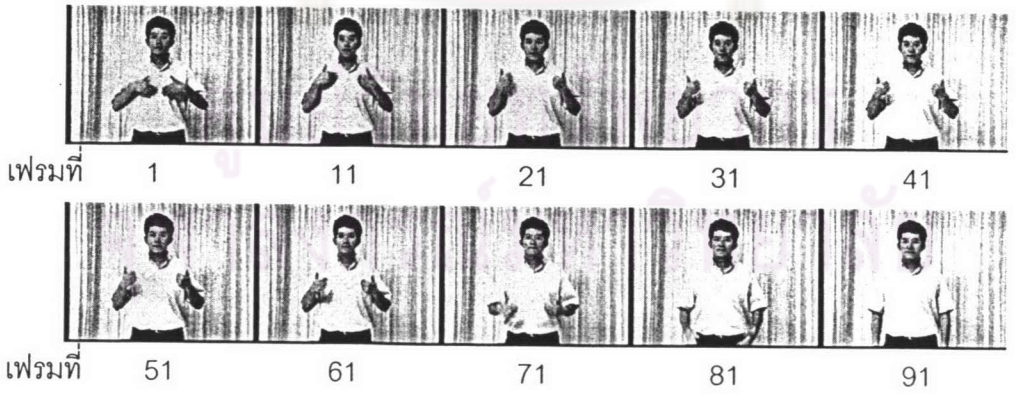
บทที่ 1

บทนำ

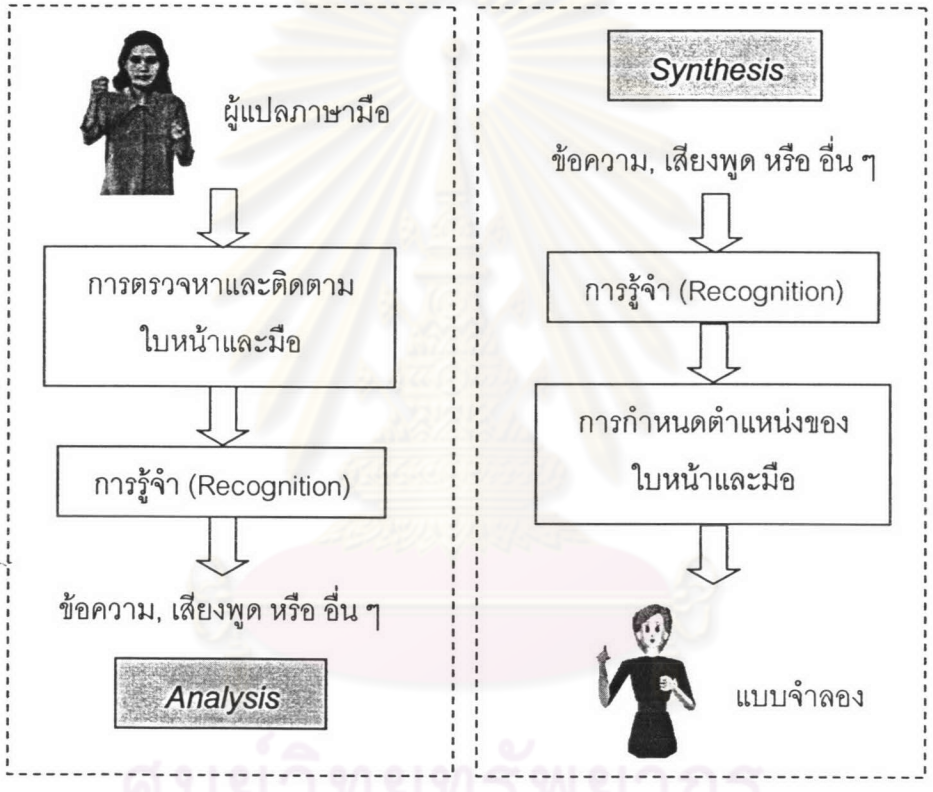
เนื่องจากปัญหาในการติดต่อสื่อสารระหว่างผู้ที่มีปัญหาด้านการได้ยินกับคนปกติ มีอุปสรรคบ่อยครั้ง สาเหตุเพราะคนปกติไม่สามารถเข้าใจภาษามือจากผู้ที่มีปัญหาด้านการได้ยินใช้สื่อสาร และผู้ที่มีปัญหาด้านการได้ยินไม่สามารถรับรู้เสียงพูดของคนปกติเช่นกัน ดังนั้น จึงต้องมีการสร้างระบบซึ่งเป็นตัวกลางในการแปลความหมายภาษามือเพื่อให้เกิดความเข้าใจที่ตรงกันแก่ทั้งสองฝ่าย ทั้งยังเป็นเครื่องมือที่ช่วยให้ผู้ที่มีปัญหาด้านการได้ยินมีโอกาสใช้ติดต่อสื่อสารกับคนปกติในสังคมภายนอกมากยิ่งขึ้น อันจะเป็นประโยชน์ต่อการสนองตอบความต้องการ หรือ นำแนวความคิดของพวกเขาเหล่านั้นมาร่วมกันพัฒนาประเทศชาติต่อไป

ในการสื่อสารโดยใช้ภาษามือของแต่ละประเทศ มีลักษณะแตกต่างกันเช่นเดียวกับภาษาพูด ด้วยเหตุนี้ระบบการแปลภาษามือที่มีในต่างประเทศ [1,2,3,4,5] จึงไม่สามารถนำมาประยุกต์ใช้เพื่อพัฒนาระบบการแปลภาษามือไทยได้ ฉะนั้น จึงมีความจำเป็นต้องสร้างระบบนี้ขึ้นเอง และจากการศึกษาเบื้องต้นเกี่ยวกับการสื่อสารโดยใช้ภาษามือ ผู้ที่มีปัญหาด้านการได้ยินนอกจากจะรับรู้ข่าวสารหรือข้อมูลของคู่สนทนาโดยการแสดงท่าทางของมือแล้ว ยังต้องอาศัยการสังเกตอากัปกิริยาของคู่สนทนาผ่านทางใบหน้าอีกด้วย ดังนั้นข้อมูลที่จำเป็นในการนำไปแปลความหมายมี 2 ส่วน คือ บริเวณใบหน้าและมือ



รูปที่ 1.1 ตัวอย่างวีดิทัศน์การใช้ภาษามือไทย

เมื่อพิจารณาการใช้เทคโนโลยีในปัจจุบัน ข้อมูลที่เป็นสื่อประสมมีความก้าวหน้าไปมากมาย และนิยมใช้กันแพร่หลาย อุปกรณ์ประเภทโทรศัพท์เคลื่อนที่หรือเครื่องคอมพิวเตอร์พกพาส่วนบุคคลก็จะมีกล้องวีดิทัศน์รวมอยู่ด้วยเป็นส่วนใหญ่ ข้อมูลที่เป็นสัญญาณวีดิทัศน์สามารถส่งไปบนโครงข่ายอินเทอร์เน็ตหรือเครือข่ายไร้สายได้อย่างง่ายดาย วิทยานิพนธ์นี้จึงมีแนวคิดที่จะสร้างกระบวนการก่อนหน้า (Pre-processing) ของระบบการแปลภาษามือไทย คือ กระบวนการตรวจหาและติดตามใบหน้าและมือมนุษย์ เพื่อประโยชน์ในการนำไปสร้างระบบแปลความหมายภาษามือไทยต่อไป



รูปที่ 1.2 ส่วนวิเคราะห์และสังเคราะห์ของระบบแปลภาษามือ

หากนำงานวิจัยเรื่องการตรวจหาและติดตามใบหน้าและมือไปพัฒนาต่อเนื่อง จะช่วยในการสร้างระบบแปลความหมายภาษามือไทยที่สมบูรณ์ ในที่นี้จะขอกล่าวถึงรายละเอียดของระบบที่สมบูรณ์โดยย่อ ดังแสดงในรูปที่ 1.2 โดยจะประกอบไปด้วย 2 ส่วนด้วยกัน คือ ส่วนแรกเป็นส่วนวิเคราะห์ ซึ่งเป็นการสื่อสารในทิศทางจากผู้ที่มีปัญหาด้านการได้ยินต้องการสื่อสารภาษามือให้คนปกติเข้าใจ กระบวนการหลัก คือ ทำการบันทึกการแสดงท่าทางภาษามือของผู้ที่มีปัญหาด้านการได้

ยินผ่านกล้องวีดิทัศน์ จากนั้นนำไปประมวลผลในระบบ (ซึ่งในระบบต้องมีส่วนการตรวจหาและติดตามใบหน้าและมือ) แล้วแสดงเป็นข้อความหรือเสียงพูดออกมาให้คนปกติสามารถเข้าใจได้ ส่วนที่สองเป็นส่วนสังเคราะห์ ซึ่งเป็นการสื่อสารในทิศทางจากคนปกติต้องการสื่อสารภาษามือกับผู้ที่มีปัญหาด้านการได้ยิน กระบวนการหลัก คือ คนปกติทำการพิมพ์หรือพูดข้อความที่ต้องการสื่อสาร แล้วให้ระบบทำการสังเคราะห์แบบจำลองมนุษย์ที่แสดงท่าทางภาษามือผ่านทางจอแสดงผล ให้ผู้ที่มีปัญหาด้านการได้ยินสามารถดูและเข้าใจได้

ประโยชน์ที่จะได้รับจากระบบแปลความหมายภาษามือไทยนั้นมีมากมาย ตัวอย่างเช่น ในชีวิตประจำวันหากผู้ที่มีปัญหาด้านการได้ยินต้องการขอความช่วยเหลือหรือสื่อสารกับคนปกติ ก็จะสามารถกระทำได้โดยอาศัยระบบแปลความหมายภาษามือไทยเป็นสื่อกลาง หรือแม้กระทั่งในด้านให้การศึกษาแก่ผู้ที่มีปัญหาด้านการได้ยิน ซึ่งบางวิชาต้องการผู้เชี่ยวชาญเฉพาะด้านไปถ่ายทอดความรู้ ซึ่งบุคคลเหล่านี้อาจไม่ทราบวิธีใช้ภาษามือ ระบบก็จะสามารถช่วยในการเรียนการสอนนั้นเป็นไปด้วยความราบรื่น เป็นต้น

1.1 ภาพรวมของงานวิจัยที่เกี่ยวข้อง

ในหัวข้อนี้จะกล่าวถึงภาพรวมของงานวิจัยที่เกี่ยวข้อง ซึ่งแบ่งออกเป็น 2 ส่วนด้วยกัน คือ

- งานวิจัยด้านการแยกสีผิวมนุษย์
- งานวิจัยด้านการติดตามใบหน้าและมือ

ส่วนแรกงานวิจัยด้านการแยกสีผิวมนุษย์ หากพิจารณาภาพรวมของงานวิจัยจะมีความแตกต่างใน 3 ส่วนหลักของกระบวนการ คือ

1. ปริภูมิสี ในแต่ละงานวิจัยจะมีการนำเสนอข้อได้เปรียบของแต่ละปริภูมิสีหลักที่มีอยู่แล้วและเหตุผลที่เลือกใช้ปริภูมิสีนั้น ๆ ปริภูมิที่ถูกเลือกมักเป็นปริภูมิสีที่มีการแยกกันระหว่างองค์ประกอบทางแสงและสี โดยจะคำนึงถึงแต่ค่าสีอย่างเดียว เพื่อมิให้เกิดความผิดพลาดในการตัดสิน เพราะขณะกำลังแสดงท่าทางจะต้องมีการเคลื่อนไหวใบหน้าและมือ ทำให้ค่าแสงที่ตกกระทบบนใบหน้าและมือมีการเปลี่ยนแปลงอยู่ตลอดเวลา นอกจากนี้ยังได้มีงานวิจัยประเภทเปรียบเทียบประสิทธิภาพเมื่อใช้ปริภูมิสีต่างกันในการแยกสีผิว และบางส่วนก็ได้เสนอปริภูมิสีใหม่ที่นำมาใช้ในการแยกสีผิว แต่จากการศึกษาพบว่าปริภูมิสีที่มักนำมาใช้สำหรับการแยกสีผิว มีอยู่ 3 ปริภูมิสี คือ นอร์แมลไลซ์ RGB [6,7], HSI [3,8] และ YCbCr [3,7,9,10,11,12,13]

2. แบบจำลองที่นำมาใช้ในการแยกสีผิวซึ่งแบ่งได้เป็น 2 แนวทาง คือ

- พิจารณาการเกาะกลุ่มกันบนระนาบ ซึ่งสามารถกำหนดบริเวณครอบคลุมส่วนเหล่านี้ได้โดยใช้ พิจารณาฮิสโทแกรมในแต่ละองค์ประกอบ (รูปสี่เหลี่ยม) [10,13] เส้นตรงหลายเส้นล้อมรอบ [9] พาราโบลา 2 เส้น [6] วงรี [12] ฯลฯ

- วิธีทางสถิติ กล่าวคือ ทำการเก็บข้อมูลจุดภาพที่เป็นสีผิวทางสถิติแล้วสร้างแบบจำลองขึ้น เช่น แบบเกาส์ [12] แบบเกาส์ผสม [8] ฯลฯ เพื่อนำแบบจำลองนี้ไปใช้ในการกำหนดรูปภาพรับเข้า (input) ว่าจุดภาพนั้น ๆ เป็นสีผิวหรือไม่

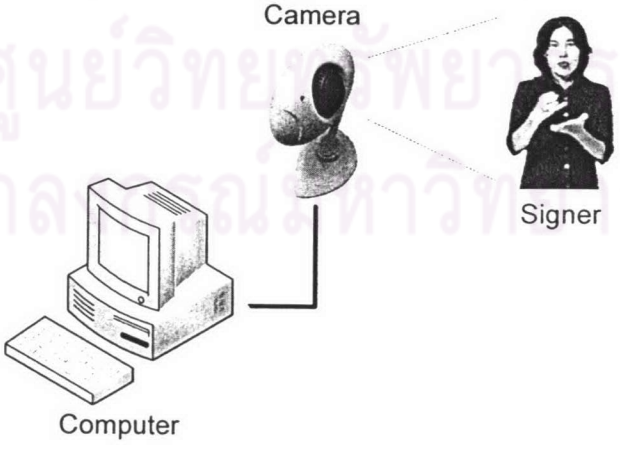
3. การนำไปประยุกต์ใช้งานที่ต่างกัน ตัวอย่างเช่น นำไปประยุกต์ใช้ในการตรวจหาใบหน้าหรือจำแนกใบหน้า [6,7,9,12,13,14,16] ระบบรักษาความปลอดภัยเมื่อมีผู้ไม่หวังดีเข้ามายังบริเวณต้องห้ามก็จะทำการเตือนภัย การติดตามมือเพื่อช่วยทำงานแทนมนุษย์ [15] เช่น การนำเครื่องบินลงจอด การพิมพ์คีย์บอร์ด การแปลความหมายภาษามือ การตีกลอง ฯลฯ

รูปที่ 1.3 กระบวนการในการแยกส่วนสีผิว

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ส่วนที่สอง งานวิจัยด้านการติดตามใบหน้าและมือ เพื่อใช้สำหรับแปลความหมายภาษามือ พิจารณางานวิจัยโดยภาพรวมของส่วนที่สอง ระบบแปลความหมายภาษามือที่พัฒนา ณ เวลานี้ จะแบ่งได้เป็น 2 ประเภท ประเภทแรกคือ ถุงมือไซเบอร์ (Cyber glove) แสดงตัวอย่างในรูปที่ 1.4 เป็นถุงมือชนิดพิเศษที่ฝังตัวรับรู้ (sensor) ไว้ สามารถอ้างอิงตำแหน่งของมือและนิ้วได้อย่างแม่นยำ เมื่อนำมาเชื่อมต่อกับคอมพิวเตอร์ ข้อมูลตำแหน่งที่ได้เหล่านั้นจะถูกนำมาใช้ในการแปลความหมายต่อไป แต่มีข้อเสียคืออุปกรณ์ประเภทนี้มีราคาแพงอีกทั้งยังไม่สะดวกต่อการพกพา สำหรับประเภทที่สองคือแบบอาศัยพื้นฐานการมองเห็น (Vision based) อุปกรณ์ที่ใช้ในระบบมีเพียงกล้องวีดิทัศน์เชื่อมต่อกับคอมพิวเตอร์ ดังแสดงในรูปที่ 1.5 ซึ่งเป็นประเภทที่วิทยานิพนธ์ฉบับนี้เลือกนำมาพัฒนา เพราะเป็นประเภทที่ใช้ต้นทุนต่ำ และ มีความยืดหยุ่นต่อการนำไปใช้งานในชีวิตประจำวันได้มากกว่า หากแต่ต้องการความสามารถของขั้นตอนวิธีในการหาตำแหน่งของใบหน้าและมือที่เหมาะสม เพื่อให้ตำแหน่งที่ได้มีความถูกต้องแม่นยำเทียบเท่าหรือดีกว่าประเภทแรก

รูปที่ 1.4 ตัวอย่างของถุงมือไซเบอร์



รูปที่ 1.5 โครงสร้างของระบบแปลภาษามือที่อาศัยพื้นฐานการมองเห็น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.2 แนวทางของวิทยานิพนธ์

วิทยานิพนธ์นี้ เสนอการตรวจหาและติดตามใบหน้าและมือมนุษย์สำหรับวีดิทัศน์ภาษามือ ในส่วนของการตรวจหา ขั้นตอนแรกทำการคัดเลือกปริภูมิสีที่เหมาะสมในการแยกส่วนสีผิว เมื่อพิจารณางานวิจัยแล้ว จึงเสนอการเปรียบเทียบโดยใช้แบบจำลองชนิดไม่มีตัวแปร (Non-parametric model) เนื่องจากเป็นแบบจำลองที่ให้ความยุติธรรมกับทุกปริภูมิ ทำการทดสอบใน 3 ปริภูมิสีย่อยที่มีความนิยม คือ นอร์แมลไลซ์ RG, HS และ CbCr หลังจากได้ปริภูมิสีย่อยที่เหมาะสมแล้ว จะสร้างแบบจำลองของสีผิวเพื่อใช้ในระบบจริง โดยแบบจำลองที่สร้างขึ้นนี้ เป็นแบบจำลองวงรีบนปริภูมิสีย่อย CbCr ซึ่งได้แสดงกรรมวิธีการครอบคลุมข้อมูลของการกระจายตัวบริเวณส่วนสีผิวเพื่อสร้างวงรีดังกล่าวด้วย ทำให้สามารถนำไปประยุกต์ใช้งานร่วมกับระบบที่ปรับเปลี่ยนแบบจำลองสีผิวขณะกำลังทำงานได้ กรณีที่มีเงื่อนไขยุ่งยากมากขึ้น เช่น ความซับซ้อนของฉากหลัง การใช้งานกับสภาพแวดล้อมภายนอก เป็นต้น นอกจากนี้ยังสามารถจำแนกว่าบริเวณใดเป็นใบหน้าหรือมือในเชิงพื้นที่ (Spatial domain) ออกมาก่อนได้ในเบื้องต้น และส่วนการติดตาม จะอาศัยสมมติฐานที่ว่าในวีดิทัศน์เฟรมที่อยู่ติดกัน มีการเปลี่ยนตำแหน่งของใบหน้าน้อยกว่ามือมาก จึงกำหนดแบบจำลองในการประมาณการเคลื่อนที่แยกกันระหว่างบริเวณใบหน้าและมือ ทำการวิเคราะห์หาแบบจำลองในการประมาณการเคลื่อนที่ให้เหมาะสมกับแต่ละส่วน การปรับปรุงประสิทธิภาพของการติดตามให้ดีขึ้น เพื่อให้ค่าความผิดพลาดออกมาน้อยที่สุด โดยมีเงื่อนไขว่าระบบต้องมีการคำนวณที่ไม่ซับซ้อนและสามารถนำไปใช้งานในเวลาจริง (Real time) ได้ นอกจากนี้ยังได้ศึกษาแนวทางการแก้ปัญหาเรื่องการผสานและแยกกันของแต่ละบริเวณ เพื่อประโยชน์เกี่ยวกับงานที่จะพัฒนาต่อไปในอนาคต

1.3 เป้าหมายและขอบเขตของวิทยานิพนธ์

1. พัฒนาขั้นตอนวิธีที่ใช้ในการตรวจหาและติดตามใบหน้าและมือของมนุษย์
2. แบบจำลองสีผิวที่สร้างขึ้นจะใช้ได้เฉพาะวีดิทัศน์ที่ทดสอบเท่านั้น
3. ขั้นตอนวิธีที่นำเสนอใช้ได้กับปัญหาที่มีผู้แสดงภาษามือเพียงคนเดียวอยู่ภายในเฟรม
4. ศึกษาแนวทางแก้ปัญหาการผสานและแยกกันของวัตถุ

1.4 ขั้นตอนและวิธีการดำเนินงาน

1. ศึกษาขั้นตอนวิธี และแบบจำลองที่ใช้ในการตรวจหาบริเวณสีผิวของมนุษย์
2. ศึกษาขั้นตอนวิธีต่าง ๆ ที่ใช้ในการตรวจหาและติดตามวัตถุ
3. ศึกษาคุณลักษณะที่เหมาะสมของวัตถุ (ใบหน้าและมือ) ที่จะทำการตรวจหาและติดตาม
4. นำขั้นตอนวิธีและเลือกคุณลักษณะที่เหมาะสมมาใช้ในการติดตามใบหน้าและมือสำหรับวีดิทัศน์ภาษามือไทย
5. ทดลองขั้นตอนวิธี แบบจำลอง และคุณลักษณะที่เลือกใช้ในการติดตามใบหน้าและมือสำหรับวีดิทัศน์ที่เป็นภาษามือไทย พร้อมทั้งวิเคราะห์ผล
6. สรุปผลการทดลองและเขียนวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ทราบถึงความรู้พื้นฐานเกี่ยวกับการตรวจหา และติดตามวัตถุ
2. สร้างแบบจำลองที่เหมาะสมในการตรวจหาบริเวณสีผิวของมนุษย์
3. พัฒนาขั้นตอนวิธีที่สำหรับการติดตามใบหน้าและมือ เพื่อนำไปสร้างระบบการแปลความหมายภาษามือไทย

1.6 เค้าโครงวิทยานิพนธ์

สำหรับเนื้อหาของวิทยานิพนธ์ฉบับนี้แบ่งรายละเอียดออกเป็น 5 บทดังต่อไปนี้

บทที่ 1 บทนำ ซึ่งได้กล่าวถึง แนวเหตุผล ภาพรวมของงานวิจัยที่เกี่ยวข้อง แนวทางของวิทยานิพนธ์ เป้าหมายและขอบเขตของวิทยานิพนธ์ ขั้นตอนและวิธีการดำเนินงาน ประโยชน์ที่คาดว่าจะได้รับ และเค้าโครงวิทยานิพนธ์

บทที่ 2 ทฤษฎีพื้นฐานและงานวิจัยที่เกี่ยวข้อง เนื้อหาในบทนี้จะกล่าวถึง ปริภูมิสีต่าง ๆ และงานวิจัยที่เกี่ยวข้อง การประมวลผลภาพดิจิทัลที่นำมาใช้ในงานวิจัย (ประกอบไปด้วย การหาความเยื้องศูนย์กลาง (Eccentricity) การวิเคราะห์ส่วนประกอบที่เชื่อมต่อกัน (Connected Component Analysis) การประมวลผลภาพเชิงสัณฐาน (Morphological Image Processing)) การตรวจหาขอบ

(Edge Detection) การวัดประสิทธิภาพของแบบจำลอง การเคลื่อนที่ของวัตถุในแนวเส้นตรง งานวิจัยที่เกี่ยวข้องด้านการติดตามและภาษามือ และ แนวทางแก้ปัญหาการผสานและแยกกันของวัตถุ

บทที่ 3 กรรมวิธีที่นำเสนอ เนื้อหาในบทนี้กล่าวถึง โครงสร้างของระบบตรวจหาและติดตามใบหน้าและมือ ซึ่งในรายละเอียดประกอบไปด้วยสองส่วนหลัก คือ ส่วนเริ่มต้น (การตรวจหา) จะสร้างแบบจำลองที่ใช้ในการแยกส่วนสีผิวมนุษย์ การเติมเต็มในส่วนที่ขาดหายและกำจัดหลุมขนาดเล็กของมาสก์ด้วยตัวดำเนินการแบบปิด การหาตำแหน่ง (Labeling and Localization) การแยกส่วนสำคัญต่าง ๆ และสำหรับในส่วนการติดตาม (Tracking) ซึ่งเสนอให้แยกการติดตามระหว่างใบหน้าและมือ (แบบจำลองที่ใช้ต่างกัน) รวมถึงเสนอกรรมวิธีในการคำนวณซ้ำเพื่อให้ตำแหน่งที่ได้มีความถูกต้องมากยิ่งขึ้น

บทที่ 4 ผลการทดลอง ซึ่งประกอบไปด้วย ผลการแยกส่วนสีผิวโดยใช้แบบจำลองชนิดไม่มีตัวแปร (Non-parametric model) เปรียบเทียบใน 3 ปริภูมิสีย่อย คือ rg, HS และ CbCr ผลการแยกส่วนสีผิวในปริภูมิสีย่อย CbCr โดยแบบจำลองค่าสูงสุดต่ำสุดแต่ละองค์ประกอบ [13] เปรียบเทียบกับการแยกด้วยแบบจำลองวงรี (Elliptical model) ที่นำเสนอ การเรียนรู้สีผิวจากกล้องโซนี่ (Sony) รุ่น EVI-D100P ผลจากตัวดำเนินการแบบปิด ผลการแยกองค์ประกอบสำคัญต่าง ๆ ที่สำคัญในร่างกายมนุษย์ อาทิ เช่น ใบหน้า คอ ตา คิ้ว จมูก ปาก และ ข้อต่อระหว่างมือกับแขน เป็นต้น ผลการวิเคราะห์หาบริเวณที่เชื่อมต่อกันรวมทั้งหาจุดศูนย์ถ่วงของแต่ละวัตถุ การกำหนดบล็อบใบหน้าและมือเชิงพื้นที่โดตัดสินจากขนาดและวัดความเป็นวงกลม ตำแหน่งของใบหน้าและมือในแต่ละเฟรม เปรียบเทียบความผิดพลาดของตำแหน่งมือทั้งสองเมื่อติดตามด้วยแบบจำลองดั้งเดิม [15] กับขั้นตอนวิธีที่นำเสนอ และข้อดีข้อด้อยของแต่ละวิธี แนวทางแก้ปัญหาการผสานและแยกกันระหว่างใบหน้าและมือ

บทที่ 5 บทสรุป ซึ่งเป็นการสรุปผลการวิจัย และ ข้อเสนอแนะเพิ่มเติมสำหรับงานวิจัยในอนาคต

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย