การพัฒนาระบบรู้จำอักษรภาษามือไทยด้วยแบบจำลองฮิดเดนมาร์คอฟ

Development of a Thai Sign Language Recognition System using Hidden Markov Models

บุษรา สกุลสุจิราภา

Budasra Sakulsujirapa

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์
มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Computer Engineering
Prince of Songkla University
2554

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์

เลขหมู่ TA1637 ส175 2554 ฉ.2
Bib Key 351274
18 ต.ค. 2554

ชื่อวิทยานิพนธ์ การพัฒนาระบบรู้จำอักษรภาษามือไทยด้วยแบบจำลองฮิดเดนมาร์คอฟ
ผู้เขียน นางสาวบุษรา สกุลสุจิราภา
สาขาวิชา วิศวกรรมคอมพิวเตอร์

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

.....
(ดร.อนันท์ ชกสุริวงค์)

คณะกรรมการสอบ

.....ประธานกรรมการ
(ดร.นิคม สุวรรณวร)

.....กรรมการ
(ดร.อนันท์ ชกสุริวงค์)

.....กรรมการ
(ดร.วชรินทร์ แก้วอภิชัย)

.....กรรมการ
(รองศาสตราจารย์ ดร.ปกรณ์ แก้วตระกูลพงษ์)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษา ตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์

.....
(ศาสตราจารย์ ดร.อมรรัตน์ พงศ์ดารา)
คณบดีบัณฑิตวิทยาลัย

ชื่อวิทยานิพนธ์ การพัฒนาระบบรู้จำอักษรภาษามือไทยด้วยแบบจำลองฮิดเดนมาร์คอฟ
ผู้เขียน นางสาวบุษรา สกุลสุจิราภา
สาขาวิชา วิศวกรรมคอมพิวเตอร์
ปีการศึกษา 2553

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้นำเสนอระบบรู้จำภาษามือสำหรับตัวอักษรภาษาไทยโดยใช้เทคนิคออปติคัล โฟลว์ ในการคัดเลือกเฟรมภาพซึ่งเป็นท่าภาษามือพื้นฐานภาษาอังกฤษจากไฟล์วีดีทัศน์ จากนั้นใช้เทคนิคการประมวลผลภาพโดยการพิจารณาจากรูปร่างของมือเพื่อระบุว่าเป็นท่าภาษามือพื้นฐานท่าใด ในขั้นตอนสุดท้ายจะใช้เทคนิคฮิดเดนมาร์คอฟในการรู้จำลำดับท่าพื้นฐานที่ได้ว่าตรงกับตัวอักษรภาษามือไทยตัวใด

ข้อมูลที่ใช้สอนระบบ เป็นไฟล์วีดีทัศน์ที่บันทึกจากผู้ทดสอบอาสาสมัครจำนวน 8 คน ซึ่งมีลักษณะของมือที่แตกต่างกัน จำนวน 840 ไฟล์วีดีทัศน์ โดยบันทึกไว้เฉพาะส่วนของมือในขณะที่อาสามสมัครกำลังแสดงท่าทางของตัวอักษรภาษามือไทย การบันทึกข้อมูลทั้งหมดได้กระทำขึ้นภายใต้สภาวะแวดล้อมคงที่ จากนั้นทำการทดสอบระบบโดยใช้ไฟล์วีดีทัศน์จำนวน 420 ไฟล์ แบ่งเป็น 2 ประเภท คือ สุ่มเลือกจากไฟล์ที่เคยผ่านการสอนระบบจำนวน 210 ไฟล์ และไฟล์ที่ไม่เคยผ่านการสอนระบบจำนวน 210 ไฟล์ โดยไฟล์วิดีทัศน์ที่ไม่เคยผ่านการสอนระบบบันทึกจากผู้ทดสอบอาสาสมัครจำนวน 5 คนที่ไม่ซ้ำกันกับผู้ทดสอบอาสาสมัครชุดแรก จากผลการทดลองพบว่า เมื่อทดสอบระบบกับข้อมูลที่เคยผ่านการสอนระบบ ระบบจะมีความถูกต้อง 81.43% และเมื่อทดสอบระบบกับชุดข้อมูลที่ไม่เคยผ่านการสอนระบบ ระบบจะมีความถูกต้อง 71.9%

ผลจากการทดลองสรุปได้ว่า การใช้เทคนิคฮิดเดนมาร์คอฟเพื่อช่วยในการรู้จำลำดับท่าพื้นฐานสามารถเพิ่มประสิทธิภาพของระบบได้มากกว่าวิธีเปรียบเทียบโดยตรงทีละลำดับท่าพื้นฐานถึง 18.69 % การพัฒนาในอนาคตควรมุ่งความสนใจไปที่การปรับปรุงการคัดเฟรมภาพ การเพิ่มข้อมูลให้แก่ระบบวิธีนี้น่าจะสามารถเพิ่มสมรรถนะของระบบให้ดีขึ้น

คำสำคัญ: ระบบรู้จำภาษามือไทย/แบบจำลองฮิดเดนมาร์คอฟ

Thesis Title	Development of a Thai Sign Language Recognition System using Hidden Markov Models
Author	Miss. Budsara Sakulsujirapa
Major Program	Computer Engineering
Academic Year	2010

ABSTRACT

This thesis proposes a sign language recognition system for Thai alphabets. Optical Flow is used to select key frames that consist of 1-3 English sign language postures from a video file. After that, using image processing techniques is employed by considering the shape of hand to identify the basic posture in each key frame and the final step uses Hidden Markov Model for recognizing basic posture sequence and translate to Thai sign alphabet.

Data for training system were 840 video files saved from 8 volunteers. Each volunteer had different characteristics of the hand capture in a controled environment. All vedio were recorded showing only part of hand. This system was tested with 420 video files that be divided into 2 categories: 210 video files randomly selected the traning set and 210 video files from obtained 5 new volunteers. The system achieved accuracies of 81.43% and 71.9% for the data selected from the training set and the new data set, respectively.

The results of testing showed that using Hidden Markov Model to recognize the basic sequence of posture increased the system performance for 18.69% over the direct posture matching method. Further work includes improving the key frame selection. Performance of the system should be increased by adding more information.

KEYWORDS: Thai Sign Language Recognition/Hidden Markov Model

กิตติกรรมประกาศ

ขอขอบพระคุณ ดร.อนันท์ ชกสุริวงค์ ประธานกรรมการที่ปรึกษางานวิจัย และ รศ.ดร. มนตรี กาญจนะเดชะที่ได้เสียสละเวลาในการให้คำปรึกษา แนวคิดในการทำวิจัย รวมถึงการช่วยเหลือแก้ไขปัญหาที่เกี่ยวกับการวิจัย ตลอดจนตรวจสอบและแก้ไขวิทยานิพนธ์ให้ดำเนินไปอย่างลุล่วงสมบูรณ์

ขอขอบพระคุณ ดร.นิคม สุวรรณวร ที่ได้กรุณาเสียสละเวลาในการให้คำปรึกษา และเป็นประธานกรรมการสอบวิทยานิพนธ์ อีกทั้งตรวจทานแก้ไขวิทยานิพนธ์ให้มีความสมบูรณ์

ขอขอบพระคุณ ดร.วชรินทร์ แก้วอภิชัย ที่กรุณาเสียสละเวลาในการให้คำปรึกษา และเป็นกรรมการสอบวิทยานิพนธ์ อีกทั้งตรวจทานและแก้ไขวิทยานิพนธ์ให้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณ รศ.ดร.ปกรณ์ แก้วตระกูลพงษ์ ที่กรุณาเสียสละเวลาเป็นกรรมการสอบวิทยานิพนธ์ อีกทั้งตรวจทานและแก้ไขวิทยานิพนธ์ให้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณครูณฐพรรณ พรมแสง ที่ได้กรุณาเสียสละเวลาให้ความรู้คำปรึกษา และตรวจสอบความถูกต้องของท่าตัวอักษรภาษามือไทย

ขอขอบพระคุณ บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่ ที่ให้การสนับสนุนทุนในการทำวิจัยและให้ความช่วยเหลือด้านการประสานงานต่างๆ

ขอขอบพระคุณ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์ ที่กรุณาให้ทุนก้นกุฏิแก่ข้าพเจ้า

ขอขอบพระคุณ คณาจารย์ บุคลากร และนักศึกษาปริญญาโทภาควิชาวิศวกรรมคอมพิวเตอร์ทุกคนที่ได้ให้คำปรึกษา และกำลังใจในการทำงานเป็นอย่างดีเสมอมา

และสุดท้ายข้าพเจ้าขอน้อมรำลึกถึงพระคุณของ บิดามารดา และครอบครัว ที่ส่งเสริมและสนับสนุนข้าพเจ้าในทุกๆเรื่องตลอดมาจนสำเร็จการศึกษา

บุษรา สกุลสุจิราภา

สารบัญ

หน้า

สารบัญ ..... (6)

รายการตาราง ..... (11)

รายการภาพประกอบ ..... (12)

สัญลักษณ์คำย่อและตัวย่อ ..... (14)

บทที่ 1 บทนำ 1

1.1 วัตถุประสงค์ 2

1.2 ขอบเขตของการวิจัย 2

1.3 ขั้นตอนและวิธีดำเนินงานวิจัย 2

1.4 ประโยชน์ที่คาดว่าจะได้รับ 3

1.5 ทรัพยากรที่ใช้ในการพัฒนาระบบ 3

1.6 ภาพรวมของระบบ 3

บทที่ 2 ทฤษฎีและหลักการ 5

2.1 ภาษามือไทย 5

2.2 องค์ประกอบของภาษามือไทย 5

2.2.1 รูปมือ (Handshape) 6

2.2.2 ตำแหน่งของมือ (Location) 6

2.2.3 การเคลื่อนไหว (Movement) 6

2.2.3.1 จังหวะการเคลื่อนไหวขั้นพื้นฐาน (Rhythms of basic movements) 6

2.2.3.2 การเคลื่อนไหวที่ซ้ำกัน (Repeated movements) 6

2.2.3.3 การเคลื่อนไหวที่มีการสั่น การขยับและการหมุนปลายแขน (Vibrating and nodding movements, and forearm rotation) 7

2.2.4 การหันของฝ่ามือ (Plam orientation) 7

สารบัญ (ต่อ)

หน้า

2.2.5 การแสดงออกท่างสีหน้า (Facial expression) 7
2.3 อักษรภาษามือไทย 7
2.4 การประมวลผลภาพ 9
2.5 ภาพดิจิทัล 10
2.5.1 รูปแบบของภาพ (Image Format) 10
2.5.1.1 JPEG (Joint Photographic Experts Group) 10
2.5.1.2 GIF (Graphic Interchange Format) 10
2.5.1.3 BMP (Bitmap) 10
2.5.2 ความละเอียดของภาพ (Image Resolution) 11
2.5.3 รูปร่างของภาพ 11
2.6 แบบจำลองสี 11
2.6.1 แบบจำลองสี RGB 11
2.6.2 แบบจำลองสี CMY 12
2.6.3 แบบจำลองสี CMYK 12
2.6.4 แบบจำลองสี HIS 12
2.7 กระบวนการทางด้านการประมวลผลภาพด้วยคอมพิวเตอร์ 12
2.7.1 ขั้นตอนพื้นฐานของการประมวลผลภาพดิจิตอล 13
2.7.1.1 การได้มาของข้อมูลภาพ (Image Acquisition) 13
2.7.1.2 การประมวลผลภาพเบื้องต้น (Image Preprocessing) 13
2.7.1.3 การแบ่งแยกข้อมูลภาพ (Image Segmentation) 14
2.7.1.4 การแสดงตัวแทนและอธิบายข้อมูล (Representation and Description) 14
2.7.1.5 การรู้จำและการแปลความหมาย (Recognition and Interpretation) 14

สารบัญ (ต่อ)

หน้า

2.7.2 เทคนิคการประมวลผลภาพเบื้องต้น (Image Preprocessing Technique).....15
2.7.2.1 การแปลงภาพสีเป็นภาพระดับสีเทา (Gray-Scale Image Transform).....15
2.7.2.2 ตัวกรองทำให้เรียบ (Smoothing Filters).....15
2.8 ออปติคัล โฟลว์ (Optical Flow).....16
2.9 การหามุมด้วยวิธีแฮร์ริส (The Harris Corner Detector).....16
2.10 Normalized cross-correlation (NCC).....18
2.11 Random Sample Consensus.....18
2.12 แบบจำลองฮิดเดนมาร์คอฟ.....18
2.12.1 การฝึกฝนแบบจำลอง (Training).....20
2.12.2 การทดสอบการจดจำของแบบจำลอง (Evaluation).....22
บทที่ 3 การออกแบบและพัฒนาระบบ.....24
3.1 แนวคิดในการออกแบบระบบ.....24
3.2 ภาพรวมการออกแบบโปรแกรม.....25
3.3 การเก็บข้อมูลที่ใช้ในการสอน และทดสอบระบบ.....26
3.4 การเลือกเฟรมหลัก.....26
3.5 การหาลักษณะเด่นของมือ และการระบุท่าพื้นฐานโดยการหามุมที่สนใจในเฟรมหลัก.....28
3.6 การหาลักษณะเด่นของมือ และการระบุท่าพื้นฐานโดยวิธีวิเคราะห์รูปร่างมือและนิ้วมือ.....29
3.6.1 การหาขอบเขตของมือ.....31
3.6.1.1 การหาจุดล่างสุดของมือ.....31
3.6.1.2 การปรับขอบเขตของมือ.....32
3.6.2 แยกประเภทของท่ามือ.....35
3.6.3 การหาปลายนิ้วมือ.....36

สารบัญ (ต่อ)

หน้า

3.6.4 ลำดับความสูงของนิ้วมือ 36

3.6.5 การวางตัวของนิ้วมือ 37

3.6.6 การหาความกว้างของนิ้วมือ 37

3.6.7 การระบุท่าพื้นฐานของเฟรมหลัก 39

3.7 การแปลเป็นตัวอักษรภาษาไทย 41

บทที่ 4 ผลการดำเนินงานและวิเคราะห์ผล 42

4.1 การเก็บข้อมูล 42

4.2 การทดสอบการเลือกเฟรมหลัก 42

4.3 การทดสอบความถูกต้องของการหาลักษณะเด่นเพื่อนำไประบุท่าพื้นฐานของเฟรมหลัก ..44

4.3.1 วิธีการหามุมที่สนใจในเฟรมหลัก 44

4.3.2 การหาลักษณะเด่นของมือ และการระบุท่าพื้นฐาน โดยวิเคราะห์รูปร่างมือและนิ้วมือ 45

4.4 การทดสอบการแปลลำดับอักขระเป็นตัวอักษรภาษาไทย 49

4.4.1 การเปรียบเทียบโดยตรง 50

4.4.2 การรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟ 50

4.5 การทดสอบประสิทธิภาพการทำงานของระบบ 53

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 57

5.1 สรุปผล 57

5.2 ผลที่ได้จากการทำวิทยานิพนธ์ชุดนี้ 58

5.3 ปัญหาและอุปสรรคของการทำวิทยานิพนธ์ 58

5.4 ข้อเสนอแนะ 58

เอกสารอ้างอิง 60

สารบัญ (ต่อ)

หน้า

ภาคผนวก ก ผลงานตีพิมพ์เผยแพร่จากวิทยานิพนธ์ 62

International Conference on Embedded Systems and Intelligent Technology 63

ประวัติผู้เขียน 66

รายการตาราง

หน้า

ตารางที่ 2-1 รายละเอียดท่าพื้นฐานของแต่ละตัวอักษรภาษามือไทย 9
ตารางที่ 3-1 ลักษณะเด่นของท่าพื้นฐานทั้ง 24 ท่า 40
ตารางที่ 4-1 ตัวอักษรภาษามือไทยทั้งหมด 42 ตัวอักษร 43
ตารางที่ 4-2 ตัวอย่างการทำงานของขั้นตอนการเลือกเฟรมหลัก 43
ตารางที่ 4-3 ผลการประเมินการทดสอบเพื่อระบุท่าพื้นฐานจากเฟรมหลักจำนวน 720 เฟรม (Confusion matrix) 46
ตารางที่ 4-4 ผลการทดสอบการแปลลำดับอักขระท่าพื้นฐานเป็นตัวอักษรภาษามือไทย 50
ตารางที่ 4-5 ตัวอย่างผลการเปรียบเทียบโดยตรง และ การรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟ 52
ตารางที่ 4-6 ตารางแสดงความผิดพลาดของระบบ 54

รายการภาพประกอบ

หน้า

ภาพประกอบ 1-1 ภาพรวมของระบบ....4
ภาพประกอบ 2-1 ความสำคัญของตำแหน่งของท่ามือ....6
ภาพประกอบ 2-2 สัญลักษณ์แทนการเคลื่อนไหว และทิศทางของการเลื่อนไหวปกติ....6
ภาพประกอบ 2-3 สัญลักษณ์แทนการเคลื่อนที่ไปในทางเดียวหลายๆ ครั้ง....6
ภาพประกอบ 2-4 สัญลักษณ์แทนการเคลื่อนที่ไปสองทิศทางที่มีการสั่น....7
ภาพประกอบ 2-5 ความแตกต่างของการหันฝ่ามือ....7
ภาพประกอบ 2-6 ภาพแสดงท่าพื้นฐานของภาษามือไทย....8
ภาพประกอบ 2-7 ภาพแสดงท่าตัวอักษร ก....8
ภาพประกอบ 2-8 ภาพแสดงท่าตัวอักษร ข....8
ภาพประกอบ 2-9 ขั้นตอนพื้นฐานของการประมวลผลภาพดิจิตอล [7]....13
ภาพประกอบ 2-10 Optical Flow....16
ภาพประกอบ 2-11 Three hidden units in HMM....19
ภาพประกอบ 2-12 ขั้นตอนการฝึกฝนแบบจำลอง (Training)....20
ภาพประกอบ 2-13 ขั้นตอนการทดสอบการจดจำของแบบจำลอง (Evaluation)....23
ภาพประกอบ 3-1 ผลของแต่ละขั้นตอน....25
ภาพประกอบ 3-2 กระบวนการการประมวลผลภาพของการเลือกเฟรมหลัก....27
ภาพประกอบ 3-3 ผลของการหามุมด้วยวิธีแฮร์ริส....28
ภาพประกอบ 3-4 การเปรียบเทียบ จุด ต่อ จุด ของภาพจากฐานข้อมูล และภาพอินพุต....28
ภาพประกอบ 3-5 การหาความน่าจะเป็นสูงสุดของแต่และมุมของภาพทั้งสองภาพ....29
ภาพประกอบ 3-6 ตัวอย่างการเลือกคำตอบของการระบุท่าหลักโดยมุมที่สนใจในเฟรมหลัก....29
ภาพประกอบ 3-7 ขั้นตอนการหาลักษณะเด่นของมือในเฟรมหลัก....30
ภาพประกอบ 3-8 ขั้นตอนผลของการหาลักษณะเด่นของมือในเฟรมหลัก....31
ภาพประกอบ 3-9 รูปแบบการวางตัวของมือ จากการแยกประเภทครั้งที่ 1....32
ภาพประกอบ 3-10 ความผิดพลาดทั้ง 4 ประเภท ของการหาขอบเขตมือ....32
ภาพประกอบ 3-11 ขั้นตอนการหา X_{right} ใหม่ ของภาพประเภทที่1....33
ภาพประกอบ 3-12 ขั้นตอนการหา Y_{bottom} ใหม่ ของภาพประเภทที่1....34

รายการภาพประกอบ(ต่อ)

หน้า

ภาพประกอบ 3-13 ขั้นตอนการหา X_{right} ใหม่ ของภาพประเภทที่2.....34
ภาพประกอบ 3-14 ขั้นตอนการหา Y_{bottom} ใหม่ ของภาพประเภทที่2.....34
ภาพประกอบ 3-15 ขั้นตอนการหา Y_{bottom} ใหม่ ของภาพประเภทที่3.....35
ภาพประกอบ 3-16 ขั้นตอนการหา X_{right} ใหม่ ของภาพประเภทที่4.....35
ภาพประกอบ 3-17 ขอบเขตมือใหม่ที่ได้จากการปรับปรุง.....35
ภาพประกอบ 3-18 ตัวอย่างรูปแบบพิกเซลปลายนิ้วมือทั้ง 4 ทิศทาง.....36
ภาพประกอบ 3-19 ตัวอย่างการกำหนดลำดับความสูงของแต่ละนิ้วมือ.....37
ภาพประกอบ 3-20 ระดับความสูงที่ใช้ในการสแกนหาการแยกของนิ้วมือ.....37
ภาพประกอบ 3-21 ท่าตัวอักษร “M” “N” “S” “T” และ “Vowel base”.....38
ภาพประกอบ 3-22 ค่า Value $_{x(left)}$ ของแต่ละพิเซล.....38
ภาพประกอบ 3-23 ภาพแสดงบริเวณจุดสัมผัสระหว่าง 2 นิ้ว.....39
ภาพประกอบ 4-1 ผลการทำงานของการระบุท่าพื้นฐานด้วยการหามุมที่สนใจในภาพ.....45
ภาพประกอบ 4-2 ลักษณะของท่าตัวอักษร “P” และตัวอักษร “Slow” ที่ถูกต้อง.....49
ภาพประกอบ 4-3 ลักษณะของท่าตัวอักษร “P” ที่ระบบระบุผิดพลาด.....49

สัญลักษณ์คำย่อและตัวย่อ

SIFT	Scale Invariant Feature Transform
HMM	Hidden Markov Model
AVI	Audio Visual Interleave standard
JPG	Joint Photographic Expers Group
GIF	Graphic Interchange Format
BMP	Bitmap
PPI	Pixel per inch
DPI	Dot per inch
NCC	Normalized cross-correlation

บทที่ 1
บทนำ

การสื่อสารเป็นสิ่งสำคัญในการดำรงชีพของมนุษย์ เพราะนอกจากจะทำให้แต่ละบุคคลมีความเข้าใจซึ่งกันและกันแล้ว ยังเป็นวิธีที่ใช้ในการถ่ายทอดความรู้จากรุ่นหนึ่งไปสู่รุ่นหนึ่ง ทำให้เกิดการพัฒนาในด้านต่างๆ มากมาย การพูดเป็นวิธีการสื่อสารของบุคคลปกติธรรมดาที่นิยมใช้กันอย่างแพร่หลายตั้งแต่อดีตจนถึงปัจจุบัน เพราะเป็นวิธีที่สะดวกรวดเร็ว และไม่ต้องสิ้นเปลืองทรัพยากรใด แต่ในความเป็นจริงสังคมเรายังประกอบด้วยบุคคลอีกประเภทหนึ่งที่ไม่สามารถใช้ภาษาพูดได้ นั่นคือ ผู้พิการทางหู บุคคลเหล่านี้จะใช้ภาษามือในการสื่อสารซึ่งนอกจากจะแตกต่างกันในเรื่องของการให้ความหมายของคำแล้ว ยังมีโครงสร้างของภาษาที่แตกต่างกัน ทำให้เกิดปัญหาในการสื่อสารระหว่างผู้พิการทางหูกับบุคคลปกติ อีกทั้งยังทำให้ผู้พิการทางหูสูญเสียโอกาสต่างๆที่ควรได้รับ

ภาษามือ คือ ภาษาที่ใช้สื่อสารของผู้พิการทางหูโดยจะใช้สีหน้า ท่าทาง การเคลื่อนไหวของมือ เพื่อสื่อความหมายแทนการใช้เสียงแบบบุคคลธรรมดาทั่วไป เนื่องด้วยผู้พิการทางหูไม่สามารถเรียนรู้เสียงจากภาษาพูดได้ ซึ่งภาษามือในแต่ละประเทศก็จะมีความแตกต่างกันไปเช่นเดียวกับภาษาพูด ขึ้นอยู่กับประเพณี วัฒนธรรมของประเทศนั้นๆ เช่น ภาษามือจีน ภาษามืออเมริกัน ภาษามือไทย ภาษามือญี่ปุ่น เป็นต้น

ปัจจุบันได้มีการนำเทคโนโลยีการประมวลผลภาพเข้ามาประยุกต์ใช้งานเพื่อสร้างระบบการติดต่อสื่อสารทางด้านภาษามือเพิ่มขึ้น หลายงานได้พัฒนาระบบรู้จำภาษามือโดยใช้เครื่องมือพิเศษเช่น ถุงมือติดเซนเซอร์ เพื่อช่วยในการหาตำแหน่ง และวิถีการเคลื่อนไหวของนิ้วและมือ [1], [2] หรือใช้ถุงมือสีเพื่อให้ง่ายต่อการหาตำแหน่งของนิ้วและมือ [3], [4] ถึงแม้ว่าการใช้ถุงมือติดเซนเซอร์ และถุงมือสี จะช่วยทำให้ประสิทธิภาพของระบบมากขึ้น แต่ก็เป็นการสิ้นเปลืองโดยเฉพาะเซนเซอร์ ที่มีราคาแพง อีกทั้งยังยุ่งยากในการใช้งานอีกด้วย เพื่อลดปัญหาดังกล่าวจึงมีผู้พัฒนาระบบโดย ปราศจากอุปกรณ์อื่นๆ ตัวอย่างเช่น ระบบแปลภาษามือ [5] ผู้พัฒนาใช้ SIFT (Scale Invariant Feature Transform) ในการหาลักษณะเด่น และรู้จำท่ามือ โดยที่ระบบสามารถรู้จำตัวอักษรภาษามือไทยได้ 12 ตัว และ สระ 3 ตัว การรู้จำลักษณะท่าทางของภาษามือสำหรับพยัญชนะไทยโดยใช้เครือข่ายประสาทเทียม [6] งานวิจัยนี้ได้ใช้ มุมที่แปลงมาจากฮิตโทแกรม ในการหาลักษณะเด่นของมือ และรู้จำโดยใช้โครงข่ายประสาทเทียม ซึ่งสามารถรู้จำได้ 15 ตัวอักษรภาษามือไทย

ในส่วนของวิทยานิพนธ์นี้ จะเป็นการศึกษาและพัฒนาระบบรู้จำอักษรภาษามือจำนวน 42 ตัวอักษรภาษามือไทย โดยใช้วิธีรับภาพท่าทางภาษามือจากกล้องไฟล์วิดีทัศน์ แล้วทำการเลือกเฟรมหลัก เพื่อนำไปหาลักษณะเด่นของมือในเฟรมนั้นๆ หลังจากนั้นจะนำข้อมูลไปผ่านกระบวนการรู้จำตัวอักษรภาษามือไทย ด้วยแบบจำลองฮิดเดนมาร์คอฟ เพื่อให้ได้ความหมายของท่าทางนั้นๆ ออกมาโดยไม่ใช้อุปกรณ์พิเศษ เพื่อช่วยในการเรียนรู้ตัวอักษรภาษามือไทย ซึ่งเป็นพื้นฐานของคำศัพท์ภาษามือไทยคำอื่นๆ และลดช่องว่างในการสื่อสารกันระหว่างบุคคลปกติ กับผู้พิการทางหู

1.1 วัตถุประสงค์

1. เพื่อวิเคราะห์วิธีที่เหมาะสมในการสร้างระบบรู้จำภาษามือไทย
2. เพื่อพัฒนาระบบรู้จำภาษามือไทยให้มีประสิทธิภาพดี ด้วยวิธีแบบจำลองฮิดเดนมาร์คอฟ

1.2 ขอบเขตของการวิจัย

1. จำกัดคำศัพท์ที่ใช้ในการรู้จำคำศัพท์ คือ ตัวอักษรภาษาไทยจำนวน 42 ตัว(ไม่มี ฃ,ฅ)
2. วิเคราะห์ภาพจากไฟล์วีดีทัศน์
3. ควบคุมสภาวะแวดล้อมขณะการใช้ระบบ ในเรื่องของ แสง ระยะจับภาพ และความเร็วของผู้ใช้ระบบ และฉากหลัง

1.3 ขั้นตอนและวิธีดำเนินงานวิจัย

ขั้นที่ 1: ศึกษาเกี่ยวกับภาษามือ, Recognizer สำหรับ Pattern Recognition,หลักการเกี่ยวกับ Hidden Markov Model (HMM) และหลักการอื่นๆ ที่เกี่ยวข้อง

ขั้นที่ 2: จัดทำโครงร่างวิทยานิพนธ์

ขั้นที่ 3: ออกแบบระบบโดยรวม และ อัลกอริธึมของโปรแกรมในการระบุเฟรมเดี่ยว

ขั้นที่ 4: นำอัลกอริธึมที่ออกแบบมาทดสอบกับข้อมูลจริง

ขั้นที่ 5: พัฒนาระบบรู้จำด้วย Hidden Markov Model (HMM)

ขั้นที่ 6: ปรับปรุงและหาเปอร์เซ็นต์ความถูกต้องของระบบที่ทำการออกแบบ

ขั้นที่ 7: รวบรวมผลการทดสอบ สรุปผล จัดทำวิทยานิพนธ์ฉบับสมบูรณ์

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. รู้จำอักษรภาษามือไทยได้ครบทั้ง 42 ตัวอักษร (ภาษามือไทยไม่มี อักษร ฃ,ฅ)
2. สามารถประมวลผลจากผู้ใช้งาน ที่ไม่เคยสอนให้กับระบบ ได้อย่างถูกต้อง
3. สร้างแนวทางให้กับผู้ทำงานวิจัยเกี่ยวกับการรู้จำภาษามือไทย

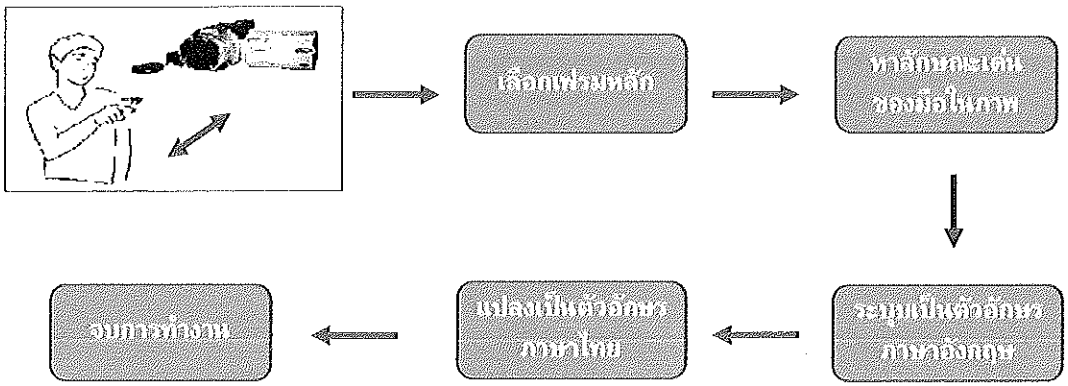
1.5 ทรัพยากรที่ใช้ในการพัฒนาระบบ

ในการวิจัยนี้เป็นการพัฒนาระบบรู้จำอักษรภาษามือไทย ซึ่งมีซอฟต์แวร์และอุปกรณ์ที่ใช้ในการออกแบบและทดสอบดังนี้

1. คอมพิวเตอร์ส่วนบุคคล (Personal Computer) ที่มีหน่วยประมวลกลาง Intel Core 2 Quad อัตราเร็ว 2.66 GHz และหน่วยความจำ RAM ขนาด 4 GB
2. กล้อง Panasonic รุ่น Nv-GS400
3. โปรแกรม Microsoft Visual C++ 6.0 และคลังโปรแกรม OpenCV สำหรับใช้ในการออกแบบและพัฒนาระบบรู้จำภาษามือไทย

1.6 ภาพรวมของระบบ

ภาพรวมของระบบที่ใช้ในการทำวิทยานิพนธ์นี้แสดงได้ดังภาพประกอบ 1-1 ซึ่งสามารถอธิบายการทำงานได้ดังนี้ ระบบจะเริ่มจากการใช้กล้องจับภาพท่าทางของผู้ใช้งาน และบันทึกจัดเก็บในรูปของไฟล์ .avi หลังจากนั้นจะส่งต่อไปยังกระบวนการเลือกเฟรมหลัก เพื่อลดจำนวนเฟรมที่ใช้ในการประมวลผล ขั้นตอนต่อไปคือ การนำเฟรมที่เลือกไว้ซึ่งอยู่ในรูปภาพนามสกุล .jpg มาหาลักษณะเด่นของมือที่อยู่ในภาพ และนำลักษณะเด่นที่ได้ไปเปรียบเทียบกับลักษณะเด่นของแต่ละท่าพื้นฐาน เพื่อระบุบว่าภาพในเฟรมนั้น เป็น ตัวอักษร ตัวเลข ภาษาอังกฤษ หรือท่าพื้นฐานตัวใด ขั้นตอนสุดท้ายคือ ทำการแปลชุดข้อมูลที่ได้จากกระบวนการที่กล่าวมาขั้นต้นเป็นตัวอักษรภาษาไทย



ภาพประกอบ 1-1 ภาพรวมของระบบ

บทที่ 2
ทฤษฎีและหลักการ

เนื้อหาบทนี้กล่าวถึงความรู้พื้นฐานและทฤษฎีต่างๆ ที่เกี่ยวข้องกับวิทยานิพนธ์ ได้แก่ ภาษามือไทย องค์ประกอบของภาษามือไทย อักษรภาษามือไทย การประมวลผลภาพ ภาพดิจิทัล รูปร่างภาพ แบบจำลองสี กระบวนการประมวลผลภาพด้วยคอมพิวเตอร์ การหาจุดเคลื่อนที่ด้วยออปติคัล โฟลว์ การหามุมด้วยวิธีแฮร์ริส การหาความคล้ายกันของจุดด้วยวิธี Normalized cross-correlation การประมาณค่าพารามิเตอร์ของแต่ละจุดด้วยวิธี Random Sample Consensus เพื่อเลือกค่าที่มีความเป็นไปได้มากที่สุด และแบบจำลองฮิดเดนมาร์คอฟ ซึ่งใช้ในการรู้จำลำดับอักขระท่าพื้นฐานของแต่ละตัวอักษรภาษามือไทย

2.1 ภาษามือไทย

ภาษามือไทยเป็นภาษาหนึ่งที่ใช้ติดต่อสื่อสารกันในชุมชนคนหูหนวก เป็นภาษาที่ใช้สายตารับรู้ความหมาย และเป็นภาษาที่มีโครงสร้างเฉพาะ แตกต่างไปจากโครงสร้างภาษาไทย ซึ่งเป็นภาษาพูดทั่วไปที่ใช้เสียงพูดสื่อความหมายและใช้หูรับรู้ความหมาย [7] ถ้าในภาษาพูด เสียงเป็นหน่วยที่เล็กที่สุด ในภาษามือก็คือ หน่วยมือ หน่วยเสียงในภาษาพูดได้แก่ หน่วยเสียงสระ และพยัญชนะ หน่วยมือในภาษามือก็ได้แก่ รูปแบบมือตำแหน่งมือและทิศทางการเคลื่อนที่มือ สำหรับภาษามือ ก็มีลักษณะคล้ายกับภาษาไทย คือมีโครงสร้างภาษาที่เป็นระบบ มีการเรียงลำดับประโยคเช่นเดียวกับภาษาพูด องค์ประกอบของภาษามือที่ทำให้เกิดความหมายแตกต่างกันออกไป คือ รูปแบบมือ ตำแหน่งฝ่ามือ หรือสันมือ และสีหน้า คนที่ได้ยินปกติสามารถใช้คำไทยได้โดยไม่ต้องคำนึงถึงภาษา และไม่จำเป็นต้องหยุดคิด แต่สำหรับภาษามือนั้นไม่เพียงใช้ภาษามือในการสื่อสาร แต่ยังใช้ร่วมกับสีหน้าและท่าทางอีกด้วย จากการวิจัยพบว่าการใช้ภาษามือของคนหูหนวกที่ได้รับการศึกษาร่วมกับคนปกติ แตกต่างไปจากของคนหูหนวกในชุมชนคนหูหนวก ซึ่งอาจทำได้การสื่อสารมีความคลาดเคลื่อนไปจากผู้ที่ต้องการจะสื่อ หรือมีความหมายคลุมเครือได้

2.2 องค์ประกอบของภาษามือไทย

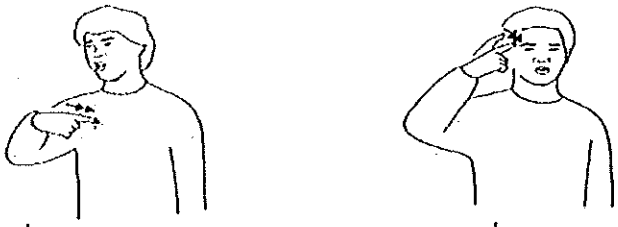
องค์ประกอบของภาษามือไทย [8] แบ่งออกเป็น 5 ประเภท

2.2.1 **รูปมือ (Handshape)**

รูปมือ คือลักษณะของมือต่างๆ เช่นการกำหมัด การแบมือ การชี้นิ้ว ในภาษามือไทย รูปมือบางท่าอาจจะถือว่าเป็นรากศัพท์ เช่น การจีบมือ การชี้นิ้ว การงอนิ้ว

2.2.2 **ตำแหน่งของมือ (Location)**

ตำแหน่งมือ คือ การใช้ท่ามือโดยที่ตำแหน่งของมืออยู่ต่างกันไป จะได้ความหมายที่แตกต่างกัน ถึงแม้ว่าลักษณะมือจะเหมือนกันก็ตาม ตัวอย่างในภาพประกอบ 2-1



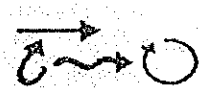
(ก) ท่ามือที่มีความหมายว่า "ฉัน" (ข) ท่ามือที่มีความหมายว่า "รู้"

ภาพประกอบ 2-1 ความสำคัญของตำแหน่งของท่ามือ

2.2.3 **การเคลื่อนไหว (Movement)**

การเคลื่อไหว คือ การเคลื่อนไหวร่างกาย, ใบหน้าและมือของผู้พูดในขณะสื่อสาร เช่นการเลื่อนมือขึ้น–ลง การเอนลำตัวเข้า-ออก จากตัวผู้ฟัง การพยักหน้า การหมุนมือเป็นต้น โดยการเคลื่อนไหวในภาษามือไทยจะแบ่งออกเป็น 3 กลุ่ม คือ

2.2.3.1 **จังหวะการเคลื่อนไหวขั้นพื้นฐาน (Rhythms of basic movements)**



ภาพประกอบ 2-2 สัญลักษณ์แทนการเคลื่อนไหว และทิศทางของการเลื่อนไหวปกติ

2.2.3.2 **การเคลื่อนไหวที่ซ้ำกัน (Repeated movements)**

ภาพประกอบ 2-3 สัญลักษณ์แทนการเคลื่อนที่ไปในทางเดียวหลายๆ ครั้ง

2.2.3.3 **การเคลื่อนไหวที่มีการสั่น การขยับและการหมุนปลายแขน (Vibrating and nodding movements, and forearm rotation)**

ภาพประกอบ 2-4 สัญลักษณ์แทนการเคลื่อนที่ไปสองทิศทางที่มีการสั่น

2.2.4 **การหันของฝ่ามือ (Plam orientation)**

การหันของฝ่ามือ คือ การที่ฝ่ามือหันเข้าหาตัวหรือหันออกจากตัว ถึงแม้ว่ามือจะมีลักษณะเดียวกันแต่การหันฝ่ามือต่างกันก็จะได้ความหมายที่ต่างกัน เช่นคำว่า "ฉัน" กับคำว่า "คุณ" ดังภาพประกอบที่ 2-5

(ก) ท่ามือที่มีความหมายว่า "ฉัน" (ข) ท่ามือที่มีความหมายว่า "คุณ"

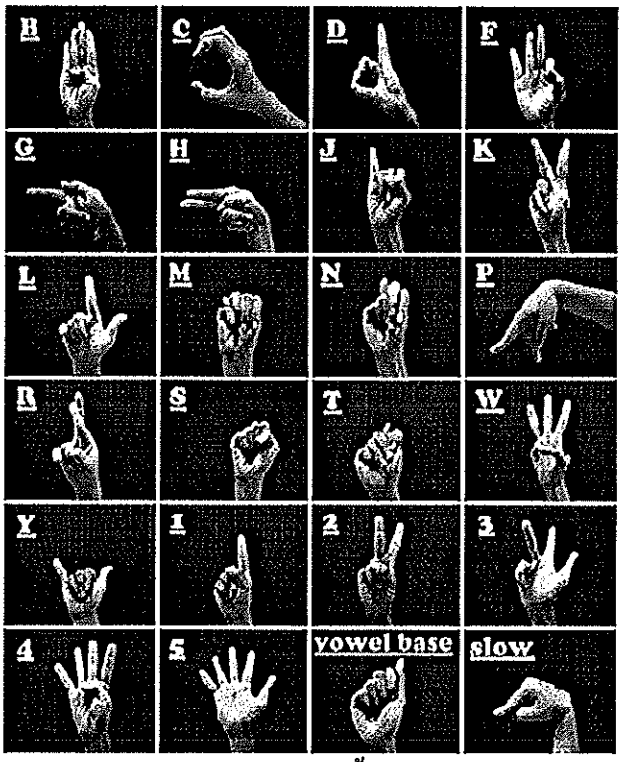
ภาพประกอบ 2-5 ความแตกต่างของการหันฝ่ามือ

2.2.5 **การแสดงออกท่างสีหน้า (Facial expression)**

การแสดงออกท่าสีหน้า คือสีหน้า แววตา ประกอบการทำท่าเช่นการเลิกคิ้ว การพองแก้ม การทำสีหน้าตกใจ เป็นต้น ซึ่งถือว่าเป็นองค์ประกอบที่สำคัญในการแสดงท่ามือ เช่น การบอกเล่า แสดงสีหน้าปกติ ประโยคคำถามแสดงสีหน้าบ่งบอกถึงความสงสัยโดยการยกคิ้วสูง หรือเป็นประโยคคำสั่งโดยการจ้องตาคู่สนทนา เป็นต้น

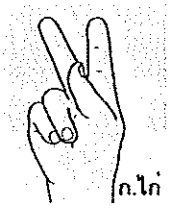
2.3 อักษรภาษามือไทย

อักษรภาษามือไทย [9] มีทั้งหมด 42 ตัว ซึ่งไม่มี "ฅ" และ "ฃ" การแสดงท่าทางของอักษรภาษามือไทยจะเป็นการนำเอาท่าของ ตัวอักษร ตัวเลข และ 2 ท่า พื้นฐานในภาษาอังกฤษมาผสมกัน มีทั้งหมด 24 ท่า ดังภาพประกอบที่ 2-6

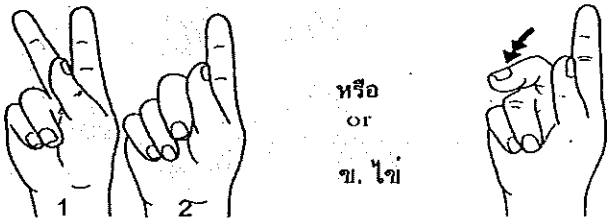


ภาพประกอบ 2-6 ภาพแสดงท่าพื้นฐานของภาษามือไทย

ตัวอย่างเช่น ตัวอักษร ก มาจากตัวอักษร "K" ในภาษาอังกฤษ ตัวอักษร "ข" มาจาก ตัวอักษร "K" ตามด้วยตัวเลข "1" และตัวอักษร "ค" จะมาจาก ตัวอักษร "K" ตามด้วยเลข "2" ดังตัวอย่างที่แสดงในรูปประกอบที่ 2-7 และ 2-8 และตารางที่ 2-1 เป็นรายละเอียดท่าพื้นฐานของแต่ละตัวอักษรภาษามือไทย



ภาพประกอบ 2-7 ภาพแสดงท่าตัวอักษร ก



ภาพประกอบ 2-8 ภาพแสดงท่าตัวอักษร ข

ตารางที่ 2-1 รายละเอียดท่าพื้นฐานของแต่ละตัวอักษรภาษามือไทย

ตัวอักษร	ประกอบด้วย (ท่าพื้นฐาน)	ตัวอักษร	ประกอบด้วย (ท่าพื้นฐาน)	ตัวอักษร	ประกอบด้วย (ท่าพื้นฐาน)
ก	K	ฒ	T+3	ม	M
ข	K+1	ณ	N+1	ย	Y
ค	K+2	ด	D	ร	R
ฆ	K+3	ต	T	ล	L
ง	N+G	ถ	T+1	ว	W
จ	J	ท	T+H	ศ	S+1
ฉ	C+H	ธ	T+H+1	ษ	S+2
ช	C+H+1	น	N	ส	S
ซ	S+Slow	บ	B	ห	H
ฌ	C+H+2	ป	P+1	ฬ	L+1
ญ	Y+1	ผ	P+2	อ	Vowel base
ฎ	D+1	ฝ	F+1	ฮ	H+1
ฏ	T+5	พ	P		
ฐ	T+2	ฟ	F		
ฑ	T+4	ภ	P+3		

การรู้จำภาษามือโดยการประมวลผลภาพจะทำการวิเคราะห์ลักษณะเด่นของมือ จากภาพเฟรมหลัก แล้วนำไปเปรียบเทียบกับลักษณะเด่นต้นแบบ เพื่อตรวจสอบว่าแต่ละเฟรมของอินพุตตรงกับตัวท่าพื้นฐานใด ซึ่งจะนำไปใช้เป็นอินพุตสำหรับกระบวนการรู้จำในขั้นตอนต่อไป

2.4 การประมวลผลภาพ

การประมวลผลภาพเป็นการนำภาพเข้าสู่การแปลงข้อมูลภาพให้อยู่ในรูปแบบข้อมูลดิจิตอลเป็นตัวเลขที่สามารถนำข้อมูลนี้ผ่านกระบวนการต่างๆ ด้วยเทคนิควิธีการประมวลผลภาพซึ่งมีเครื่องคอมพิวเตอร์ดำเนินการประมวลผลและตัดสินใจ การวิเคราะห์ข้อมูลภาพที่นำเข้ามานั้นจะอาศัยหลักวิธีการอธิบายหรือจดจำข้อมูลภาพ อินพุตจะเป็นข้อมูลภาพดิจิตอลที่ได้จากอุปกรณ์นำเข้าสัญญาณภาพดิจิตอลต่างๆ เช่น กล้องถ่ายภาพดิจิตอล กล้องวีดีทัศน์ กล้องเว็บแคม กล้องวีดีโอ ไฟล์วีดีโอ เป็นต้น และเอาท์พุตจะแสดงด้วยเครื่องหมายที่ใช้แทนข้อมูลภาพดิจิตอล

เท่านั้น ในการวิเคราะห์ภาพด้วยวิธีการประมวลผลภาพจะเลียนแบบหลักการมองเห็นของมนุษย์ นั้นคือกระบวนการทางด้านคอมพิวเตอร์วิทัศน์ (Computer Vision)

2.5 ภาพดิจิทัล

ภาพดิจิทัล คือ กลุ่มของจุดสี่เหลี่ยมขนาดเล็กที่เรียกว่า "จุดภาพ" (Pixel) ซึ่งเรียงตัวกันตามแนวตั้งและแนวนอนในลักษณะของสี่เหลี่ยมมุมฉาก โดยจะมีระดับค่าสีต่างๆกันในแต่ละจุดภาพ ทำให้เกิดเป็นภาพขึ้น

2.5.1 รูปแบบของภาพ (Image Format)

รูปแบบของภาพที่เป็นมาตรฐานของแต่ละประเภท จะเหมาะสำหรับลักษณะของภาพที่แตกต่างกัน โดยมีรูปแบบของภาพที่นิยมใช้ได้แก่

2.5.1.1 JPEG (Joint Photographic Experts Group)

ภาพประเภทนี้สามารถใช้สีได้มากถึง 16.7 ล้านสี (ความลึกของสี 24 บิต) จึงเหมาะสำหรับ ภาพถ่าย ภาพกราฟิกส์หรือภาพที่มีการไล่ระดับของสีอย่างต่อเนื่อง โดยสามารถกำหนดให้แสดงภาพแบบเบลอก่อนแล้วค่อยชัดขึ้น (Progressive) ได้ แต่ไม่สามารถใช้สร้างภาพเคลื่อนไหว หรือกำหนดให้ฉากโปร่งใสได้ JPEG เป็นมาตรฐานที่นิยมใช้มากในปัจจุบัน ภาพส่วนใหญ่สามารถบีบอัดได้ถึง 1:10 โดยคุณภาพของภาพสามารถยอมรับได้ ซึ่งสามารถทำได้หลายระดับ ยิ่งบีบอัดให้ข้อมูลมีขนาดเล็กลงเท่าใด ก็จะส่งผลให้คุณภาพของภาพน้อยลงเท่านั้น

2.5.1.2 GIF (Graphic Interchange Format)

รูปภาพประเภทนี้จะจำกัดการใช้สีสูงสุดเพียง 256 สี จึงเหมาะสำหรับภาพวาดลายเส้น ภาพการ์ตูน ภาพตัวอักษรที่มีสีทึบ หรือภาพที่มีขอบคมชัดไม่มีการไล่สี จุดเด่นของภาพประเภทนี้คือ สามารถนำไปใช้สร้างภาพเคลื่อนไหว และกำหนดให้บางส่วนของภาพโปร่งใสได้

2.5.1.3 BMP (Bitmap)

ภาพรูปแบบนี้เป็นภาพบิตแมทพื้นฐานในระบบปฏิบัติการดอส (DOS) และระบบประฏิบัติการวินโดวส์ (Windows) สามารถนำไปใช้กับโปรแกรมอื่นได้หลากหลายโปรแกรม นิยมให้กันมากกับภาพถ่ายหรือภาพวาด เพราะสามารถไล่ระดับสีเงาได้ดีที่สุด โดยสามารถกำหนดความลึกของสี ซึ่งขึ้นอยู่กับแบบจำลองสีของภาพนั้น (ความลึกของสีในแต่ละภาพ คือ จำนวนข้อมูลสีที่ปรากฏในจุดภาพของภาพจากทุกแชนเนลรวมกัน โดยมีหน่วยเป็นบิต (bit))

เนื่องด้วยวิทยานิพนธ์นี้ ข้อมูลที่ต้องการเก็บเป็นข้อมูลประเภทภาพถ่าย และต้องใช้ความละเอียดของภาพมากพอสมควรเพื่อใช้ในประมวลผล จึงได้ทำการจัดเก็บภาพในรูปแบบ JPEG

2.5.2 ความละเอียดของภาพ (Image Resolution)

ความละเอียดของภาพเป็นปัจจัยสำคัญที่ส่งผลต่อคุณภาพของภาพ การใช้รูปภาพในแต่ละงานจึงใช้ความละเอียดของภาพแตกต่างกันไปแล้วแต่ความเหมาะสม ค่าความละเอียดภาพคือ จำนวนจุดภาพต่อนิ้ว (Pixel per inch (ppi)) หรือ จำนวนจุดภาพในความกว้าง 1 นิ้ว (dot per inch (dpi)) นอกจากนี้ยังมีความหมายอีกนัยหนึ่งคือ จำนวนจุดภาพทั้งหมด (กว้าง x ยาว) ต่อ 1 ภาพ

2.5.3 รูปร่างของภาพ

รูปร่าง[10] เป็นลักษณะเฉพาะของภาพที่ใช้อธิบายถึงรูปร่าง ลักษณะ รวมถึงขนาดของวัตถุภายในภาพ ซึ่งสามารถทำให้สามารถแยกแยะระหว่างวัตถุที่มีรูปร่างแตกต่างกันออกจากกันได้ โดยการหารูปร่างของวัตถุนั้นต้องอาศัยการประมวลผลก่อนหน้า (preprocessing) หรือ ตัวกรอง (filter) ในการหาขอบของวัตถุที่สนใจภายในภาพ ซึ่งการตรวจสอบหาขอบวัตถุภายในภาพนั้นเป็นขั้นตอนหนึ่งในกระบวนการแยกส่วนภาพ (image segmentation) ปัญหาสำคัญ 2 ประการสำหรับการตรวจหารูปร่างของวัตถุ คือ การที่วัตถุถูกปิดบัง (occlusion) และ ผลกระทบจากแสง (lighting effect) ซึ่งจะส่งผลให้เกิดความผิดพลาดในการตรวจหารูปร่างได้

2.6 แบบจำลองสี

แบบจำลองสี (color model) หรือ ปริภูมิสี (color space) ถูกนิยามขึ้นโดยมีวัตถุประสงค์เพื่อใช้ในการอ้างอิงและเป็นมาตรฐานเดียวกัน แบบจำลองสีคือ ระบบพิกัด (coordinate system) และปริภูมิย่อย (subspace) ภายในระบบนั้นซึ่งถูกกำหนดขึ้นโดยแต่ละสีถูกแทนด้วยจุดๆ หนึ่งในปริภูมิสีนั้น ในวิทยานิพนธ์นี้จะกล่าวถึงแบบจำลองสี 4 แบบคือ

2.6.1 แบบจำลองสี RGB

แบบจำลองสี RGB เป็นระบบที่เกิดจากการผสมแสงสีหลัก 3 สี คือ สีแดง (Red) สีเขียว (Green) และสีน้ำเงิน (Blue) เข้าด้วยกัน ทำให้เกิดเป็นสีต่างๆขึ้น โดยแต่ละสีเกิดจากการผสมสีทั้ง 3 สีในระดับที่แตกต่างกันออกไป ซึ่งเมื่อผสมสีหลักทั้งสามที่ระดับความเข้มสูงสุดจะทำให้เกิดเป็นสีขาวขึ้น ในขณะที่เมื่อผสมสีหลักทั้งสามในระดับต่ำสุดจะเกิดเป็นสีดำ ดังนั้นจึงเรียกการผสมแบบนี้ว่า การผสมสีบวก (addtitive)

2.6.2 **แบบจำลองสี CMY**

แบบจำลองสี CMY เป็นแบบจำลองที่ใช้สำหรับการผสมหมึกพิมพ์ ประกอบด้วยสีหลัก 3 สี คือ สีฟ้า (Cyan) สีม่วงแดง (Magenta) และสีเหลือง (Yellow) ซึ่งเป็นสีเต็มเต็ม (complementary color) ของสีแดง สีเขียวและสีน้ำเงิน หลักการของการผสมสีของแบบจำลองนี้จะตรงข้ามกับแบบจำลองสี RGB โดยเรียกการผสมแบบนี้ว่าการผสมแบบลบ (subtractive) เนื่องจากเมื่อนำสีหลักทั้งสามมาผสมกันที่ระดับความเข้มสูงสุดจะได้เป็นสีดำ

2.6.3 **แบบจำลองสี CMYK**

แบบจำลองสี CMYK เป็นแบบจำลองที่ใช้สำหรับการผสมหมึกพิมพ์ ประกอบด้วยสีหลัก 3 สี คือ สีฟ้า (Cyan), สีม่วงแดง (Magenta) สีเหลือง (Yellow) และดำ (Black) โดยมีหลักการผสมสีเช่นเดียวกับแบบ CMY แต่มีสีดำเพิ่มเข้ามา เพราะเนื่องจากในทางปฏิบัตินั้นหมึกพิมพ์ไม่บริสุทธิ์ ดั้งนั้นเพื่อให้ได้สีดำสนิทจึงต้องเพิ่มสีดำเข้ามา

2.6.4 **แบบจำลองสี HIS**

แบบจำลองสี HIS เป็นแบบจำลองสีที่มีความใกล้เคียงกับระบบการมองเห็นของสายตามนุษย์ ซึ่งมีองค์ประกอบแบ่งออกเป็นสามส่วนได้แก่ Hue, Saturation และ Intensity โดยค่า Hus คือ สีสันที่สะท้อนออกมาจากวัตถุแล้วสะท้อนเข้าตามนุษย์ โดยแต่ละสีจะมีค่าตามตำแหน่งองศาในวงกลม ส่วนค่า Saturation คือ ความสดหรือความบริสุทธิ์ของสี ซึ่งจะบอกถึงสัดส่วนของสีขาวที่ผสมอยู่ในสีหลัก (Hue) และสุดท้าย Intensity คือระดับความสว่างของสี มีค่าเป็นเปอร์เซ็นตั้งแต่ 0 ถึง 100 โดยที่ค่า 0% จะมีค่าความสว่างต่ำสุดทำให้เกิดเป็นสีดำ และค่าที่ 100% คือความสว่างมากที่สุดที่จะทำให้เกิดสีตามที่กำหนดโดยค่า Hue และ Saturation

ในวิทยานิพนธ์นี้เลือกแบบจำลองสี RGB ในการประมวลผลภาพ เนื่องจากภาพดิจิทัลทั่วไปใช้แบบจำลองสีแบบ RGB ซึ่งสำหรับภาพสีขนาด 24 บิตต่อจุด แต่ละแชนแนล คือ สีแดง สีเขียวและสีน้ำเงิน จะมีระดับค่าสีได้ 256 ระดับ คือตั้งแต่ 0 ถึง 255 โดยค่าสีของแต่ละจุดภาพเกิดจากการรวมกันของทั้ง 3 แชนแนล ทำให้สามารถแสดงค่าสีได้มากถึง 16.7 ล้านสี (2^{24}) ดังนั้นก่อนนำภาพไปประมวลผลต้องนำไปทำการควอนไทซ์สีให้อยู่ในช่วงที่เหมาะสมก่อน เพื่อลดความซับซ้อน และเวลาในการประมวลผล

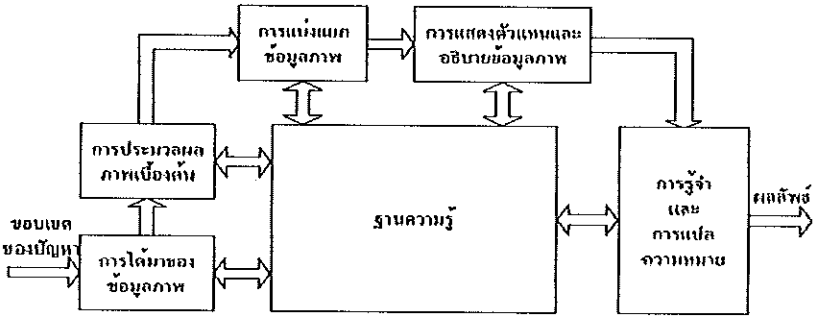
2.7 กระบวนการทางด้านการประมวลผลภาพด้วยคอมพิวเตอร์

การประมวลผลภาพด้วยคอมพิวเตอร์หรือที่นิยมเรียนกันว่า การประมวลผลภาพดิจิตอล (Digital Image Processing) เป็นกระบวนการที่มีเทคนิควิธีในการประมวลผลข้อมูลตัวเลข

ของภาพที่มีหลากหลายวิธี ซึ่งสามารถเลือกไปประยุกต์ใช้งานให้เหมาะสมกับข้อมูลภาพที่นำเข้ามาประมวลผล โดยปกติแล้วข้อมูลภาพจะมีลักษณะเด่นทางด้านรูปร่าง พื้นผิว สีสัน และโครงสร้างต่างๆ ที่แตกต่างกันไปขึ้นอยู่วัตถุและสภาพแวดล้อมโดยรอบของวัตถุ

2.7.1 ขั้นตอนพื้นฐานของการประมวลผลภาพดิจิตอล

การประมวลผลภาพดิจิตอลเป็นหลักการทำงานร่วมกันของฮาร์ดแวร์ ซอฟต์แวร์ และหลักทฤษฎีการประมวลผลภาพ โดยมีการทำงานตามขั้นตอนอย่างเหมาะสมและเป็นไปตามเทคนิคทางด้านการประมวลผลภาพ ดังภาพประกอบ 2-9



ภาพประกอบ 2-9 ขั้นตอนพื้นฐานของการประมวลผลภาพดิจิตอล [7]

2.7.1.1 การได้มาของข้อมูลภาพ (Image Acquisition)

เป็นการนำข้อมูลภาพเข้าสู่คอมพิวเตอร์ โดยอาศัยตัวรับรู้สัญญาณภาพและสามารถแปลงให้เป็นสัญญาณระบบดิจิตอลด้วยตัวรับรู้ เช่น กล้องถ่ายภาพดิจิตอล กล้องวิดีทัศน์ กล้องเว็บแคม เครื่องสแกน หรืออุปกรณ์รับสัญญาณภาพอื่นๆ ที่เหมาะสมกับระบบงานแต่ละระบบ แต่อย่างไรก็ตามรูปแบบของข้อมูลจะถูกจัดเก็บให้อยู่ในลักษณะของภาพ 2 มิติ ที่มีความสว่างของแสงหรือความคมชัดแตกต่างกันของแต่ละจุดภาพในตำแหน่งต่างๆ

2.7.1.2 การประมวลผลภาพเบื้องต้น (Image Preprocessing)

เป็นเทคนิควิธีของการปรับปรุงคุณภาพของข้อมูลภาพดิจิตอลที่ได้จากขั้นตอนการนำเข้าภาพ เพื่อให้ข้อมูลภาพมีความถูกต้องสมบูรณ์ตามความเป็นจริงก่อนนำไปประมวลผล โดยปกติแล้วการปรับปรุงคุณภาพของข้อมูลภาพดิจิตอลมีหลากหลายเทคนิค เช่น การปรับความคมชัด การปรับความสว่าง การกำจัดสัญญาณรบกวน การหมุนและการกรองช่วงความถี่ของภาพ เป็นต้น

2.7.1.3 **การแบ่งแยกข้อมูลภาพ (Image Segmentation)**

เป็นวิธีการที่จะทำให้สามารถแบ่งแยกข้อมูลภาพออกเป็นส่วนๆ หรือวัตถุ เพื่อทำให้ได้ข้อมูลที่ต้องการออกจากพื้นหลัง โดยทั่วไปผลลัพธ์ของการแบ่งแยกข้อมูลภาพจะได้เป็นข้อมูลดิบของจุดภาพที่ประกอบด้วยขอบภาพของแต่ละบริเวณหรือจุดภาพภายในบริเวณนั้น ในแต่ละกรณีจะต้องทำการแปลงข้อมูลให้มีรูปแบบที่เหมาะสมสำหรับการประมวลผลที่บังคับไว้ จะทำให้การตัดสินใจข้อมูลมีการแสดงตัวแทนของขอบภาพหรือบริเวณนั้นๆ อย่างสมบูรณ์

2.7.1.4 **การแสดงตัวแทนและอธิบายข้อมูล (Representation and Description)**

สำหรับการแสดงภาพหลังจากการแบ่งแยกข้อมูลภาพแล้ว เพื่อให้เห็นถึงลักษณะเด่นและอธิบายข้อมูลภาพของบริเวณต่างๆ ของภาพนำเข้า การเลือกตัวแทนสำหรับแสดงข้อมูลเป็นส่วนเดียวของการแก้ปัญหาสำหรับการแปลงข้อมูลดิบเป็นรูปแบบที่เหมาะสมสำหรับการประมวลผลของคอมพิวเตอร์ต่อไป วิธีการที่จะอธิบายลักษณะเด่นของข้อมูลที่สนใจถือเป็นสิ่งสำคัญ ซึ่งเรียกว่า การเลือกลักษณะเด่น (Feature Extraction) ผลลัพธ์ที่ได้จากการแยกลักษณะเด่นหรือความแตกต่างของข้อมูลที่สนใจออกจากข้อมูลอื่นๆ ก็คือ กลุ่มของวัตถุ (Class of Object) ที่ต้องการนั้นเอง

2.7.1.5 **การรู้จำและการแปลความหมาย (Recognition and Interpretation)**

เป็นขั้นตอนสุดท้ายของการประมวลผลภาพดิจิตอลหลังจากขั้นตอนการแสดงตัวแทนและอธิบายข้อมูล ก็คือ การรู้จำภาพ (Image Recognition) ซึ่งเป็นแขนงหนึ่งของการรู้จำแบบรูป (Pattern Recognition) โดยการรู้จำภาพจะต้องรู้จำแบบรูปของแต่ละภาพเป้าหมายเพื่อการให้คำตอบว่าแบบรูปของภาพนำเข้ามีความคล้ายกับแบบรูปของภาพอ้างอิงภาพใดมากที่สุด และการแปลความหมายนำไปสู่การกำหนดความหมายของชุดข้อมูลรู้จำวัตถุ การได้มาของแบบรูปอ้างอิงนั้นสามารถทำได้หลายวิธี เช่น แบบรูปอ้างอิงอาจอยู่ในรูปแบบจำลองทางคณิตศาสตร์ ซึ่งจะต้องมีวิธีเฉพาะในการเปรียบเทียบ การสร้างแบบจำลองทางคณิตศาสตร์สามารถทำได้จากขั้นตอนการฝึกฝน (Training Phrase) ซึ่งโดยทั่วไปแล้วจะต้องมีตัวอย่างภาพที่มีลักษณะเดียวกันหลายๆภาพ จากนั้นจะทำการคำนวณหาค่าลักษณะเด่นของแต่ละภาพ ซึ่งผลลัพธ์ที่ได้ก็คือ แบบรูปของภาพเหล่านั้นนั่นเอง แบบจำลองของภาพในแต่ละกลุ่มสามารถคำนวณได้จากค่าสถิติต่างๆ ของแบบรูปของภาพในกลุ่มเดียวกัน บางครั้งอาจจะอยู่ในรูปของฐานความรู้ (Knowledge Base) จำนวนมากจนมีกลุ่มข้อมูลที่เก็บไว้เป็นฐานความรู้ในรูปแบบของฐานข้อมูลความรู้ (Knowledge Database)

2.7.2 เทคนิคการประมวลผลภาพเบื้องต้น (Image Preprocessing Technique)

ขั้นตอนการประมวลภาพเบื้องต้นเป็นการปรับปรุงภาพให้มีคุณสมบัติที่เหมาะสมสำหรับนำไปสู่กระบวนการประมวลผลภาพต่อไป เทคนิควิธีการประมวลผลภาพที่นำมาประยุกต์ใช้ในงานวิจัยนี้มีหลายเทคนิคด้วยกัน

2.7.2.1 การแปลงภาพสีเป็นภาพระดับสีเทา (Gray-Scale Image Transform)

เป็นการแปลงค่าข้อมูลภาพสีให้แสดงถึงค่าความสว่างของภาพเพียงอย่างเดียว โดยปราศจากค่าข้อมูลของสีภาพ โดยทั่วไปภาพระดับสีเทาจะประกอบด้วยค่าความสว่างที่แตกต่างกัน 256 ระดับ มีค่าตั้งแต่ 0 ถึง 255 นั้นคือไล่ระดับความสว่างจากมืดไปจนขาว และความสัมพันธ์การแปลงภาพสีเป็นภาพระดับสีเทา ดังสมการที่ (2-1)

$$Gray(x,y) = 0.299R(x,y) + 0.587G(x,y) + 0.114B(x,y) \qquad (2\text{-}1)$$

โดยที่ $Gray(x,y)$ คือ ค่าระดับสีเทาที่ได้จากการคำนวณของจุดภาพของสี RGB(x , y)

$R(x,y)$ คือ ค่าสีแดงในภาพระบบสี RGB

$G(x,y)$ คือ ค่าสีเขียวในภาพระบบสี RGB

$B(x,y)$ คือ ค่าสีน้ำเงินในภาพระบบสี RGB

x, y คือ ตำแหน่งของจุดภาพ

2.7.2.2 ตัวกรองทำให้เรียบ (Smoothing Filters)

เป็นตัวกรองที่ใช้สำหรับลดภาพที่มีพื้นผิวขรุขระให้มีความเรียบขึ้นและลดสัญญาณรบกวนที่เป็นจุดเล็กๆ ข้อมูลภาพที่ผ่านตัวกรองจะมีลักษณะมัวไม่ชัด โดยในงานวิจัยนี้ใช้การกรองด้วยเกาส์เซียน (Gaussian Filtering) เป็นตัวกรองทำให้เรียบ

• **การกรองด้วยเกาส์เซียน** เป็นการกรองที่กระทำกับข้อมูลภาพขนาด 2 มิติ เพื่อการกำจัดสัญญาณรบกวนความถี่สูงและภาพผลลัพธ์จะมีลักษณะมัวไม่ชัดจะมากหรือน้อยขึ้นอยู่กับขนาดหน้าต่าง (Kernel) ที่นำมากระทำกับภาพนำเข้าในลักษณะการคอนโวลูชัน (Convolution) เช่น ขนาดหน้าต่าง 3x3, 5x5, 7x7 และ 9x9 เป็นต้น กระบวนการกรองข้อมูลด้วยวิธีนี้อาศัยหลักการฟังก์ชันเกาส์เซียนค่าเฉลี่ยยเลขคณิตเป็นศูนย์ (Zero-Mean Gaussian Function) ดังสมการที่ (2-2)

$$g(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{x^2+y^2}{2\sigma^2}} \qquad (2\text{-}2)$$

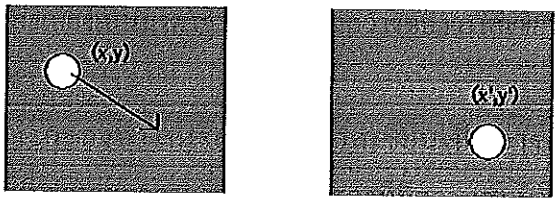
โดยที่ $g(x, y)$ คือ ภาพผลลัพธ์ของการกรองด้วยเกาส์เซียน

σ คือ ค่าเบี่ยงเบนมาตรฐานหรือช่วงกว้างของเกาส์เซียน

คุณสมบัติการกรองทำให้เรียบด้วยเกาส์เซียนเป็นการกรองความถี่ต่ำที่มีประสิทธิภาพทั้งในด้านเชิงพื้นที่ ขอบเขตความถี่ และการกรองที่มีความสมมาตรแบบหมุนหรือทุกทิศทางจะเหมือนกันจะไม่ทำการตรวจจับขอบด้านใดด้านหนึ่งมากเป็นพิเศษ

2.8 ออปติคัล โฟลว์ (Optical Flow)

ออปติคัลโฟลว์ [11] คือ เป็นการคำนวณเพื่อหาทิศทางการเคลื่อนที่ของวัตถุในหน่วยเวลาที่เปลี่ยนไป โดยใช้พิกัด x,y ของเฟรมเริ่มต้นและภาพสุดท้ายในการคำนวณ



(a)ภาพเริ่มต้น (b) ภาพสุดท้าย

ภาพประกอบ 2-10 ออปติคัล โฟลว์

2.9 การหามุมด้วยวิธีแฮร์ริส (The Harris Corner Detector)

การหามุมด้วยวิธีแฮร์ริส[13] เป็นวิธีที่ได้รับความนิยมในการนำมาใช้หาจุดที่สนใจ เนื่องมาจากมีความแน่นยำในการหา ค่าของสเกล สัญญาณรบกวน การหมุน การแปรปรวนของแสง การหามุมด้วยวิธีแฮร์ริส อาศัยฟังก์ชันความสัมพันธ์อัตโนมัติภายในของสัญญาณซึ่งสามารถวัดได้จากการ เปลี่ยนแปลงภายในของสัญญาณกับการเลื่อนโดยทิศทางที่แตกต่างกันเล็กน้อย วิธีแฮร์ริสพัฒนามาจาก Moravec [14]

ให้การเปลี่ยนแปลง $(\Delta x, \Delta y)$ และจุด (x, y) ฟังก์ชันความสัมพันธ์อัตโนมัติกำหนดค่า ได้ตามสมการที่ (2-3)

$$C(x, y) = \sum_w [I(x_i, y_i) - I(x_i + \Delta x, \Delta y_i)]^2 \qquad (2\text{-}3)$$

ที่ $I(.,.)$ แสดงฟังก์ชันของรูปภาพ (x_i, y_i) และ W คือจุดของหน้าต่างเกาเซียน มี จุดศูนย์กลางที่ (x, y)

สมการการประมาณการเปลี่ยนแปลงของภาพ

$$I(x_i + \Delta x, y_i) \approx I(x_i, y_i) + [I_x(x_i, y_i) I_y(x_i, y_i)] \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (2\text{-}4)$$

ที่ $I_x(.,.)$ และ $I_y(.,.)$ แทนด้วย x และ y ตามลำดับ
แทนสมการที่ 2-4 ลงในสมการที่ 2-3

$$\begin{aligned} C(x,y) &= \sum_w [I(x_i, y_i) - I(x_i + \Delta x, \Delta y_i)]^2 \\ &= \sum_W (I(x_i, y_i) - I(x_i, y_i) - [I_x(x_i, y_i) I_y(x_i, y_i)] \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix})^2 \\ &= \sum_W ([-I_x(x_i, y_i) I_y(x_i, y_i)] \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix})^2 \\ &= \sum_W ([I_x(x_i, y_i) I_y(x_i, y_i)] \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix})^2 \\ &= [\Delta x \Delta y] \begin{bmatrix} \sum_W (I_x(x_i, y_i))^2 & \sum_W I_x(x_i, y_i) I_y(x_i, y_i) \\ \sum_W I_x(x_i, y_i) I_y(x_i, y_i) & \sum_W (I_y(x_i, y_i))^2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \\ &= [\Delta x \Delta y] C(x,y) \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \end{aligned} \quad (2\text{-}5)$$

ที่เมตริกซ์ $C(x,y)$ วัดความเข้มของบริเวณใกล้เคียง

ให้ λ , λ_2 เป็นค่าเจาะจงของเมตริกซ์ $C(x,y)$ ซึ่งจะคงที่ แบ่งเป็น 3 กรณี

1.ถ้าทั้ง λ , λ_2 มีค่าน้อย ฟังก์ชันความสัมพันธ์อัตโนมัติภายในจะเป็นแบบราบ (เช่น การเปลี่ยนแปลงเพียงเล็กน้อยของ $C(x,y)$ ในทิศทางใดๆ) พื้นที่ของหน้าต่างรูปภาพสามารถประมาณ ได้ด้วยค่าคงที่ของความเข้ม

2.ถ้าค่าเจาะจงหนึ่งมากว่าค่าเจาะจงอีกค่าหนึ่ง ฟังก์ชันความสัมพันธ์อัตโนมัติภายในจะเป็นแบบแนวสัน นอกจากนั้นการเปลี่ยนแปลงในทิศทางหนึ่ง (ตามแนวสัน) จะเป็นสาเหตุให้เกิด การเปลี่ยนแปลงเพียงเล็กน้อยใน $C(x,y)$ และมีความหมายในการเปลี่ยนแปลงทิศทางแบบโอโท โกนอล (orthogonal direction) ซึ่งเป็นวิธีการที่ใช้ในการตัดสินขอบ

3.ถ้าค่าเจาะจงมีค่าสูงทั้งคู่ ฟังก์ชันความสัมพันธ์อัตโนมัติภายในจะเป็นแบบยอดแหลม ดังนั้นการเปลี่ยนแปลงในทิศทางใดๆ จะส่งผลในการเพิ่มความสำคัญซึ่งก็คือการชี้บอกมุม

2.10 **Normalized cross-correlation (NCC)**

Cross-correlation ในทาง signal processing คือ ตัวชี้วัดของความคล้ายคลึงกันของสัญญาณ 2 สัญญาณ โดยทั่วไปถูกใช้เพื่อหาลักษณะเฉพาะของสัญญาณที่ไม่รู้จัก เปรียบเทียบ กับสัญญาณที่รู้จัก cross-correlation เป็นฟังก์ชั่นที่แสดงความสัมพันธ์ระหว่างเวลากับสัญญาณ ที่ใช้ในการจำแนกรูปแบบ และ เทคนิคการถอดรหัส

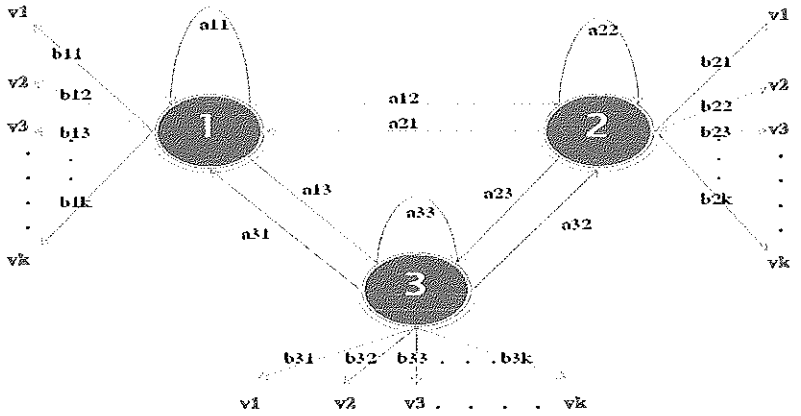
ในวิทยานิพนธ์นี้เราใช้ Cross-correlation เป็นตัวชี้วัดของความคล้ายคลึงกันของพิกเซล 2 พิกเซล เพื่อเปรียบเทียบลักษณะเฉพาะของพิกเซลในภาพที่สนใจ กับพิกเซลในภาพต้นแบบ

2.11 **Random Sample Consensus**

Random Sample Consensus คือ วิธีการประมาณค่าพารามิเตอร์ของแบบจำลองทางคณิตศาสตร์จากชุดของการสังเกตข้อมูลที่มี ค่าผิดปกติ เมื่อพิจารณาผลลัพธ์ซ้ำจะพบว่ามีเพียงบางผลลัพธ์เท่านั้นที่ความน่าจะเป็นมีแนวโน้มเพิ่มมากขึ้น ซึ่งจะเป็นการลดเป็นข้อมูลที่ไม่สอดคล้องกับรูปแบบ หรือ ข้อมูลรบกวนได้อีกด้วย

2.12 **แบบจำลองฮิดเดนมาร์คอฟ**

แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model: HMM)[15] เป็นแบบจำลองทางสถิติรูปแบบหนึ่ง สร้างอยู่บนพื้นฐานของทฤษฎีมาร์คอฟ ที่แสดงความเป็นไปได้ของการกระทำหรือว่าเหตุการณ์ที่เป็นไปได้ ฮิดเดนมาร์คอฟให้คำจำกัดความของการกระทำหรือเหตุการณ์ที่เป็นไปได้เหล่านั้นว่าเป็น สถานะซ่อน (Hidden State) ซึ่งแต่ละสถานะซ่อนจะมีเส้นโยงถึงกัน รวมทั้งมีความน่าจะเป็นกำกับ โดยการเปลี่ยนสถานะขึ้นอยู่กับความน่าจะเป็น ซึ่งมีอยู่สองชนิดคือ ค่าความน่าจะเป็นของการเปลี่ยนสถานะที่จะเกิดขึ้น (Transition Probability : a_{ij}) และ ค่าความน่าจะเป็นของการปล่อย symbol (Emission Probability : b_{ik}) เมื่อมีการย้ายสถานะ จะมีการปล่อย symbol ออกมา จากรูปที่ 13 มี 3 สถานะ a_{ij} คือ ความน่าจะเป็นของสถานะถัดไปจะเป็น j เมื่อสถานะปัจจุบันเป็น i โดยที่สามารถเปลี่ยน หรือซ้ำอยู่สถานะเดิมได้ b_{ik} คือ ความน่าจะเป็นของการปล่อยsymbol เป็น k เมื่ออยู่ที่สถานะ j และ v(k) คือ symbol ที่ปล่อยออกมาเมื่อมีการเปลี่ยนสถานะ



ภาพประกอบ 2-11 Three hidden units in HMM

เพื่อความสะดวกเราสามารถเขียน Transition Probability และ Emission Probability ให้อยู่ในรูปของเมทริกซ์ได้ดังนี้

$$A = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{23} & a_{33} \end{pmatrix} \qquad B = \begin{pmatrix} b_{11} & b_{12} & b_{13} & \cdots\cdots & b_{1k} \\ b_{21} & b_{22} & b_{23} & \cdots\cdots & b_{2k} \\ b_{31} & b_{32} & b_{33} & \cdots\cdots & b_{3k} \end{pmatrix}$$

ซึ่ง ทั้งสองเมทริกซ์ คำนวณได้จากการสอนระบบ โดยอัลกอริธึ่มที่นิยมใช้คือ Baum-Welch [12] หลังจากที่ได้แบบจำลองต้นแบบแล้ว จึงทำการทดสอบการรู้จำของแบบจำลอง โดยการคำนวณหาความน่าจะเป็นของลำดับตัวอักษรทั้ง 42 แบบจำลองตัวอักษรภาษามือไทย โดยใช้หลักการของ Forward Algorithm [12] ฮิดเดนมาร์คอฟมีการนำมาประยุกต์ ใช้กับงานทางด้านการจับคู่รูปแบบ (Pattern matching) ที่เปลี่ยนแปลงตามเวลา เช่น การรู้จำเสียงพูด รู้จำท่าทาง ฯลฯ โดยมีองค์ประกอบที่สำคัญดังนี้

- N คือ จำนวนสถานะในตัวแบบ สถานะจะเปลี่ยนไปตามเวลา t
- M คือ จำนวนของเหตุการณ์ต่อสถานะ เหตุการณ์ที่ได้จะสอดคล้องกับอินพุตที่ป้อนให้กับแบบจำลอง
- a_{ij} คือ ความน่าจะเป็นในการเปลี่ยนสถานะจากสถานะ i ไปสถานะ j
- b_{jk} คือ ค่าความน่าจะเป็นของเหตุการณ์ k ที่เกิดขึ้นในสถานะ j
- π_i ความน่าจะเป็นที่จะเกิดสถานะแรก

ส่วนประกอบต่าง ๆ ของแบบจำลองสามารถรวมเข้าด้วยกันแล้วแทนด้วยสมการ (2-3) ที่แสดงเป็นเซตของตัวแปรที่เสร็จสมบูรณ์ของแบบจำลองฮิดเดนมาร์คอฟในการจดจำ

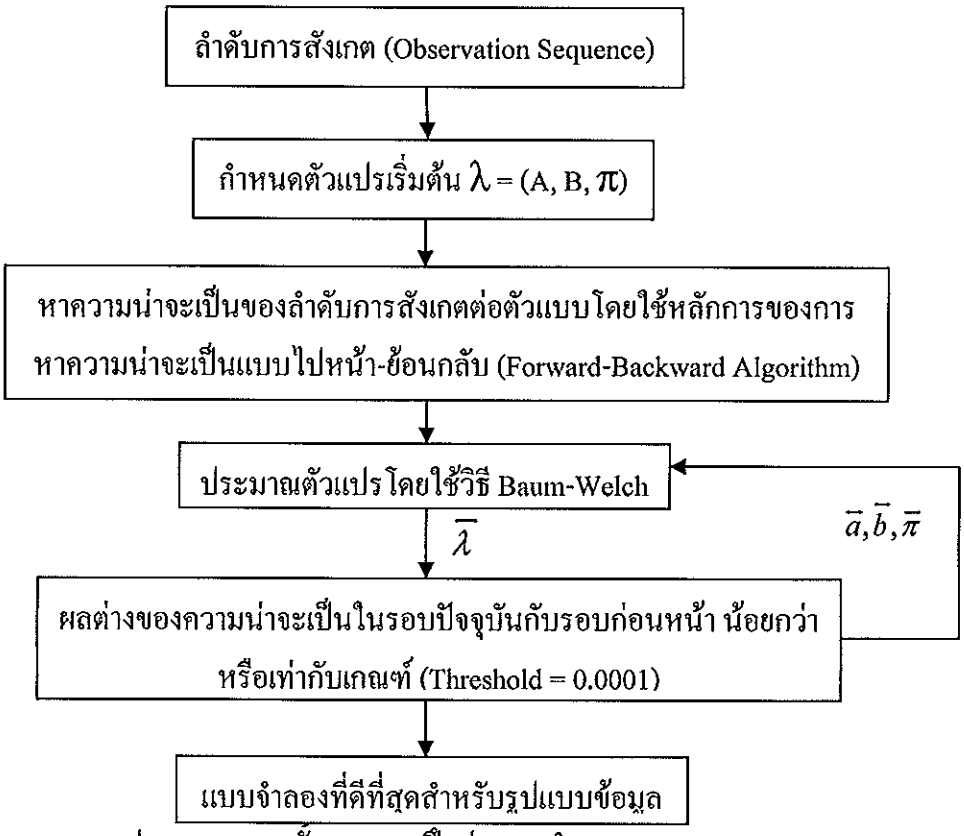
สำนักหอสมุดกลาง การเรียนรู้ตลอดชีวิตและสารสนเทศ

รูปแบบของข้อมูล ซึ่งแบบจำลองฮิดเดนมาร์คอฟสามารถแบ่งออกเป็น 2 ส่วนด้วยกันคือ การฝึกฝนแบบจำลอง (Training) และการทดสอบการจดจำของแบบจำลอง (Evaluation)

$$\lambda = (A, B, \pi) \tag{2-6}$$

2.12.1 **การฝึกฝนแบบจำลอง (Training)**

การฝึกฝนแบบจำลองให้จดจำรูปแบบข้อมูล เป็นการประมาณตัวแปรของแบบจำลองที่ดีที่สุดสำหรับรูปแบบของข้อมูลในแต่ละรูปแบบ ซึ่งขั้นตอนการฝึกฝนให้กับแบบจำลองแสดงดังภาพประกอบ 2-12 โดยอินพุตของแบบจำลอง คือ ลำดับการสังเกตของข้อมูล (Observation Sequence) จากนั้นกำหนดแบบจำลองเริ่มแรก λ ซึ่งประกอบด้วย (A, B, π) แล้วคำนวณหาค่าความน่าจะเป็นของอินพุตต่อแบบจำลองนั้นโดยใช้หลักการของการหาความน่าจะเป็นแบบไปหน้า-ย้อนกลับ (Forward-Backward Algorithm) เมื่อได้ค่าความน่าจะเป็นของอินพุตต่อแบบจำลอง จึงเข้าสู่กระบวนการประมาณตัวแปร (Parameter Estimation) โดยใช้วิธีของ Baum-Welch เพื่อทำการปรับตัวแบบจนกระทั่งได้แบบจำลองที่ดีที่สุดสำหรับอินพุตนั้น



ภาพประกอบ 2-12 ขั้นตอนการฝึกฝนแบบจำลอง (Training)

การคำนวณหาค่าความน่าจะเป็นของอินพุตหรือลำดับการสังเกตของแบบจำลอง โดยใช้ หลักการหาความน่าจะเป็นแบบไปหน้า-ย้อนกลับ มีรายละเอียดดังนี้

การหาความน่าจะเป็นแบบไปหน้า (Forward Algorithm) มีการกำหนดตัวแปรดังนี้

- $\alpha_t(i)$ เป็นตัวแปรแบบไปหน้า (Forward) ณ เวลา t ที่สถานะ i
- π_i เป็นค่าความน่าจะเป็น ณ เวลา t = 1 ที่สถานะ i
- $b_j(O_t)$ เป็นค่าความน่าจะเป็นของข้อมูล O ณ เวลา t ที่สถานะ j
- a_{ij} เป็นค่าความน่าจะเป็นในการเปลี่ยนจากสถานะ i ไปที่สถานะ j
- ขั้นเริ่มต้น

$$\alpha_1(i) = \pi_i b_i(O_1) \quad , \quad 1 \le i \le N \tag{2-7}$$

- ขั้นการวนซ้ำ

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1}) \tag{2-8}$$
$$1 \le t \le T-1 \, , \, 1 \le j \le N$$

- ขั้นสิ้นสุด

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{2-9}$$

ส่วนของกระบวนการหาความน่าจะเป็นแบบย้อนกลับ (Backward Algorithm) ได้มีการกำหนดตัวแปรเพิ่มเติมจากกระบวนการหาความน่าจะเป็นแบบไปหน้า ดังนี้

- $\beta_t(i)$ เป็นตัวแปรแบบย้อนกลับ ณ เวลา t ที่สถานะ i
- ขั้นเริ่มต้น

$$\beta_{T(i)} = 1 \, , \, 1 \le i \le N \tag{2-10}$$

- ขั้นอุปนัย

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \tag{2-11}$$
$$t = T-1, T-2, \ldots, 1 \, , \, 1 \le i \le N$$

หลังจากกระบวนการหาความน่าจะเป็นแบบไปหน้า-ย้อนกลับ แล้วจะเข้าสู่กระบวนการประมาณค่าพารามิเตอร์โดยใช้วิธี Baum-Welch โดยนำผลลัพธ์ที่ได้จากก่อนหน้านี้ มาใช้ซึ่งมีรายละเอียด ดังนี้

- กำหนดให้ $\xi_{t(i,j)}$ คือความน่าจะเป็นของการอยู่ในสถานะ i ที่เวลา t และสถานะ j ที่เวลา t+1

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \quad (2\text{-}12)$$

- กำหนดตัวแปร $\gamma_{t(i)}$ เป็นความน่าจะเป็นของการ อยู่ในสถานะ i ที่เวลา t โดยที่ตัวแบบ λ และลำดับของเหตุการณ์ O มีความสัมพันธ์กับ $\xi_{t(i,j)}$ โดยการบวกกันทุก j

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) \quad (2\text{-}13)$$

- การประมาณตัวแปรใหม่ของ A, B และ π

$$\bar{\pi}_i = \gamma_1(i) \quad (2\text{-}14)$$

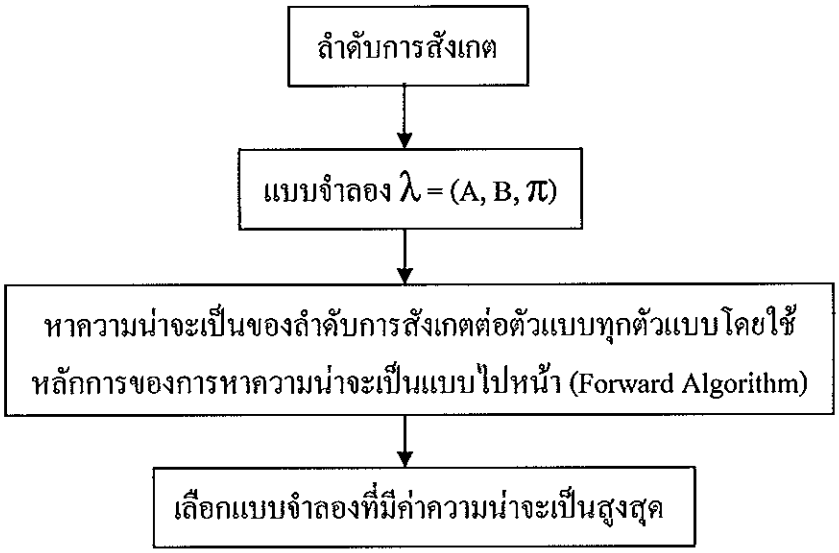
$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (2\text{-}15)$$

$$\bar{b}_j(k) = \frac{\sum_{t=1}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \frac{s,t,O_t = v_k}{} \quad (2\text{-}16)$$

2.12.2 **การทดสอบการจดจำของแบบจำลอง (Evaluation)**

หลังจากที่ได้ทำการฝึกฝนรูปแบบข้อมูลให้กับแบบจำลองแล้ว จึงทำการทดสอบการจดจำของแบบจำลองโดยการคำนวณหาความน่าจะเป็นของเหตุการณ์ต่อแบบจำลองโดยใช้หลักการหาความน่าจะเป็นแบบไปหน้ากับทุกๆ แบบจำลอง แบบจำลองใดที่ได้ค่าความน่าจะเป็น

สูงสุด ถือว่าเป็นคำตอบของลำดับสังเกตนั้น รายละเอียดของขั้นตอนการทดสอบการจดจำของแบบจำลองอธิบายได้ด้วยภาพประกอบ 2-13



ภาพประกอบ 2-13 ขั้นตอนการทดสอบการจดจำของแบบจำลอง (Evaluation)

ขั้นตอนแรกส่วนของฮิดเดนมาร์คอฟนี้ถูกสอนระบบด้วยชุดข้อมูลจากตัวอักษรภาษามือไทยทั้ง 42 ตัว เพื่อสร้างแบบจำลองของแต่ละตัวอักษรภาษาไทย ซึ่งแบบจำลองต้นแบบทั้ง 42 ต้นแบบนี้จะนำไปใช้ในการตัดสินใจแปลงจากชุดลำดับตัวอักษรท่าพื้นฐานภาษาอังกฤษเป็นอักษรภาษาไทย (หัวข้อที่ 3.7)

บทที่ 3
การออกแบบและพัฒนาระบบ

งานวิจัยนี้มีจุดประสงค์เพื่อออกแบบระบบรู้จำอักษรภาษามือไทย ซึ่งสามารถรู้จำตัวอักษรภาษามือ ทั้งหมด 42 ตัวอักษร (ยกเว้น ฃ และ ฅ) เพื่อช่วยในการเรียนรู้ตัวอักษรภาษามือไทย ซึ่งเป็นพื้นฐานของศัพท์ภาษามือไทยอื่นๆ อีกทั้งช่วยลดช่องว่างในการสื่อสารระหว่างผู้พิการทางหูและบุคคลทั่วไป โดยระบบต้องมีความสามารถรองรับความหลากหลายของผู้ใช้ระบบได้ ซึ่งในบทที่ 3 นี้จะกล่าวเน้นในส่วนของการออกแบบระบบเป็นหลัก

3.1 แนวคิดในการออกแบบระบบ

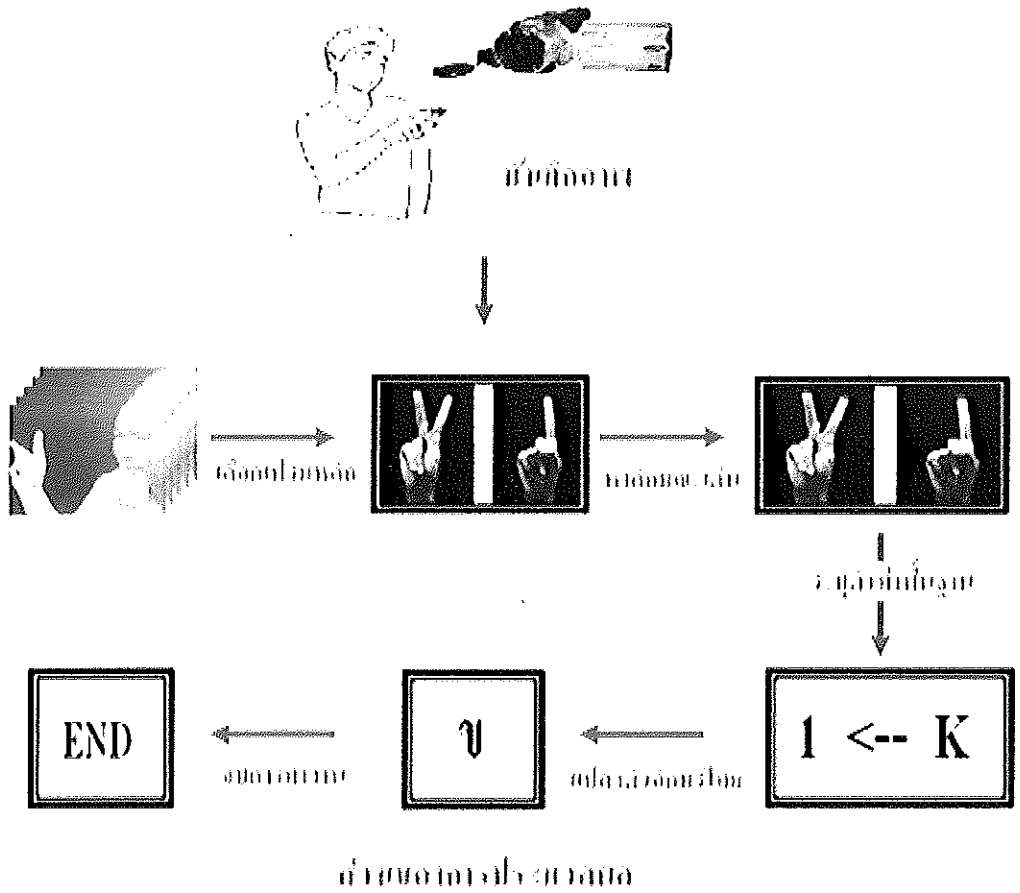
ตัวอักษรภาษามือไทย นอกจากจะใช้ในการสะกดคำเฉพาะแล้ว ยังเป็นพื้นฐานของคำศัพท์ภาษามือไทยอื่นๆ อีกด้วย ดังนั้นผู้วิจัยจึงสนใจพัฒนาระบบรู้จำตัวอักษรภาษามือไทย เพื่อเป็นพื้นฐานในการเรียนรู้ศัพท์ภาษามือไทย และลดช่องว่างในการสื่อสารระหว่างบุคคลปกติและผู้พิการ ทั้งนี้ต้นทุนของระบบควรจะต่ำ เพื่อจะได้นำไปพัฒนาต่อให้บุคคลทั่วไปสามารถนำไปใช้งานจริงได้

เนื่องจากงานวิจัยเกี่ยวกับการรู้จำภาษามือส่วนใหญ่จะใช้อุปกรณ์ช่วยในการจับและติดตามการเคลื่อนไหวของมือ เช่นการใช้ถุงมือเซนเซอร์ และการใช้ถุงมือสี ซึ่งงานวิจัยเหล่านี้จะมีความแม่นยำสูง แต่ก็ยุ่งยากในการใช้งาน โดยเฉพาะถุงมือเซนเซอร์ ต้นทุนของระบบจะสูงมาก ดังนั้นเพื่อความสะดวก และประหยัดในการใช้งาน เราได้พัฒนาระบบรู้จำภาษามือไทยขึ้น โดยไม่ใช้อุปกรณ์ช่วยในการติดตามการเคลื่อนไหวของมือ

ตัวอักษรภาษามือไทยมีทั้งหมด 42 ตัว ประกอบด้วยท่าพื้นฐาน 24 ท่าคือ ตัวอักษรภาษามืออังกฤษ 17 ตัว ตัวเลขภาษามืออังกฤษ 5 ตัว และท่ามือพื้นฐานอีก 2 ท่า คือ ได้แก่ "Vowel base" และท่า "Slow" ซึ่งแต่ละตัวอักษรภาษามือไทย จะมีจำนวน และการเรียงลำดับของท่าพื้นฐานแตกต่างกัน ดังนั้นจึงต้องเก็บข้อมูลไว้ในรูปแบบไฟล์วิดีทัศน์และเนื่องจากไฟล์วีดีทัศน์เกิดจากภาพหลายเฟรมเรียงต่อกันซึ่งจะมีทั้งเฟรมที่ ชัด ไม่ชัด และคล้ายกัน การที่จะประมวลผลทุกเฟรมจะทำให้ระบบทำงานช้า ระบบจึงต้องเลือกเฟรมที่ชัดที่สุดในแต่และท่าพื้นฐาน เพียง 1 ภาพ มาหาลักษณะเด่นเพื่อที่จะใช้ในการระบุว่าภาพแต่ละเฟรมตรงกับท่าพื้นฐานตัวใด แล้วนำชุดข้อมูลที่ได้ไปผ่านระบบรู้จำโดยใช้ฮิดเดนมาร์คอฟโมเดลเพื่อแปลเป็นตัวอักษรภาษาไทย

3.2 ภาพรวมการออกแบบโปรแกรม

ในส่วนนี้กล่าวถึงภาพรวมการออกแบบโปรแกรมที่ใช้ในงานวิจัยนี้ โดยระบบเริ่มจากการรับภาพผู้ใช้ระบบด้วยกล้อง และบันทึกภาพไว้ในรูปแบบไฟล์นามสกุล .avi เพื่อส่งไปยังขั้นตอนการประมวลผลภาพและรู้จำตัวอักษรภาษามือไทย รวมทั้งหมด 3 ขั้นตอนหลัก แต่ละขั้นตอนให้ผลดังแสดงในรูปประกอบ 3-1 ซึ่งเป็นตัวอย่างผลลัพธ์แต่ละขั้นตอนของไฟล์ "ข" ที่เกิดจากตัวอักษร "K" ตามด้วยตัวเลข "1" โดยขั้นตอนการเก็บข้อมูลเพื่อทดสอบ และ สอนระบบจะกล่าวรายละเอียดในหัวข้อ 3.3



ภาพประกอบ 3-1 ผลของแต่ละขั้นตอน

1. ส่วนการเลือกเฟรมหลักได้นำออปติคัล โฟลว์มาช่วยในการตัดสินใจเลือกภาพที่มีความชัดเจนที่สุดในแต่ละท่าพื้นฐาน ซึ่งจะกล่าวอย่างละเอียดในหัวข้อ 3.4
2. ส่วนการหาลักษณะเด่นของมือในภาพ และการระบุท่าพื้นฐาน เราได้ทำการทดลอง 2 วิธี คือการหามุมที่สนใจในเฟรมหลัก ซึ่งจะกล่าวอย่างละเอียดในหัวข้อ 3.5 และวิธี

วิเคราะห์จากรูปร่างมือและลักษณะของนิ้วมือ โดยมีลักษณะเด่นที่สำคัญ คือการวางตัวของฝ่ามือ การแยกหรือชิดกันของนิ้วมือ ลำดับความสูงต่ำของแต่ละนิ้ว และจำนวนนิ้วมือในแต่ละทิศทาง ซึ่งจะกล่าววิธีการหาอย่างละเอียดในหัวข้อ 3.6

3. ส่วนของการแปลง จากชุดข้อมูลท่าพื้นฐานจากขั้นตอนระบุท่าพื้นฐานเป็นอักษรภาษาไทย ใช้ฮิดเดนมาร์คอฟโมเดลช่วยในการรู้จำ เพื่อประสิทธิภาพที่ดีขึ้นของระบบ ซึ่งจะกล่าวอย่างละเอียดในหัวข้อ 3.7

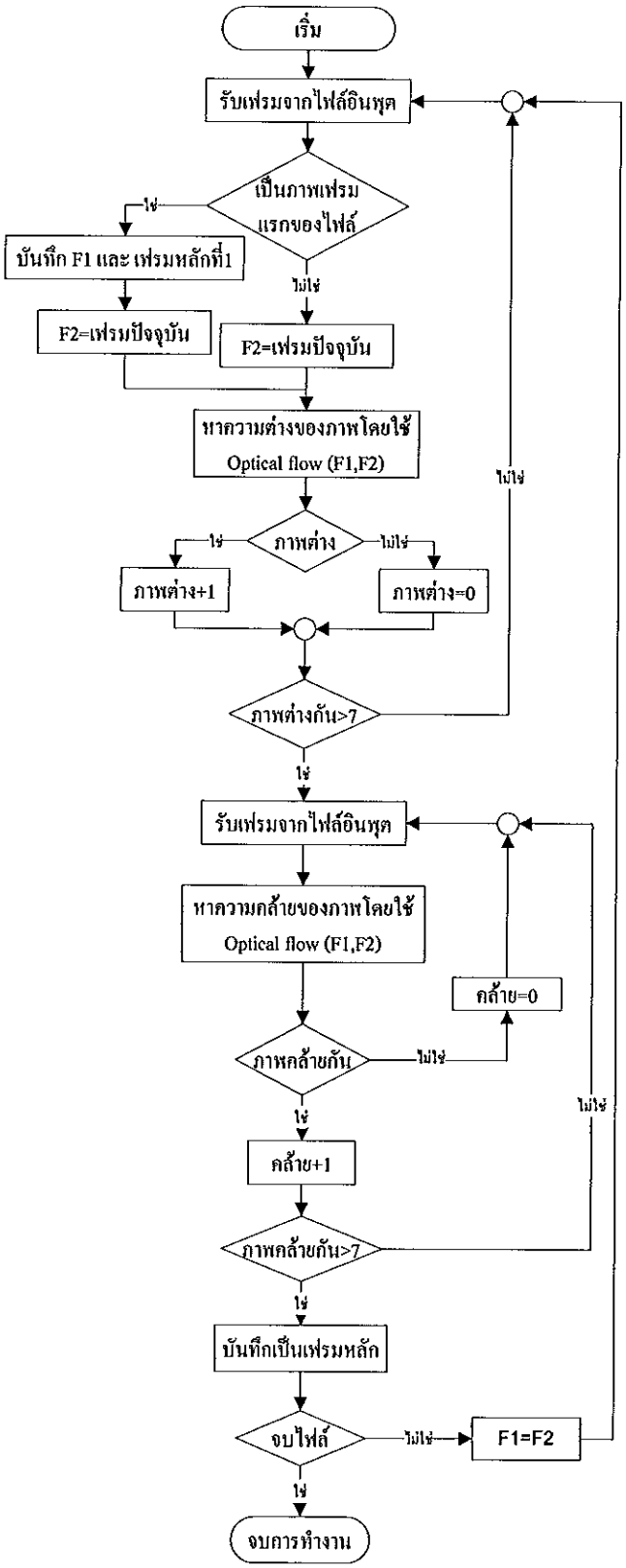
3.3 การเก็บข้อมูลที่ใช้ในการสอน และทดสอบระบบ

เริ่มจากระบบจับภาพอาสาสมัคร ด้วยกล้อง Panasonic NV-GS400 และบันทึกเป็นไฟล์วีดีทัศน์ในรูปแบบ Audio Visual Interleave standard (AVI) ขนาด 720x576 พิกเซล ความเร็วในการรับภาพ เท่ากับ 25 เฟรมต่อวินาที ระบบสี RGB โดยที่ขณะบันทึกข้อมูล จะควบคุมสภาวะแวดล้อม คือ ความสว่างของไฟ พื้นหลังเป็นสีเข้ม ตำแหน่งและระยะในการวางกล้องจะต้องสามารถจับการเคลื่อนไหวของมือได้ทั้งหมด โดยจะบันทึกภาพจากด้านหน้าของอาสาสมัคร ซึ่งจะเริ่มจากการทำท่าเตรียมตามด้วยท่าพื้นฐานของแต่ละตัวอักษรภาษามือไทย และจบด้วยท่าจบ ซึ่งเป็นท่าเดียวกับท่าเตรียม เพื่อเป็นการบอกถึงจุดเริ่มต้นและจุดบนของการใช้ระบบ

3.4 การเลือกเฟรมหลัก

จากความเร็วในการบันทึกท่าอาสาสมัครเท่ากับ 25 เฟรมต่อวินาที ในหนึ่งไฟล์ของหนึ่งตัวอักษรภาษามือไทยจึงมีหลายเฟรม ซึ่งประกอบด้วย ภาพที่ไม่ชัดเจน ภาพที่มีความชัดเจนและ ภาพที่คล้ายกัน การประมวลทุกเฟรมจะทำให้ระบบทำงานช้า ดังนั้นเพื่อเพิ่มความเร็วในการประมวลผลจึงได้ทำการเลือกบันทึกเฉพาะเฟรมที่เหมาะสมจะเป็นเฟรมหลักของแต่ละท่าไว้เพื่อส่งไปยังขั้นตอนต่อไป ซึ่งในส่วนนี้เราได้ใช้ออปติคัล โฟลว์มาช่วยในการตัดสินใจเลือกเฟรมหลัก

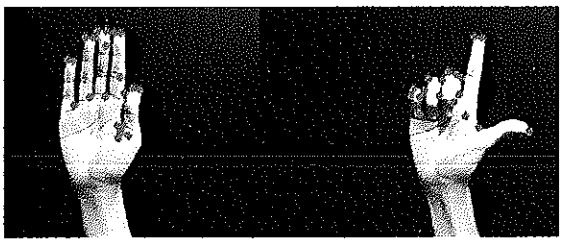
ออปติคัล โฟลว์ทำหน้าที่หาจำนวนพิกเซลที่เหมือนกันในภาพ 2 ภาพ ดังนั้นเมื่อหาจำนวนจุดเหมือนระหว่างภาพอ้างอิงและภาพต่างๆจากไฟล์อินพุต แล้วนำมาวิเคราะห์จะสามารถเลือกเฟรมมีความชัดเจนได้ โดยมีกระบวนการทำงานดังภาพประกอบ 3-2



ภาพประกอบ 3-2 กระบวนการการประมวลผลภาพของการเลือกเฟรมหลัก

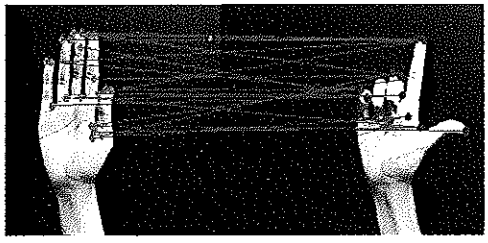
3.5 การหาลักษณะเด่นของมือ และการระบุท่าพื้นฐานโดยการหามุมที่สนใจในเฟรมหลัก

ขั้นตอนนี้ถูกพัฒนาขึ้นจากหลักการ ในหัวข้อ 2.9 - 2.11 คือ การหามุมด้วยวิธีของแฮร์ริส, Normalized cross-correlation, Random Sample Consensus โดยจะทำการเก็บภาพต้นแบบของท่าพื้นฐานทั้ง 24 ท่าไว้เพื่อใช้เป็นฐานข้อมูล เมื่อได้รับภาพอินพุตเข้ามาในระบบ ระบบจะทำการเปรียบเทียบภาพอินพุตกับทุกท่าในฐานข้อมูลและเลือกท่าที่มีความคล้ายกันมากที่สุด ซึ่งความคล้ายกันนั้นพิจารณาจากจุดแต่และจุดว่าทั้งสองภาพนั้นมีจุดใดคล้ายกันบ้าง ขั้นตอนการทำงานเริ่มจากการนำภาพอินพุตที่ได้จากการเลือกเฟรมหลัก และภาพจากฐานข้อมูล มาหาจุดมุมต่างๆ ของมือในภาพ ด้วยวิธี Harris Corner Detector ดังแสดงในภาพประกอบ 3-3 ภาพทางด้านขวาคือภาพต้นแบบ ส่วนภาพทางด้านซ้ายคือ เฟรมหลักจากขั้นตอนการเลือกเฟรมหลัก



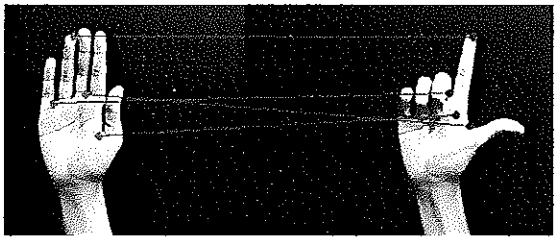
ภาพประกอบ 3-3 ผลของการหามุมด้วยวิธีแฮร์ริส

เมื่อได้มุม ณ จุดต่างๆในภาพทั้งสองภาพแล้ว ระบบจะทำการเปรียบเทียบว่าจุดใดมีความคล้ายคลึงกันบ้าง โดยใช้วิธี Normalized cross-correlation ซึ่งจะให้ผลดังตัวอย่างในภาพประกอบที่ 3-4



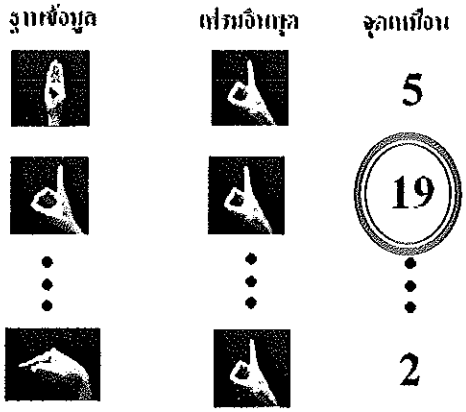
ภาพประกอบ 3-4 การเปรียบเทียบ จุด ต่อ จุด ของภาพจากฐานข้อมูล และภาพอินพุต

ขั้นตอนสุดท้ายคือการหาความน่าจะเป็นสูงสุด ที่มุมแต่และจุดในภาพอินพุต จะมีลักษณะที่เหมือนหรือใกล้เคลียกับ มุม ณ จุดๆในภาพต้นแบบ ซึ่งให้ Random Sample Consensus มาช่วยในการพิจารณาเปรียบเทียบมุมของภาพทั้งสองภาพ ดังตัวอย่างในภาพประกอบที่ 3-5



ภาพประกอบ 3-5 การหาความน่าจะเป็นสูงสุดของแต่และมุมของภาพทั้งสองภาพ

ซึ่งเมื่อได้รับภาพเข้ามา 1 ภาพ ระบบจะทำ 3 ขั้นตอนข้างต้นซ้ำๆ โดยจะเปลี่ยนภาพจากฐานข้อมูลไปเรื่อยๆจนครบ 24 ภาพ และเลือกภาพที่มีจุดเหมือนกันมากที่สุดเป็นคำตอบ ดังตัวอย่างในภาพประกอบ 3-6 ท่าที่ใกล้เคียงที่สุดคือท่าที่ 2 เนื่องจากที่จุดคล้ายมากที่สุด ดังตัวอย่างในรูปประกอบที่ 3-6

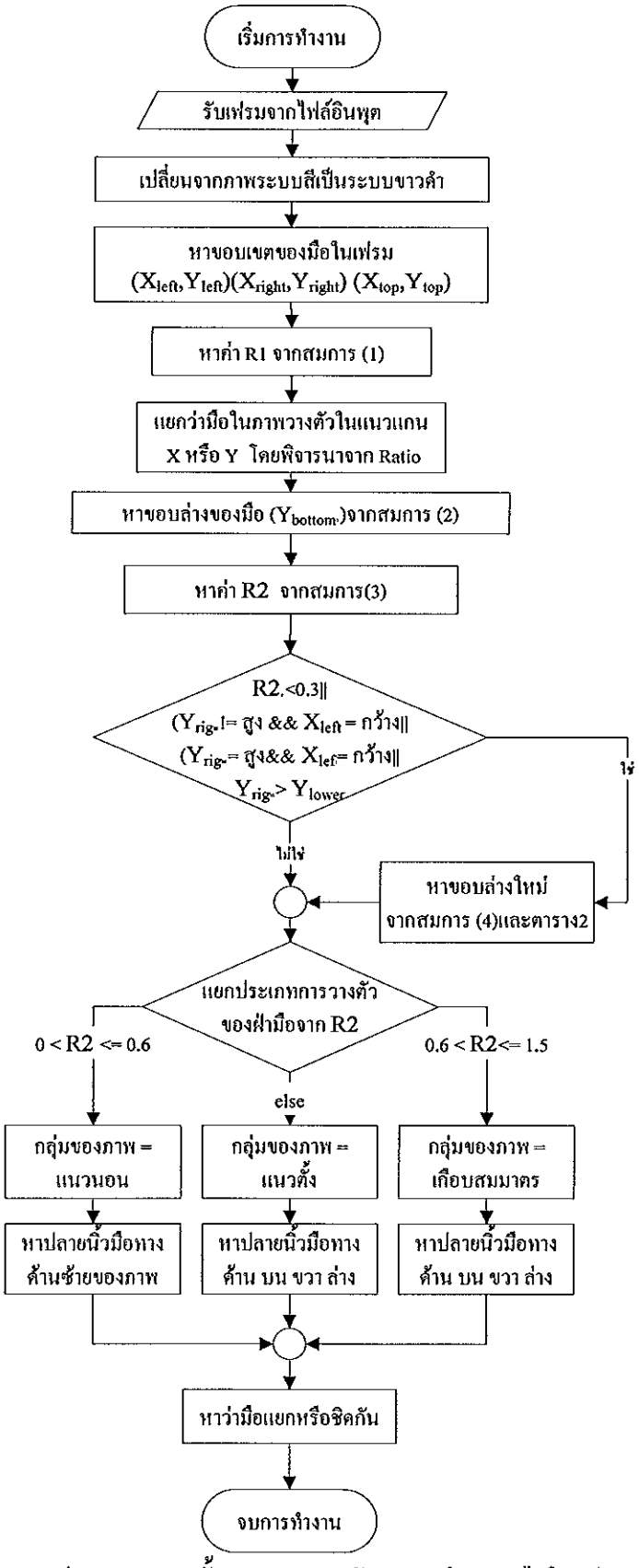


ภาพประกอบ 3-6 ตัวอย่างการเลือกคำตอบของการระบุท่าหลักโดยมุมที่สนใจในเฟรมหลัก

นอกจากการหาลักษณะเด่นของมือในภาพ และการระบุท่าพื้นฐานโดยการหามุมที่สนใจในเฟรมหลักเพื่อเพิ่มประสิทธิภาพของระบบ ในวิทยานิพนธ์นี้ยังได้ทำทดสอบการหาลักษณะเด่น และการระบุท่าพื้นฐานอีกวิธีหนึ่งคือ การหาลักษณะเด่นของมือในภาพ และการระบุท่าพื้นฐานโดยวิธีวิเคราะห์รูปร่างมือและนิ้วมือ ซึ่งจะกล่าวอย่างละเอียดในหัวข้อต่อไป

3.6 การหาลักษณะเด่นของมือ และการระบุท่าพื้นฐานโดยวิธีวิเคราะห์รูปร่างมือและนิ้วมือ

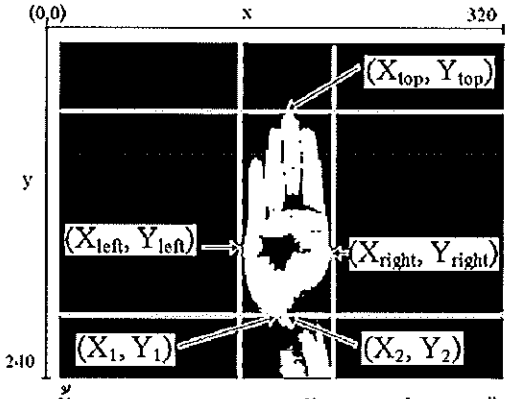
ขั้นตอนนี้เป็นการหาลักษณะเด่นของมือจากเฟรมหลักที่ได้เลือกไว้จากกระบวนการเลือกเฟรมหลัก มี 4 ขั้นตอน คือ หาขอบเขตของมือ หาการวางตัวของมือในภาพ หาปลายนิ้วมือ ระบุว่านิ้วชิดหรือกางออก โดยมีกระบวนการทำงานดังภาพประกอบ 3-7



ภาพประกอบ 3-7 ขั้นตอนการหาลักษณะเด่นของมือในเฟรมหลัก

3.6.1 การหาขอบเขตของมือ

ในขั้นตอนนี้เป็นการนำเสนอวิธีหาขอบเขตของมือ จุดที่ใช้ในการระบุขอบเขตของมือมี 4 จุดคือ จุดสูงสุด(X_{top}, Y_{top}) จุดซ้ายสุด(X_{left}, Y_{left}) จุดขวาสุด(X_{right}, Y_{right}) และ จุดล่างสุด (X_{bottom},Y_{bottom}) ดังภาพประกอบที่ 3-4 โดยเริ่มจากการเปลี่ยนภาพสีเป็นภาพขาวดำ ทำให้มือที่เราสนใจเป็นสีขาว หลังจากนั้น ทำการสแกนทุกพิกเซล ตามแนว x,y และหยุดเมื่อเจอพิกเซลที่เป็นสีขาว โดยจุดสูงสุด (X_{top}, Y_{top}) หาได้จากการสแกนตามแนวแกน x จากบนลงล่าง จุดซ้ายสุด(X_{left}, Y_{left}) หาได้จากการสแกนตามแนวแกน y จากซ้ายไปขวา และ จุดขวาสุด (X_{right}, Y_{right}) หาได้จากการสแกนตามแนวแกน y จากขวาไปซ้าย ส่วนจุด ล่างสุด (X_{bottom},Y_{bottom}) ได้จากการหาบริเวณข้อมือที่เป็นจุดต่อระหว่างฝ่ามือและแขน ซึ่งจะอธิบายในหัวข้อ 3.6.1.1



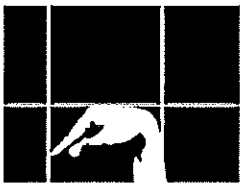
ภาพประกอบ 3-8 ขั้นตอนผลของการหาลักษณะเด่นของมือในเฟรมหลัก

3.6.1.1 การหาจุดล่างสุดของมือ

ขั้นตอนนี้เป็นการหาข้อมือ เพื่อที่จะระบุขอบเขตของมือด้านล่าง (X_{bottom},Y_{bottom}) ซึ่งจากกลางฝ่ามือลงเป็นถึงแขน ข้อมือเป็นสุดที่แคบที่สุด ขั้นแรกคือคำนวณอัตราส่วนความสูงจากปลายนิ้วถึงกลางมือต่อความกว้างของมือตามสมการ 3-1 เพื่อที่จะแยกการวางตัวของฝ่ามือ เป็นภาพแนวตั้งหรือแนวนอน ถ้าอัตราส่วนมีค่ามากกว่าหรือเท่ากับ 1 เฟรมนั้นจะเป็นภาพแนวตั้ง แต่ถ้าอัตราส่วนมีค่าน้อยกว่าหนึ่ง เฟรมนั้นจะเป็น ภาพแนวนอน ดังตัวอย่างในภาพประกอบที่ 3-9

$$R1 = \frac{2(\min(y_{left}, y_{right}) - y_{top})}{|x_{right} - x_{left}|} \qquad (3\text{-}1)$$

(ก) ภาพแนวตั้ง R1=1.2



(ข) ภาพแนวนอน R1=0.22

ภาพประกอบ 3-9 รูปแบบการวางตัวของมือ จากการแยกประเภทครั้งที่ 1

เมื่อเราแยกการวางตัวของฝ่ามือแล้ว ขั้นตอนต่อเป็นเป็นการหาจุดที่แคบที่สุด ซึ่งอยู่ในช่วงกลางฝ่ามือลงไปถึงแขนที่อยู่ล่างสุดของภาพ ด้วยสมาการที่ 3-2

$$y_{bottom} = \begin{cases} \min\limits_{y=\min(y_{left},y_{right})}^{height}(x_2 - x_1), R1 \geq 1, \\ \min\limits_{y=\max(y_{left},y_{right})}^{height}(x_2 - x_1), R1 < 1, \end{cases} \quad (3\text{-}2)$$

เมื่อ ภาพ(x_1, y) คือพิกเซลที่อยู่ทางซ้ายสุดของภาพ

ภาพ(x_2, y) คือพิกเซลที่อยู่ทางขวาสุดของภาพ

จากนั้นนำขอบเขตของมือที่ได้มาตรวจสอบความถูกต้อง ซึ่งจากการทดสอบพบว่า ด้วยวิธีการหาขอบเขตของมือด้วยวิธีข้างต้น ยังเกิดความผิดพลาดกับภาพบางภาพ ซึ่งแยกความผิดพลาดออกเป็น 4 ประเภท ดังตัวอย่างในภาพประกอบ 3-10 ซึ่งจะอธิบายรายละเอียดการปรับขอบเขตมือในหัวข้อ 3.6.1.2

(ก) (ข) (ค) (ง)

(ก) ปลายแขนอยู่ทางด้านขวาของภาพ (ข) ปลายแขนอยู่ที่มุมล่างขวาของภาพ (ค) ส่วนสูงของมือน้อยเกินไป (ง) Y_{right} อยู่ต่ำกว่า Y_{bottom}

ภาพประกอบ 3-10 ความผิดพลาดทั้ง 4 ประเภท ของการหาขอบเขตมือ

3.6.1.2 การปรับขอบเขตของมือ

ขั้นตอนแรกของการปรับของเขตมือ คือการนำขอบเขตมือที่หาขั้นตอนข้างต้นมามาหาอัตราส่วนความสูงต่อความกว้างของมือ ตามสมการที่ 3-3

$$R2 = \frac{|\, y_{top} - y_{bottom} \,|}{|\, x_{left} - x_{right} \,|} \tag{3-3}$$

จากนั้น ทำการหาและแยกประเภทความผิดพลาดในการระบุขอบเขตมือ โดยการนำค่าจุดสูงสุด (X_{top}, Y_{top}) จุดซ้ายสุด (X_{left}, Y_{left}) จุดขวาสุด (X_{right}, Y_{right}) จุดล่างสุด (X_{bottom}, Y_{bottom}) และค่า R2 มาพิจารณาตามเงื่อนไข:

$$Type = \begin{cases} 1, if \;\; y_{right} \neq 240 \; and \; x_{right} = 320, \\ 2, if \; y_{right} = 240 \; and \; x_{right} = 320, \\ 3, if \; R2 < 0.3, \\ 4, if \; y_{right} > y_{lower}, \\ 0, otherwise \, . \end{cases}$$

Type = 0: ขอบเขตมือถูกต้อง

Type = 1: ปลายแขนอยู่ทางด้านขวาของภาพ

Type = 2: ปลายแขนอยู่ที่มุมล่างขวาของภาพ

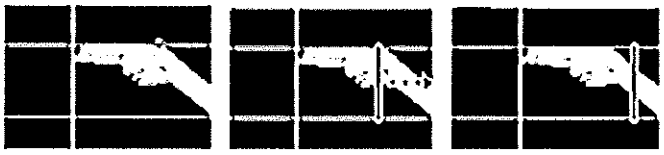
Type = 3: อัตราส่วนไม่เหมาะสม (ส่วนสูงน้อยเกินไป)

Type = 4: Y_{right} ต่ำกว่า Y_{bottom}

<u>การปรับปรุงขอบเขตมือ ของความผิดพลาดทั้ง 4 ประเภท</u>

- ประเภทที่1: ปลายแขนอยู่ทางด้านขวาของภาพ (Type1)

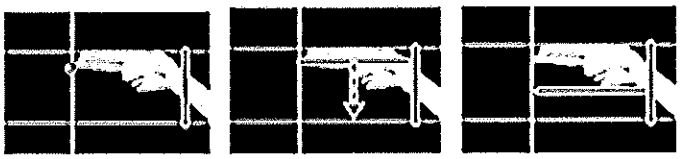
การแก้ไขปัญหานี้เราต้องกำหนดค่า X_{right} และ Y_{bottom} ใหม่ ดังตัวอย่างเช่น ภาพประกอบที่ 3-7 เป็นขั้นตอนการหาค่า X_{right} ใหม่ ภาพประกอบที่ 3-7(ก) เป็นจุดเริ่มต้นของการสแกนตามแนวแกน y โดยใช้จุด X_{top} ที่หาได้ก่อนหน้านี้ และเคลื่อนที่ไปจนสุดขอบภาพทางด้านขวา เพื่อหาจุดที่แคบที่สุด ดังแสดงใน ภาพประกอบที่ 3-11(ข)-(ค)



(ก) (ข) (ค)

ภาพประกอบ 3-11 ขั้นตอนการหา X_{right} ใหม่ ของภาพประเภทที่1

ส่วนภาพประกอบ 3-12 (ก)-(ข) เป็นภาพแสดงการหา Y_{bottom} ใหม่ โดยการสแกนตามแนวแกน x ไปด้านล่างของขอบภาพ โดยเริ่มจากค่าที่น้อยที่สุดระหว่าง Y_{left} และ Y_{right} ทำให้เราได้ X_{right} และ Y_{bottom} ใหม่ ซึ่งเป็นขอบเขตที่ถูกต้องดังแสดงใน ส่วนภาพประกอบ 3-12 (ค)



(ก) (ข) (ค)

ภาพประกอบ 3-12 ขั้นตอนการหา Y_{bottom} ใหม่ ของภาพประเภทที่1

- ประเภทที่2: ปลายแขนอยู่ที่มุมล่างขวาของภาพ (Type2)

การหาแก้ไขปัญหาปลายแขนอยู่ที่มุมล่างขวาของภาพ เราต้องกำหนดค่า X_{right} และ Y_{bottom} ซึ่งขั้นตอนการหาขอบเขตมือใหม่เหมือนกับประเภทที่1 โดยเริ่มจากการหา X_{right} และ Y_{bottom} ใหม่ ดังในในภาพประกอบ 3-13 (ก)-(ข) และภาพประกอบ 3-14 (ก)-(ข)

(ก) (ข) (ค)

ภาพประกอบ 3-13 ขั้นตอนการหา X_{right} ใหม่ ของภาพประเภทที่2

(ก) (ข) (ค)

ภาพประกอบ 3-14 ขั้นตอนการหา Y_{bottom} ใหม่ ของภาพประเภทที่2

- ประเภทที่3: อัตราส่วนไม่เหมาะสม ส่วนสูงน้อยเกินไป (Type 3)

ปัญหาอัตราส่วนไม่เหมาะสมเราทำการปรับปรุงเฉพาะ Y_{bottom} โดยการสแกนตามแนวแกน x ไปด้านล่างของขอบภาพ ซึ่งเริ่มจากค่าที่มากที่สุดระหว่าง Y_{left} และ Y_{right} เพื่อหาส่วนที่แคบที่สุดและระบุเป็น Y_{bottom} ใหม่ ดังแสดงในภาพประกอบ 3-15 (ก)-(ข)

(ก) (ข) (ค)

ภาพประกอบ 3-15 ขั้นตอนการหา Y_{bottom} ใหม่ ของภาพประเภทที่3

- ประเภทที่4: Y_{right} ต่ำกว่า Y_{bottom} (Type 4)

วิธีในการแก้ไขปัญหา Y_{right} ต่ำกว่า Y_{bottom} คือการหา X_{right} ใหม่ ซึ่งมีขั้นตอนเหมือนกับ การหา X_{right} ใหม่ ของปัญหาปลายแขนอยู่ทางด้านขวาของภาพ ดังแสดงในภาพประกอบที่ 16

(ก) (ข) (ค)

ภาพประกอบ 3-16 ขั้นตอนการหา X_{right} ใหม่ ของภาพประเภทที่4

ตัวอย่าง ผลจากการปรับปรุงขอบเขตของมือของปัญหาทั้ง 4 ประเภท แสดงในภาพประกอบ 3-17

ประเภทที่ 1 ประเภทที่ 2 ประเภทที่ 3 ประเภทที่ 4

ภาพประกอบ 3-17 ขอบเขตมือใหม่ที่ได้จากการปรับปรุง

3.6.2 แยกประเภทของท่ามือ

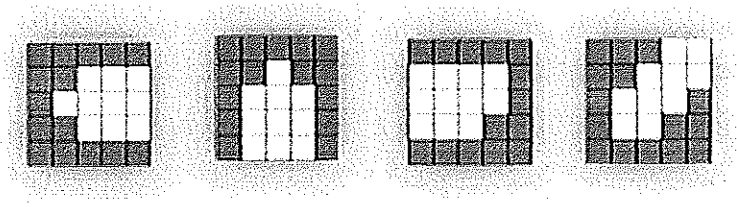
ท่าฐานพื้นฐานในตัวอักษรภาษามือไทยทั้งหมด 24 ท่า เราสามารถแยกภาพออกเป็น 3 ประเภท คือ ท่าแนวตั้ง ท่าแนวนอน ท่าสมมาตร โดยใช้ อัตราส่วนความสูงหารด้วยความกว้างขอขอบเขตภาพ เป็นตัวแยกประเภทของภาพซึ่งมีเงื่อนไขในการพิจารณา:

$$Type_{hand} = \begin{cases} \text{horizontal}, if\ 0 < R2 <= 0.6, \\ \text{symmetry}, if\ 0.6 < R2 <= 1.5, \\ \text{vertical}, \text{otherwise}, \end{cases}$$

ถ้าอัตราส่วน(R2) มีค่าอยู่ในช่วง 0 ถึง 0.6 ท่ามือในเฟรมนั้นจะเป็นประเภท แนวนอน หรือ ถ้าอัตราส่วนมีค่าอยู่ในช่วงมากกว่า 0.6 แต่ไม่ถึง 1.5 จัดเป็นประเภทสมมาตร นอนจาก 2 เงื่อนไขนี้ถือเป็นท่าแนวตั้งทั้งหมด

3.6.3 การหาปลายนิ้วมือ

ปลายนิ้วมือ หาได้จากการสแกนพิกเซลที่ละพิกเซลตามแนวขอบรูปร่างของมือ เพื่อหารูปแบบพิกเซลปลายนิ้วมือทั้ง 4 ทิศทาง ได้แก่ ปลายนิ้วทางด้านซ้าย ปลายนิ้วทางด้านบน ปลายนิ้วทางด้านขวา และปลายนิ้วทางด้านล่าง ตัวอย่างดังตัวอย่างแสดงในภาพประกอบ 3-18 ช่องสี่เหลี่ยมสีเทาแทนพิกเซลฉากหลังที่เป็นสีเข้ม และช่องสี่เหลี่ยมสีขาวแทนพิกเซลที่เป็นนิ้วมือ โดยถ้าเฟรมที่มาจากขั้นตอนก่อนหน้า เป็นภาพแนวตั้ง หรือภาพเกือบสมมาตร ระบบจะหาปลายนิ้วมือทางด้าน บน ขวา และล่าง แต่ถ้าเป็นภาพแนวนอน เราจะหาเฉพาะปลายนิ้วมือทางด้านซ้าย



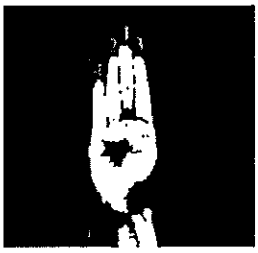
(ก) (ข) (ค) (ง)

(ก)ปลายนิ้วทางด้านซ้าย (ข)ปลายนิ้วทางด้านบน (ค)ปลายนิ้วทางด้านขวา (ง)ปลายนิ้วทางด้านล่าง

ภาพประกอบ 3-18 ตัวอย่างรูปแบบพิกเซลปลายนิ้วมือทั้ง 4 ทิศทาง

3.6.4 ลำดับความสูงของนิ้วมือ

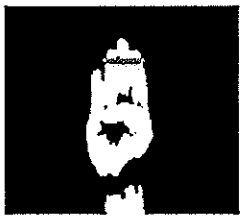
จากงานขั้นตอนก่อนหน้า เราจะได้ตำแหน่งของแต่ละปลายนิ้วมือ ซึ่งเก็บไว้ในรูปแบบคู่อันดับ (X,Y) จากนั้นทำการพิจารณาจากค่า Y ว่านิ้วใดอยู่ตำแหน่งสูงสุด พร้อมทั้งกำหนดค่าประจำนิ้วนั้นเป็น 1 ส่วนนิ้วที่มีตำแหน่งสูงรองลงมากำหนดให้เป็น 2 3 4 5 ตามลำดับเท่าที่มีการพบปลายนิ้วมือ จากนั้นนำมาเรียงเป็นลำดับของตัวเลขความสูงของนิ้วมือ จากนิ้วทางด้านขวา เช่นในภาพประกอบ 3-19 ลำดับความสูงของนิ้วมือคือ: 3124 เป็นต้น



ภาพประกอบ 3-19 ตัวอย่างการกำหนดลำดับความสูงของแต่ละนิ้วมือ

3.6.5 **การวางตัวของนิ้วมือ**

ท่าพื้นฐานของตัวอักษรภาษามือไทยบางท่าการแยกหรือเรียงชิดกันของนิ้วมือก็ให้ความหมายที่แตกต่างกัน ตัวอย่างเช่น ตัวอักษร B และตัวเลข 4 ทั้งสองท่าต่างมีลักษณะเด่นที่เหมือนกันมาก ยกเว้นการแยกกันของนิ้วมือ ดังนั้นการวางตัวของนิ้วมือจึงเป็นลักษณะสำคัญที่นำมาให้ในการตีความหมายของท่าในแต่และเฟรม การหาลักษณะเด่นว่าท่ามือในแต่ละเฟรมแยกหรือการเรียงชิดติดกันนั้นเราใช้วิธีสแกนตามแนวนอน ที่ระดับความสูง 4 ใน 5 ของขอบเขตมือ ดังตัวอย่างภาพประกอบ 3-20 เมื่อเราสแกนที่ละพิกเซลตามเส้นสีแนวนอนและนับจำนวนครั้งที่ค่าสีของพิกเซลเปลี่ยนจากสีดำเป็นสีขาวหรือจากสีขาวเป็นสีดำ จะทำให้เราสามารถสรุปได้ว่า นิ้วมือในท่ามือนั้นเรียงชิดติดกันหรือไม่ ถ้าจำนวนครั้งที่พิกเซลมีการเปลี่ยนสีน้อยกว่าหรือเท่ากับ 2 ท่ามือนั้นจะมีลักษณะนิ้วมือที่ชิดกัน ดังเช่นภาพประกอบ 3-20 (ก) แต่ถ้า จำนวนครั้งในการเปลี่ยนสีมีมากกว่า 2 ครั้ง นิ้วมือในท่านั้นจะอยู่ในลักษณะแยกกัน ดังตัวอย่างในภาพประกอบ 3-20 (ข) นอกจากตัวอักษร B กับตัวเลข 4 แล้ว ยังมีตัวอักษร R กับ ตัวเลข 2 และ ตัวอักษร K ที่ถ้าปราศจากลักษณะการวางตัวของนิ้วมือว่าแยกหรือชิดกันแล้ว ก็ไม่สามารถแยกออกจากกันได้



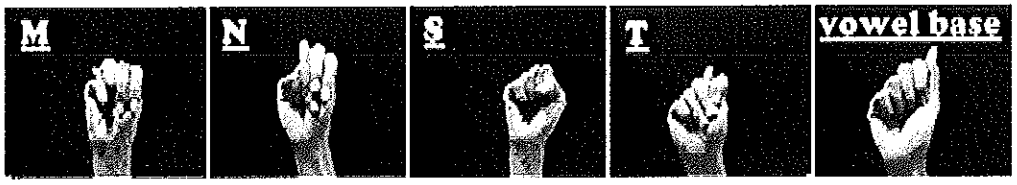
(ก)ท่าตัวอักษร B, ลักษณะนิ้วเรียงชิดติดกัน (ข) ท่าตัวเลข 4, ลักษณะนิ้วแยกห่างออกจากกัน

ภาพประกอบ 3-20 ระดับความสูงที่ใช้ในการสแกนหาการแยกของนิ้วมือ

3.6.6 **การหาความกว้างของนิ้วมือ**

ท่าพื้นฐานในภาษามือไทยเกือบทั้งหมดสามารถแยกได้ด้วยลักษณะเด่นต่างๆ ที่หาได้จากวิธีข้างต้น ยกเว้นตัวท่าตัวอักษร "M" "N" "S" "T" และ "Vowel base" เนื่องจากทั้ง 5 ท่าที่

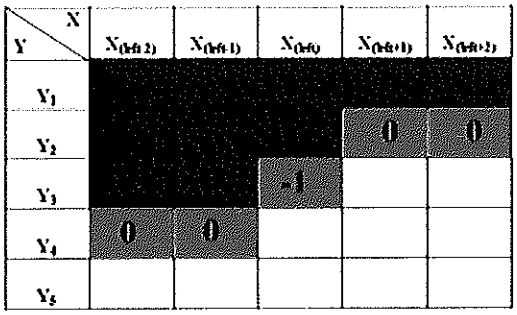
กล่าวมานั้นมีลักษณะคล้ายกันมากดังแสดงในภาพประกอบ 3-21 และทั้งหมดอยู่ในท่าประเภทสมมาตร ดังนั้นเพื่อความแม่นยำในการตีความหมายเราจึงเพิ่มลักษณะพิเศษคือ ความกว้างของนิ้วมือ ให้กับทั้ง 5 ท่าโดยเฉพาะ ซึ่งขั้นตอนนี้เราจะใช้กับภาพประเภทสมมาตร และหาเพียง 3 นิ้วจากด้านขวา



ภาพประกอบ 3-21 ท่าตัวอักษร "M" "N" "S" "T" และ "Vowel base"

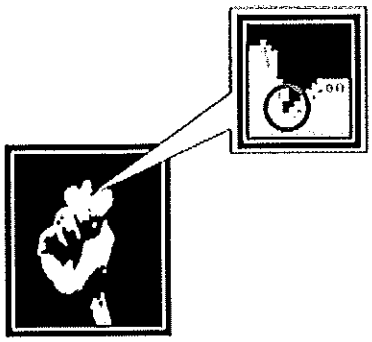
ขั้นตอนแรกของการหาความกว้างของนิ้วคือ การหาจุดสัมผัสระหว่าง 2 นิ้ว โดยคำนวนจากสมการที่ 3-4 ซึ่งเป็นการหาพารามิเตอร์ (Value $_{x(left)}$) ที่ใช้ในวิเคราะห์และการตัดสินใจว่าบริเวณใดเป็นจุดสัมผัสระหว่างนิ้ว โดยระบบจะทำการคำนวนหาค่า Value $_{x(left)}$ ประจำทุกพิเซลบนเส้นขอบที่แสดงรูปร่างรอบนอกของท่ามือ ดังตัวอย่างในภาพประกอบ 3-22

$$Value = Y_{x(left)} - Y_{x(left-1)} \tag{3-4}$$



ภาพประกอบ 3-22 ค่า Value $_{x(left)}$ ของแต่ละพิเซล

ภาพประกอบ 3-18 ช่องสีดำแทนพิกเซลที่เป็นพื้นหลัง ช่องสีขาวแทนพิกเซลพื้นที่มือ ส่วนช่องสีแดงเป็นบริเวณพิกเซลที่เป็นขอบนอกของมือ ซึ่งเป็นส่วนที่เราต้องหาค่า Value $_{x(left)}$ เพื่อนำมาพิจารณาหาสุดสัมผัสของนิ้วทั้งสองนิ้ว ซึ่งจุดที่สัมผัสคือบริเวณที่ ค่า Value $_{x(left)}$ เปลี่ยนจาก 0 หรือ จำนวนติดลบ เป็น จำนวนเต็มบวก ดังตัวอย่างในภาพประกอบที่ 3-23 และเมื่อหาจุดสัมผัสของนิ้วแต่ละคู่แล้วจะทำให้เราสามารถคำนวนความกว้างของนิ้วได้



ภาพประกอบ 3-23 ภาพแสดงบริเวณจุดสัมผัสระหว่าง 2 นิ้ว

3.6.7 **การระบุท่าพื้นฐานของเฟรมหลัก**

การระบุท่าพื้นฐานในแต่ละเฟรมหลักทำได้โดยการนำลักษณะเด่นที่หาได้จากกระบวนการหาลักษณะเด่น(หัวข้อ 3.6.1 - 3.6.6) มาเปรียบเทียบกับลักษณะเด่นต้นแบบของทุกท่าพื้นฐาน และเลือกท่าที่ใกล้เคียงมากที่สุด ซึ่งลักษณะเด่นที่สำคัญได้แก่ ประเภทท่าของมือ, ปลายนิ้วด้านบน, ปลายนิ้วด้านซ้าย, ปลายนิ้วด้านขวา, ปลายนิ้วด้านล่าง, การแยกนิ้วมือ, ลำดับของนิ้วมือ, นิ้วที่เล็กที่สุด (จากด้านขวา), สองนิ้วทางด้านซ้ายอยู่ใกล้กันน้อยกว่า ¼ ของความกว้างมือ ดังแสดงในตารางที่ 3-1

ตารางที่ 3-1 ลักษณะเด่นของท่าพื้นฐานทั้ง 24 ท่า

ท่าพื้นฐาน	ประเภทท่าของมือ	ปลายนิ้วด้านบน	ปลายนิ้วด้านซ้าย	ปลายนิ้วด้านขวา	ปลายนิ้วด้านล่าง	การแยกนิ้วมือ	ลำดับของนิ้วมือ	นิ้วที่เล็กที่สุด (จากด้านขวา)	สองนิ้วทางด้านซ้ายอยู่ใกล้กันน้อยกว่า ¼ ของความกว้างมือ
B	แนวตั้ง	4	*	*	*	ไม่แยก	*	*	*
C	สมมาตร	*	2	*	*	*	*	*	*
D,1	แนวตั้ง	1	*	*	*	*	*	*	*
F	แนวตั้ง	3	*	*	*	*	321	*	*
G	แนวนอน	*	2	*	*	*	*	*	ไม่ใช่
H	แนวนอน	*	2	*	*	*	*	*	ใช่
J	สมมาตร	4	*	*	*	ไม่แยก	1***	*	*
K,2	แนวตั้ง	2	*	*	0	แยก	*	*	*
L	สมมาตร	1	*	1	*	*	*	*	*
M	สมมาตร	*	*	*	*	ไม่แยก	**123	*	*
N	สมมาตร	*	*	*	*	ไม่แยก	*	นิ้วที่สาม	*
P	สมมาตร	<2	<2	*	2	*	*	*	*
R	แนวตั้ง	2	*	0	*	ไม่แยก	*	*	*
S	สมมาตร	4	*	*	*	ไม่แยก	4321	*	*
T	สมมาตร	*	*	*	*	ไม่แยก	**312	นิ้วที่สอง	*
W	แนวตั้ง	3	*	*	*	แยก	*1*	*	*
Y	สมมาตร	5	*	*	*	*	1***2	*	*
3	สมมาตร	2	*	1	*	แยก	*	*	*
4	แนวตั้ง	4	*	*	*	แยก	*	*	*
5	สมมาตร	4	*	1	*	แยก	*	*	*
Vowel base	สมมาตร	5	*	*	*	*	****1	*	*
Slow	แนวนอน	*	1	*	0	*	*	*	*

3.7 การแปลเป็นตัวอักษรภาษาไทย

กระบวนการในขั้นตอนนี้คือ การแปลลำดับอักขระที่ได้มาจากการระบุท่าพื้นฐานของแต่ละเฟรมหลักมาเป็นตัวอักษรภาษาไทย ซึ่งโดยปกติแล้ว เฟรมหลักแรกจะเป็นท่าเริ่มต้นแบบที่กำหนดไว้ ตามด้วย ท่าพื้นฐานในภาษามือไทยอีก 1-3 ท่า และจบด้วยท่าจบ เช่นเดียวกับท่าเริ่มต้นเป็นเฟรมหลักสุดท้าย แต่ในทางปฏิบัติ ข้อมูลที่เป็นอินพุตของส่วนนี้ไม่ถูกต้องเสมอไป เนื่องจากอาจจะเกิดความผิดพลาดในขั้นตอน การเลือกเฟรมหลัก การหาลักษณะ หรือ การระบุท่าพื้นฐานของเฟรมหลัก ซึ่งอาจมีผลทำให้จำนวนลำดับอักขระ ขาด เกิน หรือ ระบุท่าผิดพลาด การที่จะนำลำดับอักขระอินพุตนั้นมาเปรียบเทียบโดยตรงกับลำดับอักขระต้นแบบของแต่ละตัวอักษรภาษาไทยนั้น จะทำให้ระบบมีความผิดพลาดสูง ดังนั้นเราจึงได้นำ แบบจำลองฮิดเดนมาร์คอฟมาช่วยเพิ่มประสิทธิภาพการทำงานของระบบ เนื่องด้วย แบบจำลองฮิดเดนมาร์คอฟสามารถช่วยลดความผิดพลาดของปัญหาข้างต้นได้

กระบวนการรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟ เริ่มจากการสร้างแบบจำลองต้นแบบของตัวอักษรทั้งหมด 42 ตัวอักษรภาษาไทย โดยการนำลำดับอักขระที่ได้จากขั้นตอนการระบุท่าของตัวอักษรภาษาไทยแต่ละตัว ตัวละ 20 ชุดข้อมูล มาทำการสอนระบบตามรายละเอียดในหัวข้อที่ 2.12.1 ซึ่งจะทำให้ได้แบบจำลองต้นแบบของตัวอักษรภาษามือไทยทั้งหมด 42 แบบจำลอง เมื่อมีข้อมูลเข้ามาในระบบ ระบบจะทำการคำนวนค่าความน่าจะเป็นของแต่ละแบบจำลองต้นแบบ และเลือกค่าความน่าจะเป็นสุดสูดเป็นคำตอบข้อชุดข้อมูลที่เข้ามา

บทที่ 4
ผลการดำเนินงานและวิเคราะห์ผล

ในบทนี้กล่าวถึงผลการดำเนินงานและวิเคราะห์ผลการทำงานของระบบรู้จำภาษามือไทย โดยจะแบ่งการทดสอบระบบออกเป็น 5 ขั้นตอน คือ การเก็บข้อมูล การทดสอบความแม่นยำของการเลือกเฟรมหลัก การทดสอบความถูกต้องของการหาลักษณะเด่นเพื่อนำไประบุท่าพื้นฐานของเฟรมหลัก ทั้ง 2 วิธีที่กล่าวไปในหัวข้อ 3.5-3.6 การทดสอบขั้นตอนการแปลลำดับอักขระเป็นตัวอักษรภาษาไทย ซึ่งความผิดพลาดที่เกิดขึ้นจะขึ้นอยู่กับความแม่นยำของขั้นตอนก่อนหน้า เพราะอินพุตของขั้นตอนการแปลลำดับอักขระเป็นตัวอักษรภาษาไทยนั้นต้องผ่านการทำงานของขั้นตอนการเลือกเฟรมหลัก และขั้นตอนการระบุท่าพื้นฐาน โดยในส่วนนี้เราจะแยกการทดสอบเป็น 2 แบบ คือ การเปรียบเทียบลำดับอักขระข้อมูลอินพุตกับลำดับอักขระต้นแบบของอักษรภาษามือไทยโดยตรง กับการรู้จำโดยใช้ฮิดเดนมาร์คอฟ และขั้นตอนสุดท้ายคือ การทดสอบประสิทธิภาพของระบบโดยรวม ซึ่งแบ่งออกเป็น 2 ชุดข้อมูลคือ ข้อมูลที่เคยผ่านระบบ และข้อมูลที่ไม่เคยผ่านระบบ

4.1 การเก็บข้อมูล

ข้อมูลที่ใช้ในการทดลองนี้ถูกเก็บอยู่ในรูปแบบภาพถ่ายจากกล้องวีดิทัศน์นามสกุล .avi โดย ในระบบ RGB ซึ่งมีขนาดเท่ากับ 720×576 พิกเซล และควบคุมสภาวะสิ่งแวดล้อมขณะบันทึกภาพ ได้แก่ กล้องที่ใช้บันทึกคือ กล้องพานาโซนิค รุ่น NV-GS 400 ฉากหลังกำหนดให้เป็นสีเข้ม ปริมาณแสงไฟต้องเพียงพอที่จะบันทึกท่ามือของอาสาสมัครได้ชันเจน ระยะห่างจากกล้องและมือผู้ทำการทดลอง ประมาณ 60 เซ็นติเมตร และการถ่ายจะถ่ายเฉพาะส่วนมือผู้ทำการทดลอง และเพื่อความหลากหลายของชุดของมูลที่นำมาทดสอบระบบ ในงานวิจัยนี้จึงได้เก็บข้อมูลจากอาสาสมัครทั้งหมด 13 คน โดยแต่ละคนมีคุณลักษณะของมือ นิ้วมือ สีผิว และรูปแบบในการแสดงท่าทางที่ต่างกัน ซึ่งได้ถูกนำไปทดสอบความถูกต้องของระบบในส่วนต่าง โดยจะกล่าวอย่างละเอียดในหัวข้อ 4.2 - 4.4 เป็นลำดับต่อไป

4.2 การทดสอบการเลือกเฟรมหลัก

หัวข้อที่ 3.4 เสนอขั้นตอนในการเลือกเฟรมหลัก โดยใช้ออปติคัล โฟลว์ ในการตัดสินใจเลือกเฟรมหลัก ในหัวข้อนี้จึงนำเสนอผลการทดสอบของขั้นตอนดังกล่าว ระบบถูกทดสอบด้วยไฟล์ทั้งหมด 1560 ไฟล์ ซึ่งสุ่มจากไฟล์ที่ได้มาจากขั้นตอนการเก็บข้อมูล โดยข้อมูลที่

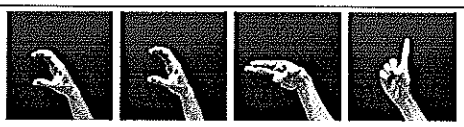
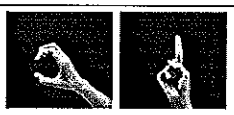
นำมาทดสอบครอบคลุมตัวอักษรภาษามือไทยทั้งหมด 42 ตัว (ยกเว้น ฃ ฅ) ดังแสดงในตารางที่ 4-1 นอกจากนี้ชุดทดสอบยังมีความหลากหลายทั้ง สีผิว ความกว้าง ยาว ของนิ้วมือ และฝ่ามือ อีกด้วย

ตารางที่ 4-1 ตัวอักษรภาษามือไทยทั้งหมด 42 ตัวอักษร

ก	ข	ค	ฆ	ง	จ	ฉ	ช	ซ	ฌ	ญ
ฎ	ฏ	ฐ	ฑ	ฒ	ณ	ด	ต	ถ	ท	ธ
น	บ	ป	ผ	ฝ	พ	ฟ	ภ	ม	ย	ร
ล	ว	ศ	ษ	ส	ห	ฬ	อ	ฮ		

ผลลัพธ์ของขั้นตอนนี้คือ เฟรมหลักท่าพื้นฐานของตัวอักษรภาษามือไทย ซึ่งประกอบไปด้วยลำดับภาพท่าพื้นฐานที่เป็นภาษามืออังกฤษ โดยจะพิจารณาเลือกเฟรมหลักจากหลักการในหัวข้อ 3.4 และบันทึกเฟรมหลักที่เลือกในรูปแบบไฟล์ภาพนามสกุล .jpg และมีขนาดเท่ากับ 320×240 พิกเซล ในกรณีที่ระบบสามารถเลือกเฟรมหลักได้ตรงตามท่าพื้นฐานของตัวอักษรภาษามือไทยนั้นๆจึงจะถือว่าระบบทำงานถูกต้อง ผลการทำงานแสดงดังตัวอย่างในตารางที่ 4-2

ตารางที่ 4-2 ตัวอย่างการทำงานของขั้นตอนการเลือกเฟรมหลัก

ลำดับ	ตัวอักษรภาษาไทย	จำนวนท่าพื้นฐาน	ผลการทำงาน	สรุปผลการทำงาน
1	ช	3		เกิน
2	ช	3		ขาด
3	ช	3		ไม่ชัด
4	ช	3		ถูก

ในตารางที่ 4-2 เป็นผลการเลือกเฟรมหลักจากวิทัศน์ 4 ไฟล์ ของตัวอักษร “ช” ซึ่งจะประกอบด้วยท่าพื้นฐานภาษาอังกฤษ 3 ท่าคือ “C” “H” และ “1” ในตัวอย่างนี้ระบบให้ผลลัพธ์เป็น 4 กรณีคือ

1. กรณีที่ 1 ระบบได้ทำการเลือกภาพตัวแทนเกินจำนวนท่าพื้นฐานซึ่งเฟรมที่ถูกเลือกคือ “C” “C” “H” และ “1” เมื่อวิเคราะห์จากข้อมูลที่นำมาทดสอบระบบพบว่าความผิดพลาดประเภทนี้ส่วนมาก เกิดจากผู้ที่ไม่ชำนาญในการใช้ภาษามือทำให้ท่าทางที่ถูกบันทึกไม่มีความต่อเนื่องกัน
2. กรณีที่ 2 ระบบเลือกเฟรมตัวแทนไม่ครบตามจำนวนท่าพื้นฐาน เฟรมที่ถูกเลือกคือ “C” และ “1” เกิดการแสดงท่าทางเร็วเกินไป ไฟล์วิดีทัศน์ที่บันทึกไว้จึงไม่มีช่วงที่นิ่งพอจะทำให้ระบบเลือกเป็นเฟรมหลักได้ ซึ่งพบทั้งในผู้ที่ชำนาญและไม่ชำนาญการใช้ภาษามือไทย
3. กรณีที่ 3 ระบบได้เลือกภาพตัวแทนได้ครบและถูกต้องตามจำนวนท่าพื้นฐาน แต่มีความผิดพลาดในเฟรมที่ 3 คือ ท่าตัวเลข “1” ภาพที่ได้ไม่ชัดเจน ซึ่งเกิดจากอาสาสมัครแสดงท่าทางช้าเกินไป ทำให้ระบบระบุช่วงที่ควรเป็นการเคลื่อนไหวเป็นช่วงที่มือหยุดนิ่งและเป็นผลทำให้เกิดการเลือกภาพตัวแทนผิดพลาด
4. กรณีที่ 4 ระบบสามารถเลือกเฟรมหลักได้อย่างถูกต้อง

อย่างไรก็ตามส่วนของขั้นตอนการเลือกเฟรมหลักยังมีโอกาสเกิดความผิดพลาดในรูปแบบอื่นๆได้ เช่น ระบบไม่สามารถเลือกเฟรมหลักได้เลย หรือ สามารถเลือกจำนวนเฟรมหลักได้ครบ และได้ภาพที่ชัดเจน แต่เป็นท่าที่ซ้ำกันซึ่งส่วนใหญ่จะขึ้นอยู่กับจังหวะในการแสดงท่าของผู้ใช้ระบบ และผลจากการทดสอบความแม่นยำในการเลือกเฟรมหลักของระบบเท่ากับ 85.5%

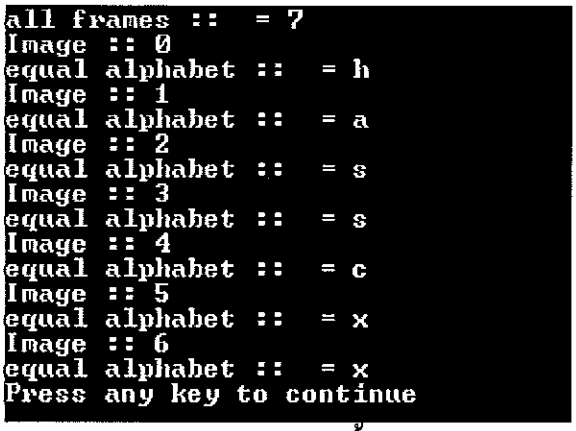
4.3 การทดสอบความถูกต้องของการหาลักษณะเด่นเพื่อนำไประบุท่าพื้นฐานของเฟรมหลัก

การทดสอบในขั้นตอนนี้แบ่งเป็นสองกรณี ตามหัวข้อ 3.5 และ 3.6 คือ การหาลักษณะเด่นของมือเพื่อนำไประบุท่าพื้นฐานโดยการหามุมที่สนใจในเฟรมหลัก และการหาลักษณะเด่นของมือนำไประบุท่าพื้นฐานโดยวิเคราะห์รูปร่างมือและนิ้วมือ ซึ่งในหัวข้อ 4.3 นี้จึงนำเสนอผลการทำงานของขั้นตอนดังกล่าว

4.3.1 วิธีการหามุมที่สนใจในเฟรมหลัก

การระบุท่าโดยการหามุมที่สนใจนี้ จะเริ่มจากการนำภาพอินพุตมาเปรียบเทียบกับ 24 ท่าพื้นฐานที่เก็บไว้เป็นฐานข้อมูลตามขั้นตอนที่ 3.5 และเลือกภาพที่มีความคล้ายคลึงกันมากที่สุด โดยระบบถูกทดสอบด้วยเฟรมหลักจำนวน 240 เฟรมที่เลือกมาจากข้อมูลที่บันทึกไว้จากหัวข้อ 4.1 และให้ผู้เชี่ยวชาญเป็นผู้เลือกเฟรมที่มีความถูกต้องของแต่ละท่า ซึ่งครอบคลุมทั้งหมด 24 ท่าพื้นฐานภาษาอังกฤษ ท่าละ 10 เฟรม แบ่งออกเป็น 2 กลุ่มคือ ชุดข้อมูลภาพที่บันทึกจากคน

เดียวกันกับที่ใช้เป็นฐานข้อมูล ท่าละ 5 เฟรม และ ชุดข้อมูลภาพที่บันทึกจากบุคคลที่ไม่ได้บันทึกเป็นภาพฐานข้อมูล ท่าละ 5 เฟรม เอาท์พุตของระบบคือตัวอักษร ดังตัวอย่างในภาพประกอบ 4-1



ภาพประกอบ 4-1 ผลการทำงานของการระบุท่าพื้นฐานด้วยการหามุมที่สนใจในภาพ

ผลจากการทดสอบพบว่า เมื่อทดสอบกับภาพที่มาจากคนเดียวกันกับฐานข้อมูล จำนวน 120 เฟรม ระบบมีความถูกต้อง 75.25% แต่เมื่อทดสอบกับชุดข้อมูลที่มาจากบุคคลอื่น ระบบมีความถูกต้อง 68% เนื่องจากลักษณะมือของแต่ละคนแตกต่างกันไปทำให้เมื่อเปลี่ยนผู้ใช้งานความถูกต้องจึงลดลง และจากการใช้ภาพเพียงภาพเดียวเป็นฐานข้อมูลในการเปรียบเทียบก็ย่อมจะทำให้ระบบมีความยืดหยุ่นน้อย ซึ่งค่าความแม่นยำที่ได้ไม่สามารถยอมรับได้ เนื่องจากจะส่งผลกระทบให้ค่าความถูกต้องในขั้นตอนต่อไปผิดพลาดด้วย ดังนั้นในงานวิจัยนี้จึงได้พัฒนาวิธีระบุท่าพื้นฐานขึ้นมาใหม่ ซึ่งจะกล่าวรายละเอียดในหัวข้อ 4.3.2

4.3.2 การหาลักษณะเด่นของมือ และการระบุท่าพื้นฐานโดยวิเคราะห์รูปร่างมือและนิ้วมือ

ระบบถูกทดสอบด้วย เฟรมหลักจำนวน 720 เฟรมที่เลือกมาจากข้อมูลที่บันทึกไว้จากหัวข้อ 4.1 โดยให้ผู้เชี่ยวชาญเป็นผู้เลือกเฟรมที่มีความถูกต้องของแต่ละท่า ครอบคลุมทั้งหมด 24 ท่าพื้นฐานที่เป็นภาษาอังกฤษ ท่าละ 30 เฟรม ซึ่งผลการทดสอบแสดงตามตารางที่ 4-3

ตารางที่ 4-3 ผลการประเมินการทดสอบเพื่อระบุท่าพื้นฐานจากเฟรมหลักจำนวน 720 เฟรม (Confusion matrix)

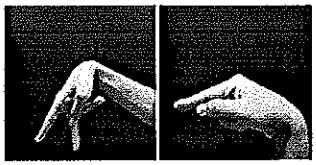
ท่าที่ผ่านการจำแนกจากระบบ	ท่าพื้นฐานในเฟรม																								
	d	k	l	c	j	y	3	2	t	b	n	m	s	Base	r	4	w	f	g	p	q	h	5	l	
d	29																								29
k		28																							28
l			30	6	2			1					1		1										41
c				19																					19
j		1			24	2	1					2					1			4					35
y					3	24																			27
3		1				2	26	2										3							34
2							1	27																	28
t							1		26	5	2	1	1		1	1	1								39
b									3	24	10		5		1										43
n										1	17	1	1		1										21
m											1	26	3	1						2					33
s						1			1				19	9						1					31
a														19											19
r															26										26
4																25	1								26
w																4	27	1							32
f							1											26							27
g																			25	2					27
p	1			2										1					2	14					20
q				2	1	1													3	7	30				44

		ท่าพื้นฐานในแฟรม																							
		d	k	l	c	j	y	3	2	t	b	n	m	s	Base	r	4	w	f	g	p	q	h	5	l	
ท่าที่ผ่านการจำแนกจากระบบ	h																						30			30
	5																							30		30
	l				1																				29	30
	นิยามไม่ได้																								1	1
	รวม																									720
	% ความผิดพลาด	3.3	6.7	0	36.7	20	20	13.3	10	13.3	20	43.3	13.3	36.7	36.7	13.3	16.7	10	13.3	16.7	53.3	0	0	0	3.3	16.7

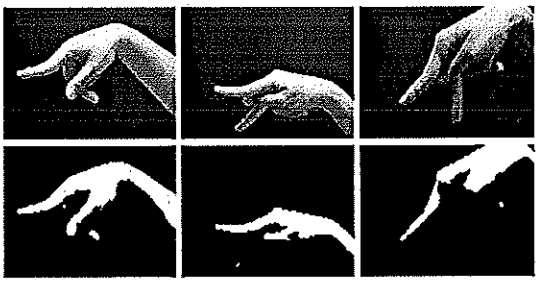
จากตารางที่ 4-3 เป็นการประเมินผลทดสอบการระบุท่าพื้นฐานของเฟรมหลัก จำนวน 720 เฟรม ซึ่งระบบระบุท่าผิดพลาด 120 เฟรมจากทั้งหมด 720 เฟรม ความถูกต้องเฉลี่ยอยู่ที่ 83.3 % เมื่อทำการวิเคราะห์การทดสอบพบว่า ความผิดพลาดส่วนใหญ่จะเกิดขึ้นกับตัวอักษรประเภทสมมาตร(หัวข้อ 3.6.2) คือ "Vowel base" "M" "N" "S" "T" และ "P" เนื่องจาก ตัวอักษรประเภทสมมาตรนั้นมีลักษณะที่ใกล้เคียงกันมาก จึงยากต่อการจำแนกประเภท ส่วนตัวอักษร P ที่เป็นท่าคว่ำมือ ทำให้เกิดความผิดพลาดได้ง่าย เนื่องจากการคว่ำมือของแต่ละคนองศาแตกต่างกันออกไป ซึ่งทำให้การดึงลักษณะเด่นที่จะนำมาใช้ระบุท่าทำได้ยาก จึงมีความผิดพลาดสูง โดยผลการทดสอบแสดงตามตารางที่ 4-3 แถวแรกคือ ท่าพื้นฐานในเฟรมหลักที่ระบุด้วยผู้เชี่ยวชาญ หลักแรกคือผลที่ได้จากการจำแนกท่าพื้นฐานในเฟรมหลักด้วยระบบ ซึ่งระบบมีความผิดพลาดรวม16.7% เมื่อพิจารณาตัวอักษรที่ระบบระบุผิดพลาดบ่อยจะพบว่า ส่วนใหญ่เป็นท่าที่มีลักษณะใกล้เคียงกันหรือ ค่าที่ใช้ในการพิจารณาเพื่อระบุท่าพื้นฐานใกล้เคียงกัน(ตารางที่ 3-1) ตัวอย่างเช่น

- ระบบระบุท่าตัวอักษร "C" เป็น ตัวอักษร "D" เนื่องด้วย เมื่อพิจารณาตามลักษณะเด่นที่นำมาใช้ในการระบุท่า ดังแสดงตารางที่ 3-1 จะพบว่า ความแตกต่างที่เด่นชัดของตัวอักษร "C" และ "D" คือตัวอักษร "C" จะเป็นท่าประเภทสมมาตร และตัวอักษร "D" จะเป็นภาพแนวตั้ง ส่วนลักษณะเด่นอื่นที่ใช้ในการระบุท่านั้นมีโอกาสที่ตัวอักษร "C" จะให้ผลตรงตามลักษณะของตัวอักษร "D" ได้ ขึ้นอยู่กับลักษณะการแสดงท่าของผู้ใช้ระบบ ดั้งนั้น ถ้าระบบแยกประเภทการวางตัวของท่า "C" ผิด ก็มีโอกาสที่ระบบจะระบุ เฟรมหลักฐาน "C" นั้นเป็นตัวอักษร "D" โดยเฉพาะกับลักษณะมือที่ผอม และนิ้วยาว
- ระบบระบุท่าตัวอักษร "B" เป็น ตัวอักษร "T" เนื่องด้วย ท่าตัวอักษรของ "B" เป็นภาพประเภทแนวตั้ง ส่วนตัว "T" เป็นภาพประเภทสมมาตร แต่ลักษณะนิ้วที่สูงที่สุดของท่า "B" และ "T" เหมือนกัน คือ นิ้วลำดับที่ 2 จากทางขวาของเฟรม ดังนั้น ถ้าระบบแยกประเภทการวางตัวของท่าผิด ก็จะทำให้ระบบระบุท่าผิดด้วยเช่นกัน
- ระบบระบุท่าตัวอักษร "S" เป็น ตัวอักษร "N" หรือ "Vowel base" เป็น "S" เนื่องด้วย ท่าตัวอักษร "S" "N" "Vowel base" อยู่ในกลุ่มท่าประเภทสมมาตร ซึ่งมีท่าที่คล้ายกันมาก จึงทำให้เกิดความผิดพลาดบ่อย
- ระบบระบุท่าตัวอักษร "P" เป็น ตัวอักษร "Slow" เนื่องด้วย ตัวอักษร "P" เป็นท่าที่อยู่ในลักษณะคว่ำมือ ทำให้บางครั้งนิ้วมือทางด้านล่างโดนเงาบัง ดังนั้นลักษณะ

ของท่าที่เหลือจะใกล้เคียงกับลักษณะของ ท่าตัวอักษร “Slow” หรือในบางกรณีองศาของนิ้วด้านล่างก็มีผลทำให้ระบบระบุท่าตัวอักษร “P” เป็นตัวอักษร “Slow” ได้ ดังตัวอย่างในภาพประกอบ 4-2 เป็นลักษณะของท่าตัวอักษร “P” และตัวอักษร “Slow” ที่ถูกต้อง และภาพประกอบ 4-3 เป็นลักษณะของท่าตัวอักษร “P” ที่ระบบระบุผิดพลาด ทั้งในรูปแบบเฟรมอินพุตและเมื่อแปลงเป็นภาพขาวดำ



ภาพประกอบ 4-2 ลักษณะของท่าตัวอักษร “P” และตัวอักษร “Slow” ที่ถูกต้อง



ภาพประกอบ 4-3 ลักษณะของท่าตัวอักษร “P” ที่ระบบระบุผิดพลาด

ผลจากหัวข้อ 4.2.1 และ 4.2.2 ทำให้เราสรุปได้ว่าวิธีที่เหมาะสำหรับการหาลักษณะเด่น และการระบุท่านั้น คือ การวิเคราะห์จากรูปร่างมือและนิ้วมือ เพราะ ระบบมีความถูกต้องสูงกว่า วิธีการหามุมที่สนใจในภาพถึง 15.3% จากนั้นข้อมูลจะถูกส่งไปให้ขั้นตอนการแปลลำดับอักขระเป็นตัวอักษรภาษามือไทย ซึ่งสามารถสรุปผลการทดสอบได้ดังรายละเอียดในหัวข้อ 4.4

4.4 การทดสอบการแปลลำดับอักขระเป็นตัวอักษรภาษาไทย

การทดสอบการแปลลำดับอักขระภาษาอังกฤษเป็นตัวอักษรไทยนั้นความผิดพลาดส่วนใหญ่จะมาจากขั้นตอนการเลือกเฟรมหลัก และการระบุท่าพื้นฐาน เนื่องจากข้อมูลที่จะนำมาประมวลผลในส่วนนี้ ต้องผ่านกระบวนการเลือกเฟรมหลัก การหาลักษณะเด่น และการระบุท่าพื้นฐาน มาก่อนเพื่อนำลำดับตัวอักษรของท่าพื้นฐานภาษาอังกฤษที่ได้มาเป็นอินพุตของขั้นตอนนี้ โดยเราได้แบ่งการทดสอบออกเป็น 2 วิธีคือ การเปรียบเทียบโดยตรง กับการรู้จำโดยใช้แบบจำลอง

ฮิดเดนมาร์คอฟ โดยใช้ไฟล์ท่าภาษามือไทยจำนวน 840 ไฟล์ ครอบคลุม 42 ตัวอักษรภาษามือไทย ตัวละ 20 ไฟล์วิทัศน์ ซึ่งจุดประสงค์ของการทดลองนี้คือ เพื่อแสดงให้เห็นว่าการพัฒนาระบบด้วยแบบจำลองฮิดเดนมาร์คอฟสามารถเพิ่มประสิทธิภาพของระบบรู้จำภาษามือไทยได้

4.4.1 **การเปรียบเทียบโดยตรง**

การเปรียบเทียบโดยตรงคือการนำ ลำดับของตัวอักษรที่ได้จากกระบวนการระบุท่าพื้นฐานเป็นภาษาอังกฤษ(หัวข้อที่ 3.6.7) มาเปรียบเทียบทีละตัว ว่าลำดับอักขระนั้นตรงกับลำดับอักขระของตัวอักษรภาษาไทยตัวใด(ตารางที่ 2-1) จากข้อมูลจำนวน 840 ท่าอักษรภาษาไทย มีความแม่นยำ 59.52%

4.4.2 **การรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟ**

การรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟ คือการนำชุดข้อมูลที่ได้จากกการระบุท่าพื้นเป็นอักขระภาษาอังกฤษมาผ่านแบบจำลองตัวอักษรภาษาไทย ที่สร้างขึ้นจากการสอนระบบตามขั้นตอนในหัวข้อ 3.7 จำนวน 42 แบบจำลอง เมื่อมีอินพุตเข้ามาระบบจะหาแบบจำลองที่มีความน่าจะเป็นสูงสุดเป็นคำตอบของช้อมูลนั้นๆ จากข้อมูลจำนวน 840 ท่าอักษรภาษาไทย ระบบมีความแม่นยำ 78.21%

จากการทดสอบระบบทั้งแบบการเปรียบเทียบโดยตรง และการรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟ พบว่าการรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟมีความยืดหยุ่นในการแปลลำดับอักขระท่าพื้นฐานเป็นตัวอักษรภาษาไทยมากกว่าการเปรียบเทียบโดยตรงถึง 18.69 % ดังเห็นได้จากตารางที่ 4-4 ซึ่งผลการทดสอบจากชุดข้อมูลเดียวกันแต่การแปลเป็นตัวอักษรภาษาไทยโดยใช้แบบจำลองฮิดเดนมาร์คอฟกลับให้ค่าความถูกต้องที่สูงกว่า

ตารางที่ 4-4 ผลการทดสอบการแปลลำดับอักขระท่าพื้นฐานเป็นตัวอักษรภาษามือไทย

ตัวอักษร	ประกอบด้วย (ท่าพื้นฐาน)	ผลการเปรียบเทียบ		ใช้แบบจำลองฮิดเดนมาร์คอฟ	
		ถูก	ผิด	ถูก	ผิด
ก	K	19	1	20	0
ข	K+1	15	5	16	4
ค	K+2	14	6	16	4
ฆ	K+3	10	10	15	5
ง	N+G	5	15	15	5
จ	J	17	3	18	2

ตัวอักษร	ประกอบด้วย (ท่าพื้นฐาน)	ผลการเปรียบเทียบ		ใช้แบบจำลองฮิดเดนมาร์คอฟ	
		ถูก	ผิด	ถูก	ผิด
ฉ	C+H	8	12	14	6
ช	C+H+1	10	10	14	6
ซ	S+Slow	8	12	19	1
ฌ	C+H+2	4	16	16	4
ญ	Y+1	12	8	15	5
ฎ	D+1	15	5	17	3
ฏ	T+5	12	8	20	0
ฐ	T+2	12	8	18	2
ฑ	T+4	7	13	19	1
ฒ	T+3	0	20	14	6
ณ	N+1	5	15	9	11
ด	D	19	1	19	1
ต	T	16	4	16	4
ถ	T+1	14	6	14	6
ท	T+H	13	7	17	3
ธ	T+H+1	0	20	14	6
น	N	12	8	12	8
บ	B	19	1	19	1
ป	P+1	7	13	13	7
ผ	P+2	5	15	15	5
ฝ	F+1	14	6	16	4
พ	P	6	14	9	11
ฟ	F	18	2	18	2
ภ	P+3	5	15	14	6
ม	M	15	5	15	5
ย	Y	14	6	14	6
ร	R	12	8	14	6
ล	L	19	1	19	1

ตัวอักษร	ประกอบด้วย (ท่าพื้นฐาน)	ผลการเปรียบเทียบ		ใช้แบบจำลองฮิดเดนมาร์คอฟ	
		ถูก	ผิด	ถูก	ผิด
ว	W	18	2	18	2
ศ	S+1	10	10	14	6
ย	S+2	16	4	16	4
ส	S	12	8	12	8
ห	H	18	2	18	2
พ	L+1	18	2	18	2
อ	Vowel base	12	8	11	9
ฮ	H+1	15	5	17	3
รวม		500	340	657	183
คิดเป็น (%)		**59.52**	40.48	**78.21**	21.79

เหตุผลที่ทำให้ค่าความถูกต้องของการแปลลำดับอักขระเป็นตัวอักษรภาษาไทยโดยใช้แบบจำลองฮิดเดนมาร์คอฟสูงกว่าวิธีการเปรียบเทียบโดยตรง เนื่องจากการรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟสามารถเรียนรู้รูปแบบของลำดับตัวของท่าพื้นฐานได้จากชุดข้อมูลที่ได้ทำการสอนไปในขั้นตอนการสร้างแบบจำลองต้นแบบ(หัวข้อที่ 3.7) และถ้าเราเพิ่มชุดข้อมูลในการสอนระบบเพื่อทำการสร้างแบบจำลองต้นแบบมากขึ้น ประสิทธิภาพในการรู้จำของระบบก็จะสูงตามไปด้วย ดังนั้นเราจึงนำแบบจำลองฮิดเดนมาร์คอฟมาช่วยเพิ่มประสิทธิภาพของงานวิจัยชิ้นนี้ ซึ่งสามารถช่วยเพิ่มความถูกต้องในบางกรณีได้ ดังตัวอย่างในตารางที่ 4-5 เป็นตัวอย่างการเปรียบเทียบการทำงานของการเปรียบเทียบโดยตรงและการรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟ

ตารางที่ 4-5 ตัวอย่างผลการเปรียบเทียบโดยตรง และ การรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟ

ตัวอักษร	ประกอบด้วย (ท่าพื้นฐาน)	ชุดข้อมูลที่ได้จาก ขั้นตอนระบุท่าพื้นฐาน	ผลการ เปรียบเทียบ	ใช้แบบจำลอง ฮิดเดนมาร์คอฟ
ง	N+G	T+G	ผิด	ถูก
ฐ	T+2	N+T+2	ผิด	ถูก
ษ	S+2	T+2	ผิด	ผิด

ในตารางนี้ประกอบด้วยข้อมูล 5 หลัก หลักแรกคือ ตัวอักษรภาษามือไทยที่นำมาทดสอบ ซึ่งถูกบันทึกไว้ในรูปไฟล์วิดีทัศน์นามสกุล .avi หลักที่สองคือท่าพื้นฐานที่ประกอบเป็นตัวอักษรภาษามือไทยในหลักแรก หลักที่สามคือ ผลที่ได้จากการทำงานของระบบในขั้นตอนการระบุท่า หลักที่สี่คือ ผลการแปลลำดับอักขระท่าพื้นฐานจากหลักที่สามโดยวิธีเปรียบเทียบโดยตรง และหลักสุดท้ายคือ ผลการแปลลำดับอักขระท่าพื้นฐานจากหลักที่สามด้วยการรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟ ซึ่งจะเห็นได้ว่าการใช้แบบจำลองฮิดเดนมาร์คอฟ มีความถูกต้องสูงกว่า แต่ทั้งนี้ความสามารถในการระบุท่าพื้นฐานของระบบ ก็ขึ้นอยู่กับชุดข้อมูลที่นำมาการสอบระบบด้วย ยกตัวอย่างเช่น “ง” ซึ่งประกอบด้วยท่าพื้นฐาน “N” ตามด้วย “G” แต่เนื่องด้วยลำดับอักขระที่นำมาสอนระบบเพื่อสร้างต้นแบบ “ง” นั้น ในขั้นตอนการระบุท่าพื้นฐาน ระบบระบุท่าตัวอักษร “N” เป็น “T” ผิดบ่อยครั้ง ทำให้เมื่อลำดับอักขระที่เป็น “T” ตามด้วย “G” เข้ามาในระบบ ระบบจึงตัดสินใจเลือก “ง” เป็นคำตอบ เนื่องจากมีค่าความน่าจะเป็นของแบบจำลอง “ง” สูงสุด แต่สำหรับตัวอักษร “ษ” ที่ประกอบจาก “S” ตามด้วย “2” นั้น เมื่อเกิดความผิดพลาดในขั้นตอนการระบุท่าฐานพื้นในเฟรมหลักตามตารางที่ 4-4 ซึ่งระบุบตัวอักษร “S” เป็นตัวอักษร “T” ทำให้ลำดับอักขระที่ได้เป็นตัวอักษร “T” ตามด้วย “2” ระบบก็จะตัดสินใจเลือก “ฐ” เป็นคำตอบ ถึงแม้ว่าลำดับอักขระที่นำมาสอนเพื่อสร้าง แบบจำลองต้นแบบของตัวอักษร “ษ” จะมี “T” ตามด้วย “2” ปะปนอยู่ด้วย แต่เมื่อนำความน่าจะเป็นจากแบบจำลอง “ษ” เปรียบเทียบกับความน่าจะเป็นจากแบบจำลอง “ฐ” แล้ว แบบจำลอง “ฐ” กลับให้ค่าความน่าจะเป็นที่สูงกว่า เนื่องด้วย ชุดข้อมูลที่นำมาสอนเพื่อสร้างแบบจำลองต้นแบบของตัวอักษร “ฐ” มีลำดับอักขระที่เป็น “T” ตามด้วย “2” มากกว่าชุดข้อมูลที่นำมาสอนเพื่อสร้างแบบจำลองต้นแบบของตัวอักษร “ษ” ทำให้ระบบทำงานผิดพลาด

เมื่อได้วิธีที่เหมาะสมในการรู้จำตัวอักษรภาษามือไทยในทุกขั้นตอนแล้ว จึงทำการทดสอบภาพรวมระบบ โดยแยกการทดสอบเป็น 2 ประเภทคือ ชุดข้อมูลที่เคยผ่านการสอนระบบ และชุดข้อมูลที่ไม่เคยผ่านการสอนระบบ

4.5 การทดสอบประสิทธิภาพการทำงานของระบบ

จากหัวข้อ ที่ 4.1-4.4 ทำให้สามารถเลือกเทคนิคที่เหมาะสมกับแต่ละขั้นตอนในระบบรู้จำภาษามือไทย ในหัวข้อที่ 4.5 นี้จะเป็นการทดสอบภาพรวมของระบบโดยระบบถูกทดสอบด้วย 420 ไฟล์วิดีทัศน์ ครอบคลุม 42 ตัวอักษรภาษามือไทย ตัวอักษรและ 10 เฟรม ซึ่งแบ่งเป็นสองประเภทประเภทละ 210 ไฟล์วิดีทัศน์ คือ ชุดข้อมูลที่เคยผ่านการสอนระบบซึ่งสุ่มเลือกจากข้อมูลที่นำไปสอนเพื่อสร้างแบบจำลองต้นแบบทั้ง 42 ตัวอักษรภาษามือไทย และชุดข้อมูลที่ไม่เคยผ่านการสอนระบบ โดยมีผลความผิดพลาดตามตารางที่ 4-5 ข้อมูลในตาราง

ประกอบด้วย ตัวอักษรภาษาไทยซึ่งตรงกับไฟล์วิดีทัศน์ที่นำมาทดสอบ หลักที่สองคือ ตัวอักษรในหลักแรกประกอบด้วยท่าพื้นฐานที่เป็นภาษาอังกฤษใดบ้าง หลักที่สามคือ จำนวนไฟล์วิทัศน์ที่ระบบทำการระบุท่าพื้นฐานผิดพลาดจากทั้งหมด 5 ไฟล์วิดีทัศน์ ในกรณีที่ไฟล์ดังกล่าวเคยผ่านการสอนระบบเพื่อสร้างแบบจำลองต้นแบบมาก่อน หลักที่สี่คือ %ความผิดพลาดที่คิดจาก (ค่าในหลักที่สาม x 100)/5 ส่วนค่าในหลักที่ห้าและหก เหมือนกับค่าในหลักที่สามและสี่ แต่เป็นข้อมูลของผู้ระบบที่ไม่เคยผ่านการสอนระบบ

ตารางที่ 4-6 ตารางแสดงความผิดพลาดของระบบ

ตัวอักษร	ประกอบด้วย	ผ่านการสอนระบบ	% ความผิดพลาด	ไม่ผ่านการสอนระบบ	% ความผิดพลาด
ก	K	0	0	2	40
ข	K+1	0	0	2	40
ค	K+2	1	20	0	0
ฆ	K+3	0	0	2	40
ง	N+G	1	20	1	20
จ	J	1	20	3	60
ฉ	C+H	0	0	0	0
ช	C+H+1	1	20	2	40
ซ	S+Slow	1	20	4	80
ฌ	C+H+2	1	20	4	80
ญ	Y+1	1	20	1	20
ฎ	D+1	0	0	0	0
ฏ	T+5	0	0	0	0
ฐ	T+2	1	20	2	40
ฑ	T+4	0	0	1	20
ฒ	T+3	2	40	2	40
ณ	N+1	2	40	3	60
ด	D	0	0	0	0
ต	T	1	20	0	0
ถ	T+1	1	20	2	40
ท	T+H	0	0	1	20
ธ	T+H+1	0	0	1	20
น	N	3	60	2	40

ตัวอักษร	ประกอบด้วย	ผ่านการสอนระบบ	% ความผิดพลาด	ไม่ผ่านการสอนระบบ	% ความผิดพลาด
บ	B	1	20	1	20
ป	P+1	1	20	1	20
ผ	P+2	2	40	1	20
ฝ	F+1	2	40	1	20
พ	P	2	40	4	80
ฟ	F	0	0	1	20
ภ	P+3	3	60	0	0
ม	M	1	20	1	20
ย	Y	0	0	2	40
ร	R	3	60	0	0
ล	L	1	20	1	20
ว	W	1	20	1	20
ศ	S+1	1	20	1	20
ษ	S+2	1	20	3	60
ส	S	1	20	1	20
ห	H	0	0	0	0
ฬ	L+1	0	0	1	20
อ	Vowel base	2	40	2	40
ฮ	H+1	0	0	1	20
รวม		39	18.57	58	28.1

จากตารางจะเห็นได้ว่าชุดข้อมูลที่เคยผ่านการสอนระบบจะมีความถูกต้องมากกว่าชุดข้อมูลที่ไม่เคยผ่านการสอนระบบ เนื่องจากระบบไม่คุ้นเคยกับลำดับอักขระของอาสาสมัครที่ไม่เคยผ่านการสอนระบบระบบ ยกเว้นตัวอักษร “ร” “พ” และ “น” ในกรณีที่ทดสอบกับชุดข้อมูลของอาสาสมัครที่เคยผ่านการสอนระบบ ในขั้นตอนการสุ่มได้สุ่มไฟล์ที่ผิดพลาดขึ้นมามาก แต่เมื่อพิจารณาระบบรวมแล้วพบว่าค่าความถูกต้องเฉลี่ยยังคงสอดคล้องกับหลักการของฮิดเดนมาร์คอฟซึ่งชุดข้อมูลที่เคยผ่านการสอนระบบมีความถูกต้องสูงกว่า ดังแสดงในตารางที่ 4-5

ความผิดพลาดของระบบรู้จำภาษามือไทยเกิดจากสองส่วนที่สำคัญคือ การเลือกเฟรมหลัก และ การหาลักษณะเด่นของมือในเฟรมหลักเพื่อระบุท่าของภาพในเฟรม สาเหตุหลักที่ทำให้เกิดความผิดพลาด ในขั้นตอนการเลือกเฟรมหลัก ส่วนใหญ่มาจากความไม่ชำนาญของผู้ใช้

ระบบ ผู้ใช้งานอาจเคลื่อนไหวมือติดขัด ซึ่งจะทำให้ได้เฟรมหลักเกินความต้องการ หรือทำท่าเร็วมากจนทำให้ระบบไม่สามารถเก็บเฟรมที่สำคัญไว้ได้ทัน ซึ่งเป็นผลทำให้ตัวอักษรภาษาไทยตัวที่ประกอบด้วยท่าพื้นฐานมากกว่า 1 ท่าขึ้นไปมีโอกาสเกิดความผิดในขั้นตอนการเลือกเฟรมหลักมากขึ้น และส่งผลกระทบต่อการทำงานในขั้นตอนถัดไปผิดพลาดเพิ่มขึ้นด้วย โดยสามารถดูผลได้จากตารางที่ 4-5 ในกรณีท่าที่แตกต่างกันชัดเจน กลุ่มตัวอักษรท่าเดี่ยวส่วนใหญ่จะมีความผิดพลาด น้อยกว่า กลุ่มตัวอักษรที่มีท่ามากกว่า 1 ท่าขึ้นไปเช่นตัวอักษร “ด” และ “ห” เป็นต้น และความผิดพลาดอีกประการหนึ่งคือ การระบุว่าท่าในท่าในเฟรมหลักคือท่าใด จากตารางที่ 4-3 ความผิดพลาดส่วนใหญ่เกิดขึ้นกับ ตัวอักษรที่ประกอบด้วยตัวอักษร “N” “S” “P” และ ตัว “Vowel base” เนื่องด้วยตัวอักษร “P” ซึ่งเป็นท่าคว่ำมือซึ่งมีการเคลื่อนไหวมาก ทำให้เกิดความผิดพลาดทั้งจากผู้ใช้ระบบและการประมวลผลได้ง่าย ส่วนตัวอักษร “N” “S” และ “Vowel base” จัดอยู่ในท่าประเภทสมมาตร ซึ่งมีลักษณะใกล้เคียงกันมากทำให้เกิดความผิดพลาดในการระบุเป็นท่าพื้นฐานบ่อยครั้ง ดังนั้นตัวอักษรที่ประกอบด้วยท่าพื้นฐานเหล่านี้จะให้ผลที่ตรงกันข้ามกับตัวอักษรอื่นคือ กลุ่มตัวอักษรมีท่าพื้นฐานเพียงเดียวที่ประกอบด้วย “N” “S” “P” และ ตัว “Vowel base” จะมีค่าความผิดพลาดสูงกว่า กลุ่มตัวอักษรที่มีท่าพื้นฐานมากกว่า 1 ท่าขึ้นไป เนื่องจากตัวอักษรภาษาไทยที่มีท่าพื้นฐานเพียงท่าเดียวนั้น เมื่อขั้นตอนระบบระบุท่าพื้นฐานทำงานผิดพลาดคำตอบในขั้นการแปลลำดับอักขระเป็นตัวอักษรภาษาไทยก็จะผิดด้วย ตัวอย่างเช่น ระบบรับอินพุตเป็นไฟลวิดีทัศน์ของตัวอักษร “น” เข้ามา ซึ่งมีท่าพื้นฐานเพียงท่าเดียวคือตัวอักษร “N” แต่ในขั้นตอนของการระบุท่าพื้นฐานได้ระบุผิดเป็นตัวอักษร “S” ทำให้เมื่อผ่านขั้นตอนการแปลลำดับอักขระเป็นตัวอักษรภาษาไทยระบบก็จะเลือกตัวอักษร “ส” เนื่องจากมีความน่าจะเป็นจากแบบจำลอง “ส” สูงที่สุด ซึ่งเป็นคำตอบที่ผิด ส่วนในกรณีที่มีท่าหลักมากกว่า 1 ท่า ถึงแม้ว่าท่าแรกระบบจะประมวลผลไม่ถูกต้อง แต่ถ้าท่าในลำดับถัดไประบบระบุถูกต้อง ฮิดเดนมาร์คอฟก็ยังสามารถนำข้อมูลนั้นไปหาความน่าจะเป็นสูงสุดในการแปลเป็นตัวอักษรภาษาไทยได้ ซึ่งเราจะเห็นได้ได้ชัดเจนในกรณี ของ ฏ (T ตามด้วย 5) และ ฑ (T ตามด้วย 4) ที่มีความผิดพลาดน้อยมาก เพราะท่าพื้นฐานที่สองของทั้งสองตัวอักษร คือ 4 และ 5 ไม่ได้ถูกใช้เพื่อประกอบเป็นอื่นตัวอักษรภาษาไทยตัวอื่นเลยทำให้ระบบสามารถตัดสินใจเลือกคำตอบได้ถูกต้อง และในกรณีที่ตัวอักษรใดประกอบด้วยท่าพื้นฐานที่ระบุยาก ความถูกต้องก็จะลดลงตามไปด้วย

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

ในบทนี้จะกล่าวถึงการสรุปผลและข้อเสนอแนะที่ได้จากการดำเนินการทำงานวิจัยตลอดจนปัญหาและอุปสรรคที่เกิดขึ้นขณะทำงานวิจัย และให้ข้อเสนอแนะแก่ผู้สนใจจะนำงานวิจัยชุดนี้ไปพัฒนาต่อ

5.1 สรุปผล

วิทยานิพนธ์ชุดนี้นำเสนอวิธีการที่เหมาะสมในการสร้างระบบรู้จำตัวอักษรภาษามือไทย โดยไม่ใชอุปกรณ์ช่วยในการตรวจจับการเคลื่อนที่ ซึ่งแบ่งการทำงานออกเป็น 3 ส่วน คือ

1. การเลือกเฟรมหลัก ซึ่งใช้ออปติคัล โฟลว์ ช่วยในการตัดสินใจเลือกเฟรมหลัก และจากการทดสอบด้วย 1,560 ตัวอักษรภาษามือไทย จากอาสาสมัคร 13 คน ระบบมีความแม่นยำ 85.5%
2. การหาลักษณะเด่นและการระบุท่าพื้นฐานของเฟรมหลัก ขั้นตอนนี้ได้แบ่งการทดสอบออกเป็น 2 กรณี คือ การวิเคราะห์หามุมที่สนใจในเฟรมหลัก ในหัวข้อ 4.3.1 และ การวิเคราะห์รูปร่างมือและนิ้วมือ ในหัวข้อ 4.3.2 ผลจากการทดสอบพบว่า การหาลักษณะเด่นและการระบุท่าพื้นฐานของเฟรมหลักโดยการวิเคราะห์รูปร่างมือและนิ้วมือ มีความแม่นยำสูงกว่า วิธีการหามุมที่สนใจในภาพถึง 15.3% ดังนั้นวิทยานิพนธ์นี้ในส่วนของการหาลักษณะเด่นของมือ และการระบุท่าพื้นฐานของเฟรมหลัก จึงใช้วิธีการวิเคราะห์รูปร่างมือและนิ้วมือ
3. การแปลลำดับอักขระเป็นตัวอักษรภาษาไทย ขั้นตอนนี้ได้ทดสอบด้วย 840 ตัวอักษรภาษามือไทย ซึ่งแบ่งการทดสอบออกเป็น 2 กรณี คือ การเปรียบเทียบโดยตรง ในหัวข้อ 4.4.1 และ การรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟ ในหัวข้อ 4.4.2 ผลจากการทดสอบพบว่า การรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟมีความยืดหยุ่นมากกว่าการเปรียบเทียบโดยตรง ซึ่งมีความแม่นยำสูงกว่าถึง 18.69 % ดังนั้นวิทยานิพนธ์นี้ในส่วนของการแปลลำดับอักขระเป็นตัวอักษรภาษาไทย จึงเลือกใช้วิธีรู้จำโดยใช้แบบจำลองฮิดเดนมาร์คอฟ เมื่อทดสอบกับชุดทดสอบจากผู้ที่เคยผ่านการสอนระบบ จำนวน 210 ท่าตัวอักษรภาษามือไทยระบบมีความแม่นยำเท่ากับ 81.43% และเมื่อทดสอบระบบ

ด้วยชุดทดสอบที่บันทึกจากผู้ที่ไม่เคยนำไปสอนระบบ จำนวน 210 ท่าตัวอักษรภาษามือไทย ระบบมีความแม่นยำเท่ากับ 71.9 %

5.2 ผลที่ได้จากการทำวิทยานิพนธ์ชุดนี้

ผู้ทำวิทยานิพนธ์ได้ศึกษาและพัฒนาระบบรู้จำภาษามือไทย ซึ่งสามารถรู้จำตัวอักษรภาษามือไทยทั้งหมด 42 ตัวอักษร (ยกเว้นฃ ฅ) โดยไม่ใช้อุปกรณ์ช่วยในการติดตามการเคลื่อนที่ เมื่อระบบถูกทดสอบด้วยชุดทดสอบที่ผ่านการสอนระบบมาแล้ว ระบบมีความถูกต้อง 80.95 % และเมื่อทดสอบด้วยชุดข้อมูลที่ไม่เคยผ่านการสอนระบบ ระบบมีความถูกต้อง 71.9 % ซึ่งอยู่ในเกณฑ์ที่ยอมรับได้

5.3 ปัญหาและอุปสรรคของการทำวิทยานิพนธ์

สิ่งแวดล้อมขณะทำการบันทึกไฟล์วิดีทัศน์มีผลต่อระบบรู้จำภาษามือไทยโดยใช้การประมวลผลภาพเป็นอย่างมาก ดังนั้นทุกครั้งที่ทำการเก็บข้อมูลจึงต้องมีการควบคุมสิ่งแวดล้อมทั้งในเรื่อง ของ แสง ระยะจับภาพ และความเร็วของผู้ใช้ระบบ และฉากหลังต้องเป็นสีเข้ม ซึ่งหลายครั้งที่ทำการเก็บข้อมูลเกิดความผิดพลาด เช่น ฉากหลังมีวัตถุอื่นปะปน ทำให้ไม่สามารถนำไฟล์นั้นไปทดสอบระบบได้ นอกจากนี้ เพื่อความยืดหยุ่นของระบบ การสอน และการทดสอบระบบ จึงต้องใช้ข้อมูลจากคนหลายคนเพื่อให้เกิดความหลากหลายของลักษณะมือ แต่การจะขอเก็บข้อมูลจากผู้พิการทางหูโดยตรงนั้นค่อนข้างยาก เนื่องจากการเก็บข้อมูลต่อหนึ่งคนนั้นต้องใช้เวลานานและทำได้เฉพาะที่ ที่สามารถควบคุมสภาวะแวดล้อมได้ ซึ่งปัญหาที่ตามมาคือ การสื่อสาร ดังนั้นเราจึงฝึกบุคคลธรรมดาใช้ท่าอักษรภาษามือไทย แล้วบันทึกไว้หลังจากนั้นจึงนำข้อมูลที่ได้ไปให้ผู้เชี่ยวชาญตรวจสอบความถูกต้อง ทำให้เสียเวลามาก

5.4 ข้อเสนอแนะ

ผลจากการทดสอบระบบพบว่ายังมีประเด็นให้นำไปพัฒนาต่อได้ดังนี้

1. ทดลองเพิ่มจำนวนข้อมูลที่นำมาสอนระบบให้มากขึ้น ซึ่งจะทำให้ระบบมีความแม่นยำมากขึ้น
2. ปรับปรุง และเพิ่มวิธีในการหาลักษณะเด่นของตัวอักษรที่ยังคงมีความผิดพลาดสูง คือ “N” “M” “S” และ “Vowel base”
3. พัฒนาระบบในส่วนของการติดต่อกับผู้ใช้งาน (User Interface) เพื่อให้สะดวกในการใช้งาน

4. พัฒนาปรับปรุงขั้นตอนการคัดเฟรมภาพ การเพิ่มข้อมูลให้แก่ระบบวิธีนี้น่าจะสามารถเพิ่มสมรรถนะของระบบให้ดีขึ้น

เอกสารอ้างอิง

[1] Visarnkuna Wuttichai, 2546, Sign Language Recognition System by Hidden Markov Models, King Mongkut's University of Technology Thonburi, Bangkok.

[2] R. Martin McGuire, Jose Hernandez-Rebollar , Thad Starner, Valerie Henderson, Helene Brashear, and Danielle S. Ross, "Towards a One-Way American Sign Language Translator", GVU Center Georgia Tech Atlanta, Engineering and Applied Science George Washington University Washington, Brain and Cognitive Sciences University of Rochester Rochester.

[3] Rini Akmeliawati, Melanie Po-Leen Ooi, Ye Chow Kuang, 2007, " Real-Time Malaysian Sign Language Translation using Colour Segmentation and Neural Network", Monash University, Malaysia.

[4] Veerasakulthong, W. "Thai Hand Sign Recognition" M. Science. Thesis, National Institute of Development Administration, Bangkok, Thailand, 2007

[5] Auephanwiriyakul, Sansanee, Nipon Theera-Umpon, Pitakwinai Suwannee, 2551, Thai Sign Language Translation System, Chiang Mai University, Chiang Mai.

[6] Kanjanapatmata Virot, 2006, Sign language gesture recognition for Thai consonant using artificial neural Networks, King Mongkut's University of Technology North Bangkok, Bangkok.

[7] จิราภา นิวาตพันธ์, 2549, การวิเคราะห์และเปรียบเทียบภาษามือไทยในเชิงภาษาศาสตร์, วิทยานิพนธ์ศิลปศาสตร์มหาบัณฑิต, กรุงเทพมหานคร

[8] สมาคมคนหูหนวกแห่งประเทศไทย. ปทานุกรมภาษามือไทย ฉบับปรับปรุงและขยายเพิ่มเติม. ไทยวัฒนาพานิช, 1990

[9] อักษรภาษามือไทย (ออนไลน์). (2552). เข้าถึงได้จาก http://www.tddf.or.th/tddf/signlang/finger.php สืบค้น 27 มกราคม 2552.

[10] Ka-Man Wong, Chun-Ho Cheung, Tak-Shing Liu, and Lai-Man Po. Dominant Color Image Retrieval using Merged Histogram. IEEE (ISCAS) (June 2003).

[11] ออปติคัลโฟลว์ (ออนไลน์). (2552). เข้าถึงได้จาก http://en.wikipedia.org/wiki/Optical_flow. สืบค้น 25 เมษายน 2552.

[12] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. International Journal of Computer Vision, 37(2):151–172, June 2000.

[13] C. Harris and M. Stephens: A combined corner and edge detector. Alvey Vision Conference (1988) 147–151.

[14] H. Moravec. Obstacle avoidance and navigation in the real world by a seeing robot rover. Technical Report CMU-RI-TR-3, Carnegie-Mellon University, Robotics Institute, 1980.

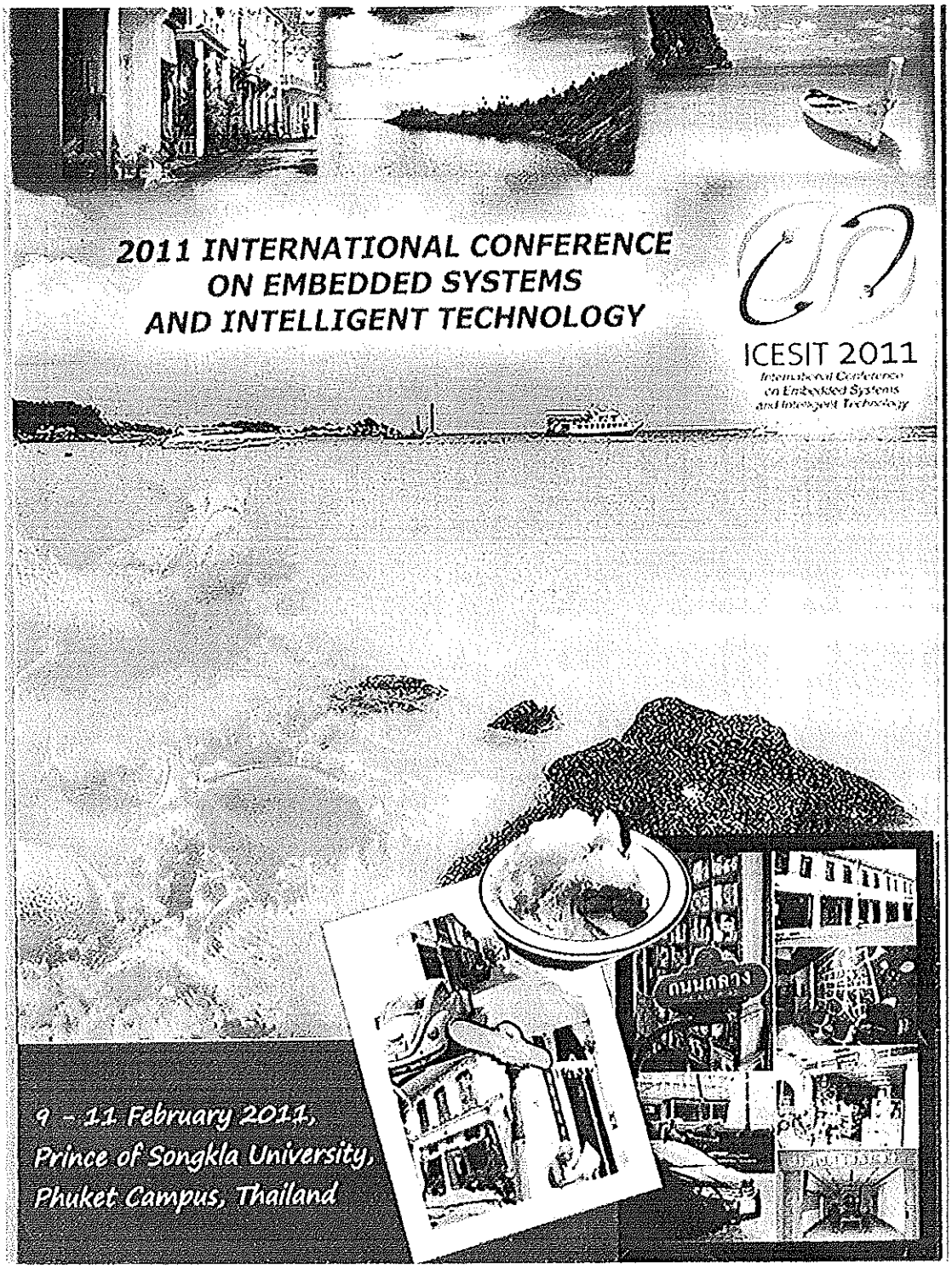
[15] Richard O. Duda, Peter E. Hart, David G.Stork. Pattern Classification Second Edition, 2000.

ภาคผนวก ก
ผลงานตีพิมพ์เผยแพร่จากวิทยานิพนธ์

International Conference on Embedded Systems and Intelligent Technology

(ICESIT 2011)

Feb 9- 11, 2011 at Prince of Songkla University (Phuket Campus), Phuket, Thailand


2011 INTERNATIONAL CONFERENCE
ON EMBEDDED SYSTEMS
AND INTELLIGENT TECHNOLOGY
ICESIT 2011
9 - 11 February 2011,
Prince of Songkla University,
Phuket Campus, Thailand

Thai Sign Language Recognition Based on Finger Pattern Analysis

Budsara Sakulsujirapa[1], Montri Karnjanadecha[2], Anant Choksuriwong[3]
Department of Computer Engineering, Faculty of Engineering
Prince of Songkla University, Hat Yai, Songkhla 90112, Thailand

E-mail: i_silli@hotmail.com[1]

Abstract- In this paper, we present an image processing method for identifying alphabets in Thai sign language. Features used for classification are based on finger pattern analysis. Key frames are selected from the video stream. Hand boundary is determined from each key frame. After that, finger pattern is analyzed using some rules. Experimental result shows that the accuracy of 83.75% is obtained.

ประวัติผู้เขียน

ชื่อ สกุล นางสาวบุษรา สกุลสุจิราภา

รหัสประจำตัวนักศึกษา 5110120058

วุฒิการศึกษา

วุฒิ	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
วิศวกรรมศาสตรบัณฑิต (วิศวกรรมคอมพิวเตอร์)	มหาวิทยาสงขลานครินทร์	2550

ทุนการศึกษา (ที่ได้รับในระหว่างการศึกษา)

- ทุนศิษย์ก้นกุฏิ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์ ประจำปีการศึกษา 2551

การตีพิมพ์เผยแพร่ผลงาน

- Budsara Sakulsujirapa, Montri Karnjanadecha and Anant Choksuriwong " Thai Sign Language Recognition Based on Finger Pattern Analysis," In International Conference on Embedded Systems and Intelligent Technology (ICESIT2011), Phuket, Thailand, 9th - 11th February 2011, pages 16.