ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

ระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอ

Human Face Recognition from Video

โดย

นายพนมพร สาคร

รหัส 45066064

H002108

อาจารย์ที่ปรึกษา

ดร. ธนารัตน์ ชลิดาพงศ์

วัน เดือน ปี 06 ม.ค. 2550
เลขทะเบียน 021108
เลขเรียกหนังสือ วพ. พ1885 ส546
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ภาคเรียนที่ 2 ปีการศึกษา 2546

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อหัวข้อ	ระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอ
นักศึกษา	นายพนมพร สาคร
อาจารย์ที่ปรึกษา	ดร. ธนารัตน์ ชลิดาพงศ์
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2546

บทคัดย่อ

ในการพัฒนาระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอนั้น สามารถที่จะแบ่งการทำงานออกเป็น 2 ส่วนใหญ่ๆ ด้วยกัน คือ การตรวจหาใบหน้ามนุษย์ (Human Face Detection) จากภาพวีดิโอ เป็นการหาตำแหน่งใบหน้ามนุษย์ที่อยู่ในภาพวีดิโอในแต่ละเฟรมออกมา เพื่อนำไปวิเคราะห์ต่อในส่วนถัดไป คือส่วนของการรู้จำใบหน้าของมนุษย์ (Human Face Recognition) เป็นส่วนที่ระบบซึ่งผ่านเรียนรู้เพื่อจำใบหน้ามนุษย์ที่มีอยู่ในระบบ แล้วนำใบหน้าที่ตรวจหาได้จากขั้นตอนแรกมาทำการวิเคราะห์ ว่าใบหน้าที่ตรวจหาได้นั้นระบบรู้จำหรือไม่ และถ้ารู้จำ บุคคลที่พบนั้นเป็นบุคคลใดในฐานข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Human Face Recognition from Video
Student	Mr. Panomporn Sakorn
Advisor	Dr. Thanarat Chalidabhongse
Level of Study	Master of Science in Information Technology
Major	Information Science
Academic Year	2003

ABSTRACT

The human face recognition system from video can be divided into two main steps. First, the position of human face in each video frame is initially detected located and Then, the human face recognition is subsequently performed by means of the supervised system to detecting and analyzing the human face whether or not the known human face. If known, the equivalent human face will be displayed.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

โครงการพัฒนาระบบงานนี้ประสบความสำเร็จลุล่วงไปด้วยความร่วมมือและความช่วยเหลือจากบุคคลหลายท่าน โดยเฉพาะอย่างยิ่งนั้น ผู้เขียนขอขอบคุณ ดร. ธนารัตน์ ชลิดาพงศ์ ซึ่งเป็นผู้ให้ความช่วยเหลือคำแนะนำที่เป็นประโยชน์ในทุกๆ ด้านในโครงการพัฒนาระบบงานนี้

ผู้เขียนขอขอบคุณ คุณพ่อคุณแม่ที่เป็นกำลังใจให้เสมอมา ผู้เขียนขอขอบคุณ คุณอลงกต น้อยใจบุญ และคุณพูลศักดิ์ จีรบุณย์ ที่คอยให้คำปรึกษาและข้อแนะนำเกี่ยวกับการพัฒนาโปรแกรม รวมทั้งเพื่อนที่ทำงานและเพื่อนที่สถาบันฯ ทุกคนที่คอยให้กำลังใจและให้ความร่วมมือในการพัฒนาระบบงานนี้ จนผู้เขียนสามารถพัฒนาระบบงานได้สำเร็จลุล่วงไปด้วยดี

ผู้เขียนหวังเป็นอย่างยิ่งว่าโครงการพัฒนาระบบงานนี้จะเป็นประโยชน์ต่อนิสิต นักศึกษา รวมทั้งผู้ค้นคว้าที่สนใจในเรื่องนี้ โดยถ้าหากว่าโครงการพัฒนาระบบงานนี้มีข้อผิดพลาดประการใดก็ตาม ผู้เขียนขอน้อมรับไว้เพื่อนำไปปรับปรุงให้สมบูรณ์ยิ่งขึ้น แต่ถ้าหากความดีที่ได้รับจากโครงการพัฒนาระบบงานนี้ ผู้เขียนขอมอบให้แก่บุพการี ครูบาอาจารย์ และผู้มีพระคุณทุกท่าน

พนมพร สาคร

กุมภาพันธ์ 2547

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VI

สารบัญภาพ VII

บทที่

1. บทนำ 1

1.1 แนวคิดในการพัฒนาระบบ 1

1.2 วัตถุประสงค์ของการพัฒนาระบบ 2

1.3 ขอบเขตของการพัฒนาระบบ 2

1.4 ทรัพยากรที่ใช้ในการพัฒนาระบบ 3

1.5 ประโยชน์ที่คาดว่าจะได้รับ 3

2. ทฤษฎีที่เกี่ยวข้องและขั้นตอนการประมวลผลของระบบ 4

2.1 ทฤษฎีและหลักการในการพัฒนาระบบตรวจหาใบหน้ามนุษย์จากภาพวีดิโอ 4

2.1.1. เทคนิคการลบฉากหลัง (Background Subtraction) 5

2.1.2 เทคนิคการตรวจหาใบหน้ามนุษย์จากรูปภาพ (Human Face Detection) 7

2.2. ทฤษฎีและหลักการในการพัฒนาระบบรู้จำใบหน้าของมนุษย์จากรูปภาพ 12

2.2.1. การรู้จำใบหน้าโดยใช้วิธี Eigenface 12

2.3. ศึกษาการใช้เครื่องมือที่ใช้ในการพัฒนาระบบ 15

2.3.1 หลักการเขียนโปรแกรมแบบเชิงวัตถุ (Object Oriented Programming) 15

2.3.2 Microsoft Vision System Development Kit 16

2.4 ขั้นตอนการประมวลผลของระบบ 29

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

3. การทดสอบระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอ 31
3.1 อุปกรณ์ที่ใช้ในการทดสอบระบบ 31
3.2 ฐานข้อมูลภาพที่ใช้ในการทดสอบ 31
3.3 การทดสอบระบบ 32
3.3.1 การทดสอบการลบภาพฉากหลัง 32
3.3.2 การทดสอบการตรวจหาใบหน้ามนุษย์ 34
3.3.3 การทดสอบการรู้จำใบหน้ามนุษย์ 35
3.4 สรุปผลการทดสอบ 36
4. บทสรุปและข้อเสนอแนะ 39
4.1 บทสรุปโครงการ 39
4.2 ปัญหาและข้อจำกัดในการทำงานของระบบ 39
4.3 ข้อเสนอแนะในการพัฒนาระบบรู้จำใบหน้ามนุษย์ต่อไป 40
บรรณานุกรม 41
ประวัติผู้เขียน 42

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้า

ตารางที่

2.1 แสดงรายละเอียดของ pixel แต่ละชนิดในกลุ่ม Gray Scale Pixel Type.....21
2.2 แสดงรายละเอียดของ pixel แต่ละชนิดในกลุ่ม RGBA Color Pixel Types.....22
2.3 แสดงรายละเอียดของ pixel แต่ละชนิดในกลุ่ม YUVA Color Pixel Types.....23

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ

หน้า

ภาพที่

2.1 ภาพแสดงตัวอย่างการนำเทคนิคการลบฉากหลัง....6

2.2 การสร้าง multi-resolution hierarchy ของภาพ โดยการ sub sampling....9

2.3 ใบหน้าตัวอย่างที่ใช้ใน knowledge-based top-down methods....9

2.4 ตัวอย่างการตรวจหาใบหน้าโดยใช้วิธี rule-based ของ Kotropoulos และ Pitas....10

2.5 องค์ประกอบทางใบหน้าทั้ง 6 องค์ประกอบตามแบบของ Yow and Cipolla model....12

2.6 แสดงตัวอย่างของ eigenface....14

2.7 แสดงโครงสร้างหลักของ MS Vision SDK....18

2.8 โครงสร้างของการเก็บภาพใน Vision SDK....19

2.9 แสดงขั้นตอนการประมวลผลของระบบ....29

3.1 แสดงภาพที่ใช้ในฐานข้อมูลภาพของระบบรู้จำใบหน้ามนุษย์....31

3.2 แสดงตัวอย่างผลการทดสอบการลบภาพฉากหลัง....33

3.3 แสดงตัวอย่างผลการทดสอบการลบภาพฉากหลังที่มีวัตถุเป้าหมายอยู่ในภาพ....34

3.4 แสดงตัวอย่างผลการทดสอบการตรวจหาใบหน้ามนุษย์จากภาพวีดิโอ....35

3.5 แสดงตัวอย่างผลการทดสอบการรู้จำใบหน้ามนุษย์จากภาพวีดิโอ....36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 แนวคิดในการพัฒนาระบบ

จากการศึกษาหลักการของ Computer Vision ในเรื่องของระบบตรวจหาใบหน้าของมนุษย์ (Human Face Detection) และระบบการรู้จำใบหน้าของมนุษย์ (Human Face Recognition) จากรูปภาพ (Image)ทำให้เกิดแนวความคิดในการนำเอาระบบทั้งสองมาประยุกต์และพัฒนาให้เป็นระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอ (Human Face Recognition) เพื่อนำไปใช้ประโยชน์ในด้านต่างๆ ได้ โดยเฉพาะอย่างยิ่งในด้านของระบบรักษาความปลอดภัย เช่น การนำมาประยุกต์ใช้กับกล้องวีดีวงจรปิดที่ติดตั้งกับอาคารหรือหน่วยงานต่างๆ ที่ต้องการระบบรักษาความปลอดภัยที่มีประสิทธิภาพ ซึ่งตัวระบบการตรวจจับใบหน้ามนุษย์จะช่วยให้สามารถตรวจจับบุคคลที่เข้าออกได้ และเมื่อสามารถตรวจจับใบหน้าจากภาพวีดิโอได้แล้ว ก็ยังสามารถนำมาประยุกต์ใช้ในเรื่องของการวิเคราะห์การรู้จำใบหน้า เพื่อหาว่าบุคคลที่สามารถตรวจจับใบหน้าได้นั้น ว่าเป็นคนในอาคารหรือหน่วยงานหรือไม่ หรือเป็นบุคคลภายนอกที่ไม่ได้รับอนุญาตให้เข้ามาภายในอาคารหรือหน่วยงาน และหากเป็นบุคคลภายนอกก็สามารถประยุกต์ระบบให้สามารถบันทึกข้อมูลการเข้า-ออก ไม่ว่าจะเป็นใบหน้าที่ตรวจจับได้ เวลาที่เข้ามาในอาคาร และเวลาที่ออกจากอาคาร เพื่อให้ผู้ที่ดูแลระบบรักษาความปลอดภัยของตัวอาคารทราบ และมีประโยชน์ในกรณีที่ต้องการตรวจหาบุคคลที่ต้องสงสัยหากมีเหตุการณ์ผิดปกติเกิดขึ้นภายในอาคาร เช่น ของหาย เป็นต้น

แต่เนื่องจากการพัฒนาระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอนั้น การทำงานของระบบจะต้องทำงานในแบบที่สามารถตอบสนองได้ในทันทีทันใด (real-time response) เนื่องจากการเคลื่อนของภาพในแต่เฟรมในภาพวีดิโอนั้นเป็นไปอย่างต่อเนื่อง ดังนั้นในเรื่องของการประมวลผลเพื่อตรวจหาตำแหน่งใบหน้าของมนุษย์ในภาพวีดิโอจำเป็นจะต้องทำงานได้อย่างรวดเร็วและถูกต้อง แต่เนื่องภาพของมนุษย์ในภาพวีดิโอ มีการเคลื่อนไหวได้ตลอดเวลา ดังนั้นการตรวจหาตำแหน่งใบหน้าของมนุษย์ในภาพอย่างเดียวจึงไม่เพียงพอที่จะนำไปใช้ในระบบรู้จำใบหน้ามนุษย์ เพราะใบหน้าที่ตรวจจับได้ อาจเป็นใบหน้าที่ไม่เหมาะสมต่อการวิเคราะห์ในระบบการรู้จำใบหน้า จึงมีความจำเป็นที่จะต้องนำใบหน้ามนุษย์ที่ตรวจหาได้ ไปวิเคราะห์เพื่อหาใบหน้าในเฟรมของภาพที่เหมาะสม เพื่อนำไปใช้ในระบบรู้จำใบหน้า ทั้งนี้เพื่อให้เกิดความถูกต้องในการทำงานของระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอมากยิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 วัตถุประสงค์ของการพัฒนาระบบ

1.2.1 ศึกษาหลักการและทฤษฎีทางด้าน Computer Vision ในเรื่องเกี่ยวกับการตรวจหาใบหน้ามนุษย์จากภาพ (Human Face Detection) การรู้จำใบหน้ามนุษย์ (Human Face Recognition) เป็นต้น

1.2.2 ศึกษาการใช้งาน Microsoft Vision System Development Kit (MS Vision SDK) เพื่อใช้ในการจัดการภาพ (Image) ในโปรแกรมประยุกต์ที่พัฒนาและศึกษาการเขียนโปรแกรมเชิงวัตถุ (Object Oriented Programming) เพื่อใช้ในการพัฒนาระบบ

1.2.3 เพื่อสร้างโปรแกรมประยุกต์ที่มีความสามารถในการรับ (Capture) ภาพต่อเนื่องเป็นเฟรม (ภาพวีดิโอ) ในแบบทันทีทันใด (real-time) บนเครื่องคอมพิวเตอร์ที่ใช้ระบบปฏิบัติการวินโดวส์ และนำภาพที่ได้ในแต่ละเฟรม ไปประมวลผลทางด้าน image processing เพื่อหาตำแหน่งใบหน้าของมนุษย์ในภาพแต่ละเฟรม และนำภาพใบหน้ามนุษย์ที่ตรวจหาได้ในแต่ละเฟรม ไปใช้ในการประมวลผลทางด้าน image processing เพื่อการค้นหาและเปรียบเทียบใบหน้ามนุษย์ที่ตรวจหาได้ กับฐานข้อมูลใบหน้ามนุษย์ที่เก็บไว้ ว่าใบหน้าที่ตรวจจับได้เหมือนกับใบหน้ามนุษย์ในภาพใดในฐานข้อมูล

1.3 ขอบเขตของการพัฒนาระบบ

ในการพัฒนาระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอนี้ ผู้พัฒนาระบบจะพัฒนาระบบโดยใช้ข้อมูลภาพวีดิโอจากกล้องวีดิโอวงจรปิด ซึ่งติดตั้ง ณ ตำแหน่งตายตัวไม่มีการเคลื่อนไหว ดังนั้นข้อมูลภาพวีดิโอ จึงไม่มีการเคลื่อนไหวหรือเปลี่ยนแปลงฉากหลังของภาพ และเนื่องจากระบบนี้มีการทำงานและการประมวลผลที่ซับซ้อนเป็นอย่างมาก อีกทั้งยังเป็นระบบที่มีความจำเป็นที่จะต้องตอบสนองการทำงานในแบบทันทีทันใด แต่เนื่องจากตัวระบบเองมีการประมวลผลค่อนข้างสูง ดังนั้นระบบที่พัฒนาขึ้นในวิชาโครงการพัฒนาระบบงานนี้ จึงจำเป็นต้องมีขอบเขตของการศึกษาและพัฒนาระบบ ดังต่อไปนี้

1.3.1 การวางตำแหน่งของกล้องตายตัว ไม่มีการเคลื่อนไหวกล้อง ดังนั้นข้อมูลภาพวีดิโอจึงไม่มีการเคลื่อนไหวหรือเปลี่ยนแปลงฉากหลังของภาพ

1.3.2 พัฒนาบนระบบปฏิบัติการ Microsoft Windows 2000

1.3.3 ขนาดของฐานข้อมูลภาพที่ใช้ในการทดลองมีจำนวน 10 ใบหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3.4 ตั้งสมมุติฐานว่าจะอยู่ในตำแหน่งตั้งตรง (Upright Position) ดังนั้น ตำแหน่งการวางศีรษะจะอยู่ด้านบนสุดเสมอ และจำนวนใบหน้าในภาพวีดิโอที่ใช้ในการประมวลผลมีเพียงใบหน้าเดียวต่อภาพวีดิโอ 1 เฟรม

1.4 ทรัพยากรที่ใช้ในการพัฒนาระบบ

1.4.1 เครื่องคอมพิวเตอร์ที่มีความเร็วของหน่วยประมวลผลกลาง 1.7 GHz และหน่วยความจำ 512 MB

1.4.2 ระบบปฏิบัติการวินโดวส์ พร้อมติดตั้ง Microsoft Visual C++ และ MS Vision SDK

1.4.3 กล้องสำหรับนำภาพเข้าสู่เครื่องคอมพิวเตอร์ (ในการพัฒนาระบบนี้ใช้กล้อง Web Cam แบบเชื่อมต่อผ่าน USB Port)

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้ความรู้เพิ่มเติมจากการศึกษาทฤษฎีทางด้าน Computer Vision ที่สามารถนำไปใช้ในการสร้างโปรแกรมประยุกต์ที่สามารถใช้งานได้จริง

1.5.2 ได้ความรู้เพิ่มเติมจากการศึกษาและใช้งาน MS Vision SDK เพื่อนำไปใช้ในการสร้างโปรแกรมประยุกต์ทางด้าน image processing ได้

1.5.3 ได้ฝึกฝนทักษะในการใช้ภาษาในแบบเชิงวัตถุในการสร้างโปรแกรมประยุกต์ ที่สามารถนำไปใช้งานได้

1.5.4 ได้ระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอที่สามารถนำไปใช้งานได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีที่เกี่ยวข้องและขั้นตอนการประมวลผลของระบบ

ในบทนี้จะกล่าวถึงทฤษฎีและหลักการที่เกี่ยวข้องกับการพัฒนาระบบรู้จำใบหน้ามนุษย์จากภาพวีดีโอวงจรปิดเพื่อใช้ในการพัฒนาระบบรักษาความปลอดภัย ในการวิเคราะห์เพื่อตรวจหาใบหน้ามนุษย์จากภาพวีดีโอวงจรปิด โดยผู้เขียนได้แยกกรณีการศึกษาออกเป็น 3 ส่วนใหญ่ๆ คือ

1. ศึกษาทฤษฎีและหลักการในการพัฒนาระบบตรวจหาใบหน้ามนุษย์จากภาพวีดิโอ
2. ศึกษาทฤษฎีและหลักการในการพัฒนาระบบรู้จำใบหน้าของมนุษย์จากรูปภาพ
3. ศึกษาการใช้เครื่องมือที่ใช้ในการพัฒนาระบบ

โดยรายละเอียดของการศึกษาทฤษฎีและหลักการที่เกี่ยวข้องต่างๆ ที่ได้ทำการศึกษาเพื่อพัฒนาโครงงานนี้ จะกล่าวในรายละเอียดต่อไป

2.1. ทฤษฎีและหลักการในการพัฒนาระบบตรวจหาใบหน้ามนุษย์จากภาพวีดิโอ

เนื่องจากข้อมูลภาพที่นำเข้ามาใช้ในการประมวลผลเพื่อตรวจหาตำแหน่งใบหน้าของมนุษย์นั้น เป็นภาพในลักษณะภาพจากกล้องวีดิโอวงจรปิด กล่าวคือฉากหลังของภาพไม่มีการเคลื่อนไหวหรือเปลี่ยนแปลง ดังนั้นเพื่อให้ขอบเขตของการตรวจหาใบหน้ามนุษย์ในภาพวีดิโอแต่ละเฟรมมีขนาดพื้นที่เล็กลง ทั้งนี้เพื่อให้การประมวลผลในการตรวจหาตำแหน่งใบหน้าของมนุษย์ในภาพแต่ละเฟรมทำได้อย่างรวดเร็วและสามารถตอบสนองต่อการทำงานของโปรแกรมในแบบทันทีทันใด (real time) ได้นั้น จะต้องอาศัยเทคนิคและวิธีการของการลบฉากหลัง (Background Subtraction) เข้ามาช่วย นอกจากเทคนิคของการลบฉากหลัง จะทำให้พื้นที่ในการตรวจหาตำแหน่งใบหน้ามนุษย์เล็กลงด้วยการแยกแยะเป้าหมายออกจากภาพฉากหลังแล้ว ยังมีส่วนทำให้การตรวจหาใบหน้ามนุษย์ของระบบมีความแม่นยำมากยิ่งขึ้น คือระบบไม่จำเป็นที่จะต้องทำการตรวจหาใบหน้ามนุษย์ในเฟรมของภาพวีดิโอที่มีแต่ฉากหลัง ไม่มีเป้าหมายอยู่ได้อีกด้วย

โดยสรุปทฤษฎีและหลักการที่ใช้ในการพัฒนาระบบตรวจหาใบหน้ามนุษย์จากภาพวีดิโอจะประกอบไปด้วย 2 ส่วนย่อย คือ

1. **เทคนิคการลบฉากหลัง** (Background Subtraction) **เพื่อแยกแยะเป้าหมายออกจากภาพฉากหลัง**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. เทคนิคการตรวจหาใบหน้ามนุษย์จากรูปภาพ (Human Face Detection) ในการตรวจหาใบหน้ามนุษย์จากภาพ ยังมีเทคนิคที่ใช้ในการตรวจหาด้วยกันหลายวิธีการ แต่ที่ทำการศึกษาเพื่อใช้ในการพัฒนาระบบงานนี้มีด้วยกัน 2 วิธี คือ

   - Knowledge-based methods

   - Feature invariant approaches

2.1.1. เทคนิคการลบฉากหลัง (Background Subtraction) (Horprasert. 1999; Seki. et.al. 2000.)

เทคนิคการลบฉากหลังเป็นเทคนิคที่ง่ายต่อการนำมาประยุกต์ใช้กับการลบฉากหลังของภาพที่ไม่มีการเคลื่อนไหวของภาพฉากหลัง หรือการเคลื่อนไหวเพียงเล็กน้อย อย่างไรก็ตาม ในความเป็นจริงนั้นการใช้เทคนิคนี้อาจจะไม่ได้ผลมากนักอันเนื่องมาจากสาเหตุดังนี้

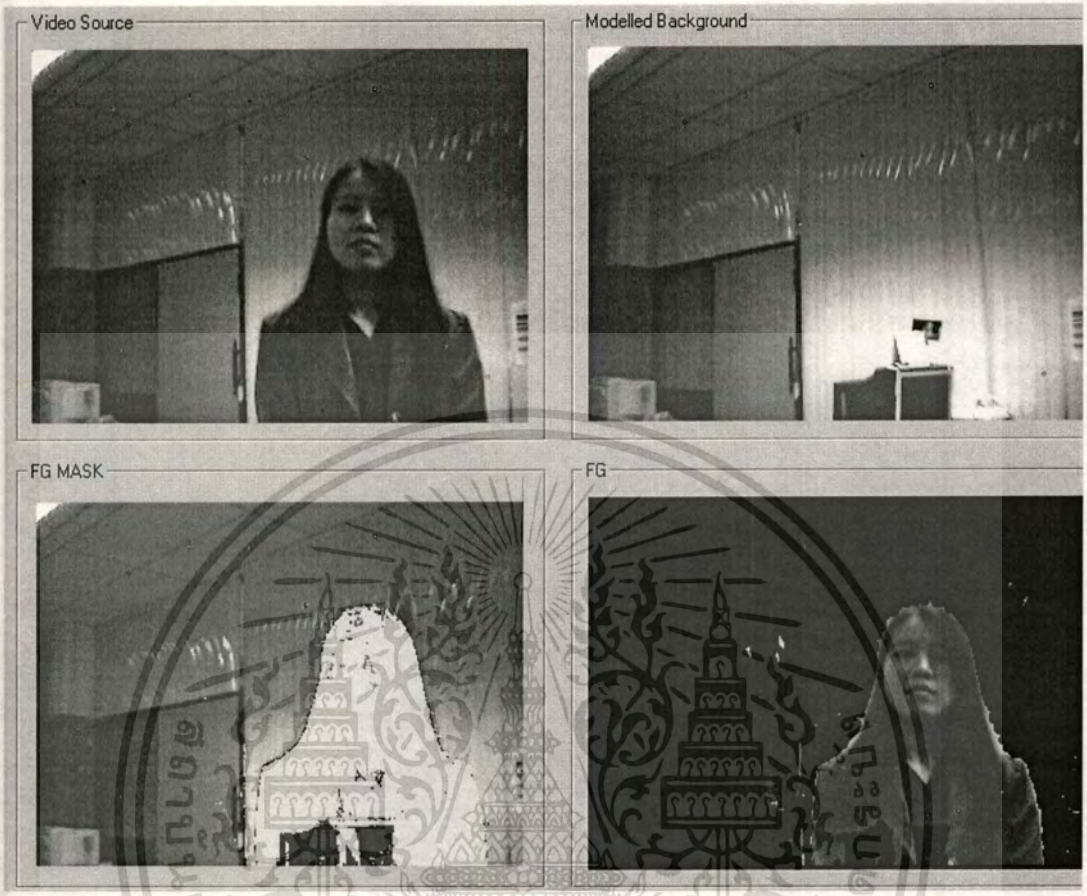
- การปรับเปลี่ยนของภาพฉากหลังอันเนื่องมาจากการเปลี่ยนแปลงของแสงสว่าง ไม่ว่าจะเป็นแสงสว่างจากดวงอาทิตย์ หรือแสงสว่างจากแหล่งกำเนิดอื่นๆ

- การเปลี่ยนแปลงโครงสร้างของภาพฉากหลัง เช่น การสั่นไหวของต้นไม้ การสะบัดไปมาของธงต่างๆ รวมถึงการเปลี่ยนแปลงที่เกิดจากตัวกล้องที่ใช้ในการเก็บภาพเอง เช่น การสั่นไหวของตัวกล้อง สัญญาณรบกวน (noise) ที่เกิดจากการจับภาพของตัวกล้องเอง

แต่ในปัจจุบันเทคนิคการลบฉากหลังได้มีการปรับปรุงและพัฒนาวิธีการในการประมวลผลให้มีประสิทธิภาพสูงมากยิ่งขึ้น เช่น มีการนำค่าความเบี่ยงเบนของแสงสว่างเข้ามาช่วย ทำให้ประสิทธิภาพในการลบฉากหลังเพื่อแยกแยะเป้าหมายออกจากวัตถุเป็นไปได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

เทคนิคการลบฉากหลังจะเริ่มจากการคำนวณเพื่อสร้างแบบจำลองของฉากหลัง (Background Model) เพื่อใช้เป็นต้นแบบในการหาความเคลื่อนไหวในภาพอื่นๆ โดยภาพที่มีการเคลื่อนไหวเราสามารถตรวจจับได้จากการที่มีจุดความแตกต่างระหว่างภาพที่เป็น input กับภาพที่เป็นแบบจำลอง อยู่เป็นจำนวนมากและต้องรวมกลุ่มกัน ซึ่งตัวเทคนิคนี้จะสามารถทำได้แค่การวิเคราะห์เพื่อหาตำแหน่งที่มีการเคลื่อนไหวของวัตถุในภาพเท่านั้น ไม่สามารถวิเคราะห์เพื่อตรวจสอบได้ว่าวัตถุที่เคลื่อนไหวอยู่ในภาพนั้นเป็นวัตถุอะไร แต่มีข้อดีคือทำให้เราสามารถรับรู้ถึงตำแหน่งของภาพที่มีการเปลี่ยนแปลงเพื่อนำข้อมูลในตำแหน่งเหล่านั้นไปวิเคราะห์ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 2.1 แสดงตัวอย่างการนำเทคนิคการลบฉากหลังมาประยุกต์ใช้เพื่อแยกเป้าหมายออกจากภาพฉากหลัง โดยภาพแรก (ซ้ายบน) จะเป็นภาพที่นำมาวิเคราะห์ ภาพที่สอง (ขวาบน) เป็นภาพแบบจำลองของภาพฉากหลัง และภาพสองภาพด้านล่างจะเป็นภาพผลลัพธ์ที่ได้จากการใช้เทคนิคการลบฉากหลัง

การใช้เทคนิคการลบฉากหลังเพื่อหาตำแหน่งที่มีการเคลื่อนไหวของวัตถุในภาพ ทำให้เราสามารถนำภาพที่ต้องการวิเคราะห์หาใบหน้ามนุษย์มาวิเคราะห์โดยสามารถเลือกการวิเคราะห์เฉพาะตำแหน่งที่มีการเคลื่อนไหวของภาพเท่านั้น ทำให้ไม่มีความจำเป็นจะต้องประมวลผลเพื่อตรวจหาใบหน้ามนุษย์จากภาพทั้งภาพ และไม่จำเป็นจะต้องตรวจหาใบหน้ามนุษย์จากภาพทุกภาพจากภาพวีดีโอวงจรปิด เนื่องจากภาพที่ไม่มีการเคลื่อนไหวของวัตถุย่อมไม่มีใบหน้ามนุษย์ในภาพ ดังนั้นเทคนิคนี้จึงมีความสำคัญต่อขั้นตอนต่อไปเป็นอย่างมากคือขั้นตอนของการตรวจหาใบหน้ามนุษย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2 เทคนิคการตรวจหาใบหน้ามนุษย์จากรูปภาพ (Human Face Detection)

การตรวจหาใบหน้ามนุษย์จากภาพที่ต้องการวิเคราะห์นั้นมีความยุ่งยากในการวิเคราะห์พอสมควรเนื่องภาพที่นำมาวิเคราะห์นั้นมีความหลากหลายของลักษณะท่าทาง มุมมองของใบหน้าอยู่เป็นจำนวนมาก ซึ่งปัจจัยหลักๆ ที่มีผลต่อการตรวจหาใบหน้าของมนุษย์จากภาพ มีดังนี้คือ

-การวางท่าทาง (Pose) เนื่องจากมนุษย์มีการเคลื่อนไหวตลอดเวลา ไม่ว่าจะหันซ้าย หันขวา หรือการเอียงใบหน้า ดังนั้นจึงไม่เสมอไปที่เราจะได้ใบหน้าตรงของมนุษย์มาใช้ในการวิเคราะห์ หรือการวางมุมมองของกล้องที่ไม่เหมาะสม คืออาจจะวางทำมุมมากเกินไปกับหน้าของมนุษย์ทำให้ภาพที่ได้ไม่เหมาะสมต่อการนำไปวิเคราะห์เพื่อหาตำแหน่งของใบหน้า

-การมีหรือหายไปของส่วนประกอบบนใบหน้า (Presence or absence of structural components) ตัวอย่างเช่น การเพิ่มขึ้น หรือหายไปของคิ้ว หนวด หรือเคราบนใบหน้า นอกจากนี้ยังมีสิ่งอื่นที่นอกเหนือจากธรรมชาติทางใบหน้าของมนุษย์ เช่น แว่นตา หรือผ้าปิดปากเป็นต้น ซึ่งสิ่งเหล่านี้มีได้ทั้งขนาด รูปร่าง และสีที่แตกต่างกันได้มากมาย

-การแสดงออกบนใบหน้า (Facial expression) ตัวอย่างเช่น การยิ้ม หัวเราะ หน้าบึ้ง หรือใบหน้าแสดงอาการตกใจเป็นต้น ซึ่งสิ่งเหล่านี้ล้วนแล้วแต่ทำให้ใบหน้าเปลี่ยนแปลงไปจากสภาวะปกติเป็นอย่างมาก

-การถูกบังจากวัตถุ (Occlusion) ตัวอย่างเช่น การถูกบังด้วยวัตถุบางอย่าง เช่น หนังสือ หรือแม้แต่การถูกบังด้วยมนุษย์ด้วยกันเอง เช่น การยืนซ้อนกัน อาจจะทำให้ส่วนใดส่วนหนึ่งของใบหน้าหายไป

- การกำหนดทิศทางของภาพ (Image orientation) คือการเปลี่ยนไปของใบหน้ามนุษย์ที่เกิดจากการปรับเปลี่ยนทิศทางหรือมุมของกล้อง

-สภาพของภาพ (Imaging Conditions) เนื่องจากภาพที่นำมาใช้ในการวิเคราะห์นั้น อาจมีสิ่งแวดล้อมที่แปรเปลี่ยนได้ เช่น แสงสว่าง หรือแม้แต่คุณสมบัติของกล้องเองก็มีผลต่อภาพที่ใช้ในการวิเคราะห์เช่นเดียวกัน เช่น ความละเอียดในการถ่ายภาพของกล้อง เป็นต้น

เทคนิคที่ใช้ในการตรวจหาใบหน้าของมนุษย์ที่ได้ทำการศึกษานั้นมีด้วยกันหลายวิธีการ แต่ที่ทำการศึกษาเพื่อใช้ในการพัฒนาระบบงานนี้ มีด้วยกัน 2 วิธีการ ดังนี้คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2.1 Knowledge-based methods (Yang. et.al. 2002)

วิธี Knowledge-based เป็นเทคนิคที่ใช้หลักการของ rule-based method ซึ่งก็คือการสร้างกฎที่ใช้ในการอธิบายโครงสร้างพื้นฐานลักษณะเฉพาะของใบหน้า และความสัมพันธ์กันขององค์ประกอบเหล่านั้น เช่น โดยปกติใบหน้าของมนุษย์จะมีตาอยู่ด้วยกัน 2 ตา โดยดวงตาทั้ง 2 ข้างจะสมมาตรกัน มีจมูก มีปาก โดยความสัมพันธ์ของสิ่งต่างๆ เหล่านี้จะแสดงได้จากระยะห่างที่มีความสัมพันธ์กันรวมถึงตำแหน่งของสิ่งเหล่านี้ด้วย โดยกระบวนการตรวจหาใบหน้าจะเริ่มจากการแยกเอาส่วนประกอบต่างๆ ของใบหน้าออกจากภาพ input image จากนั้นจะทำการตรวจสอบหาตำแหน่งที่คาดว่าจะใบหน้ามนุษย์ออกมา (face candidates) โดยอาศัยกฎของใบหน้าที่รวบรวมไว้ จากนั้นจึงมีการนำเอา face candidates เหล่านั้นมาพิสูจน์ต่อไป ทั้งนี้เพื่อลดความผิดพลาดของการตรวจหาใบหน้ามนุษย์นั่นเอง

ปัญหาหนึ่งซึ่งเป็นความยากของวิธีการนี้คือ การแปลงความรู้ของมนุษย์ไปเป็นกฎที่สามารถอธิบายถึงโครงสร้างพื้นฐานลักษณะเฉพาะทางใบหน้าและความสัมพันธ์กันขององค์ประกอบเหล่านั้นที่มีประสิทธิภาพ (well-defined) เนื่องจากถ้ามีกำหนดกฎที่มีความละเอียดมากเกินไป ก็จะเป็นการยากที่ใบหน้าที่ตรวจจับได้จะผ่านกฎได้ทั้งหมด และหากมีการกำหนดกฎที่มีความละเอียดน้อยเกินไป ก็จะทำให้ใบหน้าที่ตรวจจับได้มีความผิดพลาดมาก เช่น บางใบหน้าที่ตรวจจับได้ อาจไม่ใช่ใบหน้ามนุษย์จริงๆ นอกจากนี้การที่จะทำให้ระบบสามารถตรวจจับใบหน้ามนุษย์ได้ในทุกมุมมองนั้น ก็เป็นการยากเช่นเดียวกัน เนื่องจากการสร้างกฎที่สามารถระบุหรือบ่งบอกได้ทุกกรณีนั้นเป็นสิ่งที่ทำได้ยากมาก อีกทั้ง heuristics เกี่ยวกับใบหน้าสามารถทำการตรวจจับได้ดีในกรณีที่เป็นการตรวจจับใบหน้าตรงบนพื้นหลังที่ไม่ซับซ้อน

Yang และ Huang ใช้วิธีการ hierarchical knowledge-based ในการตรวจหาใบหน้ามนุษย์ โดยแบ่งแยกกฎที่ใช้ ออกเป็น 3 ระดับด้วยกัน คือ

-ระดับที่ 1 ทำการ scan windows บนภาพ input image และประยุกต์ใช้กลุ่มของกฎต่างๆ ในแต่ละตำแหน่งของภาพ เพื่อบอกถึงจำนวนใบหน้าทั้งหมดที่เป็นไปได้ (face candidates) กฎในระดับที่สูงกว่าจะอธิบายเรื่องทั่วๆ ไปเกี่ยวกับใบหน้า และกฎที่อยู่ในระดับต่ำกว่าจะพูดถึงรายละเอียดของลักษณะเฉพาะของใบหน้า การสร้าง multi-resolution hierarchy ของภาพ ทำโดยการ sub sampling และนำมาหาค่าเฉลี่ย ดังตัวอย่างที่แสดงในภาพที่ 2.2 กฎที่ใช้ในการอธิบายโครงสร้างพื้นฐานทางใบหน้าถูกนำมาใช้เพื่อหาตำแหน่งของ face candidates ในภาพที่มี

resolution ต่ำสุด โดยบริเวณศูนย์กลางของใบหน้า (ส่วนที่มีเงาสีเข้มในภาพที่ 2.3) จะมีเซลล์ 4 เซลล์ที่มีความเข้มเท่ากัน และส่วนบริเวณรอบๆ ทางด้านบนของใบหน้า (ส่วนที่มีสีสว่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในภาพที่ 2.3) จะมีความเข้มเท่ากัน และความแตกต่างระหว่างค่าเฉลี่ยของความเข้มในบริเวณศูนย์กลางของใบหน้าจะแตกต่างจากบริเวณส่วนรอบๆ ทางด้านบนของใบหน้ามาก จากนั้นจะนำผลลัพธ์ของ face candidates ที่ได้จากภาพที่มี resolution ต่ำสุด ไปประมวลผลต่อในภาพที่มี resolution ที่สูงขึ้น

ภาพที่ 2.2 (a) n = 1 ภาพต้นฉบับ (b) n = 4 (c) n = 8 (d) n = 16 ในภาพต้นฉบับและภาพที่มี resolution ต่ำ เซลล์สี่เหลี่ยมแต่ละเซลล์จะประกอบด้วยจำนวนของ pixels ในภาพต้นฉบับ จำนวน n x n pixels และความเข้มแต่ละเซลล์จะได้มาจากการหาค่าเฉลี่ยของความเข้มของทุก pixel ในเซลล์นั้น

ภาพที่ 2.3 ใบหน้าตัวอย่างที่ใช้ใน knowledge-based top-down methods (กฎที่สร้างขึ้นโดยพื้นฐานความรู้ของมนุษย์เกี่ยวกับคุณลักษณะเฉพาะทางใบหน้าของมนุษย์)

-ระดับที่2 จะนำเอา face candidates ที่ได้จากระดับที่ 1 มาหา local histogram equalization โดยใช้ edge detection

-ระดับที่ 3 ในพื้นที่ที่ผ่านมาถึงระดับที่ 3 นี้จะถูกนำมาตรวจสอบด้วยกฎที่บอกถึงลักษณะเฉพาะทางใบหน้า เช่น ตาและปาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Kotropoulos และ Pitas ได้เสนอวิธีในการหาตำแหน่งโดยใช้ rule-based โดยขั้นแรกจะหาตำแหน่งของลักษณะทางใบหน้า โดยวิธี projection ซึ่ง Kanade ใช้วิธีการนี้ในการหาตำแหน่งของขอบใบหน้า โดยกำหนดให้ I(x,y) แทนค่าความเข้มของภาพขนาด m x n ที่ตำแหน่ง (x,y) projection ในแนวนอน และแนวตั้งของภาพจะถูกกำหนดโดย HI(x) = $\sum_{y=1}^{n} I(x,y)$ และ VI(y) = $\sum_{x=1}^{m} I(x,y)$ โดยจะทำการ project input image ในภาพแนวนอนก่อน และจะทำการหาตำแหน่ง 2 ตำแหน่งที่ต่ำที่สุดในบริเวณรอบๆ ซึ่ง 2 ตำแหน่งนั้นก็คือขอบของส่วนหัวทางด้านซ้ายและด้านขวา ต่อจากนั้นทำการ projection input image ในแนวตั้ง และหาตำแหน่งที่ต่ำที่สุดในบริเวณรอบๆ โดยตำแหน่งต่างๆ เหล่านั้น คือตำแหน่งของริมฝีปาก ปลายจมูก และตา คุณลักษณะเฉพาะทางใบหน้าที่ตรวจจับได้นั้น จะเป็น facial candidate แต่วิธีการนี้จะไม่เหมาะสมสำหรับการตรวจจับใบหน้าในภาพที่มีพื้นหลังซับซ้อนได้ เนื่องจากวิธีการนี้สามารถตรวจจับใบหน้าในภาพที่มีพื้นหลังซับซ้อนได้ยาก และไม่สามารถตรวจจับภาพที่มีใบหน้าหลายๆ ใบหน้าอยู่ในภาพได้ ดังตัวอย่างภาพที่ 2.4

ภาพที่ 2.4 (a) และ (b) n = 8 (c) n = 4 รูปเส้นโครงสร้างในแนวแกนนอนและแนวแกนตั้ง ซึ่งสามารถตรวจสอบหาใบหน้าได้เพียงใบหน้าเดียว โดยการหาจุดที่สูงสุดจากบริเวณใกล้เคียงในรูปเส้นโครงสร้างตามแนวแกนนอนและแนวแกนตั้ง แต่อย่างไรก็ตามวิธีการเดียวกันนี้มีความยากที่จะตรวจหาใบหน้าที่มีพื้นหลังที่ซับซ้อนและมีหลายๆ ใบหน้าอยู่ในภาพเดียวกันได้ ดังแสดงในภาพ (b) และ (c)

2.1.2.2 Feature-Based Methods (Yang. et.al. 2002)

จากวิธีการแบบ knowledge-based ที่ได้กล่าวมาแล้วข้างต้น จะเห็นว่ามีการทำงานในลักษณะแบบ top-down คือ มีการหาใบหน้าของมนุษย์โดยการ scan window เพื่อหาบริเวณที่มี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลักษณะเหมือนแบบจำลองทางใบหน้าที่กำหนดไว้ ในภาพที่ความละเอียดต่ำก่อน จากนั้นจึงตรวจสอบหาใบหน้าที่มีลักษณะเฉพาะทางใบหน้าในภาพที่มีความละเอียดสูงขึ้นเรื่อยๆ

แต่ในขณะที่วิธีการแบบ feature-based จะมีการทำงานโดยพยายามค้นหากลุ่มของลักษณะเฉพาะทางใบหน้าในรูปภาพก่อน แล้วจึงนำมาวิเคราะห์เพื่อตรวจสอบใบหน้าต่อไป ภายใต้สมมติฐานที่เกิดขึ้นมาจากการสังเกตว่ามนุษย์สามารถที่จะตรวจจับใบหน้าและวัตถุต่างๆ ได้ ไม่ว่าจะเป็นใบหน้า หรือวัตถุนั้นจะวางในท่าทาง หรือแสงสว่างที่แตกต่างกันไป ดังนั้นจึงน่าจะมีคุณสมบัติ หรือลักษณะเฉาพะบางอย่างที่ไม่มีการเปลี่ยนแปลงไปตามท่าทาง หรือแสงสว่างที่เปลี่ยนไป วิธีหลายวิธีเสนอให้ตรวจจับลักษณะเฉพาะทางใบหน้าก่อน แล้วจึงสรุปว่ามีใบหน้าอยู่จริงหรือไม่ โดยการใช้การตรวจจับเส้นขอบ (edge detection) เพื่อดึงลักษณะเฉพาะทางใบหน้า เช่น คิ้ว ตา จมูก ปาก และเส้นผมออกมา จากนั้นจึงนำรูปแบบทางสถิติที่สร้างขึ้นมาจากตำแหน่งของสักษณะทางใบหน้าที่สัมพันธ์กัน มาพิสูจน์อีกครั้งว่ามีใบหน้าอยู่จริงหรือไม่ วิธีการนี้สามารถตรวจจับใบหน้าได้ในหลายมุมมอง แต่วิธีการนี้จะทำงานได้ไม่ดีนักในกรณีที่ภาพถูกรบกวนจากแสงสว่าง สัญญาณรบกวน และการถูกบังโดยวัตถุอื่นๆ

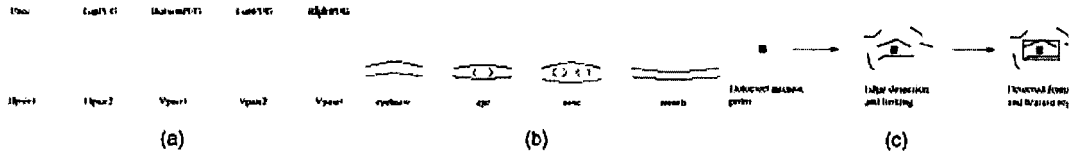
Yow และ Cipolla ได้นำเสนอวิธีการในการตรวจจับใบหน้าด้วยวิธี feature-based โดยใช้ algorithm แบบ facial features โดยมีหลักการในการทำงานคือ

-เริ่มจากการนำภาพมาผ่านกระบวนการทำ second derivative Gaussian filter เพื่อหาจุดที่สนใจ โดยดูจุดที่สูงสุดในบริเวณใกล้เคียงในภาพผลลัพธ์ที่ได้จากการทำ filter ซึ่งจุดที่สนใจนี้อาจจะเป็นตำแหน่งของลักษณะเฉพาะทางใบหน้าก็ได้

-ตรวจหาเส้นขอบรอบๆ จุดที่สนใจที่หามาได้จากขั้นตอนแรก และรวมกลุ่มของขอบและจุดเข้าด้วยกันเป็นบริเวณ จากนั้นถ้าขอบใดที่อยู่ใกล้ๆ กัน และมีการวางท่าและความหนาคล้ายคลึงกันก็ให้จัดไว้อยู่ในกลุ่มเดียวกัน (Grouping) จากนั้นทำการคำนวณหาค่าของคุณสมบัติของแต่ละบริเวณที่วัดได้ เช่น ความยาวของขอบ ความหนาของขอบ และการเปลี่ยนแปลงความเข้ม มาจัดเก็บไว้ในรูปแบบของเวกเตอร์ของคุณลักษณะเฉพาะ แล้วนำเอาเวกเตอร์ของคุณสมบัติเฉพาะที่ได้ของแต่ละบริเวณมาเปรียบเทียบกับค่าเฉลี่ย และค่า covariance matrix ของเวกเตอร์ของลักษณะเฉพาะแต่ละอันที่ได้จากข้อมูลของลักษณะเฉพาะทางใบหน้าที่ใช้ในการฝึกสอนระบบ โดยบริเวณใดๆ ในภาพจะเป็น facial feature candidate ได้ก็ต่อเมื่อ Mahalanobis distance ของเวกเตอร์ที่นำมาเปรียบเทียบกันนั้นต่ำกว่าค่า threshold ลักษณะเฉพาะเหล่านี้จะรวบรวมจัดกลุ่มเอาไว้ เพื่อนำไปพิจารณาต่อโดยแบบจำลองของความรู้ที่บอกถึงลักษณะเฉพาะอย่างใดอย่างหนึ่งควรเกิดที่บริเวณใดเมื่อเทียบกับลักษณะเฉพาะอันอื่นๆ จากนั้นมีการนำเอา Bayesian network ในการตรวจหา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใบหน้า โดยจะดูจากลักษณะต่างๆ ที่รวมกลุ่มกัน ว่าคุณลักษณะต่างๆ ที่รวมกลุ่มกันนี้เป็นใบหน้าจริงๆ หรือไม่



ภาพที่ 2.5 (a) Yow and Cipolla model ซึ่งประกอบไปด้วย 6 องค์ประกอบคือ ขนคิ้ว 2 ข้าง ตา 2 ข้าง จมูก และปาก (b) แบบจำลองเส้นขอบของทั้ง 6 องค์ประกอบ (c) กระบวนการ feature selection เริ่มจากจุดที่เราสนใจ และใช้ edge detection and linking ในการตรวจจับเส้นขอบ และตรวจสอบด้วยแบบรูปทางสถิติที่สร้างขึ้น

2.2. ทฤษฎีและหลักการในการพัฒนาระบบรู้จำใบหน้าของมนุษย์จากรูปภาพ

การศึกษา และ พัฒนาระบบรู้จำใบหน้ามนุษย์ (Face Recognition) เพื่อใช้ในการพัฒนาโครงงานนี้ ได้ศึกษาเทคนิคที่ชื่อว่า Eigenfaces ซึ่งเป็นส่วนที่จะทำให้ระบบพัฒนาที่พัฒนาขึ้นสามารถที่จะจำแนกภาพใบหน้าของมนุษย์แต่ละบุคคลได้ โดยระบบการรู้จำใบหน้านี้สามารถนำไปประยุกต์ใช้ในงานหลากหลายประเภท เช่น งานที่เกี่ยวข้องกับระบบความปลอดภัย ซึ่งต้องการความเจาะจงเฉพาะบุคคลในการเข้าถึงทรัพยากรขององค์กร เป็นต้น

2.2.1. การรู้จำใบหน้าโดยใช้วิธี Eigenface (Tou and Gonzalez. 1979; Strang. 1986)

หลักการของการรู้จำใบหน้าโดยใช้วิธี eigenface นั้น มีแนวคิดที่ว่าภาพทุกภาพในกลุ่มของภาพที่นำมาฝึกระบบนั้นสามารถแสดงในรูปแบบของการรวมกันของ principal component โดยจะให้ค่าน้ำหนักของความสำคัญของแต่ละ principal component ต่างๆ กันไป ดังนั้นจึงจำเป็นที่ต้องหาค่า principal component ของการกระจายของภาพต่างๆ หรือ หาค่า eigenvectors ของ covariance matrix ของกลุ่มของรูปภาพนั้นเอง eigenvectors เหล่านี้เปรียบกับกลุ่มของลักษณะเฉพาะซึ่งนำมาประกอบกันเพื่อที่จะแสดงคุณสมบัติของรูปแบบที่เปลี่ยนแปลงระหว่างใบหน้าต่างๆ ซึ่ง eigenvector นั้น ก็คือ eigenface นั่นเอง eigenface แต่ละรูปนั้นจะ orthonomal กัน หรือ แต่ละ eigenface นั้นไม่มีความเกี่ยวข้องกันเลย และ จำนวนของ eigenface มากที่สุดที่เป็นไปได้จะเท่ากับจำนวนรูปภาพในกลุ่มของรูปภาพที่นำมาฝึกระบบ แต่เราสามารถที่จะเลือกเฉพาะ eigenface ที่ดีที่สุด ซึ่งก็คือ eigenface ที่มีค่า eigenvalue สูงๆ เหตุผลหลักที่เลือกค่า eigenface ให้น้อยลง คือ เพื่อที่จะลดการคำนวณลงนั้นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลักการในการทำงานสามารถสรุปได้ดังนี้ คือ

1. นำภาพทั้งหมดในกลุ่มของภาพที่นำมาฝึกระบบไปคำนวณหา eigenface
2. เมื่อได้รับภาพที่ต้องการรู้จำ ให้นำภาพนั้นไปหาค่าน้ำหนักของความสำคัญของ eigenface แต่ละตัวกับภาพนั้น ซึ่งคือการ project ภาพลงบนไปยังแต่ละ eigenface
3. ตัดสินใจว่าภาพที่ได้รับมานั้นเป็นภาพที่มีใบหน้าอยู่หรือไม่
4. ถ้าภาพนั้นมีใบหน้าอยู่ ให้นำภาพนั้นไปหาว่าตรงกับ Class ใด

2.2.1.1 การคำนวณค่า Eigenfaces

การรู้จำใบหน้ามนุษย์ (Face Recognition) โดยใช้เทคนิค Eigenfaces นั้น จะใช้วิธีการของ Karhonen-Loeve (KL) โดยจะให้กลุ่มของรูปภาพที่นำมาฝึกระบบนั้นเป็น $\Gamma_1,\Gamma_2,\Gamma_3...\Gamma_M$ (ภาพในกลุ่มที่นำมาฝึกระบบมีทั้งหมด M ภาพ) จากนั้นนำภาพทุกภาพไปหาค่าเฉลี่ย โดย

$$\Psi = (\sum_{n=1}^{M} \Gamma_n) / M$$

เมื่อได้ค่าเฉลี่ยมาแล้วจะพิจารณาว่าแต่ละภาพห่างจากภาพเฉลี่ยเท่าไหร่ ซึ่งจะได้เวคเตอร์

$$\Phi_i = \Gamma_i - \Psi$$

เราต้องการที่จะหาค่า eigenface จากรูปภาพเหล่านี้ โดยขั้นตอนในการหา eigenface นั้น ต้องทำการหาค่า eigenvector (u_k) M เวคเตอร์ ที่ orthonormal กัน และ ค่า eigenvalue (λ_k) ของ covariance matrix ของ AA^T

$$C = \frac{1}{M}\sum_{n=1}^{M} \Phi_n \Phi_n^T$$

$$= AA^T$$

โดย matrix $A = [\Phi_1,\Phi_2...\Phi_M]$ แต่เนื่องจาก covariance matrix ที่ได้ออกมานั้นมีมิติที่ใหญ่มาก ขนาด N^2*N^2 (สมมติให้ภาพมีขนาดกว้าง N และ สูง N) ซึ่งทำให้การคำนวณไม่มีประสิทธิภาพ ดังนั้นจึงแก้ไขโดยหาค่า covariance matrix ของ A^TA แทน โดยมีมิติขนาด M*M ซึ่งมีขนาดเล็กกว่ามาก และนำค่า covariance นี้ ไปหาค่า eigenvectors และ eigenvalues โดย eigenvector ที่ได้ออกมานั้นจะเป็น eigenvector ของ covariance matrix ของ A^TA แต่ eigenvector ที่เราต้องการจริงนั้นคือ eigenvector ของ covariance matrix ของ AA^T ดังนั้นจึงต้องทำการแปลง eigenvector ของ A^TA (v_k)เป็น eigenvector ของ AA^T (u_k) เพื่อที่จะได้ค่า eigenface ออกมา และ เราสามารถที่จะกำหนดจำนวนของ eigenface (M') ที่ต้องการพิจารณาได้ โดยเลือกเรียงลำดับจาก eigenface ที่ดีที่สุด คือ เลือกจาก eigenface ที่มีค่า eigenvalue จากมากไปน้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาพที่ 2.6 แสดงตัวอย่างของ eigenface

เมื่อทำการหา eigenface ได้เรียบร้อยแล้วจะมาทำการสร้าง class ให้กับใบหน้าในกลุ่มของรูปภาพที่นำมาฝึกระบบ โดยจำนวน class จะเท่ากับจำนวนใบหน้าในกลุ่มของรูปภาพที่นำมาฝึกระบบ โดยในแต่ละ class จะประกอบด้วยค่าน้ำหนักของความสำคัญของ eigenface แต่ละตัว

2.2.1.2 การใช้ Eigenface ในการจัด class ให้กับภาพ

เมื่อได้สร้าง eigenfaces เรียบร้อยแล้ว จะมี eigenface จำนวน m' วางตัวอยู่เป็นส่วนย่อยๆ ของ image space ขนาด N2 ซึ่ง eigenfaces เหล่านี้จะใช้เป็นตัวจัด class ให้กับรูปภาพที่นำเข้ามา

เมื่อได้รับภาพที่ต้องการรู้จำเข้ามา จะทำการแปลงภาพนั้นเป็นอยู่ในรูปของผลรวมของ eigenface เหล่านั้นโดยจะให้ค่าน้ำหนักความสำคัญของแต่ละ eigenface ต่างๆกันไป การหาน้ำหนักทำได้โดย

$$\omega = u_k^T(\Gamma - \Psi); k = 1,\ldots,M$$

ค่าน้ำหนักที่ได้มานั้นจะอยู่ในรูปแบบของเวคเตอร์ $\Omega^T = [\omega_1\ \omega_2\ \ldots\ \omega_{M'}]$ มันจะบรรยายถึงว่าแต่ละ eigenface มีส่วนมากแค่ไหนในการแทนภาพที่นำเข้ามา และเราจะนำเวคเตอร์ที่ได้มานี้ไปตัดสินใจว่าใบหน้านั้นควรอยู่ไหน class ใด โดยนำเวคเตอร์ของน้ำหนักนี้ไปเปรียบเทียบกับเวคเตอร์ของน้ำหนักของแต่ละ class ที่เราได้เก็บไว้ในตอนฝึกระบบ และหาค่า Euclidian distance โดย

$\varepsilon_k = \| (\Omega - \Omega_k) \|$; Ω_k คือ เวคเตอร์น้ำหนักของ class k

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถ้า Euclidian distance ของ class ใดมีค่าต่ำที่สุด แสดงว่าใบหน้าที่ได้รับมานั้นน่าจะใกล้เคียงกับใบหน้าใน class นั้นมากที่สุด ดังนั้นจึงให้ใบหน้าที่ได้รับมานั้นอยู่ใน class ที่มีค่า distance ต่ำที่สุด

2.3. ศึกษาการใช้เครื่องมือที่ใช้ในการพัฒนาระบบ

ศึกษาวิธีการใช้เครื่องมือที่ใช้ในการพัฒนาระบบ และศึกษาวิธีการใช้งานเครื่องมือที่ช่วยในการสนับสนุนการพัฒนาระบบนี้ ซึ่งประกอบไปด้วย

2.3.1 หลักการเขียนโปรแกรมแบบเชิงวัตถุ (Object Oriented Programming) (Kruglinski. 1997.)

การเขียนโปรแกรมแบบ Object Oriented Programming (OOP) เป็นแนวคิดในการเขียนโปรแกรมแบบหนึ่ง โดยจะไม่มุ่งเน้นไปในการเขียนโปรแกรมเพียงอย่างเดียว แต่เป็นการมองภาพรวมของสิ่งที่อยู่รอบๆ ตัวเราโดยสามารถให้คำจำกัดความได้ดังนี้ "วัตถุแต่ละอย่างต่างก็มีลักษณะ และวิธีการใช้งานเป็นของตัวมันเอง" กล่าวคือ วัตถุแต่ละชนิดแต่ละอัน ต่างก็มีรูปร่างลักษณะ และการใช้งานที่แตกต่างกันออกไป เราจะเรียกคุณลักษณะของวัตถุนี้ว่า Attribute และเราจะเรียกวิธีการใช้งานวัตถุว่า Method ยกตัวอย่างเช่น จักรยานคันหนึ่งมีสีแดง สามารถใช้ขับขี่ได้ สีแดงจัดว่าเป็น Attribute ของจักรยาน ในขณะที่การใช้งานขับขี่ถือเป็น Method เป็นต้น

จากที่กล่าวมาข้างต้นจะเห็นได้ว่าการเขียนโปรแกรมแบบ OOP เป็นแนวคิดที่คล้ายธรรมชาติของสิ่งของสิ่งหนึ่งซึ่งเราสามารถแบ่งแยกสิ่งต่างๆ ออกเป็นประเภทๆ ได้ ถ้าเราได้นำเอาแนวคิดของ OOP มาใช้ในการเขียนโปรแกรม และการจัดการข้อมูล เราจะพบว่าโปรแกรม หรือฟังก์ชั่น จะมีความเป็นอิสระต่อกันอย่างเห็นได้ชัด อธิบายง่ายๆ ก็คือ โปรแกรม หรือฟังก์ชั่นแต่ละตัวถึงแม้จะมาจากที่เดียวกัน แต่สามารถทำงานในคนละหน้าที่หรือเก็บข้อมูลคนละค่าได้

คลาสใน OOP คือกลุ่มของวัตถุที่คุณลักษณะพื้นฐาน และมีฟังก์ชั่นพื้นฐานเหมือนกัน เช่น คลาสคน กล่าวคือไม่ว่าคนนั้นจะมีชื่อใด มีอาชีพใด หรือมีส่วนสูงเท่าไหร่ ก็จัดอยู่ในคลาสคน ซึ่งจะมีคุณลักษณะ และฟังก์ชั่นพื้นฐานเหมือนกัน โดยผู้เขียนโปรแกรมไม่สามารถนำคลาสไปใช้งานได้โดยตรง ต้องมีการสร้าง Object (หรือ Instance) ขึ้นมาก่อน ซึ่ง Object ที่สร้างจากคลาสใดจะมีคุณลักษณะพื้นฐานมาจากคลาสนั้นด้วย เพราะฉะนั้นการสร้างโปรแกรมด้วยวิธีการแบบ OOP จะทำให้การสร้างโปรแกรมมีความยืดหยุ่น และมีประสิทธิภาพสูง

คุณสมบัติพิเศษอีกอย่างหนึ่งที่ทำให้การเขียนโปรแกรมในแบบ OOP คือคุณสมบัติ การสืบทอด (Inheritance) การที่ OOP นำแนวคิดนี้มาใช้ทำให้ผู้เขียนโปรแกรมสามารถนำเอา Source Code (อาจเป็นคลาส) ที่เขียนไว้แล้วสามารถนำกลับมาใช้ใหม่ (Code Reuse) หรือมาเปลี่ยนแปลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แก้ไขให้เหมาะสมกับงานเฉพาะอย่าง ทำให้การเขียนโปรแกรมเป็นไปได้อย่างรวดเร็วและเป็นการใช้ Source Code ที่มีอยู่ให้เกิดประโยชน์สูงสุด

OOP มีการกำหนดระดับการเข้าถึงข้อมูลภายในคลาส เพื่อความปลอดภัยในการเข้าถึงข้อมูลภายในคลาส ซึ่งการกำหนดระดับการเข้าถึงข้อมูลภายในคลาสนี้เป็นคุณสมบัติอีกอย่างที่มีในแนวคิดแบบ OOP การเข้าถึงข้อมูลภายใน (Member Variable) หรือฟังก์ชั่นภายใน (Member Function หรืออาจเรียกว่า Method) แบ่งออกได้ 3 ระดับ ดังนี้คือ

- ระดับ Private เป็นระดับที่จะถูกกำหนดไว้อัตโนมัติ (Default) หากผู้เขียนโปรแกรมไม่กำหนด ระดับนี้เป็นระดับการป้องกันสูงสุด กล่าวคือเป็นการป้องกันไม่ให้กระบวนการใดๆ ที่อยู่นอกคลาสเรียกใช้สิ่งที่อยู่ในคลาสไม่ได้ การจะเข้าถึงข้อมูลภายในจะต้องผ่านจากฟังก์ชั่นที่เป็นของคลาสนั้นอนุญาตให้ใช้ได้

- ระดับ Public เป็นระดับที่ตรงข้ามกับระดับ Private กล่าวคือกระบวนการใดๆ ที่อยู่นอกคลาสนั้นสามารถเข้าถึงข้อมูลภายใน หรือฟังก์ชั่นภายในได้อย่างอิสระ ไม่จำเป็นต้องได้รับอนุญาตก่อนการเข้าใช้งาน

- ระดับ Protected คล้ายกับในระดับ Private แต่มีข้อแตกต่างคือระดับ Protected นั้น คลาสลูกที่สืบทอดคุณลักษณะจากคลาสแม่ จะสามารถเข้าถึงข้อมูลภายในคลาสแม่ได้โดยตรง

2.3.2 Microsoft Vision System Development Kit (Vision Technology Group. 2000)

Microsoft Vision System Development Kit หรือที่จักกันในนามของ MS Vision SDK ถูกพัฒนาขึ้นโดย Vision Technology Research Group เพื่อใช้ในการสนับสนุนการพัฒนาโปรแกรมทางด้าน Image Processing, Real-time Image Processing และ Computer Vision บนระบบปฏิบัติการ Microsoft Windows

MS Vision SDK มีคลาสและฟังก์ชั่นในภาษา C++ สำหรับการทำงานที่เกี่ยวข้องกับงานด้านภาพ ผู้เขียนโปรแกรมที่ติดตั้ง MS Vision SDK สามารถพัฒนาโปรแกรม โดยใช้ MS Visual C++ ในชุด Visual Studio ได้อย่างอิสระ ไม่ขึ้นกับอุปกรณ์ต่อเชื่อมที่นำสัญญาณเข้าเครื่องคอมพิวเตอร์ เพราะ MS Vision SDK อาศัย Windows Driver Model (WDM) ที่ทำหน้าที่เป็นตัวกลางในการติดต่อระหว่าง Application Software กับอุปกรณ์ต่อเชื่อม ดังนั้นผู้เขียนโปรแกรมจึงสามารถใช้อุปกรณ์ที่เป็นตัวนำข้อมูลภาพ ทั้งภาพนิ่งและภาพเคลื่อนไหวมาใช้ได้หลากหลายรูปแบบ เช่น กล้องถ่ายรูปดิจิตอล กล้องถ่ายวีดิโอดิจิตอล ที่ต่อกับอุปกรณ์คอมพิวเตอร์ผ่านทาง USB Port หรือผ่านทาง Port IEEE 1394 (Fire Wire) รวมทั้งกล้องวีดิโออนาล็อกที่ต่อผ่านทาง Capture card ที่ต่อกับคอมพิวเตอร์ผ่านทาง PCI bus เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

สรุปข้อดีและข้อเสียของ MS Vision SDK เมื่อเปรียบเทียบกับ ชุดพัฒนา Application Software ทางด้าน Image Processing อื่นๆ

ข้อดี

- สามารถประมวลผลได้รวดเร็วเหมาะสมกับงานแบบ real-time image processing

- การติดต่อระหว่าง Application กับระบบปฏิบัติการ MS Windows เป็นไปได้อย่างมีประสิทธิภาพ ในเรื่องของการใช้ image memory ร่วมกันระหว่าง processes ต่างๆ ภายใน Application รวมทั้งสามารถใช้งาน Graphic Device Interface

- ผู้ใช้งานสามารถสร้าง pixel data type ได้เอง (User-definable pixel) ซึ่งเหมาะสมกับงานวิจัยทางด้าน image processing

- ผู้เขียนโปรแกรมสามารถสร้าง Application โดยไม่ต้องคำนึงถึงอุปกรณ์ต่อเชื่อมที่ใช้ในการนำสัญญาณภาพเข้ามา ซึ่งรวมถึงสามารถต่อเชื่อมอุปกรณ์ใหม่ๆ ได้โดยไม่ต้องแก้ไขตัวโปรแกรม

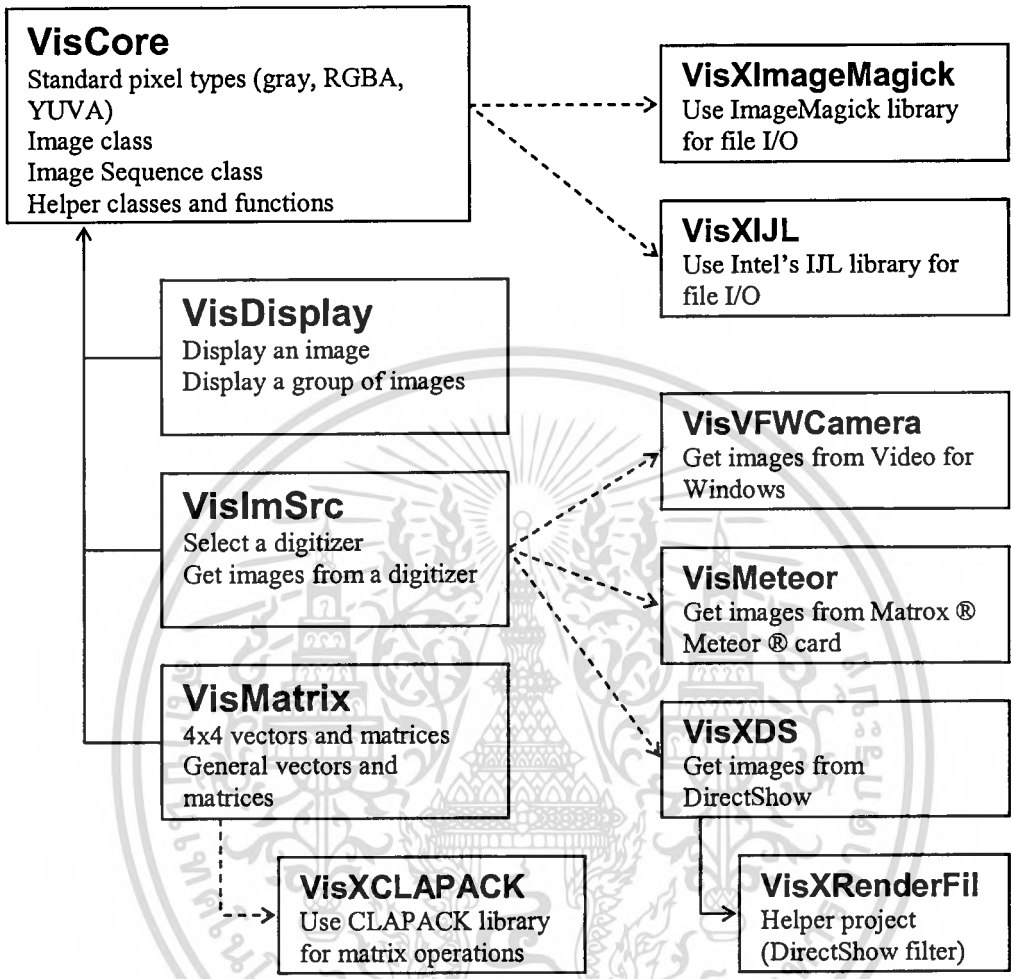
ข้อเสีย

- ข้อมูลภาพทั้งหมดจะอยู่ในหน่วยความจำ (RAM) ทำให้ไม่สามารถรองรับแฟ้มข้อมูลภาพ (Image files) ชนิดที่มีขนาดใหญ่เกินกว่าที่จะอยู่ในหน่วยความจำได้หมด MS Vision SDK สามารถรองรับขนาดข้อมูลภาพได้ประมาณ 2 GB (แต่ถ้าเป็น MS Windows 9x จะสามารถรับได้ประมาณ 1 GB)

- MS Vision SDK ไม่ได้เตรียมฟังก์ชั่นสำหรับการทำ image processing ในระดับ high-level ไว้ให้ เช่น การทำงาน threshold, erosion, dilation เป็นต้น ผู้สร้างจะต้องเขียนฟังก์ชั่นเหล่านี้ไว้ใช้งานเอง หรืออาจใช้ library ของ Intel's Image Processing Library แต่ได้เตรียมฟังก์ชั่นพื้นฐานในระดับ low-level สำหรับการใช้งานด้าน image processing ไว้แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โครงสร้างหลักของ MS Vision SDK



ภาพที่ 2.7 แสดงโครงสร้างหลักของ MS Vision SDK

ภาพที่ 2.7 แสดงโครงสร้างหลักของ MS Vision SDK ซึ่งประกอบไปด้วย header files ที่ใช้สำหรับการ include เพื่อให้ precomplier ทราบว่าจะมีการเรียกใช้ components หรือคลาสที่ได้ define ไว้มาใช้ในโปรแกรม โดยในที่นี้จะอธิบายในส่วนของ header file ที่สำคัญๆ ดังนี้คือ

- VisWin.h ซึ่งไม่ได้อยู่รวมกับ MS Vision SDK แต่จะอยู่ใน Microsoft Foundation Class (MFC) ซึ่งเป็น header file ที่ define macros พื้นฐาน และคลาสที่ต้องการใช้ใน MS Vision SDK

- VisCore.h ประกอบไปด้วยคลาสหลักๆ หลายคลาส เช่น คลาส CVisImage เป็นตัวที่เก็บข้อมูลภาพซึ่งใช้สำหรับการจัดเก็บ การนำเอาข้อมูลไปใช้ในการประมวลผล รวมทั้งการนำเสนอข้อมูลที่ผ่านการประมวลผลแล้ว คลาส CVisSequence ใช้สำหรับการจัดเก็บข้อมูลภาพในแบบต่อเนื่อง และการบันทึกภาพต่อเนื่องที่ผ่านการ Process แล้วในรูปแบบ AVI file และสุดท้าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ฟังก์ชั่น และฟังก์ชั่น DisplayInHdc เป็นการนำเสนอข้อมูลภาพที่ผ่านการ process แล้วด้วย Windows GDI

- VisImSrc.h สนับสนุนการ capture ภาพจากอุปกรณ์เชื่อมต่อ โดยมี Class CVisImageSource และฟังก์ชั่น VisFIndImageSource ในการค้นหาแหล่งที่มาของสัญญาณภาพ

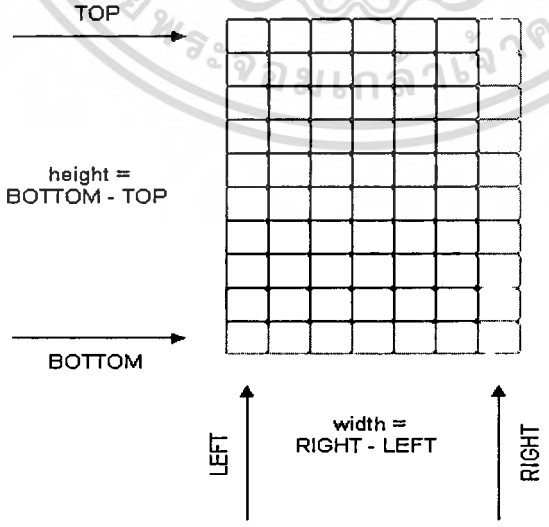
- VisDisplay.h ใช้ในการนำเสนอภาพที่ผ่านการ process แล้วเพื่อนำเสนอผ่านทางหน้าจอมอนิเตอร์ซึ่งมี Class CVisPane, CVisPaneArray รวมถึงฟังก์ชั่น VisDisplayImage

2.3.2.1 คลาส CVisImage

คลาส CVisImage เป็น คลาสหลักที่ใช้ในการทำงานของ MS Vision SDK คลาสนี้ถูกกำหนดไว้ใน Project "VisCore.h" คลาส CVisImage คล้ายคลึงกับ Windows bitmap header กล่าวคือ คลาส CVisImage หน้าที่เก็บคุณสมบัติหลายอย่างเกี่ยวกับภาพ อีกทั้งยังเก็บ pointer ที่ชี้ยังไปหน่วยความจำที่เก็บข้อมูลภาพจริง ข้อแตกต่างจาก bitmap คือ คลาสนี้ถูกเขียนโดยใช้ภาษา C++ และเขียนเป็นแบบ templates ที่สามารถเก็บและสามารถ process pixel format ชนิดต่างๆ กันได้

ข้อมูลภาพที่จัดเก็บใน Vision SDK จัดเก็บในรูป array ของ pixels ซึ่ง pixels ทั้งหมดจะมี data type ชนิดเดียวกันขึ้นอยู่กับผู้เขียนโปรแกรมว่าต้องการใช้ data type แบบใด pixel type (data type ของ pixel นั้นๆ) มีหลายแบบและมีทั้งแบบที่ MS Vision SDK กำหนดไว้แล้ว (Predefined Pixel and Image Types) และผู้เขียนโปรแกรมกำหนดเองภายหลัง (Adding a New Pixel Type)

MS Vision SDK ใช้โครงสร้างแบบ standard Windows RECT structure สำหรับการระบุข้อมูลเกี่ยวกับพื้นที่ของข้อมูลภาพ ดังแสดงในภาพที่ 2.8



ภาพที่ 2.8 โครงสร้างของการเก็บภาพใน Vision SDK

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากภาพตำแหน่ง (0, 0) คือตำแหน่ง TOP, LEFT ในขณะที่ pixel สุดท้ายของภาพคือตำแหน่ง BOTTOM, RIGHT โดยค่า height สามารถหาได้จากการนำค่า BOTTOM ลบด้วยค่า TOP และค่า width หาได้จากการนำค่า RIGHT ลบด้วยค่า LEFT

ผู้เขียนโปรแกรมสามารถอ้างถึงตำแหน่งของแต่ละ pixel ได้โดยตรง โดย MS Vision SDK จะกำหนดค่า x เป็นคอลัมน์ของภาพและค่า y เป็นแถวของภาพ โดยค่า x จะเริ่มต้นที่ 0 และเพิ่มค่าขึ้นทีละ 1 จากด้านซ้ายไปขวา ส่วนค่า y ก็จะเพิ่มค่าขึ้นทีละ 1 เหมือนค่า x ต่างกันตรงที่นับจากบนลงล่าง แต่จุด origin หรือจุด (0, 0) เป็นเพียงจุดสมมุติผู้เขียนสามารถกำหนดจุด origin ได้เองที่ตำแหน่งใดๆ ก็ได้ในภาพ ซึ่งทำให้การทำงานมีความยืดหยุ่นมากขึ้น

Object หรือ Instance ที่สร้างจากคลาส CVisImage นั้น ที่จริงแล้วไม่ได้เก็บข้อมูลของแต่ละ pixel จริงๆ แต่จะเก็บ pointer ที่ชี้ไปยัง block ข้อมูลจริงอีกทีหนึ่ง เพราะฉะนั้นผู้เขียนสามารถสร้าง objects หลายๆ objects โดยที่แต่ละ objects สามารถอ้างอิงไปที่ข้อมูลจริง โดยไม่คัดลอกข้อมูลไปยังที่ใหม่ โดยเมื่อไม่มีการใช้ object นั้นๆ แล้ว MS Vision SDK จะทำการตรวจสอบว่าถ้าไม่มีการอ้างอิงถึงข้อมูลจริงๆ อีกต่อไปแล้ว ก็จะทำการคืนหน่วยความจำส่วนนั้นไป

นอกจากนี้ในคลาส CVisImage ยังสามารถเลือกที่จะ Process เพียงแค่บางพื้นที่ของภาพเท่านั้น โดยใช้ method SubImage() ซึ่งสามารถกำหนดให้มีการ process เฉพาะตำแหน่งที่กำหนด

2.3.2.1.1 Pixel และ Image Type

CVisImage ถูกสร้างในแบบ template ในโครงสร้างของภาษา C++ ซึ่งเราสามารถกำหนดชนิดของ object ที่สร้างว่าจะให้เป็นแบบชนิดใดแล้วแต่ความต้องการ โดยใช้คำสั่ง CVisImage <pixeltype>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชนิดของพิกเซล (Pixel Types)

- Gray Scale Types แต่ละ pixel ประกอบไปด้วยค่าความสว่างของแต่ละ pixel ซึ่งจะมีระดับความสว่าง 256 ระดับ (0-255) ซึ่งใช้หน่วยความจำขนาด 8 บิต ตารางที่ 1 แสดงรายละเอียดของ pixel แต่ละชนิดในกลุ่มนี้

ตารางที่ 2.1 แสดงรายละเอียดของ pixel แต่ละชนิดในกลุ่ม Gray Scale Pixel Type

Name	Component Size	Data type	Pixel Size	Display	ImageSource
CVis(Gray)BytePixel	8 bits	Unsigned char	8 bits	Yes	Yes
CVis(Gray)CharPixel	8 bits	Signed char	8 bits	No	No
CVis(Gray)ShortPixel	16 bits	Signed short	16 bits	No	No
CVis(Gray)UShortPixel	16 bits	Unsigned short	16 bits	Yes	No
CVis(Gray)IntPixel	32 bits	Signed int	32 bits	No	No
CVis(Gray)LongPixel	32 bits	Signed long	32 bits	No	No
CVis(Gray)UIntPixel	32 bits	Unsigned int	32 bits	Yes	Yes
CVis(Gray)ULongPixel	32 bits	Unsigned long	32 bits	Yes	Yes
CVis(Gray)FloatPixel	32 bits	Float	32 bits	No	No
CVis(Gray)DoublePixel	64 bits	Double	64 bits	No	No

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- RGBA Color Pixel Types จะประกอบไปด้วย component ย่อยๆ ภายในแต่ละ pixel จำนวน 4 ค่าได้แก่ ค่าความสว่างในย่านสีแดง (method R()) ค่าความสว่างในย่านสีน้ำเงิน (method B()) ค่าความสว่างในย่านสีเขียว (method G()) และค่าอัลฟ่า (method A()) ซึ่งเป็นค่าที่กำหนดระดับการ transparency ตารางที่ 2 แสดงรายละเอียดของแต่ละ pixel ภายในกลุ่ม RGBA Color pixel types โดยลำดับในการเก็บคือ B-G-R-A ซึ่งตรงกับค่า RGBQUAD ซึ่งใช้ใน 32-bits Windows Bitmaps

ตารางที่ 2.2 แสดงรายละเอียดของ pixel แต่ละชนิดในกลุ่ม RGBA Color Pixel Types

Name	Component Size	Data type	Pixel Size	Display	ImageSource
CVisRGBABytePixel	8 bits	Unsigned char	32 bits	Yes	Yes
CVisRGBACharPixel	8 bits	Signed char	32 bits	No	No
CVisRGBAShortPixel	16 bits	Signed short	64 bits	No	No
CVisRGBAUShortPixel	16 bits	Unsigned short	64 bits	No	No
CVisRGBAIntPixel	32 bits	Signed int	128 bits	No	No
CVisRGBALongPixel	32 bits	Signed long	128 bits	No	No
CVisRGBAUIntPixel	32 bits	Unsigned int	128 bits	No	No
CVisRGBAULongPixel	32 bits	Unsigned long	128 bits	No	No
CVisRGBAFloatPixel	32 bits	Float	128 bits	No	No
CVisRGBADoublePixel	64 bits	Double	256 bits	No	No

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- YUVA Color Pixel Types คล้ายกลับกลุ่ม RGBA Color Pixel Types ที่ได้กล่าวมาแล้ว เพียงแต่ค่าความสว่างจัดเก็บในแบบ YUV component แทนที่จะเป็น RGB เท่านั้น ตารางที่ 3 แสดงรายละเอียดของ pixel แต่ละชนิดในกลุ่ม YUVA Color Pixel Types

ตารางที่ 2.3 แสดงรายละเอียดของ pixel แต่ละชนิดในกลุ่ม YUVA Color Pixel Types

Name	Component Size	Data type	Pixel Size	Display	ImageSource
CVisYUVABytePixel	8 bits	(Un)signed char	32 bits	Yes	Yes
CVisYUVACharPixel	8 bits	(Un)signed char	32 bits	Yes	Yes
CVisYUVAShortPixel	16 bits	(Un)signed short	64 bits	No	Yes
CVisYUVAIntPixel	32 bits	(Un)signed int	128 bits	No	No
CVisYUVALongPixel	32 bits	(Un)signed long	128 bits	No	No
CVisYUVAULongPixel	32 bits	(Un)signed long	128 bits	No	No
CVisYUVAUIntPixel	32 bits	(Un)signed int	128 bits	No	No
CVisYUVAFloatPixel	32 bits	Float	128 bits	No	No
CVisYUVADoublePixel	64 bits	Double	256 bits	No	No

2.3.2.1.2 การสร้าง CVisImage Object

การสร้าง CVisImage Object สามารถระบุขนาดของภาพได้โดยใช้ CVisImage constructor ในการกำหนดขนาดคอลัมน์ และแถวของภาพได้ เช่น

```
CVisByteImage image(100, 50);
```

คำสั่งข้างต้นเป็นการกำหนดขนาดของ Object ให้มีขนาด 100 คอลัมน์ และ 50 แถว หรือ กว้าง 100 pixels และสูง 50 pixels แต่ผู้เขียนโปรแกรมสามารถสร้าง object ก่อนแล้วระบุขนาดในภายหลังโดยใช้ method Allocate เช่น

```
CVisByteImage image;
image.Allocate(100, 50);
```

จากที่ได้กล่าวมาแล้วข้างต้นว่า CVisImage จะทำการเก็บ pointer ที่อ้างอิงถึงตำแหน่งของหน่วยความจำที่เก็บข้อมูลภาพจริง ในบางครั้งผู้เขียนโปรแกรมต้องการเก็บข้อมูลภาพชุดนั้นไปไว้ที่ตำแหน่งแยกกันต่างหาก method Copy() จะเป็นการสร้างหน่วยความจำที่ตำแหน่งใหม่และมีการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สร้างชุดข้อมูลใหม่สำหรับเก็บข้อมูลเดิม โดยที่การแก้ไข การ process ที่ข้อมูลชุดนี้จะไม่มีผลกระทบกับข้อมูลต้นฉบับแต่อย่างใด แต่ method Copy() จะทำงานได้เฉพาะกับข้อมูลในชนิดเดียวกัน แต่ถ้าเป็นกรณี เช่น เป็น pixel type ต่างชนิดกันต้องใช้คำสั่งอื่น คือ method CopyPixelTo() ดังตัวอย่างต่อไปนี้

```
// assign imageFaceOriginal to the subimage in the original data.
CVisByteImage imageFaceOriginal = imageOriginal.SubImage(rectFace);

// fill imageFaceCopy with a new image data block containing just the subimage
CVisByteImage imageFaceCopy.Copy(imageFaceOriginal);

// convert and copy the data into an RGBA version of the face subimage
CVisRGBAByteImage imageFaceRGBA(rectFace);
imageFaceOriginal.CopyPixelsTo(imageFaceRGBA);
```

จากตัวอย่างข้างต้นมีการใช้ method Copy() และ method CopyPixelTo() จะเห็นได้ว่า Copy() สามารถใช้ได้เฉพาะ pixel ในแบบเดียวกัน ส่วน CopyPixelTo() สามารถใช้เปลี่ยนแปลงข้ามกันไปมาได้ซึ่งคุณลักษณะนี้มีประโยชน์มากในการทำขบวนการทาง image processing อย่างการทำ binary threshold เป็นต้น

แต่ method CopyPixelTo() ไม่สามารถเปลี่ยน pixel จากกลุ่ม RGBA หรือ YUVA ไปเป็นกลุ่ม grayscale ได้ หรือแม้แต่ RGBA ไปเป็น YUVA ได้ แต่ MS Vision SDK ก็มีฟังก์ชั่นที่ใช้ในการแปลงค่าจาก RGBA ไปเป็น grayscale ได้คือ VisBrightnessFromRGBA ซึ่งมีสูตรในการคำนวณคือ

$$G = 0.299R + 0.587G + 0.114B$$

โดยที่ G คือ ค่าความสว่างในแถบ grayscale

R คือ ค่าความสว่างในย่านสีแดง

G คือ ค่าความสว่างในย่านสีเขียว

B คือ ค่าความสว่างในย่านสีน้ำเงิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.2.1.3 Operations พื้นฐานของคลาส CVisImage

method Pixel(x, y) เป็น method ในการเข้าถึงค่าความสว่างของแต่ละ pixel ในข้อมูลภาพทุกๆ ชนิด แต่ในบางครั้งการเขียนโปรแกรมต้องมีการเข้าถึงทุกๆ pixel ในภาพ method RowPointer(row) จึงมีความสะดวก และรวดเร็วกว่าในการทำงาน ดังตัวอย่างต่อไปนี้

```
float FindAverageGray(CVisFloatImage image)
{
        assert(image.NBands() == 1);
        float fltTotalIntensity = 0;
        for (int y = image.Top(); y < image.Bottom();y++)
        {
                float *pflt = image.RowPointer(y);
                for (int x = image.Left(); x < image.Right();x++)
                fltTotalIntensity += pflt[x];
        }
        return (fltTotalIntensity / image.NPoints());
}
```

จากตัวอย่างข้างต้น pflt เป็น pointer ที่รับค่าตำแหน่งของแถวของข้อมูลภาพที่แต่ละแถว เวลาต้องการเข้าถึงค่าในแต่ละ pixel จริงก็สามารถสั่งให้ pflt เลื่อนไปทีละตำแหน่งจนสุดแนวแกน x ได้

คลาส CVisByteImage และ CVisRGBABtyeImage มีความคล้ายกับ Windows Bitmaps เพราะฉะนั้น คุณสมบัติพื้นฐานบางอย่างจึงสามารถใช้ร่วมกันได้ อย่างเช่น การใช้ Windows GDI functions ในการแก้ไข การเสริมแต่งภาพได้โดยตรง การเรียกใช้ Windows GDI functions นั้นต้องมีการใช้ method Hdc เพื่อให้ได้ object ของคลาส HDC ซึ่งสามารถเรียกใช้ Windows GDI functions ได้ดังตัวอย่างต่อไปนี้

```
CVisRGBAByteImage image(100, 100);
HDC hdcImage = image.Hdc();
if (hdcImage != 0)
{
        // Note that we use (0, 0) to refer to the top-left corner of the
        // image when working with Windows GDI functions.
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
RECT rect;
rect.left = rect.top = 0;
rect.right = image.Width();
rect.bottom = image.Height();

HBRUSH hBrush = CreateSolidBrush((COLORREF) 0xff0000);
if (hBrush != 0)
{
        FillRect(hdcImage, &rect, hBrush);
        DeleteObject(hBrush);
}

DrawText(hdcImage, "Hello World!", - 1, &rect,
                DT_CENTER | DT_SINGLELINE | DT_VCENTER);
image.DestroyHdc();
}
```

จากตัวอย่างข้างต้นเป็นการสร้าง object ของคลาส CVisRGBAByteImage ขนาดความกว้าง 100 คอลัมน์ และสูง 100 แถว (100 x 100 pixels) โดยให้สีพื้นภายในเป็นสีน้ำเงิน และเขียนตัวอักษร “Hello World!” ไว้ตรงกลาง สังเกตว่ามีการใช้ Windows GDI functions เช่น DrawText() โดยต้องมีการสร้าง object ในคลาส HDC โดยรับค่าจาก method Hdc()

2.3.2.2 การแสดงข้อมูลภาพออกจากจอภาพคอมพิวเตอร์

เมื่อมีการสร้าง object ในคลาส CVisImage และมีการ process ข้อมูลภายในแล้ว การจะแสดงภาพที่ได้สามารถทำได้ 3 วิธีได้แก่

2.3.2.2.1 แสดงผลโดยใช้ Windows HDC

ถ้า object ที่สร้างจากคลาส CVisImage มีคุณสมบัติเหมือนใน Windows bitmaps ผู้เขียนโปรแกรมสามารถใช้ method DisplayInHdc() สำหรับการแสดงภาพบน Windows ได้ โดยจะต้องมีการสร้าง object ของคลาส HDC จากนั้นผ่านค่า window’s HWND ไปให้กับฟังก์ชั่น GetDC() โดยผู้เขียนโปรแกรมสามารถตรวจสอบผลลัพธ์ที่ได้ หากมีชนิดของข้อมูลภาพแตกต่างออกไป ผู้เขียนโปรแกรมสามารถใช้ method CopyPixelTo() สำหรับการเปลี่ยนแปลงชนิดของข้อมูลภาพได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชนิดของข้อมูลภาพที่ DisplayInHdc สนับสนุน

- CVisRGBAByteImage
- CVisByteImage
- CVisUShortImage (Displayed as RGB555)
- CVisUIntImage
- CVisYUVACharPixel (Displayed as gray scale)

2.3.2.2.2 แสดงผลโดยใช้คลาส CVisPane และ VisDisplayImage

คลาส CVisPane และ CVisPaneArray กำหนดอยู่ใน VisDisplay.h ผู้เขียนโปรแกรมสามารถ include VisDisplay.h และสามารถสร้าง object จากคลาส CVisPane และ CVisPaneArray โดยสามารถนำ object ที่สร้างจากคลาส CVisImage มาแสดงผลได้

2.3.2.3 การจัดการติดต่อกับอุปกรณ์ต่อพ่วง

Vision SDK สามารถติดต่อกับอุปกรณ์เชื่อมต่อจากภายนอกได้โดยที่ไม่ขึ้นกับอุปกรณ์ โดย MS Vision SDK ได้เตรียม VisImSrc DLL ไว้ให้ โดยผู้ใช้โปรแกรมสามารถเลือกอุปกรณ์เชื่อมต่อในขณะที่กำลังต่ออยู่กับคอมพิวเตอร์ได้

คลาส CVisImageSource ใน VisImSrc Project โดยสามารถใช้ฟังก์ชั่น VisFindImageSource() ซึ่ง return เป็น instance ของคลาส CVisImageSource ดังตัวอย่าง

CVisImageSource ImSrc = VisFindImageSource("");

จากตัวอย่างข้างต้น ImSrc จะทำการติดต่อกับอุปกรณ์เชื่อมต่อโดยจะเก็บข้อมูลภาพในลักษณะข้อมูลดิบเพื่อรอการ process ต่อไป ในกรณีต้องการติดต่อกับอุปกรณ์ที่เก็บภาพต่อเนื่อง เช่น กล้องถ่ายวีดิโอ สามารถใช้ method SetUseContinuousGrab() ได้

คลาส CVisSequence เป็นคลาสที่มีการใช้มากในการจัดการเกี่ยวกับ process ภาพต่อเนื่อง รวมถึงการอ่าน และการเขียนข้อมูลภาพที่เป็นภาพต่อเนื่อง เช่น AVI file คลาส CVisSequence อยู่ใน VisCore Project เหมือนกับคลาส CVisImage และมีลักษณะเป็น template เช่นเดียวกัน เพราะฉะนั้นจึงสามารถทำงานกับ data type ในชนิดต่างๆ กันได้ การทำงานของคลาส CVisSequence จะทำงานกับข้อมูลภาพที่มีลักษณะเป็นชุดโดยในแต่ละชุดจะต้องเก็บข้อมูลในประเภทเดียวกัน คลาส CVisSequence ใช้ Standard Template library deque class ในการจัดเก็บกลุ่มของข้อมูลภาพ

deque เป็นโครงสร้างข้อมูลที่เก็บ objects ไว้ภายใน และสามารถเข้าถึงข้อมูลโดยการใช้ดัชนี (index) สามารถเพิ่ม และสามารถลบข้อมูลออกจากตำแหน่งหัว หรือท้ายของ deque ได้ คุณสมบัติคล้ายในโครงสร้างข้อมูลแบบคิว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การใช้คลาส CVisSequence สามารถกำหนดขนาดของจำนวน objects ที่จะบรรจุไว้ใน deque ได้ โดยค่า default คือค่า 0 ซึ่งเวลาดึง object ไปใช้จะได้ object เป็นชุดล่าสุดที่เข้ามาใน deque การกำหนดว่าจะใช้ buffer เท่าไหร่ขึ้นกับชนิดของการทำงาน ตัวอย่างเช่น ถ้าต้องการ capture ภาพที่ได้สดๆ มาแสดงบนจอภาพคอมพิวเตอร์ควรตั้งไว้ในขนาดที่ไม่มาก คือ ประมาณ 0 ถึง 5 แต่ถ้าต้องการเก็บภาพจากส่วนอื่นของโปรแกรมไว้ก่อนทำการ process ก็ควรตั้งค่า buffer ขนาดปานกลางคือ ประมาณ 30 ถึง 200 และการตั้งค่า buffer ขนาดสูงคือ ประมาณมากกว่า 200 สำหรับการเก็บภาพแต่ละเฟรมเพื่อนำมาสร้างเป็นไฟล์วีดิโอ สำหรับ method ในการตั้งค่าขนาดของ buffer ของคลาส CVisSequence คือ SetLenghtMax() และยังมี method อื่นๆ เช่น ReadStream(), InsertStream(), AppendStream() และ WriteStream() ในการจัดการเกี่ยวกับไฟล์ AVI ด้วย

การนำเอา object แต่ละ object ที่อยู่ในคลาส CVisSequence มาทำการ process สามารถใช้ method Pop() โดยผู้เขียนสามารถสร้าง object ในคลาส CVisImage มารับไป process ต่ออีกทางหนึ่ง ดังตัวอย่างต่อไปนี้

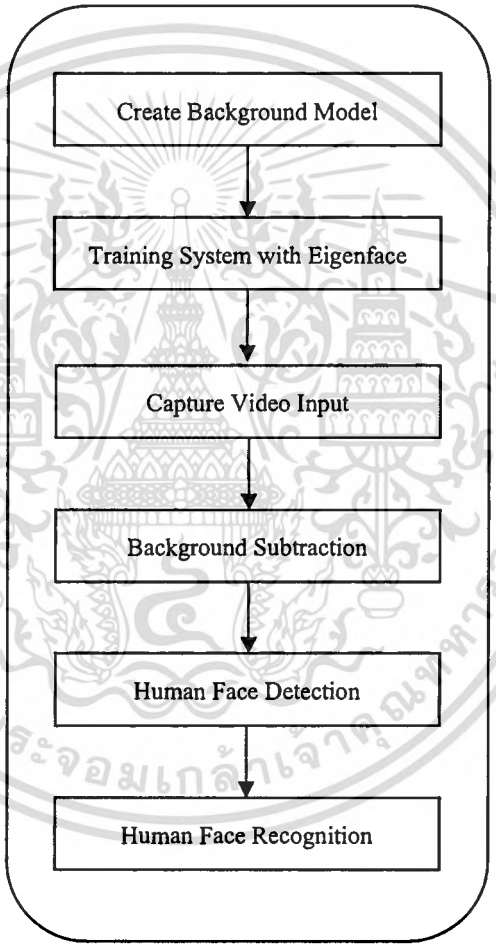
```
CVisImageSource imagesource = VisFindImageSource("");
if(imagesource.IsValid()) {
        CVisSequence<CVisRGBABytePixel> sequence;
        Sequence.ConnectToSource(imagesource, true, false);

        CVisRGBAByteImage imageT;
        if(sequence.Pop(image, 2000) {
                imageT.FWriteFile("out.bmp");
        }
        sequence.DisconnectFromSource();
}
```

จากตัวอย่างโปรแกรมข้างต้นมีการใช้ method ConnectToSource() เพื่อการเชื่อมต่อ object ในคลาส CVisSequence กับ Image Source จากนั้นจึงมีการใช้ method Pop() สำหรับการนำข้อมูลภาพภายใน deque ของ CVisSequence มาเก็บไว้ใน object ของคลาส CVisRGBAByteImage เพื่อทำการ process ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4. ขั้นตอนการประมวลผลของระบบ

การประมวลผลของระบบจะสามารถแบ่งการทำงานออกได้เป็น 2 ส่วนใหญ่ๆ ด้วยกัน ส่วนแรกคือการตรวจหาใบหน้ามนุษย์จากภาพวีดิโอ และส่วนที่สองคือส่วนของการรู้จำใบหน้ามนุษย์ ซึ่งกระบวนการในการทำงานของระบบสามารถเขียนขั้นตอนในการประมวลผลได้ดังต่อไปนี้



ภาพที่ 2.9 แสดงขั้นตอนการประมวลผลของระบบ

จากภาพที่ 2.9 สามารถอธิบายการทำงานในแต่ละขั้นได้ดังต่อไปนี้

- **Create Background Model** เป็นขั้นตอนในการสร้างแบบจำลองของภาพฉากหลัง โดยการรับภาพวีดิโอที่เป็นฉากหลังเข้ามาทำการประมวลผลเพื่อหาค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของภาพในแต่ละเฟรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

• **Training System width Eigenface** เป็นขั้นตอนในการนำเอาภาพในฐานข้อมูลที่ระบบมีมาฝึกให้กับระบบ ซึ่งเป็นขั้นตอนที่สำคัญขั้นตอนหนึ่งของ Face Recognition

• **Capture Video Input** เป็นขั้นตอนในการรับภาพวีดิโอที่ต้องการนำมาประมวลผลเข้าสู่ระบบ และแยกออกเป็นเฟรม เพื่อนำไปใช้ในขั้นตอนต่อไป

• **Background Subtraction** เป็นขั้นตอนที่นำเอาภาพวีดิโอที่แยกออกมาเป็นแต่ละเฟรมมาทำการ Subtract Background เพื่อแยกแยะเป้าหมาย (ในที่นี้คือมนุษย์) ออกจากภาพฉากหลัง

• **Human Face Detection** เป็นการนำเอาเป้าหมายที่แยกแยะได้จากขั้นตอนของ Background Subtraction มาทำการตรวจหาตำแหน่งใบหน้าของมนุษย์ว่าอยู่ในตำแหน่งใดของภาพในแต่ละเฟรม

• **Human Face Recognition** เป็นการนำเอาใบหน้ามนุษย์ที่ตรวจหาได้มาผ่านกระบวนการรู้จำใบหน้ามนุษย์เพื่อวิเคราะห์ ว่าใบหน้าที่ตรวจหาได้นั้น ตรงกับภาพที่ระบบรู้จำหรือไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

การทดสอบระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอ

3.1 อุปกรณ์ที่ใช้ในการทดสอบระบบ

3.1.1 กล้อง Web Cam ยี่ห้อ Lifeview รุ่น RoboCam ใช้ในการส่งภาพวีดิโอที่จับภาพได้เข้าสู่เครื่องคอมพิวเตอร์ที่ใช้ในการประมวลผลระบบ

3.1.2 เครื่องคอมพิวเตอร์รุ่น Pentium 4 1.7 GHz หน่วยความจำหลัก 512 MB ใช้ในการประมวลผล ซึ่งมีขั้นตอนในการประมวลผลดังที่ได้กล่าวไว้แล้วในบทที่ 2 หัวข้อที่ 2.4

3.2 ฐานข้อมูลภาพที่ใช้ในการทดสอบ

ฐานข้อมูลภาพที่ใช้ในการทดสอบระบบ ประกอบไปด้วยภาพใบหน้ามนุษย์จำนวนทั้งสิ้น 11 ใบหน้า โดยขนาดของภาพที่ใช้เป็นฐานข้อมูลมีขนาด 128 x 156 pixel

ภาพที่ 3.1 แสดงภาพที่ใช้ในฐานข้อมูลภาพของระบบรู้จำใบหน้ามนุษย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 การทดสอบระบบ

การทดสอบระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอที่พัฒนาขึ้น จำแนกการทดสอบออกเป็น 3 ส่วนหลักๆ คือ

1 การทดสอบการลบภาพฉากหลัง เพื่อแยกวัตถุเป้าหมายออกจากภาพฉากหลัง

2 การทดสอบการตรวจหาใบหน้ามนุษย์จากภาพวีดิโอในแต่ละเฟรมที่ได้ผ่านกระบวนการลบภาพฉากหลังออกแล้ว

3 การทดสอบการรู้จำใบหน้ามนุษย์ที่ตรวจหาได้

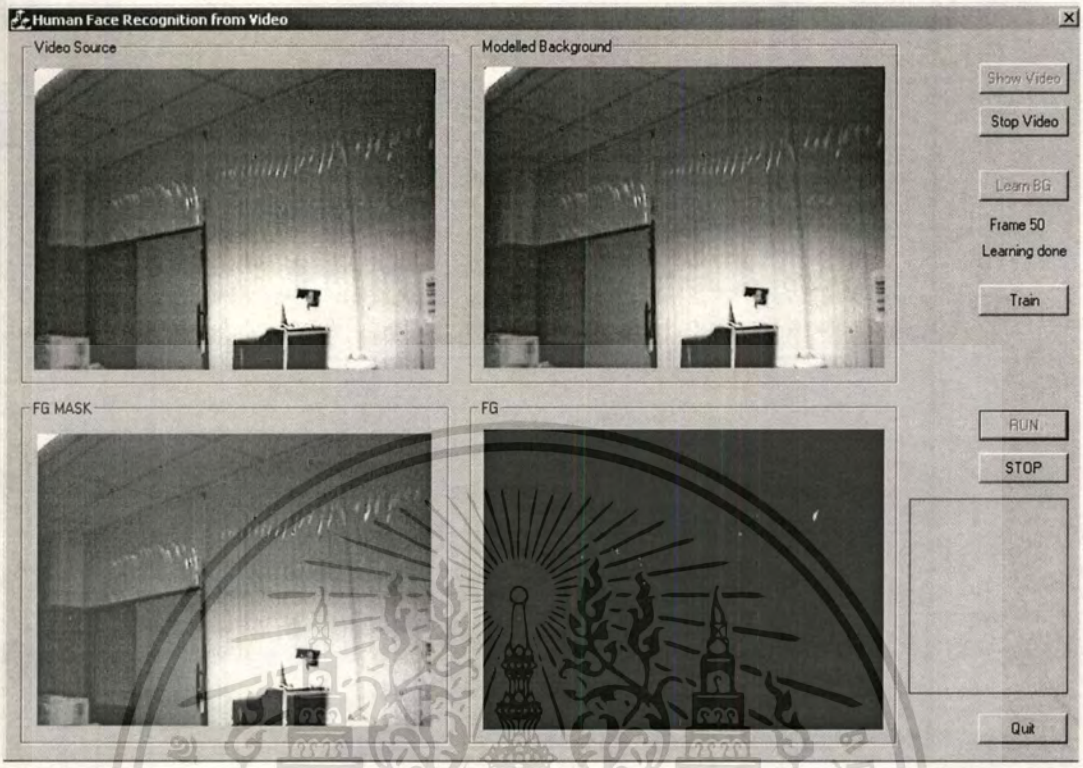
โดยขั้นตอนในการทดสอบระบบ และผลการทดสอบสามารถอธิบายได้ดังนี้

3.3.1 การทดสอบการลบภาพฉากหลัง

การทดสอบการลบภาพฉากหลังเริ่มการรับภาพวีดิโอฉากหลังมาจากกล้อง และทำการประมวลผลเพื่อสร้างแบบจำลองฉากหลัง โดยจำนวนของภาพวีดิโอฉากหลังที่นำมาใช้ในการสร้างแบบจำลองฉากหลังนี้ ใช้จำนวนภาพ 50 เฟรม

เมื่อสร้างแบบจำลองฉากหลังเสร็จเรียบร้อยแล้ว จึงทำการทดสอบแบบจำลองฉากหลังที่ได้ว่ามีคุณภาพหรือไม่ โดยการเปิดให้ระบบทำงาน ซึ่งระบบจะทำการรับภาพกล้องมาทีละเฟรม (เป็นภาพที่ยังไม่มีวัตถุเป้าหมายอยู่ในภาพ) และทำการนำภาพที่ได้ไปลบออกจากแบบจำลองฉากหลัง เพื่อทดสอบการทำงานของระบบในส่วนของการลบภาพฉากหลังว่ามีประสิทธิภาพมากเพียงพอที่จะนำไปใช้งานในระบบจริงได้หรือไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

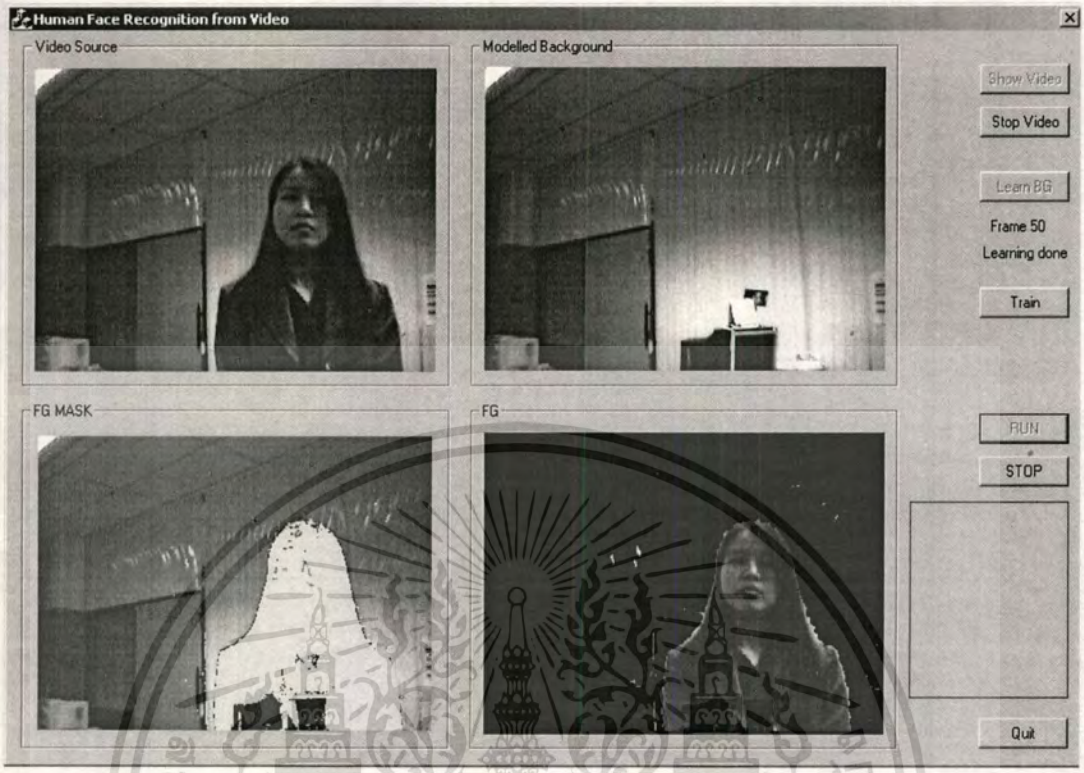


ภาพที่ 3.2 แสดงตัวอย่างผลการทดสอบการลบภาพฉากหลัง โดยภาพบนซ้ายจะเป็นภาพวีดิโอที่รับมาจากกล้องเพื่อใช้ในการประมวลผล ภาพบนขวาเป็นภาพแบบจำลองฉากหลังที่ระบบสร้างขึ้น ส่วนภาพสองภาพด้านล่าง เป็นภาพที่ผ่านกระบวนการลบฉากหลังเรียบร้อยแล้ว

จากภาพที่ 3.2 เป็นผลการทดสอบโดยการนำเอาภาพที่มีแต่ฉากหลัง ไม่มีวัตถุเป้าหมายในภาพ (ภาพซ้ายบน) มาทำการลบออกจากแบบจำลองภาพฉากหลัง (ภาพขวาบน) ซึ่งภาพผลลัพธ์ที่ผ่านกระบวนการลบภาพฉากหลังแล้วจะเป็นภาพทางด้านซ้ายล่างและด้านขวาล่าง โดยจะเห็นว่าผลลัพธ์ที่ได้มีคุณภาพดี มีสัญญาณรบกวนในภาพผลลัพธ์เพียงเล็กน้อย ทั้งนี้สัญญาณรบกวนในภาพเกิดจากการเคลื่อนไหวของม่านที่เป็นฉากหลัง

การทดสอบการลบภาพฉากหลังในขั้นตอนต่อไป คือการทดสอบการลบภาพฉากหลังที่วัตถุเป้าหมาย (มนุษย์) อยู่ในภาพฉากหลังด้วย โดยผลการทดสอบสามารถแสดงได้ดังภาพที่ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 3.3 แสดงตัวอย่างผลการทดสอบการลบภาพฉากหลังที่มีวัตถุเป้าหมายอยู่ในภาพ โดยภาพบนซ้ายจะเป็นภาพวีดิโอที่รับมาจากกล้องเพื่อใช้ในการประมวลผล ภาพบนขวาเป็นภาพแบบจำลองฉากหลังที่ระบบสร้างขึ้น ส่วนภาพสองภาพด้านล่าง เป็นภาพที่ผ่านกระบวนการลบฉากหลังเรียบร้อยแล้ว

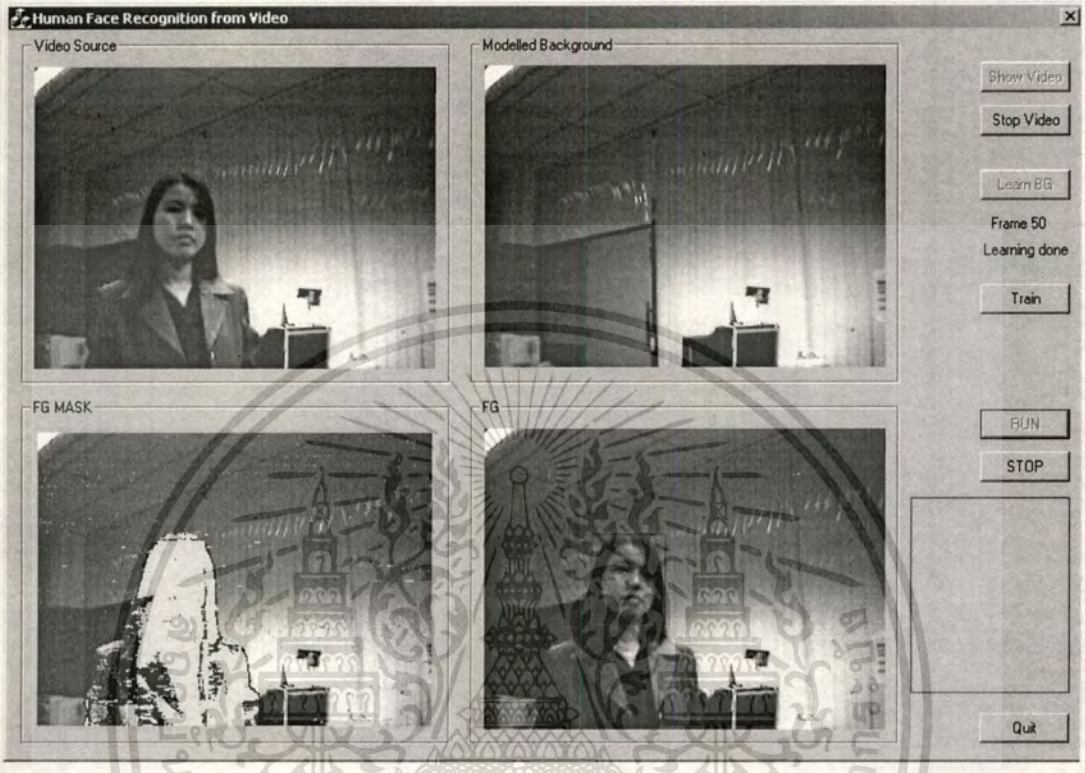
จากภาพที่ 3.3 จะเห็นว่าระบบสามารถแยกเป้าหมายออกจากภาพฉากหลังได้อย่างถูกต้อง แต่จากภาพผลลัพธ์จะเห็นว่าบริเวณด้านล่างของภาพบริเวณที่มีคนยืนอยู่นั้น ระบบยังแยกแยะฉากหลังและคน (เป้าหมาย) ได้ยังไม่ถูกต้องนัก ทั้งนี้เกิดจากแบบจำลองฉากหลังในบริเวณนั้นคือเครื่องคอมพิวเตอร์ซึ่งมีสีดำ ซึ่งคล้ายกับสีเสื้อของคนที่ยืนอยู่ในภาพ จึงทำให้ระบบไม่สามารถแยกระหว่างสีของเสื้อและสีของเครื่องคอมพิวเตอร์ที่เป็นฉากหลังได้อย่างสมบูรณ์

3.3.2 การทดสอบการตรวจหาใบหน้ามนุษย์

ในการทดสอบระบบในส่วนของการตรวจหาใบหน้ามนุษย์นั้น เป็นส่วนที่มีการทำงานต่อเนื่องจากส่วนของการลบภาพฉากหลัง กล่าวคือ เมื่อมีการนำภาพวีดิโอที่รับจากกล้องมาทำการลบฉากหลังออกแล้ว จะทำให้ระบบสามารถตรวจหาตำแหน่งของวัตถุเป้าหมายในภาพได้อย่างถูกต้องแม่นยำมากขึ้น โดยระบบนี้จะทำการตรวจสอบหาตำแหน่งใบหน้าของมนุษย์ในภาพจากการหาบริเวณตำแหน่งบนสุดของคนที่อยู่ในภาพ เมื่อพบจะทำการหาขอบเขตของศีรษะทาง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ด้านซ้าย และทางด้านขวา เพื่อให้สามารถระบบได้ว่าใบหน้าของมนุษย์ในภาพมีขอบเขตอยู่บริเวณใดของภาพ



ภาพที่ 3.4 แสดงตัวอย่างผลการทดสอบการตรวจหาใบหน้ามนุษย์จากภาพวีดิโอ

จากภาพที่ 3.4 ในภาพทางด้านขวาล่างเป็นภาพผลลัพธ์ที่ผ่านกระบวนการตรวจหาใบหน้ามนุษย์เรียบร้อยแล้ว โดยเมื่อระบบทำการตรวจหาตำแหน่ง และขอบเขตของใบหน้ามนุษย์ในภาพได้ ระบบจะทำการตีกรอบเพื่อแสดงขอบเขตของใบหน้ามนุษย์ที่ระบบตรวจหาได้

แต่จากขั้นตอนในการตรวจหาใบหน้ามนุษย์ที่ได้กล่าวมาแล้วนั้น สามารถทำงานได้ดีในกรณีที่ภาพที่ผ่านกระบวนการลบภาพฉากหลังมาแล้ว มีขนาดของสัญญาณรบกวนไม่ใหญ่มาก เพราะหากมีขนาดของสัญญาณรบกวนในภาพที่มีขนาดใหญ่มากๆ อันอาจจะเกิดจากการเปลี่ยนไปของแสงเงาของฉากหลังนั้น จะมีผลทำให้การตรวจหาตำแหน่งใบหน้ามนุษย์คลาดเคลื่อนไปได้ หากสัญญาณรบกวนดังกล่าว อยู่ในบริเวณที่สูงกว่าส่วนศีรษะของคนในภาพ

3.3.3 การทดสอบการรู้จำใบหน้ามนุษย์

การทดสอบการรู้จำใบหน้ามนุษย์ ซึ่งเป็นการทดสอบการทำงานส่วนสุดท้ายของระบบ โดยในส่วนการรู้จำใบหน้ามนุษย์ จะเป็นการนำเอาภาพใบหน้ามนุษย์ที่ตรวจหาได้ไปทำการวิเคราะห์เพื่อตรวจหาว่าตรงกับที่ระบบรู้จำหรือไม่ จะแสดงผลลัพธ์ที่วิเคราะห์ได้ออกมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาพที่ 3.5 แสดงตัวอย่างผลการทดสอบการรู้จำใบหน้ามนุษย์จากภาพวีดิโอ

จากภาพที่ 3.5 เป็นภาพแสดงตัวอย่างของการรู้จำใบหน้ามนุษย์จากภาพวีดิโอ โดยจากผลการรู้จำที่ได้จะเปลี่ยนแปลงไปตามลักษณะของใบหน้าที่ตรวจจับได้ เช่นหากตรวจจับขอบเขตของใบหน้าได้ถูกต้อง ระบบการรู้จำก็จะทำงานได้อย่างมีประสิทธิภาพด้วย แต่หากขอบเขตใบหน้าที่ตรวจจับได้ไม่ถูกต้อง ระบบการรู้จำก็จะทำงานได้ในประสิทธิภาพที่ต่ำลง

3.4 สรุปผลการทดสอบ

จากการทดสอบระบบทั้ง 3 ส่วน ตามที่ได้กล่าวมาข้างต้นแล้วนั้น สามารถสรุปผลการทดสอบได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.1 การลบภาพฉากหลัง

ประสิทธิภาพของระบบในส่วนของการลบภาพฉากหลัง สามารถแยกแยะวัตถุเป้าหมายออกจากฉากหลังได้อย่างมีประสิทธิภาพในกรณีที่แสงและเงาของฉากหลังไม่เปลี่ยนแปลงไปมากนัก เมื่อวัตถุเป้าหมายเคลื่อนที่เข้ามาในภาพวีดิโอ เพราะหากแสงและเงาเปลี่ยนแปลงไปมาก จะทำให้วัตถุเป้าหมายที่แยกได้ มีคุณภาพไม่ดีนัก เพราะมีส่วนของแสงเงาที่เปลี่ยนแปลงไปติดอยู่ในภาพที่ลบฉากหลังเรียบร้อยแล้ว กลายเป็นสัญญาณรบกวน (Noise) ภายในภาพ และหาก Noise ดังกล่าวมีขนาดใหญ่เพียงพอ จะมีผลทำให้การระบบในส่วนของการตรวจหาใบหน้ามนุษย์ทำงานผิดพลาดตามไปด้วย

จากการทดสอบระบบจะพบว่าปัจจัยที่มีผลต่อประสิทธิภาพของระบบในส่วนของการลบภาพฉากหลัง คือ คุณภาพของภาพที่ได้จากกล้องที่ใช้ในการทดสอบ เนื่องจากกล้องที่ใช้ในการทดสอบเป็นกล้องที่มีคุณภาพไม่สูงนัก มีความสามารถในการเก็บรายละเอียดของภาพไม่สูงมาก และแสงสว่างภายในห้องทดสอบที่เปลี่ยนแปลงไปเพียงเล็กน้อยก็ส่งผลให้ภาพฉากหลังเปลี่ยนแปลงไปทำให้เมื่อนำภาพฉากหลังที่ได้ไปทำการลบออกจากแบบจำลองภาพฉากหลังแล้ว จึงมี Noise เกิดขึ้นในภาพเป็นจำนวนมาก ทำให้การทำงานในส่วนอื่นๆ ของระบบที่ต้องทำงานต่อเนื่องจากส่วนของการลบภาพฉากหลังทำงานผิดพลาดไปบ้าง ตามคุณภาพของภาพที่ได้จากกล้องที่ใช้ในการทดสอบแต่ละเฟรม

3.4.2 การตรวจหาใบหน้ามนุษย์

ระบบในส่วนการตรวจหาใบหน้ามนุษย์จะทำงานได้ดีในกรณีภาพที่ผ่านกระบวนการลบภาพฉากหลังออกไปแล้ว เป็นภาพที่มีแต่วัตถุเป้าหมายเพียงอย่างเดียว หรือเป็นภาพที่มีวัตถุเป้าหมายเป็นหลักในภาพและมี Noise ภายในภาพได้ แต่ Noise ในภาพจะต้องมีขนาดเล็ก ไม่ใหญ่เกินกว่าที่ระบบกำหนดไว้ เพราะหาก Noise มีขนาดใหญ่เกินกว่าที่ระบบกำหนดไว้ ระบบตรวจหาใบหน้ามนุษย์ อาจทำงานผิดพลาดได้ คือระบบอาจวิเคราะห์บริเวณที่เป็น Noise เป็นใบหน้ามนุษย์ได้ (ในกรณีที่ Noise ขนาดใหญ่นั้น เกิดขึ้นบริเวณที่อยู่สูงกว่าศีรษะของมนุษย์ในภาพ)

3.4.3 การรู้จำใบหน้ามนุษย์

ระบบในส่วนการรู้จำใบหน้ามนุษย์เป็นส่วนที่มีการทำงานต่อเนื่องจากระบบในส่วนการตรวจหาใบหน้ามนุษย์ ดังนั้นผลจากการวิเคราะห์การรู้จำใบหน้ามนุษย์จะมีความถูกต้องมากน้อยเพียงใด ขึ้นอยู่กับใบหน้าที่ตรวจจับได้ว่ามีคุณภาพหรือไม่ เช่น ระบุขอบเขตของใบหน้าในส่วนของตำแหน่งด้านบนสุดของศีรษะ บริเวณคางซึ่งเป็นส่วนของศีรษะด้านล่าง บริเวณขอบซ้าย และ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขอบขวาของใบหน้าได้ถูกต้องตามขอบเขตที่ควรจะเป็นหรือไม่ เพราะหากขอบเขตของใบหน้าที่ตรวจหาได้ไม่เหมาะสมกับการนำไปใช้ในการรู้จำใบหน้ามนุษย์ จะส่งผลให้ระบบในส่วนของการรู้จำใบหน้ามนุษย์ทำงานผิดพลาดตามไปด้วย ซึ่งจากผลการทดสอบการทำงานของระบบ จะพบว่า การรู้จำใบหน้ามนุษย์คนๆ เดียวกันภาพวีดิโอที่มีความต่อเนื่อง แต่เป็นการรู้จำใบหน้ามนุษย์ที่ได้จากการตรวจหาใบหน้ามนุษย์ในภาพแต่ละเฟรม จะพบว่า ใบหน้ามนุษย์ที่รู้จำได้ในแต่ละเฟรม จะเปลี่ยนแปลงไปตามคุณภาพของใบหน้าที่ได้จากการตรวจหาใบหน้ามนุษย์ คือ ถ้าการตรวจหาตำแหน่งใบหน้ามนุษย์สามารถระบุขอบเขตของใบหน้าได้อย่างถูกต้อง ระบบรู้จำใบหน้ามนุษย์จะสามารถรู้จำใบหน้ามนุษย์ได้ถูกต้องด้วยเช่นกัน แต่หากตำแหน่งของใบหน้าที่ระบบตรวจหาใบหน้ามนุษย์ตรวจหาได้ระบุขอบเขตของใบหน้าได้ใหญ่หรือเล็กเกินขอบเขตที่ควรจะเป็น ระบบการรู้จำใบหน้าก็จะทำงานได้อย่างไม่มีประสิทธิภาพมากนัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

บทสรุปและข้อเสนอแนะ

4.1 บทสรุปโครงการ

โครงการพัฒนาระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอ เป็นการศึกษาทฤษฎีที่เกี่ยวข้องกับการพัฒนาระบบในส่วนต่างๆ และเป็นการสร้างระบบให้สามารถทำงานได้ตามที่กำหนดไว้ในขอบเขตของการพัฒนาระบบ โดยระบบประกอบไปด้วยส่วนหลักๆ 2 ส่วนด้วยกัน คือ การตรวจหาใบหน้ามนุษย์จากภาพวีดิโอแต่ละเฟรม และการรู้จำใบหน้ามนุษย์ แต่ในการพัฒนาระบบยังพบปัญหาและข้อจำกัดในการพัฒนาระบบหลายอย่าง ดังจะได้กล่าวในหัวข้อถัดไป

4.2 ปัญหาและข้อจำกัดในการทำงานของระบบ

ในการพัฒนาระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอ มีปัญหาและข้อจำกัดในการพัฒนาหลายอย่าง เนื่องจากระบบรู้จำใบหน้ามนุษย์จากภาพวีดิโอ มีปัจจัยและข้อจำกัดที่ส่งผลต่อการทำงานของระบบ ดังต่อไปนี้

4.2.1 คุณภาพของภาพวีดิโอที่ได้จากกล้อง Web Cam มีคุณภาพไม่สูงนัก มี Noise รบกวนภายในภาพ ความละเอียดของภาพที่ได้ไม่สูงมาก และอัตราการส่งภาพเข้าสู่เครื่องคอมพิวเตอร์ต่ำ ทำให้การประมวลผลไม่ต่อเนื่อง

4.2.2 ระบบการตรวจหาใบหน้ามนุษย์ยังไม่สามารถตรวจหาใบหน้ามนุษย์ในภาพที่มีแสงและเงาของฉากหลังเปลี่ยนแปลงไปมากๆ ได้อย่างถูกต้อง ทำให้ระบบการรู้จำใบหน้ามนุษย์ทำงานผิดพลาดไปด้วย เนื่องจากเป็นส่วนที่มีการทำงานสืบเนื่องกัน

4.2.3 ระบบตรวจหาใบหน้ามนุษย์สามารถตรวจหาใบหน้ามนุษย์ได้เพียงใบหน้าเดียวเท่านั้นในภาพวีดิโอแต่ละเฟรม ไม่สามารถตรวจหาใบหน้ามนุษย์คนอื่นได้หากในภาพมีใบหน้ามนุษย์มากกว่า 1 ใบหน้า

4.2.4 ระบบตรวจหาใบหน้ามนุษย์จะสามารถทำงานได้อย่างถูกต้อง ในกรณีที่ศีรษะของมนุษย์ในภาพวีดิโออยู่ด้านเสมอเท่านั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.5 ข้อมูลภาพในฐานข้อมูลภาพที่ใช้ในระบบรู้จำใบหน้ามนุษย์ยังไม่มีประสิทธิภาพมากนัก เนื่องจากภาพใบหน้าในฐานข้อมูลต้องมีขนาดเท่ากัน ทำให้ต้องการตกแต่งขนาดของภาพใบหน้ามีผลให้อัตราส่วนของใบหน้าผิดเพี้ยนไป

4.2.6 แสงและเงาภายในห้องที่เปลี่ยนแปลงไป มีผลให้ประสิทธิภาพของระบบรู้จำใบหน้าลดลง รวมถึงแสงและเงาของภาพในฐานข้อมูลภาพ หากภาพมีแสงและเงาที่ต่างจากความเป็นจริงมาก จะทำให้ประสิทธิภาพของระบบรู้จำลดลงเช่นกัน

4.2.7 ระบบมีทำงานแบบตอบสนองทันทีทันใด ทำให้มีการประมวลสูง ใช้ทรัพยากรต่างๆ เช่น CPU, RAM และ HARDDISK สูง จำเป็นต้องใช้เครื่องที่มีประสิทธิภาพ และมีหน่วยความจำสูงในการประมวลผลระบบ

4.3 ข้อเสนอแนะในการพัฒนาระบบรู้จำใบหน้ามนุษย์ต่อไป

4.3.1 เพิ่มเติมการตรวจจับใบหน้ามนุษย์ทุกใบหน้าในภาพวีดิโอที่มีจำนวนใบหน้ามากกว่า 1 ใบหน้าได้ ทั้งนี้เนื่องจากใบหน้ามนุษย์ในภาพวีดิโอ มีความเป็นไปได้ที่จะมีจำนวนใบหน้ามากกว่า 1 ใบหน้า

4.3.2 เพิ่มเติมการตรวจจับใบหน้ามนุษย์ในกรณีที่ส่วนศีรษะหรือใบหน้ามนุษย์ในภาพวีดิโอไม่ได้อยู่ด้านบนสุดเสมอ เพื่อป้องกันการหลบหลีกใบหน้าของผู้ไม่ประสงค์ดีที่ต้องการเข้าสู่ตัวอาคาร หรือสถานที่ที่ไม่อนุญาตให้เข้าได้

4.3.3 เพิ่มเติมการรับภาพวีดิโอที่เป็นในลักษณะของการบันทึกเทปเก็บไว้ในรูปสื่อต่างๆ เช่น วีดิโอเทป หรือเป็นไฟล์วีดิโอ เป็นต้น เพื่อประโยชน์ในการประมวลผลระบบ ในกรณีที่ไม่สามารถติดตั้งเครื่องคอมพิวเตอร์เพื่อประมวลผลระบบได้ตลอดเวลา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

Gallager, R. G. 1968. **Information Theory and Reliable Communication**. New York: Wiley.

Horprasert, T. et.al. 1999. "A statistical Approach for Real-Time Robust Background Subtraction and Shadow Detection," **Proc. IEEE ICCV'99 Framerate Workshop**, Kerkyra, Greece.

Kruglinski, D.J. 1997. **Inside Visual C++**. 4th edition. Redmond: Microsoft Press.

Papoulis, A. 1991. **Probability, Random Variables, and Stochastic Processes**, 3rd ed. New York: McGraw-Hill.

Seki, Makito. et.al. 2000. "A Robust Background Subtraction Method for Changing Background," **Industrial Electronics and Systems Lab**, pp. 207-213.

Strang, G. 1986. **Linear Algebra and its Applications**, 3rd ed. Orlando, FL: Harcourt Brace Jovanovich.

Tou, J. and Gonzalez, R. 1979. **Pattern Recognition Principles**. Reading, MA: Addison-Wesley.

Vision Technology Group, Microsoft Research, 2000. **The Microsoft Vision SDK Version 1.2**. [Online]. Available: http://research.microsoft.com/projects/VisSDK/VisSDK.doc

Yang, Ming-Hsuan. et.al. 2002. "Detecting Faces in Images: A Survey," **IEEE Transaction on Pattern Analysis and Machine Intelligence**, Vol 24, No.1.

Yow, K.C. and Cipolla, R. 1997. "Feature-Based Human Face Detection," **Image and Vision Computing**, Vol. 15, No. 9, pp. 713-735.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อผู้เขียน	นายพนมพร สาคร
วัน เดือน ปีเกิด	3 พฤศจิกายน 2519
สถานที่เกิด	ต. แสนสุข อ. เมือง จ. ชลบุรี
วุฒิการศึกษาระดับปริญญาตรี	วิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์
สถานที่สำเร็จการศึกษา	มหาวิทยาลัยบูรพา
ปีที่สำเร็จการศึกษา	2541

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้