

การตรวจจับและสกัดคุณลักษณะสำคัญของมือจากวีดิทัศน์สำหรับการสะกดนิ้วมือภาษาไทย

นายธนเดช สุขศิลป์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2555

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)

เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR) are the thesis authors' files submitted through the Graduate School.

HAND DETECTION AND FEATURE EXTRACTION FROM VIDEO FOR THAI
FINGER-SPELLING SIGN LANGUAGE

Mr. Thanadej Suksil

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2012

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การตรวจจับและสกัดคุณลักษณะสำคัญของมีเอจาก วิถีทัศน์สำหรับการสะกดนิ้วมือภาษาไทย
โดย	นายธนเดช สุขศิลป์
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.ธนรัตน์ ชลิดาพงศ์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยรับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
()

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สืบสกุล พิภพมงคล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.ธนรัตน์ ชลิดาพงศ์)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สิ้นธุฎิญา)

..... กรรมการภายนอกมหาวิทยาลัย
(ดร.ศุภกร สิทธิไชย)

ธนเดช สุขศิลา : การตรวจจับและสกัดคุณลักษณะสำคัญของมือจากวีดิทัศน์สำหรับการ
การสะกดนิ้วมือภาษาไทย. (HAND DETECTION AND FEATURE EXTRACTION FOR
THAI FINGER-SPELLING SIGN LANGUAGE) อ. ที่ปรึกษาวิทยานิพนธ์หลัก : ผศ. ดร.
ธนารัตน์ ชลิตาพงศ์, 51 หน้า.

วิทยานิพนธ์ฉบับนี้ได้นำเสนอกรอบงานทดลองจนขั้นตอนวิธีการตรวจจับและสกัด
คุณลักษณะสำคัญของมือ สำหรับระบบรู้จำภาษามือไทย โดยมีข้อมูลนำเข้าเป็นวีดิทัศน์ล่าม
ภาษามือไทย ระบบต้นแบบที่ได้พัฒนาสามารถทำงานได้อย่างอัตโนมัติและมีประสิทธิภาพ
เพียงพอต่อการนำไปประยุกต์ใช้แบบทันที

กรอบงานที่นำเสนอประกอบไปด้วย 6 ขั้นตอน คือ การสกัดบริเวณส่วนที่เป็นสีของผิวหนัง
การติดตามการเคลื่อนไหวของมือ การแยกสถานะของการแสดงท่ามือ การหาคีย์เฟรมของท่ามือ
การสกัดคุณลักษณะสำคัญของรูปร่าง และการนำคุณลักษณะสำคัญของรูปร่างมือที่สกัดได้มาทำ
การรู้จำท่ามือ ซึ่งขั้นตอนของการสกัดบริเวณส่วนที่เป็นสีของผิวหนังจะเป็นขั้นตอนในการค้นหา
บริเวณส่วนที่เป็นมือและใบหน้าของล่ามผู้แสดงภาษามือเพื่อกำหนดเป็นค่าเริ่มต้นในการติดตาม
ตำแหน่งของมือในขั้นตอนของการติดตามการเคลื่อนไหวของมือ หลังจากทำการติดตามตำแหน่ง
ของมือแล้วจึงทำการแยกแยะสถานะของการแสดงท่ามือโดยพิจารณาความเร็วในการเคลื่อนที่
ของมือ และพิจารณาอัตราการเปลี่ยนแปลงของรูปร่างมือเพื่อใช้ในการสกัดคีย์เฟรมของท่ามือ
และจากคีย์เฟรมที่สกัดได้จึงพิจารณารูปร่างของมือและสกัดคุณลักษณะสำคัญของมือ ได้แก่
จุดสำคัญของรูปร่าง ค่าฮูโมเมนต์ (Hu Moment) ของรูปร่างมือ ค่าความหนาแน่นของขอบภาพ
(Edge density) และตำแหน่งของส่วนภาพมือ เพื่อใช้ในขั้นตอนของการรู้จำท่ามือ

ผลการทดลองพบว่ากรอบงานที่นำเสนอสำหรับตรวจจับและสกัดคุณลักษณะสำคัญของ
มือสำหรับระบบรู้จำภาษามือไทยนั้นสามารถทำงานแบบทันที ภายใต้สภาวะแวดล้อมที่กำหนด
โดยมีความแม่นยำอยู่ที่ 77.78 %

ภาควิชา...วิศวกรรมคอมพิวเตอร์....	ลายมือชื่อ.....
สาขาวิชา...วิศวกรรมคอมพิวเตอร์....	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....
ปีการศึกษา.....2555.....	

5370643721 : MAJOR COMPUTER ENGINEERING

KEYWORDS : HAND FEATURE EXTRACTION/SKIN COLOR

SEGMENTATION/ HAND TRACKING/HAND POSTURE RECOGNITION

THANADEJ SUKSIL : HAND DETECTION AND FEATURE EXTRACTION FROM
VIDEO FOR THAI FINGER-SPELLING SIGN LANGUAGE . ADVISOR :
ASST.PROF. THANARAT H. CHALIDABHONGSE, Ph.D., 51 pp.

This thesis proposes framework for hand detection and feature extraction from videos of Thai finger-spelling Sign Language. The system prototype was developed and shows that it works well and efficient enough for real-time applications.

The proposed framework consists of 6 steps, skin color segmentation, hand tracking, state determination, key frame extraction, hand feature extraction, and recognition. In skin color segmentation step, we segment skin color from input image to obtain face and hand regions and use face and hand region position for tracking in tracking step. In state determination step, we calculate hand motion from centroids of hand regions. Hand motions are classified into 3 states including Neutral state, Transition state and Sign state. After extracting Sign state, key frames will be extracted by considering shape difference from the previous frames. After getting sign key frame, we then find some salient convex hull points that from the hand shape. Then, we use these points to recognize the hand gestures.

The experimental results show that the system is able to detect, track, and extract hand features for recognition in real-time with 77.78% accuracy.

Department :...Computer Engineering... Student's Signature.....

Field of Study :...Computer Engineering... Advisor's Signature.....

Academic Year :.....2012.....

กิตติกรรมประกาศ

ความสำเร็จสมบูรณ์ของวิทยานิพนธ์ฉบับนี้เกิดขึ้นได้ด้วยคำชี้แนะแนวทางการศึกษาและวิจัยอย่างทุ่มเทของ ผศ.ดร.ธนรัตน์ ชลิตาพงศ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ จึงขอขอบพระคุณอย่างสูงไว้ ณ ที่นี้

ขอขอบคุณคณาจารย์ เจ้าหน้าที่ภาควิชาฯ เจ้าหน้าที่ห้องสมุด ตลอดจนเพื่อนๆ ทั้งหลักสูตรในเวลาและภาคนอกเวลาราชการ ที่คอยสนับสนุนการทำงาน ให้ความเห็นและกำลังใจในการทำวิจัย รวมทั้ง ขอขอบพระคุณ บิดา มารดา ผู้ให้การศึกษาและเลี้ยงดูเป็นอย่างดีเสมอมา

สารบัญ

หน้า

บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ	ช
สารบัญตาราง.....	ฅ
สารบัญรูป	ญ
บทที่	
1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย	4
1.3 ขอบเขตงานวิจัย	4
1.4 วิธีการดำเนินงานวิจัย	4
1.5 ประโยชน์ที่คาดว่าจะได้รับ	4
1.6 ผลงานที่ตีพิมพ์จากวิทยานิพนธ์.....	5
2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	6
2.1 ทฤษฎีที่เกี่ยวข้อง.....	6
2.1.1 คุณลักษณะสำคัญท้องถิ่น (Local Feature)	6
2.1.2 แบบจำลองมือ (Hand Model).....	7
2.1.3 การแปลงปริภูมิสี	10
2.1.4 โมเมนต์ของภาพ (Image Moment).....	10
2.1.5 เปลือกหุ้มวัตถุ (Convex hull)	13
2.1.6 การสร้างวงรีแทนกลุ่มของข้อมูล	16
2.2 งานวิจัยที่เกี่ยวข้อง	18
2.2.1 การตรวจจับและติดตามการเคลื่อนไหวของมือ (Hand detection and hand tracking).....	18
2.2.2 การแก้ไขปัญหาการบังของวัตถุบนภาพ.....	21

2.2.3 การสกัดคุณลักษณะสำคัญของรูปร่างวัตถุบนภาพ (Shape feature extraction)	24
2.2.4 ระบบรู้จำภาษามือ (Sign language recognition system)	25
3 ขั้นตอนวิธีที่เสนอ.....	29
3.1 สกัดบริเวณส่วนที่เป็นสีผิว.....	30
3.2 ติดตามการเคลื่อนไหวของมือ.....	31
3.3 แยกสถานะของการแสดงท่ามือ	32
3.4 หาเคีย์เฟรมของท่ามือ.....	34
3.5 จุดยอดของรูปร่างเพื่อใช้ในการจำแนกท่ามือ	35
3.6 เปรียบเทียบรูปร่างของมือกับข้อมูลท่ามือ	37
4 การทดลองและผลการทดลอง	39
4.1 ทดสอบการสกัดบริเวณส่วนที่เป็นสีของผิวหนึ่งบนภาพ.....	39
4.2 การทดสอบการติดตามการเคลื่อนไหวของมือบนภาพจากวีดิทัศน์	40
4.3 ทดสอบการสกัดคุณลักษณะสำคัญของรูปร่าง และท่าทางของมือ	41
4.4 ทดสอบคุณลักษณะสำคัญของมือที่สกัดได้กับการรู้จำรูปร่างท่ามือ.....	42
5 สรุปผลการวิจัยและข้อเสนอแนะ.....	46
5.1 สรุปผลการวิจัย	46
5.2 ข้อเสนอแนะ	47
รายการอ้างอิง.....	49
ประวัติผู้เขียนวิทยานิพนธ์	51

สารบัญตาราง

ตารางที่		หน้า
3-1	ตารางจำแนกกลุ่มของท่ามือเดียว	37
4-1	เมตริกซ์ความสับสน (Confusion Matrix) ของข้อมูลท่ามือกลุ่มที่ 1	42
4-2	เมตริกซ์ความสับสน (Confusion Matrix) ของข้อมูลท่ามือกลุ่มที่ 2	43
4-3	เมตริกซ์ความสับสน (Confusion Matrix) ของข้อมูลท่ามือกลุ่มที่ 3	43
4-4	เมตริกซ์ความสับสน (Confusion Matrix) ของข้อมูลท่ามือกลุ่มที่ 4	44
4-5	เมตริกซ์ความสับสน (Confusion Matrix) ของข้อมูลท่ามือกลุ่มที่ 5	44
4-6	เมตริกซ์ความสับสน (Confusion Matrix) ของข้อมูลท่ามือกลุ่มที่ 6	44
4-7	ตารางทดสอบความแม่นยำของระบบ.....	45

สารบัญรูป

รูปที่		หน้า
1-1	คุณลักษณะสำคัญของมือโดยใช้โมเดลของมือ	2
1-2	สำคัญของมือโดยใช้สิ่งที่ปรากฏบนภาพ	2
1-3	ตัวสะกดนิ้วมือภาษาอเมริกันและตัวสะกดนิ้วมือภาษาไทย.....	3
2-1	คุณลักษณะสำคัญท้องถิ่นของรูปร่างวัตถุ	6
2-2	แสดงแบบจำลองของมือที่ใช้ในการพิจารณาการเคลื่อนไหว	8
2-3	ค่าฮูโมเมนต์ (Hu moment) ของแต่ละตัวอักษรที่มีรูปร่างแตกต่างกัน	13
2-4	จุดยอดของรูปร่างมือ	13
2-5	จุดข้อมูลรูปร่างวัตถุ	14
2-6	การกำหนดหมายเลขให้กับจุดข้อมูลรูปร่างวัตถุแบบตามเข็มนาฬิกา	14
2-7	การกำหนดค่าจุดข้อมูลรูปร่างวัตถุด้วยวิธีสามเหลี่ยม	14
2-8	ขั้นตอนวิธีสามเหลี่ยม(1).....	15
2-9	ขั้นตอนวิธีสามเหลี่ยม(2).....	15
2-10	ผลลัพธ์ของขั้นตอนวิธีสามเหลี่ยม	16
2-11	การกระจายตัวของกลุ่มข้อมูลสี่ของนิ้ว.....	19
2-12	ขั้นตอนการแยกส่วนที่เป็นมือและใบหน้าออกจากส่วนอื่นของภาพ.....	19
2-13	ผลลัพธ์ของการสกัดสี่นิ้ว	20
2-14	วงรีสมมติฐานของบริเวณส่วนที่เป็นนิ้วหนึ่ง	20
2-15	ผลลัพธ์ของการติดตามมือและใบหน้า	21
2-16	ขั้นตอนการแยกจุดภาพส่วนที่เป็นมือและใบหน้าเมื่อเกิดการบัง	22
2-17	ผลลัพธ์แยกจุดภาพบริเวณส่วนที่เป็นมือและใบหน้าเมื่อเกิดการบัง	22
2-18	การเปลี่ยนภาพรับเข้าเป็นภาพศักยะและพลังงาน	23
2-19	ภาพเปรียบเทียบการเปลี่ยนแปลงของพลังงานของภาพก่อนและหลังการถูกบัง	23
2-20	ภาพผลลัพธ์ของการแยกบริเวณที่เกิดการบังบนภาพ.....	23
2-21	ภาพการทดสอบการสกัดฮูโมเมนต์ในระยะและการหมุนที่แตกต่างกัน	24
2-22	ภาพโครงสร้างมือของการสกัดนิ้วมือ	25
2-23	ผลการตรวจจับนิ้วมือในการแสดงท่ามือ	25

รูปที่	หน้า
2-24	การกำหนดค่าเริ่มต้นของการติดตาม..... 26
2-25	การใช้แม่แบบในการแยกท่ามือที่มีการบดบัง 27
2-26	การใช้แม่แบบของส่วนของวงกลมสำหรับหาตำแหน่งข้อศอก..... 27
2-27	การรู้จำท่ามือ..... 28
2-28	การใช้เวกเตอร์คุณลักษณะสำคัญทางสถิติและเรขาคณิตในการรู้จำ 28
3-1	ขั้นตอนการทำงานของการทำงานการตรวจจับและสกัดคุณลักษณะสำคัญของมือสำหรับระบบ รู้จำภาษามือไทย 29
3-2	การค้นหาส่วนที่เป็นมือบนภาพจากกล้องวีดิทัศน์ 31
3-3	การหาตำแหน่งเริ่มต้นการติดตามการเคลื่อนไหวของใบหน้าและมือ 31
3-4	ติดตามการเคลื่อนไหวของใบหน้าและมือโดยใช้สมมติฐานของวัตถุ 32
3-5	การหาจุดศูนย์กลางมือ 33
3-6	ลำดับการเคลื่อนไหวของมือในการแสดงท่ามือ 33
3-7	ความเร็วในการเคลื่อนที่ของมือ..... 34
3-8	คีย์เฟรมของท่ามือ 34
3-9	จุดยอดของรูปร่างมือ 35
3-10	สกัดจุดยอดของรูปร่างมือสำคัญที่ใช้ในการจำแนกกลุ่มของท่ามือ..... 36
3-11	ท่ามือเดียวในท่ามือสะกดไทย..... 36
3-12	ความหนาแน่นของขอบภาพ 37
3-13	การจำแนกท่ามือภายในกลุ่มท่ามือ 38
4-1	ผลการสกัดสีผิว 39
4-2	ผลการติดตามการเคลื่อนไหวของมือจากวีดิทัศน์..... 40
4-3	ผลการติดตามการเคลื่อนไหวของมือจากวีดิทัศน์ในกรณีที่ไม่เกิดการบดบังกัน 41
4-4	ผลการติดตามการเคลื่อนไหวของมือจากวีดิทัศน์ในกรณีที่เกิดการบดบังกัน..... 41
4-5	ผลการสกัดคุณลักษณะสำคัญของมือจากวีดิทัศน์ 41
4-6	ท่ามือที่ใช้ในการทดสอบ 42
4-7	ผลการรู้จำจากคุณลักษณะสำคัญของมือที่ได้จากวีดิทัศน์..... 45
5-1	ผลการเปรียบเทียบแม่แบบ (Template Matching) ที่ไม่ถูกต้อง 47

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ระบบรู้จำภาษามืออาศัยการรับภาพผู้แสดงท่ามือจากกล้องวิดีโอที่ศน์ของผู้แสดงท่ามือ เพื่อทำการตรวจจับการเคลื่อนไหวของร่างกายสำหรับใช้ในการรู้จำความหมายของการแสดงท่าทาง โดยการรู้จำภาษามือจะพิจารณาจากการเคลื่อนไหวของมือ ใบหน้า และลำตัวส่วนบน ดังนั้นการสกัดคุณลักษณะสำคัญจากภาพของการเคลื่อนไหวของใบหน้าและมือจึงเป็นสิ่งจำเป็น แต่เนื่องจากภาพที่รับจากกล้องวิดีโอที่ศน์นั้นเป็นภาพสองมิติ ในขณะที่การเคลื่อนไหวของมนุษย์นั้นอยู่ในโลกสามมิติ อีกทั้งยังมีข้อต่อของนิ้วมือ (Finger joint) ที่มีจำนวนองศาอิสระ (Degree of freedom) มากถึง 23 องศาอิสระ (Kuch และ J.J. (1994) [1]) ทำให้มือนั้นสามารถเปลี่ยนแปลงรูปร่างไปตลอดเวลาที่มีการเคลื่อนไหว นอกจากนี้ยังมีปัญหาในเรื่องของการบดบัง (Occlusion) ทั้งในส่วนข้อมือบดบังตัวเอง เช่น การหมุนข้อมือทำให้นิ้วมือถูกบดบัง หรือ แม้กระทั่งในกรณีที่มีมือขวามือซ้าย ด้วยปัญหาดังกล่าวทำให้การสกัดคุณลักษณะสำคัญของมือนั้นเป็นสิ่งที่ยากที่จะทำการสกัดโครงสร้างรูปร่างมือ ณ ขณะนั้นออกมาได้ทั้งหมด จึงจำเป็นที่จะต้องเลือกสกัดคุณลักษณะสำคัญของมือเพียงบางส่วนที่เพียงพอสำหรับการตีความหมายของท่ามือนั้น การสกัดคุณลักษณะสำคัญของมือนั้นสามารถแบ่งออกได้เป็นสองวิธีการหลักคือ ใช้โมเดลรูปร่างข้อมือเป็นหลัก (Model base) และ ใช้สิ่งที่ปรากฏบนภาพเป็นหลัก (Appearance base) ซึ่งแต่ละวิธีก็จะมีข้อดี และข้อเสียที่แตกต่างกันออกไป

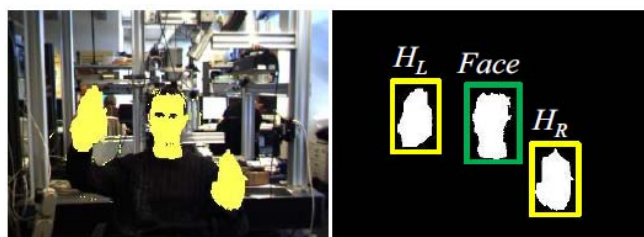
การสกัดคุณลักษณะสำคัญของมือโดยใช้โมเดลรูปร่างข้อมือเป็นหลัก (Model base) เป็นวิธีที่จะสกัดรูปร่างโครงสร้างทั้งหมดของมือจากภาพ ทำให้ได้รูปร่างโครงสร้างข้อมือทั้งหมด แต่เนื่องจากมือนั้นเป็นวัตถุสามมิติดังนั้นการที่จะสกัดโครงสร้างข้อมือทั้งหมดนั้น จำเป็นที่จะต้องมองเห็นรูปร่างข้อมือทั้งหมด เพื่อไม่ให้เกิดปัญหาการบดบัง ซึ่งจำเป็นที่จะต้องอาศัยกล้องวิดีโอที่ศน์หลายตัวจากหลายมุม เพื่อที่จะสามารถมองเห็นและสกัดโครงสร้างข้อมือได้ในทุกมุมมอง แล้วนำมุมมองของภาพที่ได้เหล่านั้นมาสร้างเป็นโมเดลสามมิติ ซึ่งวิธีนี้จำเป็นที่จะต้องอาศัยกล้องเป็นจำนวนหลายตัว อีกทั้งต้องวางในตำแหน่งที่เหมาะสม เพื่อที่จะให้เกิดปัญหาการบดบังน้อยที่สุด ซึ่งเป็นไปได้ยากสำหรับการเคลื่อนไหวท่าทางภาษามือ ที่มีการเคลื่อนไหวข้อมือที่ซับซ้อนและไม่แน่นอน ทำให้เกิดการบดบังอยู่ตลอดเวลา อีกทั้งการนำโครงสร้างข้อมือมาสร้างเป็นโมเดลสามมิตินั้นต้องใช้เวลาคำนวณสูงมาก อีกทั้งยังต้องมีการกำหนดค่าเริ่มต้นให้กับโมเดลของ

มือ (Model calibration) จึงทำให้กระบวนการทำงานช้ามาก ไม่เหมาะสมกับระบบแบบทันทีกาล (Real time)



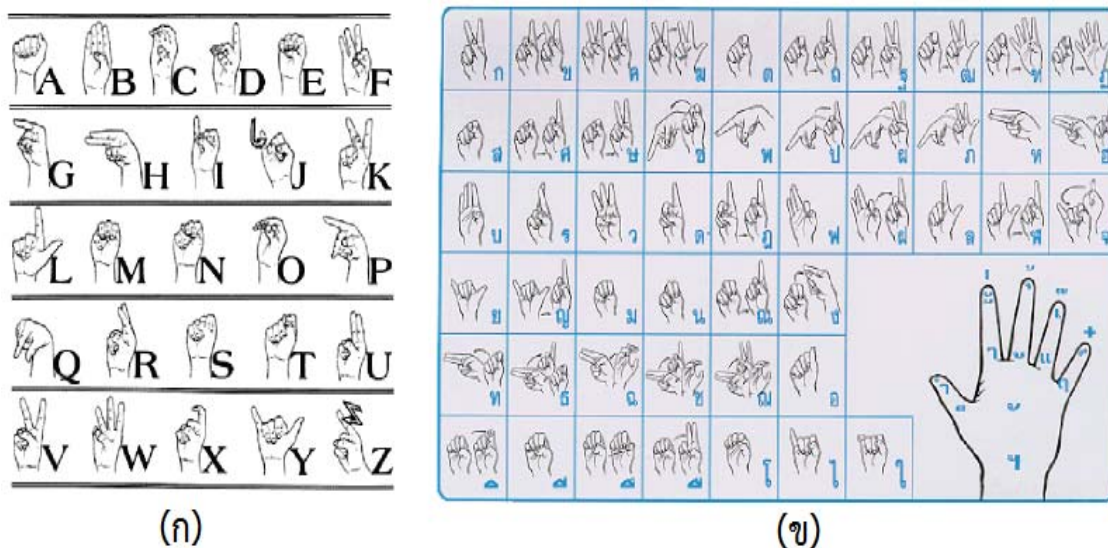
รูปที่ 1-1 คุณลักษณะสำคัญของมือโดยใช้โมเดลของมือเป็น (S. Bilal, R. Akmeliawati, M. J. El Salami, A. A. Shafie (2011)[2])

การสกัดคุณลักษณะสำคัญของมือโดยใช้สิ่งที่ปรากฏบนภาพเป็นหลัก (Appearance base) เป็นการสกัดคุณลักษณะสำคัญของสิ่งที่มองเห็นในภาพ ซึ่งเป็นการสกัดเฉพาะคุณลักษณะเด่นของมือที่ปรากฏในภาพเท่านั้น เช่น ปลายนิ้วมือ รูปร่างของมือที่ปรากฏ ซึ่งวิธีนี้แม้จะไม่สามารถที่จะทำให้ได้โครงสร้างสามมิติของมือทั้งหมด แต่ทำให้สามารถใช้แค่กล้องวิดีโอแค่ตัวเดียวหรือจากเพียงมุมมองเดียวเท่านั้น อีกทั้งยังทำให้การคำนวณเพื่อแสดงผลนั้นเป็นไปได้อย่างรวดเร็วและตรวจจับการเคลื่อนไหวของมือได้ดีกว่าแบบใช้โมเดลของมือเป็นหลัก (Model base) จึงเหมาะที่จะนำมาประยุกต์ใช้ในระบบแบบทันทีกาล (Real time) ได้



รูปที่ 1-2 คุณลักษณะสำคัญของมือที่ใช้สิ่งปรากฏบนภาพ (Omer Rashid และคณะ (2010) [3])

ในการวิเคราะห์ภาษามือไทย การสะกดนิ้วมือภาษามือไทยพัฒนามาจากการสะกดนิ้วมือภาษาอเมริกัน แต่เนื่องจากภาษามือไทยนั้นประกอบด้วยพยัญชนะ สระ วรรณยุกต์ (รูปที่ 1-3 (ข)) ทำให้มีจำนวนท่ามือตัวสะกดมากกว่าภาษามืออเมริกัน ที่มีเพียงแค่ 26 ตัวอักษร (รูปที่ 1-3 (ก)) อีกทั้งยังมีท่ามือหลายท่าประกอบกันเพื่อแสดงถึงอักษรเดียว หรือบางอักษรต้องใช้มือทั้งสองมือเพื่อแสดงท่ามือ ทำให้ท่ามือที่ใช้ในการสะกดคำในภาษามือไทยนั้นมีความยาก และมีความซับซ้อนกว่าท่ามือในภาษาอเมริกัน หรือในภาษาอื่นมาก จึงมีความจำเป็นที่จะต้องมีการกระบวนการในการสกัดคุณลักษณะสำคัญของมือของการสะกดนิ้วมือจากภาพที่แตกต่างท่ามือในภาษาอื่น



รูปที่ 1-3 (ก) ตัวสะกดนิ้วมือภาษาอเมริกัน (ข) ตัวสะกดนิ้วมือภาษาไทย

วิทยานิพนธ์ฉบับนี้จะเสนอวิธีการตรวจจับหาตำแหน่งของมือที่เคลื่อนไหวบนภาพวีดิทัศน์ภาษามือ หากจุดบริเวณที่จะสกัดคุณลักษณะสำคัญของมือในตำแหน่งเฟรมที่เหมาะสม เพื่อที่จะสกัดคุณลักษณะสำคัญของมือจากภาพ โดยการสกัดคุณลักษณะสำคัญของมือ จะใช้วิธีพิจารณาสิ่งที่ปรากฏบนภาพเป็นหลัก (Appearance base) ซึ่งเป็นวิธีที่เหมาะสมกับการแสดงภาษามือ โดยเลือกสกัดคุณลักษณะสำคัญที่เหมาะสมสำหรับการตีความหมายของท่าทาง โดยข้อมูลคุณลักษณะสำคัญของท่ามือนั้นจะมีการปรับปรุงตลอดเวลา และไม่แปรผันตามขนาด (Scale invariant) การหมุน (Rotation invariant) และ การเคลื่อนที่ (Translation invariant) อีกทั้งยังต้องแก้ปัญหาการบิดบังของมือ ซึ่งเป็นปัญหาหลักในการสกัดคุณลักษณะสำคัญของมือสำหรับการแสดงภาษามือไทยซึ่งมักเกิดปัญหาการบิดบังขึ้นบ่อยครั้งในการแสดงภาษามือโดยการพิจารณาค่าความน่าจะเป็นของตำแหน่งมือ และใช้รูปแบบของมือ (Hand template) ที่ได้จากช่วงเวลาใกล้เคียงกันในการช่วยประมาณรูปร่างมือในเฟรมปัจจุบันที่ถูกบิดบัง จึงสามารถทำการสกัดคุณลักษณะสำคัญของมือที่ปรากฏบนภาพเมื่อเกิดการบิดบังได้

1.2 วัตถุประสงค์ของงานวิจัย

เพื่อพัฒนาขั้นตอนวิธีในการตรวจจับมือและสกัดคุณลักษณะสำคัญของมือจากวีดิทัศน์ เพื่อใช้ในการรู้จำการสะกดนิ้วมือภาษาไทย ซึ่งมีการบดบังส่วนของใบหน้า หรือในส่วนที่มีมือบดบังด้วยตัวเอง

1.3 ขอบเขตงานวิจัย

สามารถพัฒนาวิธีการตรวจจับและสกัดคุณลักษณะสำคัญของมือเพื่อใช้ในระบบการสะกดนิ้วมือภาษาไทย มีขอบเขตดังนี้

1. วีดิทัศน์นำเข้าเป็นวีดิทัศน์ที่ได้ถ่ายทำในห้องที่มีสภาพแสงและฉากหลังที่เหมาะสม
2. วีดิทัศน์นำเข้าเป็นวีดิทัศน์ของกลุ่มภาษามือ 1 คน ถ่ายเห็นเฉพาะลำตัวช่วงบน (Upper body) ด้วยระยะห่างจากกล้องคงที่
3. สกัดคุณลักษณะสำคัญของมือสำหรับท่ามือพื้นฐานที่ใช้ในการแสดงพยัญชนะ สระ และวรรณยุกต์ในภาษามือไทย

1.4 วิธีการดำเนินงานวิจัย

วิธีการดำเนินงานวิจัยและการทดลอง เริ่มจากทำการบันทึกภาพวีดิทัศน์ของกลุ่มภาษามือ 1 คน โดยถ่ายเฉพาะลำตัวช่วงบน (Upper body) ด้วยระยะห่างจากกล้องคงที่ ภายในสภาพแวดล้อมที่กำหนด โดยมีขั้นตอนการทำงานตั้งแต่การรับภาพตลอดจนถึงการสกัดคุณลักษณะสำหรับรู้จำภาษามือทั้งสิ้น 6 ขั้นตอน ประกอบด้วย ขั้นตอนการสกัดบริเวณส่วนที่เป็นสีของผิวงั้น การติดตามการเคลื่อนไหวของมือ แยกสถานะของการแสดงท่ามือ หากคีย์เฟรมของการแสดงท่ามือ สกัดคุณลักษณะสำคัญของมือ และนำคุณลักษณะสำคัญของมือไปใช้เพื่อรู้จำท่ามือ โดยทำการสกัดคุณลักษณะสำคัญของมือจากวีดิทัศน์การแสดงท่ามือภาษาไทย ของการสะกดตัวอักษร พยัญชนะ สระ และวรรณยุกต์ และนำคุณลักษณะที่สกัดได้ไปทดสอบโดยผ่านกระบวนการรู้จำ

1.5 ประโยชน์ที่คาดว่าจะได้รับ

ได้ขั้นตอนวิธีในการสกัดคุณลักษณะของมือที่แก้ปัญหากล้องเคลื่อนที่บดบังในการแสดงการสะกดนิ้วมือภาษาไทย

1.6 ผลงานที่ตีพิมพ์จากวิทยานิพนธ์

- “Hand Detection and Feature Extraction for Thai Sign Language Recognition”, in Proc. 15th International Workshop on Advanced Image Technology (IWAIT), Ho Chi Minh City, Vietnam, 2012
- “Hand Detection and Feature Extraction fro Static Thai Sign Language Recognition”, in Proc. International Conference on Ubiquitous Information Management and Communication (ICUIMC 2013), Kota Kinabalu, Malaysia 2013

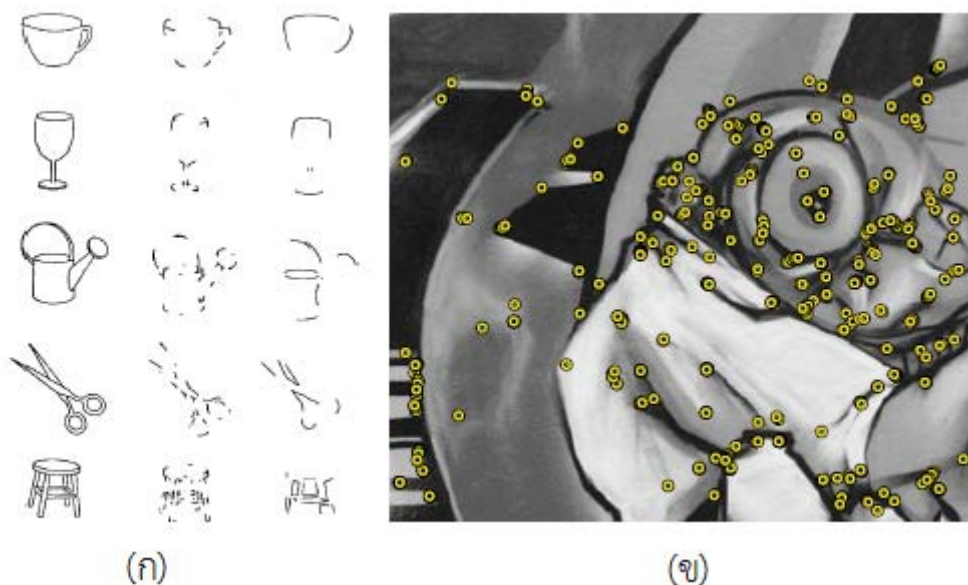
บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 คุณลักษณะสำคัญท้องถิ่น (Local Feature)

คุณลักษณะท้องถิ่นรูปแบบของรูปที่แตกต่างจากบริเวณเพื่อนบ้าน (neighborhood) ซึ่งมักจะเป็นการเปลี่ยนแปลงของคุณสมบัติของจุดภาพบนภาพ เช่น ความเข้มของจุดภาพ (intensity) สี (Color) หรือพื้นผิว (texture) จากรูปที่ 2-1 แสดงคุณลักษณะท้องถิ่น (Local Feature) ของรูปร่างวัตถุ (Contour) (รูป 2-1 (ก)) และคุณลักษณะท้องถิ่นที่เป็นความเข้มจุด (รูป 2-1 (ข)) ซึ่งคุณลักษณะท้องถิ่น (Local Feature) สามารถเป็นได้ทั้งจุด ขอบภาพ หรือส่วนของภาพก็ได้ขึ้นอยู่กับความต้องการ หรือการนำไปใช้ในแต่ละแอปพลิเคชัน



รูปที่ 2-1 (ก) คุณลักษณะสำคัญท้องถิ่นของรูปร่างวัตถุ (ข) คุณลักษณะสำคัญท้องถิ่นของความเข้มจุดภาพ (T. Tuytelaars, Krystian Mikolajczyk (2008) [4])

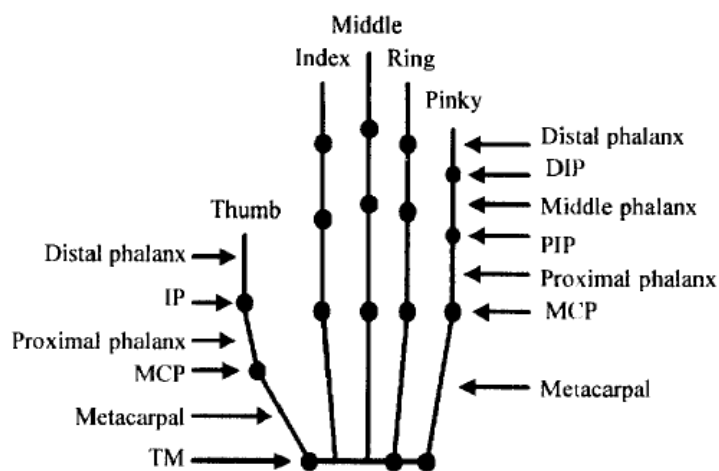
คุณลักษณะสำคัญของรูปร่างที่ดีประกอบด้วยคุณสมบัติ (T. Tuytelaars, Krystian Mikolajczyk (2008) [4]) ดังนี้

- สามารถทำซ้ำได้ (Repeatability) ในการสกัดคุณลักษณะสำคัญของรูปร่างเดียวกัน แต่อาจจะอยู่ในมุมที่ต่างหากันควรจะสกัดได้คุณลักษณะสำคัญที่เหมือนกัน

- สามารถใช้ในการจำแนกได้ (Distinctiveness/Informativeness) ข้อมูลคุณลักษณะสำคัญที่สกัดได้ควรสามารถใช้ในการจำแนกการแตกต่างของรูปร่าง และนำไปเปรียบเทียบได้
- มีความเป็นท้องถิ่น (Locality) คุณลักษณะสำคัญควรสามารถที่จะระบุตำแหน่งบนภาพ และสามารถระบุได้ว่าเป็นตำแหน่งใดในภาพที่มุมมองแตกต่างกัน
- มีปริมาณเพียงพอ (Quantity) คุณลักษณะสำคัญควรมีปริมาณที่เพียงพอที่จะใช้ในการจำแนก ทั้งนี้ขึ้นอยู่กับแอปพลิเคชันที่จะนำไปใช้
- มีความแม่นยำ (Accuracy) คุณลักษณะสำคัญควรมีความละเอียดที่แม่นยำทั้งในเรื่องของตำแหน่ง ขนาด หรือรูปร่างของวัตถุ
- มีประสิทธิภาพ (Efficiency) ประสิทธิภาพในที่ขึ้นอยู่กับแอปพลิเคชันที่จะนำไปใช้ว่ามีข้อจำกัดในเรื่องเวลาหรือไม่ เช่น ต้องการการประมวลผลแบบทันที (Real time)
- ไม่แปรผัน (Invariance) ในกรณีที่มีการสร้างคุณลักษณะสำคัญเป็นโมเดลทางคณิตศาสตร์ในการประมาณการรูปร่างของวัตถุบนภาพ คุณลักษณะสำคัญที่ดีควรไม่ได้รับผลกระทบจากการเปลี่ยนรูปทางคณิตศาสตร์ (Mathematical transformations)
- มีความทนทาน (Robustness) คือ ความทนทานต่อสิ่งที่จะมากระทบต่อภาพ เช่น สิ่งรบกวน (noise) สภาพแสงที่แตกต่าง ความคมชัดของภาพ ฯลฯ ซึ่งสิ่งเหล่านี้จะทำให้ความแม่นยำลดลง คุณลักษณะสำคัญที่ดีควรมีแม่นยำลดลงน้อยที่สุด

2.1.2 แบบจำลองมือ (Hand Model)

ในการศึกษาการเคลื่อนไหวของมือในการทำภาษามือนั้น จำเป็นที่จะต้องเข้าใจโครงสร้างและรูปร่างของมือ ขอบเขตการเคลื่อนไหวของข้อต่อ ขอบเขตและความเป็นไปได้ของการเคลื่อนไหวของมือ เพื่อใช้ในการอนุมานถึงส่วนต่างๆ ของมือภายในภาพ อีกทั้งยังเป็นการลดขอบเขตบริเวณส่วนภาพในการค้นหาและลดเวลาในการค้นหามือในภาพได้



รูปที่ 2-2 ภาพแสดงแบบจำลองของมือที่ใช้ในการพิจารณาการเคลื่อนไหว (Kuch และ J.J. (1994) [1])

เนื่องจากแบบจำลองในปัจจุบันนั้นใช้จุดควบคุม (control point) กว่า 300 จุดสำหรับอ้างอิงตำแหน่งบนมือทั้งหมดทำให้เสียเวลาในการคำนวณส่งผลให้ไม่สามารถแสดงผลแบบทันทีทันใด ดังนั้นงานวิจัย Kuch และ J.J. (1994) [1] ได้เสนอรูปร่างมือ (Hand model) ได้ใช้การบีสไปไลน์ (B-Splines) ในการอ้างอิงพื้นผิวของมือและนิ้วมือซึ่งเป็นการลดจำนวนของจุดควบคุม (control point) ส่งผลให้ระยะเวลาในการคำนวณน้อยลงทำให้สามารถทำงานแบบทันทีทันใด (Real time) ได้

การจะกล่าวถึงโมเดลของมือนั้นจะต้องกล่าวถึงองศาอิสระ (Degree of freedom) ซึ่งเราจะใช้องศาอิสระ ในการพิจารณาการหมุนของมือในแนวแกนต่างๆ ว่าอยู่ในลักษณะใด ซึ่งจากการศึกษาทำให้ทราบว่าองศาอิสระ ทั้งสิ้น 23 องศาอิสระ (Kuch และ J.J. (1994) [1]) ดังนี้ ตรงส่วนนิ้วชี้ถึงนิ้วก้อยจะมีองศาอิสระนิ้วละ 4 องศาอิสระ คือ DIP 1 PIP 1 และ MCP 2 ตรงส่วนนิ้วหัวแม่มือจะมี 5 องศาอิสระ คือ TM 2 MCP 2 และ IP 1 และสุดท้ายคือส่วนฐานล่างสุดของนิ้วนางแล้วนิ้วก้อย 2 องศาอิสระ สำหรับการเริ่มต้นในการสร้างโมเดล จะทำการสร้างตำแหน่งของข้อต่อโดยดูตามโครงสร้างของมือขึ้นมาก่อน จากนั้นจะทำการปรับขนาดให้เข้ากับมือจริงๆ ที่ใช้โมเดลนี้

จากรูปที่ 2-2 ความสามารถของข้อนิ้วมือ (MCP) ในการกางออกและการหดเข้าสู่นิ้วมือ โดยเมื่อนิ้วมืองอเข้าหาฝ่ามือทำให้นิ้วมือสามารถในการกางออกของนิ้วมือน้อยลง โดยสามารถคำนวณได้ดังสมการต่อไปนี้

$$MCP^{\lim}(a/a) = \frac{MCP(f/e)}{90} * (MCP^{converge} - MCP^{s_lim}(a/a)) + MCP^{s_lim}(a/a) \quad \dots(1)$$

โดยที่

$MCP^{\lim}(a/a)$ คือ ความสามารถในการกางนิ้วมือ

$MCP(f/e)$ คือ องศาของการงอนิ้วมือเข้าหาฝ่ามือ

$MCP^{converge}$ คือ เงื่อนไขบังคับคงที่ (Static constrain) ของข้อต่อ

$MCP^{s_lim}(a/a)$ คือ ความสามารถในการกางนิ้วมือที่มากที่สุดที่เป็นไปได้

จากสมการสรุปได้ว่า เมื่อนิ้วมืองอเข้าหาฝ่ามือ ($MCP(f/e)$) มีองศาตั้งแต่ 90 องศา นิ้วมือจะไม่สามารถกางออกจากกันได้ แต่หากนิ้วมือไม่มีการงอเข้าหาฝ่ามือ หรือ มีองศาการงอนิ้วมือเป็น 0 นิ้วมือจะสามารถกางออกได้อย่างเต็มที่

อัตราส่วนขอบเขตการเคลื่อนไหวของส่วนข้อต่อนิ้วมือตรงส่วน PIP (Proximal interphalangeal) และ DIP (Distal interphalangeal) คือ

$$PIP = \frac{3}{2} DIP \quad \dots(2)$$

อัตราส่วนขอบเขตการเคลื่อนไหวของข้อต่อของนิ้วมือตรงส่วน MCP (Metacarpophalangeal) และ PIP (Proximal interphalangeal) คือ

$$MCP(f/e) = \frac{1}{2} PIP \quad \dots(3)$$

2.1.3 การแปลงผันปริภูมิสี (Color conversion)

ในการสกัดคุณลักษณะสำคัญเพื่อการรู้จำในระบบภาษามือ มักมีความจำเป็นการสกัดพื้นที่บริเวณที่เป็นสีของผิวหนังเพื่อให้ได้พื้นที่บริเวณมือและใบหน้า จากงานวิจัย Nariman Habili, Cheng Chew Lim และ Alireza Moini (2004) [7] พบว่าปริภูมิสีที่มีความเหมาะสมในการสกัดสีผิวคือเวกซ์ซีบีซีอาร์ (YCbCr) เนื่องจากเป็นปริภูมิสีที่มีความทนทานต่อการเปลี่ยนแปลงของแสง แต่เนื่องจากภาพที่รับจากกล้องมักเป็นภาพที่อยู่ในปริภูมิสีอาร์จีบี (RGB) ดังนั้นก่อนนำภาพไปประมวลผลเพื่อการสกัดสีผิวหนังจึงมีความจำเป็นต้องแปลงปริภูมิสีก่อน ดังนี้

$$[Y \ Cb \ Cr] = [R \ G \ B] \begin{bmatrix} 0.299 & -0.168935 & 0.99813 \\ 0.587 & -0.331665 & -0.418531 \\ 0.114 & 0.50059 & -0.081282 \end{bmatrix} + [16 \ 128 \ 128] \quad \dots(4)$$

2.1.4 โมเมนต์ของภาพ (Image Moment)

โมเมนต์เป็นวิธีการหนึ่งในการเปรียบเทียบลักษณะของรูปร่างวัตถุ (Shape contour) ในภาพ โดยการคำนวณโมเมนต์ของรูปร่างวัตถุ (Contour moments) ซึ่งโมเมนต์แต่ละชนิดจะบอกคุณลักษณะของรูปร่างวัตถุที่แตกต่างกัน โดยมีการคำนวณดังนี้

$$m_{p,q} = \sum_{i=1}^n I(x,y) x^p y^q \quad \dots(5)$$

โดยที่

$m_{p,q}$ คือ โมเมนต์ของภาพ

n คือ จำนวนของจุดภาพรอบรูปร่างของวัตถุ (Contour)

$I(x,y)$ คือ ค่าความเข้มของจุดภาพที่ตำแหน่ง (x,y)

p คือ ลำดับของ x (x -order)

q คือ ลำดับของ y (y -order)

หาก p และ q มีค่าเท่ากับ 0 (m_{00}) จะหมายถึงความยาวเส้นรอบรูปวัตถุ นอกจากนี้ยังสามารถหาจุดศูนย์กลางของวัตถุจากโมเมนต์ได้จาก

$$\left(\bar{x}, \bar{y} \right) = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad \dots(6)$$

โดยที่

(\bar{x}, \bar{y}) คือ จุดศูนย์กลางของวัตถุบนภาพ

m_{00} คือ ความยาวเส้นรอบรูปวัตถุ

m_{10} คือ ผลรวมของค่าพิกัด x ของเส้นรอบรูปร่างของวัตถุบนภาพ

m_{01} คือ ผลรวมของค่าพิกัด y ของเส้นรอบรูปร่างของวัตถุบนภาพ

เนื่องจากโมเมนต์ที่คำนวณเบื้องต้นเพื่อบอกคุณลักษณะของวัตถุนั้นไม่เพียงพอต่อการเปรียบเทียบรูปร่างระหว่างรูปร่างวัตถุสองสิ่ง เนื่องจากไม่มีความทนทานต่อการเปลี่ยนแปลงตำแหน่ง (translation invariant) จึงได้มีการคำนวณโมเมนต์ศูนย์กลาง (Central moment) ซึ่งมีความทนทานต่อการเปลี่ยนแปลงตำแหน่ง เนื่องจากมีการทำให้เป็นมาตรฐาน (normalize) กับจุดศูนย์กลางของวัตถุ

$$\mu_{p,q} = \sum_{i=0}^n I(x, y)(x - \bar{x})^p (y - \bar{y})^q \quad \dots(7)$$

โดยที่

$\mu_{p,q}$ คือ โมเมนต์ศูนย์กลาง (central moment)

(\bar{x}, \bar{y}) คือ จุดศูนย์กลางของวัตถุบนภาพ

$I(x, y)$ คือ ค่าความเข้มของจุดภาพที่ตำแหน่ง (x, y)

p คือ ลำดับของ x (x -order)

q คือ ลำดับของ y (y -order)

และเพื่อให้โมเมนต์มีความทนทานต่อการขยายขนาด (Scale Invariant) และมีความทนทานต่อการหมุน (Rotation invariant) จึงได้มีการนำโมเมนต์ศูนย์กลางมาหารกับค่าโมเมนต์ที่เหมาะสม ($m_{00} :^*$)

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{00}^{(1+\frac{p+q}{2})}} \quad \dots(8)$$

โดยที่

$\eta_{p,q}$ คือ โมเมนต์ที่ถูกทำให้เป็นมาตรฐาน (Normalized moment)

$\mu_{p,q}$ คือ โมเมนต์ศูนย์กลาง (Central moment)

$\mu_{00}^{(1+\frac{p+q}{2})}$ คือ โมเมนต์ศูนย์กลางมาหารกับค่าโมเมนต์ที่เหมาะสม (m_{00}^*)

ต่อมาได้มีการสร้างเซตของโมเมนต์ 7 ค่าเพื่อใช้ในการเปรียบเทียบรูปร่างวัตถุในภาพ คือ ฮูโมเมนต์ (Hu invariant moment) ซึ่งสร้างมาจากโมเมนต์ศูนย์กลาง (Central moment) และ โมเมนต์ที่ถูกทำให้เป็นมาตรฐาน (Normalized moment) ทำให้ฮูโมเมนต์ได้คุณสมบัติของการทนทานต่อการเปลี่ยนตำแหน่ง (Translation invariant) ทนทานต่อการขยายขนาด (Scale invariant) และทนทานต่อการหมุน (Rotation invariant) ดังนี้

$$h_1 = \eta_{20} + \eta_{02} \quad \dots(9)$$

$$h_2 = (\eta_{20} - \eta_{02})^2 + (2\eta_{11})^2 \quad \dots(10)$$

$$h_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad \dots(11)$$

$$h_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad \dots(12)$$

$$h_5 = (\eta_{30} - \eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad \dots(13)$$

$$h_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad \dots(14)$$

$$h_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad \dots(15)$$

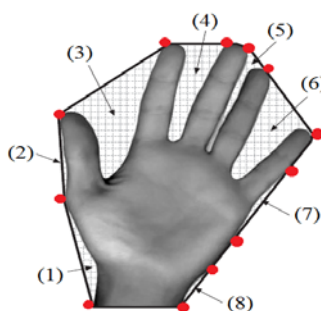
โดยฮูโมเมนต์จะมีทั้งหมด 7 ค่า โดยค่าที่มีลำดับสูงขึ้นไปจะมีแนวโน้มที่จะมีค่าน้อยลง (รูปที่ 2-3) เนื่องจากค่าฮูโมเมนต์ลำดับที่สูงขึ้น จะได้จากโมเมนต์ที่ถูกทำให้เป็นมาตรฐาน (Normalized moment) ที่มีกำลังสูงซึ่งมีค่าน้อย ทำให้ได้ค่าผลรวมที่น้อย

	h_1	h_2	h_3	h_4	h_5	h_6	h_7
A	2.837e-1	1.961e-3	1.484e-2	2.265e-4	-4.152e-7	1.003e-5	-7.941e-9
I	4.578e-1	1.820e-1	0.000	0.000	0.000	0.000	0.000
O	3.791e-1	2.623e-4	4.501e-7	5.858e-7	1.529e-13	7.775e-9	-2.591e-13
M	2.465e-1	4.775e-4	7.263e-5	2.617e-6	-3.607e-11	-5.718e-8	-7.218e-24
F	3.186e-1	2.914e-2	9.397e-3	8.221e-4	3.872e-8	2.019e-5	2.285e-6

รูปที่ 2-3 ค่าฮูโมเมนต์ของแต่ละตัวอักษรที่มีรูปร่างแตกต่างกัน (G.Gradski, A.Kaehler (2008) [5])

2.1.5 เปลือกหุ้มวัตถุ (Convex Hull)

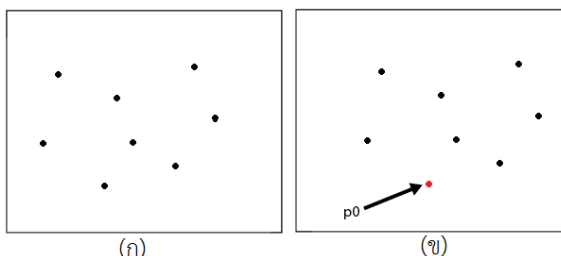
การคำนวณหาเปลือกหุ้มรูปร่างวัตถุ (Convex hull) ล้อมรอบรูปร่างของวัตถุที่พิจารณา จากนั้นทำการคำนวณหาจุดที่ทำให้เปลือกหุ้มวัตถุเกิดการหักเหทำให้เกิดจุดยอดของรูปร่างวัตถุ (Convexity defects) (จากรูปที่ 2-4 คือจุดบนภาพ) การจำแนกรูปร่างด้วยเปลือกหุ้มรูปร่างวัตถุ (Convex hull) จะใช้ได้ดีกับการจำแนกรูปร่างของวัตถุที่มีความแตกต่างของรูปร่างชัดเจน เช่น รูปร่างมือที่มีการกำมือ และการแบมือ ซึ่งจะมีความแตกต่างของจุดที่ทำให้เปลือกหุ้มวัตถุเกิดการหักเห (Convexity defects) แตกต่างกันทั้งจำนวน องศา อย่างชัดเจน ซึ่งสามารถใช้จำนวน และองศาของจุดเหล่านี้ในการจำแนกรูปร่างของมือที่แตกต่างกันออกจากกันได้



รูปที่ 2-4 จุดยอดของรูปร่างมือ G. Gradski และ A. Kaehler (2008) [5])

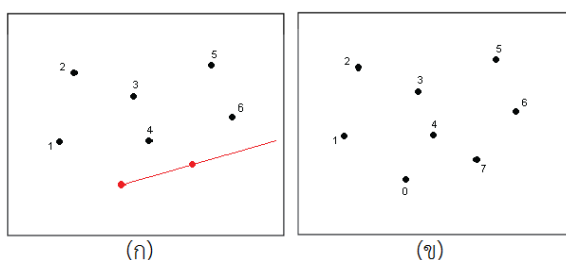
โดยกระบวนการในการคำนวณหาเปลือกหุ้มรูปร่างของวัตถุ (Convex hull) จะใช้วิธีการสามเหรียญ (3-coins algorithm) (Aloupic และ Greg and Bohdan Kalu (1999)[6]) โดยมีลำดับการทำงานดังนี้

1. เลือกจุดสุดขีด (extremal point) ของรูปร่าง เช่น จุดที่มีค่า y ต่ำที่สุด และกำหนดให้เป็นจุด p_0



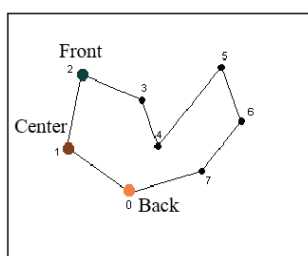
รูปที่ 2-5 (ก) จุดข้อมูลรูปร่างวัตถุ (ข) จุดสุดขีด (extremal point) ของรูปร่าง (Aloupic และ Greg and Bohdan Kalu (1999) [6])

2. เรียงลำดับจุดที่เหลือโดยการกวาดตามเข็มนาฬิกา โดยใช้จุด p_0 เป็นศูนย์กลาง



รูปที่ 2-6 (ก) การกวาดตามเข็มนาฬิกาโดยมีจุด p_0 เป็นศูนย์กลาง (ข) กำหนดลำดับของจุด (Aloupic และ Greg and Bohdan Kalu (1999) [6])

3. กำหนดจุด p_0 , p_1 และ p_2 เปรียบเสมือนเหรียญสามเหรียญ โดยกำหนดให้ p_0 คือ จุดหลัง (back) p_1 คือ ศูนย์กลาง (Center) และ p_2 คือ จุดหน้า (front) ตามลำดับ

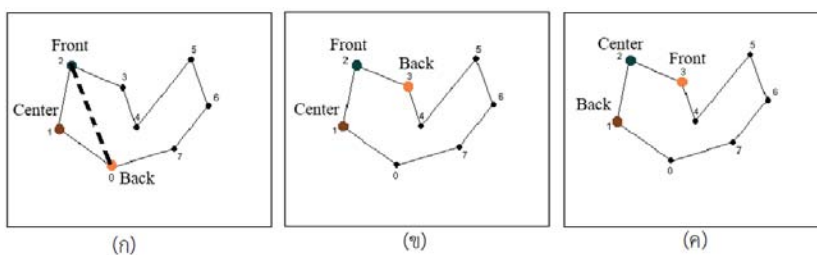


รูปที่ 2-7 กำหนดค่าจุดด้วยวิธีการสามเหรียญ (3-Coins algorithm) (Aloupic และ Greg and Bohdan Kalu (1999) [6])

4. เป็นขั้นตอนการกำหนดการเคลื่อนที่ของเหรียญดังนี้

ถ้าเหรียญทั้งสามมีทิศทางการวางตัวโดยมีทิศทางไปทางซ้าย (form a left turn) หรืออยู่บนเส้นตรงเดียวกัน ให้

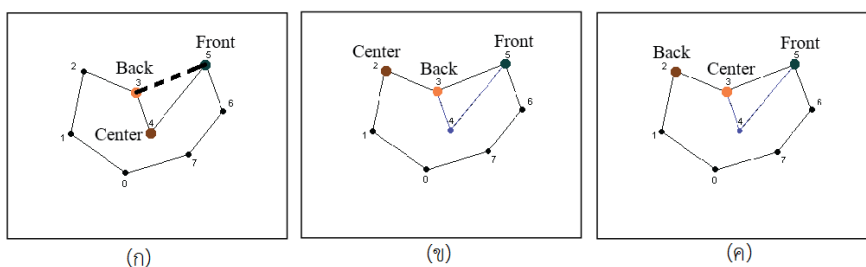
- เลื่อนจุดหลัง (back) ไปยังตำแหน่งจุดถัดไปจากจุดหน้า (front) (รูปที่ 2-8 (ข))
- กำหนดค่าจุดใหม่ (Re label) โดยเปลี่ยนจุดหลัง (back) เป็นจุดหน้า (front), จุดหน้า (front) เป็นศูนย์กลาง (center), และศูนย์กลาง (center) เป็นจุดหลัง (back) (รูปที่ 2-8 (ค))



รูปที่ 2-8 (ก) เหรียญมีทิศทางไปทางซ้าย (ข) เลื่อนตำแหน่งจุดด้านหลัง (ค) กำหนดค่าจุดใหม่ (Aloupic และ Greg and Bohdan Kalu (1999) [6])

ถ้าเหรียญทั้งสามเหรียญมีการวางตัวโดยมีทิศทางไปทางขวา (form a right turn) กำหนดให้

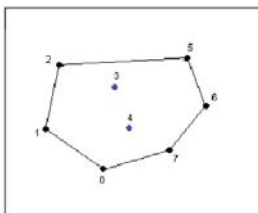
- เปลี่ยนตำแหน่งของศูนย์กลาง (center) ไปอยู่ในตำแหน่งก่อนหน้าของจุดหลัง (back)
- ลบศูนย์กลาง (center) เดิม (จุดเดิมก่อนการย้ายตำแหน่ง) ออกจากลำดับของจุด
- กำหนดค่าจุดใหม่ (Re label) โดยเปลี่ยนศูนย์กลาง (center) เป็นจุดหลัง (back) และจุดหลัง (back) ศูนย์กลาง (center)



รูปที่ 2-9 (ก) เหรียญมีทิศทางไปทางขวา (ข) เลื่อนตำแหน่งจุดศูนย์กลางและลบจุดเดิม (ค) กำหนดค่าจุดใหม่ (Aloupic และ Greg and Bohdan Kalu (1999) [6])

ดำเนินการไปจนกระทั่งกลับมายังตำแหน่งเริ่มต้น (p_0) หรือจนกระทั่งตำแหน่ง p_0 มีค่าเป็นจุดหน้า (front)

5. เชื่อมจุดที่ได้เข้าด้วยกันเพื่อสร้างเป็นเปลือกหุ้มวัตถุ (Convex hull)



รูปที่ 2-10 เชื่อมจุดที่เหลื่อเข้าด้วยกันเพื่อสร้างเปลือกหุ้มวัตถุ (Convex hull) (Aloupic และ Greg and Bohdan Kalu (1999) [6])

2.1.6 การสร้างวงรีแทนกลุ่มของข้อมูล

วิธีสร้างวงรีแทนกลุ่มของข้อมูลจากภาพซึ่งเป็นข้อมูลสองมิติโดยเป็นพิกัดของจุดบนระนาบ เช่น $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ เป็นต้น เริ่มต้นจากการหาเมตริกซ์ความแปรปรวนร่วม (Covariance matrix) ของข้อมูล

$$\Sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix} \quad \dots(17)$$

โดยที่

σ_{xx} คือ ความแปรปรวน (Variance) ของ x

σ_{yy} คือ ความแปรปรวน (Variance) ของ y

σ_{xy} คือ ความแปรปรวนร่วม (Covariance) ของ x และ y

โดยจุดศูนย์กลางวงรีคือค่าเฉลี่ยของข้อมูล (\bar{x}, \bar{y})

ความยาวของแกนวงรีตามแนวแกน X หาได้จาก

$$\alpha = \sqrt{\lambda_1} \quad \dots(18)$$

โดยที่

α คือ ความยาวของแกนวงรีตามแนวแกน X

λ_1 สามารถคำนวณได้จาก

$$\lambda_1 = \frac{\sigma_{xx} + \sigma_{yy} + \Lambda}{2} \quad \dots(19)$$

โดย Λ คำนวณได้จาก

$$\Lambda = \sqrt{(\sigma_{xx} - \sigma_{yy})^2 - 4\sigma_{xy}^2} \quad \dots(20)$$

และคำนวณความยาวของแกนวงรีตามแนวแกน Y หาได้จาก

$$\beta = \sqrt{\lambda_2} \quad \dots(21)$$

β คือ ความยาวของแกนวงรีตามแนวแกน Y

λ_2 สามารถคำนวณได้จาก

$$\lambda_2 = \frac{\sigma_{xx} + \sigma_{yy} - \Lambda}{2} \quad \dots(22)$$

มุมของวงรีสามารถคำนวณได้จาก

$$\theta = \tan^{-1} \left(\frac{-\sigma_{xy}}{\lambda_1 - \sigma_{yy}} \right) \quad \dots(23)$$

โดยที่

θ คือ มุมของวงรี

นอกจากนี้ยังสามารถตรวจสอบได้ว่าข้อมูลจุดภาพใดๆ อยู่ภายในวงรีหรือไม่โดยการคำนวณระยะทางดังนี้

$$D(p, h) = \sqrt{\vec{v} \cdot \vec{v}} \quad \dots(24)$$

โดยที่

$D(p, h)$ คือ ระยะทางจากจุดถึงวงรี

\vec{v} สามารถคำนวณได้จาก

$$\vec{v} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{pmatrix} \frac{x - x_c}{\alpha} \\ \frac{y - y_c}{\beta} \end{pmatrix} \quad \dots(25)$$

หาก $D(p, h)$ มีค่าน้อยกว่า 1 แสดงว่าจุดอยู่ในวงรี

$D(p, h)$ มีค่าเท่ากับ 1 แสดงว่าจุดอยู่บนขอบของวงรี

$D(p, h)$ มีค่ามากกว่า 1 แสดงว่าจุดอยู่นอกวงรี

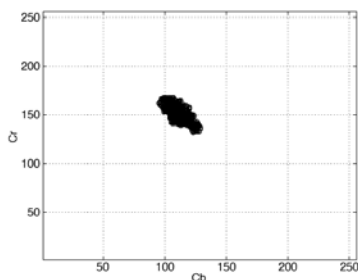
2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 การตรวจจับและติดตามการเคลื่อนไหวของมือ (Hand detection and Hand tracking)

จุดประสงค์ของการตรวจจับและติดตามการเคลื่อนไหวของมือในภาษามือคือ การตรวจจับและติดตามการเคลื่อนไหวของมือของผู้แสดงท่ามือในภาพ เพื่อใช้ในการติดตามหาตำแหน่ง ทิศทางการเคลื่อนที่ และรูปร่างของมือบนภาพ

Nariman Habili, Cheng Chew Lim และ Alireza Moini (2004) [7] ได้นำเสนอวิธีการสกัดสีผิวบริเวณที่เป็นมือและใบหน้าออกจากภาพต่อเนื่อง (Image sequence) โดยได้แบ่งวิธีการออกเป็นสองส่วนคือ การสกัดบริเวณสีผิว ณ เฟรมปัจจุบัน และ ตรวจสอบการเปลี่ยนแปลงของพื้นที่ หรือการเคลื่อนที่ของจุดภาพ จากนั้นนำสองส่วนมารวมกันเพื่อสร้างเป็นโมเดลสำหรับสกัดสีผิวบริเวณส่วนที่เป็นมือและใบหน้าบนภาพ

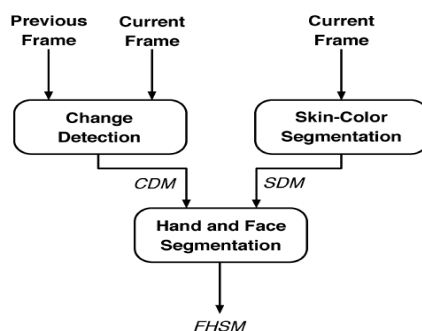
การสกัดส่วนภาพบริเวณสีผิว ทำการสกัดส่วนภาพบริเวณสีผิวภายใต้ปริภูมิสีวายซีอาร์ซีบี (YCrCb Color space) ซึ่งเป็นปริภูมิสีที่มีความทนทานต่อการเปลี่ยนแปลงของแสง เนื่องจากมีการแยกช่องสัญญาณ (Channel) ของความสว่างคือส่วนของช่องสัญญาณวาย (Y channel) ออกจากช่องสัญญาณของสี (CrCb Channel) ทำให้การเปลี่ยนแปลงของความสว่างส่งผลต่อการเปลี่ยนแปลงของสีน้อย โดยการสกัดสีผิวนั้นจะใช้เฉพาะช่องสัญญาณของสีหรือซีอาร์ซีบี (CrCb) เท่านั้น โดยเริ่มจากการเรียนรู้สีผิวเพื่อทำการสร้างโมเดลของผิวก่อน เพื่อทำการสร้างเป็นหน้ากากรตรวจจับสีผิวหนัง (Skin-color Detection Mask : SDM)



รูปที่ 2-11 แสดงสีของผิวหนังบนแกนของ Cr และ Cb (Nariman Habili, Cheng Chew Lim และ Alireza Moini (2004) [7])

การตรวจสอบการเปลี่ยนแปลงของจุดภาพ ทำการแบ่งภาพเป็นบล็อกจากนั้นทำการหาตัวแทนของความเร็วของบล็อกของภาพนั้นๆ ภายใต้สมมติฐานว่าการเคลื่อนไหวของมือและใบหน้าจะเป็นกลุ่มของการเคลื่อนไหวที่มีค่าไม่น้อยกว่าค่าที่กำหนด ซึ่งเป็นค่าที่ได้จากการทดลองเพื่อสร้างเป็นหน้ากากตรวจจับการเปลี่ยนแปลง (Change Detection Mask : CDM)

จากนั้นทำการรวมหน้ากากการตรวจจับสีผิว และหน้ากากการตรวจจับการเปลี่ยนแปลง เข้าด้วยกันเพื่อทำการสร้างเป็นหน้ากากการตรวจจับมือและใบหน้า (Face and Hand Segmentation Mask : SHSM) เพื่อใช้ในการสกัดบริเวณส่วนที่เป็นมือและใบหน้า



รูปที่ 2-12 ขั้นตอนการแยกส่วนที่เป็นมือและใบหน้าออกจากส่วนอื่นของภาพ (Nariman Habili, Cheng Chew Lim และ Alireza Moini (2004) [7])

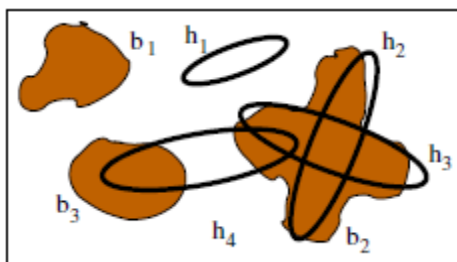
ผู้วิจัยได้ทำการทดสอบการตรวจจับมือและใบหน้าภายในสภาพแวดล้อมที่กำหนด และภายใต้สิ่งแวดลอมใดๆ ซึ่งผลการสกัดบริเวณส่วนที่เป็นมือและใบหน้าออกมาได้เป็นอย่างดี



รูปที่ 2-13 ผลลัพธ์ของการสกัดสีผิว (Nariman Habibi, Cheng Chew Lim และ Alireza Moini (2004) [7])

A. A. Argyros และ M. I. A. Lourakis (2004) [8] ได้นำเสนอวิธีการติดตามตำแหน่งของวัตถุที่มีสีผิว ได้แก่ มือและใบหน้า ซึ่งเป็นวิธีการที่ได้กระทำบนภาพที่ผ่านการสกัดบริเวณสีส่วนที่เป็นผิวออกจากภาพแล้ว ภายใต้สมมติฐานที่ว่าวัตถุที่ทำการติดตามนั้น จะมีลักษณะใกล้เคียงกับรูปร่างของวงรี โดยวิธีการดังกล่าวได้แบ่งปัญหาของการติดตามวัตถุออกเป็น 3 ข้อ คือ การสร้างสมมติฐานตำแหน่งของวัตถุ การติดตามตำแหน่งของวัตถุ และการเลิกการติดตามวัตถุเมื่อวัตถุหลุดออกจากภาพ

การสร้างสมมติฐานตำแหน่งของวัตถุ สร้างวงรีครอบตำแหน่งบริเวณสีของผิวที่สกัดได้ ณ ตำแหน่งเริ่มต้นที่กำหนด โดยสามารถสร้างวงรีได้โดยดูจุดศูนย์กลางและการกระจายตัวของสีผิวจากเมตริกความแปรปรวนร่วมเกี่ยว (Covariance matrix) ของบริเวณจุดภาพที่เป็นสีของผิว



รูปที่ 2-14 วงรีสมมติฐานของบริเวณส่วนที่เป็นผิวหนึ่ง (A. A. Argyros และ M. I. A. Lourakis (2004) [8])

การติดตามตำแหน่งของวัตถุ เมื่อได้วงรีสมมติฐานของวัตถุแล้ว เมื่อมีการเปลี่ยนแปลงของเวลาจะทำการตรวจสอบการเปลี่ยนแปลงของวัตถุ โดยดูการเปลี่ยนแปลงตำแหน่งของจุดภาพบริเวณที่เป็นสีของผิวหนึ่งเทียบกับสมมติฐานของวัตถุ (Object hypothesis) ภายใต้กฎ 2 ข้อ คือ

1. หากจุดภาพสีผิวที่อยู่ภายในสมมติฐานของวัตถุ (Object hypothesis) ในจุดภาพนั้นคือจุดภาพของสมมติฐานของวัตถุ (Object hypothesis) นั้นๆ

2. หากจุดภาพสีผิว ไม่ได้อยู่ภายในสมมติฐานของวัตถุ (Object hypothesis) ใดๆ จุดภาพสีผิวนั้นคือจุดภาพของสมมติฐานของวัตถุ (Object hypothesis) ที่อยู่ใกล้ที่สุด

เมื่อมีการกำหนดข้อมูลจุดภาพที่เป็นสีผิวให้กับสมมติฐานของวัตถุ (Object hypothesis) ใหม่แล้ว ทำการสร้างสมมติฐานของวัตถุ (Object hypothesis) ใหม่จากข้อมูลที่กำหนดใหม่

การยกเลิกการติดตามวัตถุเมื่อวัตถุหลุดออกจากภาพ โดยการตรวจสอบจำนวนจุดภาพที่อยู่ภายในสมมติฐานของวัตถุ (Object hypothesis) หากไม่มีจุดภาพสีผิวนั่งอยู่เลย จะทำการยกเลิกการติดตามวัตถุนั้น

โดยการทดลองการติดตามการเคลื่อนไหวของมือและใบหน้า ให้ความเร็วการติดตามอยู่ในระดับทันการ (Real time) และสามารถติดตามมือและใบหน้าได้แม้เกิดการบังขึ้น



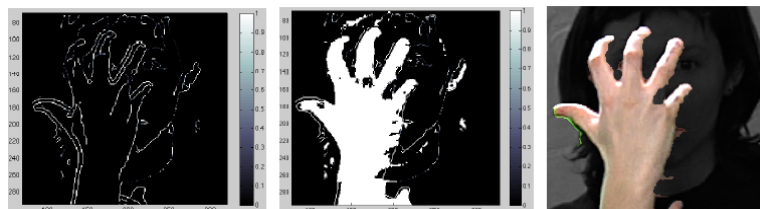
รูปที่ 2-15 ผลลัพธ์ของการติดตามมือและใบหน้า (A. A. Argyros และ M.I.A. Lourakis (2004) [8])

2.2.3 การแก้ไขปัญหาการบังของวัตถุบนภาพ (Occlusion)

การบังของวัตถุบนภาพสามารถแบ่งออกเป็น 2 สาเหตุหลักๆ คือ การบังตัวเอง และการถูกบังโดยวัตถุอื่น ซึ่งในการแสดงท่าภาษามือนั้นสามารถเกิดการบังขึ้นได้จากทั้งสองสาเหตุ

Gonzalez, M., C. Collet และ R. Dubot (2010) [9] ได้นำเสนอวิธีการแก้ไขการบังในกรณีที่มือมีการบังกับใบหน้า โดยใช้คุณลักษณะของสี (Pixel color) และการกำหนดทิศทางของขอบ (Edge orientation) โดยในขั้นตอนแรกทำการสกัดสีผิวหนึ่งจากภาพ ทำการติดตามการเคลื่อนไหวบริเวณใบหน้า และเก็บแม่แบบของใบหน้า (Face template) ไว้เมื่อไม่เกิดการบัง และเมื่อเกิดการบังขึ้น ก็ทำการนำแม่แบบของใบหน้าที่เก็บไว้ มาคำนวณหาขอบภาพโดยใช้การกำหนดทิศทางความลาดชันท้องถิ่น (Local gradient orientations) ของภาพที่ถูกบังเทียบ

กับภาพแม่แบบเพื่อหาตำแหน่งของใบหน้า จากนั้นคำนวณหาการเปลี่ยนแปลงของสีและแยกจุดภาพของใบหน้าและมือจากแม่แบบของใบหน้า



รูปที่ 2-16 ขั้นตอนการแยกจุดภาพส่วนที่เป็นมือและใบหน้าเมื่อเกิดการบดบัง

(Gonzalez, M., C. Collet และ R. Dubot (2010) [9])

ผู้วิจัยได้รายงานว่ วิธีการนี้มีความทนทานต่อการเปลี่ยนแปลงของแสง และให้ข้อมูลรูปร่างของมือที่ละเอียด ซึ่งจากผลการทดลองให้ค่าความถูกต้องของการแยกจุดภาพของมือและใบหน้าออกจากกันเฉลี่ยสูงถึง 96.61%



รูปที่ 2-17 ผลลัพธ์แยกจุดภาพบริเวณส่วนที่เป็นมือและใบหน้าเมื่อเกิดการบดบัง

(Gonzalez, M., C. Collet และ R. Dubot (2010) [9])

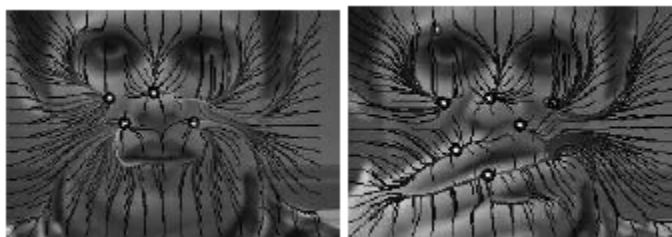
P. Smith, N.D.V. Lobo และ M. Shah (2007) [10] ได้นำเสนอการแก้ไขปัญหการบดบังระหว่างมือและใบหน้าบนภาพโดยใช้เทคนิคภาพสนามแรง (Force field Image) ซึ่งเป็นเทคนิคในการหาพลังงานของแต่ละจุดภาพบนภาพ เพื่อให้เห็นทิศทางของพลังงานของจุดภาพบนภาพ ทำให้สามารถได้คุณลักษณะสำคัญใหม่เพื่อใช้ในการแก้ปัญหการบดบังได้ โดยในขั้นตอนแรกทำการเปลี่ยนภาพสองมิติธรรมดาให้เป็นภาพศักยะและพลังงาน (Potential and Force Image)



รูปที่ 2-18 การเปลี่ยนภาพรับเข้าเป็นภาพศักยะและพลังงาน (Potential and Force Image) โดยภาพแรกเป็นภาพรับเข้า ภาพที่สองเป็นภาพศักยะ (Potential Image) ภาพที่สามเป็นภาพขนาดของสนามพลัง (Magnitude of the force field) และภาพที่สี่เป็นภาพทิศทางของสนามพลัง (Direction of the force field)

(P. Smith, N.D.V. Lobo และ M. Shah (2007) [10])

จากนั้นทำการคำนวณหาพลังงานในแต่ละจุดภาพเพื่อให้ได้ทิศทางของแรงในภาพ และสกัดคุณลักษณะสำคัญของภาพจากจุดยอดของทิศทางแรงที่ได้ภายในภาพเพื่อใช้ในการแยกพื้นที่บริเวณที่ถูกบดบัง โดยทำการเปรียบเทียบการเปลี่ยนแปลงของคุณลักษณะสำคัญ และทิศทางแรงในภาพก่อนถูกบดบังและหลังถูกบดบัง



รูปที่ 2-19 ภาพเปรียบเทียบการเปลี่ยนแปลงของพลังงานของภาพก่อนและหลังการถูกบดบัง

(P. Smith, N.D.V. Lobo และ M. Shah (2007) [10])

ผู้วิจัยได้รายงาน ว่า ผลการทดลองให้ค่าความถูกต้องในการแยกจุดภาพส่วนที่เป็นมือและใบหน้าเฉลี่ยอยู่ที่ 80.04% จากภาพ 44 เฟรม และ 79.53% จาก 150 เฟรม ซึ่งข้อเสียของวิธีนี้คือใช้การคำนวณที่สูง อีกทั้งยังให้ผลลัพธ์การแยกจุดภาพเมื่อเกิดการบดบังที่ไม่ละเอียดเพียงพอกับการรู้จำภาษามือ

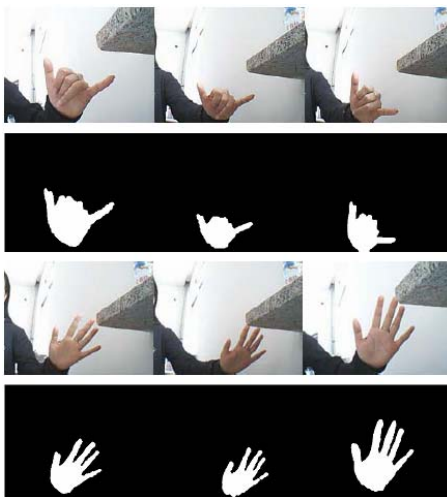


รูปที่ 2-20 ภาพผลลัพธ์ของการแยกบริเวณที่เกิดการบดบังบนภาพ (P. Smith, N.D.V. Lobo และ M. Shah (2007) [10])

2.2.3 การสกัดคุณลักษณะสำคัญของรูปร่างวัตถุบนภาพ (Shape feature extraction)

จุดประสงค์ของการสกัดคุณลักษณะสำคัญของมือคือ เพื่อที่จะสามารถรู้จำท่าทางของมือเพื่อใช้ในการตีความหมายของท่าทางนั้นๆ ได้

W. Tan, C. Wu, S. Zhao และ S. Chen (2007) [11] ได้นำเสนอการสกัดคุณลักษณะสำคัญของมือคือ ฮูโมเมนต์ (Hu moment) ซึ่งเป็นค่าตัวเลข 7 ค่าที่ใช้แทนรูปร่างของวัตถุ โดยได้มีการสกัดสี่บริเวณที่เป็นสีผิว เพื่อให้ได้รูปร่างมือ จากนั้นนำรูปร่างของมือที่ได้ไปสกัดหาค่า ฮูโมเมนต์

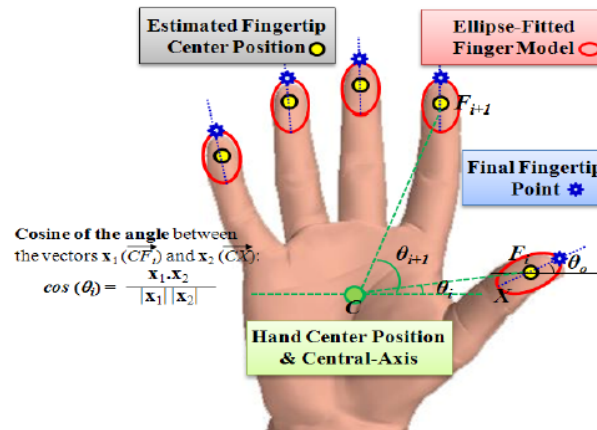


รูปที่ 2-21 ภาพการทดสอบการสกัดฮูโมเมนต์ในระยะและการหมุนที่แตกต่างกัน (W. Tan, C. Wu, S. Zhao และ S. Chen (2007) [11])

โดยได้ทำการทดลองกับการแสดงท่ามือตัวเลขตั้งแต่ 0 ถึง 10 โดยมีการหมุนในมุมที่แตกต่างกันเพื่อทดสอบการแปรผันต่อการหมุน และมีความใกล้เคียงที่แตกต่างกัน เพื่อทดสอบการแปรผันต่อการขยายขนาด ซึ่งจากผลการทดลองสรุปได้ว่า ค่าของฮูโมเมนต์เพียง 3 ค่าแรกก็เพียงพอต่อการจำแนกท่ามือที่มีความแตกต่างกันมาก อีกทั้งยังมีความทนทานต่อการหมุน และการขยายขนาด

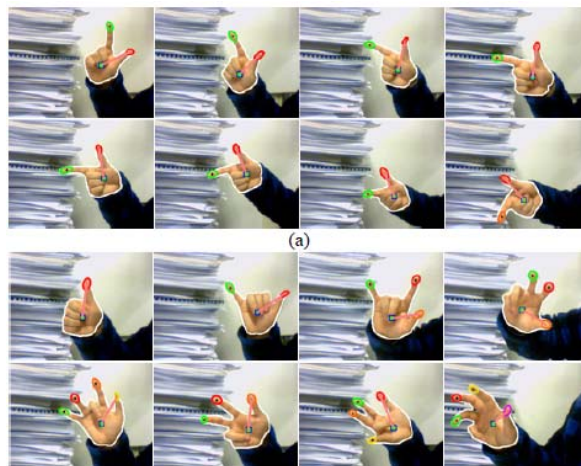
Lae-Kyoung Lee, Su-Yong An และ Se-Young Oh (2012) [12] ได้นำเสนอการสกัดมือและนิ้วมือบนภาพเพื่อการรู้จำการแสดงท่าทางของนิ้วมือแบบทันที โดยในขั้นตอนแรกทำการสกัดหาพื้นที่บริเวณที่เป็นผิวหนังในปริภูมิสีวายซีอาร์ซีบี (YCrCb) จากนั้นทำการติดตามตำแหน่งของมือ และสกัดบริเวณที่เป็นนิ้วมือ โดยการแปลงระยะทาง (Distance transform) และ

คุณสมบัติทางเรขาคณิตของมือ (Geometrical feature of hand) และจากการคำนวณการหมุนและตำแหน่งของมือ ทำให้สามารถคำนวณตำแหน่งและการหมุนของนิ้วมือได้



รูปที่ 2-22 โครงสร้างมือของการสกัดนิ้วมือ (Lae-Kyoung Lee, Su-Yong An และ Se-Young Oh (2012) [12])

จากผลการทดลองแสดงให้เห็นว่าวิธีการนี้สามารถสกัดนิ้วมือได้อย่างรวดเร็ว และมีความทนทานต่อสภาพแสง และยังไม่มีการแปรผันต่อการหมุน (Rotation invariant) และการขยายขนาด (Scale invariant) อีกด้วย และให้ความแม่นยำในการตรวจจับนิ้วมือสูงถึง 95.88 %



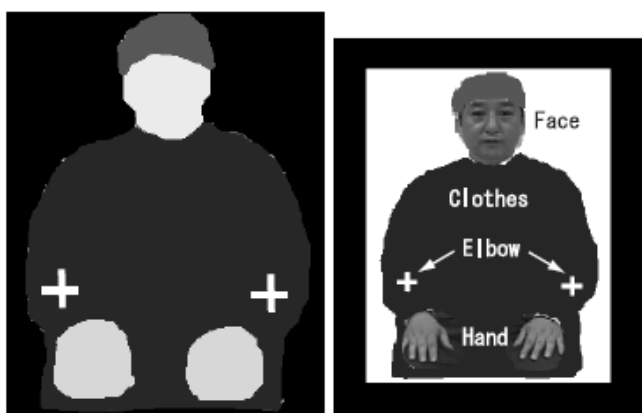
รูปที่ 2-23 ผลการตรวจจับนิ้วมือในการแสดงท่ามือ (Lae-Kyoung Lee, Su-Yong An และ Se-Young Oh (2012) [12])

2.2.4 ระบบรู้จำภาษามือ (Sign language recognition system)

N. Tanibata, N. Shimada และ Y. Shirai (2002) [13] ได้นำเสนอกรอบงานในการสกัดคุณลักษณะสำคัญของมือจากภาพ เพื่อใช้ในการรู้จำคำในการแสดงท่าทางภาษามือ สำหรับ

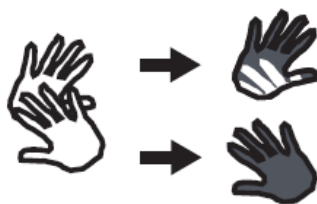
ภาษามือที่ใช้ในการรู้จำท่าทางจะเป็นภาษามือญี่ปุ่น โดยระบบได้แบ่งขั้นตอนการทำงานของกรอบงานออกเป็น 5 ขั้นตอน คือ การกำหนดค่าเริ่มต้นของการติดตาม (Initialization of tracking) การติดตามการเคลื่อนไหวของใบหน้า มือ และข้อศอก (Tracking of face, hand and elbows) การสกัดคุณลักษณะสำคัญ (Feature extraction) และการรู้จำการแสดงท่าทางของคำ (Recognition of JSL words)

การกำหนดค่าเริ่มต้นของการติดตาม (Initialization of tracking) ในขั้นตอนนี้จะเป็นขั้นตอนเริ่มต้นของการติดตาม โดยผู้แสดงภาษามือจะต้องนำมือและใบหน้าให้ไปอยู่ในตำแหน่งเริ่มต้นของการแสดงท่ามือ ซึ่งตำแหน่งดังกล่าวได้ถูกเริ่มต้นให้เป็นตำแหน่งเริ่มต้นของการติดตาม มีการสกัดใบหน้า มือ และเสื้อผ้าออกจากพื้นหลังโดยใช้วิธีการลบฉากพื้นหลัง (Background subtraction) และคัดแยกบริเวณที่เป็นมือและใบหน้าออกจากส่วนที่เป็นเสื้อผ้าโดยวิธีการสกัดบริเวณที่เป็นสีของผิวหนัง (Skin color segmentation)



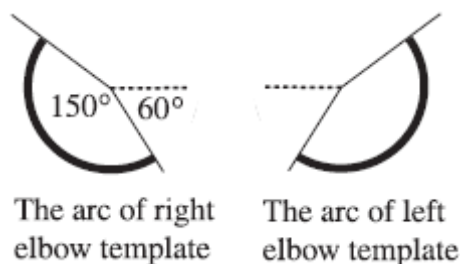
รูปที่ 2-24 การกำหนดค่าเริ่มต้นของการติดตาม (N. Tanibata, N. Shimada และ Y. Shirai (2002) [13])

การติดตามการเคลื่อนไหวของใบหน้า มือ และข้อศอก (Tracking of face, hand and elbows) การติดตามจะใช้การคำนวณหาพื้นที่บริเวณผิวหนังที่อยู่ใกล้เคียงที่สุดกับตำแหน่งที่ได้จากการตั้งค่าเริ่มต้นของการติดตาม และมีการตรวจสอบการบิดบังโดยจะมีการคำนวณพื้นที่ของบริเวณส่วนที่เป็นสีผิว ภายใต้สมมติฐานที่ว่า หากพื้นที่ของสีผิวที่ติดตามมีการเพิ่มขึ้นอย่างรวดเร็ว แสดงว่าวัตถุที่ทำการติดตามเกิดการซ้อนทับกัน โดยการเกิดการซ้อนทับกันขึ้นจะมีการใช้แม่แบบ (template) ของรูปร่างท่ามือก่อนหน้าที่เกิดการบิดบังในการแยกพื้นที่ส่วนที่บิดบังออกจากกัน



รูปที่ 2-25 การใช้แม่แบบ (template) ในการแยกท่ามือที่มีการบิดบัง (N. Tanibata, N. Shimada และ Y. Shirai (2002) [13])

นอกจากนี้ยังมีการหาบริเวณส่วนที่เป็นข้อศอกโดยการใช้แม่แบบของส่วนของวงกลมสำหรับหาตำแหน่งข้อศอก (Arc of elbow template) ในการกวาดหาบริเวณส่วนที่เป็นข้อศอกจากภาพ



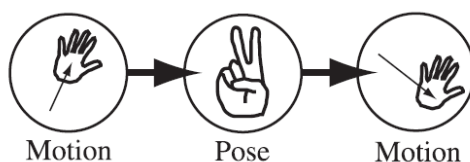
รูปที่ 2-26 การใช้แม่แบบของส่วนของวงกลมสำหรับหาตำแหน่งข้อศอก (N. Tanibata, N. Shimada และ Y. Shirai (2002) [13])

การสกัดคุณลักษณะสำคัญ (Feature extraction) ระบบได้สกัดคุณลักษณะสำคัญจากภาพที่ใช้ในการรู้จำการแสดงท่าทางภาษามือมาทั้งหมด 6 คุณลักษณะ

1. ความเรียบของพื้นที่ส่วนที่เป็นมือ (Flatness of hand region)
2. ตำแหน่งของมือ (Gravity center position of the hand region)
3. พื้นที่บริเวณมือ (Area of the hand region)
4. ทิศทางของการเคลื่อนที่มือ (Direction of hand motion)
5. ทิศทางของมือ (Direction of hand region)
6. จำนวนของสิ่งที่ยื่นออกมาจากพื้นที่มือ (Number of protrusions) ได้แก่ นิ้วมือ

การรู้จำการแสดงท่าทางของคำ (Recognition of JSL words) งานวิจัยนี้ได้ใช้รูปแบบการรู้จำแบบมาคอฟ (Markov Hidden model) ในการรู้จำคำในการแสดงท่าทางภาษามือ โดยได้

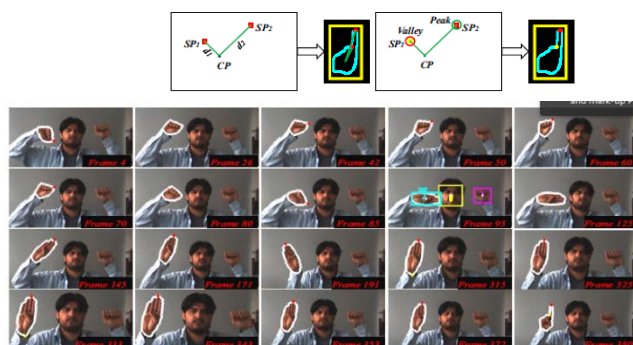
ทำการตัดคำจากการเคลื่อนไหวของมือซ้ายและมือขวา จากนั้นนำลำดับการเคลื่อนไหวของมือซ้ายและมือขวามาตรวจหาความน่าจะเป็นของท่าทาง เพื่อใช้ในการอ้างถึงคำด้วยการรู้จำแบบมาคอฟ โดยเลือกคำที่มีความน่าจะเป็นสูงสุดจากลำดับท่าทางที่แสดง



รูปที่ 2-27 การรู้จำท่ามือ (N. Tanibata, N. Shimada และ Y. Shirai (2002) [13])

งานวิจัยที่นำเสนอกรอบงานสำหรับระบบรู้จำภาษามือญี่ปุ่นนี้ ผู้วิจัยได้รายงานว่าสามารถใช้ในการรู้จำคำในการแสดงท่าทางภาษามือได้ทั้งหมด 65 ท่ามือ สามารถใช้ได้กับพื้นหลังใดๆ (Complex background) และมีการประมวลผลแบบทันที (Real time)

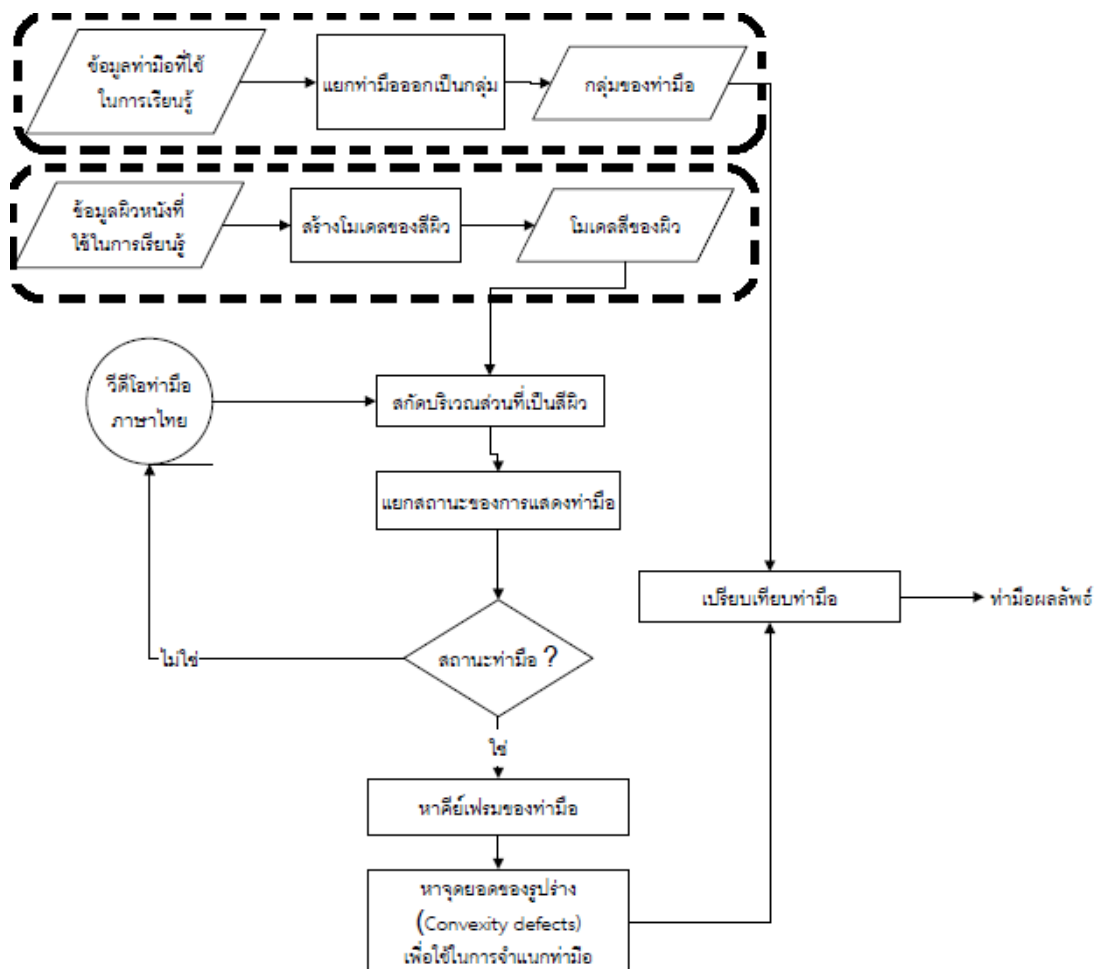
Omer Rashid และคณะ (2010) [3] ได้นำเสนอระบบรู้จำท่าทางภาษามือ โดยใช้ในการแปลงความหมายท่ามือของตัวอักษรภาษาอังกฤษ (A-Z) และตัวเลข (0-9) โดยใช้คุณลักษณะสำคัญคือฮูโมเมนต์รวมกับคุณลักษณะสำคัญเชิงเรขาคณิต ได้แก่ รูปร่างของวงกลมและสามเหลี่ยมในการช่วยในการรู้จำท่ามือ โดยทำการแบ่งหมวดหมู่ของท่ามือตามจำนวนของปลายนิ้วมือที่หาได้ภายในภาพ ซึ่งให้ผลในการรู้จำสูงถึง 98.6 %



รูปที่ 2-28 ระบบรู้จำโดยใช้เวกเตอร์คุณลักษณะสำคัญทางสถิติและเรขาคณิตในการรู้จำแบบ SVM (Omer Rashid และคณะ (2010) [3])

บทที่ 3 ขั้นตอนวิธีที่เสนอ

ในการตรวจจับและสกัดคุณลักษณะสำคัญของมือนั้น จำเป็นที่จะต้องตรวจจับตำแหน่งของมือบนภาพให้ได้ก่อน จากนั้นจึงนำส่วนภาพที่เป็นบริเวณมือที่สกัดได้ ไปทำการสกัดคุณลักษณะสำคัญของภาพมือเพื่อใช้ในการรู้จำต่อไป แต่สำหรับการแสดงท่ามือที่มีความต่อเนื่องนั้น ข้อมูลนำเข้าเป็นวิดีโอที่คนภาษามือซึ่งประกอบด้วยเฟรมภาพหลายเฟรมจึงจำเป็นต้องจำแนกเฟรมภาพให้ได้ว่าเฟรมใดของวิดีโอที่มีการแสดงท่ามือ เฟรมใดที่ไม่มีการแสดงท่ามือ



รูปที่ 3-1 ขั้นตอนการทำงานของ การตรวจจับและสกัดคุณลักษณะสำคัญของมือสำหรับระบบรู้จำภาษามือไทย

ซึ่งในการจำแนกช่วงเฟรมของการแสดงท่ามือนั้นสามารถใช้ความเร็วของมือในการพิจารณาเพื่อจำแนกช่วงของการแสดงท่ามือนั้นได้ เมื่อได้ช่วงของการแสดงท่ามือนั้นแล้ว ต่อไปคือการหาคีย์เฟรมของการแสดงท่ามือนั้น สามารถทำได้โดยการเปรียบเทียบรูปร่างของท่ามือนั้นในปัจจุบันและในเฟรมก่อนหน้า เพื่อหาอัตราการเปลี่ยนแปลงของท่ามือ ซึ่งช่วงที่อัตราการเปลี่ยนแปลงรูปร่างของมือคงที่แสดงว่ามือมีการหยุดนิ่งซึ่งแสดงว่ามีการแสดงท่ามือ ก็จะถือว่าเป็นคีย์เฟรมของการแสดงท่ามือ ด้วยวิธีการดังกล่าวทำให้สามารถจำแนกคีย์เฟรมออกมาได้จากนั้นนำคีย์เฟรมที่ได้ไปสกัดหาคุณลักษณะสำคัญของมือบนภาพโดยการจำแนกกลุ่ม และเปรียบเทียบรูปร่างของมือภายในกลุ่มต่อไป ซึ่งขั้นตอนการทำงานได้แบ่งเป็นขั้นตอน (รูปที่ 3-1)

งานวิจัยนี้ต้องการศึกษาค้นคว้าเพื่อหากระบวนการวิธีการตรวจจับและสกัดคุณลักษณะสำคัญของมือเพื่อใช้ในการรู้จำภาษามือไทย โดยขั้นตอนการทำงานในการตรวจจับและสกัดคุณลักษณะสำคัญของมือนี้ได้แบบขั้นตอนการทำงานออกเป็น 5 ขั้นตอน ต่อไปนี้

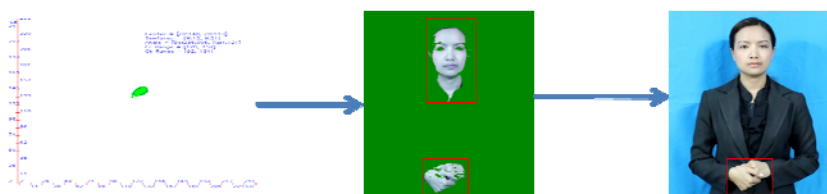
1. สกัดบริเวณส่วนที่เป็นสีของผิวหนัง
2. ติดตามการเคลื่อนไหวของมือ
3. แยกสถานะของการแสดงท่ามือ
4. หาคีย์เฟรมของท่ามือ
5. สกัดคุณลักษณะสำคัญของรูปร่าง และท่าทางของมือ
6. นำคุณลักษณะสำคัญของรูปร่างมือที่สกัดได้มาทำการรู้จำท่ามือ

3.1 สกัดบริเวณส่วนที่เป็นสีผิว

เริ่มต้นรับภาพซึ่งอยู่ในปริภูมิสีอาร์จีบี (RGB) โดยทำการแปลงให้อยู่ในปริภูมิสีวายซีอาร์ซีบี (YCrCb) (สมการที่ 4) เนื่องจากเป็นปริภูมิสีที่เหมาะสมต่อการสกัดแยกส่วนภาพที่เป็นสีของผิวหนัง (Nariman Habili, Cheng Chew Lim และ Alireza Moini (2004)) เนื่องจากมีการแยกช่องสัญญาณส่วนที่เป็นความสว่าง และสีออกจากกัน ทำให้มีความทนทานต่อการเปลี่ยนแปลงสภาพแสงระดับหนึ่ง จากนั้นภาพที่ได้มาสกัดบริเวณส่วนที่เป็นสีผิวจากรูปแบบของสีผิว (Skin color model) ที่ได้ทำการเรียนรู้ไว้ โดยการเรียนรู้รูปแบบของสีผิว (Skin color model) จะศึกษาการกระจายตัวของค่าสีในช่องสัญญาณของซีอาร์ (Cr Channel) และซีบี (Cb Channel) โดยจะ

สามารถสังเกตได้ว่าจะมีการกระจายตัวอยู่ในรูปแบบของวงรีดังรูปที่ 3-2 (ซ้าย) จากนั้นทำการสร้างสมการวงรีเพื่อเป็นตัวแทนของการกระจายตัวของสีผิวหนึ่ง เพื่อสร้างเป็นรูปแบบของสีผิวหนึ่ง (Skin color model)

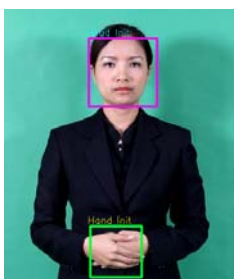
นำรูปแบบการกระจายตัวของสีผิวหนึ่ง (Skin color model) ที่ได้ ไปทำการสกัดบริเวณที่เป็นสีผิวหนึ่งบนภาพ โดยการเปรียบเทียบจุดภาพกับสมการรูปแบบของสีผิว (Skin color model) เพื่อสกัดส่วนภาพที่เป็นสีผิว ดังรูปที่ 3-2



รูปที่ 3-2 การค้นหาส่วนภาพที่เป็นสีผิวจากกล้องวิดีโอ

3.2 ติดตามการเคลื่อนไหวของมือ

หลังจากสกัดสีผิวจะเหลือเฉพาะส่วนภาพที่เป็นพื้นที่ของสีผิว จากนั้นทำการค้นหาบริเวณส่วนที่เป็นใบหน้า โดยใช้วิธีการหาคุณลักษณะสำคัญฮาร์ไลค์ (Haar like feature) เมื่อได้ส่วนภาพที่เป็นใบหน้า ส่วนภาพที่เหลือจะถูกกำหนดว่าเป็นพื้นที่ของมือ เป็นการกำหนดตำแหน่งเริ่มต้นของใบหน้าและมือ เพื่อเริ่มการติดตามการเคลื่อนไหว โดยการสร้างสมมติฐานของวัตถุ (Object hypothesis) แทนส่วนภาพมือและใบหน้า ดังแสดงในรูปที่ 3-3



รูปที่ 3-3 ผลการหาตำแหน่งเริ่มต้นเพื่อใช้ในการติดตามการเคลื่อนไหวของส่วนภาพใบหน้าและมือ

เมื่อได้สมมติฐานของวัตถุ (Object hypothesis) ของมือและใบหน้าแล้ว เมื่อมีการเปลี่ยนแปลงของเวลาจะทำการตรวจสอบการเปลี่ยนแปลงของมือและใบหน้า โดยดูการ

เปลี่ยนแปลงตำแหน่งของจุดภาพบริเวณที่เป็นสีผิวเทียบกับสมมติฐานของวัตถุ (Object hypothesis) ภายใต้กฎ 2 ข้อ (A. A. Argyros, M. I. A. Lourakis (2004) [8]) คือ

1. หากจุดภาพสีผิวที่อยู่ภายในสมมติฐานของวัตถุ (Object hypothesis) ใด จุดภาพนั้นคือจุดภาพของสมมติฐานวัตถุ (Object hypothesis) นั้นๆ

2. หากจุดภาพสีผิว ไม่ได้อยู่ภายในสมมติฐานของวัตถุ (Object hypothesis) ใดๆ จุดภาพสีผิวนั้นคือจุดภาพของสมมติฐานของวัตถุ (Object hypothesis) ที่อยู่ใกล้ที่สุด

เมื่อมีการกำหนดข้อมูลจุดภาพที่เป็นสีผิวให้กับสมมติฐานของวัตถุ (Object hypothesis) ใหม่แล้ว ทำการสร้างสมมติฐานของวัตถุ (Object hypothesis) ใหม่จากข้อมูลที่กำหนดใหม่ ทำให้เราได้ตำแหน่งของใบหน้า และมือ ณ ปัจจุบัน ดังรูปที่ 3-4



รูปที่ 3-4 ผลการติดตามการเคลื่อนไหวของส่วนภาพใบหน้าและมือโดยใช้สมมติฐานของวัตถุ (Object hypothesis)

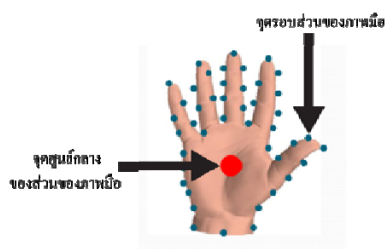
3.3 แยกสถานะของการแสดงท่ามือ

เมื่อทำการติดตามการเคลื่อนไหวของส่วนภาพมือและใบหน้าได้แล้ว ทำการคำนวณหาความเร็วในการเคลื่อนที่ของมือ โดยคำนวณจากจุดศูนย์กลางของมือ เปรียบเทียบผลต่างของตำแหน่งมือ ณ ตำแหน่งปัจจุบัน และตำแหน่งมือของเฟรมก่อนหน้า โดยสามารถหาตำแหน่งศูนย์กลางของมือได้จาก

$$X_c = \frac{1}{N} \sum_{i=1}^N x_i \quad \dots(26)$$

$$Y_c = \frac{1}{N} \sum_{i=1}^N y_i$$

...(27)



รูปที่ 3-5 การหาจุดศูนย์กลางข้อมือ

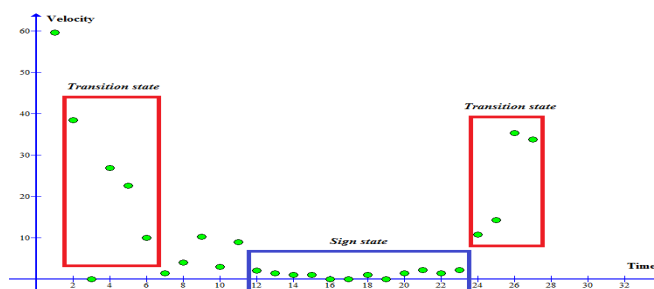
จากการศึกษาการเคลื่อนไหวข้อมือในการแสดงท่ามือสังเกตได้ว่า สามารถใช้ความเร็วของมือในการแยกแยะการแสดงท่ามือได้ โดยก่อนที่จะมีการแสดงท่ามือผู้แสดงท่ามือจะแสดงท่าทางหยุดนิ่ง ซึ่งมือจะไม่มีเคลื่อนไหว เมื่อผู้แสดงท่ามือจะทำการแสดงท่าทางของมือ ช่วงนี้การเคลื่อนไหวข้อมือจะเป็นไปอย่างรวดเร็ว และมีการหยุดนิ่งอีกครั้งเมื่อมีการแสดงท่ามือ จากการสังเกตนี้ทำให้เราสามารถแบ่งสถานะของการแสดงท่ามือออกเป็น 3 สถานะ คือ สถานะหยุดนิ่ง สถานะเปลี่ยนแปลง และสถานะแสดงท่ามือ ดังรูปที่ 3-6



รูปที่ 3-6 ผลการวิเคราะห์เฟรมภาพลำดับการเคลื่อนไหวข้อมือในการแสดงท่ามือซึ่งแบ่งออกเป็น 3 กลุ่ม คือ สถานะเริ่มต้น (กรอบเส้นจุด) สถานะเปลี่ยนแปลง (กรอบเส้นประ) และสถานะท่ามือ (กรอบเส้นทึบ)

จากรูปที่ 3-6 คือการจำแนกสถานะของท่ามือจากความเร็วในการเคลื่อนที่ข้อมือ โดยได้ทำการแบ่งสถานะของท่ามือออกเป็น 3 กลุ่มคือ สถานะเริ่มต้น (Start and stop State) (กรอบเส้นจุด) สถานะเปลี่ยนแปลง (Transition State) (กรอบเส้นประ) และสถานะท่ามือ (Sign State) (กรอบเส้นทึบ) โดยหากมืออยู่ ณ ตำแหน่งเริ่มต้นและไม่มีการเคลื่อนไหว จะพิจารณาว่า เป็น

สถานะเริ่มต้น (Start and stop State) หากมือมีการเคลื่อนที่จากสถานะเริ่มต้น และมีความเร็วมากกว่าค่าหนึ่งจะพิจารณาว่าเป็นสถานะเปลี่ยนแปลง (Transition State) และหากท่ามือมีความเร็วน้อยกว่าค่าหนึ่ง และมีรูปร่างของมือคงที่ ณ ช่วงเวลาหนึ่ง จะพิจารณาว่าเป็นสถานะท่ามือ (Sign State)

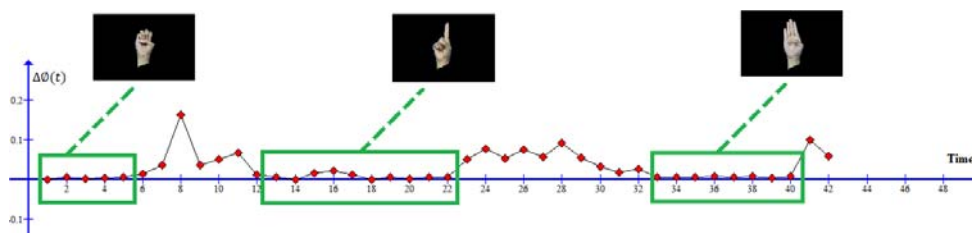


รูปที่ 3-7 กราฟแสดงความเร็วในการเคลื่อนที่ของมือ แกนนอนคือเวลา ส่วนแกนตั้งคือความเร็วของส่วนภาพมือ

จากรูปที่ 3-7 ความเร็วของมือในสถานะเปลี่ยนแปลง (Transition state) จะมีความเร็วที่สูงมากเมื่อเทียบกับสถานะท่ามือ (Sign state) ซึ่งมีความเร็วที่ค่อนข้างคงที่ ทำให้เราสามารถที่จะจำแนกสถานะได้

3.4 หาคีย์เฟรมของท่ามือ

หลังจากทำการแยกสถานะของท่ามือแล้ว เราจะทำการหาคีย์เฟรมของท่ามือในสถานะท่ามือ โดยทำการสกัดค่าฮูโมเมนต์ของท่ามือ แล้วทำการเปรียบเทียบค่าฮูโมเมนต์ของรูปร่างท่ามือในเฟรมปัจจุบันกับค่าฮูโมเมนต์ของท่ามือในเฟรมก่อนหน้า เพื่อหาอัตราการเปลี่ยนแปลงรูปร่างของท่ามือ

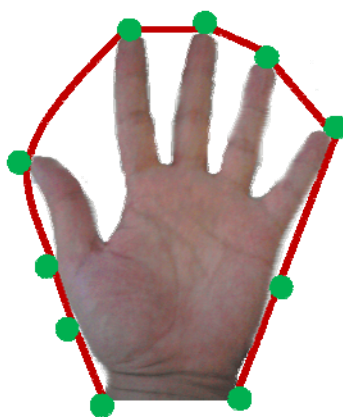


รูปที่ 3-8 คีย์เฟรมของท่ามือ

โดยหากท่ามือนี้อัตราการเปลี่ยนแปลงของรูปร่าง หรือค่าผลต่างของค่าฮูโมเมนต์ต่ำกว่าค่าที่กำหนดต่อเนื่องกันเป็นระยะเวลาหนึ่ง ซึ่งหมายความว่ารูปร่างของมือเปลี่ยนแปลงไปไม่มาก เรา จะทำการสรุปว่าเฟรมนั้นเป็นคีย์เฟรมของท่ามือนั้นๆ จากรูปที่ 3-8 สังเกตได้ว่า ช่วงที่มีการแสดง ท่ามือ ขณะที่มีการแสดงท่ามือนี้อัตราการเปลี่ยนแปลงรูปร่างจะคงที่ เนื่องจากรูปร่างของมือไม่มี การเปลี่ยนแปลง หรือเปลี่ยนแปลงเพียงเล็กน้อย ทำให้เราสามารถนำช่วงที่ท่ามือนี้อัตราการ เปลี่ยนแปลงน้อยมาใช้ในการหาคีย์เฟรมได้ โดยสกัดคีย์เฟรมจากช่วงที่มีอัตราการเปลี่ยนแปลง ของรูปร่างน้อย มาเป็นคีย์เฟรม

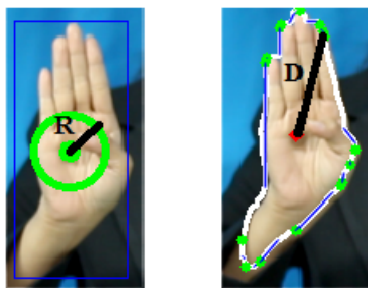
3.5 หาจุดยอดของรูปร่าง (Convexity defects) เพื่อใช้ในการจำแนกท่ามือ

หลังจากได้คีย์เฟรมของท่ามือ ก็ทำการหาจุดยอดรูปร่างของมือ สำหรับการจำแนกรูปร่าง ของท่ามือเป็นกลุ่ม ซึ่งจุดยอดของรูปร่างมือคือที่ทำให้เปลือกหุ้ม (Convex hull) ของมือเกิดการ หักเหทำให้เกิดจุดยอดของรูปร่างท่ามือ (จุด)



รูปที่ 3-9 จุดยอดของรูปร่างมือ

เมื่อได้จุดยอดของรูปร่างมือแล้ว เราจะทำการสกัดเอาเฉพาะจุดที่มีคุณลักษณะเด่น สำหรับนำมาพิจารณาแบ่งรูปร่างเท่านั้น ซึ่งเกณฑ์ในการพิจารณาคือระยะห่างจากตำแหน่งของ มือเป็นระยะหนึ่ง และอยู่ภายในองศาที่กำหนดเท่านั้น



รูปที่ 3-10 สกัดจุดยอดของรูปร่างมือสำคัญที่ใช้ในการจำแนกกลุ่มของท่ามือ
 พิจารณาสกัดจุดยอดของรูปร่างมือสำคัญ (Salient Convexity defects) ดังสมการ

$$H_s = \{H_{convex}(X); 0 < \theta < 180; D > R\} \quad \dots(28)$$

โดยที่

- H_s คือ เซตของจุดคอนเวกซ์สำคัญ
- H_{convex} คือ จุดคอนเวกซ์ที่สกัดได้จากรูปร่างของมือ
- D คือ ระยะทางจากจุดตำแหน่งของมือไปยังจุดคอนเวกซ์
- R คือ ระยะทางน้อยที่สุดจากจุดตำแหน่งของมือไปยังจุดคอนเวกซ์

จุดยอดของรูปร่างมือสำคัญ (Salient Convexity defects) เหล่านี้ใช้ในการจำแนกกลุ่ม
 ของท่ามือโดยใช้จำนวน ของจุดยอดของรูปร่างมือสำคัญในการแบ่งกลุ่มของท่ามื่อดังรูป ซึ่งได้
 แบ่งออกเป็น 6 กลุ่ม ดังตารางที่ 3-1



รูปที่ 3-11 ท่ามือเดี่ยวในท่ามือสะกดไทย

ตารางที่ 3-1 ตารางการจำแนกกลุ่มของท่ามือเดียว

กลุ่ม	จำนวนจุดคอนเวกซ์	หมายเลขรูปร่างท่ามือ
1	0	1,2,3,4,5,6,7,8,9,10
2	1	11,12,13,14,15,16,17,18
3	2	19,20,21,22,23,24,25
4	3	26,27,28,29,30,31,32
5	4	33,34
6	5	35

3.6 เปรียบเทียบรูปร่างของมือกับข้อมูลท่ามือ

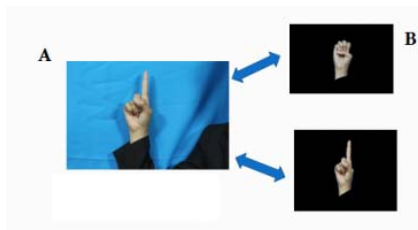
ทำการเปรียบเทียบคุณลักษณะสำคัญของมือที่สกัดมาได้จากข้อมูลรูปร่างท่ามือที่เก็บไว้ ซึ่งได้ทำการแบ่งข้อมูลออกเป็น 6 กลุ่ม โดยจะเปรียบเทียบค่าฮูโมเมนต์ (Hu moment) และความหนาแน่นของขอบภาพ (Edge density) (รูปที่ 3-12) ของรูปร่างคุณลักษณะสำคัญของมือที่สกัดได้จากกลุ่มข้อมูลท่ามือที่เก็บไว้



รูปที่ 3-12 ความหนาแน่นของขอบภาพ (Edge density)

$$I(A, B) = \sum_{i=1}^8 \left| \frac{1}{m_i^A - m_i^B} \right| \quad \dots(29)$$

โดยจะเลือกค่าที่น้อยที่สุดซึ่งแสดงว่ารูปร่างของคุณลักษณะของมือที่สกัดได้จากข้อมูลท่ามือใกล้เคียงกันมากที่สุด



รูปที่ 3-13 การจำแนกท่ามือภายในกลุ่มของท่ามือ

บทที่ 4

การทดลองและผลการทดลอง

ภาพจากกล้องวีดิทัศน์ที่ใช้ในการทดลองได้บันทึกในสภาพแวดล้อมปิด เพื่อควบคุมปัจจัยที่ส่งผลต่อภาพวีดิทัศน์ เช่น สภาพแสง ระยะห่างระหว่างกล้องและล่ามภาษามือ โดยอุปกรณ์ที่ใช้ในการทดลองประกอบด้วย

- เครื่องคอมพิวเตอร์แล็ปท็อป

ซีพียู	Intel core i5
หน่วยความจำ	4 GB
ระบบปฏิบัติการ	Window 7 x86

- กล้องรับภาพความละเอียด 1.3 ล้านพิกเซล สำหรับบันทึกวีดิทัศน์ล่ามภาษามือ โดยผู้วิจัยได้ดำเนินการทดสอบดังต่อไปนี้

4.1 การทดสอบการสกัดบริเวณส่วนที่เป็นสีของผิวหนังบนภาพ

การสกัดสีผิวจากวีดิทัศน์ของล่ามภาษามือไทยจากวีดิทัศน์ภายใต้สภาพแวดล้อมที่กำหนด โดยโมเดลสีผิว



รูปที่ 4-1 ผลการสกัดสีผิวจากวีดิทัศน์

4.2 การทดสอบการติดตามการเคลื่อนไหวของมือบนภาพจากวีดิทัศน์

ได้ทำการทดสอบการติดตามตำแหน่งของมือของล่ามภาษามือไทย 1 คนจากวีดิทัศน์ ภายใต้สภาพแวดล้อมที่กำหนดขณะแสดงท่าทางภาษามือใน กรณีสถานการณ์มือซ้ายและมือขวาไม่บดบังกัน และกรณีที่มือซ้ายหรือมือขวาเกิดการบดบังกัน



รูปที่ 4-2 ผลการติดตามการเคลื่อนไหวของมือจากวีดิทัศน์
- การทดสอบกรณีมือซ้าย มือขวา และมือขวาไม่บดบังกัน



รูปที่ 4-3 ผลการติดตามการเคลื่อนไหวของมือจากวีดิทัศน์ในกรณีที่ไม่เกิดการบดบังกัน
- การทดสอบกรณีมือซ้ายและมือขวาเกิดการบดบังกัน



รูปที่ 4-4 ผลการติดตามการเคลื่อนไหวของมือจากวีดิทัศน์ในกรณีที่มีมือเกิดการบดบังกัน
- การทดสอบกรณีใบหน้าและมือเกิดการบดบังกัน

4.3 ทดสอบการสกัดคุณลักษณะสำคัญของรูปร่าง และท่าทางของมือ

คุณลักษณะสำคัญของมือที่สกัดจากรูปร่างของมือคือ จุดสำคัญของรูปร่างมือเพื่อใช้ในการแยกกลุ่มท่ามือ ค่าฮิสโตแกรมต์เพื่อใช้ในการจำแนกรูปร่างของท่ามือ ความหนาแน่นของขอบภาพ (Edge density) เพื่อใช้ในการจำแนกหน้ามือและหลังมือ



รูปที่ 4-5 ผลการสกัดคุณลักษณะสำคัญของมือจากวีดิทัศน์

4.4 ทดสอบคุณลักษณะสำคัญของมือที่สกัดได้กับการรู้จำรูปร่างท่ามือ

ได้ทำการทดสอบกับท่ามือพยัญชนะ สระและวรรณยุกต์ภาษาไทยจากวีดิทัศน์ภาษามือไทย 87 ท่าที่แตกต่างกัน โดยสามารถจำแนกออกเป็นท่ามือเดี่ยวได้ทั้งสิ้น 35 ท่ามือ ซึ่งได้ทำการทดลองกับวีดิทัศน์ภาษามือไทยจำนวน 126 วีดิทัศน์ สามารถสกัดคีย์เฟรมได้ทั้งสิ้น 216 คีย์เฟรม โดยมีความแม่นยำรวมในการรู้จำท่ามืออยู่ทั้งสิ้น 77.78%



รูปที่ 4-6 ท่ามือภาษาไทยที่ใช้ในการทดสอบ

การรู้จำกลุ่มท่ามือที่ 1 ความแม่นยำเฉลี่ย 63.3%

ตารางที่ 4-1 เมตริกซ์ความสับสน (Confusion Matrix) ของข้อมูลท่ามือกลุ่มที่ 1

ท่ามือ	1	2	3	4	5	6	7	8	9	10
1	0.67	0	0	0	0.165	0	0	0.165	0	0
2	0	0.5	0	0	0.5	0	0	0	0	0
3	0	0	0.83	0	0	0	0.17	0	0	0
4	0.165	0	0	0.5	0.33	0	0	0	0	0
5	0	0	0	0	0.83	0	0	0	0	0.17
6	0	0.25	0	0.25	0	0.5	0	0	0	0
7	0	0	0	0	0.5	0.17	0.33	0	0	0
8	0	0	0	0	0.33	0	0	0.67	0	0
9	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0.5	0	0	0	0	0.5

การรู้จำกลุ่มท่ามือที่ 2 83.25%

ตารางที่ 4-2 เมตริกซ์ความสับสน (Confusion Matrix) ของข้อมูลท่ามือกลุ่มที่ 2

ท่ามือ	11	12	13	14	15	16	17	18
11	0.83	0	0	0	0	0.17	0	0
12	0	0.67	0	0.33	0	0	0	0
13	0	0	0.83	0	0.17	0	0	0
14	0	0.165	0.165	0.67	0	0	0	0
15	0	0	0.17	0	0.83	0	0	0
16	0.17	0	0	0	0	0.83	0	0
17	0	0	0	0	0	0	1	0
18	0	0	0	0	0	0	0	1

การรู้จำกลุ่มท่ามือที่ 3 ความแม่นยำเฉลี่ย 85.57%

ตารางที่ 4-3 เมตริกซ์ความสับสน (Confusion Matrix) ของข้อมูลท่ามือกลุ่มที่ 3

ท่ามือ	19	20	21	22	23	24	25
19	0.83	0	0	0	0	0	0.17
20	0	0.83	0	0	0.17	0	0.17
21	0	0	1	0	0	0	0
22	0	0	0	1	0	0	0
23	0	0.33	0	0	0.67	0	0
24	0	0	0	0	0	0.83	0.17
25	0	0	0	0	0	0.17	0.83

การรู้จักกลุ่มท่ามือที่ 4 ความแม่นยำเฉลี่ย 75.0%

ตารางที่ 4-4 เมตริกซ์ความสับสน (Confusion Matrix) ของข้อมูลท่ามือกลุ่มที่ 4

ท่ามือ	26	27	28	29	30	31	32
26	0.87	0	0	0	0	0	0.17
27	0	0.87	0	0.17	0	0	0
28	0	0	1	0	0	0	0
29	0	0.5	0	0.5	0	0	0
30	0	0	0	0	0.67	0.165	0.165
31	0	0	0	0	0	0.67	0.33
32	0	0	0	0	0	0.33	0.67

การรู้จักกลุ่มท่ามือที่ 5 ความแม่นยำเฉลี่ย 91.5%

ตารางที่ 4-5 เมตริกซ์ความสับสน (Confusion Matrix) ของข้อมูลท่ามือกลุ่มที่ 5

ท่ามือ	33	34
33	0.83	0.17
34	0	1

การรู้จักกลุ่มท่ามือที่ 6 ความแม่นยำ 100% เนื่องจากมีท่ามืออยู่ในกลุ่มเพียงท่ามือเดียว
ดังนั้นหากสามารถแบ่งกลุ่มท่ามือได้อย่างถูกต้องจะสามารถรู้จักได้ถูกต้องอย่างแน่นอน

ตารางที่ 4-6 เมตริกซ์ความสับสน (Confusion Matrix) ของข้อมูลท่ามือกลุ่มที่ 6

ท่ามือ	35
35	1

ตารางที่ 4-7 ตารางผลความทดสอบความแม่นยำการทำงานของระบบ

การทดสอบ	ความแม่นยำ
ความสามารถในการแยกกลุ่มท่ามือ	87.50%
ความสามารถในการเปรียบเทียบรูปร่างในกลุ่มท่ามือ	88.89%
ความสามารถในการรู้จำโดยรวม	77.78%



รูปที่ 4-7 ผลการรู้จำจากคุณลักษณะสำคัญของมือที่ได้จากวีดิทัศน์

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

เป้าหมายของวิทยานิพนธ์ฉบับนี้คือ การตรวจจับและสกัดคุณลักษณะสำคัญของมือ สำหรับการสะกดนิ้วมือภาษาไทย ผู้วิจัยได้ดำเนินการพัฒนาขั้นตอนวิธีตามที่ได้กล่าวไว้เบื้องต้น โดยผลการทดสอบคือ สามารถติดตามตำแหน่งของมือได้อย่างถูกต้องที่ความเร็วของภาพนำเข้า (frame rate) ไม่ต่ำกว่า 25 เฟรมต่อวินาที หากความเร็วของภาพนำเข้า (frame rate) ต่ำกว่า 25 เฟรมต่อวินาที จะทำให้การติดตามตำแหน่งไม่ถูกต้องเนื่องจากตำแหน่งของวัตถุในแต่ละเฟรมแตกต่างกันมาก ทำให้การกำหนดสมมติฐานตำแหน่งของวัตถุไม่ถูกต้องและคลาดเคลื่อน ทำให้ผลการติดตามคลาดเคลื่อนไป

การทดสอบการติดตามตำแหน่งของมือกับพื้นหลังที่แตกต่างกันนั้น สามารถกระทำการติดตามได้ดีในพื้นที่ที่มีสี หรือความต่างของสีแตกต่างกับสีของผิวหนัง เนื่องจากใช้การติดตามตำแหน่งจากสีของผิวหนังเป็นหลัก ดังนั้นสภาพแสงที่เหมาะสมคือ สภาพแสงภายในอาคาร ซึ่งให้ผลลัพธ์ของการสกัดสีผิวหนังที่ดี

การทดสอบการติดตามตำแหน่งของมือเมื่อเกิดการบิดบัง สามารถสกัดรูปร่างของมือ ขณะเกิดการบิดบังได้ดี หากรูปร่างของมือก่อนการบิดบังและหลังการบิดบังเปลี่ยนแปลงไปไม่มาก เนื่องจากมีการเก็บรูปร่างของมือ (hand template) ก่อนถูกบิดบังไว้ ภายใต้สมมติฐานที่ว่า รูปร่างของมือขณะถูกบิดบังจะมีการเปลี่ยนแปลงไปไม่มากเมื่อเทียบกับรูปร่างของมือก่อนถูกบิดบัง แต่หากรูปร่างของมือมีการเปลี่ยนแปลงขณะถูกบิดบัง จะทำให้การสกัดรูปร่างของมือคลาดเคลื่อน (รูปที่ 4-2)

ผลการสกัดคุณลักษณะสำคัญของมือ ได้ทำการทดลองกับท่ามือตัวอักษรไทยจำนวน 126 วิธีทัศน ซึ่งสามารถจำแนกออกเป็นท่ามือหนึ่งได้ 35 ท่ามือ ซึ่งจากการทดสอบสามารถสกัดคีย์เฟรมจากไฟล์วิธีทัศน 126 วิธีทัศน ออกมาได้ทั้งหมด 216 คีย์เฟรม โดยความแม่นยำในการสกัดแยกกลุ่มท่ามือจากคุณลักษณะจุดสำคัญของรูปร่างคือ 88.89% โดยความผิดพลาดที่เกิดขึ้นมา

จากการใช้การเปรียบเทียบแม่แบบ (Template Matching) เนื่องจากการเก็บบันทึกแม่แบบ (Template) ของรูปร่างมือจะเกิดขึ้นก่อนเกิดการบิดงอโดยคาดการณ์ว่ารูปร่างของท่ามือจะเปลี่ยนแปลงไปไม่มากระหว่างเกิดการบิดงอ ดังนั้นในกรณีที่รูปร่างของมือมีการเปลี่ยนแปลงขณะที่มีการบิดงอจะทำให้แม่แบบ (Template) ที่เก็บไว้ไม่ถูกต้อง ทำให้การสกัดหาคุณลักษณะสำคัญของรูปร่างท่ามือไม่ถูกต้อง (รูปที่ 5-1) ทำให้การแบ่งกลุ่มท่ามือ และการรู้จำท่ามือไม่ถูกต้อง



รูปที่ 5-1 ผลการเปรียบเทียบแม่แบบ (Template Matching) ที่ไม่ถูกต้อง

จากเมตริกซ์ความสับสนเพื่อจำแนกความสามารถในการรู้จำของแต่ละท่ามือ พบว่าท่ามือในกลุ่มที่ 1 (ตาราง 4-1) จะให้ค่าความแม่นยำในการจำแนกท่ามือน้อยที่สุด เนื่องจากเป็นกลุ่มที่ท่ามือมีรูปร่างที่ใกล้เคียงกันมาก ทำให้คุณลักษณะสำคัญที่สกัดได้ซึ่งได้แก่ ฮูโมเมนต์และจุดสำคัญของรูปร่างที่ใช้จำแนกรูปร่างท่ามือนั้นไม่สามารถแยกแยะท่ามือได้เท่าที่ควร โดยให้ความแม่นยำโดยเฉลี่ยอยู่ที่ 63.3% แต่อย่างไรก็ตามในกลุ่มที่ท่ามือมีรูปร่างที่แตกต่างกันมาก เช่น ท่ามือในกลุ่มที่ 3 (ตาราง 4-3) คุณลักษณะสำคัญของรูปร่างก็ให้ความแม่นยำที่สูง โดยความแม่นยำโดยเฉลี่ยอยู่ที่ 85.57% โดยความแม่นยำในการรู้จำท่ามือรวมของทุกกลุ่มท่ามืออยู่ที่ 77.78 %

5.2 ข้อเสนอแนะ

ข้อเสนอแนะสำหรับนำวิทยานิพนธ์ฉบับนี้ไปพัฒนาต่อไป

- การแยกสถานะของการแสดงท่ามือ

ในขั้นตอนของการแยกแยะการแสดงท่ามือจะพิจารณาเฉพาะความเร็วของมือทำให้สามารถพิจารณาได้เฉพาะท่ามือนิ่งเท่านั้น ดังนั้นหากต้องการกรอบงานสามารถพิจารณาท่ามือที่มีการเคลื่อนไหวได้ควรใช้ทิศทางในการเคลื่อนที่ของมือในการพิจารณาท่ามือด้วย เพื่อสามารถใช้จำความหมายของท่ามือที่มากขึ้น

- การสกัดค่าจากการสกัดท่ามือ

ผลลัพธ์ที่ได้จากกรอบงานคือคุณลักษณะสำคัญของท่ามือสำหรับใช้ระบุท่าทางภาษามือไทยของพยัญชนะ สระ และวรรณยุกต์ ซึ่งสามารถนำไปใช้ในการสกัดเป็นค่าเพื่อใช้ในการติดต่อสื่อสารได้

รายการอ้างอิง

- [1] Kuch, J.J., Human computer interaction via the human hand: a hand model, 2(1994): 1256-1256.
- [2] S. Bilal, R. Akmeliawati, M. J. El Salami, A. A. Shafie, Vision-based Hand Posture Detection and Recognition for Sign Language-A study, ICOM, 2011.
- [3] Omer Rashid et all, Posture Recognition using Combined Statistical and Geometrical Feature Vectors based on SVM, International Journal of Information and Mathematical Sciences 6(2010):1.
- [4] T.Tuytelaars, K. Mikolajczyk, Local Invariant Feature Detectors: A Survey, Foundation and Trends in Computer Graphics and Vision,3,3(2007):177-280.
- [5] G. Gradski, A. Kaehler, Learning OpenCV Computer vision with the OpenCV Library,2008.
- [6] Aloupic, Greg and Bohdan Kaluany. The Three Coins Algorithm for Convex Hulls of Polygons, <http://cgm.cs.mcgill.ca/~beezer/cs507/main.html>]. 1999.
- [7] N. Habili, Cheng Chew Lim, A. Moini, Segmentation of the Face and Hands in Sign Language Video Sequences using Color and Motion Cues, TCSVT, 14, 8(2004):1086-1097.
- [8] A. A. Argyros, M. I. A. Lourakis, Real-Time Tracking of Multiple Skin-Colored Objects with a Possibly Moving Camera, ECCV,2004.
- [9] Gonzalez, M., C. Collet and R. Dubot (2010). Head Tracking and Hand Segmentation during Hand over Face Occlusion in Sign Language Workshop on Sign Gesture and Activity (SGA), 11th European Conference on Computer Vision (ECCV), 2010.
- [10] P. Smith, N.D.V. Lobo, M. Shah: Resolving hand over face occlusion, 25, 9(2007): 1432 – 1448.
- [11] W. Tan, C. Wu, S. Zhao, S. Chen, Hand Extraction Using Geometric Moments Based on Active Skin Color Model, ICIS, 4(2009):468-471.

- [12] Lae-Kyoung Lee, Su-Yong An, Se-Young Oh, Robust fingertip extraction with improved skin color segmentation for finger gesture recognition in Human-robot interaction, Evolutionary Computation (CEC), 2012.
- [13] N. Tanibata, N. Shimada, Y. Shirai, Extraction of Hand Features for Recognition of Sign Language Words, International on Vision Interface, 2002:319-398.
- [14] N. Soontranon, S. Aramvith, T.H. Chalidabhongse, Face and hands localization and tracking for language recognition, ISCIT, 2(2004):1246-1251.
- [15] N. Soontranon and S. Aramvith, T.H. Chalidabhongse, Improved Face and Hand Tracking for Sign Language Recognition, the International Conference on Information Technology: Coding and Computing (ITCC'05), 2005.
- [16] U.S. Rokade, D. Doye, M. Kokare, Hand Gesture Recognition Using Object Based Key Frame Selection, ICDIP, 2009:288-291.
- [17] Hanning Zhou, D.J. Lin, T.S. Huang, Static Hand Gesture Recognition based on Local Orientation Histogram Feature Distribution Model, CVPR, 2004:161
- [18] K.W.Kwok, K.C.Lo, Recognition of curved shapes using geometric invariants, ICOSP, 2(1998):1096-1099
- [19] D.M. Gavrilu, The Visual Analysis of Human Movement: A Survey. Computer Vision and Image Understanding, 73, 1(1999):82-98.
- [20] D. Comaniciu, V. Ramesh, P. Meer, Real-Time Tracking of Non-Rigid Objects using Mean Shift, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2(2000):142-151.
- [21] G.R. Bradski, Computer Vision Face Tracking For Use in a Perceptual User Interface, Intel Technology Journal, 2, 2(1998):15-26.
- [22] Ahsan, M.R., Electromyography (EMG) signal based hand gesture recognition using artificial neural network (ANN), ICOM, 2011:1-6.

ประวัติผู้เขียนวิทยานิพนธ์

นายธนเดช สุขศิลป์ เกิดเมื่อวันที่ 27 มกราคม 2530 ในจังหวัดสุราษฎร์ธานี จบการศึกษาชั้นมัธยมศึกษาที่โรงเรียนสุราษฎร์ธานี จังหวัดสุราษฎร์ธานี และจบการศึกษาระดับปริญญาบัณฑิตจากภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้า ธนบุรี เมื่อปีการศึกษา 2552