการที่ 3.56

เมื่อนำสมการที่ 3.58 , 3.59 และ 3.60 แทนลงไปในสมการที่ 3.50 จะได้สมการที่ใช้ในการปรับปรุงค่า Weight จาก Input Node ไปยัง Hidden Node ดังสมการที่ 3.61

$$W_{ij}(new) = W_{ij} + \eta \left(\left[\sum_{k=1}^Q \delta_k W_{jk} \right] f'(net_j^h) X_i \right) \quad (3.61)$$

โดยที่ X คือ Input Data
 W_{ij} คือค่า Weight ระหว่าง Hidden Layer กับ Input Layer
 δ_k หาได้จากสมการที่ 2.31
 Q คือจำนวนของ Output Node
 η คืออัตราการเรียนรู้ (Learning rate)

3.7 สหสัมพันธ์ (ส่วน สายยศ และอังคณา สายยศ. 2540)

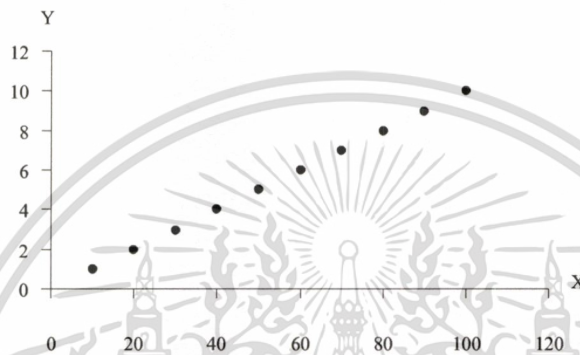
ในการแสดงความสัมพันธ์ระหว่างข้อมูลที่ได้มาจากสิ่งๆ เดียวกัน สามารถแสดงได้โดยการใช้กราฟที่เรียกว่าสแคตเตอร์ไดอะแกรม (Scatter Diagram) และสามารถแสดงค่าความสัมพันธ์ระหว่างข้อมูลได้จากค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ซึ่งมีรายละเอียดดังต่อไปนี้

3.7.1 การแปลสหสัมพันธ์จากกราฟ

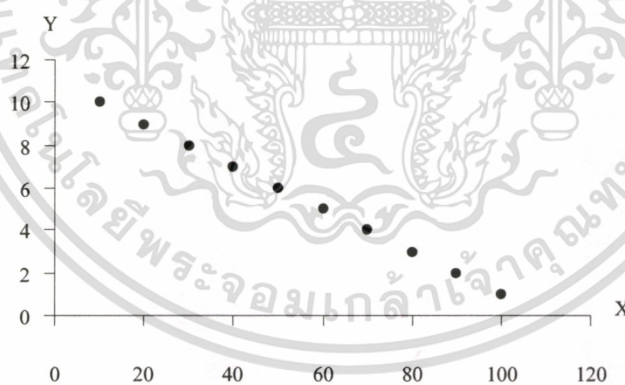
ในการพิจารณาความสัมพันธ์ระหว่างข้อมูลจากกราฟนี้ทำให้เห็นแนวโน้มความสัมพันธ์หรือจุดร่วมระหว่างข้อมูลเป็นลักษณะต่างๆ ทำให้คาดคะเนได้ว่าข้อมูลคู่หนึ่งสัมพันธ์กันอย่างไร และสัมพันธ์กันในทิศทางใด ซึ่งความสัมพันธ์ของกราฟสามารถแบ่งออกเป็นลักษณะต่างๆ ได้ดังต่อไปนี้

3.7.1.1 ข้อมูลสัมพันธ์กันอย่างสมบูรณ์

สามารถแบ่งออกเป็น 2 ประเภทคือแบบไปทางบวก (Positive) เช่นถ้าค่า X มาก ค่า Y จะมากด้วย เป็นต้น ซึ่งเป็นลักษณะข้อมูลที่มีค่าคล้อยตามกันทำให้จุดรวมของข้อมูล X และ Y เป็นเส้นตรงเริ่มจากค่าต่ำไปหาค่าสูง จึงทำให้เกิดความสัมพันธ์อย่างสมบูรณ์ไปทางบวก และอีกประเภทคือแบบไปทางลบ (Negative) เช่นถ้าค่า X มาก ค่า Y จะน้อย เป็นต้น ซึ่งเป็นลักษณะข้อมูลที่เป็นความสัมพันธ์กลับไม่คล้อยตามกัน จึงทำให้เกิดความสัมพันธ์อย่างสมบูรณ์ไปทางลบ



รูปที่ 3.22 กราฟแสดงสหสัมพันธ์อย่างสมบูรณ์เป็นไปทางบวก

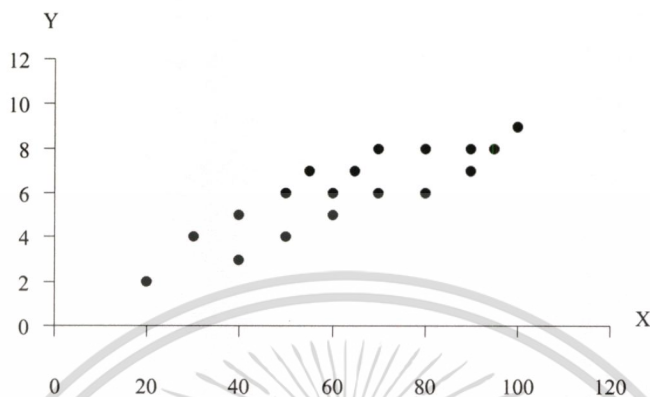


รูปที่ 3.23 กราฟแสดงสหสัมพันธ์อย่างสมบูรณ์เป็นไปทางลบ

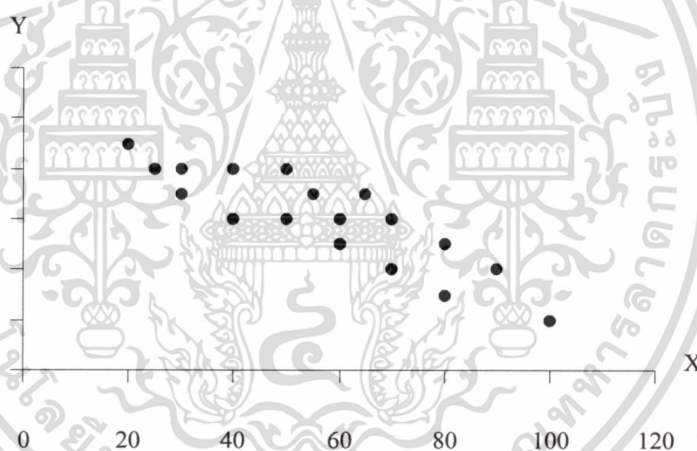
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7.1.2 ข้อมูลสัมพันธ์กันสูงแต่ไม่สมบูรณ์

ข้อมูลลักษณะนี้มีค่าของข้อมูลทั้งสองจะคล้อยตามกันแต่ไม่ถึงกับเป็นเส้นตรง สามารถแบ่งออกเป็น 2 ประเภทเช่นกันคือแบบไปทางบวกและแบบไปทางลบ



รูปที่ 3.24 กราฟแสดงสหสัมพันธ์ค่อนข้างสูงเป็นไปทางบวก

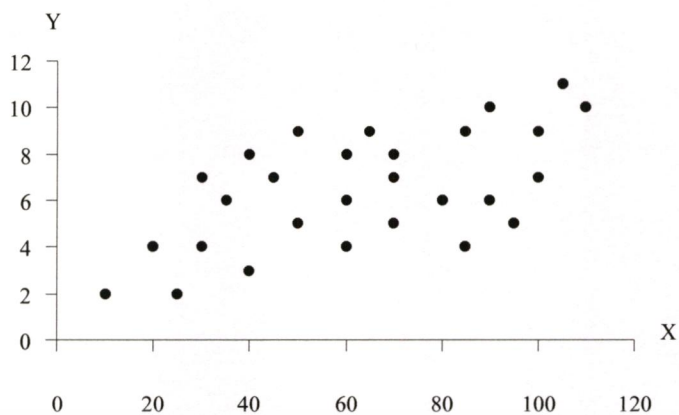


รูปที่ 3.25 กราฟแสดงสหสัมพันธ์ค่อนข้างสูงเป็นไปทางลบ

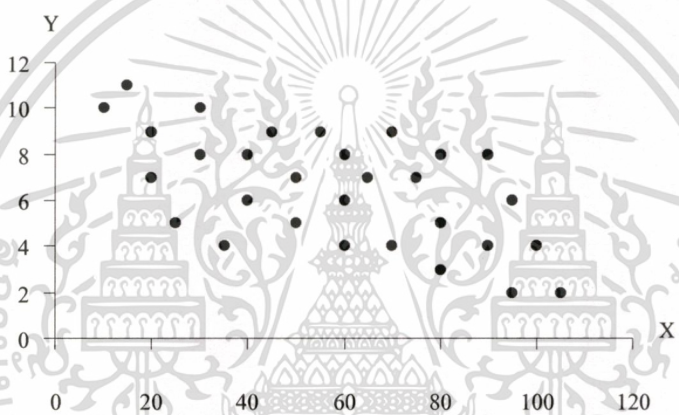
3.7.1.3 ข้อมูลสัมพันธ์กันต่ำ

ข้อมูลลักษณะนี้มีค่าของข้อมูลทั้งสองจะคล้อยตามกันน้อย สามารถแบ่งออกเป็น 2 ประเภทเช่นกันคือแบบไปทางบวกและแบบไปทางลบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



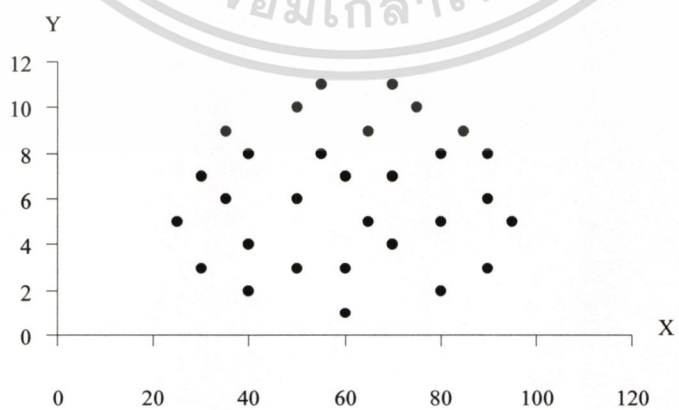
รูปที่ 3.26 กราฟแสดงสหสัมพันธ์ค่อนข้างต่ำเป็นไปทางบวก



รูปที่ 3.27 กราฟแสดงสหสัมพันธ์ค่อนข้างต่ำเป็นไปทางลบ

3.7.1.4 ข้อมูลสัมพันธ์เป็นศูนย์

ข้อมูลลักษณะนี้มีค่าของข้อมูลไม่เกี่ยวกันเลยคือกระจายไปทั่วจนเป็นรูปวงกลม



รูปที่ 3.28 กราฟแสดงสหสัมพันธ์เป็นศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7.2 การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์

ค่าสัมประสิทธิ์สหสัมพันธ์ (r) จะทำให้เห็นว่าข้อมูล X และ Y มีความสัมพันธ์กันเท่าใด โดยสมการที่จะนำมาใช้ในการคำนวณจะเป็นสมการที่สามารถนำข้อมูลดิบมาใช้ในการคำนวณได้ โดยง่ายดังต่อไปนี้

$$r = \frac{N \sum XY - \sum X \sum Y}{\sqrt{(N \sum X^2 - (\sum X)^2)(N \sum Y^2 - (\sum Y)^2)}} \quad (3.62)$$

โดย N แทนจำนวนข้อมูล

$\sum X$ แทนผลรวมทั้งหมดของข้อมูล X

$\sum Y$ แทนผลรวมทั้งหมดของข้อมูล Y

$\sum X^2$ แทนผลรวมทั้งหมดของข้อมูล X แต่ละตัวยกกำลังสอง

$\sum Y^2$ แทนผลรวมทั้งหมดของข้อมูล Y แต่ละตัวยกกำลังสอง

$\sum XY$ แทนผลรวมทั้งหมดของข้อมูล X และ Y คูณกันแต่ละคู่

บทที่ 4

กรรมวิธีที่นำเสนอ

ดังที่กล่าวมาข้างต้นถึงความสำคัญของการพัฒนาระบบรักษาความปลอดภัยและระบบเฝ้าระวังอัตโนมัติ ซึ่งความสามารถในการระบุตัวบุคคลจากสัญญาณวิดีโอทัศน์ เป็นหัวใจหลักของระบบดังกล่าว ระบบระบุตัวบุคคลจากสัญญาณวิดีโอทัศน์ที่ได้มีการพัฒนาขึ้นก่อนหน้านี้มักจะใช้การรู้จำเอกลักษณ์ใดเอกลักษณ์หนึ่ง เช่น ใบหน้า ท่าเดิน ลายม่านตา เป็นต้น แต่เนื่องจากในสภาพแวดล้อมใช้งานจริง บางครั้งเราไม่สามารถติดตั้งกล้องหรือกำหนดเส้นทางการเดินของบุคคล เพื่อให้ได้ภาพที่สามารถสกัดเอาเอกลักษณ์ที่ต้องการมาวิเคราะห์เพื่อระบุตัวบุคคลได้ วิทยานิพนธ์ฉบับนี้จึงเสนอวิธีการในการระบุตัวบุคคลโดยใช้หลายเอกลักษณ์และหลายกล้อง เพื่อให้กล้องทำงานร่วมกันในการระบุตัวบุคคลจากเอกลักษณ์ที่สกัดได้จากมุมมองของแต่ละกล้อง เอกลักษณ์ที่ใช้ในการรู้จำตัวบุคคลในวิทยานิพนธ์ฉบับนี้ ได้แก่ ใบหน้า สีผิว ท่าเดิน สีของเครื่องแต่งกาย และข้อมูลเวลา

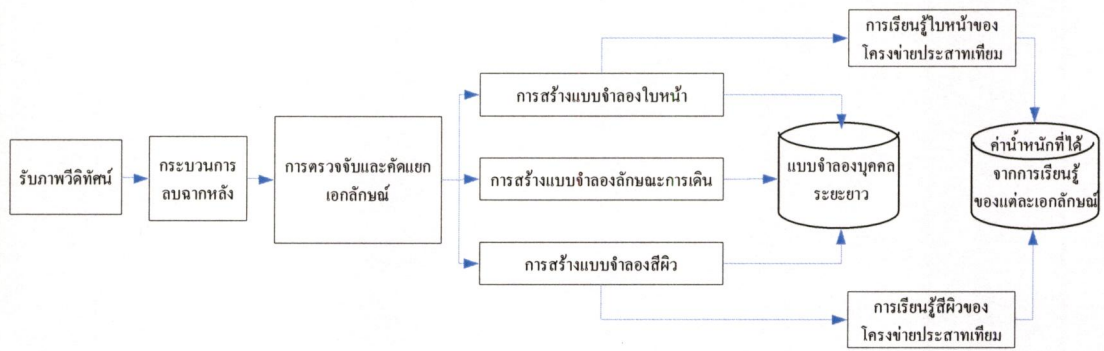
ในบทนี้จะกล่าวถึงรายละเอียดของกรรมวิธีที่นำเสนอ ได้แก่ การสร้างจำลองบุคคล และการรู้จำบุคคลจากเอกลักษณ์ต่างๆ รวมถึงการทำงานร่วมกันระหว่างกล้อง โดยระบบแบ่งออกเป็น 2 ส่วน

1. ส่วนสร้างแบบจำลองบุคคล (Person modeling)

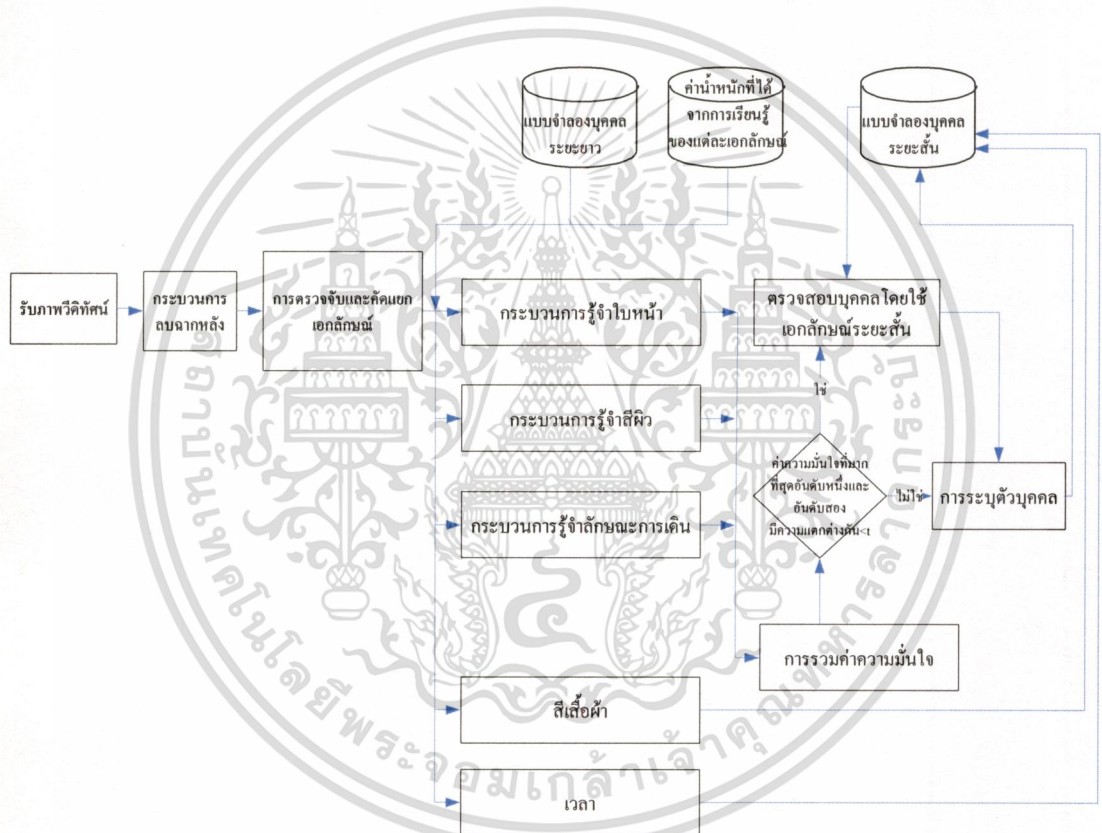
กระบวนการสร้างแบบจำลองบุคคล แสดงดังรูปที่ 4.1 เริ่มต้นจากการนำสัญญาณวิดีโอทัศน์รับเข้า มาทำการตัดแยกส่วนของบุคคลออกจากฉากหลัง โดยใช้วิธีการลบฉากหลังดังที่ได้กล่าวในบทที่ 3 จากนั้นทำการตรวจจับและตัดแยกส่วนภาพของบุคคลเพื่อนำมาสกัดหาเอกลักษณ์บุคคล ได้แก่ ภาพใบหน้า สีผิว พารามิเตอร์ที่บ่งบอกท่าเดิน เป็นต้น เมื่อสกัดได้เอกลักษณ์ต่างๆแล้ว ก็จะนำมาสร้างแบบจำลอง โดยใช้วิธีการดังที่จะกล่าวต่อไปในบทนี้

2. ส่วนระบุตัวบุคคล (Person identification)

ส่วนนี้เป็นส่วนระบุตัวบุคคลโดยเปรียบเทียบเอกลักษณ์กับแบบจำลองบุคคลที่ได้สร้างไว้ แสดงดังรูปที่ 4.2 เริ่มต้นจากการนำสัญญาณวิดีโอทัศน์รับเข้า มาทำการตัดแยกส่วนของบุคคลออกจากฉากหลัง จากนั้นทำการตรวจจับและตัดแยกส่วนภาพของบุคคลเพื่อนำมาสกัดหาเอกลักษณ์บุคคล ได้แก่ ภาพใบหน้า สีผิว พารามิเตอร์ที่บ่งบอกท่าเดิน เป็นต้น เมื่อสกัดได้เอกลักษณ์ต่างๆแล้ว ก็จะนำเข้าสู่กระบวนการรู้จำโดยใช้เอกลักษณ์ต่างๆ จากนั้นนำผลการรู้จำมารวมกันเพื่อแสดงผลลัพธ์สุดท้าย ในส่วนการระบุตัวบุคคลนี้ จะมีการสร้างแบบจำลองส่วนขยายด้วยเพื่อช่วยในการระบุตัวบุคคล



รูปที่ 4.1 แสดงขั้นตอนการสร้างแบบจำลองบุคคล



รูปที่ 4.2 แสดงขั้นตอนการระบุตัวบุคคล

4.1 แบบจำลองบุคคล (Person Model)

แบบจำลองบุคคลที่ใช้ในการระบุตัวบุคคลที่เสนอในวิทยานิพนธ์ฉบับนี้แบ่งออกเป็น 2 ส่วนด้วยกัน โดยในส่วนแรกจะเป็นแบบจำลองบุคคลระยะยาว (Long-term Person Model) ส่วนที่สองจะเป็นแบบจำลองบุคคลระยะสั้น (Short-term Person Model) และจะมีการปรับตัวตลอดเวลา (Adaptive)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.1 แบบจำลองบุคคลระยะยาว

แบบจำลองบุคคลระยะยาวเป็นแบบจำลองหลัก ประกอบไปด้วยเอกลักษณ์ 3 เอกลักษณ์ ได้แก่ ใบหน้า สีผิว และท่าเดิน แบบจำลองบุคคลระยะยาวของบุคคลที่ i สามารถเขียนได้ดังสมการที่ 4.1

$$P_i^L = \langle F_i, S_i, G_i \rangle \quad (4.1)$$

- โดยที่ F คือ ภาพของใบหน้าที่ใช้ในการเรียนรู้ของโครงข่ายประสาทเทียมสำหรับรู้จำใบหน้า
- S คือ สีผิวเฉลี่ยของบุคคล กำหนดในรูปแบบ Cb, Cr และใช้จากการเรียนรู้ของโครงข่ายประสาทเทียมรู้จำสีผิว
- G คือ ค่าความกว้างในการก้าวขาสูงสุดเชิงสถิติของบุคคล

4.1.2 แบบจำลองบุคคลระยะสั้น

แบบจำลองบุคคลระยะสั้นเป็นแบบจำลองบุคคลเพิ่มเติม ที่ใช้ช่วยในกระบวนการรู้จำบุคคลในช่วงเวลาสั้นๆ เช่น ภายใน 1 วัน ประกอบด้วย สีเสื้อ และ ข้อมูลเชิงเวลา ของบุคคล ตัวอย่างเช่น เมื่อกำลังตัวหนึ่งในระบบได้ทำการรู้จำบุคคลหนึ่งที่ตรวจจับได้ และได้ทำการบันทึกสีเสื้อและเวลาที่พบในระบบ ในเวลาต่อมากล้องตัวอื่นสามารถใช้ข้อมูลดังกล่าว ช่วยในกระบวนการรู้จำได้ เพื่อเพิ่มความแม่นยำในการรู้จำ โดยแบบจำลองบุคคลระยะสั้นของบุคคลที่ i สามารถเขียนได้ดังสมการที่ 4.2

$$P_i^S = \langle C_i, T_i \rangle \quad (4.2)$$

- โดยที่ C คือ สีเสื้อเฉลี่ยล่าสุดที่ตรวจจับได้ของบุคคล กำหนดในรูปแบบ r,g (Normalized R, Normalized G)
- T คือ ข้อมูลบันทึกเวลาที่กล้องแต่ละตัวในระบบ ตรวจจับและรู้จำบุคคลได้

4.2 การสร้างแบบจำลองบุคคล (Person Modeling)

4.2.1 การสร้างแบบจำลองบุคคลระยะยาว

ในการสร้างแบบจำลองบุคคลระยะยาวนั้น ระบบจะทำการสกัดเอกลักษณ์ออกมาจากภาพวิดิทัศน์รับเข้าได้ แก่ ภาพใบหน้า สีผิว และลักษณะการเดิน เมื่อได้ส่วนที่เป็นใบหน้า และ สีผิวเป็นที่เรียบร้อยแล้วนั้น (ซึ่งจะกล่าวอย่างละเอียดในส่วนถัดไป) จะนำข้อมูลที่ได้ออกไปให้โครงข่ายประสาทเทียมเรียนรู้เพื่อใช้ในการระบุตัวตนของบุคคลนั้น เมื่อผู้ดูที่เห็นภาพหรือได้ยินเสียงนั้นๆ ไม่รู้ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมีให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประสาทเทียมทำการเรียนรู้ ในส่วนของลักษณะการเดินจะเก็บเป็นข้อมูลความกว้างในการก้าวขา สูงสุดในเชิงสถิติ ดังแสดงในรูปที่ 4.3



รูปที่ 4.3 ภาพแบบจำลองบุคคลระยะยาว

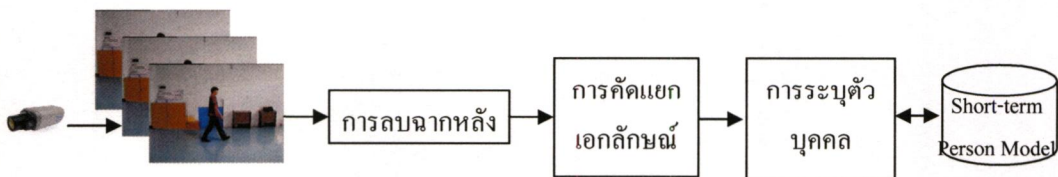
ในการระบุตัวบุคคลหนึ่งคน จะใช้ข้อมูลเอกลักษณ์ ใบหน้า สีผิว และลักษณะการเดิน ซึ่งสามารถแทนให้อยู่ในรูปของสมการ ที่ 4.3

$$P_i = \langle F_i, S_i, G_i \rangle \quad (4.3)$$

โดยที่ P คือ บุคคล
 F คือ เอกลักษณ์ใบหน้า
 S คือ เอกลักษณ์สีผิว
 G คือ เอกลักษณ์การเดิน
 i คือ จำนวนบุคคล

4.2.2 การสร้างแบบจำลองบุคคลระยะสั้น

แบบจำลองบุคคลระยะสั้นนั้นจะเก็บข้อมูลเอกลักษณ์ของบุคคลหลังจากที่ได้ผ่านกระบวนการระบุตัวบุคคลมาแล้วว่าเป็นบุคคลใด โดยจะเก็บเอกลักษณ์ที่เป็นส่วนประกอบของบุคคลนั้น เช่น สีเสื้อที่สวมใส่ และ เวลาที่ปรากฏในบริเวณนั้น ๆ ในขณะที่ระบบกำลังตรวจสอบว่าบุคคลนั้นเป็นบุคคลใด ระบบก็จะทำการสกัดส่วนที่เป็นสีเสื้อไว้ พอระบุว่าเป็นบุคคลใดได้ระบบจะทำการเพิ่มข้อมูลเข้าไปในส่วน of แบบจำลองบุคคลส่วนขยาย ดังแสดงในรูปที่ 4.4



รูปที่ 4.4 แสดงการรู้จำและแบบจำลองระยะสั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อระบบจะทำการระบุตัวบุคคลระบบจะนำข้อมูลเอกลักษณ์จากแบบจำลองหลัก และแบบจำลองส่วนขยายเข้ามาใช้ในการระบุตัวบุคคล ซึ่งสามารถแทนให้อยู่ในรูปของสมการ ที่ 4.2

$$P_i = \langle F_i, S_i, G_i \rangle + \langle C_i, T_i \rangle \quad (4.4)$$

โดยที่

- P คือ บุคคล
- F คือ เอกลักษณ์ใบหน้า
- S คือ เอกลักษณ์สีผิว
- G คือ เอกลักษณ์การเดิน
- C คือ สีเสื้อ
- T คือ เวลา
- i คือ จำนวนบุคคล

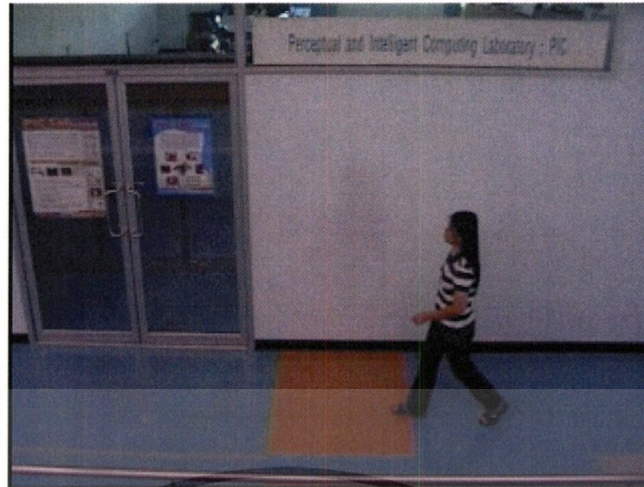
4.3 การเก็บภาพวิถีทัศนตัวอย่าง

การระบุตัวบุคคลโดยใช้ภาพจากวิถีทัศนตำแหน่งของการติดตั้งกล้องจะมีความสำคัญ เพื่อให้ได้เอกลักษณ์ตามที่ต้องการจึงต้องติดตั้งไว้ในตำแหน่งที่เหมาะสม ซึ่งในงานวิจัยนี้ได้เลือกมุมกล้องที่จะติดตั้งไว้ 2 มุมกล้อง โดยมุมกล้องแรกนั้นจะเป็นมุมที่บุคคลเดินเข้าหา กล้อง ซึ่งจะทำได้เอกลักษณ์ใบหน้า สีผิว และ สีเสื้อ และในมุมที่สองบุคคลจะเดินหันด้านข้างให้กับกล้องจะทำให้ได้เอกลักษณ์สีผิว ลักษณะการเดิน และ สีเสื้อ ดังแสดงในรูปที่ 4.5 และ 4.6 ตามลำดับ



รูปที่ 4.5 แสดงวิถีทัศนของบุคคลที่เดินเข้าหากกล้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ภายในห้องปฏิบัติการเท่านั้น เมื่อผู้ดูแลระบบใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.6 แสดงวิถีทัศน์ของบุคคลเดินหันข้างกับกล้อง

4.4 เอกลักษณ์ที่ใช้ในการระบุตัวบุคคล

การที่จะตรวจจับและคัดแยกเอกลักษณ์ให้ได้อย่างมีประสิทธิภาพนั้นจึงต้องทำการลดบริเวณที่ใช้ในการค้นหาเอกลักษณ์โดยทำการลบบริเวณที่เป็นฉากหลังออกไปเสียก่อน เพื่อให้เหลือเฉพาะบริเวณที่เป็นบุคคลเท่านั้นดังแสดงในรูปที่ 4.7



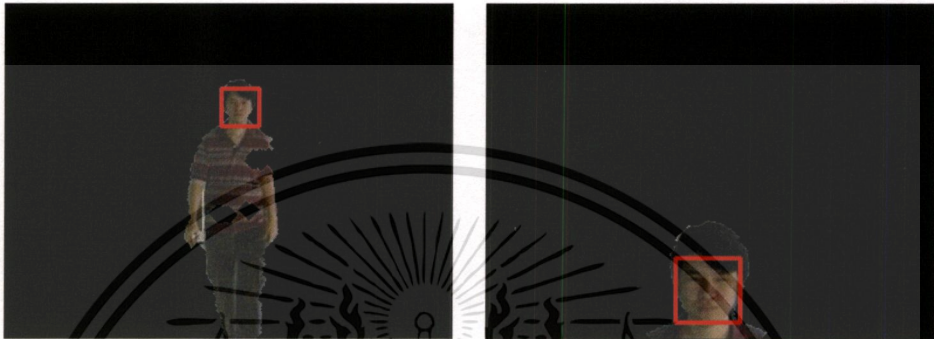
รูปที่ 4.7 แสดงผลการลบฉากหลังในมุมมองทั้งสอง

เมื่อได้เฉพาะบริเวณที่เป็นตัวบุคคลแล้วจึงทำการคัดแยกเอกลักษณ์ต่าง ๆ ออกมาเพื่อนำไปใช้ในการรู้จำตัวบุคคลต่อไป ซึ่งมีรายละเอียดดังนี้

4.4.1 ภาพใบหน้า

เมื่อบุคคลปรากฏขึ้นภายในบริเวณที่เฝ้าระวัง ระบบจะทำการค้นหาบริเวณที่เป็นใบหน้าโดยใช้วิธีการตรวจจับภาพใบหน้าด้วย ฮาร์-ฟีเจอร์ ซึ่งได้นำเสนอโดย Viola et al. (2001) โดยตัวแยกประเภทจะทำการเรียนรู้ภาพใบหน้าจากรูปภาพที่ขนาดเท่ากันทั้งหมด เสียก่อนจึงจะใช้ในการเอกสารเป็นเอกสารทสงวนเวลาสำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค้นหาบริเวณที่เป็นใบหน้าได้ (Viola ใช้ขนาด 20×20 เพื่อใช้ในการตรวจจับใบหน้า) โดยกระบวนการตรวจจับนั้นจะทำการเลื่อนกรอบค้นหา (Search window) ว่าบริเวณใดในภาพที่มีลักษณะใกล้เคียงหรือเหมือนกับใบหน้าที่มากที่สุด สำหรับการค้นหาภาพใบหน้าที่มีขนาดแตกต่างกันนั้น สามารถทำได้เนื่องจากตัวแยกประเภทนั้นมีความสามารถในการหดหรือขยาย (scale) ได้เป็นอย่างดี ดังแสดงในรูปที่ 4.8



รูปที่ 4.8 แสดงการค้นหาใบหน้า

จากรูปที่ 4.8 เมื่อได้บริเวณที่เป็นใบหน้า ภาพของใบหน้าที่จะถูกแยกออกมาเพื่อนำไปใช้ในการรู้จำ และเนื่องจากเป็นการค้นหาภาพใบหน้าที่จากระบบวิดิทัศน์จึงทำให้ขนาดของภาพใบหน้าที่แต่ละภาพที่ได้มีขนาดไม่เท่ากัน เมื่อบุคคลอยู่ห่างจากกล้องภาพของใบหน้าที่ตรวจจับได้จะมีขนาดเล็กกว่าภาพของใบหน้าที่อยู่ใกล้กล้อง เนื่องจากบุคคลมีทิศทางการเดินเข้าหากล้องจึงทำให้ขนาดของภาพใหญ่ขึ้น โดยขนาดของภาพใบหน้าที่จะนำมาใช้นั้นจะเลือกขนาดภาพของใบหน้าที่มีขนาดใหญ่กว่า 20×20 จุดภาพ จากนั้นนำภาพใบหน้าทุก ๆ ภาพมาปรับขนาดให้มีขนาดเท่ากับ 20×20 จุดภาพ เนื่องจากขั้นตอนในการรู้จำใบหน้าด้วยโครงข่ายประสาทเทียม ขนาดของข้อมูลขาเข้าจะต้องมีขนาดที่คงที่ตลอด ดังแสดงในรูปที่ 4.9

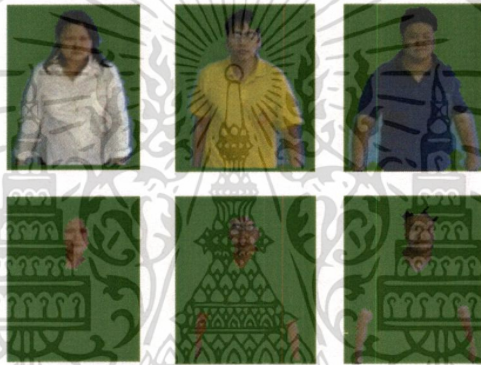


รูปที่ 4.9 แสดงตัวอย่างภาพใบหน้า

ภาพที่ได้จะนำไปใช้ในการการเรียนรู้ของโครงข่ายประสาทเทียมเพื่อสร้างเป็นแบบจำลอง
 เอกสาร ใบหน้าและทำการสกัดภาพใบหน้าจากวิดิทัศน์รับเข้าเพื่อใช้ในการระบุตัวบุคคล ใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.2 ภาพสีผิว

การตรวจจับสีผิวจะใช้วิธีที่เรียกว่าการตรวจจับสีผิวโดยใช้แบบจำลองชนิดวงรี (Elliptic Boundary Model) โดยการสร้างแบบจำลองสีผิวที่ใช้ในการตรวจจับลักษณะเป็นวงรี ซึ่งการแยกแยะสีผิวโดยวิธีนี้จะนำปริภูมิสี YCbCr มาใช้ในการหาโดยปริภูมิตั้งนี้ก็จะแก้ไขปัญหาในเรื่องของแสงที่ตกกระทบบนใบหน้าและมือซึ่งมีค่าเปลี่ยนแปลงอยู่ตลอดเวลา ซึ่งในการพิจารณาจะพิจารณาเพียงองค์ประกอบของสี (Cb,Cr) โดยละทิ้งองค์ประกอบทางแสงสว่าง(Y) โดยจะพิจารณาทุก ๆ จุดภาพว่าจุดภาพใดมีค่า Cb และ Cr อยู่ในขอบเขตของวงรีหรือไม่ ถ้าอยู่ในขอบเขตของวงรีแสดงว่าบริเวณนั้นเป็นสีผิว ถ้าอยู่นอกขอบเขตวงรีแสดงว่าไม่ใช่สีผิว เพื่อการแสดงผลจึงแสดงผลในปริภูมิสี RGB ดังแสดงในรูปที่ 4.10



รูปที่ 4.10 แสดงตัวอย่างภาพสีผิวในระบบสี RGB

เมื่อได้บริเวณที่เป็นสีผิวแล้ว จะนำค่าสีผิวมาทำการหาค่าเฉลี่ย โดยค่าของสีผิวที่จะนำไปใช้ในการรู้จำตัวบุคคลนั้นจะเป็นค่าของ \bar{Cb} และ \bar{Cr} โดยไม่นำค่าความสว่าง (Y) มาใช้ในการรู้จำตัวบุคคล จากนั้นนำค่าของ \bar{Cb} และ \bar{Cr} ที่ได้ไปผ่านกระบวนการเรียนรู้ของโครงข่ายประสาทเทียมเพื่อสร้างเป็นแบบจำลองสีผิว และทำการสกัดข้อมูลสีผิวจากวิดีโอที่ศึ้นรับเข้าเพื่อใช้ในการระบุตัวบุคคล

4.4.3 ภาพลักษณะการเดิน

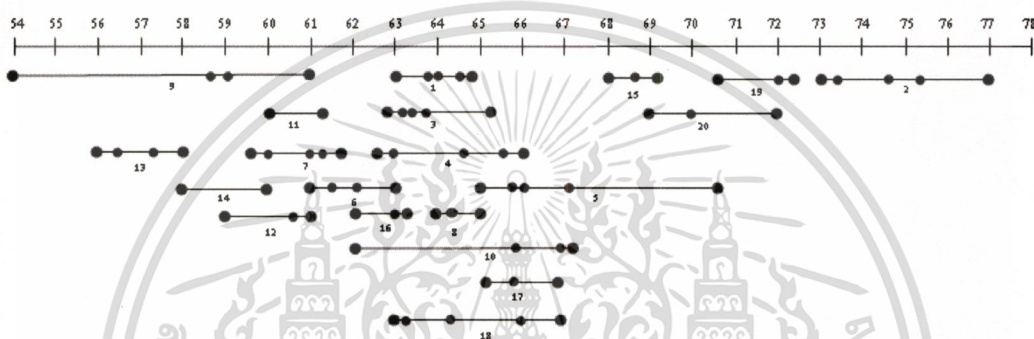
ลักษณะการเดินที่ใช้ในการระบุตัวบุคคลนั้นจะใช้ความกว้างในการก้าวขาที่มากที่สุด โดยทำการจับการเคลื่อนที่ของบุคคลจากวิดีโอที่ศึ้นด้านข้างเป็นจำนวน 60 ภาพที่ต่อเนื่องกัน ซึ่งจะใช้เวลาในการตรวจจับไม่นานนัก เนื่องจากความเร็วของภาพวิดีโอที่ทำการเก็บนั้นใช้ความเร็วที่ 25 ภาพ/วินาที จากนั้นนำชุดภาพที่ได้มาทำการหาค่าความกว้างในการก้าวขาที่มากที่สุดเฉลี่ย โดยนำค่าความกว้างในการก้าวขาที่มากที่สุดในแต่ละช่วงมาทำการหาค่าเฉลี่ยโดยขั้นตอนการหาค่าความกว้างสูงสุดนั้นจะกล่าวในส่วนของกระบวนการระบุตัวบุคคลด้วยลักษณะการเดิน รูปที่ 4.11 แสดงภาพวิดีโอ

ที่ศึ้นเดินหันข้าง
เอกสารนี้เป็นเอกสารสงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.11 ภาพตัวอย่างการเดินหัดข้างกับกล้อง

ค่าความกว้างในการก้าวขาสูงสุดนั้นจะเก็บอยู่รูปของข้อมูลทางสถิติ ดังแสดงในรูปที่ 4.12



รูปที่ 4.12 ผลทางสถิติของความกว้างในการก้าวขาของบุคคล 20 คน

ในส่วนต่อไปจะเป็นส่วนที่สอง คือส่วนของการระบุตัวบุคคลโดยใช้แบบจำลองบุคคลที่ได้สร้างไว้ในส่วนแรก และส่วนของแบบจำลองส่วนขยายที่ใช้ในการเก็บเอกลักษณ์สี่เสือ และเวลาของแต่ละบุคคลไว้ชั่วคราว โดยรายละเอียดดังนี้

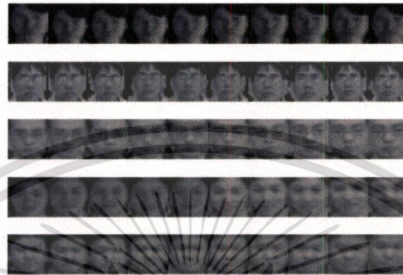
4.5 การรู้จำและการปรับปรุงแบบจำลองส่วนขยาย

ในการรู้จำตัวบุคคล ระบบจะทำการค้นหาเอกลักษณ์จากนั้นนำเอกลักษณ์ที่ได้ไปวิเคราะห์และนำไปเปรียบเทียบกับข้อมูลในแบบจำลองบุคคลที่สร้างไว้ ซึ่งค่าที่ได้จากการตรวจสอบจะเรียกว่าค่าความมั่นใจ คือเป็นค่าที่ระบบมีความมั่นใจในเอกลักษณ์นั้น ๆ ว่าเป็นบุคคลใด จากนั้นนำค่าความมั่นใจที่ได้จากเอกลักษณ์ทั้งหมดมารวมกันเพื่อนำมาหาค่าความมั่นใจรวมของบุคคลที่มากที่สุด ก่อนที่ระบบจะระบุตัวบุคคลออกมาว่าบุคคลนั้นเป็นบุคคลใด และเมื่อระบุตัวบุคคลได้แล้วจะเพิ่มเอกลักษณ์ส่วนขยายได้แก่ สี่เสือ และ เวลา ของบุคคลนั้น ๆ ไว้ในแบบจำลองส่วนขยายเพื่อใช้เป็นส่วนช่วยในการระบุตัวบุคคล ดังแสดงในรูปที่ 4.4 โดยแต่ละขั้นตอนมีรายละเอียดดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

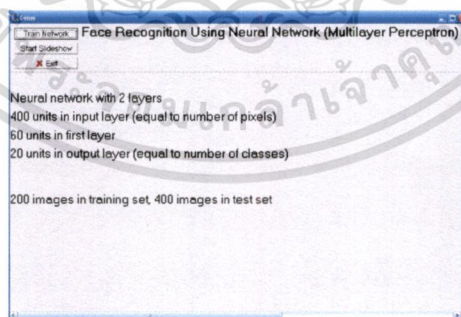
4.5.1 การระบุตัวบุคคลด้วยใบหน้า

การระบุตัวบุคคลด้วยใบหน้าจะนำวิธีการรู้จำตัวบุคคลด้วยโครงข่ายประสาทเทียมเข้ามาช่วยในการรู้จำภาพใบหน้าของแต่ละบุคคล โดยภาพข้อมูลขาเข้านั้นจะเป็นภาพขาวดำ (gray scale) เนื่องจากภาพที่ได้มาจากระบบวิดิทัศน์ซึ่งอยู่ในระบบสี RGB จึงต้องนำภาพมาปรับเปลี่ยนระบบสีเสียก่อน ดังแสดงในรูปที่ 4.13



รูปที่ 4.13 แสดงภาพใบหน้าที่ผ่านการปรับสี

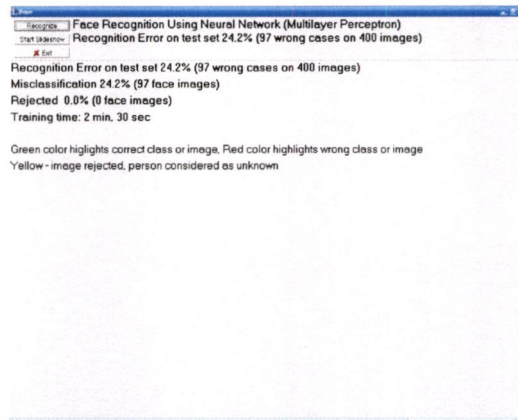
ข้อมูลภาพที่ใช้ในการเรียนรู้ของโครงข่ายประสาทเทียมนั้นจะมีทั้งหมด 20 กลุ่ม โดยแต่ละกลุ่มจะใช้ 1 วิดิทัศน์ในการเรียนรู้และใช้ภาพในการเรียนรู้กลุ่มละ 10 ภาพ ดังนั้นข้อมูลภาพที่ใช้ในการเรียนรู้ของระบบโครงข่ายประสาทเทียมจะมีทั้งสิ้น 200 ข้อมูลภาพ และข้อมูลภาพที่ใช้ในการทดสอบจะมาจากวิดิทัศน์ 5 วิดิทัศน์ วิดิทัศน์ละ 4 ข้อมูลภาพ ทำให้มีข้อมูลภาพที่ใช้ในการทดสอบทั้งสิ้น 400 ภาพ และกำหนดให้จำนวนรอบของการเรียนรู้ของโครงข่ายประสาทเทียมอยู่ที่ 100 รอบ ดังแสดงในรูปที่ 4.14



รูปที่ 4.14 แสดงการเรียนรู้ของโครงข่ายประสาทเทียม

เมื่อระบบทำการเรียนรู้เป็นที่เรียบร้อยแล้ว ระบบจะทำการทดสอบว่ามีอัตราความผิดพลาดของการรู้จำเป็นอัตราเท่าใด รู้จำผิดพลาดกี่กรณี หรือไม่สามารถระบุได้กี่กรณี ดังแสดงในรูปที่ 4.15

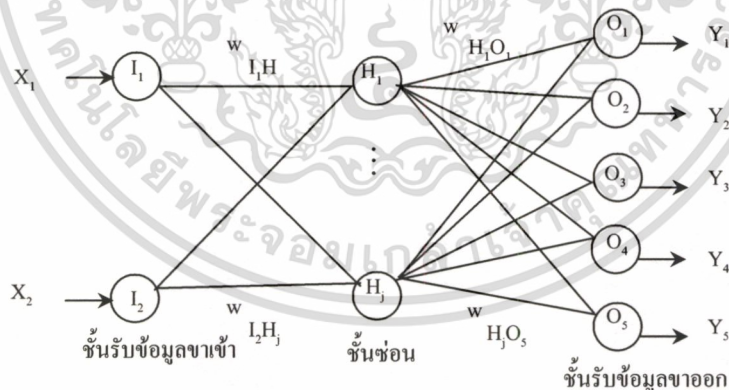
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.15 แสดงผลการรู้จำด้วยโครงข่ายประสาทเทียม

4.5.2 การระบุตัวบุคคลด้วยสีผิว

การรู้จำตัวบุคคลด้วยสีผิวนั้นก็จะใช้ระบบโครงข่ายประสาทเทียมในการรู้จำตัวบุคคลแต่จะมีความแตกต่างกับการระบุตัวบุคคลด้วยใบหน้าคือ ระบบที่ใช้ในการรู้จำสีผิวนั้น จะทำการแปลงค่าระบบสีจาก RGB ให้อยู่ในระบบสี YCbCr โดยละทิ้งค่าความสว่าง Y ไปเพื่อไม่ให้ค่าความสว่างผลต่อการรู้จำ โดยระบบโครงข่ายประสาทเทียมที่ใช้ที่มีข้อมูลนำเข้า 2 นิวรอน ตามลักษณะของระบบสีที่ใช้ คือ Cb และ Cr มีชั้นซ่อน 1 ชั้น และข้อมูลขาออกเป็น 20 นิวรอนตามจำนวนของบุคคลที่นำมารู้จำ ดังแสดงในรูปที่ 4.16



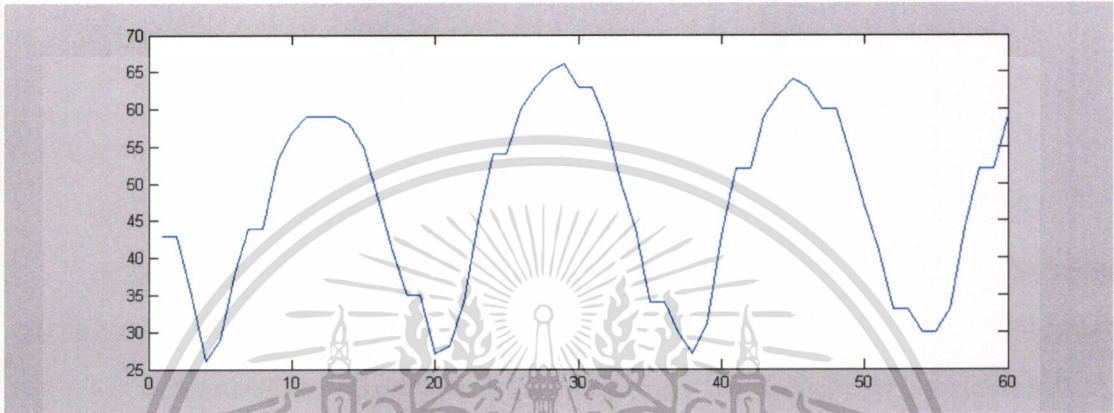
รูปที่ 4.16 ตัวอย่างโครงร่างของโครงข่ายประสาทเทียมแบบแบคพรอพพาเกชัน

4.5.3 การระบุตัวบุคคลด้วยลักษณะการเดิน (Gait)

การนำลักษณะการเดินหรือเกทเข้ามาช่วยในการรู้จำบุคคลนั้น จะต้องใช้ข้อมูลที่เป็นชุดข้อมูลภาพเข้ามาประมวลผลซึ่งอย่างน้อยที่สุดนั้นต้องใช้ประมาณ 60 ภาพ ดังนั้นมุกกล้องที่ใช้ในการรับภาพวีดิทัศน์นั้นควรจะเป็นบริเวณทางเดินยาว หรือบริเวณห้องโถงโล่ง ๆ เพื่อให้สามารถจับภาพบุคคลได้ครบเพื่อที่จะนำไปใช้ในการวิเคราะห์ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลักษณะการเดินที่ใช้ในงานวิจัยนี้จะใช้ความกว้างในการก้าวขา มาช่วยในการระบุตัวบุคคล ซึ่งเป็นการใช้ลักษณะทางกายภาพที่ปรากฏให้เห็น โดยยังมีลักษณะทางกายภาพอื่น ๆ อีกหลายชนิดที่ยังไม่ได้นำมาใช้ในการรู้จำบุคคล เมื่อใช้ความกว้างในการก้าวขามาช่วยรู้จำตัวบุคคล ภาพวิถีทัศนที่ใช้จะต้องใช้ภาพวิถีทัศนที่ถ่ายจากด้านข้างเท่านั้น ถึงจะนำความกว้างในการก้าวขามาช่วยในการระบุตัวบุคคลได้ โดยลักษณะของความกว้างในการเดินของบุคคลจะแสดงดังรูปที่ 4.17



รูปที่ 4.17 ความกว้างของการก้าวขาในการเดิน

จากรูปที่ 4.17 จะเป็นการแสดงลักษณะการเดินของบุคคลเป็นจำนวน 60 ข้อมูลภาพ เมื่อได้ข้อมูลภาพเข้ามาแล้ว จะนำไปทำการค้นหาความกว้างในการก้าวในภาพนั้น แล้วเก็บไว้ ทำซ้ำจนกว่าจะครบตามจำนวนข้อมูลภาพที่ต้องการ จากนั้นนำความกว้างที่ได้มาพล็อตกราฟ จะได้กราฟการเดินที่มีจังหวะกว้างและแคบ ต่อเนื่องสลับกันไป จากนั้น เราจะทำการหาค่าที่สูงที่สุดของกราฟแต่ละลูก ด้วยวิธี Local Maxima เพื่อหาว่าที่ข้อมูลภาพใดมีความกว้างมากที่สุดในช่วงนั้น ดังแสดงในรูปที่ 4.18 เมื่อได้ค่าความกว้างครบแล้วก็จะนำมาหาค่าเฉลี่ยแล้วนำไปเปรียบเทียบกับข้อมูลที่ได้เคยเก็บไว้ก่อนแล้ว



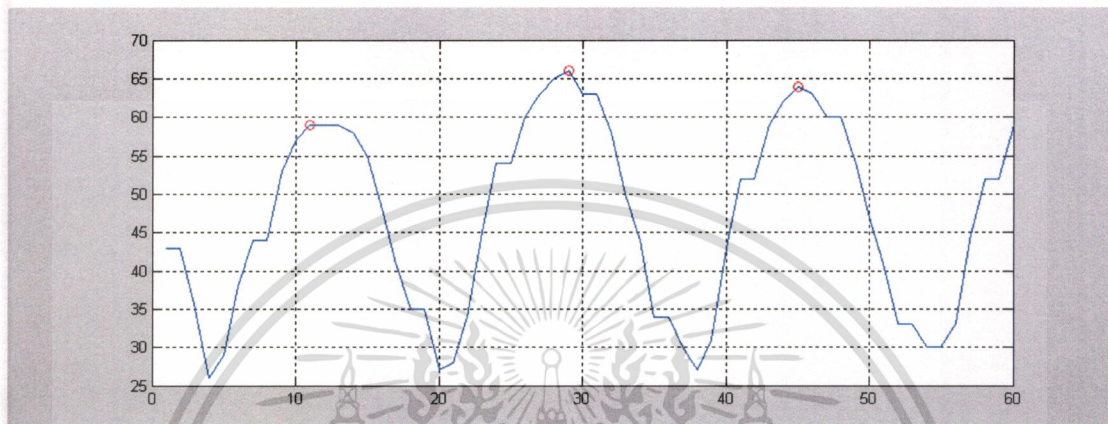
f=11



f=29



f=45



รูปที่ 4.18 แสดงการหาค่าความกว้างสูงสุด

ในส่วนต่อไปจะเป็นการกล่าวถึงในส่วนของแบบจำลองส่วนขยาย ซึ่งจะใช้ในการเก็บเอกลักษณ์ของบุคคลนั้น ๆ ไว้ชั่วระยะเวลาหนึ่ง เช่น วันต่อวัน โดยจะลบค่านี้ออกทิ้งเมื่อเปลี่ยนวันเป็นต้น ซึ่งในส่วนนี้ของแบบจำลองบุคคลส่วนขยายนี้จะมีเอกลักษณ์อยู่ 2 เอกลักษณ์ คือ สีเสื้อ และเวลา โดยจะมีรายละเอียดดังนี้

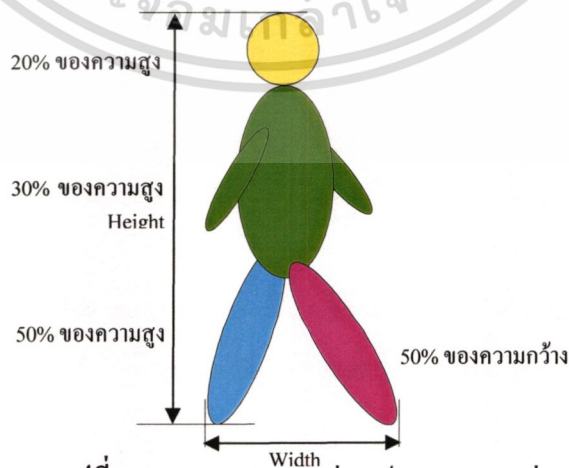
4.5.4 การระบุตัวบุคคลด้วยสีเครื่องแต่งกาย

สีของเสื้อผ้าจะนำมาช่วยในการระบุตัวบุคคลในกรณีที่ระบบได้รวมค่าความมั่นใจออกมาแต่ค่าความมั่นใจอันดับหนึ่งและอันดับสองมีความแตกต่างกันน้อยกว่าค่าขีดแบ่ง τ โดยสีของเสื้อผ้านั้นจะเก็บไว้ในขั้นตอนลงทะเบียน เช่น เมื่อมีบุคคลเดินผ่านประตูทางเข้าเข้ามา ซึ่งจะเป็นบริเวณที่สามารถจับเอกลักษณ์ต่าง ๆ ได้ครบสมบูรณ์ เมื่อระบบทำการระบุตัวบุคคลได้แล้ว ระบบจะทำการเพิ่มเอกลักษณ์ทางสีเสื้อผ้าเข้าไปไว้ในส่วนขยายของแบบจำลองบุคคลส่วนขยาย เพื่อเป็นส่วนช่วยในการระบุตัวบุคคลได้อีกทางหนึ่ง โดยสีของเสื้อผ้าที่นำมาช่วยในการระบุตัวบุคคลนี้จะใช้แบบวันต่อวัน และระบบสีที่ใช้นี้จะใช้ระบบสีนอร์มอลไลซ์ RGB เพื่อลดปัญหาทางด้านความแตกต่างกันของความสว่าง โดยลักษณะของเสื้อผ้าแสดงในรูปที่ 4.19



รูปที่ 4.19 ลักษณะสีเสื้อผ้า

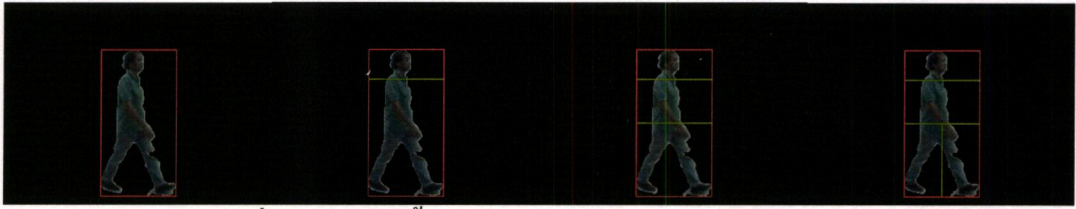
จากที่ได้กล่าวไปแล้วว่าสีของเครื่องแต่งกายจะถูกเก็บไว้ในส่วนขยายของแบบจำลองบุคคล โดยเริ่มแรกนั้นภายในแบบจำลองจะไม่มีข้อมูลของสีเครื่องแต่งกายอยู่เลย แต่เมื่อมีบุคคลผ่านบริเวณที่ฝ้าระวางตรงทางเข้า ซึ่งบริเวณนั้นสามารถได้รับข้อมูลของการรู้จำต่าง ๆ ได้ครบถ้วน ไม่ว่าจะเป็น ใบหน้า สีผิว ในระหว่างที่ระบบกำลังวิเคราะห์ที่อยู่ว่าเป็นบุคคลใด ระบบจะทำการคัดแยกส่วนที่เป็นสีของเสื้อผ้าเตรียมไว้ เมื่อระบบสามารถระบุตัวบุคคลได้แล้ว ระบบก็จะทำการเพิ่มข้อมูลของเครื่องแต่งกายไว้ในฐานข้อมูลส่วนขยาย ในกรณีที่ระบบได้รวมค่าความมั่นใจออกมาแต่ค่าความมั่นใจอันดับหนึ่งและอันดับสองมีความแตกต่างกันน้อยกว่าค่าขีดแบ่ง τ ระบบจะนำข้อมูลของสีเสื้อผ้าเข้ามาช่วยในการตัดสินใจ โดยทำการเปรียบเทียบสีเสื้อและทำการปรับค่าความมั่นใจ จะทำให้ระบบสามารถระบุได้ว่าบุคคลนั้นเป็นบุคคลใด โดยเราสามารถทำการแยกร่างกายออกเป็นส่วนตัวต่าง ๆ ได้โดยนำข้อมูลภาพที่ผ่านการลบฉากหลังเรียบร้อยแล้ว จากนั้นทำการหากรอบโครงร่างของบุคคลที่ตรวจจับได้ (Detected human bounding box) ระบบจะทำการวิเคราะห์เพื่อทำการแยกส่วนประกอบของร่างกายออกเป็นส่วนตัวต่างๆ โดยใช้แบบจำลองมนุษย์คาร์ดบอร์ด (Human cardboard model) เสนอ โดย Ju et al. (1996) ดังแสดงในรูปที่ 4.20



รูปที่ 4.20 แสดงการหาส่วนประกอบของร่างกาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อทราบถึงวิธีการค้นหาส่วนประกอบของร่างกายแล้วเราสามารถนำมาค้นหา ส่วนประกอบของร่างกายในภาพวิดีโอ ซึ่งผลของการค้นหาแสดงในรูปที่ 4.21



รูปที่ 4.21 แสดงขั้นตอนการค้นหาส่วนประกอบของร่างกาย

ซึ่งเมื่อทำตามวิธีการตามที่ได้กล่าวไปแล้วนั้น จะทำให้ได้บริเวณที่เป็นส่วนประกอบต่าง ๆ บนร่างกายอย่างคร่าว ๆ ดังแสดงในรูปที่ 4.21 ซึ่งจะต้องทำการปรับปรุงให้มีขนาดที่พอดีกับขนาด ที่แท้จริงของส่วนประกอบนั้น ๆ โดยจะเริ่มทำกับบริเวณที่เป็นส่วนหัวก่อน จากนั้นก็จะเป็นบริเวณ ลำตัว ส่วนบริเวณขา นั้นจะทำเป็นขั้นตอนสุดท้าย เพื่อจะได้มีความถูกต้องมากยิ่งขึ้นเมื่อนำไปใช้ในการ รู้จำ ดังแสดงในรูปที่ 4.22



รูปที่ 4.22 แสดงขั้นตอนการปรับขนาด

เมื่อได้ขนาดของกรอบที่พอดีกับตัวบุคคลแล้ว ระบบจะทำการตัดภาพออกเป็นส่วนๆ เพื่อให้ได้ บริเวณสี่ของเครื่องแต่งกายที่จะนำไปช่วยเสริมการระบุตัวบุคคล ดังแสดงในรูปที่ 4.23



รูปที่ 4.23 แสดงส่วนของเครื่องแต่งกายที่ถูกคัดแยก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพื่อให้มีความถูกต้องในการนำสีเสื้อผ้ามาช่วยในการรู้จำตัวบุคคล ระบบจะตรวจจับบริเวณที่เป็นสีผิว และทำการลบออกไปเท่าที่จะทำได้ เพื่อให้เหลือเพียงแค่บริเวณที่เป็นสีเสื้อ ดังแสดงในรูปที่ 4.24



รูปที่ 4.24 แสดงการกำจัดบริเวณที่ไม่ใช่สีเสื้อ

เมื่อได้บริเวณที่เป็นสีเสื้อแล้วจากนั้นจะนำจุดภาพ ทุก ๆ จุดภาพในแต่ละช่องสีมาหาค่าเฉลี่ยจะได้ $\bar{R}, \bar{G}, \bar{B}$ จากนั้นทำการแปลงระบบสีให้อยู่ในระบบสีนอร์มอลไลซ์ RGB

ในการเปรียบเทียบสีเสื้อนั้นจะใช้วิธีการหาความแตกต่างโดยใช้วิธีการวัดระยะทางแบบยูคลิดีเนียน (Euclidean Distance) ซึ่งถ้าค่าสีมีความเหมือนหรือใกล้เคียงกันค่าที่ได้จากการวิเคราะห์จะมีค่าน้อย ส่วนในกรณีที่ค่าสีมีความแตกต่างกันมากค่าที่ได้จะมีค่าที่มากขึ้น จากนั้นนำค่าที่ได้ซึ่งเป็นน้ำหนักมาทำการปรับลดค่าความมั่นใจของทุก ๆ บุคคลที่ระบบระบุออกมาซึ่งจะกล่าวในส่วนของการรวมค่าความมั่นใจ

4.5.5 เวลา

ในที่นี้จะนำเวลาเข้ามาช่วยในการระบุตัวบุคคล เนื่องจากระบบรักษาความปลอดภัยในปัจจุบันจะมีการติดกล้องวงจรปิดเป็นระบบเครือข่าย โดยจะติดกล้องไว้ตามจุดสำคัญต่าง ๆ เพื่อที่จะทำให้สามารถเฝ้าระวังได้ทั่วทั้งอาคารสถานที่ จึงได้มีความคิดที่ว่า เวลาสามารถนำมาช่วยกรองบุคคลได้ เนื่องจากเรารู้ตำแหน่งของกล้องที่อยู่ทุกตัวภายในอาคาร ระบบจะทำการบันทึกตำแหน่งล่าสุดของบุคคลนั้น ๆ ว่าปรากฏอยู่ที่กล้องใด เวลาใด เช่น เมื่อเวลา A ระบบตรวจได้ตรวจสอบและพบว่า นาย ก ได้ปรากฏอยู่ในกล้องทางออก และในเวลาเดียวกัน ที่ชั้น 5 ระบบได้ทำการตรวจสอบและพบว่า ขณะนี้มีบุคคลอยู่ที่ชั้น 5 คือ นาย ก และ นาย ข ซึ่งระบบไม่สามารถระบุได้แน่ชัดจากเอกลักษณ์ที่มีอยู่ในขณะนั้น แต่เมื่อระบบตรวจสอบกลับพบว่า ระบบได้พบ นาย ก ครั้งล่าสุดที่บริเวณทางออก ระบบจึงนำองค์ความรู้นี้มาช่วยในการตัดสินใจว่าบุคคลที่ปรากฏอยู่บนชั้น 5 นั้นคือ นาย ข เป็นต้น

โดยในที่นี้จะใช้ค่าความต่างของเวลามาเป็นค่าน้ำหนักในการปรับค่าความมั่นใจ ถ้าความต่างของเวลามีค่ามากกว่าค่าขีดแบ่ง τ ค่าที่ได้จะมีค่าน้อย ถ้าน้อยกว่าจะทำให้ได้ค่าน้ำหนักที่มากขึ้น

4.5.6 การปรับปรุงแบบจำลองระยะสั้น

แบบจำลองส่วนขยายจะเป็นที่เก็บข้อมูลเอกลักษณ์ของบุคคลไว้ชั่วคราวเมื่อถึงเวลาที่กำหนดระบบจะทำการลบข้อมูลของบุคคลที่ตรงตามเงื่อนไขออกจากแบบจำลองชั่วคราว ดังเช่น สี่เสื่อ ได้กำหนดให้ระบบทำการลบทิ้งเมื่อเปลี่ยนวัน ในส่วนของเวลาจะมีการปรับปรุงอยู่ตลอดเวลาตามที่ระบบได้ทำการตรวจสอบว่าบุคคลนี้ปรากฏขึ้น ณ เวลาใด ระบบจะทำการปรับปรุงเวลาให้เป็นเวลาล่าสุดที่ระบบตรวจพบบุคคลคนนั้น

4.6 การรวบรวมผลของการระบุตัวบุคคล

เนื่องจากการระบุตัวบุคคล ได้ใช้เอกลักษณ์หลายเอกลักษณ์เข้ามาใช้ในการระบุตัวบุคคล ผลที่ได้จากการระบุตัวบุคคลนั้นจะเป็นค่าความมั่นใจที่ระบบมีต่อบุคคลนั้น ๆ แล้วนำค่าความมั่นใจที่ได้จากแต่ละเอกลักษณ์มารวมกันเพื่อตรวจสอบดูว่าระบบให้ค่าความมั่นใจกับบุคคลใดมากที่สุด

4.6.1 ค่าความมั่นใจของเอกลักษณ์ใบหน้าและสรีระ

ค่าความมั่นใจของเอกลักษณ์ใบหน้าของแต่ละบุคคลนั้นจะอยู่ในช่วง 0-1 โดยแสดงได้ดังรูปที่ 4.25

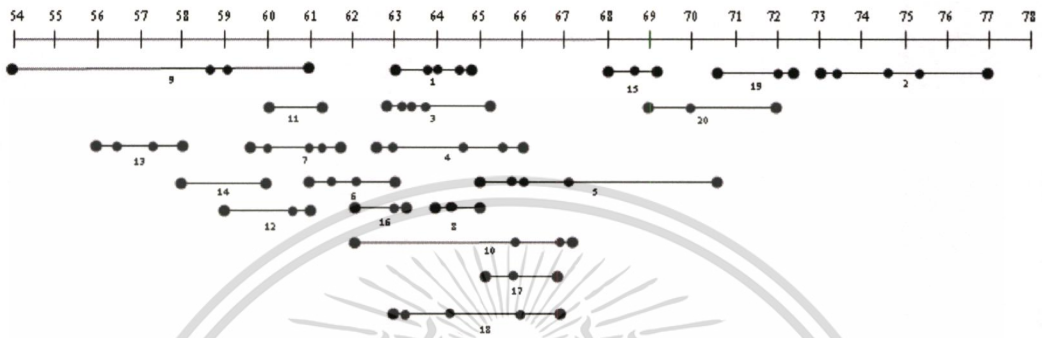
	ค่าความมั่นใจ	ค่าความมั่นใจ	ค่าความมั่นใจ	ค่าความมั่นใจ	ค่าความมั่นใจ
นาย ก	1	1	1	1	0
นาย ข	0	0	0	1	0
นาย ค	0	1	1	1	0
นาย ง	0	0	0	1	0
นาย จ	0	0	0.5	1	0
	(ก)	(ข)	(ค)	(ง)	(จ)

รูปที่ 4.25 แสดงค่าความมั่นใจที่ได้

จากรูปที่ 4.25(ก) ค่าความมั่นใจที่ระบบระบุได้นั้นจะมีความหมายว่าเป็นนาย ก 100% ในรูปที่ 4.25(ข) นั้นจะระบุว่าเป็นนาย ก 100 % และ นาย ค 100 % รูปที่ 4.25(ค) นั้นจะระบุว่าเป็นนาย ก 100% นาย ค 100% และนาย จ 50 % รูปที่ 4.25(ง) ค่าความมั่นใจที่ได้ของทุกบุคคลจะมีค่า 100% รูปที่ 4.25(จ) ระบบไม่สามารถระบุตัวบุคคลได้

4.6.2 ค่าความมั่นใจที่ได้จากลักษณะการเดิน

จะเป็นการนำค่าความกว้างในการก้าวขาที่สูงที่สุดมาเปรียบเทียบกับข้อมูลความกว้างในการก้าวขาที่สูงที่สุดซึ่งเป็นการเก็บข้อมูลในเชิงสถิติ โดยเปรียบเทียบว่าความกว้างนั้นมีขนาดเท่าใด และมีบุคคลใดบ้างที่มีช่วงก้าวขาอยู่ในบริเวณนั้น ดังแสดงในรูปที่ 4.26



รูปที่ 4.26 ผลทางสถิติของความกว้างในการก้าวขาของบุคคล 20 คน

ในการระบุค่าความมั่นใจของลักษณะการเดินนั้นจะพิจารณาว่าค่าความกว้างที่ได้อยู่ในช่วงบริเวณใด และมีบุคคลใดบ้างที่มีช่วงการก้าวขาอยู่ในบริเวณนั้น และค่าที่ได้ใกล้เคียงกับการเดินของบุคคลใดมากที่สุด ระบบก็จะให้ค่าความมั่นใจออกมา เช่น ค่าความกว้างที่ได้มีค่าเป็น 64 เมื่อสังเกตจากผลทางสถิติที่ได้บันทึกไว้จะมีบุคคลอยู่ 6 บุคคล ที่ระยะก้าวขาอยู่ในช่วงดังกล่าว และเมื่อพิจารณาถึงค่าความกว้างในการก้าวที่ใกล้เคียง นาย ก มีระยะก้าวขา 64 นาย ข มีระยะก้าวขา 63.33 และ นาย ค มีระยะก้าวขา 64.66 จะสังเกตได้ว่า นาย ก มีระยะก้าวขาที่เท่ากับค่าที่สังเกต นาย ข มีระยะการก้าวแตกต่างอยู่ 0.67 และนาย ค มีระยะการก้าวแตกต่างอยู่ 0.66 เมื่อนำมาคิดเป็นค่าความมั่นใจจะได้ว่า นาย ก มีค่าความมั่นใจ 100% นาย ข มีค่าความมั่นใจ 33% และ นาย ค จะมีค่าความมั่นใจ 34%

4.6.3 การรวมค่าความมั่นใจ

เมื่อได้ค่าความมั่นใจในแต่ละเอกลักษณ์มาแล้วนั้น จะนำค่าที่ได้มาทำการรวมเข้าด้วยกัน ดังแสดงในรูปที่ 4.27 โดยค่าความมั่นใจระหว่างใบหน้ากับผิวหนังและเวลานั้นหาได้จากสมการที่ 4.1

$$\text{Conf}(P_i) = \text{confFace}(P_i) + \text{confSkin}(P_i) \quad (4.1)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าความมั่นใจระหว่างสีผิวกับเกทและเวลานั้นหาได้จากสมการที่ 4.2

$$Conf(P_i) = confSkin(P_i) + confGait(P_i) \quad (4.2)$$



รูปที่ 4.27 แสดงขั้นตอนการรวมค่าความมั่นใจ

4.6.3.1 การปรับค่าความมั่นใจด้วยเวลา

เวลาที่นำมาใช้ในการระบุตัวบุคคลนั้นจะเป็นการนำมาปรับค่าความมั่นใจที่ได้จากการรวมค่าความมั่นใจของใบหน้ากับสีผิว หรือ ใบหน้ากับลักษณะการเดิน โดยเมื่อรวบรวมค่าความมั่นใจแล้วนั้น ระบบจะทำการตรวจสอบว่าบุคคลที่ได้ระบุค่าความมั่นใจออกมาแล้วนั้น ระบบเคยตรวจพบอยู่ในบริเวณอื่น ๆ ด้วยกล้องตัวอื่นหรือไม่ ถ้าบุคคลใดเคยตรวจพบ ระบบก็จะนำเวลาที่เคยตรวจพบมาทำการหาความแตกต่างกับเวลาปัจจุบัน ถ้ามีความแตกต่างกันน้อยกว่าค่าขีดแบ่ง (ในที่นี้ใช้ $\tau = 15$ วินาที ซึ่งจะขึ้นอยู่กับระยะห่างระหว่างกล้องในการเดินปกติ) ค่าความมั่นใจของบุคคลนั้นก็จะถูกลดลงครึ่งหนึ่ง ดังแสดงในสมการที่ 4.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$Conf(P_i) = \begin{cases} Conf(P_i)/2; & \text{if } (TimePCam_{present} - TimePCam_j) < \tau \\ Conf(P_i); & \text{else} \end{cases} \quad (4.3)$$

สมมุติให้ที่เวลา t ระบบตรวจสอบได้ว่า นาย ก อยู่ที่บริเวณกล้องที่ 1 และที่เวลา $t+5$ ระบบกำลังตรวจสอบบุคคลอยู่ว่าเป็นบุคคลใด และได้ระบุค่าความมั่นใจรวมออกมา เมื่อระบบทำการเปรียบเทียบค่าเวลาที่เก็บไว้ในแบบจำลองบุคคลส่วนขยายของทุกบุคคล ปรากฏว่ามีเวลาของนาย ก เก็บไว้และปรากฏที่บริเวณกล้องที่ 1 เมื่อเวลา t ซึ่งน้อยกว่าค่าขีดแบ่ง จึงทำให้ระบบทำการปรับลดค่าความมั่นใจที่ระบุออกมาในตอนแรก โดยทำการลดค่าความมั่นใจของนาย ก ลงครึ่งหนึ่ง

4.6.3.2 การปรับค่าความมั่นใจด้วยสีเครื่องแต่งกาย

ค่าความมั่นใจที่ได้จากสีเครื่องแต่งกายนั้นจะเป็นตัวปรับค่าน้ำหนัก ของการรวมค่าความมั่นใจต่าง ๆ เข้าด้วยกัน โดยเป็นการเปรียบเทียบค่าสีของบริเวณที่เป็นเสื้อผ้าของบุคคลที่ปรากฏอยู่ในวิดิทัศน์กับสีเสื้อผ้าที่เก็บอยู่ในแบบจำลองบุคคล โดยถ้าสีที่กำลังตรวจสอบอยู่นั้นแตกต่างกับสีเสื้อผ้าที่เก็บอยู่ในแบบจำลองบุคคลค่าที่ได้ก็จะสูง แสดงให้เห็นว่าไม่น่าจะเป็นคนเดียวกัน แต่ถ้าค่าที่ได้มีความแตกต่างกันน้อยก็แสดงให้เห็นว่าน่าจะเป็นคนเดียวกัน โดยปริภูมิสีที่นำมาใช้การเปรียบเทียบนั้นเป็นแบบปริภูมิสีนอร์มอล โลซ์ RGB โดยเปรียบเทียบความแตกต่างแบบ วิธีการวัดระยะทางแบบอูคลีเดียน (Euclidean Distance)

ในขั้นตอนการรวมค่าความมั่นใจนั้นจะแบ่งออกได้เป็นสองกรณีคือ การรวมค่าความมั่นใจระหว่างใบหน้ากับสีผิว และสีผิวกับลักษณะการเดิน เมื่อได้ค่าความมั่นใจรวมจากระบบมาแล้ว จากนั้นทำการคำนวณหาค่าความมั่นใจที่มากที่สุดอันดับที่หนึ่งและอันดับที่สอง แล้วนำมาหาความแตกต่าง ถ้ามีค่าความแตกต่างเกินค่าขีดแบ่ง (Threshold) ที่ได้กำหนดไว้ ระบบจะทำการระบุตัวบุคคลโดยใช้ค่าความมั่นใจที่มากที่สุด แต่ถ้าไม่เกินค่าขีดแบ่งระบบก็จะทำการนำสีเครื่องแต่งกายเข้ามาช่วยในการตรวจสอบเพื่อเพิ่มความถูกต้องในการระบุตัวบุคคลให้ดียิ่งขึ้น ดังสมการที่ 4.4

$$Conf(P_i) = \begin{cases} Conf(P_i) - confCloth_{color}(P_i); & \text{if } (\max_1(Conf(P_i)) - \max_2(Conf(P_i)) < \tau \\ 0; & \text{else} \end{cases} \quad (4.4)$$

เมื่อระบบได้ทำการปรับค่าความมั่นใจด้วยสีของเครื่องแต่งกายแล้ว จากนั้นระบบจะระบุตัวบุคคลโดยดูจากค่าความมั่นใจที่มากที่สุด

บทที่ 5

ผลการทดลอง

ในงานวิจัยนี้ได้ทำการทดลองสร้างระบบที่สามารถระบุตัวบุคคล โดยใช้เอกลักษณ์ต่าง ๆ เข้ามาช่วยในการระบุตัวบุคคลโดยจะแบ่งการทดสอบออกเป็นสองส่วนคือ ในส่วนแรกจะเป็นการทดสอบเอกลักษณ์เดี่ยวๆ ว่าเอกลักษณ์ใดบ้างที่จะสามารถนำมาใช้ในการรู้จำบุคคลได้ และในส่วนที่สองจะเป็นการนำเอกลักษณ์ที่ได้ทั้งหมดมาทำการรวมเข้าด้วยกันเพื่อใช้ในการระบุตัวบุคคล โดยจำนวนบุคคลที่ใช้ในการทดลองนี้มีจำนวน 20 คน ภาพวิดีโอที่สนับตัวอย่างที่ใช้ในการทดสอบนั้นจะประกอบไปด้วยภาพวิดีโอที่สนับใบหน้า สีผิว ภาพวิดีโอที่สนับของการเดิน และภาพวิดีโอที่สนับของเครื่องแต่งกายที่เปลี่ยนไปในแต่ละวัน โดยใช้ระยะเวลาในการเก็บข้อมูล 1 สัปดาห์

5.1 ผลของการลบฉากหลังออกจากภาพวิดีโอที่สนับ

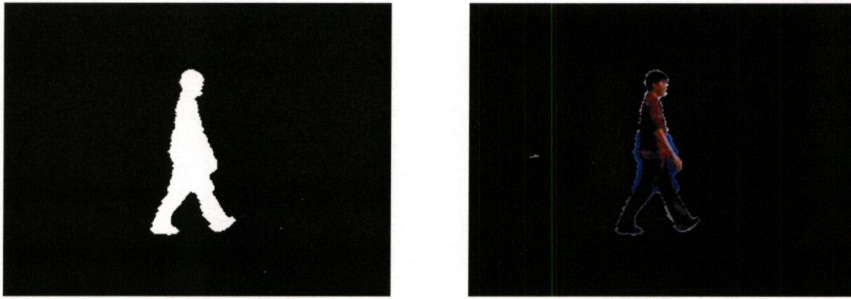
การที่จะทำให้ฉากหลังนั้นหายออกไปจากภาพโดยสมบูรณ์ตลอดเวลานั้นเป็นเรื่องที่ยาก อย่างเช่น เมื่อแสงเปลี่ยนไปอย่างมาก หรือ สิ่งของบางอย่างภายในฉากนั้นหายไป หรือ เพิ่มเติมเข้ามา ก็จะทำให้กระบวนการลบฉากหลังนั้นมีปัญหาเกิดขึ้นได้ ซึ่งถ้ากระบวนการลบฉากหลังไม่มีประสิทธิภาพเพียงพอ ก็จะทำให้กระบวนการขั้นต่อไปยากขึ้น ผลของกระบวนการลบฉากหลังจึงมีความสำคัญมาก ดังแสดงในรูปที่ 5.1



รูปที่ 5.1 แสดงผลของกระบวนการลบฉากหลัง

จากรูปที่ 5.1 จะเห็นได้ว่าภาพที่ภาพกระบวนการลบฉากหลังนั้นจะคงยังมีสิ่งรบกวนหลงเหลืออยู่ ดังนั้นจะต้องมีการประมวลผลภาพเพื่อที่จะทำให้สิ่งรบกวนนั้นลดน้อยลงไป โดยในที่นี้ได้ใช้วิธีทางมอร์ฟอลยีเข้ามาช่วย คือ Closing โดยจะทำการ Closing เพื่อกำจัดสิ่งรบกวนออกไปก่อนที่จะส่งไปทำการวิเคราะห์ต่อไป ดังแสดงในภาพที่ 5.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.2 แสดงผลของการกำจัดสิ่งรบกวน

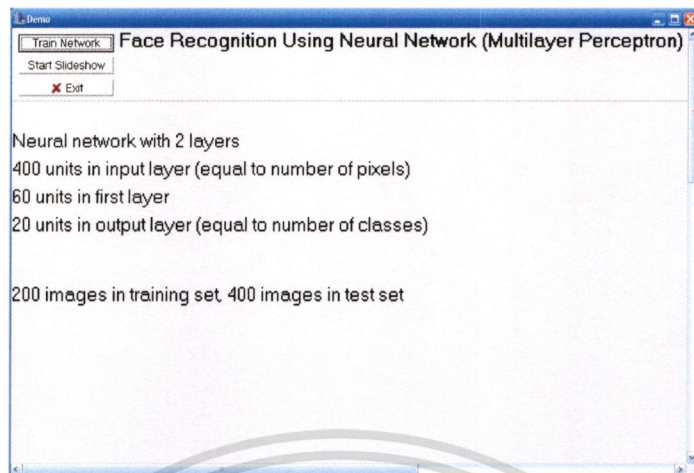
5.2 ผลของการรู้จำใบหน้า

ในการทดลองการรู้จำใบหน้านั้น จะทำการทดลองกับบุคคล 20 คน โดยแต่ละคนจะมีภาพทั้งหมดคนละ 30 ภาพ โดยใช้ 1 วิดีทัศน์ในการเรียนรู้ของโครงข่ายประสาทเทียม โดยใช้ภาพสำหรับเรียนรู้ 10 ภาพ และภาพที่ใช้ในการทดสอบโครงข่ายประสาทเทียม 20 ภาพ โดยภาพที่ใช้ในการทดสอบจะมาจากวิดีโอต่าง ๆ กัน 5 วิดีทัศน์ โดยใช้วิดีโอแต่ละ 4 ภาพ จึงทำให้มีภาพที่ใช้ในการเรียนรู้ทั้งหมด 200 ภาพ และภาพที่ใช้ในการทดสอบ 400 ภาพ เนื่องจากภาพในระบบ วิดีทัศน์นั้นจะอยู่ในระบบสี RGB แต่ในการรู้จำใบหน้านั้นจะใช้อยู่ในระบบขาวดำ (Gray Scale) ดังนั้นจึงต้องทำการปรับเปลี่ยนระบบสีให้เหมาะสมเสียก่อนจึงจะนำไปเรียนรู้และรู้จำได้ โดยชุดภาพที่ใช้ทดสอบแสดงดังรูปที่ 5.3



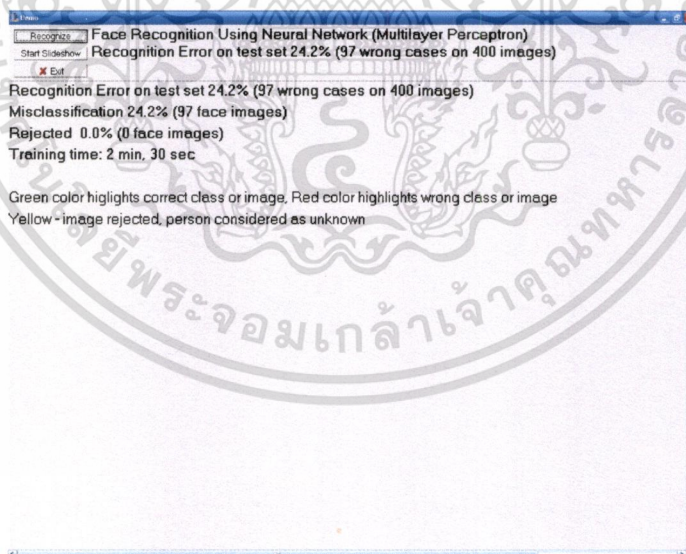
รูปที่ 5.3 ตัวอย่างข้อมูลใบหน้าที่ใช้ในการเรียนรู้

เมื่อได้ภาพแล้วนำไปให้ระบบโครงข่ายประสาทเทียมทำการเรียนรู้เสียก่อน โดยภาพใบหน้านี้จะมีขนาด ความกว้างเป็น 20 จุดภาพ และความสูงเป็น 20 จุดภาพ จึงทำให้ระบบโครงข่ายประสาทเทียมในชั้นรับข้อมูลขาเข้าจะมีจำนวน 400 นิวรอน (20×20) และประกอบไปด้วยชั้นซ่อนหนึ่งชั้นและชั้นรับข้อมูลขาออกโดยจะมีจำนวน 20 นิวรอน ด้วยกัน และจำนวนรอบที่ใช้ในการเรียนรู้นั้นจะใช้จำนวนรอบทั้งหมด 100 รอบ การเรียนรู้ของระบบโครงข่ายประสาทเทียมเอกสารแสดงดังรูปที่ 5.4 วนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.4 การเรียนรู้ของระบบโครงข่ายประสาทเทียม

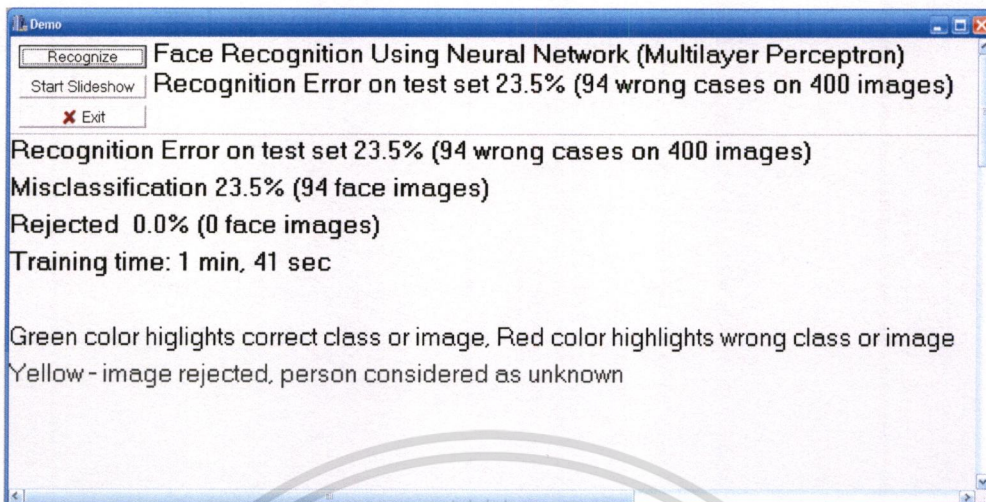
เมื่อระบบทำการเรียนรู้เสร็จแล้ว ขั้นตอนต่อไปก็จะเป็นการทดสอบระบบโครงข่ายประสาทเทียมว่าที่เรียนรู้ไปแล้วนั้นสามารถระบุใบหน้าได้แม่นยำมากน้อยเพียงไร ต้องเพิ่มข้อมูลในการเรียนรู้หรือไม่ ต้องทำการเพิ่มจำนวนรอบในการเรียนรู้หรือไม่ โดยผลของการเรียนรู้ที่แสดงดังรูปที่ 5.5



รูปที่ 5.5 แสดงผลของการเรียนรู้ของระบบโครงข่ายประสาทเทียม

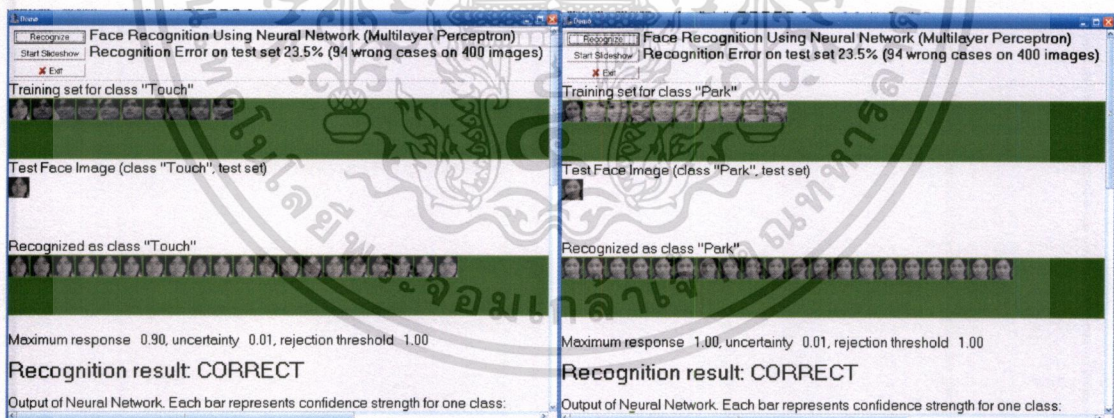
จากรูปที่ 5.5 ผลการเรียนรู้ของโครงข่ายประสาทเทียมนั้นมีอัตราความผิดพลาดอยู่ที่ 24.2% คิดเป็นจำนวนภาพที่ผิดพลาดได้ 97 ภาพ จากทั้งหมด 400 ภาพ ซึ่งเป็นอัตราความผิดพลาดที่สูงพอสมควร จึงได้มีการปรับจำนวนนิเวรอนในชั้นซ่อนให้มีความเหมาะสมมากยิ่งขึ้น จากนั้นทำการเรียนรู้อีกครั้ง ผลที่ได้แสดงดังภาพที่ 5.6

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ซึ่งการเขียนขึ้นเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



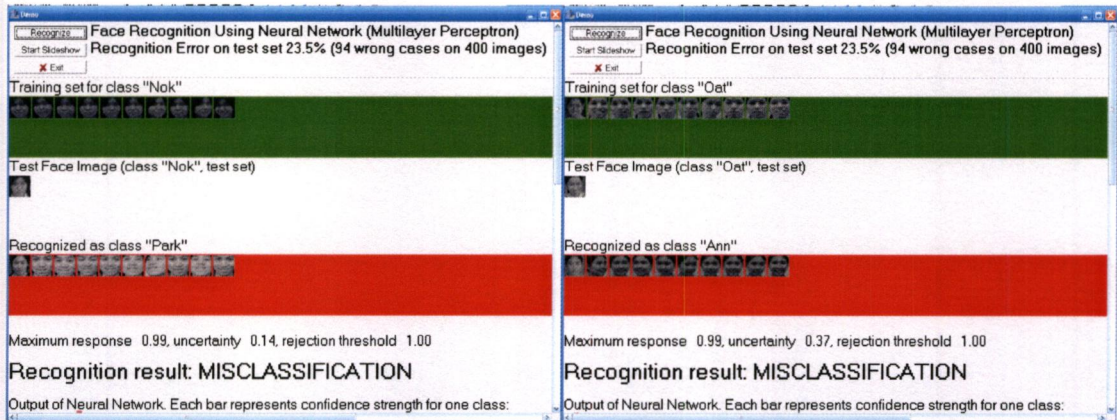
รูปที่ 5.6 แสดงผลการเรียนรู้ของโครงข่ายประสาทเทียมครั้งที่สอง

ผลที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียมในครั้งที่สองนั้นอัตราความผิดพลาดนั้นลดลงเหลือ 23.5% โดยรู้จำข้อมูลภาพผิดพลาดไป 94 ข้อมูลภาพ ซึ่งไม่ว่าจะพยายามปรับค่าตัวแปรต่าง ๆ อย่างไรก็ตาม ก็ไม่สามารถทำให้อัตราความผิดพลาดนั้นลดน้อยลงกว่านี้ได้ ซึ่งอัตราความผิดพลาดที่ได้นั้นก็อยู่ในระดับที่ยอมรับได้ ผลของการทดสอบแสดงดังรูปที่ 5.7



รูปที่ 5.7 แสดงผลของการระบุตัวบุคคลได้อย่างถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.8 แสดงผลของการระบุตัวบุคคลที่ผิดพลาด

5.3 การรู้จำบุคคลด้วยสีผิว

จากที่ได้กล่าวไปแล้วว่าก่อนที่จะทำการรู้จำด้วยสีผิวได้นั้นระบบจะต้องทำการปรับเปลี่ยนระบบสีจาก RGB ให้อยู่ในระบบสี YCbCr ก่อน โดยละทิ้งค่าความสว่าง Y เพื่อลดปัญหาที่สีผิวผิดเพี้ยนไปเนื่องจากค่าความสว่าง การรู้จำบุคคลด้วยสีผิวนั้นก็จะใช้ระบบโครงข่ายประสาทเทียมเข้ามาช่วยในการรู้จำ โดยข้อมูลที่ใช้ในการรู้จำนั้นจะเป็นค่าสี Cb และ Cr ซึ่งเป็นค่าสีที่ได้ทำการหาค่าเฉลี่ย จึงทำให้ชั้นรับข้อมูลขาเข้าของระบบโครงข่ายประสาทเทียมมีสองนิวรอนและชั้นซ่อนหนึ่งชั้น และชั้นรับข้อมูลขาออก 20 นิวรอน โดยมีรอบการเรียนรู้อยู่ที่ 100 รอบ จำนวนข้อมูลสีผิวที่ใช้ในการเรียนรู้นั้นจะมีคนละ 20 ค่า จึงมีข้อมูลที่ใช้ในการเรียนรู้ 400 ค่า และข้อมูลที่ใช้ทดสอบคนละ 10 ค่า ก็จะมีข้อมูลที่ใช้ในการทดสอบ 200 ค่า โดยผลของการรู้จำด้วยสีผิว แสดงในตารางที่ 5.1



รูปที่ 5.9 ตัวอย่างสีผิวในระบบสี RGB

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.1 ผลของการรู้จำด้วยสีผิว

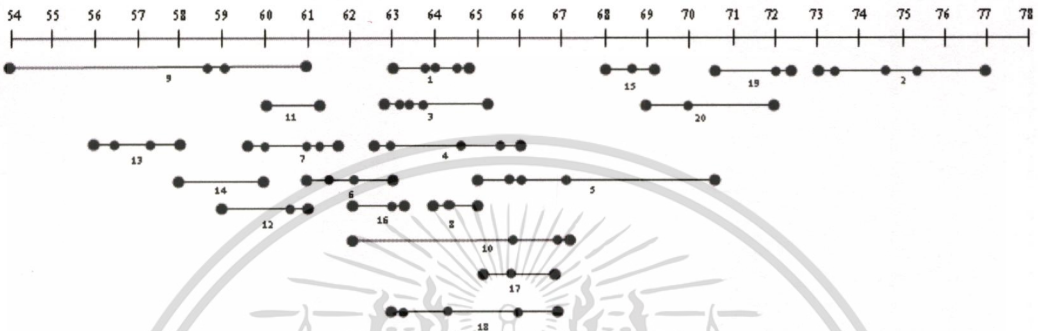
	B	D	It	P	W	A	N	O	S	T	Pe	Ta	Ao	Ak	Bo	K	Ki	Pa	Pp	To
B	0		10																	
D	10	0																		
It			10																	
P				10																
W					10															
A						4					2							1	2	1
N						2	0				4								3	1
O								10												
S									0	1							8			1
T										2							1		2	5
Pe						3					3							4		
Ta						2					2	0						4	2	
Ao													0				7			3
Ak						4								0	2		3		1	
Bo										1					1				8	
K						3									1	0		1	5	
Ki											9						1			
Pa						1												1	8	
Pp																			10	
To																	10			0

จากตารางที่ 5.1 จะพบว่าอัตราความผิดพลาดของการรู้จำด้วยสีผิวนั้นอยู่ที่ 69% ซึ่งเป็นอัตราความผิดพลาดที่สูงมาก เนื่องจากวิธีการวิเคราะห์ในการตรวจจับบริเวณที่เป็นสีผิวนั้นไม่สามารถแยกบริเวณที่ไม่เป็นสีผิวได้หมด ดังสังเกตได้จากรูปที่ 5.9 จะพบว่าสิ่งที่ปรากฏอยู่บนใบหน้าไม่ว่าจะเป็น ดวงตา คิ้ว ผม หรือ แขนงตา จะรวมอยู่ในบริเวณที่เป็นสีผิว ไม่สามารถที่จะกำจัดสิ่งเหล่านี้ออกไปได้ด้วยวิธีการตรวจจับสีผิวที่ใช้ในงานวิจัยนี้ จึงอาจทำให้ผลของสีผิวที่ได้เกิดความผิดพลาดขึ้น จึงทำให้เอกลักษณ์สีผิวไม่สามารถนำมาใช้ในการรู้จำตัวบุคคลโดยใช้เพียงเอกลักษณ์เดียวได้ (ภาพที่ใช้ในการทดลองจะเป็นภาพที่มาจากกล้องเดียวกัน ในเวลาที่ใกล้เคียงกัน) แต่พอที่จะสามารถนำมาใช้เป็นตัวช่วยในการรู้จำตัวบุคคลได้โดยนำไปใช้ร่วมกับเอกลักษณ์อื่น ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.4 การรู้จำบุคคลด้วยลักษณะการเดิน

การรู้จำบุคคลด้วยลักษณะการเดินนั้นจะนำความกว้างในการก้าวเดินของแต่ละบุคคลมาใช้ในการรู้จำ ซึ่งจะเป็นการเก็บข้อมูลทางสถิติไว้ว่าบุคคลใดเดินด้วยความกว้างเท่าใดบ้าง ดังแสดงในภาพที่ 5.10



รูปที่ 5.10 ผลทางสถิติของความกว้างในการก้าวของบุคคล 20 คน

ในการทดสอบการรู้จำตัวบุคคลโดยใช้ลักษณะการเดินนั้นจะทำการทดสอบกับวิดิทัศน์จำนวนสามชุด ซึ่งผลการทดลองแสดงในตารางที่ 5.2

ตารางที่ 5.2 ผลการรู้จำด้วยเกท

	B	D	It	P	W	A	N	O	S	T	Pe	Ta	Ao	Ak	Bo	K	Ki	Pa	Pp	To
B	1		1		1															
D		3																		
It	1		1	1																
P	2			0		1														
W					3															
A					3	0														
N							2		1											
O	1			1				1												
S							1	2	0											
T	1									2										
Pe								1			2									
Ta										1		2								
Ao													3							
Ak				1								1		1						

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.2 ผลการรู้จำด้วยเกท (ต่อ)

	B	D	It	P	W	A	N	O	S	T	Pe	Ta	Ao	Ak	Bo	K	Ki	Pa	Pp	To
Bo															3					
K						1	1									1				
Ki			1														2			
Pa	1																1	1		
Pp																			3	
To															1				1	1

จากตารางที่ 5.2 จะพบว่าอัตราความผิดพลาดในการรู้จำบุคคลด้วยเกทนั้นอยู่ที่ 46.66% ซึ่งเป็นอัตราความผิดพลาดที่สูงพอสมควร แต่ยังคงนำมาเป็นตัวช่วยในการรู้จำตัวบุคคลได้เมื่อนำไปใช้ร่วมกับเอกลักษณ์อื่น ๆ ในการระบุตัวบุคคล

5.5 การรู้จำตัวบุคคลด้วยใบหน้า สีผิว เวลาและสีเสื้อ

ในการทดลองนี้จะเป็นการนำค่าความมั่นใจที่ได้จากเอกลักษณ์ใบหน้ากับสีผิวมารวมเข้าด้วยกันเพื่อใช้ในการหาค่าความมั่นใจ โดยที่ผลการรู้จำตัวบุคคลโดยใช้สองเอกลักษณ์แสดงในตารางที่ 5.3

ตารางที่ 5.3 ผลการรู้จำด้วยใบหน้ากับสีผิว

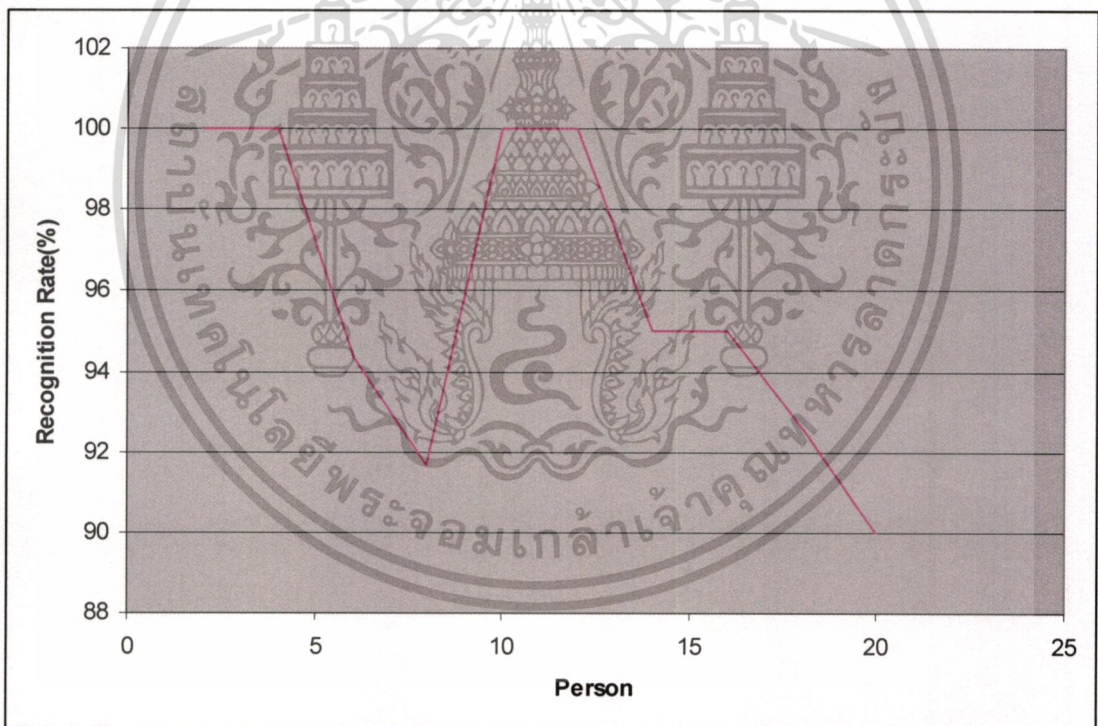
	B	D	It	P	W	A	N	O	S	T	Pe	Ta	Ao	Ak	Bo	K	Ki	Pa	Pp	To
B	3																			
D		1			1												1			
It	1		2																	
P				3																
W					3															
A						3														
N							3													
O								3												
S									3											
T										3										
Pe											3									
Ta												3								
Ao													3							
Ak														3						

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมีเหตุดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.3 ผลการรู้จำด้วยใบหน้ากับสปีผิว (ต่อ)

	B	D	It	P	W	A	N	O	S	T	Pe	Ta	Ao	Ak	Bo	K	Ki	Pa	Pp	To	
Bo															3						
K						1							1			1					
Ki																	3				
Pa																		3			
Pp																			3		
To				1																	2

จากตารางที่ 5.3 จะพบว่าเมื่อนำหลายเอกลักษณ์เข้ามาช่วยในการรู้จำตัวบุคคลจะทำให้ อัตราความผิดพลาดนั้นอยู่ที่ 10% ซึ่งมีค่าน้อยกว่าการใช้เอกลักษณ์ใบหน้าหรือสปีผิวเพียงอย่างเดียว โดยอัตราความถูกต้องในการรู้จำตัวบุคคลเมื่อเทียบกับจำนวนตัวบุคคลที่เพิ่มขึ้น แสดงดังรูปที่ 5.11



รูปที่ 5.11 แสดงอัตราความถูกต้องกับจำนวนบุคคลที่เพิ่มขึ้น

จากรูปที่ 5.11 จะเห็นว่าค่าอัตราความถูกต้องที่ได้จากกราฟจะมีการแกว่งของอัตราความถูกต้อง อาจจะมีสาเหตุอันเนื่องมาจากการเรียนรู้ของระบบโครงข่ายประสาทเทียม ที่ได้กำหนดเพียงค่าการเรียนรู้ไว้ที่ 100 รอบเพียงอย่างเดียว โดยไม่ได้กำหนดอัตราความผิดพลาดที่ยอมรับได้ในการเรียนรู้ของโครงข่ายประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.6 การรู้จำตัวบุคคลด้วยสีผิว ลักษณะการเดิน เวลา และสีเสื้อ

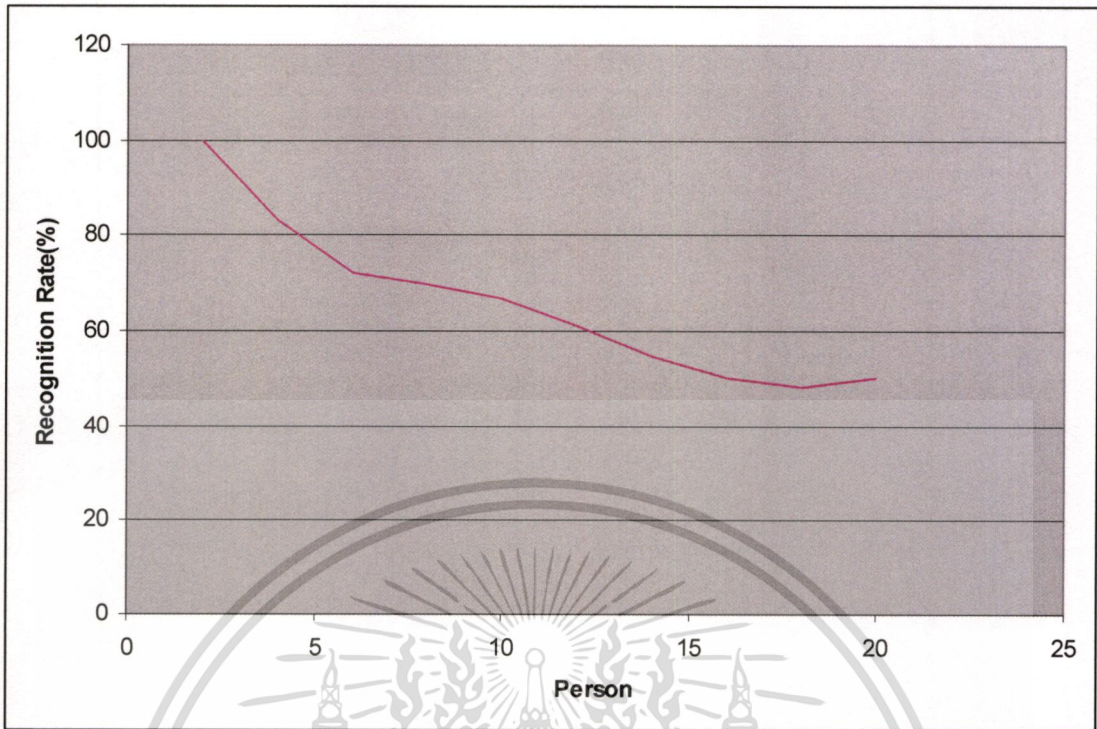
ในการทดลองนี้จะเป็นการนำค่าความมั่นใจที่ได้จากเอกลักษณ์สีผิวกับลักษณะการเดิน เวลา และสีเสื้อ มารวมเข้าด้วยกันเพื่อใช้ในการหาค่าความมั่นใจ เนื่องจากเป็นภาพวิทัศน์ ด้านข้างทำให้เอกลักษณ์ใบหน้าไม่สามารถนำมาใช้ในการระบุตัวบุคคลได้ โดยที่ผลการรู้จำตัวบุคคลโดยใช้สามเอกลักษณ์แสดงในตารางที่ 5.4

ตารางที่ 5.4 แสดงผลการรู้จำด้วยสีผิวกับเกทและสีเสื้อ

	B	D	It	P	W	A	N	O	S	T	Pe	Ta	Ao	Ak	Bo	K	Ki	Pa	Pp	To
B	2		1																	
D		2			1															
It	1		2																	
P				0																
W					2										1					
A						2										1				
N							1					2								
O								2												
S									0	1			1			1				
T	1		1							0									1	
Pe							2				0	1								
Ta												3								
Ao													3							
Ak												1		2						
Bo															3					
K						1	2										0			
Ki										1							2			
Pa	1					1		1										0		
Pp														1						2
To															1					2

จากตารางที่ 5.4 จะพบว่าเมื่อนำหลายเอกลักษณ์เข้ามาช่วยในการรู้จำตัวบุคคลจะทำให้ อัตราความผิดพลาดนั้นอยู่ที่ 50% ซึ่งมีค่าน้อยกว่าการใช้เอกลักษณ์สีผิวหรือเกทเพียงอย่างเดียว โดย อัตราความถูกต้องเมื่อเทียบกับจำนวนบุคคลที่เพิ่มขึ้น แสดงดังรูปที่ 5.12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.12 แสดงอัตราความถูกต้องกับจำนวนบุคคลที่เพิ่มขึ้น

จากรูปที่ 5.12 จะสังเกตเห็นว่าอัตราความถูกต้องนั้นจะลดลงเรื่อยๆ เมื่อมีจำนวนบุคคลเพิ่มมากขึ้น เนื่องจากวิธีการวิเคราะห์โดยใช้สีผิวจะมีประสิทธิภาพลดลง อันเนื่องมาจากข้อมูลของสีผิวจะเกาะกลุ่มกันแน่นขึ้นเรื่อยๆ เมื่อจำนวนของบุคคลเพิ่มมากขึ้น รวมไปถึงลักษณะการเดินโดยใช้ความกว้างสูงสุดในการก้าวขาจะมีการเกาะกลุ่มของช่วงก้าวขาที่ใกล้เคียงกัน ซึ่งเป็นผลที่ทำให้อัตราการรู้จำลดลง มีข้อสังเกตอยู่ประการหนึ่งก็คืออัตราการรู้จำเมื่อมีจำนวนบุคคล 20 คน ค่าของอัตราการรู้จำนั้นกลับเพิ่มขึ้น ซึ่งเป็นผลมาจากความถูกต้องในการระบุตัวบุคคลด้วยลักษณะการเดิน ซึ่งบุคคลที่ 19 และ 20 นั้น มีลักษณะการเดินที่แยกออกมาจากกลุ่มของข้อมูล จึงทำให้อัตราการรู้จำเพิ่มกลับขึ้นไปอยู่ที่ 50% ดังนั้นอาจจะต้องมีการปรับปรุงวิธีการในการวิเคราะห์เอกลักษณ์สีผิว และลักษณะการเดิน รวมไปถึงอาจจะต้องนำเอกลักษณ์ชนิดอื่นๆ เข้ามาช่วยในการระบุตัวบุคคลจากวิทัศน์ด้านข้าง เพื่อให้ผลที่ได้มีความถูกต้องมากยิ่งขึ้นไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยชิ้นนี้ได้นำเสนอวิธีการทางคอมพิวเตอร์วิชัน เพื่อใช้ในการวิเคราะห์และค้นหาเอกลักษณ์ต่าง ๆ จากภาพวิดีโอเพื่อนำไปใช้กับระบบระบุตัวบุคคลโดยใช้หลายเอกลักษณ์ โดยทดสอบกับบุคคลจำนวน 20 คน โดยใช้เอกลักษณ์ใบหน้า เอกลักษณ์สีผิว เอกลักษณ์การเดิน เวลาและสีเสื้อ เข้ามาประกอบรวมกันกลายเป็นระบบระบุตัวบุคคลโดยอัตโนมัติขึ้นมา โดยในการทดลองจะทำการทดลองกับข้อมูลทดสอบที่ได้เตรียมไว้ โดยแบ่งออกเป็น 2 ช่วงด้วยกัน คือ ในช่วงแรกนั้นจะเป็นการสร้างแบบจำลองหรือองค์ความรู้ให้กับระบบ ได้แก่การเรียนรู้ภาพใบหน้า และสีผิว ด้วยโครงข่ายประสาทเทียม เพื่อที่จะนำมาใช้ในการรู้จำ และข้อมูลลักษณะการเดินเชิงสถิติเพื่อใช้ในการรู้จำด้วยลักษณะการเดิน ในส่วนที่สองก็จะเป็นการทดสอบเมื่อใช้เอกลักษณ์ใดเอกลักษณ์หนึ่งจะมีความถูกต้องแม่นยำเท่าใด และเมื่อนำแต่ละเอกลักษณ์เข้ามารวมกันผลของการรู้จำที่ได้จะเป็นอย่างไร ซึ่งผลการทดลองสามารถสรุปได้ดังนี้

6.1 สรุปผลการวิจัย

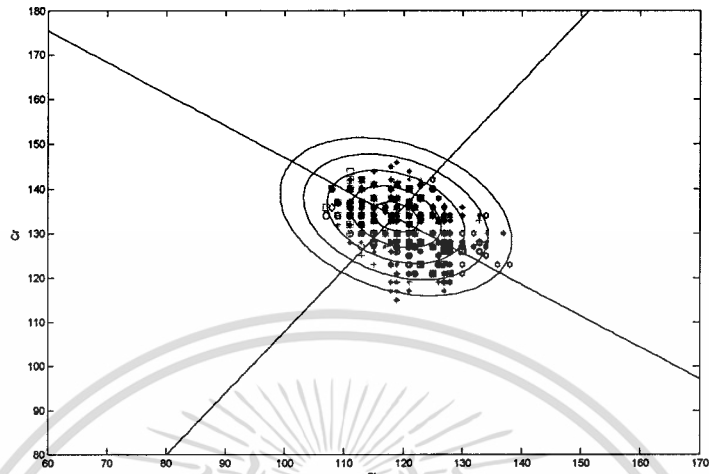
6.1.1 การระบุตัวบุคคลด้วยใบหน้า

การระบุตัวบุคคลด้วยใบหน้านั้นถือว่ามีมีความถูกต้องแม่นยำเป็นอันดับหนึ่ง โดยมีอัตราความผิดพลาดอยู่ที่ 23.5% ซึ่งผลที่ได้อยู่ในเกณฑ์ที่น่าพอใจแต่มีข้อสังเกตว่าภาพใบหน้าที่จะใช้ระบุตัวบุคคลให้ได้อย่างถูกต้องและแม่นยำนั้น ภาพของใบหน้านั้นควรจะเป็นภาพที่ค่อนข้างตรง หรือ เอียงเพียงเล็กน้อย เนื่องจากถ้าภาพใบหน้ามีการเอียง หรือ หันด้านข้างให้กับกล้อง ระบบจะทำการระบุตัวบุคคลผิดพลาดได้ และชุดข้อมูลภาพที่นำมาใช้ในการเรียนรู้ก็มีความสำคัญ ถ้าชุดข้อมูลภาพที่นำมาใช้ในการเรียนรู้ไม่ครอบคลุมเพียงพอ ระบบก็อาจจะเกิดความผิดพลาดในการระบุตัวบุคคลได้

6.1.2 การระบุตัวบุคคลด้วยสีผิว

การระบุตัวบุคคลด้วยสีผิวนั้นมีความถูกต้องแม่นยำเป็นอันดับที่สามารถจากใบหน้าและลักษณะการเดิน โดยมีอัตราความผิดพลาดอยู่ที่ 69% ซึ่งผลที่ได้ไม่เป็นที่น่าพอใจนัก เนื่องจากการตรวจจับสีผิวนั้นไม่สามารถแยกสิ่งที่ไม่ได้เป็นสีผิวซึ่งอยู่บนใบหน้าได้เช่น คิ้ว หรือ แวนตา เป็นต้น และเมื่อสังเกตผลจากการนำค่าสีผิวมาทำการสร้างกราฟระหว่างค่าสี Cb กับ Cr จะเห็นว่าค่าสีผิวจะเกาะกันเป็นกลุ่ม ซึ่งเมื่อจำนวนของบุคคลยังมีน้อยอยู่ค่าสียังไม่เกาะกลุ่มกันอย่างหนาแน่นนัก จึงทำให้สามารถระบุตัวบุคคลได้ แต่พอจำนวนบุคคลเพิ่มมากขึ้นทำให้ค่าของสีผิว

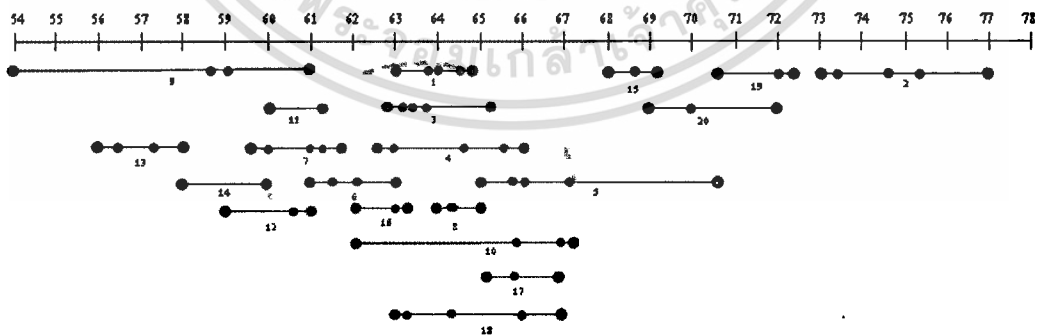
เกาะกลุ่มกันมากขึ้น ดังแสดงในรูปที่ 6.1 ทำให้การระบุตัวบุคคลด้วยสีผิวนั้นมีความผิดพลาดมากขึ้นตามไปด้วย



รูปที่ 6.1 แสดงการเกาะกลุ่มของสีผิว

6.1.3 การระบุตัวบุคคลด้วยลักษณะการเดิน

การระบุตัวบุคคลด้วยลักษณะการเดินนั้นจะมีความถูกต้องแม่นยำเป็นอันดับที่สองรองจากใบหน้า โดยมีอัตราความผิดพลาดอยู่ที่ 46.66% ซึ่งเมื่อสังเกตจากการเก็บข้อมูลความกว้างในการเดินนั้นลักษณะของข้อมุลนั้นจะไปรวมกันอยู่ตรงกลาง (ความกว้างในการก้าวขาในช่วง 60 – 67) ดังแสดงในรูปที่ 6.2 ซึ่งมีเมื่อมีบุคคลมากขึ้นก็จะทำให้อัตราความถูกต้องในการระบุตัวบุคคลนั้นลดลง



รูปที่ 6.2 แสดงการเกาะกลุ่มของลักษณะการเดิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6.1.4 การระบุตัวบุคคลด้วยใบหน้า สีผิว เวลา สีเสื้อ

เป็นการนำค่าความมั่นใจของเอกลักษณ์ใบหน้า สีผิว และเวลา มารวมเข้าด้วยกัน ซึ่งผลของการระบุตัวบุคคลนั้นจะมีอัตราความผิดพลาดอยู่ที่ 10% ซึ่งเป็นผลมาจากความถูกต้องของเอกลักษณ์ใบหน้าโดยที่มีสีผิวเข้ามาช่วยอีกทางหนึ่ง

6.1.5 การระบุตัวบุคคลด้วยสีผิว ลักษณะการเดิน เวลา และสีเสื้อ

เป็นการนำค่าความมั่นใจของเอกลักษณ์สีผิว ลักษณะการเดิน และเวลา มารวมเข้าด้วยกัน โดยมีสีเสื้อเป็นค่านำหนักในกรณีที่ระบบไม่มั่นใจว่าบุคคลนั้น ๆ เป็นใคร โดยจะนำสีเสื้อเข้ามาปรับค่าความมั่นใจให้ลดลง ในกรณีที่สีเสื้อไม่ใกล้เคียงกับข้อมูลที่เก็บไว้ ซึ่งเมื่อดูจากผลการทดลองนั้น การระบุตัวบุคคลด้วยลักษณะการเดินนั้นจะให้ค่าความมั่นใจออกมาค่อนข้างสูง เช่น ในกรณีที่ระบุตัวบุคคลได้ถูกต้องนั้น ค่าความมั่นใจของบุคคลที่ไม่ถูกต้องก็จะมีค่าที่ค่อนข้างสูงหรือเกือบใกล้เคียงกับค่าที่ถูกต้อง ส่วนในกรณีที่ระบุตัวบุคคลไม่ถูกต้องนั้นค่าความมั่นใจที่ได้ก็มีค่าที่สูงอีกเช่นกันจึงทำให้เมื่อนำไปรวมกับเอกลักษณ์อื่น ๆ แล้วอาจจะไปทำให้ค่าความมั่นใจที่ต่ำ ซึ่งเป็นค่าที่ไม่ถูกต้อง กลับมีค่าสูงขึ้นเกินกว่าค่าที่ถูกต้องเดิม จึงต้องมีการหาปรับวิธีการหานำหนักให้เหมาะสมมากกว่านี้

6.2 ข้อเสนอแนะในการทำวิจัยต่อไป

6.2.1 ในกระบวนการรู้จำใบหน้านั้นมีจำนวนข้อมูลขาเข้าที่มีขนาดใหญ่ ซึ่งควรจะลดขนาดของข้อมูลภาพเสียก่อน เพื่อช่วยลดเวลาในการวิเคราะห์คำนวณลง

6.2.2 พัฒนาระบบที่สามารถขยายภาพ (Zoom) เข้าไปยังบริเวณที่มีเอกลักษณ์ที่สนใจปรากฏอยู่ เพื่อให้ได้ภาพข้อมูลที่ชัดเจนมากยิ่งขึ้น เช่น ใบหน้า เป็นต้น

6.2.3 ระบบการระบุตัวบุคคลด้วยสีผิวจะไม่สามารถใช้ระบุตัวบุคคลได้ถ้าจำนวนบุคคลมากเกินไปควรหาเอกลักษณ์อื่น ๆ เข้ามาทดแทน

6.2.3 ระบบการระบุตัวบุคคลด้วยลักษณะการเดินยังไม่เหมาะสม ควรปรับปรุงหรือใช้วิธีการวิเคราะห์ลักษณะการเดินวิธีอื่น เช่น ซิมิลาริตีพล็อต (Similarity Plots) เป็นต้น

บรรณานุกรม

จรัญ จันทลักษณ์ และอนันต์ชัย เขื่อนธรรม. 2529. สถิติเบื้องต้นแบบประยุกต์. พิมพ์ครั้งที่ 2
กรุงเทพฯ : ไทยวัฒนาพานิช.

ล้วน สายยศ และอังคณา สายยศ. 2540. สถิติวิทยาทางการวิจัย. พิมพ์ครั้งที่ 3. กรุงเทพฯ:
สุวีริยาสาส์น

สนั่น ศรีสุข. “การลดมิติข้อมูลสำหรับระบบรู้จำ ภาพใบหน้าโดยใช้วิธี PCA” เอกสารรวมเล่มการ
ประชุมวิชาการทางวิศวกรรมไฟฟ้า ครั้งที่ 21 พ.ศ. 2541.

A. Bobick and A. Johnson, “Gait recognition using static, activity-specific parameters,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2001, pp. (I)423–(I)430.

A. Johnson and A. Bobick, “A multi-view method for gait recognition using static body parameters,” in *Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication*, 2001, pp. 301–311.

A. Kale, A. Rajagopalan, N. Cuntoor, and V. Kruger, “Gait-based recognition of humans using continuous HMMs,” in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, Washington, DC, 2002, pp. 336–341.

C. BenAbdelkader, R. Cutler, H. Nanda, and L. Davis, “EigenGait: motion- based recognition of people using image self-similarity,” in *Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication*, 2001, pp. 312–317.

C. BenAbdelkader, R. Culter, and L. Davis, “Motion-based recognition of people in eigengait space,” in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, Washington, DC, 2002, pp. 267–274.

C. Y. Yam, M. S. Nixon, and J. N. Carter, “Extended model-based automatic gait recognition of walking and running,” in *Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication*, 2001, pp. 278–283.

C. Y. Yam, M. S. Nixon, and J. N. Carter, “Gait recognition by walking and running: a model-based approach,” in *Proc. Asia Conf. Computer Vision*, Melbourne, Australian, 2002, pp. 1–6.

- D. Cunado, M. S. Nixon, and J. N. Carter, "Using gait as a biometric: via phase-weighted magnitude spectra," in *Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication*, 1997, pp. 95–102.
- D. Cunado, M. S. Nixon, and J. N. Carter, "Extracting a human gait model for use as a biometric," in *Proc. Inst. Elect. Eng. (IEE) Colloq. Computer Vision for Virtual Human Modeling*, 1998, pp. 11/1–11/4.
- G. W. Cottrell and M. Fleming, "Face recognition using unsupervised feature extraction," in *Proc. Int. Neural Network Conf.*, vol. 1, Paris, France, July 9–13, 1990, pp. 322–325.
- H. Bourlard and Y. Kamp, "Auto-association by multilayer perceptrons and singular value decomposition," *Biological Cybern.*, vol. 59, pp. 291–294, 1988.
- H. Murase and R. Sakai, "Moving object recognition in eigenspace representation: gait analysis and lip reading," *Pattern Recognit. Lett.*, vol. 17, no. 2, pp. 155–162, 1996.
- I. Haritaoglu, D. Harwood, and L. S. Davis. 1998. "A Real Time System for Detecting and Tracking People", *International Conference on Face and Gesture Recognition*, Nara, Japan.
- J. D. Shutler, M. S. Nixon, and C. J. Harris, "Statistical gait recognition via temporal moments," in *Proc. IEEE Southwest Symp. Image Analysis and Interpretation*, 2000, pp. 291–295.
- J. Sullivan and S. Carlsson. 2002. "Recognizing and Tracking Human Action", *Proc. European Conf. Computer Vision*.
- MH. Yang, D.J.Kriegman, and N. Ahuja. 2002. "Detecting Faces in Images: A Survey" 24(1): 34–58. *IEEE Tran. Pattern Analysis and Machine Intelligence (PAMI)*.
- R. Culter and L. Davis, "Robust real-time periodic motion detection, analysis and applications," *IEEE Trans. Pattern Recognit. Machine Intell.*, vol. 13, pp. 129–155, Feb. 2000.
- Ross Cutler and Matthew Turk. "View-based Interpretation of Real-time Optical Flow for Gesture Recognition," *IEEE International Conference on Automatic Face and Gesture Recognition*, April 1998, Nara, Japan.
- R. C. Gonzalez and R. E. Woods. 2001. *Digital Image Processing*. 2nd Ed. Prentice-Hall.
- R. Jain, R. Kasturi, B. G. Schunck. 1995. *Machine Vision*. McGraw-Hill.
- R. O. Duda, P. E. Hart and D. G. Stork. 2001. *Pattern Classification*. 2nd Ed. A Wiley-Interscience.

- Rowley, H.A.; Baluja, S.; Kanade, T., Rotation Invariant Neural Network-Based Face Detection Computer Vision and Pattern Recognition, 1998. **Proceedings. 1998 IEEE Computer Society Conference on 23-25 June 1998 Page(s):963 - 963**
- S. A. Niyogi and E. H. Adelson, “Analyzing gait with spatio-temporal surface,” in *Proc. IEEE Workshop Motion of Non-Rigid and Articulated Objects*, 1994, pp. 64–69.
- T. Horprasert, D. Harwood, and L. S. Davis.1999. “A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection”, **Proc. IEEE Frame Rate Workshop, Greece.**
- Vezhnevets V., Sazonov V., Andreeva A. “A Survey on Pixel-Based Skin Color Detection Techniques.” Proc. Graphicon-2003, pp. 85-92, Moscow, Russia, September 2003
- W.Hu, T.Tan, L. Wang and S. Maybank, “A Survey on Visual Surveillance of Object Motion and Behaviors.” In **Proc. IEEE System, Man, And Cybernetics part c: application and reviews, vol.34 No.3 August 2004.**
- WY. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld. 2003. “Face recognition: A literature survey” 35(4): 399-458, **ACM Compute. Survey.**
- Y. Sugaya and K. Kanatani. 2004 “Extracting Moving Objects from a Moving Camera Video Sequence” 279-284. **Proc. 10th Symposium on Sensing via Image Information (SSII 2004), Japan.**



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

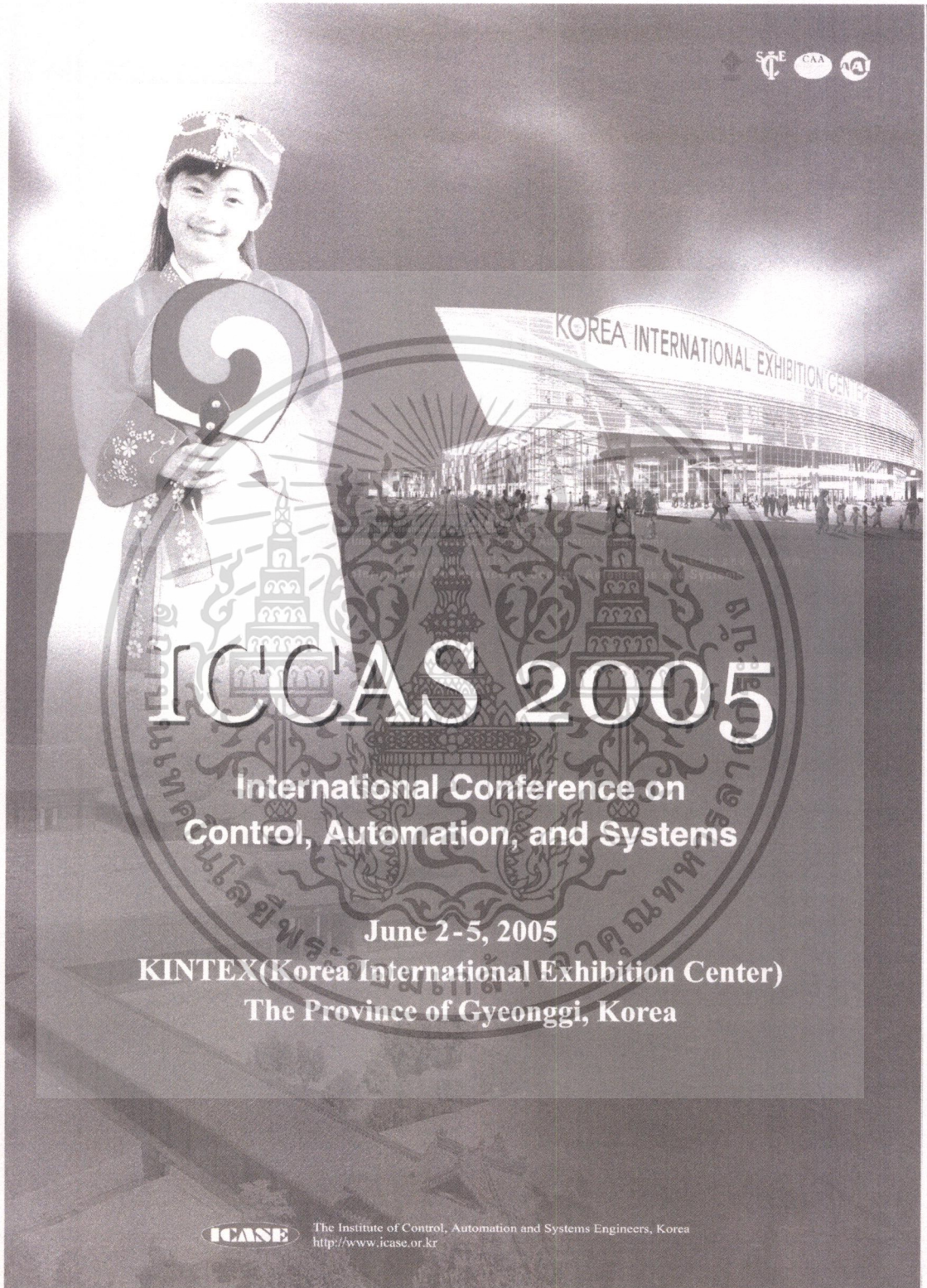
ภาคผนวก ก.

ผลงานวิจัยที่ได้รับการตีพิมพ์เผยแพร่

1. D. Swangpol, T. Chalidabhongse, “Person Identification using Multiple Cues”, Proc. ICCAS2005, 2005.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ICCAS2005

June 2-5, KINTEX, Gyeonggi-Do, Korea

Automatic Person Identification using Multiple Cues

Danuwat Swangpol and Thanarat Chalidabhongse

Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
(Tel: +66-2-737-2551; E-mail: s4067410@kmitl.ac.th, thanarat@it.kmitl.ac.th)

Abstract: This paper describes a method for vision-based person identification that can detect, track, and recognize person from video using multiple cues: height and dressing colors. The method does not require constrained target's pose or fully frontal face image to identify the person. First, the system, which is connected to a pan-tilt-zoom camera, detects target using motion detection and human cardboard model. The system keeps tracking the moving target while it is trying to identify whether it is a human and identify who it is among the registered persons in the database. To segment the moving target from the background scene, we employ a version of background subtraction technique and some spatial filtering. Once the target is segmented, we then align the target with the generic human cardboard model to verify whether the detected target is a human. If the target is identified as a human, the cardboard model is also used to segment the body parts to obtain some salient features such as head, torso, and legs. The whole body silhouette is also analyzed to obtain the target's shape information such as height and slimmness. We then use these multiple cues (at present, we use shirt color, trousers color, and body height) to recognize the target using a supervised self-organization process. We preliminary tested the system on a set of 5 subjects with multiple clothes. The recognition rate is 100% if the person is wearing the clothes that were learned before. In case a person wears new dresses the system fail to identify. This means height is not enough to classify persons. We plan to extend the work by adding more cues such as skin color, and face recognition by utilizing the zoom capability of the camera to obtain high resolution view of face; then, evaluate the system with more subjects.

Keywords: Person Identification, Active Camera, Pattern Recognition, Image Analysis.

1. INTRODUCTION

As the need of intelligent system for security and surveillance grows rapidly, it becomes more urgent and important to establish viable methods for automatic person identification that can detect, track and recognize persons in real-time. Our goal is to develop a person identification system that can detect, track, and recognize people from video using multiple cues such as face, shape, clothes, skin color and gender. The system must not require constrained target's pose or fully frontal face image to identify the person. Our underlying motivation is to develop an automatic system to guard a watched area such as offices, dormitories, apartments, etc. The system is responsible to detect target that approaches the watched area, then identify the target. If the target is known person, the system put the recognition result and time stamp on the entry log file. Otherwise, it sends alarm signal to an authorized person via instant message over the network or SMS message over the mobile phone.

The main focus of this paper is on the person identification from unconstrained video part. First, the system, which is connected to a pan-tilt-zoom camera, detects target using motion detection and human cardboard model. The system keeps tracking the moving target while it is trying to identify whether it is a human and identify who it is among the registered persons in the database. This continuous identifying while tracking can provide a significant benefit as mentioned in [1] as *continuity of identity*. To segment the moving target from the background scene, we employ a version of background subtraction technique [2] and some spatial filtering. Once the target is segmented, we then align the target with the generic human cardboard model to verify whether the detected target is a human. If the target is identified as a human, the cardboard model is also used to segment the body parts to obtain some salient features such as head, torso, and legs. The whole body silhouette is also analyzed to obtain the target's shape information such as height and slimmness. We then use these multiple cues (at present, we use shirt color, trousers color,

and body height) to recognize the target using a supervised self-organization process.

The organization of this paper is as follows: Section 2 reviews the literature related to our work. Section 3 describes our proposed methods in segmenting and identifying person from images. Section 4 presents the experiments and results. Finally, we conclude the work and discuss about future works in Section 5.

2. LITERATURE SURVEY

In this section, we review existing techniques relevant to the problem of detecting, tracking, and identifying people in video. We divide into 3 subsections: first reviews person identification techniques, second reviews methods in detecting and tracking human in video, and the last subsection discusses about works that tracking and identifying human using active camera(s).

2.1 Person Identification

As one of the most successful applications of image analysis and understanding, person identification in video has recently received significant attention, especially during the past several years. To identify a person, most of the early works focused on biometrics such as fingerprint, iris, and face recognition. However, fingerprint and iris require the subject to directly interact with the sensors. This limits the domain of applications. Face recognition seems to be more feasible in the sense that it uses passive sensor. For exhaustive surveys of face analysis techniques, the reader is referred to Yang et al. [3] for survey on face detection and Zhao et al. [4] for face recognition survey. However, most of the typical systems work on static high-resolution frontal face images. These are not feasible to some application domain such as surveillance. In surveillance, people might not need to know they are observed, and cameras are usually placed in distance from the subjects. In this circumstance, obtaining high-resolution frontal face images are very difficult. There are typically two approaches solving this

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

problem. One uses other cues such as shape and gait [5,6] to recognizing person, the other trying to model the face in multiple views [7-9].

2.2 Human Detection and Tracking

Human detection and tracking have been very active area of research in the past two decades. Two comprehensive surveys of computer vision-based human motion analysis have been published by Gavrilu in 1999 [10], and by Moeshund and Granum in 2001 [11]. More recently, Kentaro et al. [12] presented a new exemplar-based probabilistic paradigm for visual tracking. Their approach, called Metric Mixture (M^2), combines the advantages of exemplar-based models [13] with a probabilistic framework introduced in [14] into a single probabilistic exemplar-based paradigm. Sullivan and Carlsson [15] presented a method in recognizing and tracking human in action. In their approach, view-based activity recognition serves as an input to a human body location tracker. By recognizing the image of a person's posture as corresponding to a particular stored key frame, they then map body locations from the key frames to actual frames using shape matching algorithm based on appearance similarity.

2.3 Human Tracking using Active Cameras

Recently, many researchers have been working on active and multiple cooperative cameras. This approach takes benefit of multi-scale imaging. Using wide-angle or stereo cameras to segment and track target in video, then use active camera(s) to zoom-in to capture face region for identification. The works in this category include work done by Peixoto et al. [16]. They use a wide-angle camera plus a ground plane assumption to estimate 3D location of the object. Then, the object is tracked using a binocular active camera. Another work done by Collins et al. [17] uses multiple cooperative cameras to detect, track, and recognize person. With multi-view images from multiple calibrated cameras, they can estimate 3D location of the target using triangulation and groundplane assumption. The closest related work in the literature we found is the one proposed by Hampapur et al. [1]. Their goal is to build a system that can answer the "who is where" question. Their system uses two static cameras for wide baseline stereo to estimate 3D position of the subject's head using triangulation technique. Then another set of two Pan-Tilt-Zoom (PTZ) cameras is used to zoom in on the moving target and catalog the face. However, they did not mentioned in detail about their method in identifying.

background subtraction technique proposed by Horprasert et al. [2] is used.

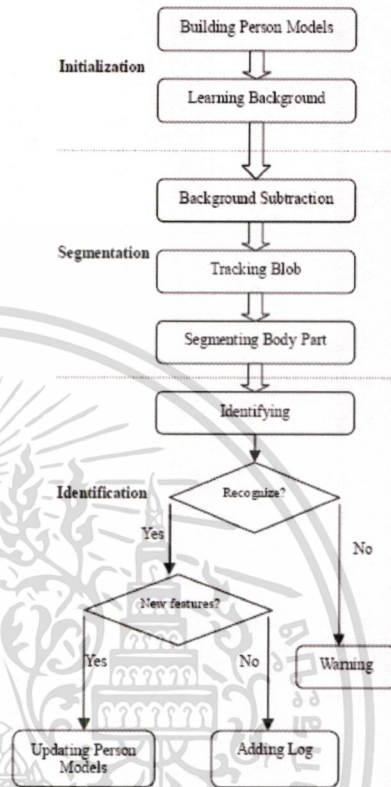


Fig. 1 Block diagram of our system.

3. OUR APPROACH

Our approach use only one PTZ camera connected to the PC. The camera is mounted and points to the entrance area of the lab. Once it detects motion, it will keep tracking while identifying the moving object. Fig. 1 shows a block diagram of our system. The processes are divided into three phases: initialization, segmentation, and identification phases.

3.1 Initialization Phase

In this phase, personal models and background models are constructed. To build background model, the camera is panned in certain step while capturing background scene; a simple panorama background model is then constructed. Fig. 2 shows the background model. In this work, background modeling and

To construct initial person models, we capture video sequence of each subject while walking into an observed area. The system then subtracts the input video from the background model. Result of the subtraction is shown in Fig.3. Then, the cardboard model is used to segment body parts (details in next subsection). The color of torso, color of legs, and height are then registered as feature of the person. We then iterate the process, until we register all the known subjects.

3.2 Segmentation Phase

This phase concerns with detecting and segmenting the target. At running mode, the camera is initially set to watch the entrance area of the room. Whenever, there is a person come in the scene, the system detects it using background subtraction. A spatial median filtering is employed to clean some erroneous segmentation. Connected component analysis is then



Fig. 2 Background model.



Fig. 3 Subtraction results.

performed. Only big regions remain. A simple cardboard model [18] is then used to verify whether the segmented region is person-like. While identifying, the camera keeps tracking the subject with Pan-Tilt capability. This enhances the identifying efficiency. The cardboard is also used to segment salient body parts such as head, torso, and legs. Subject's height, torso color, and leg color are then computed, taken as a feature vector for this subject. Fig. 4 shows a result of body part segmentation.



Fig. 4 Body part segmentation result.

3.3 Identification Phase

To recognize the subject, we utilize feedforward backpropagation neural networks. At current version, feature vector composes of 7 values: $H, T_R, T_G, T_B, L_R, L_G, L_B$, where H, T , and L are height, torso, and leg orderly. R, G, B are typical color dimension. The networks compose of 3 layers with 7 input nodes which correspond to the feature vector. The hidden layer contains hidden nodes that are fully linked to the input and output nodes. The number of output nodes are $N+1$, where N is the number of register persons. The other node is for "unknown person" answer.

4. DISCUSSION

The preliminary system was implemented using C/C++ and OpenCV library running on a Pentium4 PC. The camera is Sony EVI-D100 connected to the PC via a frame grabber. To

evaluate the system, we tested it on a set of 5 subjects with multiple clothes. Fig. 5 shows some of the training set of data and the segmentation results. The recognition rate is 100% if the person is wearing the clothes that were learned before. In case a person wears new dresses the system fail to identify. This means height is not enough to classify persons. Other available collateral information that we can extract from image such as skin color, gender, face, or speech may be used in enhancing recognition.

We plan to extend the work as follows:

- Working on more sophisticate background subtraction method from moving camera video sequence. Example of works in the literature include Sugaya and Kanatani [18], and Bartoli et al. [19].
- Adding more cues; skin color and face recognition by utilizing the zoom capability of the camera to obtain high-resolution view of face.
- Evaluating the system with more subjects.

REFERENCES

- [1] A. Hampapur, S. Pankanti, A. Senior, Y.L.Tian, L. Brown, and R. Boll, "Face Cataloger: Multi-Scale Imaging for Relating identity to Location", Proc. IEEE Intl. Conf. Advanced Video and Signal Based Surveillance (AVSS 2003), IEEE Press, 2003, pp.13-20.
- [2] T. Horprasert, D. Harwood, and L. S. Davis, "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection", Proc. IEEE Frame Rate Workshop, Greece, 1999.
- [3] M.H. Yang, D.J.Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey", IEEE Tran. Pattern Analysis and Machine Intelligence (PAMI), 24(1): 34-58, 2002.
- [4] WY. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, "Face recognition: A literature survey", ACM Comput. Survey, 35(4): 399-458, 2003.
- [5] A. Kale, A. K. R. Chowdhury, and R. Chellappa, "Towards a View Invariant Gait Recognition Algorithm", dddd
- [6] S. Sarkar, P.J. Phillips, Z. Liu, I. Robledo, P. Grother, and K. Bowyer, "The Human ID Gait Challenge Problem: Data Sets, Performance, and Analysis", IEEE Transaction on Pattern Analysis and Machine Intelligence, 27(2): 162-177, 2005.
- [7] Y. Li, S. Gong, and H. Liddell, "Support Vector Regression and Classification Based Multi-view Face Detection and Recognition", Proc. Intl. Conf. Face and Gesture Recognition (FG 2000), France, 2000.
- [8] C. Sanderson, and S. Bengio, "Extrapolating Single View Face Models for Multi-View Recognition", Proc. Intl. Conf. Intelligent Sensors, Sensor Networks, and Information Processing, Australia, 2004.
- [9] Y. Gao, S.C. Hui, and A.C.M. Fong, "A Multiview Facial Analysis Technique for Identity Authentication", IEEE Pervasive Computing, IEEE Communication Society, 2003.
- [10] D.M. Gavrila, "The Visual Analysis of Human Movement: A Survey", Computer Vision and Image Understanding (CVIU), 1999.
- [11] T.B. Moeslund, and E. Granum, "A Survey of Computer Vision-based Human Motion Capture", Computer Vision and Image Understanding (CVIU), 2001.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [12] K. Toyama, and A. Blake, "Probabilistic Tracking with Exemplars in a Metric Space", *Intl. Journal of Computer Vision*, 48(1): 9-19, 2002.
- [13] D. Gavrila, and V. Philomin, "Real-time Object Detection for Smart Vehicles", *Proc. Intl. Conf. Computer Vision*, pp:87-93, 1999.
- [14] B. Frey and N. Jojic, "Learning Graphical Models of Images, Videos, and their Spatial Transformations", *Proc. Conf. Uncertainty in Artificial Intelligence*, 2000.
- [15] J. Sullivan and S. Carlsson, "Recognizing and Tracking Human Action", *Proc. European Conf. Computer Vision*, 2002.
- [16] Peixoto, Batista, and Araujo, "A Surveillance System Combining Peripheral and Foveated Motion Tracking", *Proc. Intl. Conf. Pattern Recognition*, 1998.
- [17] R.T. Collins, A.J. Lipton, H. Fujiyoshi, and T. Kanade, "Algorithms for Cooperative Multisensor Surveillance", *Proceedings of the IEEE*, 89(10):1456-1477, 2001.
- [18] Y. Sugaya and K. Kanatani, "Extracting Moving Objects from a Moving Camera Video Sequence", *Proc. 10th Symposium on Sensing via Image Information (SSII 2004)*, Japan, 2004, pp:279-284.
- [19] A. Bartoli, N. Dalai, B. Bose, and R. Horaud, "From Video Sequence to Motion Panoramas", *Proc. the Workshop on Motion and Video Computing (MOTION'02)*, IEEE Press, 2002.



Fig. 5 More segmentation result images

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ	นายคุณวัฒน์ แสงผล
วัน เดือน ปีเกิด	9 มกราคม พ.ศ.2522
ที่อยู่	เลขที่ 222/117 หมู่ 7 ต.คลองสวนพลู อ.พระนครศรีอยุธยา จ.พระนครศรีอยุธยา 13000
ประวัติการศึกษา	2543 วิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีมหานคร



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้