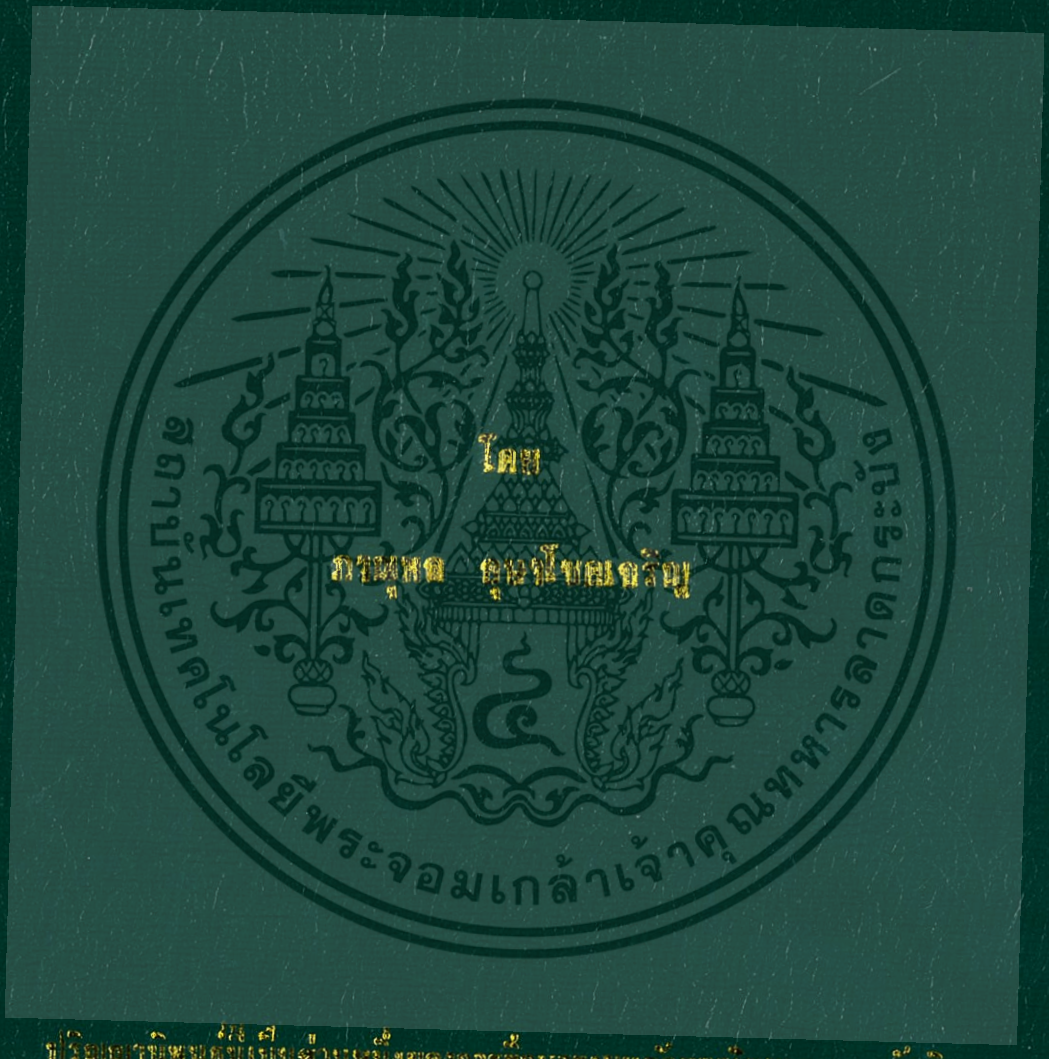
การจำแนกภาษามือไทยโดย KINECT

SIGN LANGUAGE RECOGNITION USING KINECT



โดย

ภานุพงศ์ ดุจฟ้อมดวิญญ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2557

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การจำแนกภาษามือโดยใช้ KINECT

SIGN LANGUAGE RECOGNITION USING KINECT

T144532

โดย

ภาณุพล อุษาโชคเจริญ

PANUPON USACHOKCHAROEN

อาจารย์ที่ปรึกษา

ผู้ช่วยศาสตราจารย์ ดร. กิติ์สุชาต พสุภา

b00268122

เลขหมู่.....
เลขทะเบียน 144532
วัน,เดือน,ปี 25 พ.ย. 2559

b. 12-8133569
i.....

รายงานนี้เป็นส่วนหนึ่งของการจัดการศึกษารายวิชาสหกิจศึกษา

หลักสูตรปริญญาวิทยาศาสตรบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2557

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การจำแนกภาษามือโดยใช้ KINECT

SIGN LANGUAGE RECOGNITION USING KINECT

โดย

ภาณุพล อุษาโชคเจริญ

PANUPON USACHOKCHAROEN

อาจารย์ที่ปรึกษา

ผู้ช่วยศาสตราจารย์ ดร. กิติ์สุชาต พสุภา

รายงานนี้เป็นส่วนหนึ่งของการจัดการศึกษารายวิชาสหกิจศึกษา

หลักสูตรปริญญาวิทยาศาสตรบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2557

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SIGN LANGUAGE RECOGNITION USING KINECT

PANUPON USACHOKCHAROEN

A REPORT SUBMITTED IN PARTIAL FULLFILLMENT

OF THE REQUIREMENT FOR COOPERATING EDUCATION PROGRAM

THE DEGREE OF BACHELOR OF SCIENCE PROGRAM IN

INFORMATION TECHNOLOGY

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2/2014

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2015

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใบรับรองรายงานโครงการสหกิจศึกษา ประจำปีการศึกษา 2557

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การจำแนกภาษามือโดยใช้ KINECT

SIGN LANGUAGE RECOGNITION USING KINECT

ผู้จัดทำ

นายภาณุพล อุษาโชคเจริญ รหัสนักศึกษา 54070070

..... อาจารย์ที่ปรึกษา

(ผู้ช่วยศาสตราจารย์ ดร. กิติ์สุชาต พสุภา)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อโครงงาน	การจำแนกภาษามือโดยใช้ KINECT
นักศึกษา	นายภาณุพล อุษาโชคเจริญ รหัสนักศึกษา 54070070
ปริญญา	วิทยาศาสตรบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
ปีการศึกษา	2557
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. กิติ์สุชาต พสุภา

บทคัดย่อ

ในช่วงหลายปีที่ผ่านมา เทคโนโลยีที่ช่วยเหลือผู้พิการถูกพัฒนาขึ้นมาอย่างต่อเนื่อง รวมไปถึงเทคโนโลยีที่ช่วยในการจำแนกภาษามือ ซึ่งสามารถช่วยผู้พิการในการสื่อสารกับบุคคลอื่น ๆ ในงานวิจัยนี้ได้ศึกษาเกี่ยวกับการจำแนกภาษามือโดยใช้อุปกรณ์คิเน็กท์ (Kinect) จากบริษัทไมโครซอฟท์ งานวิจัยดั่งเดิมได้ใช้ข้อมูลความลึกและข้อมูลการเคลื่อนไหว ส่วนงานวิจัยนี้ได้พัฒนาโดยการใช้ข้อมูลสี ข้อมูลต่างๆเหล่านี้ได้รับมาจากเซนเซอร์ของอุปกรณ์คิเน็กท์เพื่อสกัดหาลักษณะเด่นและนำไปเรียนรู้ในระบบการเรียนรู้ ท่าทางภาษามือที่ศึกษาในงานวิจัยนี้ได้แก่ชื่อ (Name), ไม่ (No), เท่าไหร่ (How many), ขอบคุณ (Thank you), อะไร (What), ที่ไหน (Where), ใช่ (Yes) และของคุณ (Your) ซึ่งผลที่ได้จากงานวิจัยพบว่า ประสิทธิภาพของการจำแนกภาษามือโดยใช้ข้อมูลการเคลื่อนไหวและข้อมูลสีร่วมกันมีความแม่นยำสูงสุดที่ร้อยละ 95

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Project Title	Sign Language Recognition Using Kinect
Student	Mr. Panupon Usachokcharoen Student ID 54070070
Degree	Bachelor of Science
Program	Information Technology
Academic Year	2014
Advisor	Asst. Prof. Dr. Kitsuchart Pasupa

ABSTRACT

In the last few years, many technologies for helping differently-abled people have been developed continuously including technology for recognizing sign languages which make differently-abled people to be able to communicate with others. In this research, we studied about recognizing sign languages by using the Microsoft Kinect. The conventional method utilized depth and motion information and our proposed method improved by using color information. These information acquired from the sensor to extract feature and learn in multi-class Support Vector Machine. Considered sign languages were Name, No, Thank you, How many, What, Where, Yes and Your. The experimental result showed that the performance of recognizing sign language by using the proposed method which uses motion and color information together could achieve 95 percent of accuracy.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้สำเร็จไปได้ด้วยดี ด้วยความกรุณาของอาจารย์ที่ปรึกษาโครงงาน ผู้ช่วยศาสตราจารย์ ดร.กิตติ์สุชาต พสุภา และ Asst. Prof. Dr. Washizawa Yoshikazu จากประเทศญี่ปุ่น ที่ได้สละเวลาให้ความช่วยเหลือ คำแนะนำในการปรับปรุงแก้ไขปัญหาต่าง ๆ ที่เกิดขึ้นตลอดจนให้ความรู้และประสบการณ์ที่ดีในการทำงาน และขอขอบพระคุณอย่างสูง

ขอกราบขอบพระคุณ The University of Electro-Communications (UEC) ประเทศญี่ปุ่น ผู้สนับสนุนทุนการศึกษาตลอดโครงการที่ประเทศญี่ปุ่นเป็นระยะเวลา 1 ปี

ขอกราบขอบพระคุณคณาจารย์คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังและ UEC ประเทศญี่ปุ่น ทุกๆท่านที่ได้ประสิทธิ์ประสาทวิชาให้กับข้าพเจ้า คอยให้คำปรึกษาและชี้แนะแนวทางในการแก้ไขปัญหาต่างๆให้สำเร็จลุล่วงไปได้

ขอขอบคุณพี่ เพื่อน และน้อง ทั้งที่คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังและ UEC ประเทศญี่ปุ่น ทุกคนที่ให้ความช่วยเหลือ คำแนะนำ และกำลังใจเสมอมา

ขอกราบขอบพระคุณ บิดา มารดา และครอบครัวของข้าพเจ้า ที่ให้กำลังใจ และให้การสนับสนุนในทุกๆเรื่อง ทำให้ข้าพเจ้าสามารถทำโครงงานฉบับนี้สำเร็จลุล่วงด้วยดี

คุณค่าและประโยชน์อันพึงมาจากโครงงานฉบับนี้ ข้าพเจ้าขอมอบแด่ผู้มีพระคุณทุกท่าน

ภาณุพล อุษาโชคเจริญ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VII

สารบัญรูป VIII

บทที่

1. บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 ความมุ่งหมายและวัตถุประสงค์ของการวิจัย 1

1.3 สมมติฐานของการวิจัย 2

1.4 ขอบเขตการวิจัย 2

1.5 ขั้นตอนการปฏิบัติงานวิจัย 2

1.6 ประโยชน์ที่คาดว่าจะได้รับ 2

1.7 สถานที่ปฏิบัติงาน 3

2. ทฤษฎีและหลักการ 4

2.1 ภาษามือ. 4

2.1.1 ภาษามือสากล. 4

2.1.2 ภาษามือผู้พิการ. 4

2.1.3 ภาษามือทางทหาร 5

2.2 ไมโครซอฟท์คิเน็กท์ (Microsoft Kinect) 5

2.2.1 ส่วนประกอบที่สำคัญของ Microsoft Kinect 6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

2.3 การประมวลผลภาพดิจิตอล (Digital Image Processing). 6

2.3.1 ภาพดิจิตอล (Digital Image). 6

2.3.2 การแยกส่วนที่สนใจออกจากสิ่งแวดล้อม 8

2.3.3 การแปลงภาพสีให้เป็นภาพระดับสีเทา. 10

2.3.4 Image Histogram 11

2.4 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) 13

2.5 งานวิจัยที่เกี่ยวข้อง 14

2.6 เครื่องมือที่ใช้ทำการดำเนินงาน 15

3. การดำเนินงานวิจัย 16

3.1 การสกัดข้อมูล 16

3.1.1 ข้อมูลความลึก (Depth Information) 16

3.1.2 ข้อมูลการเคลื่อนไหว (Motion Information) 17

3.1.3 ข้อมูลสี (Colour Information) 18

3.2 การจำแนกภาษามือด้วยการรวมหลายลักษณะเด่น 20

3.3 วิธีการประเมินคุณภาพ 21

3.3.1 การคำนวณหาค่าความถูกต้อง 21

3.3.2 การวัดประสิทธิภาพกับงานวิจัยที่เกี่ยวข้อง 21

3.4 เทคนิค 5-fold cross-validation 21

3.5 เทคนิค Normalization 22

3.6 กระบวนการทำงานของซัพพอร์ตเวกเตอร์แมชชีน 22

3.6.1 ขั้นตอนการสอนให้ระบบเรียนรู้ (Training) 22

3.6.2 ขั้นตอนการทดสอบระบบ (Testing) 23

4. ผลการวิจัย 24

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

4.1 การดำเนินการทดลอง 24

4.2 ผลการทดลอง 24

4.3 สรุปผลการทดลอง 25

5. สรุปและข้อเสนอแนะ 26

5.1 สรุปผลการทดลอง 26

5.2 ข้อเด่นหรือข้อด้อยของวิธีการที่พัฒนาขึ้นใหม่ 26

5.3 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต 26

5.4 ปัญหาของงานวิจัย 27

บรรณานุกรม 28

ประวัติผู้เขียน 29

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้า

ตารางที่

2.1 ตารางแสดงข้อมูลทั้งหมดของฮิสโตแกรมสีที่แบ่งออกเป็น 4 ช่วง 12
3.1 ตารางกำหนดค่า Class Index ของแต่ละภาษามือ 23
4.1 ค่า C ของแต่ละข้อมูล 24
4.2 ผลการทดลอง 25

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

หน้า

รูปที่

2.1 ท่าทางการตบมือ 4
2.2 ตัวอย่างภาษามือผู้พิการ 5
2.3 ตัวอย่างภาษามือทางการทหาร 5
2.4 ส่วนประกอบของคิเน็กท์ 6
2.5 ภาพไบนารี 7
2.6 ภาพระดับสีเทา 7
2.7 รูปภาพเฟรมที่ต่อเนื่องกัน 8
2.8 ภาพการเคลื่อนไหวของวัตถุในภาพ 9
2.9 ภาพและฮิสโตแกรมก่อนลบพื้นหลัง 9
2.10 ภาพและฮิสโตแกรมหลังลบพื้นหลัง 10
2.11 การแปลงภาพสีให้เป็นภาพระดับสีเทา 10
2.12 แสดงฮิสโตแกรมของภาพระดับสีเทา 11
2.13 แสดงฮิสโตแกรมก่อนการทำ Histogram Equalization 13
2.14 แสดงฮิสโตแกรมหลังการทำ Histogram Equalization 13
2.15 แสดงเวคเตอร์ในสเปซของ 2 กลุ่มข้อมูลขนาด 3 มิติ 14
3.1 กระบวนการหาขอบเขตที่สนใจของข้อมูลความลึก 17
3.2 ตัวอย่างข้อมูลใน 1 เซลล์ของข้อมูลความลึก 17
3.3 กระบวนการหาขอบเขตที่สนใจของข้อมูลการเคลื่อนไหว 18
3.4 ตัวอย่างข้อมูลใน 1 เซลล์ของข้อมูลการเคลื่อนไหว 18
3.5 ถุงมือสี 19
3.6 กระบวนการหาขอบเขตที่สนใจของข้อมูลสี 19
3.7 ตัวอย่างข้อมูลใน 1 เซลล์ของข้อมูลสี 20
3.8 การจัดวางเวกเตอร์ข้อมูลโดยการรวม 2 ลักษณะเด่น 20
3.9 การจัดวางเวกเตอร์ข้อมูลโดยการรวม 3 ลักษณะเด่น 21
3.10 Cross-validation 22

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันมีผู้ที่มีความบกพร่องทางด้านต่าง ๆ เป็นจำนวนมาก เช่น บุคคลที่บกพร่องในการมองเห็น บุคคลที่บกพร่องทางกายหรือสุขภาพ บุคคลที่บกพร่องทางสติปัญญา รวมไปถึงบุคคลที่บกพร่องในการพูดการได้ยิน สำหรับผู้ที่ไม่สามารถสื่อสารกับผู้อื่นได้โดยปกติ จำเป็นต้องใช้ภาษาที่เรียกว่า "ภาษามือ" ในการสื่อสาร

การเรียนภาษามือนั้นไม่เป็นที่แพร่หลายในปัจจุบัน ผู้ที่เรียนภาษามือส่วนใหญ่เป็นผู้พิการหรือผู้ที่ต้องการติดต่อกับผู้พิการโดยตรง ซึ่งถือว่ายังเป็นส่วนน้อย ทำให้คนธรรมดาที่ต้องการสื่อสารกับผู้พิการเป็นไปได้ยาก หรือแม้กระทั่งผู้ที่ไม่ใช่ผู้พิการโดยกำเนิดอันเนื่องมาจากสาเหตุบางประการทำให้ไม่สามารถสื่อสารโดยการใช้เสียงได้ จำเป็นต้องเรียนภาษามือใหม่ซึ่งเป็นเรื่องที่ยาก

ในปัจจุบันมีการพัฒนาเทคโนโลยีที่ช่วยในการสื่อสารกับผู้ใช้ภาษามืออย่างต่อเนื่อง แต่ยังไม่มีประสิทธิภาพเพียงพอ อาทิเช่นมีการนำอุปกรณ์คิเน็กท์ (Kinect) ซึ่งเป็นอุปกรณ์ที่พัฒนาโดยบริษัทไมโครซอฟท์ที่มีความสามารถในการรับรู้ระยะห่างระหว่างวัตถุกับเซนเซอร์ มาใช้ในการจำแนกภาษามือ Biswas และ Basu (2011) ใช้ข้อมูลความลึกและข้อมูลการเคลื่อนไหวที่สกัดจากระยะห่างระหว่างวัตถุจากอุปกรณ์คิเน็กท์ในการจำแนกภาษามือ [1] จากนั้น Agarwal และ Thakur (2013) ได้ทำการต่อยอดโดยการเพิ่มความเร็วในการประมวลผลและใช้เคอร์เนลชนิดอื่นเพื่อเพิ่มประสิทธิภาพ [2] แต่ยังจำแนกภาษามือได้ไม่แม่นยำ ในงานวิจัยนี้จึงเล็งเห็นว่าอุปกรณ์คิเน็กท์มีเซนเซอร์สีที่สามารถถ่ายภาพวีดีโอสีได้ จึงได้มีแนวคิดที่จะพัฒนาหลักการเดิมที่มีอยู่จากที่มีเพียงการใช้ข้อมูลความลึกและข้อมูลการเคลื่อนไหวจะมีการเพิ่มการใช้ข้อมูลสีด้วย โดยการพัฒนาถุงมือสีเพื่อเพิ่มประสิทธิภาพให้ดีมากยิ่งขึ้น

1.2 ความมุ่งหมายและวัตถุประสงค์ของการวิจัย

งานวิจัยนี้มุ่งหวังเพื่อศึกษาการจำแนกภาษามือโดยใช้ข้อมูลระยะห่างระหว่างวัตถุกับอุปกรณ์คิเน็กท์ ซึ่งสามารถสกัดลักษณะเด่นได้ 2 ชนิดคือ ข้อมูลความลึก (Depth Information) และข้อมูลการเคลื่อนไหวของภาพ (Motion Information) ต่อจากนั้นจะใช้ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) ในการเรียนรู้รูปแบบต่างๆของภาษามือที่กำหนดไว้ นอกจากนี้จะพัฒนาการจำแนกให้มีประสิทธิภาพมากยิ่งขึ้นโดยการใช้สีจากถุงมือสี (Color Information) เพื่อเพิ่มความแม่นยำในการจำแนกแต่ละท่าทางของภาษามือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 สมมติฐานของการวิจัย

การใช้ความลึกของภาพในการจำแนกภาษามือนั้น ยังไม่เพียงพอต่อการจำแนกภาษามือ ทั้งในเรื่องของระยะห่างระหว่างผู้ใช้กับอุปกรณ์คิเน็กท์ที่ไม่เท่ากัน ซึ่งจะทำให้การจำแนกมีความผิดพลาดได้ง่าย จึงได้มีการใช้ข้อมูลสีในการปรับปรุงการจำแนกให้มีประสิทธิภาพมากยิ่งขึ้น เนื่องจากอุปกรณ์คิเน็กท์ มีเซนเซอร์หลายตัวในการจับภาพ เช่น อินฟราเรดเซนเซอร์ (Infrared sensor), เซนเซอร์ความลึก (Depth sensor) และ เซนเซอร์สี (Color sensor) ทำให้สามารถใช้ความหลากหลายของภาพในการเพิ่มความแม่นยำในการจำแนกภาษามือได้

1.4 ขอบเขตของการวิจัย

1.4.1 ภาพที่นำมาทำการทดสอบเป็นภาพดิจิตอลขนาด 320x240 พิกเซล (Pixels)

1.4.2 จำนวนวีดีโอที่ใช้ทดสอบมีจำนวนทั้งหมด 10 วีดีโอต่อ 1 ภาษามือ

1.4.3 ภาษามือที่ใช้จะเป็นรูปแบบของภาษามือของประเทศสหรัฐอเมริกาซึ่งมีทั้งหมด 8 รูปแบบได้แก่ ชื่อ (Name), ไม่ (No), เท่าไหร่ (How many), ขอบคุณ (Thank you), อะไร (What), ที่ไหน (Where), ใช่ (Yes) และของคุณ (Your)

1.4.4 เปรียบเทียบประสิทธิภาพลักษณะเด่นของวีดีโอด้วยข้อมูลความลึก, ข้อมูลการเคลื่อนไหวและข้อมูลสี

1.4.5 ด้านหน้าของผู้แสดงท่าทางภาษามือไม่มีสิ่งกีดขวางใดๆ

1.4.6 สามารถจำแนกภาษามือได้เพียงคนเดียว

1.5 ขั้นตอนการปฏิบัติงานวิจัย

1.5.1 ศึกษางานวิจัยเดิมที่มีอยู่แล้วเกี่ยวกับการจำแนกภาษามือและการใช้อุปกรณ์คิเน็กท์

1.5.2 พัฒนาโปรแกรมที่ใช้ข้อมูลความลึกและการเคลื่อนไหวของภาพในการจำแนกภาษามือ

1.5.3 วิเคราะห์เปรียบเทียบข้อมูลและพัฒนาประสิทธิภาพให้ดีมากยิ่งขึ้นโดยใช้ข้อมูลสี

1.5.4 สรุปผลการวิจัยและนำเสนอ

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1.6.1 สามารถจำแนกภาษามือได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

1.6.2 เป็นแนวทางให้ผู้ที่สนใจได้ค้นคว้าและศึกษาเกี่ยวกับการจำแนกภาษามือโดยใช้อุปกรณ์คิเน็กท์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.7 สถานที่ปฏิบัติงาน

สถานที่ปฏิบัติงานอยู่ที่ The University of Electro-Communications โดยการควบคุมดูแลของ Asst. Prof. Dr. Washizawa Yoshikazu และมีระยะเวลาตั้งแต่ 1 เมษายน 2557 ถึง 1 เมษายน 2558

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและหลักการ

2.1 ภาษามือ

ภาษามือคือภาษาสำหรับผู้พิการที่ไม่สามารถสื่อสารได้โดยการใช้เสียง ซึ่งต้องใช้การเคลื่อนไหวของมือ แขน ร่างกายหรือการแสดงความรู้สึกบนใบหน้าในการสื่อสารความคิดของผู้สื่อ รูปแบบของการใช้ภาษามือแบ่งออกเป็น 3 แบบคือ

2.1.1 ภาษามือสากล

ภาษามือสากลเป็นภาษามือที่คนทั่วไปสามารถเข้าใจได้ เช่นการตบมือ ซึ่งบ่งบอกถึงความประทับใจหรือความชื่นชอบในสิ่งนั้น ๆ ดังรูปที่ 2.1

รูปที่ 2.1 ท่าทางการตบมือ [3]

2.1.2 ภาษามือผู้พิการ

ภาษามือผู้พิการเป็นภาษามือสำหรับผู้พิการที่ไม่สามารถสื่อสารได้โดยการใช้เสียง โดยในแต่ละประเทศจะใช้รูปแบบของภาษามือไม่เหมือนกัน ในงานวิจัยนี้ได้พิจารณาภาษามือของประเทศสหรัฐอเมริกาเพราะเป็นสากล ยกตัวอย่างเช่นสวัสดี (Hello), ลาก่อน (Goodbye), ขอร้อง (Please), ด้วยความยินดี (You're Welcome), ใช่ (Yes), ขอโทษ (Sorry), ไม่ (No), ขอบคุณ (Thanks) ดังรูปที่ 2.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 2.2 ตัวอย่างภาษามือผู้พิการ [4]

2.1.3 ภาษามือทางทหาร

ภาษามือทางการทหารเป็นภาษามือที่ใช้ในกลุ่มทหารที่ปฏิบัติหน้าที่ในสถานที่ที่ไม่สามารถสื่อสารด้วยวิธีอื่นได้ เช่นหยุด (Stop), มอง (Look) และกลับไป (Go back) ดังรูปที่ 2.3

รูปที่ 2.3 ตัวอย่างภาษามือทางทหาร [5]

2.2 ไมโครซอฟท์คิเน็กท์ (Microsoft Kinect)

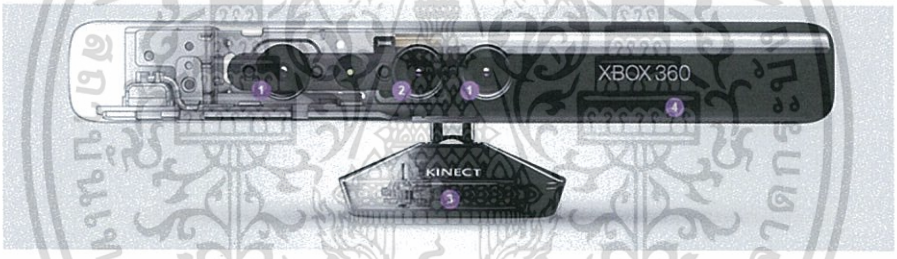
คืออุปกรณ์ตรวจจับการเคลื่อนไหวที่ถูกพัฒนาขึ้นโดยบริษัทไมโครซอฟท์ ซึ่งผู้ใช้สามารถมีปฏิสัมพันธ์กับเครื่องได้โดยใช้ท่าทางหรือการเคลื่อนไหวร่างกาย โดยอุปกรณ์คิเน็กท์สามารถตรวจจับการเคลื่อนไหวได้โดยใช้เซนเซอร์ความลึก เซนเซอร์สีและไมโครโฟน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.1 ส่วนประกอบที่สำคัญของ Microsoft Kinect

จากรูปที่ 2.4 สามารถแบ่งส่วนประกอบของคิเน็กท์ หลักๆได้ 4 ประเภทดังนี้

1. Infrared optics (Depth sensor) ซึ่งประกอบไปด้วย อินฟราเรดโปรเจคเตอร์และกล้อง 1 ตัว ซึ่งอินฟราเรดเซนเซอร์ทำหน้าที่ฉายแสงอินฟราเรดซึ่งมองด้วยตาเปล่าไม่เห็นออกมาเป็นแพทเทิร์นจุด และมีกล้องอีกตัวที่ทำหน้าที่ในการรับแสงอินฟราเรดที่ถูกฉายออกไป เพื่อใช้ในการสร้างภาพที่บอกระยะห่างวัตถุกับอุปกรณ์คิเน็กท์ (Depth map)
2. RGB camera เป็นกล้องถ่ายภาพเคลื่อนไหวใช้รับข้อมูลสี ความละเอียดของภาพสูงสุดคือ 1280x960 พิกเซล
3. Motorized tilt ใช้ในการปรับการมองเห็นของคิเน็กท์ ให้เงยขึ้นหรือก้มลงได้ 27 องศา
4. Multi-array microphone มีไมโครโฟนทั้งหมดจำนวน 4 ตัว ซึ่งมีหน้าที่ในการรับเสียงและมีส่วนช่วยในการระบุตำแหน่งของผู้ใช้ว่าอยู่ที่ไหน



รูปที่ 2.4 ส่วนประกอบของคิเน็กท์ [6]

2.3 การประมวลผลภาพดิจิตอล (Digital Image Processing)

การประมวลผลภาพดิจิตอลเป็นการใช้อัลกอริทึมในการแปลงภาพดิจิตอลให้อยู่ในรูปแบบที่เหมาะสมในการนำไปวิเคราะห์ต่อได้ง่าย ซึ่งการประมวลผลภาพดิจิตอลถูกนำไปประยุกต์ใช้อย่างกว้างขวาง เช่น ภาพเอ็กซเรย์ทางการแพทย์ การตรวจสอบการผลิตในโรงงานหรืองานด้านหุ่นยนต์เป็นต้น

2.3.1 ภาพดิจิตอล (Digital Image)

ภาพดิจิตอลคือรูปที่เก็บข้อมูลอยู่ในรูปแบบดิจิตอลซึ่งมีลักษณะเป็นสองมิติ ภาพดิจิตอลสามารถนิยามเป็นฟังก์ชั่นสองมิติคือ $f(x,y)$ โดยที่ x คือพิกัดแนวนอนของภาพและ y คือพิกัดแนวตั้งของภาพ ที่ตำแหน่ง (x,y) ใด ๆ ข้อมูลภายในภาพคือค่าความเข้มแสง (Intensity) ณ ตำแหน่งนั้น ๆ และสามารถเรียกตำแหน่ง (x,y) นั้น ๆ ว่าพิกเซล (Pixel) ถ้ากำหนดให้ภาพ มี

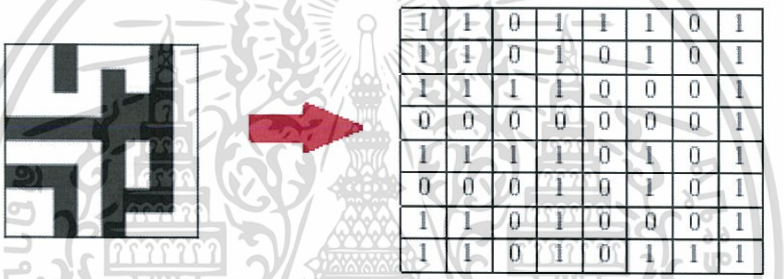
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขนาด i แถวและ j คอลัมน์ และถ้าพิกัดของจุดกำเนิดของภาพคือที่ตำแหน่ง $(x, y) = (0,0)$ แล้ว จะสามารถเขียนสมการให้อยู่ในรูปเมทริกซ์ได้ดังสมการที่ 2.1

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,j\text{-}1) \\ f(1,0) & f(1,1) & \dots & f(1,j\text{-}1) \\ \vdots & \vdots & \vdots & \vdots \\ f(i\text{-}1,0) & f(i\text{-}1,1) & \dots & f(i\text{-}1,j\text{-}1) \end{bmatrix} \quad (2.1)$$

2.3.1.1 ภาพไบนารี (Binary Image)

ภาพไบนารีเป็นภาพดิจิตอลที่มีค่าแค่ 2 ค่าคือ 0 แทนพิกเซลสีดำ และ 1 แทนพิกเซลสีขาว บางครั้งเรียกว่าภาพสองสี (Bi-tonal) ดังรูปที่ 2.5



รูปที่ 2.5 ภาพไบนารี

2.3.1.2 ภาพระดับสีเทา (Gray Scale Image)

ภาพระดับสีเทา (Grayscale Image) คือรูปภาพที่มีระดับความเข้ม 256 ระดับ ซึ่งมีค่าตั้งแต่ 0 ถึง 255 ซึ่งเรียงจากมืดไปสว่างหรือจากดำไปขาว ดังรูปที่ 2.6

รูปที่ 2.6 ภาพระดับสีเทา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.1.3 ภาพสี (Color Image or RGB Image)

มี 3 เลเยอร์ (Layer) ได้แก่สีแดง สีเขียวและสีน้ำเงิน จะใช้ในการแสดงผลบนจอภาพเช่น จอคอมพิวเตอร์หรือการสแกนภาพ ในภาพสีมี 24 บิตโดยแต่ละแชนแนลมี 8 บิต ในอีกความหมายคือ ภาพจะประกอบไปด้วย 3 ภาพ โดยแต่ละภาพสามารถเก็บค่าความเข้ม ความสว่างของภาพตั้งแต่ 0 ถึง 255

2.3.2 การแยกส่วนที่สนใจออกจากสิ่งแวดล้อม (Image Subtraction)

เป็นขั้นตอนในการสกัดส่วนที่สนใจโดยกระบวนต่างๆ เพื่อที่จะสามารถนำไปประมวลผลต่อไปได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

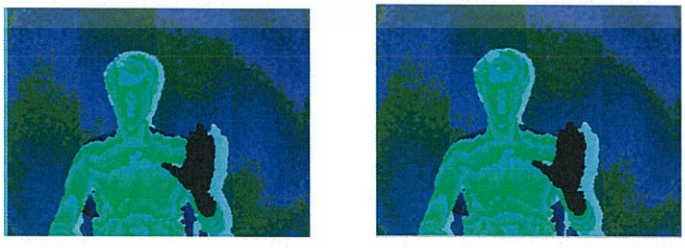
2.3.2.1 การหาความแตกต่างระหว่างเฟรม (Frame Differencing)

การหาความแตกต่างระหว่างเฟรมเป็นวิธีการหาการเคลื่อนไหวของวัตถุที่อยู่ในภาพ จะสามารถรับรู้ได้จากค่าพิกเซลที่แตกต่างกันระหว่าง 2 เฟรมที่ต่อเนื่องกันดังรูปที่ 2.7 ซึ่งจะได้ภาพที่เป็นสีเทา จากนั้นใช้เทรชโฮลด์ (Threshold) ในการทำให้ภาพเป็นไบนารีคือมีค่าเพียงสีขาวและสีดำเท่านั้นดังรูปที่ 2.8 ซึ่งบริเวณที่มีความแตกต่างกันนั้นจะเป็นบริเวณที่มีการเคลื่อนไหวของวัตถุ

สมการของเทรชโฮลด์ สามารถเขียนได้ดังนี้

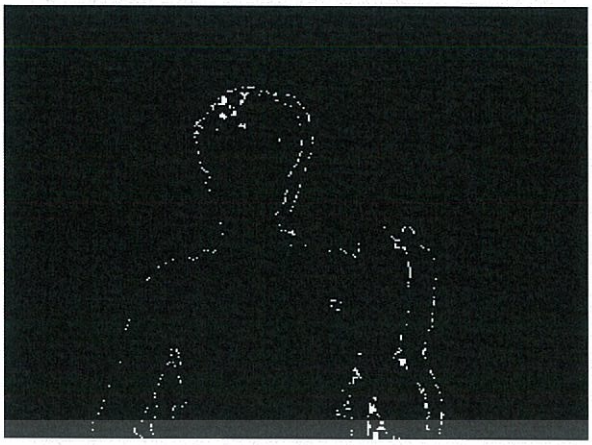
$$IBM = \begin{cases} 1 & when \;\; Pixel\,value > Threshold\,value \\ 0 & when \;\; Pixel\,value < Threshold\,value \end{cases} \tag{2.2}$$

โดยที่ *IBW* คือค่าใหม่ของพิเซลหลังใช้เทรชโฮลด์; *Pixel value* คือค่าของพิกเซล ณ ตำแหน่งนั้นๆ; *Threshold value* คือค่าเทรชโฮลด์ที่กำหนด;



รูปที่ 2.7 รูปภาพเฟรมที่ต่อเนื่องกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

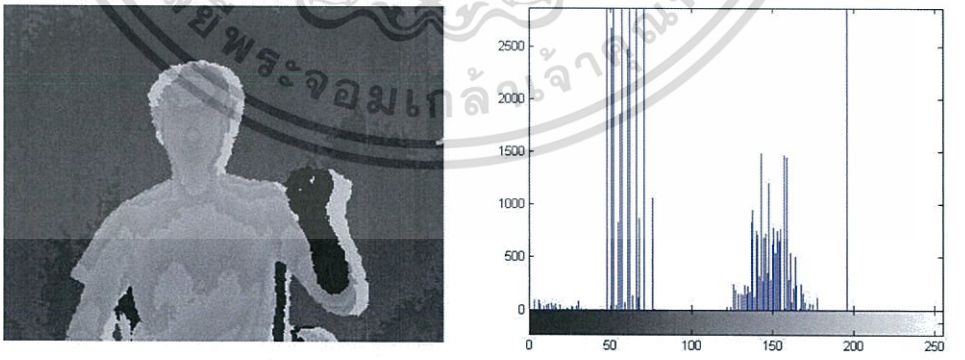


รูปที่ 2.8 ภาพการเคลื่อนไหวของวัตถุในภาพ

2.3.2.2 การลบฉากหลัง (Background Subtraction)

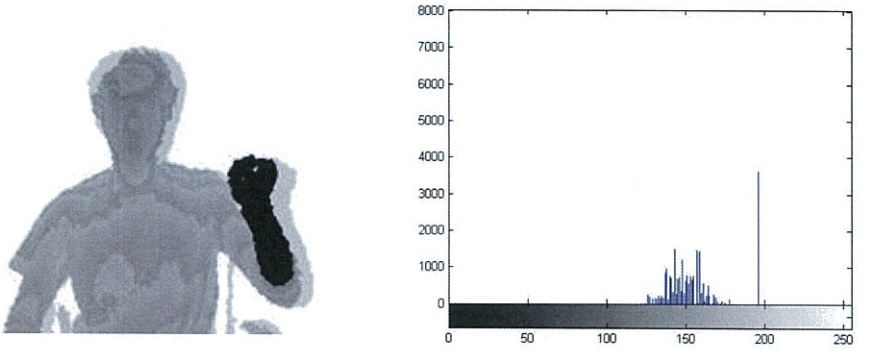
การลบฉากหลังเป็นการลบส่วนด้านหลังที่ไม่สนใจออกให้เหลือแต่วัตถุที่สนใจที่อยู่ด้านหน้าเท่านั้น ในปัจจุบันวิธีการลบฉากหลังมีมากมายหลายวิธี แต่ในงานวิจัยนี้เน้นการลบฉากหลังโดยใช้ภาพข้อมูลความลึก เนื่องจากอุปกรณ์คิเน็กท์สามารถหาระยะของวัตถุที่อยู่ในภาพได้ ซึ่งภาพข้อมูลความลึกนั้นควรเป็นภาพระดับสีเทา เพื่อง่ายต่อการลบฉากหลังเนื่องจากถ้าเป็นภาพความลึกที่เป็นสีจะมี 3 เลเยอร์ได้แก่สีแดง สีเขียว และสีน้ำเงิน ซึ่งมีขนาดของข้อมูลมากกว่า

กระบวนการการลบฉากหลังโดยใช้ภาพข้อมูลความลึกนั้น เริ่มต้นให้หาฮิสโตแกรมของข้อมูลความลึก จากนั้นให้พิจารณาค่าเฉพาะกลุ่มที่มีค่าในช่วงมากกว่า 100 และค่า 0 ดังรูปที่ 2.9 เพราะวัตถุอยู่ใกล้เซนเซอร์มากที่สุดคือมือจะมีสีดำสนิทและลำตัวจะมีสีสว่างกว่าพื้นหลัง



รูปที่ 2.9 ภาพและฮิสโตแกรมก่อนลบพื้นหลัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.10 ภาพและฮิสโตแกรมหลังลบพื้นหลัง

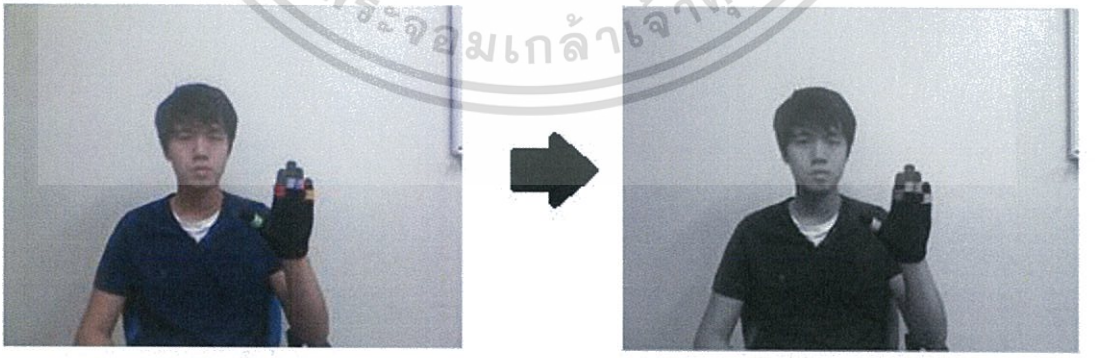
2.3.3 การแปลงภาพสี (Color Image) ให้เป็นภาพระดับสีเทา (Grayscale Image)

ทำได้โดยการหาค่าเฉลี่ยของแม่สีทั้งสามสีได้แก่สีน้ำเงิน สีเขียวและสีแดง ดังสมการต่อไปนี้

$$f_{\text{Gray}}(x,y) = \frac{f_{\text{Red}}(x,y) + f_{\text{Blue}}(x,y) + f_{\text{Green}}(x,y)}{3} \quad (2.3)$$

อย่างไรก็ตามตาของคนเราสามารถตอบสนองต่อแม่สีทั้งสามสีได้ต่างกัน จึงเปลี่ยนมาใช้สมการดังต่อไปนี้ [7]

$$f_{\text{Gray}}(x,y) = 0.229 \times f_{\text{Red}}(x,y) + 0.114 \times f_{\text{Blue}}(x,y) + 0.587 \times f_{\text{Green}}(x,y) \quad (2.4)$$



รูปที่ 2.11 การแปลงภาพสีให้เป็นภาพระดับสีเทา

จากรูปที่ 2.11 คือตัวอย่างการแปลงภาพสีไปเป็นภาพระดับสีเทา

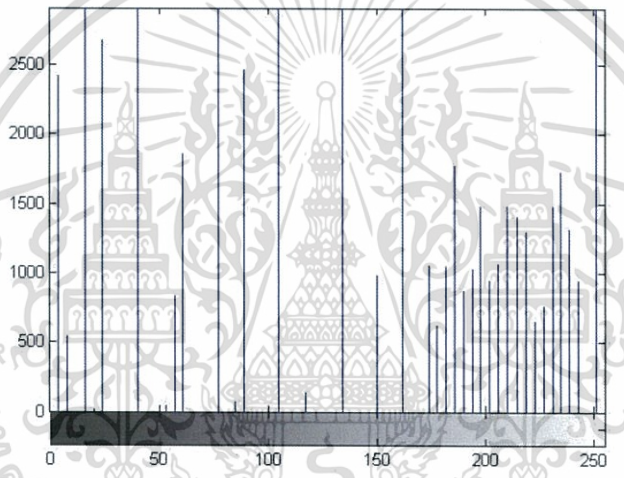
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.4 Image Histogram

เป็นกราฟที่แสดงจำนวนพิกเซลในแต่ละความสว่างต่าง ๆ หรือข้อมูลค่าสีแดง, สีเขียว และสีน้ำเงิน ของรูปภาพดิจิตอล

2.3.4.1 ฮิสโตแกรมของภาพระดับสีเทา (Grayscale histogram)

คือความถี่ของความเข้มในแต่ระดับสีในรูป โดยในแกน x จะแสดงระดับสีตั้งแต่ 0 ถึง 255 ส่วนแกน y จะแสดงความถี่ของความเข้มในแต่ระดับสี โดยทางซ้ายของกราฟจะมีค่าความสว่างน้อย ภาพจะมีสีเข้มเข้าใกล้สีดำ ส่วนทางขวาจะมีค่าความสว่างสูง ภาพจะสว่างเข้าใกล้สีขาว ส่วนบริเวณตรงกลางภาพจะเป็นระดับสีเทา จากรูปที่ 2.12



รูปที่ 2.12 แสดงฮิสโตแกรมของภาพระดับสีเทา

2.3.4.2 ฮิสโตแกรมสี (Color histogram)

ฮิสโตแกรมสีจะแสดงจำนวนของพิกเซลที่มีค่าสีอยู่ในช่วงของสีที่กำหนดได้แก่ ช่วงของสีแดง สีเขียวและสีน้ำเงิน ยกตัวอย่างเช่นถ้าต้องการแบ่งความถี่ของระดับความเข้มของสีออกเป็น 4 ช่วง (4 bins) จะได้ช่วงของสีแดง สีเขียว และสีน้ำเงิน ดังนี้ ช่วงที่ 1 (Bin0) คือค่าที่อยู่ในช่วง 0 ถึง 63 ช่วงที่ 2 (Bin1) คือ ค่าที่อยู่ในช่วง 64 ถึง 127 ช่วงที่ 3 (Bin2) คือ ค่าที่อยู่ในช่วง 128 ถึง 191 และช่วงที่ 4 (Bin3) คือ ค่าที่อยู่ในช่วง 192 ถึง 255 ของแต่ละสี ดังตารางที่ 2.1

ตารางที่ 2.1 จะแสดงข้อมูลทั้งหมดของฮิสโตแกรมสีที่แบ่งเป็น 4 ช่วงโดยแต่ละค่าก็คือจำนวนพิกเซล (Pixel count) ที่มีค่าสีอยู่ในช่วงต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.1 ตารางแสดงข้อมูลทั้งหมดของฮิสโตแกรมสีที่แบ่งออกเป็น 4 ช่วง

Red	Green	Blue	Pixel count
0-64 (Bin0)	0-64 (Bin0)	0-64 (Bin0)	10
0-64 (Bin0)	0-64 (Bin0)	65-127 (Bin1)	0
0-64 (Bin0)	0-64 (Bin0)	128-191 (Bin2)	1
0-64 (Bin0)	0-64 (Bin0)	192-255 (Bin3)	2
0-64 (Bin0)	65-127 (Bin1)	0-64(Bin0)	4
0-64 (Bin0)	65-127 (Bin1)	65-127 (Bin1)	0
0-64 (Bin0)	65-127 (Bin1)	128-191 (Bin2)	1
0-64 (Bin0)	65-127 (Bin1)	192-255 (Bin3)	6
.	.	.	.
.	.	.	.
.	.	.	.
192-255 (Bin3)	192-255 (Bin3)	192-255 (Bin3)	28

2.3.4.3 Histogram Equalization

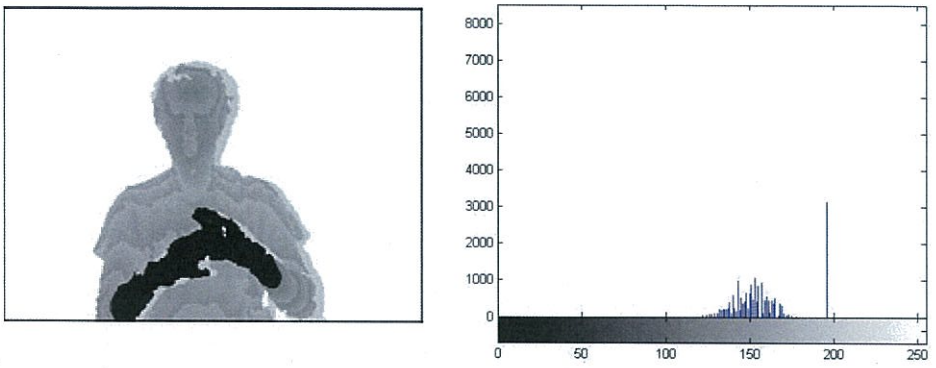
เป็นกระบวนการที่ใช้เพิ่มความคมชัดของภาพด้วยการกระจายค่าความถี่ของความเข้มในฮิสโตแกรมดังรูปที่ 2.13 และรูปที่ 2.14 โดยให้ค่าความเข้มเสมอกันเท่าที่จะเป็นไปได้ ซึ่งทำให้ภาพดูไม่สมจริง แต่จะเป็นประโยชน์ในทางวิทยาศาสตร์เช่นภาพ X-ray หรือภาพความร้อนจากดาวเทียม เพื่อช่วยในการวิเคราะห์ข้อมูลจากรูปภาพได้มากยิ่งขึ้น

โดยสมการของ Histogram Equalization สามารถเขียนได้ดังนี้

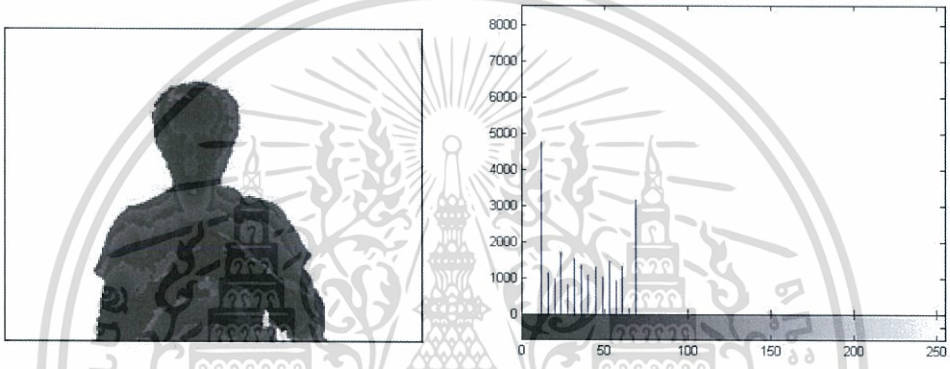
$$h(v) = \text{round}\left(\frac{cdf(v) - cdf_{\min}}{(i \times j) - cdf_{\min}} \times (L-1)\right) \tag{2.5}$$

จากสมการที่ (2.5) $h(v)$ คือค่าระดับสีเทาใหม่ ณ ตำแหน่ง v ของค่าระดับสีเทาที่พิจารณา ; $cdf(v)$ คือค่าความถี่สะสม (Cumulative Distribution Function) ณ ตำแหน่ง v ของค่าระดับสีเทาที่พิจารณา; cdf_{min} คือค่าความถี่สะสมต่ำสุด; L คือค่าระดับสีเทา; i คือจำนวนแถว; j คือจำนวนคอลัมน์; round คือการปัดเศษ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.13 แสดงฮิสโตแกรมก่อนการทำ Histogram Equalization

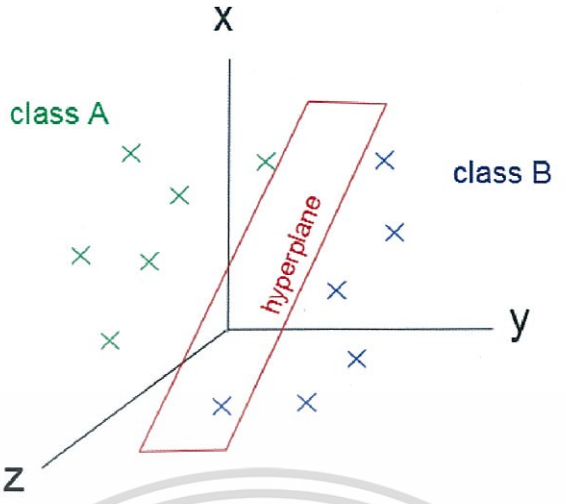


รูปที่ 2.14 แสดงฮิสโตแกรมหลังการทำ Histogram Equalization

2.4 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

เป็นเทคนิคที่ถูกพัฒนาเพื่อใช้ในการจำแนกข้อมูลออกเป็น 2 กลุ่ม (Class)โดยใช้ไฮเปอร์เพลน (Hyperplane) ในการแบ่งแยกดังรูปที่ 2.15 โดยหลักการของซัพพอร์ตเวกเตอร์แมชชีนคือการให้ข้อมูลที่นำเข้ามาในระบบการเรียนรู้เป็นเวคเตอร์ในสเปซขนาด M มิติ โดย M คือจำนวนลักษณะเด่นทั้งหมด จากนั้นสร้างไฮเปอร์เพลนที่สามารถแยกกลุ่มของข้อมูลที่นำเข้ามาเรียนรู้ ข้อเด่นของ SVM คือการทำการเก็บแมพ (Map) ของเวคเตอร์ให้เข้าสู่ฟีเจอร์สเปซ (Feature space) โดยใช้ฟังก์ชันหรือเคอร์เนล (Kernel) ชนิดต่างๆเช่นเคอร์เนลเชิงเส้นหรือเรเดียลเป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.15 แสดงเวคเตอร์ในสเปซของ 2 กลุ่มข้อมูลขนาด 3 มิติ

สมการของ SVM สามารถเขียนได้ดังนี้

$$wx_i + b \geq 1 \; when \; y_i = +1 \tag{2.6}$$

$$wx_i + b \leq 1 \; when \; y_i = -1 \tag{2.7}$$

จากสมการที่ (2.6) และ (2.7) เมื่อ w คือความชัน; x_i คือค่าลักษณะเด่น; b คือค่าคงที่; y_i คือประเภทข้อมูลลักษณะเด่น

2.5 งานวิจัยที่เกี่ยวข้อง

งานวิจัยของ Biswas และ Basu (2011) เป็นการใช้คิเน็กท์เพื่อสร้างข้อมูลระยะห่างระหว่างวัตถุกับอุปกรณ์คิเน็กท์และข้อมูลการเคลื่อนไหว เพื่อเรียนรู้ในซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) [1]

งานวิจัยของ Ridler และ Calvard (1978) เป็นการลบจุดรบกวน (Noise) ของภาพโดยการเลือกค่าเทรชโฮลด์ (Threshold) ที่เหมาะสม ถ้าค่าพิกเซลบนภาพมีค่ามากกว่าเทรชโฮลด์ก็จะเป็นฉากหน้า (Foreground) แต่ถ้าค่าพิกเซลบนภาพมีค่าน้อยกว่าเทรชโฮลด์ (Threshold) ก็จะเป็นฉากหลัง (Background) [8]

งานวิจัยของ Agarwal และ Thakur (2013) เป็นงานวิจัยที่อ้างอิงมาจาก ของ Biswas และ Basu [6] ซึ่งพัฒนาให้มีความเร็วในการประมวลผลและจำแนกภาษามือได้ดีมากยิ่งขึ้นโดยการใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดับเบิ้ลเทรด (Double Thread) และเรเดียลเบสิสฟังก์ชั่นเคอร์เนล (Radial Basis Function Kernel) [2]

งานวิจัยของ Sun และ Zhange และ Bao และ Xu (2013) เป็นงานวิจัยที่จำแนกภาษามือโดยใช้ Latent SVM Model โดยใช้สมมติฐานที่ว่าในวีดีโอของภาษามือ จะมีบางเฟรมในวีดีโอที่สามารถจำแนกภาษามือได้ [9]

จากงานวิจัยที่กล่าวมาจะเห็นได้ว่าการจำแนกภาษามือยังสามารถพัฒนาประสิทธิภาพหรือความแม่นยำให้มากยิ่งขึ้นได้ เนื่องจากข้อมูลความลึกและข้อมูลการเคลื่อนไหวต้องคำนึงถึงปัจจัยต่างๆเช่นระยะห่างระหว่างผู้แสดงท่าทางภาษามือกับอุปกรณ์คิเน็กท์หรือวัตถุที่มีการเคลื่อนไหวอยู่ด้านหลังผู้แสดงท่าทาง งานวิจัยนี้ได้เล็งเห็นว่าอุปกรณ์คิเน็กท์ยังมีเซนเซอร์สีที่สามารถบันทึกวีดีโอสี จึงคิดสมมติฐานที่จะใช้ข้อมูลสีมาช่วยในการวิเคราะห์และจำแนกภาษามือ

2.6 เครื่องมือในการดำเนินงาน

ซอฟท์แวร์ (Software)

1. Visual C# 2010 ใช้ในการบันทึกภาพที่เป็นความลึกและสี
2. แมตแล็บ (MATLAB) ใช้ในการวิเคราะห์ภาพ
3. LIBSVM ใช้เป็นส่วนเสริมของโปรแกรมแมตแล็บ เพื่อการเรียนรู้ระบบ (Machine Learning)

ฮาร์ดแวร์ (Hardware)

1. คิเน็กท์ อุปกรณ์ในการรับภาพความลึกและสี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การดำเนินงานวิจัย

เนื้อหาของงานวิจัยในบทนี้จะกล่าวถึงขั้นตอนการจำแนกภาษามือโดยใช้ SVM ซึ่งได้นำทฤษฎีและวิธีการในบทที่ 2 มาใช้ในงานวิจัย ซึ่งมีขั้นตอนการดำเนินงานวิจัยดังต่อไปนี้

3.1 การสกัดข้อมูล

ในขั้นตอนนี้จะทำการสกัดข้อมูลในรูปแบบเวกเตอร์ทั้งหมด 3 รูปแบบได้แก่ข้อมูลความลึก (Depth Information), ข้อมูลการเคลื่อนไหว (Motion Information) และข้อมูลสี (Color Information) เพื่อที่จะนำไปเรียนรู้ใน SVM ซึ่งการประมวลผลข้อมูลจะกระทำทีละเฟรมและรูปภาพของแต่ละข้อมูลจะมีขนาดความละเอียดของภาพที่ 320x240 พิกเซล

3.1.1 ข้อมูลความลึก (Depth Information)

ข้อมูลความลึกคือข้อมูลที่สามารถรับรู้ระยะห่างระหว่างเซนเซอร์กับวัตถุที่อยู่ในภาพได้ โดยภาพความลึกสามารถกำหนดสีของแต่ละระยะห่างเพื่อสร้างเป็นข้อมูลความลึก โดยการกำหนดสีสามารถกำหนดได้ทั้งเป็นภาพสีหรือเป็นภาพระดับสีเทา โดยภาพสีจะมีขอบเขตมากกว่าภาพระดับสีเทาเพราะมีจำนวนสีในการกำหนดระยะห่างมากกว่า จากนั้นหาขอบเขตที่สนใจของข้อมูลความลึก โดยเริ่มจากการแปลงภาพจากภาพสีเป็นภาพระดับสีเทาแล้วลบพื้นหลังโดยหาฮิสโตแกรมของข้อมูลความลึก จากนั้นพิจารณาค่าเฉพาะกลุ่มที่มีค่าในช่วงมากกว่า 100 และค่า 0 เพราะวัตถุอยู่ใกล้เซนเซอร์มากที่สุดคือมือจะมีสีดำสนิทและลำตัวจะมีสีสว่างกว่าพื้นหลัง ต่อจากนั้นใช้หลักการ Equalization Histogram ดังที่ได้อธิบายไว้ในบทที่ 2 เพื่อทำให้ภาพคมชัดขึ้นและครอบตารางขนาด 23x20 เซลล์บนตัวคนเพื่อจำกัดขอบเขตที่สนใจดังรูปที่ 3.1

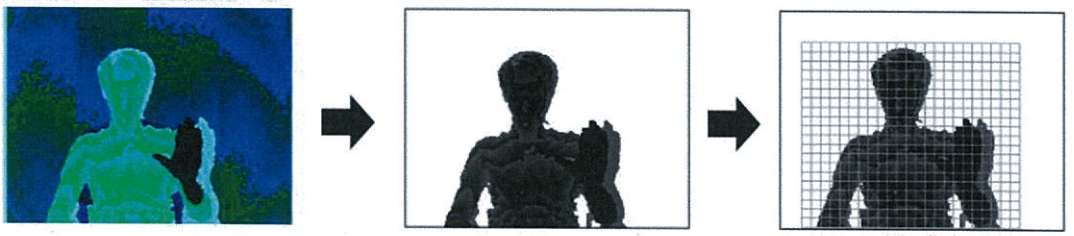
การสกัดข้อมูลในแต่ละเซลล์ของตารางทำได้โดยการใช้ Histogram 10 bins ซึ่งคือการแบ่งช่วงของระดับสีเทา (0 ถึง 255) เป็น 10 ช่วงโดยมีช่วงดังต่อไปนี้ 0-10, 11-20, 21-30, 31-40, 41-50, 51-70, 71-100, 101-120, 121-155 และ 156-255 [1] ซึ่งจากช่วงดังกล่าวจะสังเกตเห็นได้ว่ามีระยะห่างบางช่วงไม่เท่ากันเนื่องจากช่วง 71-255 เป็นช่วงที่เป็นพื้นหลังซึ่งไม่ใช่ช่วงที่สนใจ ดังนั้นช่วงระหว่าง 0-70 จึงมีความถี่ของช่วงมากกว่า หลังจากนั้นนับจำนวนพิกเซลที่อยู่ในช่วงดังกล่าวซึ่งผลลัพธ์ในแต่ละเซลล์จะมีจำนวนข้อมูล 10 ค่าคือจำนวนพิกเซลที่อยู่ใน 10 ช่วงของภาพระดับสีเทา

จากนั้นทำการสกัดข้อมูลทุก ๆ เซลล์ในตารางโดยใช้หลักการข้างต้น และนำข้อมูลในแต่ละเซลล์มาต่อกันเพื่อให้ได้ข้อมูลของ 1 เฟรม โดยข้อมูลใน 1 เฟรมของข้อมูลความลึกจะมีจำนวน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

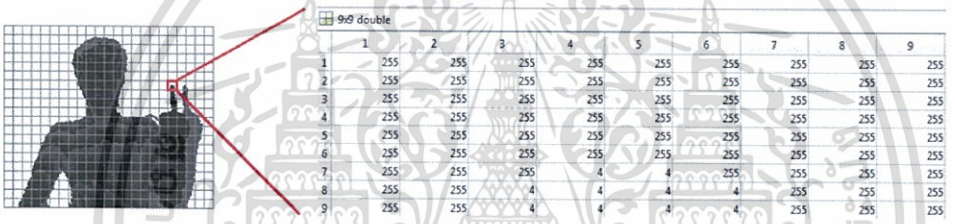
สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

23×20×10 = 4,600 ค่าและใช้หลักการนี้กระทำกับทุกเฟรมในวีดีโอจะได้ข้อมูลทั้งหมด 4,600×60 = 276,000 ค่าสำหรับ 1 วีดีโอของข้อมูลความลึก โดยที่ 60 คือจำนวนเฟรมใน 1 วีดีโอ



ภาพความตื้นลึกที่ได้จากอุปกรณ์คิเน็กท์ | ภาพความตื้นลึกที่เป็นภาพระดับสีเทาหลังลบพื้นหลัง | ภาพหลังจากครอบตารางเพื่อจำกัดขอบเขตที่สนใจของข้อมูลความลึก

รูปที่ 3.1 กระบวนการหาขอบเขตที่สนใจของข้อมูลความลึก



9×9 double

	1	2	3	4	5	6	7	8	9
1	255	255	255	255	255	255	255	255	255
2	255	255	255	255	255	255	255	255	255
3	255	255	255	255	255	255	255	255	255
4	255	255	255	255	255	255	255	255	255
5	255	255	255	255	255	255	255	255	255
6	255	255	255	255	255	255	255	255	255
7	255	255	255	4	4	255	255	255	255
8	255	255	4	4	4	4	255	255	255
9	255	255	4	4	4	4	255	255	255

ตารางขนาด 23 × 20 เซลล์ | ตารางแสดงข้อมูลใน 1 เซลล์ของข้อมูลความลึก

รูปที่ 3.2 ตัวอย่างข้อมูลใน 1 เซลล์ของข้อมูลความลึก

3.1.2 ข้อมูลการเคลื่อนไหว (Motion Information)

ข้อมูลการเคลื่อนไหวคือข้อมูลที่เกิดจากการเคลื่อนไหวท่าทางภาษามือ ซึ่งใช้หลักการการหาความแตกต่างระหว่างเฟรม โดยรับรู้ได้จากค่าพิกเซลที่แตกต่างกันระหว่าง 2 เฟรมที่ต่อเนื่องกัน ซึ่งจะได้ภาพที่เป็นสีเทา จากนั้นใช้เทรชโฮลด์ (Threshold) ในการทำให้ภาพเป็นไบนารีคือมีค่าเพียงสีขาวและสีดำเท่านั้นดังที่ได้อธิบายไว้ในบทที่ 2 การหาขอบเขตที่สนใจของภาพการเคลื่อนไหวทำได้โดยครอบตารางขนาด 23x20 เซลล์บนตัวคนเพื่อจำกัดขอบเขตที่สนใจดังรูปที่ 3.3

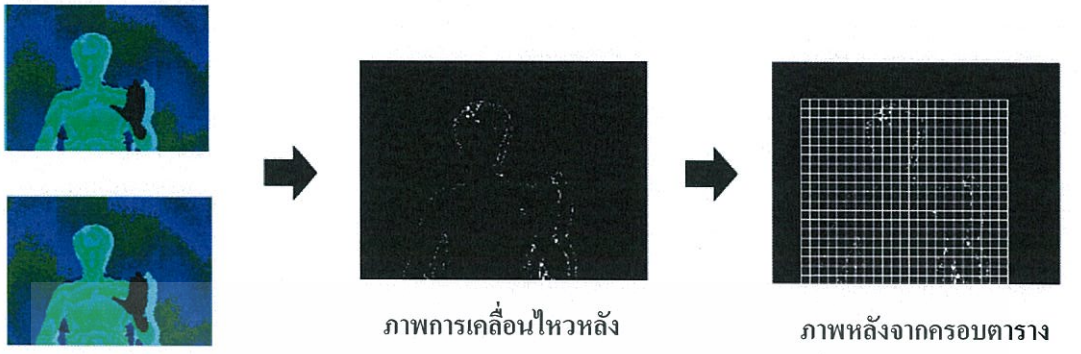
การสกัดข้อมูลจากขอบเขตที่สนใจของข้อมูลการเคลื่อนไหวทำโดยการนับพิกเซลที่มีค่าเป็น 255 (สีขาว) ในแต่ละเซลล์ของตารางดังรูปที่ 3.4 และในแต่ละเซลล์จะมีจำนวนข้อมูล 1 ค่าคือจำนวนพิกเซลสีขาวหรือจำนวนตำแหน่งที่มีการเคลื่อนไหว

จากนั้นทำการสกัดข้อมูลทุก ๆ เซลล์ในตารางโดยใช้หลักการข้างต้น และนำข้อมูลในแต่ละเซลล์มาต่อกันเพื่อให้ได้ข้อมูลของ 1 เฟรม โดยข้อมูลใน 1 เฟรมของข้อมูลการเคลื่อนไหวจะมี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

144532

จำนวน 23×20×1 = 460 ค่าและใช้หลักการนี้กระทำกับทุกเฟรมในวีดีโอจะได้ข้อมูลทั้งหมด 460×60 = 27,600 ค่าสำหรับ 1 วีดีโอของข้อมูลการเคลื่อนไหว โดยที่ 60 คือจำนวนเฟรมใน 1 วีดีโอ



ภาพความต้นลึก 2 ภาพที่ต่อเนื่องกัน

ภาพการเคลื่อนไหวหลังใช้เทรชโฮลด์

ภาพหลังจากครอบตารางเพื่อจำกัดขอบเขตที่สนใจของข้อมูลการเคลื่อนไหว

รูปที่ 3.3 กระบวนการหาขอบเขตที่สนใจของข้อมูลการเคลื่อนไหว

ตารางขนาด 23 × 20 เซลล์

9x9 double

	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
5	0	0	0	255	255	0	0	0	0
6	0	0	255	0	0	0	255	0	0
7	0	0	255	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0

ตารางแสดงข้อมูลใน 1 เซลล์ของข้อมูลการเคลื่อนไหว

รูปที่ 3.4 ตัวอย่างข้อมูลใน 1 เซลล์ของข้อมูลการเคลื่อนไหว

3.1.3 ข้อมูลสี (Color Information)

ข้อมูลสีคือข้อมูลที่มาจากสีบนถุงมือ ซึ่งมีด้วยกันข้างละ 12 สีดังรูปที่ 3.5 โดยการพิจารณาการเลือกสีนั้น พิจารณาโดยไม่ให้สีซ้ำกับสีของผิวหนัง เพื่อสะดวกในการสกัดข้อมูลสี

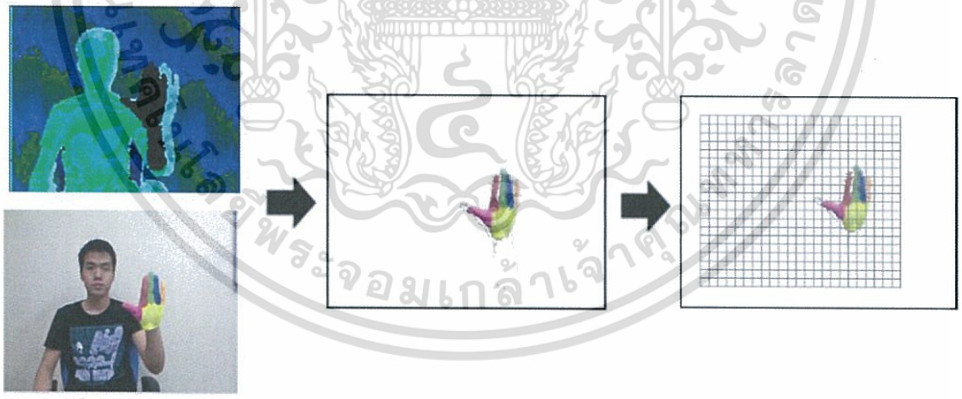
การหาขอบเขตที่สนใจของภาพสีทำได้โดยการใช้ภาพของข้อมูลความลึกและภาพสีร่วมกัน โดยสังเกตได้ว่าภาพของข้อมูลความลึกจะสามารถแยกแยะระหว่างลำตัวกับมือได้ เนื่องจากมืออยู่ใกล้กับเซนเซอร์วัดความลึกมากกว่าลำตัวทำให้มีสีดำ ส่วนลำตัวมีสีเขียวและพื้นหลังมีสีน้ำเงิน จากนั้นจึงพิจารณาเฉพะตำแหน่งพิกเซลของภาพสีที่มีตำแหน่งเดียวกันกับพิกเซลที่มีสีดำของภาพข้อมูลความลึก เมื่อได้ตำแหน่งที่พิจารณาของภาพสีก็จะขจัดส่วนที่ไม่เกี่ยวข้องกับสีที่ถุงมือเช่นสีของผิวหนังโดยกำหนดช่วงของสีที่ไม่ใช่สีเนื้อดังรูปที่ 3.6 จากนั้นครอบตาราง ขนาด 23x20 เซลล์ที่ตำแหน่งเดียวกับข้อมูลความลึกเพื่อกำหนดขอบเขตที่สนใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การสกัดข้อมูลจากขอบเขตที่สนใจของข้อมูลสีทำได้โดยการสกัดข้อมูลทีละเซลล์ของตาราง เหมือนกับข้อมูลความลึกแต่เนื่องจากข้อมูลสี แต่ละพิกเซลมี 3 เลเยอร์ได้แก่สีแดง สีเขียว และสีน้ำเงินดังรูปที่ 3.7 จึงได้ใช้หลักการ Color histogram 4bins ดังที่ได้กล่าวไว้ในบทที่ 2 ซึ่งในแต่ละเซลล์จะมีจำนวนข้อมูล 64 ค่า

จากนั้นทำการสกัดข้อมูลทุก ๆ เซลล์ในตารางโดยใช้หลักการข้างต้น และนำข้อมูลในแต่ละเซลล์มาต่อกันเพื่อให้ได้ข้อมูลของ 1 เฟรม โดยข้อมูลใน 1 เฟรมของข้อมูลสีจะมีจำนวน $23 \times 20 \times 64 = 29,440$ ค่าและใช้หลักการนี้กระทำกับทุกเฟรมในวีดีโอจะได้ข้อมูลทั้งหมด $29,440 \times 60 = 1,766,400$ ค่าสำหรับ 1 วีดีโอของข้อมูลสี โดยที่ 60 คือจำนวนเฟรมใน 1 วีดีโอ

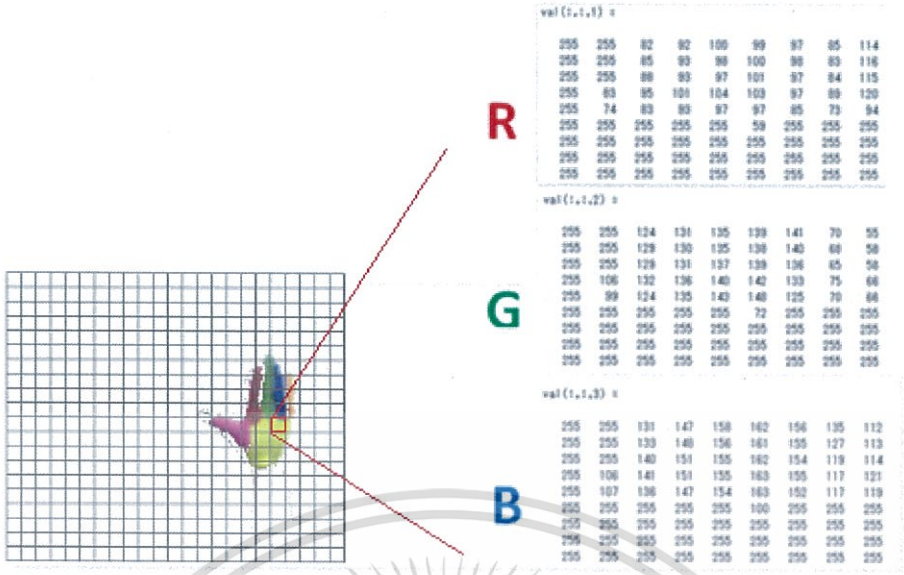
รูปที่ 3.5 ถุงมือสี



ภาพความลึกและภาพสี ณ ตำแหน่งเดียวกัน | **ภาพหลังจากลบสีของผิวหนังเพื่อสนใจเฉพาะสีบนถุงมือ** | **ภาพหลังจากครอบตารางเพื่อจำกัดขอบเขตที่สนใจของข้อมูลสี**

รูปที่ 3.6 กระบวนการหาขอบเขตที่สนใจของข้อมูลสี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



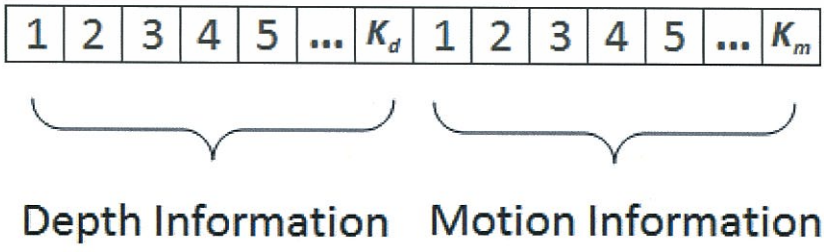
ตารางขนาด 23 × 20 เซลล์

ข้อมูลใน 1 เซลล์ของข้อมูลสี โดยแบ่งเป็นสีแดง สีเขียวและสีน้ำเงิน

รูปที่ 3.7 ตัวอย่างข้อมูลใน 1 เซลล์ของข้อมูลสี

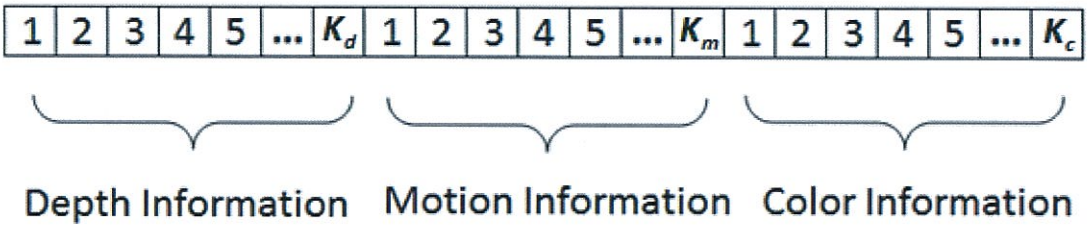
3.2 การจำแนกภาษามือด้วยการรวมหลายลักษณะเด่น

กระบวนการทำงานของการรวมหลายลักษณะเด่นมี 2 ประเภท ได้แก่การรวมเอาลักษณะเด่นทั้ง 3 แบบเข้าด้วยกัน ได้แก่ข้อมูลความลึก ข้อมูลการเคลื่อนไหวและข้อมูลสี และการรวมเอาลักษณะเด่นทั้ง 2 แบบเข้าด้วยกันเช่นข้อมูลความลึกและข้อมูลการเคลื่อนไหว โดยการใช้ข้อมูลในหัวข้อ 3.1 มาจัดวางในรูปแบบเวกเตอร์ $1 \times M$ มิติ โดย M คือจำนวนข้อมูลทั้งหมด; K_d คือจำนวนข้อมูลความลึก; K_m คือจำนวนข้อมูลการเคลื่อนไหว; K_c คือจำนวนข้อมูลสี ดังแสดงในรูปที่ 3.8 และ 3.9



รูปที่ 3.8 การจัดวางเวกเตอร์ข้อมูลโดยการรวม 2 ลักษณะเด่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.9 การจัดวางเวกเตอร์ข้อมูลโดยการรวม 3 ลักษณะเด่น

3.3 วิธีการประเมินคุณภาพ

การจำแนกภาษามือโดยใช้ซัพพอร์ตเวกเตอร์แมชชีนจำเป็นต้องประเมินประสิทธิภาพของงานวิจัย เพื่อให้เป็นมาตรฐานชี้วัดคุณภาพการเรียนรู้และเปรียบเทียบประสิทธิภาพกับงานวิจัยที่เกี่ยวข้อง

3.3.1 การคำนวณหาค่าความถูกต้อง

การคำนวณหาค่าความถูกต้องทำได้โดยการเปรียบเทียบข้อมูลลักษณะเด่นของภาษามือที่ต้องการทดสอบกับข้อมูลลักษณะเด่นของภาษามือทั้งหมดที่เรียนรู้แล้วดังสมการที่ 3.1

$$\text{ค่าความถูกต้อง } (accuracy) = \frac{\text{จำนวนข้อมูลภาษามือที่ตรวจจับได้ถูกต้อง}}{\text{จำนวนข้อมูลภาษามือที่ใช้ในการทดลอง}} \text{ x } 100 \qquad (3.1)$$

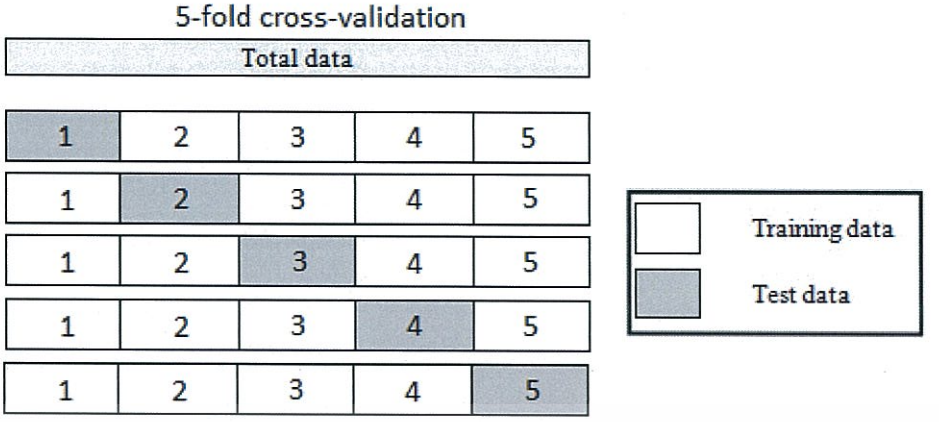
3.3.2 การวัดประสิทธิภาพกับงานวิจัยที่เกี่ยวข้อง

การวัดประสิทธิภาพกับงานวิจัยที่เกี่ยวข้องทำได้โดยการเปรียบเทียบกับงานวิจัยของ Biswas และ Basu [1] ซึ่งได้ทำการวิจัยโดยใช้ข้อมูลเฉพาะข้อมูลความลึกและข้อมูลการเคลื่อนไหว โดยงานวิจัยนี้ได้มีการเพิ่มประสิทธิภาพการจำแนกภาษามือโดยใช้ข้อมูลสีเพิ่มเข้ามา

3.4 เทคนิค 5-fold cross-validation

ในขั้นตอนการเตรียมข้อมูลจะใช้เทคนิค 5-fold cross-validation [10] ซึ่งเป็นวิธีการวัดประสิทธิภาพในการทำนายตัวอย่างของโมเดล เทคนิคคือการสุ่มตัวอย่าง (re-sampling) โดยแบ่งชุดข้อมูลออกเป็นส่วนๆ (fold) ในงานวิจัยนี้เลือกสุ่มข้อมูล 5 กลุ่ม ซึ่งแบ่งข้อมูลออกเป็น 5 ชุดเท่า ๆ กัน และทำการคำนวณค่าความแม่นยำจากการทำนาย 5 รอบ โดยแต่ละรอบจะมีการสร้างโมเดลจำแนกประเภทหนึ่งตัว จากข้อมูลเรียนรู้ 4 ชุด และใช้ข้อมูลทดสอบ 1 ชุด (ชุดที่ไม่ได้นำมาเรียนรู้) ดังรูปที่ 3.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.10 Cross-validation [10]

3.5 เทคนิค Normalization

ในขั้นตอนการเตรียมข้อมูลจะมีการทำ Normalization ข้อมูลคือการปรับค่าที่สกัดได้ให้อยู่ในมาตราส่วนเดียวกันโดยใช้สมการที่ 3.2

$$NL = \frac{d - \mu}{\sigma} \tag{3.2}$$

โดยที่ d คือข้อมูลทั้งหมด; μ คือค่าเฉลี่ยของข้อมูลทั้งหมด; σ คือค่าเบี่ยงเบนมาตรฐาน; NL คือค่าที่ผ่านการ Normalization แล้ว

3.6 กระบวนการทำงานของซัพพอร์ตเวกเตอร์แมชชีน

3.6.1 ขั้นตอนการสอนให้ระบบเรียนรู้ (Training)

กระบวนการทำงานของซัพพอร์ตเวกเตอร์แมชชีนในขั้นตอนการสอนให้ระบบเรียนรู้คือการใช้ลักษณะเด่นของข้อมูลมาเรียนรู้ในระบบการเรียนรู้ (Machine Learning) โดยในขั้นตอนนี้ได้นำค่าของเวกเตอร์ของข้อมูลที่ได้จากการดึงลักษณะเด่นจากหัวข้อที่ 3.1.1 3.1.2 และ 3.1.3 มาสอนให้กับระบบการเรียนรู้ซึ่งจะได้ข้อมูล x_i จากนั้นจึงกำหนดค่า y_i ซึ่งเป็นตัวกำหนดประเภทข้อมูลลักษณะเด่น โดยค่าของ y_i จะมีทั้งหมด 8 ค่าคือ 1 2 3 4 5 6 7 และ 8 ซึ่งเป็นการกำหนดท่าทางภาษามือของแต่ละท่าดังตารางที่ 3.1 การแบ่งกลุ่มแบบนี้เป็นการแบ่งกลุ่มข้อมูลแบบหลายกลุ่ม (Multiclass Support Vector Machine) ชนิด 1-VS-All นอกจากนี้ยังใช้เคอร์เนล (Kernel) เข้ามาใช้ในการแบ่งกลุ่มของข้อมูล ซึ่งเคอร์เนลที่ใช้มี 2 ชนิดคือแบบเชิงเส้น (Linear) และเรเดียนเบสิสฟังก์ชัน (Redial basis function)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ตารางกำหนดค่า Class index ของแต่ละภาษามือ

ภาษามือ	Class index
ชื่อ (Name)	1
ไม่ (No)	2
เท่าไหร่ (How many)	3
ขอบคุณ (Thank you)	4
อะไร (What)	5
ที่ไหน (Where)	6
ใช่ (Yes)	7
ของคุณ (Your)	8

3.6.2 ขั้นตอนการทดสอบระบบ (Testing)

กระบวนการทดสอบระบบจะทำโดยนำข้อมูลที่ดึงลักษณะเด่นเรียบร้อยที่ต้องการทดสอบมาทดสอบเพื่อเปรียบเทียบกับข้อมูลที่เรียนรู้แล้ว (Model) ซึ่งจะสามารถหาค่าความแม่นยำในการจำแนกภาษามือได้จากหัวข้อที่ 3.3.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ผลการวิจัย

4.1 การดำเนินการทดลอง

การดำเนินการทดลองจะแบ่งการทดลองเป็นการใช้ข้อมูลประเภทเดียว และการทดลองโดยใช้ข้อมูลร่วมกัน ซึ่งการทดลองโดยใช้ข้อมูลประเภทเดียวคือข้อมูลความลึก ข้อมูลการเคลื่อนไหวและข้อมูลสี ส่วนการทดลองโดยนำข้อมูลมารวมกันคือข้อมูลความลึกรวมกับข้อมูลการเคลื่อนไหวและข้อมูลการเคลื่อนไหวรวมกับข้อมูลสีและข้อมูลความลึกรวมกับข้อมูลการเคลื่อนไหวและข้อมูลสี โดยการเรียนรู้และทดสอบใน SVM มีการปรับค่า C ตั้งแต่ 10^{-16} ถึง 10^{16} โดยใช้ค่า C ที่ให้ค่าความแม่นยำสูงสุดในแต่ละข้อมูลดังตารางที่ 4.1

ตารางที่ 4.1 ค่า C ของแต่ละข้อมูล

ข้อมูล	Kernel	ค่า C
เฉพาะข้อมูลความลึก	Linear	0.0001
	RBF	10
เฉพาะข้อมูลการเคลื่อนไหว	Linear	0.01
	RBF	1
เฉพาะข้อมูลสี	Linear	0.01
	RBF	1000
ข้อมูลความลึกรวมกับข้อมูลการเคลื่อนไหว	Linear	0.001
	RBF	10
ข้อมูลการเคลื่อนไหวรวมกับข้อมูลสี	Linear	0.01
	RBF	1000
ข้อมูลความลึกรวมกับข้อมูลการเคลื่อนไหวและข้อมูลสี	Linear	0.001
	RBF	0.0001

4.2 ผลการทดลอง

ผลการทดลองจะเป็นค่าความแม่นยำของแต่ละภาษามือซึ่งได้แก่ชื่อ (Name) ไม่ (No) เท่าไหร่ (How many) ขอบคุณ (Thank you) อะไร (What) ที่ไหน (Where) ใช่ (Yes) และของคุณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Your) โดยทำการหาค่าความแม่นยำโดยใช้เคอร์เนล 2 ประเภทได้แก่เคอร์เนลเชิงเส้น (Linear kernel) และเคอร์เนลเรเดียนเบสิสฟังชัน (RBF kernel) ดังตารางที่ 4.2

ตารางที่ 4.2 ผลการทดลอง

ภาษามือ	ค่าความแม่นยำ (%)											
	เฉพาะข้อมูลความลึก		เฉพาะข้อมูลการเคลื่อนไหว		เฉพาะข้อมูลสี		ข้อมูลความลึกรวมกับข้อมูลการเคลื่อนไหว		ข้อมูลการเคลื่อนไหวรวมกับข้อมูลสี		ข้อมูลความลึกรวมกับข้อมูลการเคลื่อนไหวและข้อมูลสี	
	Linear	RBF	Linear	RBF	Linear	RBF	Linear	RBF	Linear	RBF	Linear	RBF
ชื่อ	91.6	95.6	96.5	94.6	96.1	96.0	**99.6**	99.5	98.5	98.6	**99.6**	**99.6**
ไม่	70.6	70.6	78.8	84.0	94.1	95.1	96.8	96.8	**98.6**	**98.6**	96.6	96.6
เท่าไหร่	97.1	97.8	95.8	95.6	90.3	89.3	**98.5**	98.1	92.1	92.1	98.1	98.1
ขอบคุณ	88.0	90.1	93.6	96.0	93.6	91.1	80.5	81.3	**96.6**	**96.6**	80.6	80.6
อะไร	91.5	87.0	90.3	92.6	98.0	98.0	93.8	93.6	**98.3**	**98.3**	95.6	95.6
ที่ไหน	83.1	77.0	69.1	77.5	89.1	89.1	80.0	80.6	**89.3**	**89.3**	80.1	80.1
ใช่	65.0	71.0	73.3	78.1	87.1	87.8	82.8	75.1	**88.6**	**88.6**	83.6	83.6
ของคุณ	79.0	77.3	78.8	78.5	**99.6**	**99.6**	96.0	96.0	**99.6**	**99.6**	96.0	96.0
ค่าเฉลี่ย	83.2	83.3	84.5	87.1	93.5	93.2	91.0	90.1	**95.2**	**95.2**	91.3	91.3

4.3 สรุปผลการทดลอง

จากผลการทดลองโดยเปรียบเทียบกับงานวิจัยที่เกี่ยวข้อง พบว่าค่าความแม่นยำเมื่อใช้ข้อมูลสีเพิ่มเข้ามาจากเดิมที่ใช้แค่ข้อมูลความลึกและข้อมูลการเคลื่อนไหวซึ่งเป็นสมมติฐานของงานวิจัยนี้ มีค่าความแม่นยำสูงกว่า โดยการใช้ข้อมูลการเคลื่อนไหวและข้อมูลสีร่วมกันมีค่าความแม่นยำสูงสุดที่ร้อยละ 95.2 อันเนื่องมาจากข้อมูลความลึกต้องคำนึงถึงระยะห่างระหว่างผู้แสดง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ท่าทางกับอุปกรณ์คิเน็กท์ ทำให้การทำนายมีโอกาสผิดพลาดมากกว่า และขอบเขตของงานนี้คือไม่มีวัตถุเคลื่อนไหวอยู่ด้านหลังผู้แสดงท่าทาง ทำให้การทำนายโดยใช้ข้อมูลการเคลื่อนไหวและข้อมูลสีมีโอกาสแม่นยำสูงกว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปและข้อเสนอแนะ

5.1 สรุปผลการทดลอง

งานวิจัยนี้ได้นำเสนอวิธีการจำแนกภาษามือโดยใช้อุปกรณ์คิเน็กท์เป็นตัวรับข้อมูลเพื่อนำมาสกัดข้อมูลหาลักษณะเด่นในแต่ละชนิด โดยข้อมูลที่นำมาสกัดได้แก่ข้อมูลความลึก ข้อมูลการเคลื่อนไหวและข้อมูลสี ซึ่งงานวิจัยที่เกี่ยวข้องของ Biswas และ Basu (2011) ได้ใช้ข้อมูลความลึกและข้อมูลการเคลื่อนไหวร่วมกันในการสกัดหาลักษณะเด่น [1] ส่วนงานวิจัยนี้ได้เพิ่มประสิทธิภาพโดยการใช้ข้อมูลสีและทดลองการใช้ข้อมูลแต่ละประเภทเดี่ยว ๆ ในการเรียนรู้และทดสอบใน SVM เนื่องมาจากงานวิจัยที่เกี่ยวข้องทดลองเฉพาะการใช้ข้อมูลความลึกร่วมกับข้อมูลการเคลื่อนไหวเท่านั้น

ผลการทดลองเปรียบเทียบค่าความแม่นยำระหว่างงานวิจัยที่เกี่ยวข้องพบว่าค่าความแม่นยำมีค่าสูงขึ้นเมื่อเพิ่มข้อมูลสี โดยการใช้ข้อมูลการเคลื่อนไหวและข้อมูลสีร่วมกันมีค่าความแม่นยำสูงสุดที่ร้อยละ 95.2 เนื่องจากข้อมูลความลึกต้องคำนึงถึงระยะห่างระหว่างผู้แสดงท่าทางกับอุปกรณ์คิเน็กท์ ทำให้การทำนายมีโอกาสผิดพลาดมากกว่า

ผลการทดลองโดยเฉลี่ยพบว่าภาษามือใช่ (Yes) มีค่าความแม่นยำต่ำสุดที่ร้อยละ 88.6 และภาษามือชื่อ (Name) และภาษามือของคุณ (Your) มีค่าความแม่นยำสูงสุดที่ร้อยละ 99.6

5.2 ข้อเด่นและข้อด้อยของวิธีการที่พัฒนาขึ้นใหม่

ข้อเด่นของงานวิจัยนี้คือการเพิ่มข้อมูลสีเพื่อช่วยเพิ่มประสิทธิภาพหรือความแม่นยำในการจำแนกภาษามือ โดยช่วยในเรื่องของการรับรู้ตำแหน่งของมือและนิ้วมือในแต่เฟรม แต่มีข้อด้อยคือขนาดของข้อมูลสีมีขนาดใหญ่มากและมีข้อมูลที่ไม่จำเป็นเป็นจำนวนมากเช่นเซลล์ที่ไม่มีข้อมูลของตารางของข้อมูลสี

5.3 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

งานวิจัยนี้ได้เพิ่มข้อมูลสีโดยสามารถพัฒนาประสิทธิภาพของการจำแนกภาษามือได้ดีขึ้น แต่ข้อมูลสีมีขนาดใหญ่กว่าข้อมูลความลึกและข้อมูลการเคลื่อนไหว ทำให้มีค่าที่ไม่จำเป็นในข้อมูลสีเป็นจำนวนมาก เพื่อที่จะสามารถแยกแยะภาษามือได้ดีขึ้น ควรที่จะลดขนาดของข้อมูลสีโดยพิจารณาแค่บริเวณสีส่วนมือที่ต้องการ ซึ่งจะสามารถทำให้มีประสิทธิภาพมากยิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.4 ปัญหาของงานวิจัย

ปัญหาที่พบในงานวิจัยนี้คือการแสดงท่าทางภาษามือต้องไม่มีสิ่งกีดขวางอยู่ด้านหน้าเพราะจะไม่สามารถแยกแยะระหว่างสิ่งกีดขวางกับตัวคนได้และพบปัญหาสำคัญของข้อมูลความลึกที่จำเป็นต้องคำนึงถึงระยะห่างระหว่างผู้แสดงท่าทางภาษามือกับอุปกรณ์คิเน็กท์ เนื่องจากงานวิจัยนี้ได้แสดงท่าทางภาษามือในบริเวณที่จำกัด ถ้าต้องการให้มีประสิทธิภาพดีขึ้นในระยะไกลขึ้นจำเป็นต้องเก็บข้อมูลมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

[1] Biswas, K. K., Basu, S. K. **"Gesture recognition using Microsoft KINECT®"** In: *Proceeding of 5th International Conference on Automation, Robotics and Applications (ICARA 2011)*, 6-8 Dec 2011, Wellington, New Zealand, pp 100-103.

[2] Anant, A., Manish, K. T. **"Sign Language Recognition using Microsoft Kinect"** In: *Proceeding of 6th International Conference on Contemporary Computing (IC3 2013)*, 8-10 Aug 2013, Noida, pp 181-185.

[3] Duveau **"Clap art"** [Online].Available: http://www.clker.com/clipart-49611.html. 2010

[4] Nidcd **"American Sign Language"** [Online].Available: http://www.nidcd.nih.gov/staticresources/images/NIDCD-ASL-Phrases.jpg. 2014

[5] Paul, T. **"Hand Signals"** [Online].Available: https://www.pinterest.com/pin/427771664579743658. 2010

[6] Daniel, H. **"Microsoft Kinect Sensor"** [Online].Available: http://gamelab.hkdm.de/wp-content/uploads/2014/04/fmx_2014_009_kinect.jpg. 2014

[7] Thiang, Andre, T. G., Resmana, L. **"Type of Vehicle Recognition Using Template Matching Method"** In: *Proceeding of the International conference on Electrical, Electronics, Communication and Information (CECI 2001)*, 7-8 March 2001, Jakarta.

[8] Ridler, T. W., Calvard **"Picture Thresholding Using an Iterative Selection Method"** In: *Proceeding of Systems, Man and Cybernetics (SMC 1978)*, Aug 1978, pp 630-632.

[9] Chao, S., Tianzhu, Z., Bing-Kun, B., Changesheng, X. **"Latent Support Vector Machine for sign language recognition with Kinect"** In: *Proceeding of 20th IEEE International Conference on Image Processing (ICIP 2013)*, 15-18 Sept 2013, Melbourne, pp 4190-4194.

[10] Chinnapat, K. **"k-fold cross-validation"** [Online]. Available: http://scriptslines.com/blog/k-fold- cross-validation. 2013

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ – นามสกุล	นายภาณุพล อุษาโชคเจริญ
วัน เดือน ปีเกิด	15 กรกฎาคม 2536
ที่อยู่	110 ซอยปทุมคงคา ถนนตรีมิตร เขต สัมพันธวงศ์
	อำเภอ สัมพันธวงศ์ จังหวัดกรุงเทพมหานคร 10100
โทรศัพท์	085-151-4187
อีเมล	panupon.usa@gmail.com

ประวัติการศึกษา

2557	วิทยาศาสตรบัณฑิต สาขาเทคโนโลยีสารสนเทศ
	คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การจำแนกภาษามือโดยใช้ Kinect

ภาณุพล อุษาโชคเจริญ[1] และ ผู้ช่วยศาสตราจารย์ ดร. กิติ์สุชาต พสุภา[1]

[1]*คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ*

Emails: panupon.usa@gmail.com, kitsuchart@it.kmit.ac.th

บทคัดย่อ

ในช่วงหลายปีที่ผ่านมา เทคโนโลยีที่ช่วยเหลือผู้พิการถูกพัฒนาขึ้นมาอย่างต่อเนื่อง รวมไปถึงเทคโนโลยีที่ช่วยในการจำแนกภาษามือ ซึ่งสามารถช่วยผู้พิการในการสื่อสารกับบุคคลอื่น ๆ ในงานวิจัยนี้ได้ศึกษาเกี่ยวกับการจำแนกภาษามือโดยใช้อุปกรณ์คิเน็กท์ (Kinect) จากบริษัทไมโครซอฟท์ งานวิจัยดั่งเดิมได้ใช้ข้อมูลความลึกและข้อมูลการเคลื่อนไหว ส่วนงานวิจัยนี้ได้พัฒนาโดยการใช้ข้อมูลสี ข้อมูลต่างๆเหล่านี้ได้รับมาจากเซนเซอร์ของอุปกรณ์คิเน็กท์เพื่อสกัดหาลักษณะเด่นและนำไปเรียนรู้ในระบบการเรียนรู้ ท่าทางภาษามือที่ศึกษาในงานวิจัยนี้ได้แก่ชื่อ (Name), ไม่ (No), เท่าไหร่ (How many), ขอบคุณ (Thank you), อะไร (What), ที่ไหน (Where), ใช่ (Yes) และของคุณ (Your) ซึ่งผลที่ได้จากงานวิจัยพบว่า ประสิทธิภาพของการจำแนกภาษามือโดยใช้ข้อมูลการเคลื่อนไหวและข้อมูลสีร่วมกันมีความแม่นยำสูงสุดที่ร้อยละ 95

คำสำคัญ — Sign Language Recognition; Machine Learning; Colour Gloves; Support Vector Machine

1. บทนำ

ในปัจจุบันมีผู้ที่มีความบกพร่องทางด้านต่าง ๆ เป็นจำนวนมาก เช่น บุคคลที่บกพร่องในการมองเห็น บุคคลที่บกพร่องทางกายหรือสุขภาพ บุคคลที่บกพร่องทางสติปัญญา รวมไปถึงบุคคลที่บกพร่องในการพูดการได้ยิน สำหรับผู้ที่ไม่สามารถสื่อสารกับผู้อื่นได้โดยปกติ จำเป็นต้องใช้ภาษาที่เรียกว่า “ภาษามือ” ในการสื่อสาร

การเรียนภาษามือนั้นไม่เป็นที่แพร่หลายในปัจจุบัน ผู้ที่เรียนภาษามือส่วนใหญ่เป็นผู้พิการหรือผู้ที่ต้องการติดต่อกับผู้พิการโดยตรง ซึ่งถือว่ายังเป็นส่วนน้อย ทำให้คนธรรมดาที่ต้องการสื่อสารกับผู้พิการเป็นไปได้ยาก หรือแม้กระทั่งผู้ที่ไม่ใช่ผู้พิการโดยกำเนิดอันเนื่องมาจากสาเหตุบางประการทำให้ไม่สามารถสื่อสารโดยการใช้เสียงได้จำเป็นต้องเรียนภาษามือใหม่ซึ่งเป็นเรื่องที่ยาก

ในปัจจุบันมีการพัฒนาเทคโนโลยีที่ช่วยในการสื่อสารกับผู้ใช้ภาษามืออย่างต่อเนื่อง แต่ยังไม่มีประสิทธิภาพเพียงพอ อาทิเช่นมีการนำอุปกรณ์คิเน็กท์ (Kinect) ซึ่งเป็นอุปกรณ์ที่พัฒนาโดยบริษัทไมโครซอฟท์ที่มีความสามารถในการรับรู้ระยะห่างระหว่างวัตถุกับเซนเซอร์ มาใช้ในการจำแนกภาษามือ Biswas และ Basu (2011) ใช้ข้อมูลความลึกและข้อมูลการเคลื่อนไหวที่สกัดจากระยะห่างระหว่างวัตถุจากอุปกรณ์คิเน็กท์ในการจำแนกภาษามือ [1] จากนั้น Agarwal และ Thakur (2013) ได้ทำการต่อยอดโดยการเพิ่มความเร็วในการประมวลผลและใช้เคอร์เนลชนิดอื่นเพื่อเพิ่มประสิทธิภาพ [2] แต่ยังจำแนกภาษามือได้ไม่แม่นยำ ในงานวิจัยนี้จึงเล็งเห็นว่าอุปกรณ์คิเน็กท์มีเซนเซอร์สีที่สามารถถ่ายภาพวีดีโอสีได้ จึงได้มีแนวคิดที่จะพัฒนาหลักการเดิมที่มีอยู่จากที่มีเพียงการใช้ข้อมูลความลึกและข้อมูลการเคลื่อนไหว จะมีการเพิ่มการใช้ข้อมูลสีด้วย โดยการพัฒนาถุงมือสีเพื่อเพิ่มประสิทธิภาพให้ดีมากยิ่งขึ้น

2. ทฤษฎีที่เกี่ยวข้อง

2.1. ภาษามือ

ภาษามือคือภาษาสำหรับผู้พิการที่ไม่สามารถสื่อสารได้โดยการใช้เสียง ซึ่งต้องใช้การเคลื่อนไหวของมือ แขน ร่างกาย หรือการแสดงความรู้สึกบนใบหน้าในการสื่อสารความคิดของผู้สื่อ รูปแบบของการใช้ภาษามือแบ่งออกเป็น 3 แบบ คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.1. ภาษามือสากล

ภาษามือสากลเป็นภาษามือที่คนทั่วไปสามารถเข้าใจได้ เช่น การตบมือ ซึ่งบ่งบอกถึงความประทับใจหรือความชื่นชอบในสิ่งนั้น ๆ

2.1.2. ภาษามือผู้พิการ

ภาษามือผู้พิการเป็นภาษามือสำหรับผู้พิการที่ไม่สามารถสื่อสารได้โดยการใช้เสียง โดยในแต่ละประเทศจะใช้รูปแบบของภาษามือไม่เหมือนกัน ในงานวิจัยนี้ได้พิจารณาภาษามือของประเทศสหรัฐอเมริกาเพราะเป็นสากล ยกตัวอย่างเช่น สวัสดี (Hello), ลาก่อน (Goodbye), ขอร้อง (Please), ด้วยความยินดี (You're Welcome), ใช่ (Yes), ขอโทษ (Sorry), ไม่ (No) และ ขอบคุณ (Thanks)

2.1.3. ภาษามือทางทหาร

ภาษามือทางการทหารเป็นภาษามือที่ใช้ในกลุ่มทหารที่ปฏิบัติหน้าที่ในสถานที่ที่ไม่สามารถสื่อสารด้วยวิธีอื่นได้ เช่น หยุด (Stop), มอง (Look) และกลับไป (Go back)

2.2. ไมโครซอฟท์คิเน็กท์ (Microsoft Kinect)

ไมโครซอฟท์คิเน็กท์ คือ อุปกรณ์ตรวจจับการเคลื่อนไหวที่ถูกพัฒนาขึ้นโดยบริษัทไมโครซอฟท์ ซึ่งผู้ใช้สามารถมีปฏิสัมพันธ์กับเครื่องได้โดยใช้ท่าทางหรือการเคลื่อนไหวร่างกาย โดยอุปกรณ์คิเน็กท์สามารถตรวจจับการเคลื่อนไหวได้โดยใช้เซนเซอร์ความลึก เซนเซอร์สีและไมโครโฟน ส่วนประกอบที่สำคัญของคิเน็กท์ดังรูปที่ 1 ได้แก่

2.2.1. Infrared optics (Depth sensor)

ซึ่งประกอบไปด้วย อินฟราเรดเซนเซอร์ทำหน้าที่ฉายแสงอินฟราเรด ซึ่งมองด้วยตาเปล่าไม่เห็นออกมาเป็นแพทเทิร์นจุด และมีกล้องอีกตัวที่ทำหน้าที่ในการรับแสงอินฟราเรดที่ถูกฉายออกไป เพื่อใช้ในการสร้างภาพที่บอกระยะห่างวัตถุกับอุปกรณ์คิเน็กท์ (Depth map)

2.2.2. RGB camera

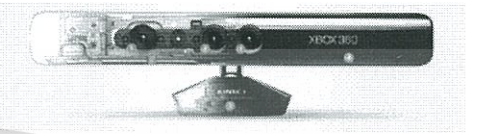
เป็นกล้องถ่ายภาพเคลื่อนไหวใช้รับข้อมูลสี ความละเอียดของภาพสูงสุดคือ 1280x960 พิกเซล

2.2.3. Motorized tilt

ใช้ในการปรับการมองเห็นของคิเน็กท์ ให้เงยขึ้นหรือก้มลงได้ 27 องศา

2.2.4. Multi-array microphone

มีไมโครโฟนทั้งหมดจำนวน 4 ตัว ซึ่งมีหน้าที่ในการรับเสียงและมีส่วนช่วยในการระบุตำแหน่งของผู้ใช้ว่าอยู่ที่ไหน



รูปที่ 1. ส่วนประกอบของคิเน็กท์ [3]

2.3. การประมวลผลภาพดิจิตอล (Digital Image Processing)

การประมวลผลภาพดิจิตอลเป็นการใช้อัลกอริทึมในการแปลงภาพดิจิตอลให้อยู่ในรูปแบบที่เหมาะสมในการนำไปวิเคราะห์ต่อได้ง่าย ซึ่งการประมวลผลภาพดิจิตอลถูกนำไปประยุกต์ใช้อย่างกว้างขวาง เช่น ภาพเอ็กซเรย์ทางการแพทย์ การตรวจสอบการผลิตในโรงงานหรืองานด้านหุ่นยนต์เป็นต้น

2.3.1. การแยกส่วนที่สนใจออกจากสิ่งแวดล้อม (Image Subtraction)

1) การหาความแตกต่างระหว่างเฟรม (Frame Differencing)

การหาความแตกต่างระหว่างเฟรมเป็นวิธีการหาการเคลื่อนไหวของวัตถุที่อยู่ในภาพ จะสามารถรับรู้ได้จากค่าพิกเซลที่แตกต่างกันระหว่าง 2 เฟรมที่ต่อเนื่องกันดังรูปที่ 2 ซึ่งจะได้ภาพที่เป็นสีเทา จากนั้นใช้เทรชโฮลด์ (Threshold) ในการทำให้ภาพเป็นไบนารีคือมีค่าเพียงสีขาวและสีดำเท่านั้นดังรูปที่ 3 ซึ่งบริเวณที่มีความแตกต่างกันนั้นจะเป็นบริเวณที่มีการเคลื่อนไหวของวัตถุ

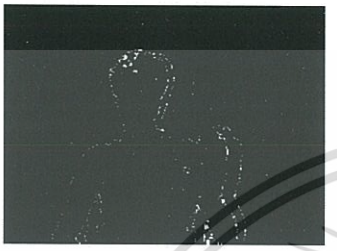
สมการของเทรชโฮลด์ สามารถเขียนได้ดังนี้

$$IBW = \begin{cases} 1 \ when \ Pixel\,value > Threshold\,value \\ 0 \ when \ Pixel\,value < Threshold\,value \end{cases} \quad (1)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่ *IBW* คือค่าใหม่ของพิเซลหลังใช้เทรชโฮลด์; *Pixel value* คือค่าของพิกเซล ณ ตำแหน่งนั้นๆ; Threshold value คือค่าเทรชโฮลด์ที่กำหนด;

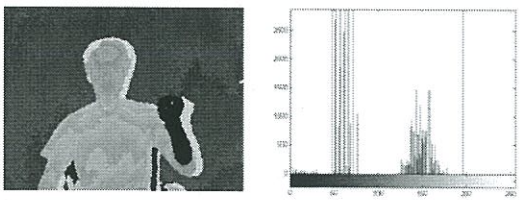
รูปที่ 2. รูปภาพเฟรมที่ต่อเนื่องกัน



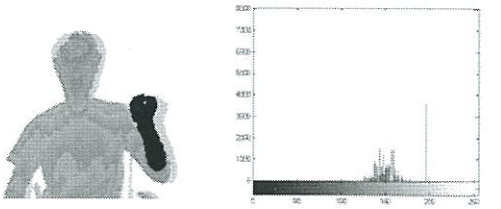
รูปที่ 3. ภาพการเคลื่อนไหวของวัตถุในภาพ

2) การลบพื้นหลัง (Background Subtraction)
การลบฉากหลังเป็นการลบส่วนด้านหลังที่ไม่สนใจออกให้เหลือแต่วัตถุที่สนใจที่อยู่ด้านหน้าเท่านั้น ในปัจจุบันวิธีการลบฉากหลังมีมากมายหลายวิธี แต่ในงานวิจัยนี้เน้นการลบฉากหลังโดยใช้ภาพข้อมูลความลึก เนื่องจากอุปกรณ์คิเน็กท์สามารถหาระยะของวัตถุที่อยู่ในภาพได้ ซึ่งภาพข้อมูลความลึกนั้นควรเป็นภาพระดับสีเทา เพื่อง่ายต่อการลบฉากหลัง เนื่องจากถ้าเป็นภาพความลึกที่เป็นสีจะมี 3 เลเยอร์ได้แก่สีแดง สีเขียว และสีน้ำเงิน ซึ่งมีขนาดของข้อมูลมากกว่า

กระบวนการการลบฉากหลังโดยใช้ภาพข้อมูลความลึกนั้น เริ่มต้นให้หาฮิสโตแกรมของข้อมูลความลึก จากนั้นให้พิจารณาค่าเฉพาะกลุ่มที่มีค่าในช่วงมากกว่า 100 และค่า 0 ดังรูปที่ 4 เพราะวัตถุอยู่ใกล้เซนเซอร์มากที่สุดคือมือจะมีสีดำสนิทและลำตัวจะมีสีสว่างกว่าพื้นหลัง



รูปที่ 4. ภาพและฮิสโตแกรมก่อนลบพื้นหลัง



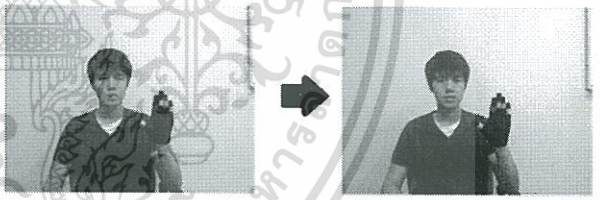
รูปที่ 5. ภาพและฮิสโตแกรมหลังลบพื้นหลัง

3) การแปลงภาพสี (Colour Image) ให้เป็นภาระดับสีเทา (Grayscale Image)
ทำได้โดยการหาค่าเฉลี่ยของแม่สีทั้งสามสีได้แก่สีน้ำเงิน สีเขียวและสีแดง ดังสมการ (2)

$$f_{Gray}(x,y)=\frac{f_{Red}(x,y)+f_{Blue}(x,y)+f_{Green}(x,y)}{3} \quad (2)$$

อย่างไรก็ตามตาของคนเราสามารถตอบสนองต่อแม่สีทั้งสามสีได้ต่างกัน จึงเปลี่ยนมาใช้สมการ (3)

$$f_{Gray}(x,y)=0.229f_{Red}(x,y)+0.114f_{Blue}(x,y) +0.587f_{Green}(x,y) \quad (3)$$



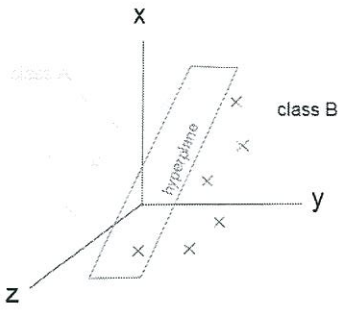
รูปที่ 6. การแปลงภาพสีให้เป็นภาพระดับสีเทา

2.4. ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

เป็นเทคนิคที่ถูกพัฒนาเพื่อใช้ในการจำแนกข้อมูลออกเป็น 2 กลุ่ม (Class) โดยใช้ไฮเปอร์เพลน (Hyperplane) ในการแบ่งแยกดังรูปที่ 7 โดยหลักการของซัพพอร์ตเวกเตอร์แมชชีนคือการให้ข้อมูลที่นำเข้ามาในระบบการเรียนรู้เป็นเวคเตอร์ในสเปซขนาด *M* มิติ โดย *M* คือจำนวนลักษณะเด่นทั้งหมด จากนั้นสร้างไฮเปอร์เพลนที่สามารถแยกกลุ่มของข้อมูลที่นำเข้ามาเรียนรู้ ข้อเด่นของ SVM คือการทำการเก็บแมพ (Map) ของเวคเตอร์ให้เข้าสู่ฟีเจอร์สเปซ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Feature space) โดยใช้ฟังก์ชันหรือเคอร์เนล (Kernel) ชนิดต่างๆเช่นเคอร์เนลเชิงเส้นหรือเรเดียลเป็นต้น



รูปที่ 7. แสดงเวคเตอร์ในสเปซของ 2 กลุ่มข้อมูลขนาด 3 มิติ

สมการของ SVM สามารถเขียนได้ดังนี้

$$wx_i + b \geq 1 \; when \; y_i = +1 \quad (4)$$

$$wx_i + b \leq 1 \; when \; y_i = -1 \quad (5)$$

จากสมการที่ (4) และ (5) เมื่อ w คือความชัน; x_i คือค่าลักษณะเด่น; b คือค่าคงที่; y_i คือประเภทข้อมูลลักษณะเด่น

3. การดำเนินงานวิจัย

3.1. การสกัดข้อมูล

ในขั้นตอนนี้จะทำการสกัดข้อมูลในรูปแบบเวกเตอร์ทั้งหมด 3 รูปแบบได้แก่ข้อมูลความลึก (Depth Information), ข้อมูลการเคลื่อนไหว (Motion Information) และข้อมูลสี (Color Information) เพื่อที่จะนำไปเรียนรู้ใน SVM ซึ่งการประมวลผลข้อมูลจะกระทำทีละเฟรมและรูปภาพของแต่ละข้อมูลจะมีขนาดความละเอียดของภาพที่ 320x240 พิกเซล

3.1.1. ข้อมูลความลึก (Depth Information)

ข้อมูลความลึกคือข้อมูลที่สามารถรับรู้ระยะห่างระหว่างเซนเซอร์กับวัตถุที่อยู่ในภาพได้ โดยภาพความลึกสามารถกำหนดสีของแต่ละระยะห่างเพื่อสร้างเป็นข้อมูลความลึก โดยการกำหนดสีสามารถกำหนดได้ทั้งเป็นภาพสีหรือเป็นภาพระดับสีเทา โดยภาพสีจะมีขอบเขตมากกว่าภาพระดับสีเทาเพราะมีจำนวนสีในการกำหนดระยะห่างมากกว่า จากนั้นหาขอบเขตที่สนใจของข้อมูลความลึก โดยเริ่มจากการแปลงภาพจากภาพสีเป็นภาพระดับสีเทาแล้วลบพื้นหลังโดยหาฮิสโตแกรมของข้อมูลความลึก จากนั้นพิจารณาค่าเฉพาะกลุ่มที่มีค่าในช่วงมากกว่า 100 และค่า 0 เพราะวัตถุอยู่ใกล้เซนเซอร์มากที่สุดคือมือจะมีสีดำสนิทและลำตัวจะมีสีสว่างกว่าพื้นหลัง ต่อจากนั้นใช้หลักการ Equalization Histogram เพื่อทำให้ภาพคมชัดขึ้นและครอบตารางขนาด 23x20 เซลล์บนตัวคนเพื่อจำกัดขอบเขตที่สนใจดังรูปที่ 8

การสกัดข้อมูลในแต่ละเซลล์ของตารางทำได้โดยการใช้ Histogram 10 bins ซึ่งคือการแบ่งช่วงของระดับสีเทา (0 ถึง 255) เป็น 10 ช่วงโดยมีช่วงดังต่อไปนี้ 0-10, 11-20, 21-30, 31-40, 41-50, 51-70, 71-100, 101-120, 121-155 และ 156-255 [1] ซึ่งจากช่วงดังกล่าวจะสังเกตเห็นได้ว่ามีระยะห่างบางช่วงไม่เท่ากันเนื่องจากช่วง 71-255 เป็นช่วงที่เป็นพื้นหลังซึ่งไม่ใช่ช่วงที่สนใจ ดังนั้นช่วงระหว่าง 0-70 จึงมีความถี่ของช่วงมากกว่า หลังจากนั้นนับจำนวนพิกเซลที่อยู่ในช่วงดังกล่าว ซึ่งผลลัพธ์ในแต่ละเซลล์จะมีจำนวนข้อมูล 10 ค่าคือจำนวนพิกเซลที่อยู่ใน 10 ช่วงของภาพระดับสีเทา

จากนั้นทำการสกัดข้อมูลทุก ๆ เซลล์ในตารางโดยใช้หลักการข้างต้น และนำข้อมูลในแต่ละเซลล์มาต่อกันเพื่อให้ได้ข้อมูลของ 1 เฟรม โดยข้อมูลใน 1 เฟรมของข้อมูลความลึกจะมีจำนวน 23×20×10 = 4,600 ค่าและใช้หลักการนี้กระทำกับทุกเฟรมในวีดีโอจะได้ข้อมูลทั้งหมด 4,600×60 = 276,000 ค่าสำหรับ 1 วีดีโอของข้อมูลความลึก โดยที่ 60 คือจำนวนเฟรมใน 1 วีดีโอ

ภาพความลึกที่ได้จากอุปกรณ์ | **ภาพระดับสีเทาหลังลบพื้นหลัง** | **ครอบตารางเพื่อจำกัดขอบเขตที่สนใจ**

รูปที่ 8. กระบวนการหาขอบเขตที่สนใจของข้อมูลความลึก

3.1.2. ข้อมูลการเคลื่อนไหว (Motion Information)

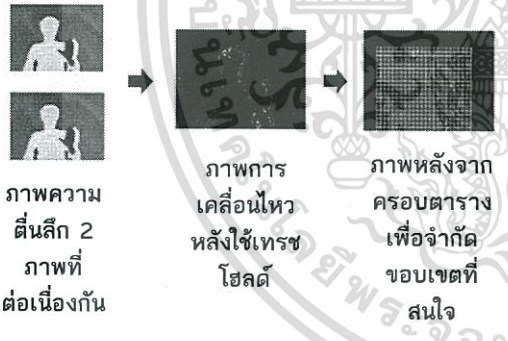
ข้อมูลการเคลื่อนไหวคือข้อมูลที่เกิดจากการเคลื่อนไหวท่าทางภาษามือ ซึ่งใช้หลักการการหาความแตกต่างระหว่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เฟรม โดยรับรู้ได้จากค่าพิกเซลที่แตกต่างกันระหว่าง 2 เฟรมที่ต่อเนื่องกัน ซึ่งจะได้ภาพที่เป็นสีเทา จากนั้นใช้เทรชโฮลด์ (Threshold) ในการทำให้ภาพเป็นไบนารีคือมีค่าเพียงสีขาวและสีดำเท่านั้น การหาขอบเขตที่สนใจของภาพการเคลื่อนไหวทำได้โดยครอบตารางขนาด 23x20 เซลล์บนตัวคนเพื่อจำกัดขอบเขตที่สนใจดังรูปที่ 9

การสกัดข้อมูลจากขอบเขตที่สนใจของข้อมูลการเคลื่อนไหวทำโดยการนับพิกเซลที่มีค่าเป็น 255 (สีขาว) ในแต่ละเซลล์ของตาราง และในแต่ละเซลล์จะมีจำนวนข้อมูล 1 ค่าคือจำนวนพิกเซลสีขาวหรือจำนวนตำแหน่งที่มีการเคลื่อนไหว

จากนั้นทำการสกัดข้อมูลทุก ๆ เซลล์ในตารางโดยใช้หลักการข้างต้น และนำข้อมูลในแต่ละเซลล์มาต่อกันเพื่อให้ได้ข้อมูลของ 1 เฟรม โดยข้อมูลใน 1 เฟรมของข้อมูลการเคลื่อนไหวจะมีจำนวน 23×20×1 = 460 ค่าและใช้หลักการนี้กระทำกับทุกเฟรมในวีดีโอจะได้ข้อมูลทั้งหมด 460×60 = 27,600 ค่าสำหรับ 1 วีดีโอของข้อมูลการเคลื่อนไหว โดยที่ 60 คือจำนวนเฟรมใน 1 วีดีโอ



รูปที่ 9. กระบวนการหาขอบเขตที่สนใจของข้อมูลการเคลื่อนไหว

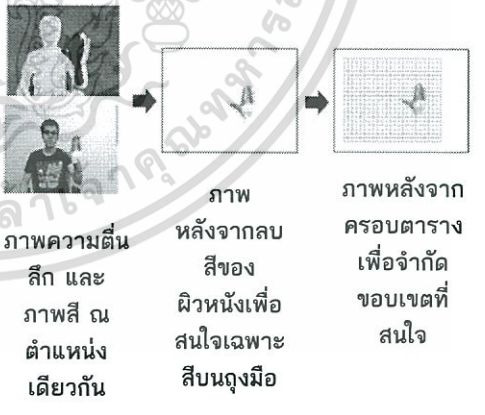
3.1.3. ข้อมูลสี (Colour Information)

รูปที่ 10. ถุงมือสี

ข้อมูลสีคือข้อมูลที่มาจากสีบนถุงมือ ซึ่งมีด้วยกันข้างละ 12 สีดังรูปที่ 10 โดยการพิจารณาการเลือกสีนั้น พิจารณาโดยไม่ให้สีซ้ำกับสีของผิวหนัง เพื่อสะดวกในการสกัดข้อมูลสี

การหาขอบเขตที่สนใจของภาพสีทำได้โดยการใช้ภาพของข้อมูลความลึกและภาพสีร่วมกัน โดยสังเกตได้ว่าภาพของข้อมูลความลึกจะสามารถแยกแยะระหว่างลำตัวกับมือได้ เนื่องจากมืออยู่ใกล้กับเซนเซอร์วัดความลึกมากกว่าลำตัวทำให้มีสีดำ ส่วนลำตัวมีสีเขียวและพื้นหลังมีสีน้ำเงิน จากนั้นจึงพิจารณาเฉพะตำแหน่งพิกเซลของภาพสีที่มีตำแหน่งเดียวกันกับพิกเซลที่มีสีดำของภาพข้อมูลความลึก เมื่อได้ตำแหน่งที่พิจารณาของภาพสีก็จะขจัดส่วนที่ไม่เกี่ยวข้องกับสีที่ถุงมือเช่นสีของผิวหนังโดยกำหนดช่วงของสีที่ไม่ใช่สีเนื้อ จากนั้นครอบตาราง ขนาด 23x20 เซลล์ที่ตำแหน่งเดียวกับข้อมูลความลึกเพื่อกำหนดขอบเขตที่สนใจดังรูปที่ 11

การสกัดข้อมูลจากขอบเขตที่สนใจของข้อมูลสีทำได้โดยการสกัดข้อมูลทีละเซลล์ของตาราง เหมือนกับข้อมูลความลึกแต่เนื่องจากข้อมูลสี แต่ละพิกเซลมี 3 เลเยอร์ได้แก่ สีแดง สีเขียวและสีน้ำเงิน จึงได้ใช้หลักการ Color histogram 4bins ซึ่งในแต่ละเซลล์จะมีจำนวนข้อมูล 64 ค่า



รูปที่ 11. กระบวนการหาขอบเขตที่สนใจของข้อมูลสี

จากนั้นทำการสกัดข้อมูลทุก ๆ เซลล์ในตารางโดยใช้หลักการข้างต้น และนำข้อมูลในแต่ละเซลล์มาต่อกันเพื่อให้ได้ข้อมูลของ 1 เฟรม โดยข้อมูลใน 1 เฟรมของข้อมูลสีจะมีจำนวน 23×20×64 = 29,440 ค่าและใช้หลักการนี้กระทำกับทุกเฟรมในวีดีโอจะได้ข้อมูลทั้งหมด 29,440×60 =

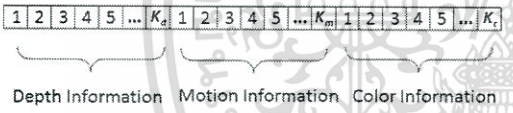
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1,766,400 ค่าสำหรับ 1 วีดีโอของข้อมูลสี โดยที่ 60 คือจำนวนเฟรมใน 1 วีดีโอ

3.2. การจำแนกภาษามือด้วยการรวมหลายลักษณะเด่น

กระบวนการทำงานของการรวมหลายลักษณะเด่นมี 2 ประเภท ได้แก่การรวมเอาลักษณะเด่นทั้ง 3 แบบเข้าด้วยกันได้แก่ข้อมูลความลึก ข้อมูลการเคลื่อนไหวและข้อมูลสี และการรวมเอาลักษณะเด่นทั้ง 2 แบบเข้าด้วยกันเช่น ข้อมูลความลึกและข้อมูลการเคลื่อนไหว โดยการใช้ข้อมูลในหัวข้อ 3.1 มาจัดวางในรูปแบบเวกเตอร์ 1 x M มิติ โดย M คือจำนวนข้อมูลทั้งหมด; K_d คือจำนวนข้อมูลความลึก; K_m คือจำนวนข้อมูลการเคลื่อนไหว; K_c คือจำนวนข้อมูลสี ดังแสดงในรูปที่ 12 และ 13

รูปที่ 12. การจัดวางเวกเตอร์ข้อมูลโดยการรวม 2 ลักษณะเด่น



รูปที่ 13. การจัดวางเวกเตอร์ข้อมูลโดยการรวม 3 ลักษณะเด่น

3.3. เทคนิค Normalization

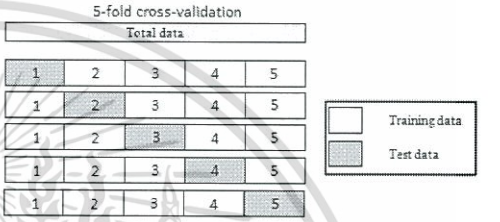
ในขั้นตอนการเตรียมข้อมูลจะมีการทำ Normalization ข้อมูลคือการปรับค่าที่สกัดได้ให้อยู่ในมาตราส่วนเดียวกัน โดยใช้สมการ (6)

$$NL = \frac{d - \mu}{\sigma} \qquad (6)$$

โดยที่ d คือข้อมูลทั้งหมด; μ คือค่าเฉลี่ยของข้อมูลทั้งหมด; σ คือค่าเบี่ยงเบนมาตรฐาน; NL คือค่าที่ผ่านการ Normalization แล้ว

3.4. เทคนิค 5-fold cross-validation

ในขั้นตอนการเตรียมข้อมูลจะใช้เทคนิค 5-fold cross-validation [5] ซึ่งเป็นวิธีการวัดประสิทธิภาพในการทำนายตัวอย่างของโมเดล เทคนิคคือการสุ่มตัวอย่าง (re-sampling) โดยแบ่งชุดข้อมูลออกเป็นส่วนๆ (fold) ในงานวิจัยนี้เลือกสุ่มข้อมูล 5 กลุ่ม ซึ่งแบ่งข้อมูลออกเป็น 5 ชุดเท่า ๆ กัน และทำการคำนวณค่าความแม่นยำจากการทำนาย 5 รอบ โดยแต่ละรอบจะมีการสร้างโมเดลจำแนกประเภทหนึ่งตัว จากข้อมูลเรียนรู้ 4 ชุด และใช้ข้อมูลทดสอบ 1 ชุด (ชุดที่ไม่ได้นำมาเรียนรู้) ดังรูปที่ 14



รูปที่ 14. Cross-validation

3.5. กระบวนการทำงานของซัพพอร์ตเวกเตอร์แมชชีน

3.5.1. ขั้นตอนการสอนให้ระบบเรียนรู้ (Training)

กระบวนการทำงานของซัพพอร์ตเวกเตอร์แมชชีนในขั้นตอนการสอนให้ระบบเรียนรู้คือการใช้ลักษณะเด่นของข้อมูลมาเรียนรู้ในระบบการเรียนรู้ (Machine Learning) โดยในขั้นตอนนี้ได้นำค่าของเวกเตอร์ของข้อมูลที่ได้จากการดึงลักษณะเด่น มาสอนให้กับระบบการเรียนรู้ซึ่งจะได้ข้อมูล x_i จากนั้นจึงกำหนดค่า y_i ซึ่งเป็นตัวกำหนดประเภทข้อมูลลักษณะเด่น โดยค่าของ y_i จะมีทั้งหมด 8 ค่าคือ 1 2 3 4 5 6 7 และ 8 ซึ่งเป็นการกำหนดท่าทางภาษามือของแต่ละท่าดังตารางที่ 1 การแบ่งกลุ่มแบบนี้เป็นการแบ่งกลุ่มข้อมูลแบบหลายกลุ่ม (Multiclass Support Vector Machine) ชนิด 1-VS-All นอกจากนี้ยังใช้เคอร์เนล (Kernel) เข้ามาใช้ในการแบ่งกลุ่มของข้อมูล ซึ่งเคอร์เนลที่ใช้มี 2 ชนิดคือแบบเชิงเส้น (Linear) และเรเดียนเบสิสฟังก์ชัน (Redial basis function)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 1. ตารางกำหนดค่า Class index

ภาษามือ	Class index
ชื่อ (Name)	1
ไม่ (No)	2
เท่าไหร่ (How many)	3
ขอบคุณ (Thank you)	4
อะไร (What)	5
ที่ไหน (Where)	6
ใช่ (Yes)	7
ของคุณ (Your)	8

3.5.1 ขั้นตอนการทดสอบระบบ (Testing)

กระบวนการทดสอบระบบจะทำโดยนำข้อมูลที่ดึงลักษณะเด่นเรียบร้อยที่ต้องการทดสอบมาทดสอบเพื่อเปรียบเทียบกับข้อมูลที่เรียนรู้แล้ว (Model) ซึ่งจะสามารถหาค่าความแม่นยำได้

4. ผลการวิจัย

4.1. การดำเนินการทดลอง

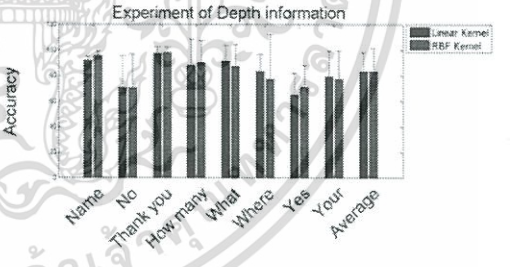
การดำเนินการทดลองจะแบ่งการทดลองเป็นการใช้ข้อมูลประเภทเดียว และการทดลองโดยใช้ข้อมูลร่วมกัน ซึ่งการทดลองโดยใช้ข้อมูลประเภทเดียวคือข้อมูลความลึก ข้อมูลการเคลื่อนไหวและข้อมูลสี ส่วนการทดลองโดยนำข้อมูลมารวมกันคือข้อมูลความลึกรวมกับข้อมูลการเคลื่อนไหวและข้อมูลการเคลื่อนไหวรวมกับข้อมูลสีและข้อมูลความลึกรวมกับข้อมูลการเคลื่อนไหวและข้อมูลสี โดยการเรียนรู้และทดสอบใน SVM มีการปรับ C ตั้งแต่ 10^{-16} ถึง 10^{16} โดยใช้ค่า C ที่ให้ค่าความแม่นยำสูงสุดในแต่ละข้อมูลดังตารางที่ 2

ตารางที่ 2. ค่า C ของแต่ละข้อมูล

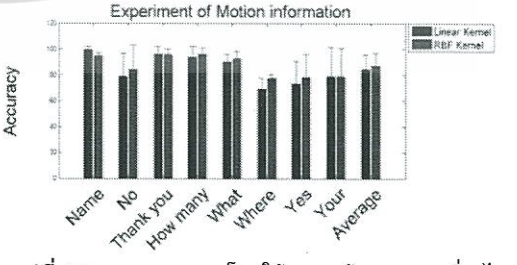
ข้อมูล	Kernel	ค่า C
เฉพาะข้อมูลความลึก	Linear	0.0001
	RBF	10
เฉพาะข้อมูลการเคลื่อนไหว	Linear	0.01
	RBF	1
เฉพาะข้อมูลสี	Linear	0.01
	RBF	1000
ข้อมูลความลึกรวมกับข้อมูลการเคลื่อนไหว	Linear	0.001
	RBF	10
ข้อมูลการเคลื่อนไหวรวมกับข้อมูลสี	Linear	0.01
	RBF	1000
ข้อมูลความลึกรวมกับข้อมูลการเคลื่อนไหวและข้อมูลสี	Linear	0.001
	RBF	0.0001

4.2. ผลการทดลอง

ผลการทดลองจะเป็นค่าความแม่นยำของแต่ละภาษามือซึ่งได้แก่ชื่อ (Name) ไม่ (No) เท่าไหร่ (How many) ขอบคุณ (Thank you) อะไร (What) ที่ไหน (Where) ใช่ (Yes) และของคุณ (Your) โดยทำการหาค่าความแม่นยำโดยใช้เคอร์เนล 2 ประเภทได้แก่เคอร์เนลเชิงเส้น (Linear kernel) และเคอร์เนลเรเดียนเบสิสฟังชัน (RBF kernel) ดังรูปที่ 15-20 และรูปที่ 21 แสดงการเปรียบเทียบค่าความแม่นยำของข้อมูลแต่ละประเภท

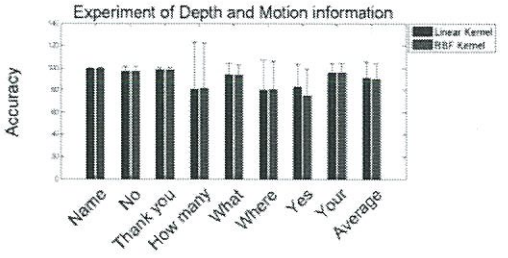


รูปที่ 15. ผลการทดลองโดยใช้เฉพาะข้อมูลความลึก

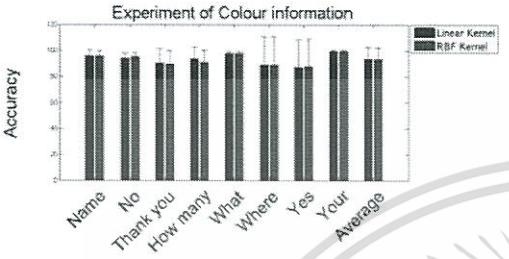


รูปที่ 16. ผลการทดลองโดยใช้เฉพาะข้อมูลการเคลื่อนไหว

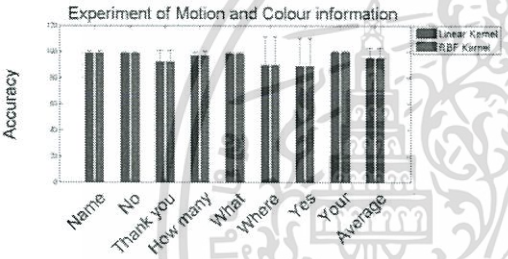
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



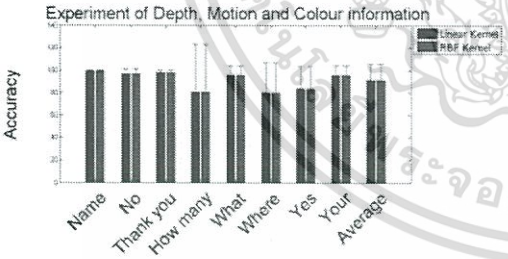
รูปที่ 17. ผลการทดลองโดยใช้ข้อมูลความลึกและข้อมูลการเคลื่อนไหว



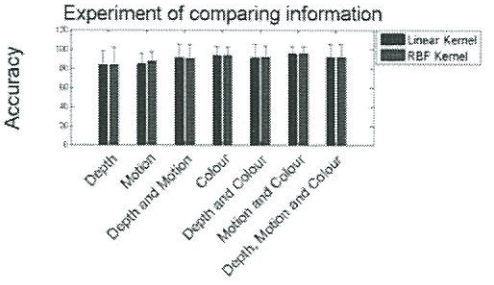
รูปที่ 18. ผลการทดลองโดยใช้เฉพาะข้อมูลสี



รูปที่ 19. ผลการทดลองโดยใช้ข้อมูลการเคลื่อนไหวและข้อมูลสี



รูปที่ 20. ผลการทดลองโดยใช้ข้อมูลความลึกและข้อมูลการเคลื่อนไหวและข้อมูลสี



รูปที่ 21. ผลการทดลองเปรียบเทียบค่าความแม่นยำของแต่ละข้อมูล

5. สรุปผลการทดลอง

งานวิจัยนี้ได้นำเสนอวิธีการจำแนกภาษามือโดยใช้อุปกรณ์คิเน็กท์เป็นตัวรับข้อมูลเพื่อนำมาสกัดข้อมูลหาลักษณะเด่นในแต่ละชนิด โดยข้อมูลที่นำมาสกัดได้แก่ข้อมูลความลึก ข้อมูลการเคลื่อนไหวและข้อมูลสี ซึ่งงานวิจัยที่เกี่ยวข้องของ Biswas และ Basu (2011) ได้ใช้ข้อมูลความลึกและข้อมูลการเคลื่อนไหวร่วมกันในการสกัดหาลักษณะเด่น [1] ส่วนงานวิจัยนี้ได้เพิ่มประสิทธิภาพโดยการใช้ข้อมูลสีและทดลองการใช้ข้อมูลแต่ละประเภทเดี่ยว ๆ ในการเรียนรู้และทดสอบใน SVM เนื่องมาจากงานวิจัยที่เกี่ยวข้องทดลองเฉพาะการใช้ข้อมูลความลึกร่วมกับข้อมูลการเคลื่อนไหวเท่านั้น

ผลการทดลองเปรียบเทียบค่าความแม่นยำระหว่างงานวิจัยที่เกี่ยวข้องพบว่าค่าความแม่นยำมีค่าสูงขึ้นเมื่อเพิ่มข้อมูลสี โดยการใช้ข้อมูลการเคลื่อนไหวและข้อมูลสีร่วมกันมีค่าความแม่นยำสูงสุดที่ร้อยละ 95.2 เนื่องจากข้อมูลความลึกต้องคำนึงถึงระยะห่างระหว่างผู้แสดงท่าทางกับอุปกรณ์คิเน็กท์ ทำให้การทำนายมีโอกาสผิดพลาดมากกว่า

ผลการทดลองโดยเฉลี่ยพบว่าภาษามือใช่ (Yes) มีค่าความแม่นยำต่ำที่สุดที่ร้อยละ 88.6 และภาษามือชื่อ (Name) และภาษามือของคุณ (Your) มีค่าความแม่นยำสูงที่สุดที่ร้อยละ 99.6

เอกสารอ้างอิง

[1] Biswas, K. K., Basu, S. K. **"Gesture recognition using Microsoft KINECT®"** In: *Proceeding of 5th International Conference on Automation, Robotics and Applications (ICARA 2011)*, 6-8 Dec 2011, Wellington, New Zealand, pp 100-103.

[2] Anant, A., Manish, K. T. **"Sign Language Recognition using Microsoft Kinect"** In: *Proceeding of 6th International Conference on Contemporary Computing (IC3 2013)*, 8-10 Aug 2013, Noida, pp 181-185.

[3] Daniel, H. **"Microsoft Kinect Sensor"** [Online].Available: http://gamelab.hkdm.de/wp-content/uploads/2014/04/fmx_2014_009_kinect.jpg. 2014

[4] Thiang, Andre, T. G., Resmana, L. **"Type of Vehicle Recognition Using Template Matching Method"** In: *Proceeding of the International conference on Electrical, Electronics, Communication and Information (CECI 2001)*, 7-8 March 2001, Jakarta.

[5] Chinnapat, K. **"k-fold cross-validation"** [Online].Available:http://scriptslines.com/blog/k-fold-cross-validation. 2013

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้