การรู้จำท่าทางการเตะและการชกของบุคคลโดยใช้กล้องมุมมองเดียว

นางสาวภาวินี จรูญพันธุ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2555

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย
บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository(CUIR)
are the thesis authors' files submitted through the Graduate School.

SINGLE VIEW HUMAN KICKING AND PUNCHING RECOGNITION

Miss Pawinee Jaroonphan

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Engineering Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2012

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การรู้จำท่าทางการเตะและการชกของบุคคลโดยใช้กล้องมุมมองเดียว
โดย	นางสาวภาวินี จรูญพันธุ์
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ นงลักษณ์ โควาวิสารัช

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

…….. คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.บุญสม เลิศหิรัญวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

……ประธานกรรมการ
(รองศาสตราจารย์ ดร.วิวัฒน์ วัฒนาวุฒิ)

…….. อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ นงลักษณ์ โควาวิสารัช)

…….. กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สืบสกุล พิภพมงคล)

……กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.บัณฑิต ทิพากร)

ภาวินี จรูญพันธุ์ : การรู้จำท่าทางการเตะและการชกของบุคคลโดยใช้กล้องมุมมองเดียว. (SINGLE VIEW HUMAN KICKING AND PUNCHING RECOGNITION) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : รศ.นงลักษณ์ โควาวิสารัช, 53หน้า.

งานวิจัยนี้นำเสนอวิธีการประมวลผลภาพดิจิทัลสำหรับการรู้จำท่าทางของบุคคล ได้แก่ การเตะและการชก โดยใช้กล้องถ่ายภาพวีดิทัศน์เพียงมุมมองเดียว ในฉากหลังทั้งแบบที่ไม่วุ่นวายและฉากหลังที่มีความหลากหลาย ซึ่งการเคลื่อนที่ของท่าทางมีลักษณะเป็นวิถีไปกลับแบบซ้ำแนวเดิม โดยหาบริเวณที่มีการเคลื่อนที่แบบคร่าวๆ ในขั้นแรกจากการฉายเงาจากภาพความต่างระหว่างเฟรม และหาเฉพาะบริเวณที่มีอวัยวะเคลื่อนที่โดยการเปรียบคู่แม่แบบ ขั้นต่อมาผู้วิจัยได้เลือกใช้วิธีการแสดงการเคลื่อนที่โดยใช้ MHI เพื่อแสดงการเคลื่อนที่ของแต่ละท่าทางในภาพเดียว แต่ MHI จะแสดงการเคลื่อนที่ที่เกิดการซ้อนทับวิถีเดิมเนื่องจากลักษณะของข้อมูล จากนั้นแก้ปัญหาด้วยการหาเฟรมกึ่งกลางของการเคลื่อนที่ เพื่อที่จะแบ่งการเคลื่อนที่ออกเป็นส่วนไม่ให้ทับซ้อนแนวกัน และใช้ซัพพอร์ตเวคเตอร์แมชชีน (SVM) ในการรู้จำข้อมูล

จากผลการทดลองการรู้จำท่าทางการเตะและการชก โดยมีท่าทางอื่นเป็นตัวทดสอบได้แก่ การถีบและการผลัก ซึ่งมีการทดลองโดยแยกท่าทางที่สนใจ 1 ชนิดออกจากท่าทางอื่น 1 ชนิดและแยกท่าทางที่สนใจ 1 ชนิด ออกจากท่าทางอื่นๆ 3 ชนิด มีการเพิ่มลดอัตราส่วนข้อมูลที่ใช้สอนระบบ การแบ่งส่วนการเคลื่อนที่ และการกำหนดลำดับการเคลื่อนที่ ผลที่ได้สามารถบอกได้ว่า การแยกออกจากท่าทาง 1 ชนิด ได้ค่าความถูกต้องสูงสุดทั้งการเตะและการชกคือ 98% และ 92% ตามลำดับ ค่าความถูกต้องสูงสุดจากการแยกออกจากท่าทางอื่นๆ 3 ชนิด ได้แก่ การเตะ 89% และการชก 84%

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

ปีการศึกษา 2555 .....

5270445721 : MAJOR COMPUTER ENGINEERING
KEYWORDS : GESTURE RECOGNITION/ KICKING/ PUNCHING/ MOTION HISTORY IMAGE/ SUPPORT VECTOR MACHINE/ REPEATING PATH TRAJECTORIES

PAWINEE JAROONPHAN: SINGLE VIEW HUMAN KICKING AND PUNCHING RECOGNITION. ADVISOR: ASSOC. PROF. NONGLUK COVAVISARUCH, 53 pp.

This research purposes a digital image processing approach for gesture recognition, from single view video sequence at static and clutter background. Specifically, repeating path trajectories kicking and punching. At first, found the rough's motion area by project sum of difference frame profile both of vertical and horizontal. Next, we used template matching for extract especially limb motion and then representing the ROI's movement by MHI. The MHI can be encoded movement sequence into single frame but limited of MHI is cannot present movement that have repeating path trajectory type. We solved by searching the mid-action frame for separate video sequence into parts of video sequence and recognize kicking and punching by SVM.

Our experiments are used other gestures for test recognition approach, there are foot thrusting and pushing. We tested by classify the target movement from 1 type of difference movement, and in addition to classify from 3 types of other gestures, after that ratio of input data for training, dividing movement to create parts of MHI and labeling history number are influence for the success rate, too. Resultant are showed 1 type of difference movement are highly rate that have highest recall rate of kicking and punching are 98% and 92%, respectively. The recall rate from 3 other gestures classification are 89% of kicking and 84% of punching.

Department : Computer Engineering Student's Signature .....
Field of Study : Computer Engineering Advisor's Signature .....
Academic Year : 2012 .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงเป็นอย่างดีเพราะความช่วยเหลือและสนับสนุนจากอาจารย์ที่ปรึกษาวิทยานิพนธ์ รศ.นงลักษณ์ โควาวิสารัช ซึ่งได้สละเวลาให้คำแนะนำปรึกษา เสนอแนะแนวทางการวิจัย และได้ให้ความช่วยเหลือดูแลในด้านต่าง ๆ เป็นอย่างดี ขอกราบขอบพระคุณอาจารย์เป็นอย่างสูง

ขอขอบพระคุณ ผศ.ดร. วิวัฒน์ วัฒนาวุฒิ ประธานกรรมการสอบวิทยานิพนธ์ กรรมการสอบวิทยานิพนธ์ ผศ.ดร. สืบสกุล พิภพมงคล และ ผศ.ดร. บัณฑิต ทิพากร ที่ให้ความกรุณาเสนอคำแนะนำ ข้อคิดเห็น และแนะแนวทางในการพัฒนางานวิจัยนี้ ตลอดจนถึงตรวจสอบ และแก้ไขข้อบกพร่องต่าง ๆ ของวิทยานิพนธ์ฉบับนี้

ขอขอบคุณ เพื่อนๆ พี่ๆ น้องๆ สมาชิกภายในห้องปฏิบัติการวิจัย CGCI ที่น่ารักทุกคนที่คอยให้คำแนะนำ ความช่วยเหลือ ในการเก็บข้อมูลเพื่อใช้ทดสอบ และเป็นกำลังใจตลอดการวิจัย

ขอขอบพระคุณผู้ที่มีส่วนช่วยเหลือและให้กำลังใจทุกท่านที่ทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี

และท้ายที่สุดขอกราบขอบพระคุณครอบครัว พ่อ แม่ และญาติพี่น้องทุกคน ที่ให้การอุปการะเลี้ยงดู เป็นกำลังใจ สนับสนุนด้านการศึกษาและในด้านต่างๆ จนวิทยานิพนธ์เล่มนี้สำเร็จลุล่วงไปด้วยดี

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ญ

สารบัญรูป ..... ฏ

เรื่อง

บทที่ 1 บทนำ 1

1. ความเป็นมาและความสำคัญของปัญหา 1
2. วัตถุประสงค์ 2
3. ขอบเขตงานวิจัย 2
4. ขั้นตอนการทำงาน 2
5. ประโยชน์ที่คาดว่าจะได้รับ 2
6. โครงสร้างวิทยานิพนธ์ 2

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 4

ทฤษฎีและหลักการ 4

1.ทฤษฎีสำหรับการสกัดคุณลักษณะท่าทาง (Feature Extraction) 4

1.1 การหักลบภาพฉากหลัง (Background Subtraction) 4

1.2 การหาความแตกต่างระหว่างเฟรม (Frame Difference) 5

1.3 การฉายเงาภาพ (Projection Profile) 6

1.4 Motion History Image (MHI) 8

2.ทฤษฎีสำหรับการการรู้จำหรือจัดประเภทข้อมูล (Recognition) 9

2.1 ซัพพอร์ตเวคเตอร์แมชชีน (Support Vector Machine: SVM) 9

2.2 การตรวจสอบแบบนำตัวหนึ่งออกไป (Leave-one-out Cross -validation) 10

2.3 Confusion Matrix 11

งานวิจัยที่เกี่ยวข้อง 12

1. งานวิจัยด้านการรู้จำท่าทางของบุคคล 12

1.1 ชุดข้อมูล(dataset) 12

เรื่อง หน้า

1.1.1 ชุดข้อมูลที่มีฉากหลังคงที่และว่าง 13
1.1.2 ชุดข้อมูลที่มีฉากหลังมีความหลากหลาย 14
1.2 ฟีเจอร์สำหรับท่าทางของบุคคล 15
1.2.1 ฟีเจอร์รูปร่าง 15
1.2.2 ฟีเจอร์การเคลื่อนที่ 16
1.2.3 ฟีเจอร์รูปร่างและการเคลื่อนที่ประกอบกัน 17
2. งานวิจัยที่แก้ปัญหาการเคลื่อนที่ที่มีวิถีไปกลับและซ้ำแนวเดิมโดยใช้ MHI เป็นตัวแสดงการเคลื่อนที่ 18
บทที่ 3 การรู้จำท่าทางการเตะและการชก 19
1. การหาบริเวณที่เกิดการเคลื่อนที่ (Human's Gesture Detection) 22
1.1 การหาบริเวณการเคลื่อนที่อย่างคร่าวๆ (Rough ROI's Detection) 22
1.1.1 การหาความแตกต่างระหว่างเฟรม (Frame Difference) 22
1.1.2 การรวมผลที่ได้จากความแตกต่างระหว่างเฟรม (Summation All of Frame Difference) 23
1.1.3 การฉายเงาภาพ (Projection Profile) 24
1.2 การหาเฟรมกึ่งกลางของชุดภาพการเคลื่อนที่ (Mid-action Frame Detection) 24
1.3 การหาบริเวณการเคลื่อนที่จากการเปรียบคู่แม่แบบ (ROI's from Template Matching) 25
2. การแสดงการเคลื่อนที่ในรูปแบบ MHI (MHI Feature Extraction) 26
3. การรู้จำและแยกประเภทท่าทาง (Action Classification) 28
บทที่ 4 การทดลองและผลการทดลอง 29
1. การรู้จำการเตะและการชกกับท่าทางที่ต่างกันชนิดละ 1 ท่าทาง 30
1.1 วิธีการทดลอง 31
1.2 ผลการทดลอง 31
1.3 วิเคราะห์ผลการทดลอง 34
2. การรู้จำการเตะและการชกกับท่าทางที่ต่างกันชนิดละ 3 ท่าทาง 35
2.1 วิธีการทดลอง 36
2.2 ผลการทดลอง 36
2.3 วิเคราะห์ผลการทดลอง 38

เรื่อง หน้า
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 39
สรุปผลการวิจัย 39
ข้อเสนอแนะ 41
รายการอ้างอิง 42
ภาคผนวก 44
ภาคผนวก ก 45
ภาคผนวก ข 48
ประวัติผู้เขียนวิทยานิพนธ์ 53

สารบัญตาราง

ตารางที่		หน้า
1	ผล Recall rate ของการรู้จำท่าทางการเตะและการชกกับท่าทางที่ต่างกัน 1 ท่าทาง	32
2	ผล Prediction rate ของการรู้จำท่าทางการเตะและการชกกับท่าทางที่ต่างกัน 1 ท่าทาง	33
3	ผล Accuracy rate ของการรู้จำท่าทางการเตะและการชกกับท่าทางที่ต่างกัน 1 ท่าทาง	34
4	ผล Recall rate ของการรู้จำท่าทางการเตะและการชกกับท่าทางที่ต่างกัน 3 ท่าทาง	37
5	ผล Prediction rate ของการรู้จำท่าทางการเตะและการชกกับท่าทางที่ต่างกัน 3 ท่าทาง	37
6	ผล Accuracy rate ของการรู้จำท่าทางการเตะและการชกกับท่าทางที่ต่างกัน 3 ท่าทาง	37
ก-1	แสดงจำนวนเฟรมของแต่ละท่าทางในหนึ่งชุดภาพวีดิทัศน์ ที่ฉากหลังคงที่	46
ก-2	แสดงจำนวนเฟรมของแต่ละท่าทางในหนึ่งชุดภาพวีดิทัศน์ ที่ฉากหลังที่มีความหลากหลาย	47
ข-1	ตัวอย่างภาพ Motion History Image (MHI) ที่แสดงผลการแยกด้วย Confusion Matrix	49

สารบัญรูป

รูปที่		หน้า
1	ตัวอย่างภาพการหักลบ ฉากหลัง	5
2	ตัวอย่างภาพการหาความแตกต่างระหว่างเฟรม	6
3	ตัวอย่างการฉายเงาภาพ	7
4	ตัวอย่างภาพที่เกิดจากการใช้ MHI แสดงการเคลื่อนที่	9
5	แสดงการจัดกลุ่มข้อมูลของซัพพอร์ตเวคเตอร์แมชชีน (SVM)	10
6	ตัวอย่างตาราง Confusion Matrix	11
7	ภาพตัวอย่างของชุดข้อมูล KTH ที่ประกอบด้วยการแสดง 6 ท่าทาง ใน 4 สถานการณ์	13
8	ภาพตัวอย่างของชุดข้อมูล Weizmann ประกอบด้วยการแสดง 10 ท่าทาง	14
9	ภาพตัวอย่างท่าทางของชุดข้อมูล IXMAS โดยแสดงในมุมกล้องที่ต่างกัน	14
10	ภาพตัวอย่างของชุดข้อมูล CMU	15
11	ตัวอย่างข้อมูลภาพวีดิทัศน์ที่แสดงท่าทางทั้ง 4 ท่าทาง ในฉากหลังคงที่	19
12	แผนผังลำดับงานการรู้จำท่าทางการเตะและการชก	20
13	ตัวอย่างข้อมูลภาพวีดิทัศน์ที่แสดงท่าทางทั้ง 4 ท่าทาง ในฉากหลังที่มีความหลาหลาย	21
14	ตัวอย่างภาพความแตกต่างระหว่างเฟรม	22
15	ผลรวมของภาพความแตกต่างระหว่างเฟรม	23
16	ตัวอย่างการฉายภาพและการเลือกขอบเขตของบริเวณที่สนใจคร่าวๆ	24
17	กราฟพื้นที่ของแต่ละเฟรมและเฟรมกึ่งกลางของการเคลื่อนที่ที่ถูกเลือก	25
18	การหาบริเวณที่สนใจจากการเปรียบคู่แม่แบบ	26
19	ตัวอย่างภาพ MHI ของท่าทางต่างๆ	27
20	ภาพตัวอย่าง MHI หลังจากแบ่งชุดภาพออกเป็น 2 ส่วน	27
21	ตัวอย่าง MHI จากการกำหนดลำดับเหตุการณ์เป็นตัวเลข	28

บทที่ 1

บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

งานวิจัยและศึกษาด้านการประยุกต์ใช้ภาพวีดิทัศน์ในปัจจุบันมีหลายแขนง เนื่องจากอุปกรณ์ที่ใช้เก็บภาพหาได้ง่ายและสามารถประยุกต์ใช้กับงานด้านต่างๆ ได้หลากหลาย เช่น การรู้จำวัตถุ การแยกประเภทวัตถุ การสำรวจ การระบุตัวตน การมองเห็นของหุ่นยนต์ เป็นต้น และการรู้จำวัตถุก็เป็นหัวข้อที่น่าสนใจในการประยุกต์ใช้กับภาพวีดิทัศน์ เพราะเป็นการค้นหาลักษณะหรือรูปแบบของวัตถุที่สนใจจากภาพวีดิทัศน์นั้น ซึ่งการรู้จำท่าทางของบุคคล หรือของมือ เป็นเรื่องที่นิยมนำมาศึกษามากที่สุด

การศึกษาด้านการรู้จำท่าทางของบุคคลเป็นการเรียนรู้ท่าทางบุคคลและหาวิธีการรู้จำท่าทางบุคคล เพื่อนำวิธีการไปประยุกต์ใช้ในงานที่ต้องการตรวจสอบท่าทางบุคคล โดยงานหรือสถานการณ์จะต่างกันออกไป และการเลือกใช้ท่าทางก็ขึ้นอยู่กับวัตถุประสงค์ของงานด้วย งานด้านการรู้จำท่าทางของบุคคลสามารถนำไปใช้กับงานหลายชนิด ตัวอย่างเช่น ด้านความปลอดภัยโดยมีการตรวจจับท่าทางที่อาจก่อให้เกิดอันตรายในสถานที่สาธารณะ หรือด้านกีฬาที่เป็นการตรวจจับท่าทางของนักกีฬาในขณะทำการแข่งขัน เป็นต้น เนื่องจากวัตถุประสงค์ของงานที่ใช้การรู้จำท่าทางบุคคลมีหลากหลาย ทำให้ท่าทางที่นำมาเรียนรู้มีความหลากหลายเช่นกัน เช่น ท่าทางการใช้โทรศัพท์ ท่าทางการชี้บอก ท่าทางการเดิน ท่าทางการชก ท่าทางการเตะ เป็นต้น

การตรวจจับการทำร้ายร่างกายหรือกีฬาประเภทที่ต้องตรวจจับท่าทางประกอบการให้คะแนนมักมีปัญหาคือ ไม่สามารถตรวจจับท่าทางได้ทันทีหรือเกิดการแสดงท่าทางเร็วมากจนมองไม่ทัน จึงนำการรู้จำท่าทางบุคคลจากภาพมาประยุกต์ใช้เพื่อช่วยในการตรวจจับท่าทางบุคคล ท่าทางที่เลือกศึกษาสำหรับงานวิจัยนี้ คือ ท่าทางการเตะและการชก เนื่องจากการเตะและการชกเป็นการแสดงท่าทางของแขนและขาที่สามารถเห็นการเคลื่อนที่ในส่วนของแขนและขาชัดเจน ซึ่งลักษณะการเคลื่อนที่ของท่าทางเหล่านี้จะเป็นการเคลื่อนที่ที่มีวิถีไปกลับและซ้ำแนวเดิม เก็บข้อมูลเป็นภาพวีดิทัศน์โดยกำหนดให้นักแสดงแสดงท่าทางขณะหันหน้าไปด้านข้างในฉากหลังที่คงที่ไม่วุ่นวายและฉากหลังที่มีความหลากหลาย โดยใช้กล้องวีดิทัศน์ถ่ายในแนวระดับเดียวกับสายตาเพียงมุมมองเดียว

2. วัตถุประสงค์

เพื่อศึกษาวิธีการรู้จำท่าทางการเตะและการชกในฉากหลังที่ซับซ้อนและมีการเคลื่อนที่ โดยจับภาพมุมมองเดียวในระดับสายตา

3. ขอบเขตของงานวิจัย

3.1 ข้อมูลเป็นภาพวิดีโอที่ถ่ายในระดับสายตา เพียงมุมมองเดียว และใช้กล้องตัวเดียว

3.2 เก็บภาพในช่วงเวลากลางวันที่มีแสงสว่างเพียงพอ เพื่อการถ่ายภาพที่ชัดเจน

3.3 ภาพที่ใช้ทดสอบต้องเห็นภาพเต็มตัว เห็นตั้งแต่ศีรษะจนถึงเท้า

3.4 ท่าทางการเตะจะแสดงการเตะแบบเตะขาเหยียดตรงและแกว่งขาสูง โดยสูงสุดที่ระดับเอวทั้ง ขาซ้ายและขวา

3.5 ท่าทางการชกเป็นการยืนชกแบบไปข้างหน้าจากระดับหน้าอก ทั้งแขนซ้ายและขวา

4. ขั้นตอนการทำงาน

4.1 ศึกษางานวิจัยที่เกี่ยวกับการรู้จำท่าทางบุคคล

4.2 ศึกษาข้อมูลอัลกอริทึมที่เกี่ยวกับการรู้จำท่าทางบุคคล

4.3 เก็บภาพข้อมูลเพื่อนำมาใช้ในการทดสอบ

4.4 ออกแบบอัลกอริทึมเพื่อใช้ในการพัฒนาโปรแกรม

4.5 ทดสอบนำอัลกอริทึมมาใช้กับข้อมูลที่เก็บมา

4.6 ประเมินผลจากการทดสอบ

4.7 วิเคราะห์ผลการทดสอบ

4.8 สรุปผลและจัดทำวิทยานิพนธ์

5. ประโยชน์ที่ได้รับ

5.1 สามารถนำไปประยุกต์ใช้กับงานด้านการตรวจจับท่าทางที่มีการเตะและการชกเกี่ยวข้อง เช่น กีฬาเทควันโด มวยไทย เป็นต้น

5.2 สามารถนำไปประยุกต์ใช้กับงานด้านการรักษาความปลอดภัยได้

6. โครงสร้างวิทยานิพนธ์

วิทยานิพนธ์นี้มีทั้งหมด 5 บท ประกอบด้วย บทที่ 1 จะกล่าวถึงความเป็นมาและความสำคัญของปัญหา วัตถุประสงค์งานวิจัย ขอบเขตงานวิจัย ขั้นตอนการทำงาน และประโยชน์ที่ได้รับ ดังที่ได้กล่าวมาแล้ว บทต่อมา จะกล่าวถึงทฤษฎีที่เกี่ยวข้อง และงานวิจัยที่เกี่ยวข้องกับงานวิจัย บทที่ 3 เป็นขั้นตอนการดำเนินงาน

สำหรับการรู้จำท่าทางการเตะและการชก ต่อมาการทดลองและผลการทดลองแสดงในบทที่ 4 และบทสุดท้ายบทที่ 5 คือสรุปผลการวิจัยและข้อเสนอแนะต่างๆ

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

บทนี้จะกล่าวถึงทฤษฎีที่จะนำมาประยุกต์ใช้สำหรับงานวิจัยนี้ ซึ่งในแต่ละทฤษฎีอาจจะมีการปรับเปลี่ยนหรือประยุกต์ใช้ในหลากหลายรูปแบบกับงานอื่นๆ แต่ในบทนี้จะกล่าวถึงเฉพาะส่วนที่นำมาใช้งานกับงานวิจัยนี้เท่านั้น ประกอบไปด้วยทฤษฎีที่ใช้สำหรับสกัดคุณลักษณะที่สนใจออกมาและทฤษฎีสำหรับการแยกประเภทข้อมูล แสดงใน “ทฤษฎีที่เกี่ยวข้อง”

นอกจากนี้ยังมีการศึกษางานวิจัยอื่นๆ ที่มีการศึกษาด้านการรู้จำบุคคลที่คล้ายคลึงกับงานวิจัยนี้ โดยแบ่งออกเป็น 2 ส่วน ประกอบด้วยงานวิจัยที่มีการศึกษาด้านการรู้จำท่าทางของบุคคลโดยทั่วไป และงานวิจัยที่มีการศึกษาด้านการรู้จำบุคคลโดยใช้ MHI (Motion History Image) ซึ่งเน้นงานวิจัยที่แก้ปัญหาที่เกิดจากการเคลื่อนที่ซ้ำเส้นทางเดิมของ MHI เนื่องจากถ้ามีวิถีการเคลื่อนที่ไปกลับและซ้ำแนวทางเดิม MHI จะไม่สามารถแสดงวิถีการเดินทางทั้งหมดได้ เพราะจะมีการซ้อนทับกันของข้อมูล และงานวิจัยที่กล่าวถึงเหล่านี้จะแสดงใน “งานวิจัยที่เกี่ยวข้อง”

ทฤษฎีที่เกี่ยวข้อง

งานด้านการรู้จำท่าทางของบุคคลจะประกอบไปด้วยสองส่วนหลักๆ คือ การสกัดคุณลักษณะของท่าทางออกจากภาพวีดิทัศน์ และการรู้จำหรือแยกคุณลักษณะของท่าทางที่สนใจ และในส่วนนี้จะแยกทฤษฎีออกเป็น 2 กลุ่มเช่นกัน คือ ทฤษฎีที่ใช้สำหรับสกัดคุณลักษณะของท่าทาง และทฤษฎีที่ใช้สำหรับการรู้จำหรือจัดประเภทข้อมูล ได้แก่

1. ทฤษฎีสำหรับการสกัดคุณลักษณะท่าทาง (Feature Extraction)

1.1 การหักลบภาพฉากหลัง (Background Subtraction)

เป็นการสกัดภาพฉากหน้า (Foreground) ออกจากภาพอ้างอิง โดยที่ภาพที่เก็บมานั้นได้จากกล้องที่มีตำแหน่งคงที่ ซึ่งในหัวข้อนี้ได้ให้ภาพฉากหลังหรือภาพที่มีสภาพแวดล้อมคงที่ภาพหนึ่งเป็นภาพอ้างอิง โดยการสร้างภาพพื้นหลังนั้นสามารถสร้างขึ้นมาได้หลายวิธี อาจได้มาจากการสอนระบบจากภาพหลายภาพในตำแหน่งที่ระบุว่าเป็นภาพฉากหลังโดยไม่มีสิ่งรบกวน เพื่อความทนทานต่อการเปลี่ยนแปลงของแสงและสิ่งแวดล้อมในภาพ หรืออาจได้จากการกำหนดภาพใดภาพหนึ่งที่อยู่ในตำแหน่งเดียวกันให้เป็นภาพฉากหลัง จากนั้นเมื่อต้องการตรวจหาว่ามีวัตถุหรือฉากหน้าเพิ่มเติมเข้ามาหรือไม่ก็สามารถใช้ภาพในเวลาปัจจุบันหักลบกับภาพฉากหลังนั้นก็จะสามารถตรวจสอบได้จากค่าความแตกต่าง

ของความเข้มแสงถ้ามากกว่าค่าขีดแบ่งที่กำหนดก็แสดงว่ามีวัตถุหรือฉากหน้าเพิ่มเข้ามาจากพื้นหลังเดิม ซึ่งสมการการหักลบภาพพื้นหลังหาได้จาก

$$|I(x,y,t) - B(x,y,t)| > Th \quad (1)$$

$B(x,y,t)$ คือ ภาพฉากหลัง $I(x,y,t)$ คือภาพที่สนใจที่มีค่าความเข้มแสงที่ตำแหน่ง (x,y) ณ เฟรม t และ Th คือค่าขีดแบ่งของผลความแตกต่าง ผลที่ได้จากการหักลบฉากหลังเมื่อนำมาเปรียบเทียบกับค่าขีดแบ่งจะได้วัตถุหรือฉากหน้าที่เพิ่มเข้ามา ดังแสดงตัวอย่างในรูปที่ 1 และหากเลือกใช้ภาพพื้นหลังเพียงภาพเดียวเป็นภาพอ้างอิง การเปลี่ยนแปลงของแสงและสภาพแวดล้อมในเวลาที่ต่างออกมาอาจจะทำให้มีข้อมูลที่นอกเหนือจากวัตถุหรือฉากหน้าที่สนใจปะปนเข้ามาด้วยได้

(ก) (ข) (ค)

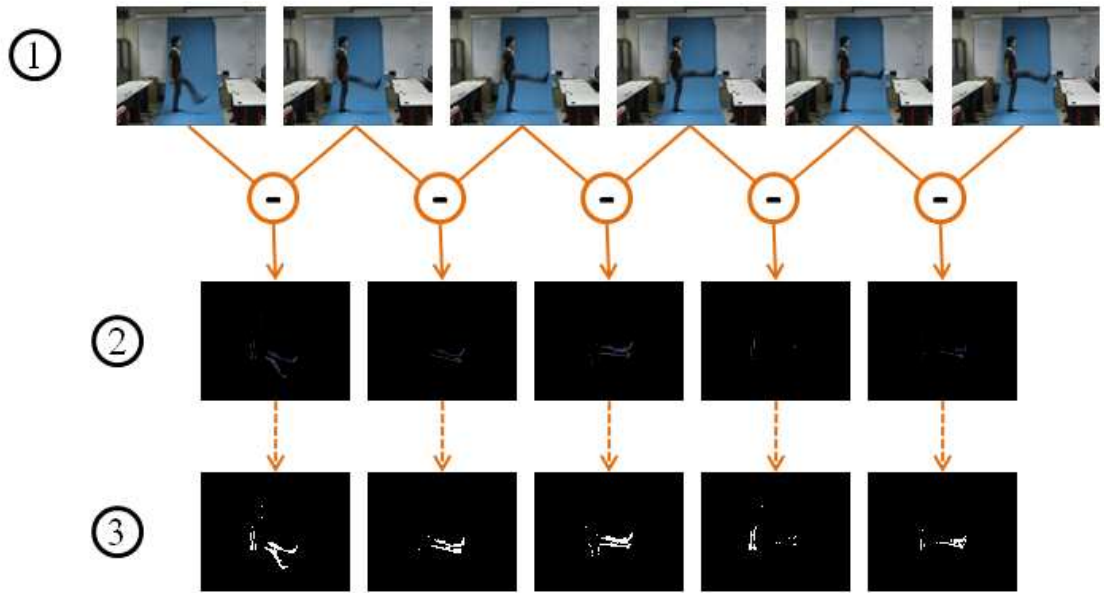
รูปที่ 1 ตัวอย่างภาพการหักลบฉากหลัง
(ก) ภาพเฟรมที่สนใจ (ข) ภาพพื้นหลัง (ค) ภาพวัตถุที่ได้จากการหักลบพื้นหลัง

1.2 การหาความแตกต่างระหว่างเฟรม (Frame Difference)

เป็นวิธีการพื้นฐานอีกวิธีหนึ่งของการหักลบภาพฉากหลังที่ใช้ในการความแตกต่างระหว่างภาพที่สนใจกับภาพอ้างอิง โดยให้ภาพที่เฟรมปัจจุบันเป็นภาพที่สนใจและภาพอ้างอิงคือเฟรมที่ต่อเนื่องกัน ซึ่งจะใช้ตรวจสอบว่ามีการเคลื่อนที่เกิดขึ้นระหว่างช่วงเวลานั้นหรือไม่ หากไม่มีการเคลื่อนที่หรือภาพในเฟรมยังคงเป็นภาพเดิมความแตกต่างของค่าความเข้มแสงจะน้อยมากหรือไม่มีเลย แต่ถ้ามีการเคลื่อนที่เกิดขึ้นค่าความแตกต่างของความเข้มแสงในบริเวณที่เกิดการเคลื่อนที่นั้นจะมีผลต่างเกิดขึ้น ซึ่งการหาค่าความแตกต่างนั้นจะพิจารณาที่ค่าความเข้มแสงในแต่ละจุดภาพเมื่อเกิดการเคลื่อนที่ค่าความเข้มแสงในตำแหน่งนั้นจะเปลี่ยนไป และหากเปรียบเทียบกับเฟรมก่อนหน้าที่ตำแหน่งเดียวกันจะเกิดผลต่างของค่าความเข้มแสง ดังนั้นผลต่างนี้จะเป็นตัวบอกว่ามีการเปลี่ยนแปลงหรือการเคลื่อนที่ที่บริเวณนั้นๆ โดยสมการที่ใช้ในการคำนวณค่าความแตกต่างระหว่างเฟรมคือ

$$|I(x,y,t) - I(x,y,t-1)| > Th \quad (2)$$

$I(x,y,t)$ คือค่าความเข้มแสงที่ตำแหน่ง (x,y) เฟรม t ส่วน $I(x,y,t-1)$ เป็นเฟรมก่อนหน้า และ Th คือค่าขีดแบ่งของผลความแตกต่าง ผลที่ได้จากการหาความแตกต่างระหว่างเฟรมแสดงดังตัวอย่างรูปที่ 2



รูปที่ 2 ตัวอย่างภาพการหาความแตกต่างระหว่างเฟรม (1) เฟรมภาพที่ต่อเนื่องกัน (2) ผลจากการหาความแตกต่างระหว่างเฟรม (3) แปลงเป็นภาพขาว-ดำ

ถ้าพิจารณาจากความเร็วของการเคลื่อนที่แบบเฟรมต่อเฟรม จะพบว่าบริเวณที่มีการเคลื่อนที่เร็วจะส่งผลให้เกิดบริเวณของความแตกต่างระหว่างเฟรมมาก และหากเคลื่อนที่ช้าบริเวณของความแตกต่างจะน้อยตามไปด้วย แสดงว่าบริเวณของความแตกต่างที่เกิดจากการเคลื่อนที่จะแปรผันตรงกับความเร็วด้วย นอกจากความเร็วในการเคลื่อนที่จะมีผลต่อบริเวณของความแตกต่าง ขนาดของวัตถุก็มีผลต่อบริเวณของความแตกต่างเหมือนกัน

1.3 การฉายเงาภาพ (Projection Profile)

การที่แสงส่องที่วัตถุจะทำให้เกิดเงาของรูปทรงนั้นบนพื้นอ้างอิง ซึ่งเงาของวัตถุจะเปรียบเสมือนจำนวนของจุดภาพที่อยู่ในแนวที่ตั้งฉากกับแกนอ้างอิง ซึ่งแกนอ้างอิงแบ่งเป็นสองแกนคือ แกนตั้งและแกนนอน โดยที่การฉายเงาภาพในแนวแกนตั้ง หรือ Vertical Projection Profile และการฉายเงาภาพในแนวแกนนอน หรือ Horizontal Projection Profile โดยที่จะเป็นการสะสมค่าในแต่ละจุดภาพตามแนวที่ต้องการ ซึ่งการฉายเงาภาพในแนวแกนตั้งจะเป็นการสะสมค่าทุกๆ แถว (row) ในแต่ละสดมภ์

(column) แสดงดังรูปที่ 3(ข) และการฉายเงาภาพในแนวแกนนอนจะเป็นการสะสมค่าทุกๆ สดมภ์ในแต่ละแถว แสดงดังรูปที่ 3(ค) และสมการที่ 3 จะแสดงการคำนวณของการฉายเงาภาพ โดย $V[i]$ คือการฉายเงาภาพตามแนวแกนตั้ง ส่วน $H[i]$ เป็นการฉายเงาภาพตามแนวแกนนอน n, m คือจำนวนของจุดภาพตามแนวแกนตั้งและแกนนอนตามลำดับ [1]

$$H[j] = \sum_{j=0}^{m} B[i,j] \tag{3}$$

$$V[i] = \sum_{i=0}^{n} B[i,j] \tag{4}$$

$B[i,j]$ คือภาพระดับเทาที่มีมีขนาดความสูงเป็น m และ ความกว้างเป็น n จุดภาพ อย่างไรก็ตามการฉายเงาภาพระดับเทาจะมีข้อเสียคือจะได้ข้อมูลที่ไม่มีเอกลักษณ์ของภาพ โดยภาพมากกว่าหนึ่งภาพสามารถมีเงาภาพฉายที่เหมือนกันได้ นอกจากนี้ถ้าเป็นภาพฐานสอง (Binary Image) การหาภาพเงาการฉายได้โดยสามารถนับจำนวนจุดภาพได้เลย เนื่องจากแต่ละจุดภาพจะมีค่าเท่ากับ 1 และผลที่ได้จะมีค่าเท่ากับจำนวนของจุดภาพที่อยู่ตามแนวแกนนั้น

(ก)

(ค)

(ข)

รูปที่ 3 ตัวอย่างการฉายเงาภาพ
(ก) ตัวอย่างภาพระดับเทาที่นำมาทำการฉายเงาภาพ
(ข) การฉายเงาภาพในแนวแกนตั้ง
(ค) การฉายเงาภาพในแนวแกนนอน

1.4 Motion History Image (MHI)

Motion history Image หรือ MHI [2, 3, 4] เป็นวิธีการหนึ่งที่แสดงตำแหน่งของการเคลื่อนที่ของวัตถุจากแต่ละเฟรมในภาพวีดิทัศน์หนึ่งชุดภาพตามลำดับเหตุการณ์มารวมกันไว้ในภาพเดียว และแสดงค่าแต่ละบริเวณนั้นด้วยค่าความเข้มแสงต่าง ๆ กันเพื่อสื่อถึงลำดับการเกิดเหตุการณ์ก่อนหรือหลัง บริเวณที่มีความเข้มแสงที่มากที่สุด (หรือสว่างที่สุด) สื่อถึงบริเวณที่มีการเคลื่อนที่ผ่านล่าสุดหรือปัจจุบันที่สุด ส่วนบริเวณที่มีการเคลื่อนที่ผ่านเฟรมก่อนหน้าก็จะแสดงด้วยค่าความเข้มแสงที่ลดลงตามลำดับการเกิดเหตุการณ์ โดยที่เฟรมที่แสดงการเคลื่อนที่ในลำดับแรกสุด (คือปัจจุบันน้อยที่สุด) อาจจะมีความเข้มแสงบริเวณที่มีการเคลื่อนที่ผ่านน้อยมากๆ หรือเป็นสีดำก็ได้

การสร้าง MHI ซึ่งแสดงบริเวณที่มีการเคลื่อนที่ตามลำดับเหตุการณ์ที่เกิดก่อนหลังด้วยภาพเพียงภาพเดียว ทำได้โดยปรับเปลี่ยนบริเวณที่มีการเคลื่อนที่ไปตามลำดับของข้อมูลการเคลื่อนที่ที่เกิดขึ้น บริเวณที่มีการเคลื่อนที่ล่าสุด (คือปัจจุบันที่สุด) จะถูกกำหนดให้มีค่าความเข้มแสงสูงสุด ส่วนบริเวณที่มีการเคลื่อนที่ที่เกิดขึ้นก่อนหน้าก็จะถูกลดค่าความเข้มแสงลงไปตามลำดับเหตุการณ์ก่อนหลัง

เนื่องจากแต่ละเฟรมในชุดภาพวีดิทัศน์จะต้องถูกหาบริเวณที่มีการเคลื่อนที่มาก่อน แล้วค่อยปรับเปลี่ยนบริเวณการเคลื่อนที่ล่าสุดกับ MHI ด้วยฟังก์ชันที่ใช้ปรับค่า (Update function) ซึ่งในงานวิจัยนี้จะหาบริเวณและตำแหน่งที่อวัยวะมีการเคลื่อนที่จากภาพแรกด้วยการหักลบภาพฉากหลังจากภาพแต่ละเฟรม โดยให้ภาพแรกของแต่ละชุดภาพวีดิทัศน์เป็นพื้นหลังหรือภาพอ้างอิง

$$H(x,y,t) = |I(x,y,t) - I(x,y,1)| \tag{5}$$

$H(x,y,t)$ คือ ภาพที่ได้จากการหักลบภาพฉากหลังที่เวลา t และ $I(x,y,t)$ เป็นภาพที่สนใจที่เวลา t และ $I(x,y,1)$ คือภาพแรกของแต่ละชุดภาพ บริเวณที่ได้จากการหักลบภาพฉากหลังนี้จะถูกนำไปเป็นฟังก์ชันที่ใช้ปรับค่าตามลำดับเหตุการณ์ โดยฟังก์ชันที่ใช้ปรับค่าหาได้จากสมการที่ (6)

$$\psi(x,y,t) = \begin{cases} 0, & H(x,y,t) < \xi \\ 1, & H(x,y,t) \geq \xi \end{cases} \tag{6}$$

จากสมการ $\psi(\mathrm{x},\mathrm{y},\mathrm{t})$ คือ ฟังก์ชันที่ใช้ปรับค่า (Update function) ที่ได้ผลเป็นภาพขาว-ดำ โดยใช้ค่าขีดแบ่ง (ξ) ที่ได้จาก Otsu's Method เมื่อได้ฟังก์ชันสำหรับปรับค่าที่เป็นภาพขาว-ดำแล้ว ก็นำมาใช้ในการปรับปรุง MHI ตามสมการ (7)

$$MHI(x,y,t) = \begin{cases} \tau & if\ \psi(x,y,t) = 1 \\ \max(0, MHI(x,y,t-1) - \delta) & otherwise \end{cases} \quad (7)$$

จะพบว่า MHI จะมีการปรับค่าไปเรื่อยๆ ตามฟังก์ชันที่ใช้ปรับค่า บริเวณใดในภาพขาว-ดำของฟังก์ชันปรับค่าเป็น 1 จะให้บริเวณนั้นแทนด้วยค่า τ ซึ่งในที่นี้คือค่าความเข้มแสงสูงสุดคือสีขาว และ δ เป็นตัวแปรที่คอยลดค่าความเข้มแสงลงเรื่อยๆ ในบริเวณที่ไม่ใช่บริเวณการเคลื่อนที่ล่าสุด ภาพตัวอย่างของ MHI แสดงดังรูปที่ 4 ซึ่งบริเวณล่าสุดของการเคลื่อนที่จะแสดงด้วยสีขาว และลำดับของเหตุการณ์ก่อนหน้าจะมีการไล่ระดับสีเทาลงเรื่อยๆ จนถึงสุดท้ายเป็นสีดำ

รูปที่ 4 ตัวอย่างภาพที่เกิดจากการใช้ MHI แสดงการเคลื่อนที่

2. ทฤษฎีสำหรับการรู้จำหรือจัดประเภทข้อมูล (Recognition)

ในส่วนนี้จะกล่าวถึงทฤษฎีที่ใช้ในการรู้จำหรือช่วยในการจัดประเภทข้อมูล ซึ่งการรู้จำท่าทางที่สนใจมักจะใช้การเรียนรู้ของเครื่อง (Machine Learning) เป็นเครื่องมือช่วยในการสอนระบบให้รู้จักท่าทางนั้นและจัดประเภทข้อมูล เนื่องจากต้องการแยกประเภทข้อมูล 2 ชนิด จึงเลือก ซัพพอร์ตเวคเตอร์แมชชีน (Support Vector Machine: SVM) ช่วยในการจัดประเภท ส่วนของการทดสอบระบบจะใช้วิธีการทดสอบแบบนำตัวหนึ่งออกไป (Leave-one-out Cross-validation) และการวัดความถูกต้องจะพิจารณาจาก Confusion Metrix

2.1 ซัพพอร์ตเวคเตอร์แมชชีน (Support Vector Machine: SVM)

การแยกประเภทข้อมูลด้วยซัพพอร์ตเวคเตอร์แมชชีน หรือ SVM [5, 6] เป็นวิธีที่ใช้อย่างแพร่หลายสำหรับการรู้จำรูปแบบ (Pattern Recognition) เพราะเป็นวิธีที่เหมาะสำหรับการจดจำรูปแบบ เมื่อมีข้อมูลที่ต้องการเรียนรู้รูปแบบ (Training Data) เข้ามา SVM จะทำการสร้างไฮเปอร์เพลน (Hyperplane) ที่เหมาะสมเพื่อที่จะแบ่งข้อมูลออกเป็นกลุ่ม ซึ่งจะพยายามแยกข้อมูลเพื่อให้ได้ระยะห่างระหว่างกลุ่มให้มากที่สุดและสร้างไฮเปอร์เพลนระหว่างกลุ่ม โดยเว้นระยะห่างระหว่างทั้งสองกลุ่มเท่ากัน

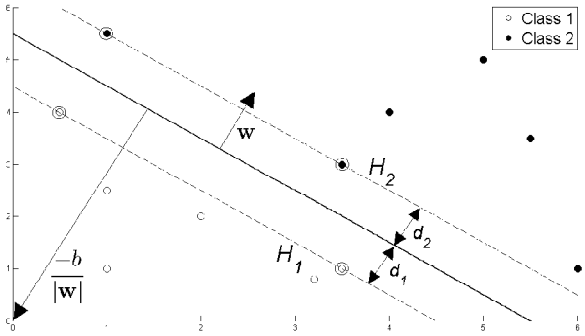
ในบทนี้จะกล่าวถึงหลักการโดยคร่าวๆ ของซัพพอร์ตเวคเตอร์แมชชีน หากต้องการเรียนรู้เพิ่มเติมสามารถศึกษาได้จาก [7, 8] และข้อมูลที่ต้องการเรียนรู้จะมีลักษณะดังนี้

$$\{x_i, y_i\} \quad where \ i = 1,2,\dots,m \ ; \ y_i \in \{-1,1\} ; \ x_i \in \Re^D \tag{8}$$

เมื่อ x_i เป็นข้อมูลที่ต้องการเรียนรู้ชุดที่ i โดยที่ใน x_i จะมีข้อมูลที่ต้องการเรียนรู้ทั้งหมด D จำนวน และ y_i คือค่าที่บอกว่าข้อมูล x_i จัดอยู่ในกลุ่มใด และจำนวนข้อมูลมีทั้งหมด m ชุด

การสร้างไฮเปอร์เพลนจะสร้างโดยอาศัยสมการ (9) ซึ่งจะพยายามแยกข้อมูลออกเป็นสองกลุ่ม ได้ w เป็นระยะห่างระหว่าง x_i กับไฮเปอร์เพลนและ $\frac{b}{\|w\|}$ เป็นระยะห่างจากไฮเปอร์เพลนถึงจุดกำเนิด เมื่อได้ไฮเปอร์เพลนที่เหมาะสมจะพบว่าจะได้ระยะห่างระหว่างกลุ่มมากที่สุด โดยข้อมูลที่อยู่ใกล้ไฮเปอร์เพลนที่สุดของทั้งสองกลุ่มจะเรียกซัพพอร์ตเวคเตอร์ (Support Vector) ซึ่งมีระยะห่างจากไฮเปอร์เพลนคือ d

$$x \cdot w + b = 0 \tag{9}$$



รูปที่ 5 แสดงการจัดกลุ่มข้อมูลของซัพพอร์ตเวคเตอร์แมชชีน (SVM) [6]

2.2 การตรวจสอบแบบนำตัวหนึ่งออกไป (Leave-one-out Cross -validation)

การทดสอบระบบการเรียนรู้ข้อมูลจากการเรียนรู้ของเครื่อง สามารถทำได้โดยมีข้อมูลสำหรับสอนระบบ (Training Data) และข้อมูลสำหรับทดสอบระบบ (Testing Data) ซึ่งวิธีการนี้จะเป็นการเลือกข้อมูลทดสอบออกจากข้อมูลทั้งหมดมา 1 ตัว และให้ที่เหลือเป็นข้อมูลสอนระบบ สับเปลี่ยนข้อมูลทดสอบโดยไม่ให้ซ้ำกับตัวเดิมจนครบทุกตัวในข้อมูลที่มีทั้งหมด และนำผลที่ได้จากการทดสอบแต่

ละครั้งมาเฉลี่ยด้วยจำนวนของข้อมูลทั้งหมด ผลเฉลี่ยจากการทดสอบด้วยการตรวจสอบแบบนำตัวหนึ่งออกไปแสดงตามสมการ [9]

$$E = \frac{1}{n}\sum_{i=1}^{n} E_i \tag{10}$$

E คือผลเฉลี่ยจากการทดสอบ โดยที่มีจำนวนข้อมูลทั้งหมด n ข้อมูล และ E_i เป็นผลที่ได้จากการทดสอบที่ข้อมูลในลำดับที่ i

2.3 Confusion Matrix [10]

เป็นการวัดประสิทธิภาพของการเรียนรู้ของระบบ โดยจะมีการเปรียบเทียบผลจากระบบที่ทำนายมา (Predicted) กับข้อมูลที่แท้จริง (Actual) ตามรูปที่ 6

		Actual Class	
		Yes	No
Predicted Class	Yes	TP	FP
	No	FN	TN

รูปที่ 6 ตัวอย่างตาราง Confusion Matrix

จากรูปที่ 6 จะสามารถอธิบายตัวแปรในตาราง confusion ได้ดังนี้

TP (True Positive) คือผลจากการทำนายทายว่า “ใช่” และผลจริง “ใช่”

FP (False Positive) คือผลจากการทำนายทายว่า “ใช่” แต่ผลจริง “ไม่ใช่”

FN (False Negative) คือผลจากการทำนายทายว่า “ไม่ใช่”แต่ผลจริง “ใช่”

TN (True Negative) คือผลจากการทำนายทายว่า “ไม่ใช่” และผลจริง “ไม่ใช่”

และการคำนวณค่าประสิทธิภาพสามารถคำนวณได้หลายค่าแล้วแต่ว่าต้องการเลือกใช้ค่าไหนโดยในงานวิจัยนี้จะกล่าวถึง 4 ค่าคือ

$$Accuracy\ rate = \frac{TP + TN}{TP + FP + FN + TN} \quad (11)$$

$$Pricision\ rate = \frac{TP}{TP + FP} \quad (12)$$

$$Recall\ rate\ = \frac{TP}{TP + FN} \quad (13)$$

แต่ละสมการจะอธิบายประสิทธิภาพจากการทดสอบระบบดังนี้

Accuracy rate	เป็นอัตราส่วนการทำนายข้อมูลทั้งใช่และไม่ใช่ตรงกับข้อมูลจริงจากทั้งหมด
Precision rate	เป็นอัตราส่วนการทำนายข้อมูลที่ใช่ตรงกับข้อมูลจริงกับการทำนายว่าใช่ทั้งหมด
Recall rate	เป็นอัตราส่วนการทำนายข้อมูลที่ใช่ตรงกับข้อมูลจริงกับข้อมูลจริงที่ใช่ทั้งหมด

งานวิจัยที่เกี่ยวข้อง

เนื่องจากงานวิจัยนี้สนใจการตรวจหาท่าทางของการเตะและการชก โดยจะนำเสนอการเคลื่อนที่ที่เกิดขึ้นเป็นภาพของ Motion History Image (MHI) และ MHI จะแสดงข้อมูลของประวัติการเคลื่อนที่ของท่าทางในแต่ละเฟรมภาพไว้ในภาพเดียวกัน ดังนั้นการแสดงภาพแบบนี้จะไม่เหมาะกับการเคลื่อนที่ที่มีวิถีไปกลับและซ้ำแนวเดิมผู้วิจัยจึงหาวิธีที่จะแก้ปัญหานี้เพื่อให้สามารถประยุกต์ใช้กับ MHI และตรวจจับท่าทางที่สนใจจาก MHI ได้ จึงได้แบ่งงานวิจัยที่เกี่ยวข้องออกเป็น 2 กลุ่ม คือ งานวิจัยที่สนใจด้านการรู้จำท่าทางของบุคคล และงานวิจัยที่แก้ปัญหาการเคลื่อนที่ที่มีวิถีไปกลับและซ้ำแนวเดิมโดยใช้ MHI เป็นตัวแสดงการเคลื่อนที่

1. งานวิจัยด้านการรู้จำท่าทางของบุคคล

จากการศึกษาค้นคว้าพบว่าองค์ประกอบสำคัญในงานวิจัยด้านการรู้จำท่าทางบุคคลมี 2 องค์ประกอบหลัก คือ ชุดข้อมูล และฟีเจอร์สำหรับท่าทางของบุคคล

1.1 ชุดข้อมูล(dataset)

ชุดข้อมูลที่ใช้ในงานวิจัยนำแบ่งฉากหลัง (Background) ออกเป็น 2 แบบ คือ ฉากหลังคงที่และว่าง (Static and Clean background) และฉากหลังซับซ้อนและมีการเคลื่อนไหวของวัตถุ (Clutter and Dynamic background)

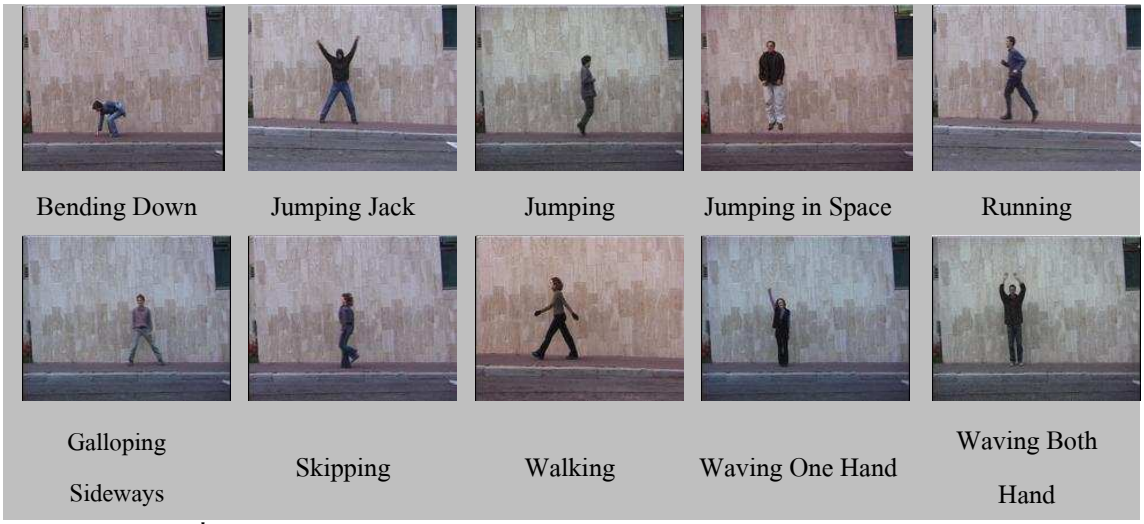
1.1.1 ชุดข้อมูลที่มีฉากหลังคงที่และว่าง ข้อมูลที่มีฉากหลังคงที่และว่างจะมีคนแสดงท่าทางโดยมีฉากหลังคงที่ คือมีส่วนประกอบของรายละเอียดฉากหลังน้อย หรือฉากสีเดียวกันเกือบทั้งหมด ตัวอย่างชุดข้อมูลที่มีฉากหลังคงที่ ได้แก่ ชุดข้อมูลของ KTH [5] ชุดข้อมูลของ Weizmann [11] และ ชุดข้อมูลของ IXMAS[12]

- *ชุดข้อมูล KTH* (KTH ย่อมาจาก Kungliga Tekniska Högskolan: Royal Institute of Technology) (2004) [5] ผู้วิจัยเก็บข้อมูลภาพจาก 25 คน คนละ 6 ท่าทาง คือ ชกมวย (Boxing) โบกมือ (Hand-waving) ปรบมือ (Hand-clapping) วิ่งช้าๆ (Jogging) วิ่ง (Running) และเดิน (Walking) โดยเก็บใน 4 สถานการณ์ ได้แก่ นอกสถานที่ (Outdoors) นอกสถานที่ที่มีขนาดของคนต่างกัน (Outdoors with scale variation) นอกสถานที่โดยเสื้อผ้าต่างกัน (Outdoors with difference clothes) และในร่ม (Indoors) ได้ภาพวีดิทัศน์ทั้งหมด 599 วีดิทัศน์ ดังตัวอย่างในรูปที่ 7

รูปที่ 7 ภาพตัวอย่างของชุดข้อมูล KTH[5] ที่ประกอบด้วยการแสดง 6 ท่าทาง ใน 4 สถานการณ์ โดยที่ s1 คือนอกสถานที่ s2 คือนอกสถานที่ที่มีขนาดของคนต่างกัน s3 คือนอกสถานที่โดยเสื้อผ้าต่างกัน และ s4 คือในร่ม

- *ชุดข้อมูล Weizmann* (Weizmann มาจาก Weizmann Institute of Science(2005)) [11] ชุดข้อมูลนี้เป็นชุดข้อมูลมาตรฐานของท่าทางบุคคล (Standard benchmark human action dataset) ที่นำมาใช้ทั่วไป ประกอบด้วย 93 วีดิทัศน์ จาก 9 คน แสดงท่าทางคนละ 10 ท่าทาง ได้แก่ ก้มตัวลง (Bending down) กระโดดตบมือเหนือศีรษะ (Jumping jack) กระโดด (Jumping) กระโดดตัวลอยอยู่กับที่ (Jumping in space) กระโดดขาเดียว (Skipping) วิ่ง (Running) กระโดดไป

ด้านข้าง (Galloping sideways) เดิน (Walking) โบกมือข้างเดียว (Waving one hand) และโบกมือสองข้าง (Waving both hands) ดังตัวอย่างในรูปที่ 8



รูปที่ 8 ภาพตัวอย่างของชุดข้อมูล Weizmann[11] ประกอบด้วยการแสดง 10 ท่าทาง

- *ชุดข้อมูล IXMAS* (INRIA Xmas Motion Acquisition Sequences(2007)) [12] เป็นชุดข้อมูลที่เก็บจากกล้อง 5 ตัว 5 มุมมอง ในสภาพแวดล้อมที่สร้างขึ้น จาก 11 คน แสดงท่าทางคนละ 15 ท่าทาง ได้แก่ ยืนเฉย (Nothing) ดูนาฬิกาข้อมือ (Check watch) กอดอก (Cross arms) เกาศีรษะ (Scratch head) นั่งลงบนพื้น (Sit down) ลุกขึ้นยืน (Get up) หมุนรอบตัว (Turn around) เดิน (Walk) โบกมือ (Wave) ชก (Punch) เตะ (Kick) ชี้นิ้ว (Point) หยิบของขึ้น (Pick up) ขว้างของข้ามศีรษะ (Throw -over head) และโยนของขึ้นข้างบน (Throw -from bottom up) ดังตัวอย่างในรูปที่ 9

-

1) Camera1 2) Camera2 3) Camera3 4) Camera4 5) Camera5

รูปที่ 9 ภาพตัวอย่างท่าทางของชุดข้อมูล IXMAS[12] โดยแสดงในมุมกล้องที่ต่างกัน

1.1.2 ชุดข้อมูลที่มีฉากหลังมีความหลากหลาย ข้อมูลที่มีฉากหลังหลากหลายจะมีรายละเอียดในฉากหลังมาก เช่น ฉากหลังเป็นร้านขายผลไม้ที่มีสีสันจากผลไม้ ฉากหลังที่มีการเคลื่อนไหวของวัตถุจะประกอบด้วยการเคลื่อนไหวทั้งจากคนและจากยานพาหนะ ตัวอย่างชุดข้อมูลที่ฉากหลังซับซ้อนและมีการเคลื่อนไหวของวัตถุ ได้แก่ ชุดข้อมูลของ CMU[13]

- *ชุดข้อมูล CMU* (CMU ย่อมาจาก Carnegie Mellon University(2007)) [13] เป็นชุดข้อมูลที่เก็บในสถานการณ์จริงที่มีการเคลื่อนที่ของคนและยานพาหนะเป็นฉากหลัง ประกอบด้วย 5 ท่าทาง ได้แก่ การโบกมือข้างเดียว (One-hand waving) โบกมือสองข้าง (Two-hand waving) การหยิบของขึ้นมา (Picking up) กดปุ่มลิฟต์ (Pushing an elevator button) และกระโดดตบมือเหนือศีรษะ (Jumping-jacks) ใน 110 เหตุการณ์ที่ฉากหลังไม่ซ้ำกันเลย ดังตัวอย่างในรูปที่ 10

(ก) (ข)

รูปที่ 10 ภาพตัวอย่างของชุดข้อมูล CMU[13]
(ก) ภาพการโบกมือข้างเดียว (ข) ภาพการก้มลงเก็บวัตถุที่อยู่บนพื้น

1.2 ฟีเจอร์สำหรับท่าทางของบุคคล

การเลือกฟีเจอร์ (Feature) สำหรับท่าทางของบุคคล (Human action) มาใช้ในการเปรียบคู่กับแม่แบบ (Template matching) มี 3 รูปแบบ ได้แก่ ฟีเจอร์รูปร่าง (Shape feature) ฟีเจอร์การเคลื่อนที่ (Motion feature) และฟีเจอร์รูปร่างและการเคลื่อนที่ประกอบกัน (Both shape and motion feature)

การสกัดฟีเจอร์เพื่อนำมาใช้ในการเปรียบคู่แม่แบบ ได้นำเอาตัวแสดงแทนฟีเจอร์ (Represent feature)ใช้แสดงแทนฟีเจอร์ที่สกัดออกมา ตัวแสดงแทนฟีเจอร์จะแสดงแทนรูปร่างของคนที่สกัดได้จากฟีเจอร์รูปร่าง และแสดงแทนบริเวณที่มีการเคลื่อนที่สำหรับฟีเจอร์การเคลื่อนที่ ตัวแสดงแทนฟีเจอร์มีหลายแบบ เช่น แบบจุด (Points) แบบกล่อง (Boxes) แบบภาพเงาร่าง (Silhouettes) แบบบล็อบ (Blobs) แบบบาร์ (Bars) เป็นต้น งานวิจัยที่เกี่ยวข้องกับฟีเจอร์ทั้ง 3 รูปแบบมีดังนี้

1.2.1 ฟีเจอร์รูปร่าง

ฟีเจอร์รูปร่างเป็นการนำท่าทางของคนสกัดเป็นโครงร่าง และใช้ภาพโครงร่างสำหรับการเปรียบคู่แม่แบบ

ปี 1996 Kuno และ Watanabe [14] เสนอการใช้ภาพเงาร่างในการรู้จำท่าทางบุคคล และแปลงให้อยู่ในรูปแบบ project histogram เพื่อใช้ในการเปรียบคู่ ประกอบด้วยท่าทาง 3 แบบ

คือ การเดินผ่านหน้า (Crossing) การเดินเข้าใกล้ (Approaching) และการเดินเข้าใกล้แบบเฉียง (Obliquely approaching) และเก็บข้อมูลในฉากหลังคงที่ที่สร้างเอง แต่ข้อจำกัดของงานนี้คือต้องควบคุมสภาพแวดล้อมโดยเฉพาะแสง เพราะถ้ามีแสงไม่เพียงพอค่าความต่างของฉากหลังกับวัตถุอาจจะน้อย ทำให้เมื่อสกัดภาพวัตถุออกมาอาจมีข้อมูลรบกวนในภาพมากเกินไป

ต่อมาในปี 2006 Dedeo□lu และคณะ [15] พัฒนาการใช้ภาพเงาร่างเป็นตัวแสดงแทนฟีเจอร์ ในฉากหลังคงที่ที่ใช้แสงจากธรรมชาติ โดยสกัดจากการหักล้างฉากหลังในภาพออก (Background subtraction) และเปรียบคู่โดยใช้ภาพเงาร่าง ข้อมูลที่ใช้เป็นข้อมูลของท่าทางทั้งหมด 3 แบบ คือ การเดิน (Talking) การชก (Boxing) และการเตะ (Kicking) ได้ผลความถูกต้องในการรู้จำท่าทางดีมากเมื่อใช้ข้อมูลในมุมมองปกติ ถ้าเปลี่ยนมุมมองของกล้องผลความถูกต้องจะลดลง

ปี 2007 Lv และ Nevatia [16]ได้พัฒนาการเอาภาพเงาร่างที่ใช้เป็นตัวแสดงแทนฟีเจอร์แปลงเป็นภาพสามมิติเพื่อช่วยในการรู้จำท่าทาง เมื่อสกัดได้ภาพเงาร่างก็นำมาพิจารณาสร้างเป็นรูปร่างแบบสามมิติโดยใช้ key pose ใช้เทคนิค Action net ช่วยหาภาพสามมิติของภาพเงาร่างในเฟรมถัดไป และเปรียบคู่โดยใช้ key pose นี้ ชุดข้อมูลที่นำมาใช้คือ IXMAS[12]

นอกจากการใช้ภาพเงาร่างแล้วยังมีงานวิจัยของ Niebles และ Lee [17] ในปี 2007 ใช้จุดเป็นตัวแสดงแทนฟีเจอร์รูปร่าง โดยมีการระบุตำแหน่งของจุดบนภาพขอบ (Edge image) จับกลุ่มจุดที่สัมพันธ์กันด้วยวงรี และเชื่อมวงรีทั้งหมดเข้าด้วยกัน จากนั้นจึงใช้ Hierarchical model ช่วยในการแยกประเภทท่าทางบุคคล และชุดข้อมูลที่ใช้ในการทดสอบคือ Weizmann[11]

งานวิจัยที่ใช้ฟีเจอร์รูปร่าง ส่วนใหญ่จะทำงานได้ดีเมื่อภาพฉากหลังมีความคงที่หรือไม่ซับซ้อน และหากมีท่าทางของบุคคลคล้ายกันอาจเกิดความผิดพลาดในการรู้จำท่าทางและแยกประเภทท่าทางได้

1.2.2 ฟีเจอร์การเคลื่อนที่

ฟีเจอร์การเคลื่อนที่จะพิจารณาจากภาพที่ต่อเนื่องกัน เพื่อดูว่าบริเวณใดมีการเคลื่อนที่และสกัดออกมา อาจอยู่ในรูปของทิศทางหรือขนาดของการเคลื่อนที่ในบริเวณที่มีการเคลื่อนที่นั้น

ในปี 2009 Yang และคณะ [18] ได้ใช้ Patch-based motion feature แสดงบริเวณที่มีการเคลื่อนที่ของร่างกายในการรู้จำท่าทางของบุคคล โดยที่แต่ละ patch จะมีน้ำหนักในการคำนวณเพื่อใช้ในการเปรียบคู่ตามการเคลื่อนที่ที่เกิดขึ้นใน patch นั้น และใช้ Transferable distance function มาช่วยในการสอนระบบเพื่อการคำนวณค่าถ่วงน้ำหนักให้ได้ค่าที่เหมาะสม ชุดข้อมูลที่นำมาทดสอบ ได้แก่ KTH [5] Weizmann [11] และ CMU[13]

การใช้ optical flow เป็นตัวแสดงแทนฟีเจอร์การเคลื่อนที่ เป็นวิธีที่นิยมใช้มาก Castellanos และคณะ [19] ได้ทำงานวิจัยในปี 2010 โดยนำ optical flow ใช้ในการระบุขนาด

และ ทิศทางการเคลื่อนที่ของท่าทาง โดยมีเหตุการณ์ที่สนใจ 2 แบบคือ การเคลื่อนที่ในทิศทางตรงกันข้าม (Opposing Flow) และการวิ่ง (Person Running) ชุดข้อมูลที่ใช้คือ TRECVID (2008) [20] เป็นข้อมูลที่เก็บในเหตุการณ์จริงในสนามบิน

งานวิจัยที่ใช้ฟีเจอร์การเคลื่อนที่ ถ้าฉากหลังมีการเคลื่อนที่อาจเกิดความผิดพลาดในการรู้จำท่าทางได้ เนื่องจากมีข้อมูลรบกวนมากเกินไป

1.2.3 ฟีเจอร์รูปร่างและการเคลื่อนที่ประกอบกัน

บางกรณีการใช้ฟีเจอร์ใดฟีเจอร์หนึ่งเพียงอย่างเดียวอาจไม่เพียงพอต่อการรู้จำท่าทางบุคคล เช่น การใช้ฟีเจอร์รูปร่างในฉากหลังที่ซับซ้อน อาจได้ภาพของรูปร่างไม่สมบูรณ์หรือรูปร่างผิดเพี้ยนทำให้ผลความถูกต้องลดลง หรือการใช้ฟีเจอร์การเคลื่อนที่ในฉากหลังที่มีการเคลื่อนไหวของวัตถุ เมื่อสกัดการเคลื่อนที่ออกมาอาจมีข้อมูลรบกวนจากฉากหลัง ทำให้การคำนวณทิศทางและขนาดผิดพลาดได้เหมือนกัน ดังนั้นจึงนำเอาทั้งสองฟีเจอร์มาประกอบกันเพื่อแก้ปัญหาและเพิ่มความถูกต้องในการรู้จำท่าทางบุคคลมากขึ้น

ในปี 2007 Ke และคณะ [13] ใช้ภาพเงาร่างเป็นตัวแสดงแทนฟีเจอร์รูปร่างและ flow เป็นตัวแสดงแทนฟีเจอร์การเคลื่อนที่ สร้างเป็นแบบจำลอง (Model) ในการเปรียบคู่แม่แบบจะใช้แบบจำลองที่แบ่งออกเป็นส่วนๆ (Parts-based models) แล้วนำมาเปรียบคู่ และใช้แบบจำลองทั้งตัว (Whole-based models) แล้วนำมาเปรียบคู่ พร้อมทั้งเปรียบเทียบผลจากการเปรียบคู่ของแบบจำลองทั้งสอง โดยชุดข้อมูลคือ CMU[13]

ต่อมาในปี 2010 Yao และ Zhu [21] ได้ใช้ชุดข้อมูล KTH [5] และ CMU [13] ในการทดสอบงานวิจัย โดยเลือกตัวแสดงแทนฟีเจอร์รูปร่างเป็นแบบบาร์ และใช้ตัวแสดงแทนฟีเจอร์การเคลื่อนที่เป็นแบบกล่อง ได้ shape และ flow primitives การสอนระบบจะเลือกท่าทางที่เหมือนกันจากวิดีโออื่นๆ ด้วย dynamic space-time warping algorithm และปรับข้อมูลไปเรื่อยๆ ด้วย shared pursuit algorithm ครอบคลุมทุกวิดีโอที่ใช้สอนระบบ จากนั้นก็เปรียบคู่แม่แบบโดยใช้ primitives ที่เก็บไว้

นอกจาก [13] [21] ในปี 2010 Guo และคณะ [22] ใช้ภาพเงาร่างที่สกัดจาก mask-based shape method เป็นตัวแสดงแทนฟีเจอร์รูปร่าง และใช้ flow ที่ได้จาก optical flow เป็นตัวแสดงแทนฟีเจอร์การเคลื่อนที่ โดยเก็บฟีเจอร์ทั้งสองไว้ในรูปของ binary action template เพื่อใช้ในการเปรียบคู่แม่แบบ โดยชุดข้อมูลที่ใช้ทดสอบ ได้แก่ KTH[11] Weizmann[12] และ CMU[14]

งานวิจัยที่ใช้ทั้งสองฟีเจอร์รวมกันให้ความถูกต้องในการรู้จำท่าทางบุคคลค่อนข้างสูง และประยุกต์ใช้กับภาพฉากหลังได้ทั้งสองแบบ แต่ใช้ทรัพยากรและเวลาในการคำนวณมากกว่าการใช้ฟีเจอร์เดียว

2. งานวิจัยที่แก้ปัญหาการเคลื่อนที่ที่มีวิถีไปกลับและซ้ำแนวเดิมโดยใช้ MHI เป็นตัวแสดงการเคลื่อนที่

ในปี 2004 Valstar และคณะ [24, 25, 26] ได้ใช้ MHI กับการแสดงออกของใบหน้า (Facial Action) และเสนอวิธีการแก้ปัญหาด้วยการใช้ Multiple-level MHI (MMHI) โดยให้ *'จำนวนของลำดับเหตุการณ์' (History levels)* ซึ่งเป็นตัวระบุค่าของ *'ประวัติการเกิดเหตุการณ์' (History)* ให้กับแต่ละตำแหน่งการเคลื่อนที่ในภาพ โดยที่ n คือ จำนวนของลำดับเหตุการณ์ และจำนวนของเฟรมภาพเป็น n+1 จากนั้นใช้ s เป็นตัวกำหนดค่าความต่างของความเข้มแสงระหว่างเหตุการณ์แต่ละลำดับที่ต่อเนื่องกัน (ตัวอย่างเช่น กำหนดให้ s = 255/n เป็นต้น) ซึ่งจะเป็นการระบุค่าความเข้มแสงให้กับแต่ละลำดับเฟรม แต่ผลที่ได้ไม่ค่อยดีนักและให้ค่าความถูกต้องน้อยกว่าการใช้ MHI อย่างเดียว [26]

ต่อมาในปี 2008 Ahad และคณะ [27] ได้ใช้ MHI กับท่าทางการเคลื่อนที่ของมนุษย์ โดยนำ Directional MHI (DMHI) มาแก้ปัญหาที่วิถีการเคลื่อนที่ซ้ำแนวเดิมนี้ โดยจะพยายามนำทิศทางในแกนต่างๆ ของการเคลื่อนที่มาประยุกต์ใช้ ซึ่งวิธีที่นำมาใช้คือ Gradient-based Optical Flow ที่จะแยกทิศทางการเคลื่อนที่ที่ได้จากภาพเฟรมต่อเฟรมออกเป็น 4 ทิศ คือ แนวตั้ง (ขึ้นและลง $\pm y$) กับแนวนอน(ซ้ายและขวา $\pm x$) วิธีนี้สามารถแก้ปัญหาการเคลื่อนที่ไปกลับทับแนววิถีเดิมได้ดี แต่ข้อมูลที่แยกเป็น 4 แบบจะมีขนาดใหญ่และใช้เวลาในการคำนวณเพิ่มมากขึ้น

ในปีเดียวกัน Kellokumpu และคณะ [28] เลือกที่จะใช้ Local Binary Pattern (LBP) histogram ที่ได้จาก MHI และ MEI สำหรับสร้างโมเดลฟีเจอร์ของท่าทางในการแก้ปัญหา หาได้จากการแบ่ง MHI และ MEI ออกเป็น 4 ส่วน โดยยึดเอาจุดศูนย์กลางของโครงร่างเป็นตัวแบ่ง และ LBP จะสร้างฮิสโตแกรมของแต่ละส่วน จากนั้นจึงใช้ HMMs (Hidden Markov Model) เป็นตัวสร้างโมเดล ซึ่งวิธีนี้จะแก้ปัญหาวิถีการเคลื่อนที่ซ้ำแนวเดิมได้ดี แต่การแบ่งส่วนข้อมูลแบบนี้อาจไม่เหมาะสำหรับทุกท่าทางก็ได้ [4]

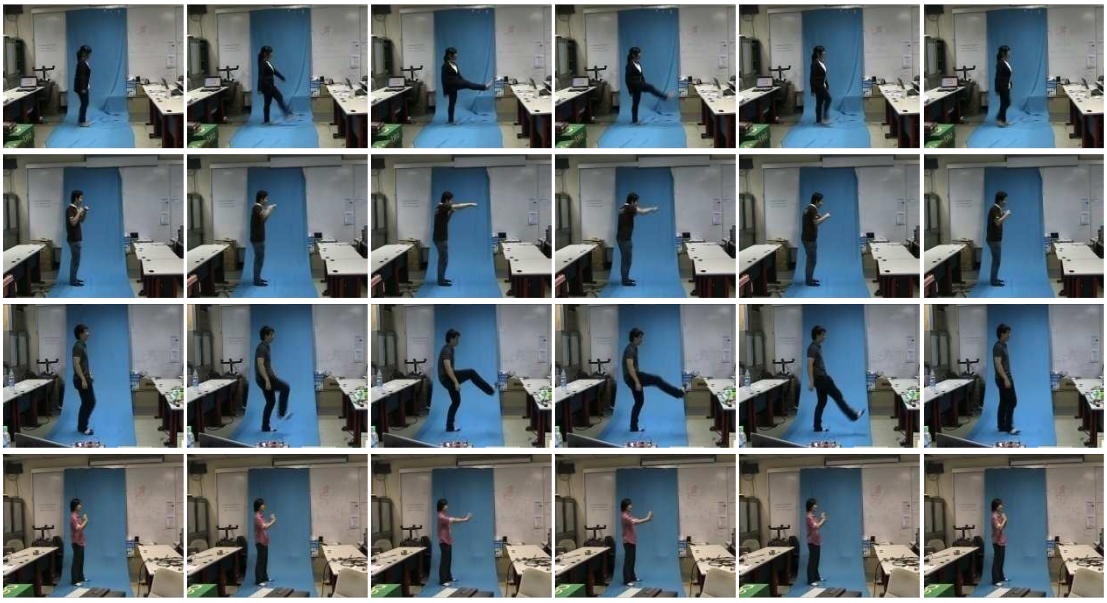
จากงานวิจัยที่กล่าวมาข้างต้นจะพบว่า งานเหล่านี้สนใจแก้ปัญหาการเคลื่อนที่ที่มีวิถีซ้ำแนวเดิมด้วยการปรับปรุง MHI [24, 25, 26, 27] ซึ่งมีทั้งผลดีและผลเสีย ส่วนงานของ Kellokumpu และคณะ [28] เป็นการแก้ไขโดยเปลี่ยนตัวแทนฟีเจอร์จาก MHI เป็นลักษณะอื่น อย่างไรก็ตามเนื่องจากข้อดีของ MHI ที่มีคุณสมบัติในการคำนวณง่ายและสามารถใช้ได้ดีกับภาพที่มีความละเอียดต่ำ ผู้วิจัยจึงเลือกที่จะปรับวิธีการใช้งาน MHI เพื่อให้เหมาะกับการเคลื่อนของมนุษย์ที่มีวิถีซ้ำแนวเดิมโดยการแบ่งชุดภาพวีดิทัศน์เป็นสองส่วนที่แยกการเคลื่อนที่ที่มีแนวทางซ้อนทับกันออกจากกัน

บทที่ 3

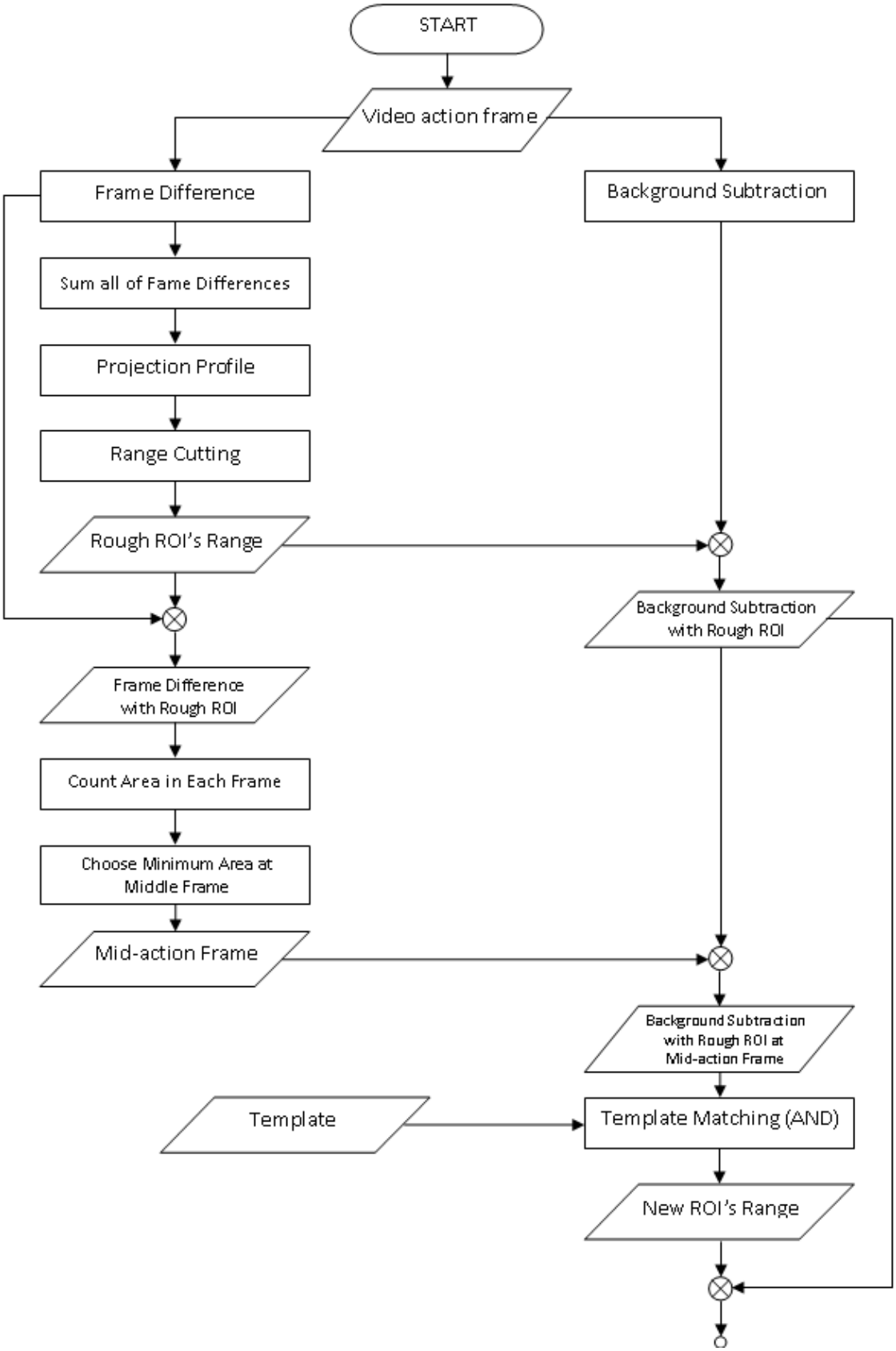
การรู้จำท่าทางการเตะและการชก

งานวิจัยนี้ต้องการที่จะศึกษาเกี่ยวกับการรู้จำท่าทางการเคลื่อนที่ของบุคคล โดยท่าทางที่สนใจคือ การเตะและการชก เป็นท่าทางที่มีลักษณะการเคลื่อนที่เป็นวิถีที่มีการเคลื่อนที่ไปกลับแบบซ้ำเส้นทางเดิม จากภาพวีดิทัศน์ ซึ่งเก็บท่าทางมาทั้งหมด 4 ท่าทาง ประกอบด้วย การเตะ การชก การถีบ และการผลัก โดยนำภาพวีดิทัศน์ที่ได้มาสกัดเป็นภาพนิ่งโดยมีอัตราเฟรม (Frame rate) 15 เฟรมต่อวินาที พร้อมทั้งแยกข้อมูลออกจากกันเป็นชุดภาพของท่าทาง ซึ่ง 1 ชุดภาพจะประกอบด้วยคนแสดง 1 คน แสดง 1 ท่าทางต่อ 1 ครั้ง แล้วนำมาเข้าขั้นตอนต่างๆ ในกระบวนการเพื่อรู้จำท่าทางที่สนใจ ขั้นตอนหลักๆ ของการดำเนินงานจะประกอบด้วย 3 ขั้นตอน คือ การหาบริเวณที่เกิดการเคลื่อนที่ การแสดงการเคลื่อนที่ในรูปแบบ MHI และการรู้จำและแยกประเภทท่าทาง โดยจะแสดงแผนผังลำดับงานตามรูปที่ 12

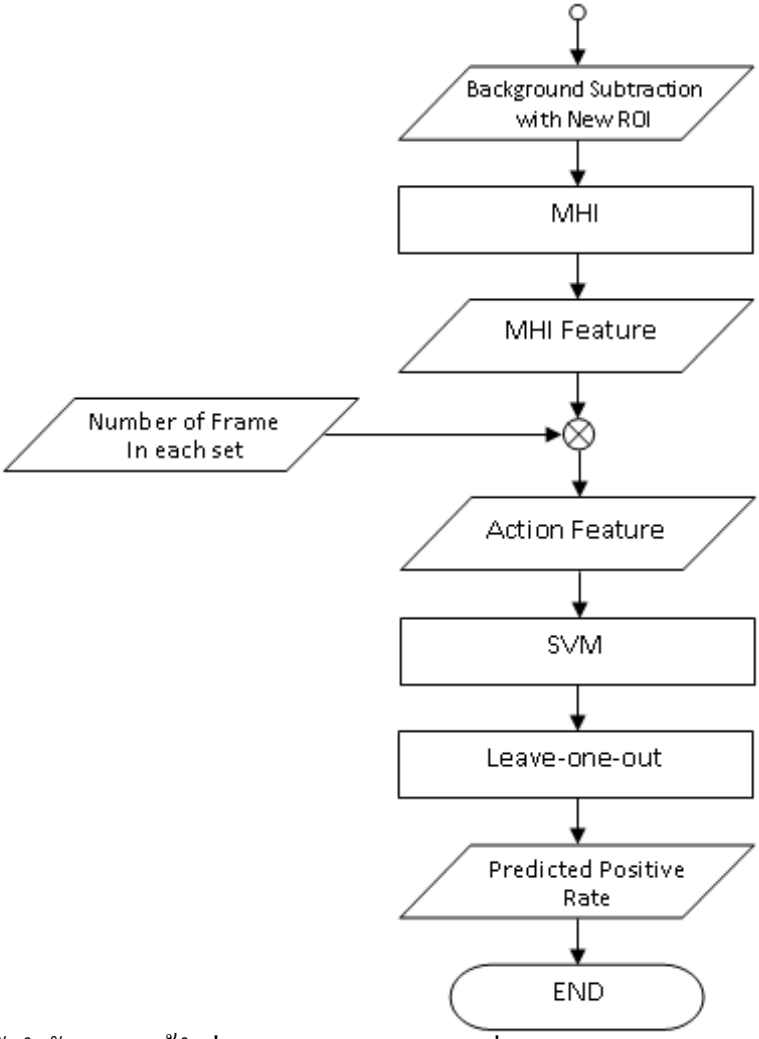
ข้อมูลที่เก็บมาทั้งหมดจะประกอบด้วย 4 ท่าทาง คือ การเตะ การชก การถีบและการผลัก โดยมีนักแสดงทั้งหมด 11 คน แสดงทั้งด้านซ้ายและขวา ในสถานการณ์ต่างกันทั้งฉากหลังที่คงที่และฉากหลังที่มีความหลากหลาย ตัวอย่างของข้อมูลแสดงในรูปที่ 11 และรูปที่ 13



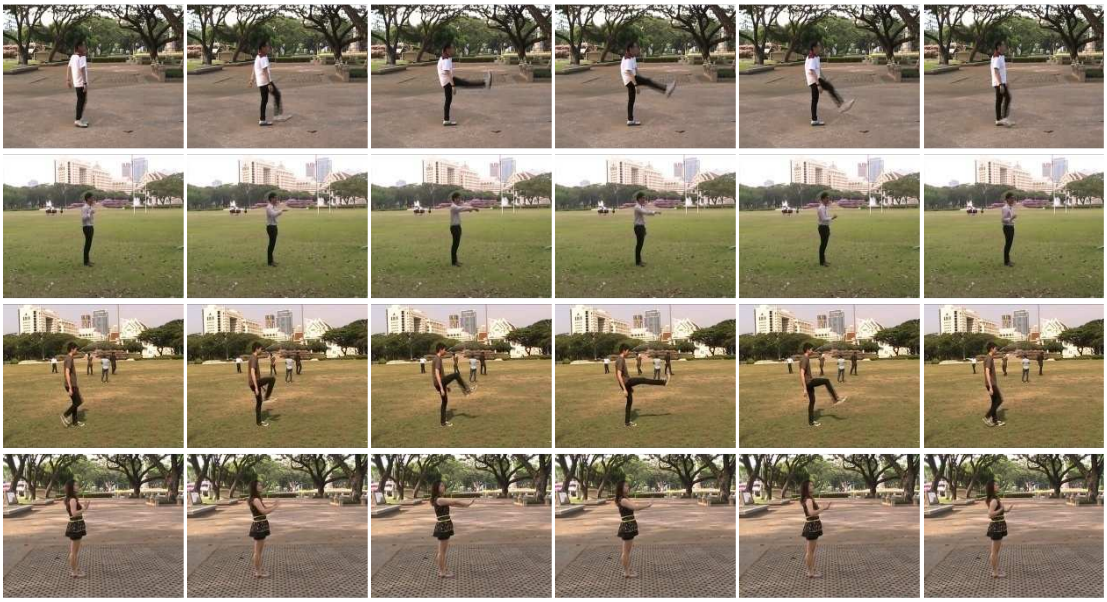
รูปที่ 11 ตัวอย่างข้อมูลภาพวีดิทัศน์ที่แสดงท่าทางทั้ง 4 ท่าทาง ในฉากหลังคงที่



รูปที่ 12 แผนผังลำดับงานการรู้จำท่าทางการเตะและการชก



รูปที่ 12 แผนผังลำดับงานการรู้จำท่าทางการเตะและการชก (ต่อ)



รูปที่ 13 ตัวอย่างข้อมูลภาพวีดิทัศน์ที่แสดงท่าทางทั้ง 4 ท่าทาง ในฉากหลังที่มีความหลากหลาย

1. การหาบริเวณที่เกิดการเคลื่อนที่ (Human's Gesture Detection)

เนื่องจากท่าทางที่สนใจมีลักษณะการเคลื่อนที่เฉพาะบางส่วนของร่างกายและเป็นการแสดงท่าทางที่มีจุดยืนไม่เปลี่ยนตำแหน่ง ดังนั้นจึงสามารถหาบริเวณที่เกิดการเคลื่อนที่โดยที่ขอบเขตของบริเวณที่สนใจ (Region of Interest: ROI) เป็นบริเวณเดียวกันทั้งชุดภาพได้ และการหาบริเวณที่สนใจนี้แบ่งออกเป็น 3 ส่วน คือ การหาบริเวณการเคลื่อนที่อย่างคร่าวๆ การหาเฟรมกึ่งกลางของชุดภาพการเคลื่อนที่ และการหาบริเวณการเคลื่อนที่จากการเปรียบคู่แม่แบบ

1.1 การหาบริเวณการเคลื่อนที่อย่างคร่าวๆ (Rough ROI's Detection)

เนื่องจากชุดภาพของการเคลื่อนที่ของท่าทางที่ได้จากภาพวีดิทัศน์ เป็นการเก็บข้อมูลจากภาพพื้นหลังที่มีทั้งแบบไม่คงที่วุ่นวายและมีความซับซ้อนบ้างเล็กน้อย การหาบริเวณการเคลื่อนที่จึงได้มีการหาอย่างคร่าวๆ ก่อนว่าอยู่บริเวณไหน แล้วค่อยเจาะจงเลือกเฉพาะบริเวณที่เป็นการเคลื่อนที่ของอวัยวะส่วนใดส่วนหนึ่ง และการหาอย่างคร่าวๆนี้ ประกอบด้วย

1.1.1 การหาความแตกต่างระหว่างเฟรม (Frame Difference)

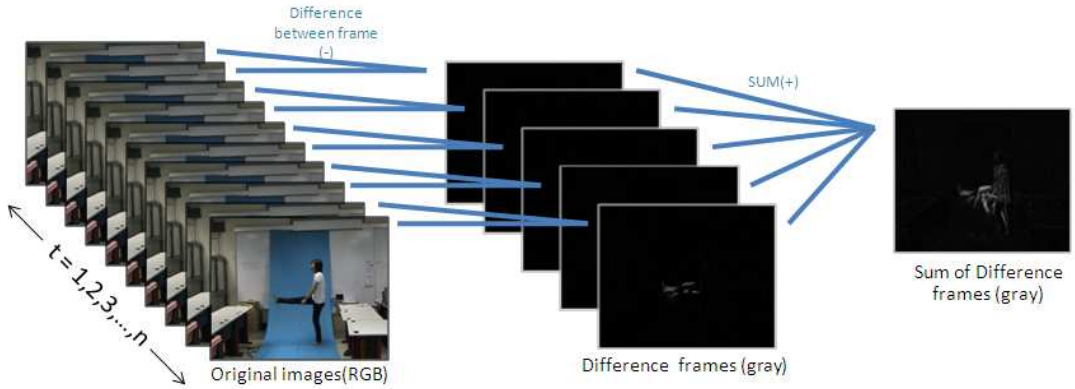
ใน 1 ชุดภาพ จะมี 1 ท่าทางของ 1 คน ซึ่งเป็นชุดภาพที่มีภาพการเคลื่อนที่ของท่าทางที่มีความต่อเนื่องกัน เมื่อพิจารณาแล้วจะพบว่าบริเวณที่เกิดการเคลื่อนที่จะมีการเปลี่ยนตำแหน่งของอวัยวะที่แสดงท่าทางแบบต่อเนื่อง ค่าความเข้มแสงที่จุดภาพที่มีการเคลื่อนที่ผ่านก็จะเปลี่ยนไปเรื่อยๆ ตามบริเวณที่อวัยวะเปลี่ยนตำแหน่งไป ผู้วิจัยจึงเลือกใช้การหาความแตกต่างระหว่างเฟรมเป็นวิธีการสกัดบริเวณที่เกิดการเคลื่อนที่อย่างคร่าวๆ เพราะความแตกต่างของความเข้มแสงที่เกิดขึ้นระหว่างเฟรมจะเป็นตัวแสดงบริเวณที่มีการขยับจากเฟรมก่อนหน้ามายังเฟรมปัจจุบัน บริเวณที่มีการเคลื่อนที่ผ่านจะมีความแตกต่างมากๆ ส่วนบริเวณที่เคลื่อนที่น้อยหรือไม่มีการเคลื่อนที่จะมีความแตกต่างน้อยมาก การคำนวณหาความแตกต่างระหว่างเฟรมดูได้จากสมการ (2) และตัวอย่างภาพที่ได้จากการหาความแตกต่างระหว่างเฟรมแสดงตามรูปที่ 14

(ก) (ข) (ค)

รูปที่ 14 ตัวอย่างภาพความแตกต่างระหว่างเฟรม(ก) ภาพเฟรมก่อนหน้า (ข) ภาพเฟรมปัจจุบัน (ค) ผลของการหาความแตกต่างระหว่างเฟรม

1.1.2 การรวมผลที่ได้จากความแตกต่างระหว่างเฟรม (Summation All of Frame Difference)

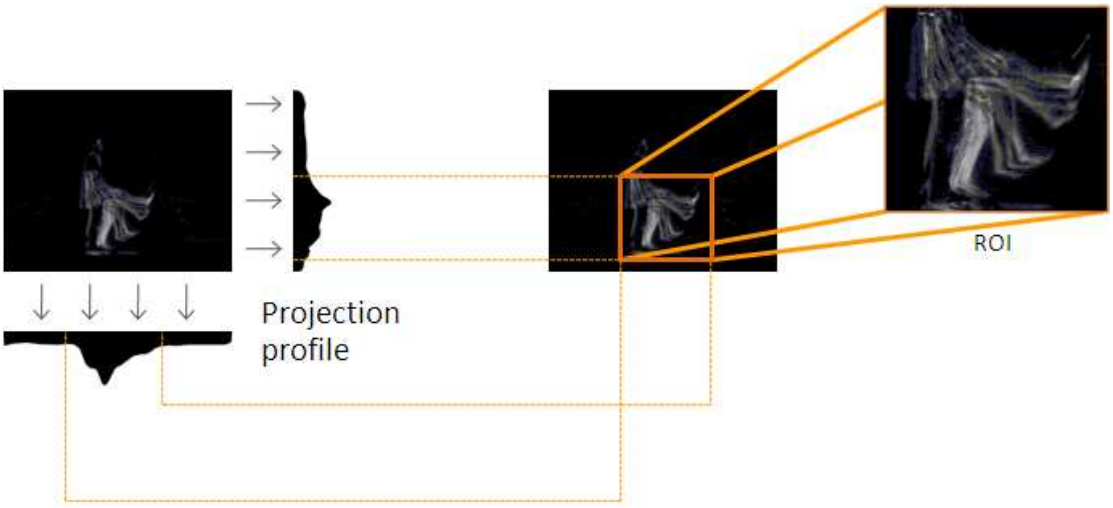
เมื่อพิจารณาจากภาพผลของการหาความแตกต่างระหว่างเฟรมจะพบว่าเป็นค่าความเข้มแสงบริเวณที่เกิดการเคลื่อนที่ผ่านจะมีความเข้มมากกว่าบริเวณอื่นในภาพเดียวกัน และถ้าพิจารณาที่ความต่อเนื่องของภาพผลต่าง บริเวณที่มีค่าความแตกต่างที่ให้ความเข้มแสงมากๆจะมีความต่อเนื่องกันตามลำดับการเคลื่อนที่และครอบคลุมบริเวณที่เกิดการเคลื่อนที่ทั้งหมดในชุดภาพนั้น ดังนั้นจึงนำภาพผลของการหาความแตกต่างระหว่างเฟรมที่ได้ทั้งหมดจากชุดภาพนั้นรวมกัน ซึ่งจะแสดงบริเวณที่เกิดการเคลื่อนที่ทั้งหมดในชุดภาพนั้นไว้ในภาพเพียงภาพเดียว แสดงดังรูปที่ 15



รูปที่ 15 ผลรวมของภาพความแตกต่างระหว่างเฟรม

1.1.3 การฉายเงาภาพ (Projection Profile)

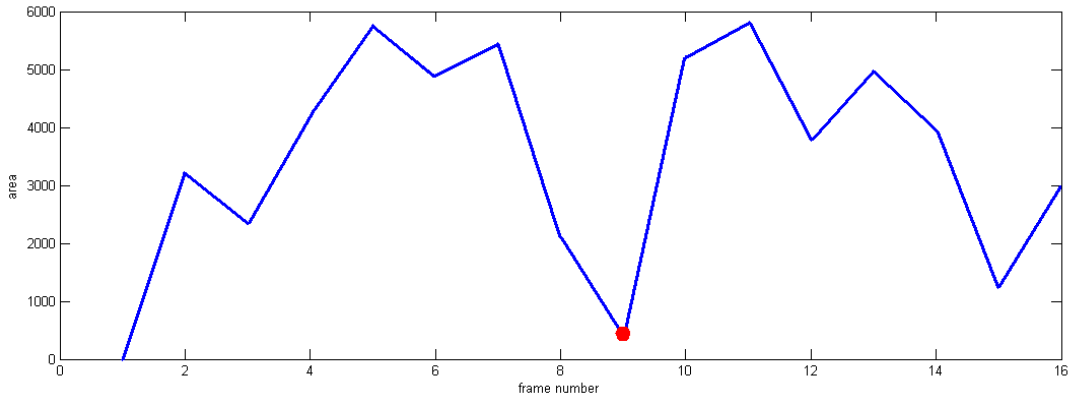
จากรูปที่ 15 จะพบว่าภาพผลรวมของความแตกต่างระหว่างเฟรมจะมีความเข้มแสงบริเวณที่มีการเคลื่อนที่มากกว่าบริเวณอื่น จึงใช้การฉายเงาภาพช่วยในการพิจารณาว่าบริเวณไหนน่าจะเป็นการเคลื่อนที่ เพราะการฉายภาพจะเป็นการเก็บค่าความเข้มแสงในจุดภาพสะสมตามแนวแกนทั้งแกนตั้งและแกนนอน ซึ่งบริเวณที่เป็นการเคลื่อนที่ ค่าสะสมความเข้มแสงในบริเวณนั้นจะมีมากกว่าบริเวณอื่นเมื่อพิจารณาตามแนวแกนทั้งสองแกน และใช้ลักษณะที่เกิดขึ้นนี้เป็นตัวกำหนดขอบเขตคร่าวๆ ของบริเวณที่สนใจ การฉายภาพและเลือกขอบเขตของบริเวณที่สนใจแสดงตามรูปที่ 16



รูปที่ 16 ตัวอย่างการฉายภาพและการเลือกขอบเขตของบริเวณที่สนใจคร่าวๆ

1.2 การหาเฟรมกึ่งกลางของชุดภาพการเคลื่อนที่ (Mid-action Frame Detection)

เนื่องจากท่าทางการเคลื่อนที่ที่นำมาศึกษาเป็นการเคลื่อนที่ที่มีวิถีไปกลับแบบซ้ำเส้นทางเดิม แต่ละชุดภาพจะมีเฟรมหนึ่งเฟรมที่เป็นตำแหน่งที่อวัยวะเคลื่อนที่มาหยุดและเปลี่ยนทิศจากเดิมที่ออกจากจุดเริ่มต้นเป็นกลับเข้าหาจุดเริ่มต้น ดังนั้นตำแหน่งนั้นอวัยวะจะเคลื่อนที่ออกจากร่างกายมากที่สุด ขนานกับพื้นมากที่สุด และครอบคลุมบริเวณการเคลื่อนที่ในแนวแกนนอนมากที่สุดด้วย การหาเฟรมกึ่งกลางของชุดภาพการเคลื่อนที่เป็นส่วนที่สำคัญ เนื่องจากอันดับแรกการหาบริเวณที่สนใจจากการเปรียบคู่แม่แบบ จะเป็นการเปรียบเทียบแม่แบบที่เป็นสี่เหลี่ยมจัตุรัสกับภาพที่เฟรมกึ่งกลางของจุดภาพ (ภาพที่อวัยวะขนานกับพื้นที่สุด) เพื่อตัดเอาเฉพาะส่วนที่เป็นอวัยวะที่เคลื่อนที่เท่านั้น และอันดับสองคือตำแหน่งที่เฟรมกึ่งกลางของชุดภาพนี้จะเป็นส่วนที่ช่วยแก้ปัญหาการแสดงภาพที่เป็น MHI เนื่องจากจะเกิดการทับซ้อนของข้อมูล



รูปที่ 17 กราฟพื้นที่ของแต่ละเฟรมและเฟรมกึ่งกลางของการเคลื่อนที่ที่ถูกเลือก

1.3 การหาบริเวณการเคลื่อนที่จากการเปรียบคู่แม่แบบ (ROI's from Template Matching)

จากหัวข้อ 1.1 จะพบว่าบริเวณการเคลื่อนที่ที่ได้จะเป็นบริเวณอย่างคร่าวๆ ซึ่งเมื่อนำไปทำเป็น MHI อาจจะมีข้อมูลที่เป็นส่วนอื่นที่ไม่ใช่อวัยวะที่เคลื่อนที่ทำให้ข้อมูลคลาดเคลื่อน นอกจากนี้ขนาดของบริเวณที่สนใจในแต่ละชุดภาพจะไม่เท่ากัน แต่ภาพที่จะนำมาสอนระบบการรู้จำจำเป็นที่จะต้องมีขนาดเท่ากัน ดังนั้นในส่วนนี้จึงต้องการที่จะตัดบริเวณที่มีการเคลื่อนที่เฉพาะส่วนที่เป็นอวัยวะที่เคลื่อนที่เท่านั้นและตัดบริเวณการเคลื่อนที่ใหม่ให้มีขนาดเท่ากัน ขั้นตอนการทำงานมีดังนี้

1) หาขอบเขตตามแนวแกนนอนที่คาดว่าเป็นขอบเขตของอวัยวะที่แสดงท่าทาง โดยการเลือกจุดภาพสีขาวจุดแรกที่เจอ จากทั้งซ้ายและขวา

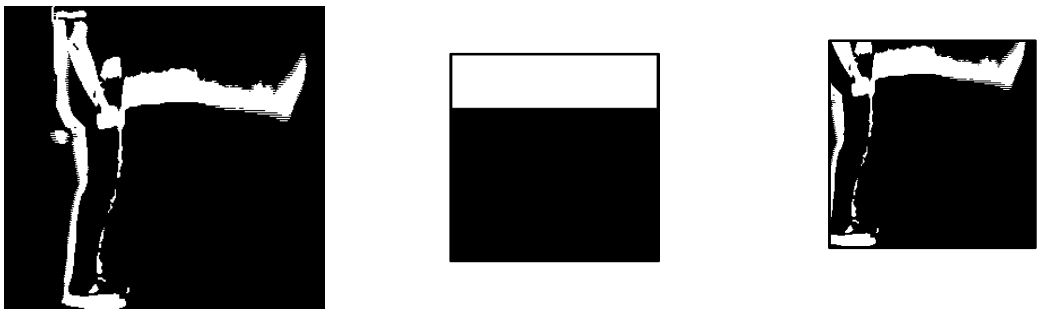
2) สร้างแม่แบบที่มีลักษณะตามรูปที่ 17(ข) ที่เป็นสี่เหลี่ยมจัตุรัส ตามขนาดของขอบเขตจากข้อ 1)

3) นำภาพที่ได้จากการทำการหักลบพื้นหลังที่ตำแหน่งเฟรมกึ่งกลาง มาเปรียบคู่กับแม่แบบโดยการ AND กัน

4) เลือกตำแหน่งที่ได้ผลจากการ AND มากที่สุด

5) ตัดขอบเขตของทุกเฟรมในชุดภาพตามขนาดของแม่แบบที่ตำแหน่งนั้น จะได้บริเวณการเคลื่อนที่ใหม่ที่เป็นสี่เหลี่ยมจัตุรัสสำหรับทุกเฟรมในชุดภาพเดียวกัน

6) ขนาดของเฟรมในแต่ละชุดภาพไม่เท่ากัน จึงนำขนาดบริเวณการเคลื่อนที่ใหม่ของทุกชุดภาพมาเฉลี่ยกันและปรับทุกชุดภาพให้มีขนาดเท่ากันทั้งหมด โดยใช้วิธีการประมาณค่าแบบไบคิวบิก (Bicubic Interpolation)



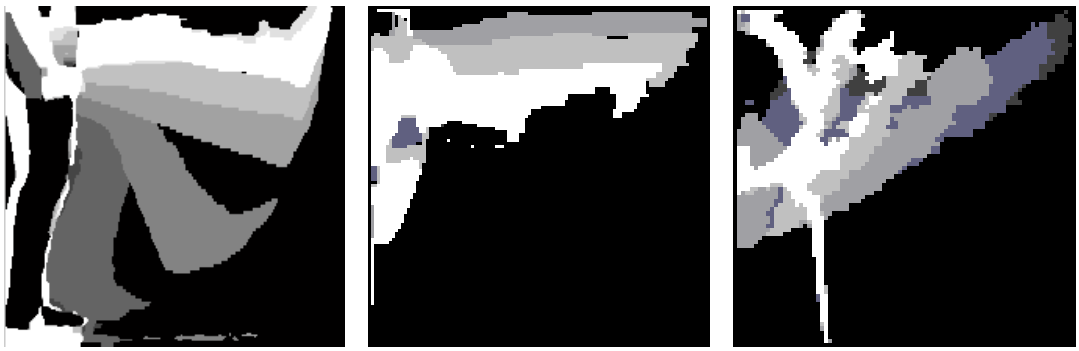
(ก) (ข) (ค)

รูปที่ 18 การหาบริเวณที่สนใจจากการเปรียบคู่แม่แบบ
(ก) ภาพจากการหักลบพื้นหลังขนาดเท่ากับบริเวณการเคลื่อนที่คร่าวๆ ที่ตำแหน่งเฟรมกึ่งกลางของการเคลื่อนที่
(ข) แม่แบบ (template)
(ค) ภาพที่ได้จากการตัดบริเวณที่สนใจใหม่โดยมีขนาดเท่ากับแม่แบบ

2. การแสดงการเคลื่อนที่ในรูปแบบ MHI (MHI Feature Extraction)

การสร้าง MHI สามารถทำได้จากสมการ (5-7) โดยที่สมการ (5) เป็นสมการเริ่มต้นในการสกัดส่วนที่เคลื่อนที่ออกจากภาพพื้นหลังด้วยวิธีการหักลบภาพออกจากพื้นหลัง ซึ่งกำหนดให้ภาพแรกของทุกชุดภาพเป็นภาพพื้นหลังเนื่องจากเป็นการเริ่มต้นของท่าทางต่างๆ โดยส่วนใหญ่จะเริ่มจากการยืนตรง จึงเลือกให้ภาพแรกของทุกชุดภาพเป็นภาพพื้นหลัง

เมื่อได้ผลจากการหักลบแล้วนั้นก็จะเข้าสมการ (6) เพื่อสร้างฟังก์ชันที่ใช้ปรับค่า (Update Function) โดยค่าขีดแบ่งจะหาจากการใช้วิธีการของโอทสึ (Otsu's Method) ภาพที่ได้จะเป็นภาพขาว-ดำ และนำฟังก์ชันที่ใช้ปรับค่าไปปรับค่าใน MHI ตามสมการ (7) ซึ่งจะเข้ามาตามลำดับการเคลื่อนที่ที่ได้จากชุดภาพ โดยภาพขาว-ดำที่เข้ามาล่าสุดจะสำคัญที่สุด ตำแหน่งที่เป็นสีขาวจะแสดงเป็นค่าความเข้มแสงที่สว่างที่สุดใน MHI ส่วนตำแหน่งอื่นๆ ในภาพขาว-ดำล่าสุดที่เป็นสีดำ ณ ตำแหน่งเดียวกันที่ภาพ MHI จะถูกลดค่าลงตามตัวแปร δ เพื่อใช้ค่าความเข้มแสงเป็นตัวบอกลำดับการเคลื่อนที่ที่อยู่ใน MHI สว่างสุดคือล่าสุดและค่าความเข้มที่ลดหลั่นกันคือลำดับก่อนหน้า ซึ่ง MHI จะปรับไปเรื่อยๆ ตามลำดับของฟังก์ชันที่เข้ามาจนกว่าไม่มีการเปลี่ยนแปลงที่ฟังก์ชันปรับค่า ตัวอย่างของการใช้ MHI แสดงตามรูปที่ 19



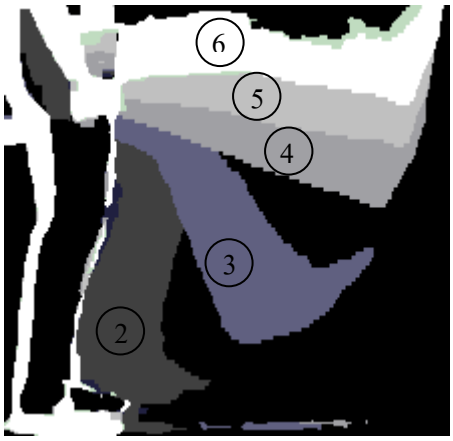
รูปที่ 19 ตัวอย่างภาพ MHI ของท่าทางต่างๆ

เนื่องจากภาพวีดิทัศน์ที่มีเป็นการเคลื่อนที่ที่มีวิถีแบบไปกลับและซ้ำเส้นทางเดิม ดังนั้นถ้านำมาสร้าง MHI จะเกิดการซ้อนทับกันของการเคลื่อนที่ที่เกิดก่อนหน้าถ้าวิถีการเคลื่อนที่วนกลับมาซ้ำแนวเดิม ลำดับการเคลื่อนที่จะเปลี่ยนแปลงไปด้วย จึงได้นำตำแหน่งเฟรมกึ่งกลางของชุดภาพมาแบ่งชุดภาพออกเป็นสองส่วน เพื่อป้องกันการซ้อนทับกันของการเคลื่อนที่ที่วนกลับมาทิศทางเดิม

(ก) (ข)

รูปที่ 20 ภาพตัวอย่าง MHI หลังจากแบ่งชุดภาพออกเป็น 2 ส่วน
(ก) ส่วนแรกของการเตะ (ข) ส่วนหลังของการเตะ

นอกจากนี้การกำหนดค่าความเข้มแสงที่แสดงลำดับการเคลื่อนที่ใน MHI จะใช้จำนวนของเฟรมที่จะใช้สร้าง MHI เป็นตัวกำหนดช่วงของค่าความเข้มแสง ถ้าจำนวนของเฟรมในการสร้าง MHI ต่างกัน ค่าความเข้มแสงของ MHI ในแต่ละครั้งจะต่างกันด้วย ดังนั้นจึงแก้ปัญหาด้วยการใส่ลำดับการเคลื่อนที่แทน โดยให้ลำดับล่าสุดมีค่ามากที่สุดและลำดับก่อนหน้าลดลงทีละ 1



รูปที่ 21 ตัวอย่าง MHI จากการกำหนดลำดับเหตุการณ์เป็นตัวเลข

3. การรู้จำและแยกประเภทท่าทาง (Action Classification)

การรู้จำท่าทางที่สนใจ ได้แก่ การเตะ และการชก โดยต้องแยกออกจากท่าทางที่เก็บมาทดลองทั้งหมด 4 ท่า คือ การเตะ การชก การถีบ และการผลัก ซึ่งจะแยกออกเป็นสองกลุ่ม โดยแยกการเตะและการถีบให้อยู่ในกลุ่มเดียวกัน และที่เหลืออยู่อีกกลุ่มหนึ่ง เป็นการเคลื่อนที่ส่วนล่างและส่วนบนตามลำดับ ซึ่งแบ่งการเคลื่อนที่ส่วนล่างกับส่วนบนโดยพิจารณาที่จุดศูนย์กลางของบริเวณการเคลื่อนที่ถ้าสูงกว่ากึ่งกลางภาพตามแนวตั้ง จะเป็นการเคลื่อนที่ส่วนบน ถ้าต่ำกว่าเป็นส่วนล่าง

เมื่อแบ่งเป็นการเคลื่อนที่เป็นกลุ่มส่วนบนและส่วนล่างแล้ว นำแต่ละส่วนที่สร้างเป็น MHI เรียบร้อยแล้ว และทดสอบระบบด้วยการใช้การนำตัวหนึ่งออกไป (Leave-one-out Cross-validation) โดยดึงเอาหนึ่งในจำนวนของเฟรมออก แล้วใช้ที่เหลือในแต่ละชุดภาพไปสอนระบบโดยใช้ซัพพอร์ตเวคเตอร์แมชชีน (SVM) ที่เหมาะสำหรับการแยกข้อมูลสองประเภทที่สุด และผลที่ได้จะนำมาเฉลี่ยจากจำนวนข้อมูลทั้งหมด ซึ่งจะใช้ทดสอบการแยกท่าทางทั้งแยกการเตะออกจากการถีบ และแยกการชกออกจากการผลัก พร้อมทั้งรายงานผลการรู้จำจากการแยกประเภทด้วยผลของ Predicted Positive rate เพราะต้องการทราบผลการรู้จำท่าทางที่สนใจเท่านั้น ดังนั้นจึงสนใจส่วนของการหาข้อมูลที่ถูกต้องเท่านั้น

บทที่ 4

การทดลองและผลการทดลอง

งานวิจัยนี้นำเสนอการรู้จำท่าทางของบุคคลด้วยการใช้ MHI เป็นตัวแสดงการเคลื่อนที่ของแต่ละท่าทาง โดยแบ่งการทดลองและผลการทดลองออกเป็น 2 ส่วน คือ การรู้จำการเตะและการชกกับท่าทางที่ต่างกันชนิดละ 1 ท่าทาง และการรู้จำการเตะและการชกกับท่าทางที่ต่างกันชนิดละ 3 ท่าทาง ซึ่งในแต่ละส่วนจะอธิบายรายละเอียดเกี่ยวกับ ข้อมูลกับชนิดของข้อมูลที่ใช้ทดลอง การกำหนดค่าตัวแปรในการทดลอง วิธีการทดลอง ผลการทดลอง และการวิเคราะห์ผลการทดลอง

ก่อนจะกล่าวถึงข้อมูลในแต่ละส่วนจะกล่าวถึงข้อมูลของท่าทางที่มีทั้งหมด ในลักษณะของข้อมูลที่เก็บมาเป็นภาพวีดิทัศน์ โดยมีอัตราเฟรมในการสกัดจากภาพวีดิทัศน์เป็นเฟรมภาพ 15 เฟรมต่อวินาที จากนั้นแยกเป็นชุดภาพและจำนวนเฟรมในแต่ละชุดภาพไม่เท่ากัน ลักษณะของภาพวีดิทัศน์ที่เก็บมามีดังนี้

- ภาพคนแสดงท่าทางเต็มตัว
- มีวิถีการเคลื่อนที่แบบไปกลับและซ้ำเส้นทางเดิม
- หนึ่งชุดภาพมีหนึ่งคนแสดงท่าทางหนึ่งครั้ง
- ใช้กล้องเก็บภาพเพียงมุมมองเดียว ตัวเดียว
- ระดับการถ่ายภาพอยู่ในระดับสายตา
- ในเฟรมภาพจะต้องมีคนแสดงท่าทางที่สนใจเพียงคนเดียว
- ไม่มีการแสดงท่าทางร่วมกับบุคคลอื่น
- แสดงโดยหันด้านข้างทั้งด้านซ้ายและขวา

โดยมีจำนวนนักแสดงทั้งหมด 11 คน แสดงท่าทาง 4 ท่าทาง คือ การเตะ การชก การถีบ และการผลัก นอกจากนี้ยังมีฉากหลังของการเก็บภาพที่เป็นอีกตัวแปรสำหรับการทดลองนี้ คือ ฉากหลังคงที่ไม่ซับซ้อน และฉากหลังที่มีความหลากหลาย โดยฉากหลังคงที่ไม่ซับซ้อนจะให้นักแสดงทั้งหมดแสดงท่าทางทั้ง 4 ท่าทาง โดยแสดงการเตะและการชกคนละ 6 ครั้ง การถีบและการผลักคนละ 3 ครั้ง รวมทั้งหมด 369 ชุดภาพ ส่วนฉากหลังที่มีความหลากหลายจะแสดงการเตะและการชกคนละ 3 ครั้ง การถีบและการผลักคนละ 3 ครั้ง รวมทั้งหมด 264 ชุดภาพ แสดงตามรูปที่ 11 และ 13

1. การรู้จำการเตะและการชกกับท่าทางที่ต่างกัน 1 ท่าทาง

การทดลองในส่วนนี้มีจุดประสงค์ที่จะรู้จำการเตะและการชก โดยที่การเตะและการชกจะถูกรวมอยู่กับการถีบและการผลัก ซึ่งทั้ง 4 ท่าทางจะถูกแสดงอยู่ในรูปของ MHI และต้องการทดลองว่าถ้าใช้ซัพพอร์ตเวคเตอร์แมชชีน (SVM) ในการแยกประเภทข้อมูล 2 ประเภทออกจากกัน เป็นการแยกท่าทางแบบ 1 ชนิดต่อ 1 ชนิด โดยเปรียบเทียบระหว่างการแยกท่าทางที่ต้องการรู้จำออกจากท่าทางที่มีความใกล้เคียงกับท่าทางที่สนใจ โดยท่าทางที่มีความใกล้เคียงกันได้แก่ การเตะกับการถีบ และ การชกกับการผลัก ในการทดลองจึงได้จับคู่การทดลองดังนี้

- แยกประเภทระหว่างการเตะกับการถีบ (รู้จำการเตะ)
- แยกประเภทระหว่างการชกกับการผลัก (รู้จำการชก)

นอกจากการเปรียบเทียบตามลักษณะท่าทางการเคลื่อนที่แล้ว ตัวแปรอื่นที่มีการกำหนดสำหรับการทดลองนี้ ได้แก่ อัตราส่วนจำนวนของชุดภาพที่นำมาทดลอง การแบ่งส่วนชุดภาพสำหรับ MHI และการกำหนดลำดับของการเคลื่อนที่ที่แสดงใน MHI และการกำหนดค่าตัวแปรสำหรับการทดลองมีดังนี้

1) *<u>อัตราส่วนจำนวนของชุดภาพที่นำมาทดลอง</u>* จะมีการกำหนดให้อัตราส่วนของจำนวนชุดภาพท่าทางที่จะรู้จำต่อจำนวนชุดภาพที่นำมาทดสอบเพิ่มขึ้นเรื่อยๆ เป็น 1:1 , 2:1 และ 3:1

- อัตราส่วน 1:1
  - รู้จำการเตะ การเตะ 22 ชุดภาพ การถีบ 22 ชุดภาพ
  - รู้จำการชก การชก 22 ชุดภาพ การผลัก 22 ชุดภาพ
- อัตราส่วน 2:1
  - รู้จำการเตะ การเตะ 44 ชุดภาพ การถีบ 22 ชุดภาพ
  - รู้จำการชก การชก 44 ชุดภาพ การผลัก 22 ชุดภาพ
- อัตราส่วน 3:1
  - รู้จำการเตะ การเตะ 66 ชุดภาพ การถีบ 22 ชุดภาพ
  - รู้จำการชก การชก 66 ชุดภาพ การผลัก 22 ชุดภาพ

2) *<u>การแบ่งส่วนชุดภาพสำหรับ MHI</u>* เนื่องจาก MHI จะเน้นที่การเคลื่อนที่ล่าสุด และการแบ่งชุดภาพเป็น 2 ส่วนจากเฟรมกึ่งกลางของชุดภาพ จะทำให้ MHI เน้นที่ภาพที่ตำแหน่งกึ่งกลางและภาพที่ลำดับสุดท้าย แต่ท่าทางที่นำมาทดลองอาจมีลักษณะเด่นที่ตำแหน่งอื่น จึงเพิ่มการแบ่งชุดภาพอีก จากเดิมที่มี 2 ส่วน ได้แบ่งย่อย 2 ส่วนนั้นด้วยค่ากลางของแต่ละส่วน เป็น 4 ส่วน ดังนั้นในการทดลองจะมีการ

แบ่งชุดภาพแล้วแปลงเป็น MHI 2 แบบ คือ MHI 2 ส่วน กับ MHI 4 ส่วน (4 ส่วนจะทดลองกับการรู้จำการเตะเท่านั้น เพราะการชกมีจำนวนเฟรมน้อยไม่สามารถแบ่งเป็น 4 ส่วนได้)

3) *การกำหนดลำดับของการเคลื่อนที่* จากการที่เปลี่ยนจากใช้ค่าความเข้มแสงในการกำหนดลำดับการเคลื่อนที่ใน MHI เป็นการบอกลำดับการเคลื่อนที่โดยให้ลำดับล่าสุดเป็นค่าที่มากที่สุด แล้วลำดับก่อนหน้าก็ลดหลั่นกันมาเรื่อยๆ ถ้าต้องการเห็นทุกลำดับเหตุการณ์ที่เกิดขึ้นและให้ลำดับแรกของเหตุการณ์หมดที่ 1 พอดี จะต้องกำหนดลำดับให้พอดีกับจำนวนเฟรมที่จะสร้าง MHI และจากข้อ 2) ที่มีการแบ่งชุดภาพเป็น 2 และ 4 ส่วน จึงได้มีการกำหนดค่าของลำดับเป็นจำนวนเฟรมที่เท่ากับเฟรมกึ่งกลาง เนื่องจากชุดภาพค่าเฟรมกึ่งกลางไม่คงที่ แต่ต้องมีการกำหนดให้เท่ากันทุกชุดภาพ จึงได้มีการทดสอบว่าถ้ากำหนดให้ใช้ค่าที่จำนวนเฟรมน้อยที่สุดเป็นลำดับ คือ 6 และค่าเฉลี่ยของค่าเฟรมกึ่งกลาง คือ 8 ค่าไหนมีผลต่อการรู้จำมากกว่ากัน

1.1 วิธีการทดลอง

1) เตรียมชุดภาพท่าทางที่จะทดลองตามอัตราส่วนที่กล่าวไว้ (1:1 , 2:1 และ 3:1) ที่อยู่ในรูปของ MHI และปรับค่าของลักษณะ MHI ตามตัวแปรที่กล่าวไว้
2) รู้จำโดยใช้ซัพพอร์ตเวคเตอร์แมชชีน (SVM)
3) ทดสอบความถูกต้องโดยการใช้การตรวจสอบแบบนำตัวหนึ่งออกไป (Leave-one-out Cross-validation) ในการวัดความถูกต้อง
4) พิจารณาค่าความถูกต้องที่ Recall rate หรือ ค่าการค้นหาข้อมูลที่ถูกต้อง, Precision rate หรือ ค่าความแม่นยำ และ Accuracy rate หรือ ค่าความถูกต้องในการแยกข้อมูล

1.2 ผลการทดลอง

ผลจากการทดลองโดยปรับค่าตัวแปรแสดงตามตารางด้านล่างโดยตารางที่ 1 เป็นการแสดงผล Recall rate ของการรู้จำท่าทางการเตะและการชก ตารางที่ 2 เป็นการแสดงผล Prediction rate และตารางที่ 3 จะแสดงผล Accuracy rate

ตารางที่ 1 ผล Recall rate ของการรู้จำท่าทางการเตะและการชกกับท่าทางที่ต่างกัน 1 ท่าทาง

ท่าทาง	อัตราส่วน	ส่วนของชุดภาพ	ฉากหลังคงที่		ฉากหลังที่มีความหลากหลาย	
			MHI 6 ลำดับ	MHI 8 ลำดับ	MHI 6 ลำดับ	MHI 8 ลำดับ
เตะ	1:1	2 part	0.68	0.77	0.77	0.81
		4 part	0.86	0.81	0.95	0.95
	2:1	2 part	0.90	0.93	0.93	0.93
		4 part	0.93	0.93	0.95	0.95
	3:1	2 part	0.95	0.95	0.98	0.98
		4 part	0.93	0.93	0.98	0.98
ชก	1:1	2 part	0.50	N/A	0.54	N/A
		4 part	N/A	N/A	N/A	N/A
	2:1	2 part	0.65	N/A	0.88	N/A
		4 part	N/A	N/A	N/A	N/A
	3:1	2 part	0.83	N/A	0.92	N/A
		4 part	N/A	N/A	N/A	N/A

ตารางที่ 2 ผล Prediction rate ของการรู้จำท่าทางการเตะและการชกกับท่าทางที่ต่างกัน 1 ท่าทาง

ท่าทาง	อัตราส่วน	ส่วนของชุดภาพ	ฉากหลังคงที่		ฉากหลังที่มีความหลากหลาย	
			MHI 6 ลำดับ	MHI 8 ลำดับ	MHI 6 ลำดับ	MHI 8 ลำดับ
เตะ	1:1	2 part	0.60	0.62	0.77	0.90
		4 part	0.86	0.85	0.91	0.91
	2:1	2 part	0.80	0.80	0.89	0.91
		4 part	0.89	0.91	0.93	0.93
	3:1	2 part	0.88	0.88	0.90	0.94
		4 part	0.92	0.92	0.95	0.95
ชก	1:1	2 part	0.52	N/A	0.75	N/A
		4 part	N/A	N/A	N/A	N/A
	2:1	2 part	0.65	N/A	0.79	N/A
		4 part	N/A	N/A	N/A	N/A
	3:1	2 part	0.80	N/A	0.84	N/A
		4 part	N/A	N/A	N/A	N/A

ตารางที่ 3 ผล Accuracy rate ของการรู้จำท่าทางการเตะและการชกกับท่าทางที่ต่างกัน 1 ท่าทาง

ท่าทาง	อัตราส่วน	ส่วนของชุดภาพ	ฉากหลังคงที่		ฉากหลังที่มีความหลากหลาย	
			MHI 6 ลำดับ	MHI 8 ลำดับ	MHI 6 ลำดับ	MHI 8 ลำดับ
เตะ	1:1	2 part	0.61	0.65	0.77	0.86
		4 part	0.86	0.84	0.93	0.93
	2:1	2 part	0.78	0.80	0.87	0.89
		4 part	0.87	0.89	0.92	0.92
	3:1	2 part	0.87	0.87	0.90	0.94
		4 part	0.89	0.89	0.95	0.95
ชก	1:1	2 part	0.52	N/A	0.68	N/A
		4 part	N/A	N/A	N/A	N/A
	2:1	2 part	0.54	N/A	0.77	N/A
		4 part	N/A	N/A	N/A	N/A
	3:1	2 part	0.72	N/A	0.81	N/A
		4 part	N/A	N/A	N/A	N/A

1.3 วิเคราะห์ผลการทดลอง

จากผลการทดลองเปรียบเทียบการรู้จำท่าทางกับท่าทางที่ต่างกัน 1 ท่าทาง ผล Recall rate, Precision rate และ Accuracy rate แสดงผลตามตาราง ซึ่งผลที่ได้จาก Recall rate บอกได้ว่าค่าการค้นหาข้อมูลที่ถูกต้องสามารถหาข้อมูลที่ถูกต้องของการเตะได้สูงถึง 95% และ 98% สำหรับฉากหลังคงที่และฉากหลังที่มีความหลากหลายตามลำดับ ให้ผล 83% และ 92% สำหรับการชก ในส่วนของความแม่นยำ (Precision rate) การเตะให้ผลสูงประมาณ 93% การชกประมาณ 82% และความถูกต้องของการแยกประเภทข้อมูล (Accuracy rate) การเตะจะได้ผลประมาณ 93% การชกประมาณ 75% สามารถวิเคราะห์ได้ว่า กรณีของ 3 ตัวแปรที่เพิ่มเข้ามาเพื่อทดลอง คือ อัตราส่วนระหว่างท่าทางที่สนใจต่อท่าทางที่นำมาทดสอบ การแบ่งส่วนของชุดภาพสำหรับ MHI และการกำหนดลำดับการเคลื่อนที่ พบว่าเมื่อเพิ่มค่าให้กับแต่ละตัวแปร จะสามารถเพิ่มความสามารถในการหาข้อมูล ความแม่นยำ และความถูกต้องในการแยกประเภท ดังนั้นหากต้องการเพิ่มค่าของ Recall rate, Precision rate และ Accuracy rate การเพิ่มจำนวนของข้อมูล หรือแบ่งส่วนภาพเพื่อเน้นจุดเด่นเพิ่มขึ้น หรือ เลือกค่าที่กำหนดลำดับที่เหมาะสมกับจำนวนลำดับของข้อมูล จะช่วยเพิ่มค่าเหล่านั้นได้

ผลการรู้จำระหว่างการเตะและการชกพบว่าการเตะจะมีค่าสูงกว่าการชกทั้ง Recall rate, Precision rate และ Accuracy rate เนื่องจากขนาดของพื้นที่การเคลื่อนที่ วิถีการเคลื่อนที่ที่ชัดเจน และรายละเอียดที่แสดงการเคลื่อนที่ของ MHI เพราะการเตะเป็นการเคลื่อนที่ของขาที่มีขนาดใหญ่กว่าการชกที่ใช้แขน การที่การเตะมีพื้นที่มากกว่า ทำให้เครื่องมือแสดงรายละเอียดการเคลื่อนที่อย่าง MHI จะแสดงความละเอียดของลำดับการเคลื่อนที่และช่วงการเคลื่อนที่ได้ดี ส่งผลให้มีวิถีการเคลื่อนที่ชัดเจน ส่วนการชกเป็นการเคลื่อนที่ของอวัยวะขนาดเล็กที่เร็ว MHI จะแสดงรายละเอียดการเคลื่อนที่ด้วยลำดับที่ไม่ละเอียดเพราะเมื่อเคลื่อนที่เร็ว จำนวนเฟรมจะน้อย ลำดับการเคลื่อนที่จึงน้อยด้วย ช่วงของการเคลื่อนที่อย่างการชกจะแสดงวิถีการเคลื่อนที่ค่อนข้างแคบ ทำให้การแสดงการเคลื่อนที่ของการชกด้วย MHI แสดงได้ไม่ชัดเจนเท่าการเตะ พร้อมทั้งท่าทางที่นำมาทดสอบคือการผลัก ซึ่งมีความใกล้เคียงกันมากทำให้ผลต่างๆ ออกมาน้อยกว่าการเตะ

นอกจากนี้ผลที่ได้จากฉากหลังที่ต่างกันได้ผลคือ ในส่วนของฉากหลังที่มีความคงที่จะให้ผลของทั้ง Recall rate, Precision rate และ Accuracy rate น้อยกว่าฉากหลังที่มีความหลากหลาย เนื่องจากความละเอียดของลำดับการแสดงท่าทาง ผู้แสดงท่าทางในฉากหลังที่มีความหลากหลายแสดงได้ชัดเจนกว่า ซึ่งหมายถึงขนาดของอวัยวะที่เคลื่อนที่ วิถีการเคลื่อนที่และรายละเอียดของการเคลื่อนที่ทั้งความเร็วกับช่วงของการเคลื่อนที่ที่แสดงใน MHI ดังนั้นจะบอกได้ว่า ถ้าการเคลื่อนที่ที่ชัดเจนของผู้แสดงขนาดของอวัยวะที่แสดงท่าทาง วิถีการเคลื่อนที่และรายละเอียดของการเคลื่อนที่ที่แสดงใน MHI ชัดเจน การรู้จำจะสามารถทำได้ดี แม้ในฉากหลังที่มีความหลากหลาย

2. การรู้จำการเตะและการชกกับท่าทางที่ต่างกัน 3 ท่าทาง

การทดลองในส่วนนี้มีจุดประสงค์ที่จะรู้จำการเตะและการชก โดยทั้งการเตะและการชกจะถูกรวมอยู่กับการถีบและการผลัก ซึ่งทั้ง 4 ท่าทางจะถูกแสดงอยู่ในรูปของ MHI และต้องการทดลองว่าถ้าใช้ซัพพอร์ตเวคเตอร์แมชชีน (SVM) ในการแยกประเภทข้อมูล 2 ประเภทออกจากกัน เป็นการแยกท่าทางแบบ 1 ชนิดต่อ 3 ชนิด ซึ่งจะทำการเลือกท่าทางที่ต้องการรู้จำออกจากท่าทางอื่นอีก 3 ชนิด คือ ต้องการรู้จำการเตะ ท่าทางที่เหลือจะเป็นท่าทางที่ใช้ทดสอบ เมื่อต้องการรู้จำการชก จะให้อีก 3 ท่าทางที่เหลือเป็นตัวทดสอบ ในการทดลองจึงได้จับคู่การทดลองดังนี้

- แยกประเภทระหว่างการเตะกับท่าทางอื่นที่ไม่ใช่การเตะ(การถีบ การชก การผลัก) (รู้จำการเตะ)
- แยกประเภทระหว่างการชกกับท่าทางอื่นที่ไม่ใช่การชก (การเตะ การถีบ การผลัก) (รู้จำการชก)

สำหรับอัตราส่วนจำนวนของชุดภาพที่นำมาทดลอง การแบ่งส่วนชุดภาพสำหรับ MHI และ การกำหนดลำดับของการเคลื่อนที่ที่แสดงใน MHI และการกำหนดค่าตัวแปรสำหรับการทดลองมีดังนี้

1) *อัตราส่วนจำนวนของชุดภาพที่นำมาทดลอง* จะมีการกำหนดให้อัตราส่วนของจำนวนชุดภาพท่าทางที่จะรู้จำต่อจำนวนชุดภาพที่นำมาทดสอบเป็น 1:1

- อัตราส่วน 1:1
  - รู้จำการเตะ การเตะ 66 ชุดภาพ ไม่ใช่การเตะ 66 ชุดภาพ (การถีบ 22 ชุดภาพ การชก 22 ชุดภาพ การผลัก 22 ชุดภาพ)
  - รู้จำการชก การชก 66 ชุดภาพ ไม่ใช่การชก 66 ชุดภาพ (การเตะ 22 ชุดภาพ การถีบ 22 ชุดภาพ การผลัก 22 ชุดภาพ)

2) *การแบ่งส่วนชุดภาพสำหรับ MHI* ในการทดลองจะมีการแบ่งชุดภาพแล้วแปลงเป็น MHI 2 แบบ คือ MHI 2 ส่วน กับ MHI 4 ส่วน สำหรับการรู้จำการเตะเท่านั้น ส่วนของการรู้จำการชกเนื่องจากจำนวนของเฟรมน้อยกว่ามาก การแบ่งเป็น 4 ส่วน จึงทำได้ยาก ดังนั้นจะใช้การแบ่ง MHI เป็น 2 ส่วนเท่านั้นสำหรับการทดลอง

3) *การกำหนดลำดับของการเคลื่อนที่* จากการที่เปลี่ยนจากใช้ค่าความเข้มแสงเป็นการบอกลำดับการเคลื่อนที่โดยให้ลำดับล่าสุดเป็นค่าที่มากที่สุด แล้วลำดับก่อนหน้าก็ลดหลั่นกันมาเรื่อยๆ และจากข้อ 2) ที่มีการแบ่งชุดภาพเป็น 2 และ 4 ส่วน จึงได้มีการกำหนดค่าของลำดับเป็นจำนวนเฟรมที่เท่ากับเฟรมกึ่งกลาง เนื่องจากชุดภาพค่าเฟรมกึ่งกลางไม่คงที่ แต่ต้องมีการกำหนดให้เท่ากันทุกชุดภาพ จึงได้มีกำหนดให้ใช้ 6 และ8 เป็นค่าลำดับ

2.1 วิธีการทดลอง

1) เตรียมชุดภาพท่าทางที่จะทดลองตามอัตราส่วนที่กล่าวไว้ (1:1) ที่อยู่ในรูปของ MHI และปรับค่าของลักษณะ MHI ตามตัวแปรที่กล่าวไว้
2) รู้จำโดยใช้ซัพพอร์ตเวคเตอร์แมชชีน (SVM)
3) ทดสอบความถูกต้องโดยการใช้การตรวจสอบแบบนำตัวหนึ่งออกไป (Leave-one-out Cross-validation) ในการวัดความถูกต้อง
4) พิจารณาค่าความถูกต้องที่ Recall rate, Precision rate และ Accuracy rate ของผลที่ได้

2.2 ผลการทดลอง

ผลจากการทดลองโดยปรับค่าตัวแปรแสดงตามตารางด้านล่างโดยตารางที่ 4 เป็นการแสดงผล Recall rate ของการรู้จำท่าทางการเตะและการชกเมื่อเทียบกับท่าทางที่ไม่ใช่การเตะ 3 ชนิด (การถีบ การชก การผลัก) ตารางที่ 5 เป็นการแสดงผล Precision rate โดยตารางที่ 6 แสดงผล Accuracy rate

ตารางที่ 4 ผล Recall rate ของการรู้จำท่าทางการเตะและการชกกับท่าทางที่ต่างกัน 3 ท่าทาง

ท่าทาง	อัตราส่วน	ส่วนของชุดภาพ	ฉากหลังคงที่		ฉากหลังที่มีความหลากหลาย	
			MHI 6 ลำดับ	MHI 8 ลำดับ	MHI 6 ลำดับ	MHI 8 ลำดับ
เตะ	1:1	2 part	0.84	0.81	0.86	0.84
		4 part	0.86	0.89	0.83	0.83
ชก	1:1	2 part	0.72	N/A	0.84	N/A
		4 part	N/A	N/A	N/A	N/A

ตารางที่ 5 ผล Prediction rate ของการรู้จำท่าทางการเตะและการชกกับท่าทางที่ต่างกัน 3 ท่าทาง

ท่าทาง	อัตราส่วน	ส่วนของชุดภาพ	ฉากหลังคงที่		ฉากหลังที่มีความหลากหลาย	
			MHI 6 ลำดับ	MHI 8 ลำดับ	MHI 6 ลำดับ	MHI 8 ลำดับ
เตะ	1:1	2 part	0.82	0.80	0.85	0.91
		4 part	0.86	0.86	0.94	0.94
ชก	1:1	2 part	0.68	N/A	0.76	N/A
		4 part	N/A	N/A	N/A	N/A

ตารางที่ 6 ผล Accuracy rate ของการรู้จำท่าทางการเตะและการชกกับท่าทางที่ต่างกัน 3 ท่าทาง

ท่าทาง	อัตราส่วน	ส่วนของชุดภาพ	ฉากหลังคงที่		ฉากหลังที่มีความหลากหลาย	
			MHI 6 ลำดับ	MHI 8 ลำดับ	MHI 6 ลำดับ	MHI 8 ลำดับ
เตะ	1:1	2 part	0.83	0.81	0.85	0.88
		4 part	0.86	0.87	0.89	0.89
ชก	1:1	2 part	0.69	N/A	0.79	N/A
		4 part	N/A	N/A	N/A	N/A

2.3 วิเคราะห์ผลการทดลอง

จากผลการทดลองเปรียบเทียบการรู้จำท่าทางการเตะกับท่าทางอื่นที่แตกต่างกัน 3 ชนิด และการชกกับท่าทางอื่นที่แตกต่างกันอีก 3 ชนิด ซึ่งในส่วนนี้จะทดสอบว่าหากข้อมูลที่เข้ามามีการกระจายตัว ผลการจัดกลุ่มของข้อมูลพบว่าผล Recall rate, Precision rate และ Accuracy rate แสดงผลตามตาราง ซึ่งผลที่ได้จาก Recall rate บอกได้ว่าค่าการค้นหาข้อมูลที่ถูกต้องสามารถหาข้อมูลที่ถูกต้องของการเตะได้สูงถึง 89% และ 83% สำหรับฉากหลังคงที่และฉากหลังที่มีความหลากหลายตามลำดับ ให้ผล 72% และ 84% สำหรับการชก ในส่วนของความแม่นยำ (Precision rate) การเตะให้สูงประมาณ 90% การชกประมาณ 72% และความถูกต้องของการแยกประเภทข้อมูล (Accuracy rate) การเตะจะได้ประมาณ 88% การชกประมาณ 74% ซึ่งสามารถวิเคราะห์ได้ว่าข้อมูลที่มีการกระจายตัวจะทำให้ประสิทธิภาพในการรู้จำลดลง เนื่องจากการใช้ซัพพอร์ตเวคเตอร์แมชชีนเป็นเครื่องมือที่เหมาะสำหรับการแยกข้อมูล 2 กลุ่ม และจะได้ผลดีเมื่อข้อมูลของแต่ละกลุ่มมีความเหมือนกันภายในกลุ่มชัดเจนทั้งสองกลุ่ม แต่ข้อมูลที่ทดสอบเป็นข้อมูลที่มีความสัมพันธ์ภายในกลุ่มชัดเจนหนึ่งกลุ่ม และความสัมพันธ์ที่แปรปรวนอีกหนึ่งกลุ่มการสร้างไฮเปอร์เพลนเพื่อแยกข้อมูลจึงทำได้ไม่ดีเท่าแยกข้อมูลที่มีความเหมือนกันภายในกลุ่มชัดเจนทั้งสองกลุ่ม

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

บทนี้จะกล่าวถึงการสรุปผลการวิจัยในการรู้จำท่าทางการเตะและการชก โดยใช้กล้องมุมมองเดียวและมีฉากหลังหลากหลายแบบ พร้อมทั้งข้อเสนอแนะดังนี้

สรุปผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์ที่จะรู้จำท่าทางการเตะและการชก ที่มีวิถีการเคลื่อนที่เป็นแบบไปกลับและซ้ำแนวเดิม โดยจะต้องสามารถรู้จำท่าทางทั้งสองได้ และแยกออกจากท่าทางอื่นที่นำมาทดสอบระบบ เช่น การถีบ และการผลักได้ ในการทดลองระบบการรู้จำจะมีการเลือกท่าทางที่ใกล้เคียงกับท่าทางที่สนใจ เพื่อที่จะบอกได้ว่าสามารถรู้จำได้ดีแค่ไหน และการทดลองกับฉากหลังหลายแบบก็เป็นตัวชี้วัดได้ว่าระบบมีความทนทานต่อสภาพแวดล้อมที่เปลี่ยนได้มากน้อยแค่ไหน

ข้อมูลที่เก็บมานั้นเป็นการเก็บภาพวีดิทัศน์ที่มีคนแสดงท่าทางทั้งหมด 11 คน แต่จะแสดงครั้งละคนแบบเต็มตัวในภาพและเป็นท่าทางที่หันด้านข้าง ซึ่งแต่ละคนนั้นจะแสดงท่าทางทั้งหมด 4 ท่าทาง ในสภาพแวดล้อมที่มีทั้งฉากหลังคงที่ไม่วุ่นวาย (ในร่ม) และฉากหลังที่มีความหลากหลาย (กลางวัน) โดยใช้กล้องตัวเดียวและมุมมองเดียว จากนั้นนำภาพวีดิทัศน์ที่ได้มาสกัดเป็นเฟรมภาพโดยมีอัตราเฟรม 15 เฟรมต่อวินาที พร้อมทั้งจัดเป็นชุดภาพตามท่าทางที่แสดงโดยมี 1 ท่าทาง ต่อ 1 ชุดภาพ จะได้ชุดภาพทั้งหมดคือฉากหลังคงที่ไม่วุ่นวายจำนวน 396 ชุดภาพ และฉากหลังที่มีความหลากหลายจำนวน 264 ชุดภาพ

ขั้นตอนการดำเนินงานจะประกอบด้วย 3 ขั้นตอนหลักๆ คือ การหาบริเวณที่เกิดการเคลื่อนที่ การแสดงการเคลื่อนที่ในรูปแบบ MHI และการรู้จำและแยกประเภทท่าทาง ในส่วนของการหาบริเวณการเคลื่อนที่จะมีการแบ่งย่อยขั้นตอนสำหรับการหาบริเวณที่เกิดการเคลื่อนที่โดยหาบริเวณอย่างคร่าวๆ ก่อนด้วยการใช้การหาความแตกต่างระหว่างเฟรมเป็นภาพระดับเทาเพื่อหาบริเวณที่มีการเคลื่อนที่ จากนั้นนำภาพผลต่างทั้งหมดมารวมกันจะได้บริเวณที่มีค่าความเข้มแสงมากๆ ซึ่งตั้งสมมตฐานว่าบริเวณนั้นน่าจะเป็นบริเวณที่มีการเคลื่อนที่ ต่อมาใช้การฉายภาพทั้งแนวแกนตั้งและแกนนอนเพื่อหาขอบเขตของบริเวณการเคลื่อนที่ เมื่อเลือกขอบเขตอย่างคร่าวๆ ได้แล้ว ก็นำภาพความแตกต่างระหว่างเฟรมที่หาไว้ในรูปที่ตัดขอบเขตอย่างคร่าวๆ แล้วมาแปลงเป็นภาพขาว-ดำ และนับจำนวนจุดภาพสีขาวที่เป็นพื้นที่ของการเคลื่อนที่ในแต่ละเฟรม จากนั้นเลือกเฟรมในช่วงตำแหน่งกึ่งกลาง 3 เฟรม เพื่อหาเฟรมที่มีจำนวนจุดภาพสีขาวน้อยที่สุดเป็นเฟรมกึ่งกลางของการเคลื่อนที่ในชุดภาพนั้น เพื่อที่จะนำไปทำการหาบริเวณการเคลื่อนที่ที่เลือกเฉพาะอวัยวะส่วนที่เคลื่อนที่เท่านั้น และใช้เป็นตำแหน่งในการแบ่งส่วนชุดภาพก่อนสร้าง MHI เพื่อแก้ปัญหาการทับซ้อนกันของข้อมูล สำหรับการหาบริเวณการเคลื่อนที่ที่สนใจเฉพาะอวัยวะที่เคลื่อนที่นั้นได้ใช้วิธีเปรียบคู่แม่แบบในการหาอวัยวะและตำแหน่งที่เหมาะสมที่สุดสำหรับการตัดขอบเขตของ

บริเวณที่มีการเคลื่อนที่และเปรียบคู่กับภาพที่ตำแหน่งเฟรมกึ่งกลางของการเคลื่อนที่ ซึ่งแม่แบบที่นำมาจะเป็นสี่เหลี่ยมจัตุรัสที่มีลักษณะเป็นภาพขาว-ดำ ที่ส่วนบนประมาณ 20 เปอร์เซ็นต์ของความยาวเป็นสีขาว และส่วนล่างเป็นสีดำ โดยขนาดของแม่แบบจะเปลี่ยนตามขนาดของขอบเขตในแนวแกนนอนที่หาได้จากการ OR ข้อมูลในแนวแกนตั้งและเลือกจุดภาพสีขาวจุดแรกที่เจอทั้งจากซ้ายและขวา และปรับขนาดของแม่แบบตามระยะของจุดภาพสีขาวจากซ้ายไปขวา แม่แบบที่ได้มานั้นจะ AND กับภาพขาว-ดำที่ได้จากการหักลบภาพฉากหลัง เมื่อได้ตำแหน่งที่ให้ผลการ AND สูงสุดจะทำการตัดขอบเขตที่บริเวณนี้ตามขนาดของแม่แบบ และเฟรมอื่นๆ ในชุดภาพเดียวกันที่เป็นภาพขาว-ดำจากการหักลบภาพฉากหลังเช่นกันออกมา จะได้ภาพขาว-ดำที่ผ่านการหักลบภาพฉากหลังและตัดขอบเขตได้บริเวณที่เป็นการเคลื่อนที่เฉพาะส่วน

ส่วนของการแสดงการเคลื่อนที่ในรูปแบบ MHI เป็นการสร้างภาพที่แสดงการเคลื่อนที่ทั้งหมดในชุดภาพนั้นตามลำดับการเคลื่อนที่ โดยการเคลื่อนที่ล่าสุดจะมีความเข้มแสงสูงสุดหรือมีค่าสูงสุดตามที่กำหนด แต่เนื่องจากการเคลื่อนที่ที่มีวิถีการเคลื่อนที่ไปกลับแบบซ้ำเส้นทางเดิมจะทำให้เกิดการทับซ้อนของข้อมูลจึงได้ใช้เฟรมกึ่งกลางของการเคลื่อนที่เป็นตัวแบ่งชุดภาพออกเป็นสองส่วน และสร้าง MHI จากแต่ละส่วนของชุดภาพนั้น ได้มีการกำหนดลำดับการเคลื่อนที่แทนการใช้ค่าความเข้มแสง โดยให้การเคลื่อนที่ลำดับล่าสุดมีค่าสูงสุดลำดับก่อนหน้าจะลดหลั่นตามลำดับ แต่ก่อนที่จะนำไปแยกประเภทหรือรู้จำท่าทาง ขนาดของ MHI ที่จะมีขนาดเท่ากับขอบเขตจากการตัดบริเวณที่เป็นการเคลื่อนที่เฉพาะส่วนแต่ละชุดภาพนั้นไม่เท่ากัน จึงมีการปรับขนาดของภาพให้เท่ากันทั้งหมดโดยใช้ค่าเฉลี่ยของขนาดทั้งหมดเป็นขนาดมาตรฐานและปรับขนาดของ MHI ตามค่าเฉลี่ยนั้นด้วยวิธีการประมาณค่าแบบไบคิวบิก

ส่วนสุดท้ายคือการรู้จำท่าทางและแยกออกจากท่าทางอื่นโดยใช้ซัพพอร์ตเวคเตอร์แมชชีน (SVM) โดยที่ลักษณะของข้อมูลจะประกอบด้วย MHI ที่เป็นเครื่องมือที่แสดงการเคลื่อนที่ และจำนวนเฟรมในชุดภาพ จะแยกข้อมูลเป็น 2 ส่วน คือ ใช่ท่าทางที่ต้องการ และไม่ใช่ท่าทางที่ต้องการ มีวิธีการทดสอบคือใช้การตรวจสอบแบบนำตัวหนึ่งออกไป (Leave-one-out Cross-validation) เป็นตัวทดสอบ และนำผลที่ได้มาคำนวณค่า Recall rate, Precision rate และ Accuracy rate เป็นผลที่จะบอกว่าระบบสามารถรู้จำท่าทางที่สนใจได้เท่าไหร่ ระบบมีความแม่นยำเท่าไหร่ และระบบมีความสามารถแยกประเภทข้อมูลได้ดีแค่ไหน

จากการทดลองการรู้จำท่าทางทั้งการเตะและการชก โดยสนใจค่า Recall rate, Precision rate และ Accuracy rate ที่คำนวณได้จากตาราง Confusion matrix พบว่าเมื่อมีการเพิ่มตัวแปรในการทดลองไป 3 ตัวแปร คือ อัตราส่วนของข้อมูลที่นำมาทดลอง การกำหนดลำดับการเคลื่อนที่ของ MHI และการแบ่งส่วนการเคลื่อนที่ของ MHI จะส่งผลให้ค่าทั้ง 3 ที่สนใจเพิ่มขึ้นด้วย โดยเฉพาะการแบ่งส่วนการเคลื่อนที่ที่จะช่วยแยกการเตะและการถีบได้ดีขึ้น เพราะการถีบเมื่อแบ่งเป็น 4 ส่วนจะสามารถแสดงรายละเอียดการเคลื่อนที่และช่วยแยกออกจากการเตะได้ดี ดังนั้นการเพิ่มค่าให้กับตัวแปรทั้ง 3 จะช่วยให้ได้ผลการรู้จำดีขึ้น และถ้าต้องการเน้นรายละเอียดของท่าทางอื่นๆ ที่มีลักษณะเด่นของท่าทางอยู่ตำแหน่งใดๆ ในช่วงของ

การเคลื่อนที่ การแบ่งส่วนของการเคลื่อนที่ให้เน้นที่ตำแหน่งนั้นแล้วค่อยแสดงด้วย MHI จะสามารถช่วยให้แยกท่าทางได้ดีขึ้น

ผลการรู้จำระหว่างการเตะและการชกพบว่าการเตะจะมีค่าสูงกว่าการชก เนื่องจากขนาดของอวัยวะที่แสดงท่าทาง (ขา, แขน) วิถีของท่าทาง และรายละเอียดของการเคลื่อนที่ที่แสดงด้วย MHI เพราะการเตะ เป็นการเคลื่อนที่ของขา ที่มีวิถีการเคลื่อนที่ที่ชัดเจน พร้อมทั้งมีรายละเอียดที่แสดงโดย MHI ทั้งลำดับของการเคลื่อนที่ที่มีจำนวนที่ค่อนข้างมาก และช่วงการเคลื่อนที่ที่กว้าง ผลการแสดงท่าทางที่ได้จึงมีรายละเอียดของการเคลื่อนที่ที่มีความชัดเจน ส่วนการชกเป็นการเคลื่อนที่ของแขนที่มีขนาดเล็กกว่าขามาก วิถีการเคลื่อนที่ค่อนข้างสั้น เพราะเคลื่อนที่เร็ว ระยะทางการเคลื่อนที่น้อย จำนวนเฟรมที่จะเป็นตัวบอกลำดับก็น้อยด้วย ทำให้รายละเอียดของการแสดงท่าทางไม่ชัดเจนเท่ากับการเตะ อีกทั้งการเตะเป็นการแยกออกจากการถีบที่มีการแบ่งส่วนการเคลื่อนที่มาช่วยในการแยกประเภททำให้ผลออกมาดีกว่าการชกที่จะต้องแยกประเภทออกจากการผลักที่มีลักษณะการเคลื่อนที่ใกล้เคียง ดังนั้น หากต้องการให้ระบบเรียนรู้และแยกท่าทางได้ดีการเน้นที่ขนาดของวัตถุที่เคลื่อนที่ มีวิถีที่ชัดเจน และแสดงรายละเอียดการเคลื่อนที่มากพอ จะช่วยเพิ่มประสิทธิภาพของการแยกข้อมูลมากขึ้น

สำหรับการใช้ฉากหลังที่ต่างกันทั้งสองแบบคือ ฉากหลังที่คงที่ และฉากหลังที่มีความหลากหลาย (เป็นฉากช่วงเวลากลางวันในสถานที่ต่างๆ โดยที่ด้านหลังผู้แสดงมีการเคลื่อนที่ของวัตถุขนาดที่เล็กกว่าผู้แสดงแต่ไม่วุ่นวาย หรือซับซ้อนมาก หรือไม่มีการเคลื่อนที่ของวัตถุขนาดใหญ่ผ่านขณะแสดงท่าทาง) ระบบจะสามารถสกัดเอาบริเวณที่เป็นการเคลื่อนที่ที่สนใจออกมาได้ดีทั้งสองแบบ แต่ส่วนที่ส่งผลต่อการรู้จำที่สำคัญคือ ขนาดของอวัยวะที่แสดงการเคลื่อนที่ วิถีการเคลื่อนที่ และรายละเอียดของการแสดงเคลื่อนที่ใน MHI จะมีผลให้การรู้จำสามารถทำได้ดี สำหรับกรณีที่ต้องการแยกท่าทางที่มีลักษณะเด่นอยู่ในช่วงใดช่วงหนึ่งของการเคลื่อนที่ การแบ่งส่วนของการเคลื่อนที่สำหรับ MHI ก็จะเป็นอีกลักษณะหนึ่งที่เพิ่มเข้ามาเพื่อช่วยเน้นลักษณะเด่นของท่าทางนั้น

ข้อเสนอแนะ

1. การสกัดบริเวณที่มีการเคลื่อนที่ยังใช้งานกับฉากหลังที่มีความหลากหลายยังไม่ดีพอ อาจจะมีการเพิ่มวิธีการสร้างโมเดลของฉากหลัง เพื่อความทนทานต่อสภาพแวดล้อมทั่วไป

2. การสร้าง MHI ของทั้งการเตะและการชก พื้นที่ของอวัยวะจะมีผลต่อการแสดงท่าทางทั้งหมด และการชกเป็นการใช้แขนที่มีขนาดพื้นที่น้อยกว่าขา วิถีและรายละเอียดการเคลื่อนที่อาจไม่ชัดเจนเท่าการเตะ ดังนั้นอาจมีการปรับเพิ่มทางด้านการเก็บข้อมูลให้มีขนาดของแขนที่ใหญ่ขึ้นจากการปรับระยะห่างระหว่างกล้องกับคนแสดง หรืออาจเพิ่มวิธีอื่นที่จะได้ความละเอียดการเคลื่อนที่ของการชกมากขึ้น

3. การใช้ MHI เพียงอย่างเดียวอาจไม่เพียงพอสำหรับการแสดงท่าทางที่ใกล้เคียงกัน (การชกและการผลัก) ดังนั้นอาจจะมีการเพิ่มลักษณะเด่นอื่นๆ เพิ่มเข้ามาอีก

รายการอ้างอิง

[1] Gonzalez, R.C. and Woods, R.E., Digital Image Processing. U.S.A.: Prentice Hall, 2002.

[2] Bobick, A. and Davis, J. , An appearance-based representation of action. *Proceedings of the 13th International Conference on Pattern Recognition*, vol.1, no., pp.307-312 ,1996.

[3] Bobick, A.F. and Davis, J.W., The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.23, no.3, pp.257-267, Mar 2001.

[4] Ahad Md.A.R., Tan J.K., Kim H. and Ishikawa S., Motion history image: its variants and applications. *Machine Vision and Applications*, v.23 n.2, pp.255-281, March 2012.

[5] Schuldt , C., Laptev,I. and Caputo, B., Recognizing Human Actions: A Local SVM Approach. *17th International Conference on Pattern Recognition(ICPR'04)*, pp. 32-36, August 23-26 2004.

[6] Fletcher, T., Support Vector Machines Explained, 2008.

[7] Cristianini, N.and Taylor, J., An Introduction to Support VectorMachines and Other Kernel-based Learning Methods.: *Cambridge UP*, 2000.

[8] Vapnik,V., Statistical Learning Theory, 1998.

[9] Evgeniou, T.,Pontil, E.M. and Elisseeff, A., Leave-one-out error, stability, and generalization of voting combinations of classifiers.: *Mach. Learning*, vol. 55, no. 1, pp. 71 -97, 2004.

[10] Kohavi, R., and Provost, F., Glossary of terms.: Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process, 30(2–3), 1998.

[11] Blank, M., Gorelick, L., Shechtman, E., Irani, M. and Basri. R., Actions as Space-Time Shapes. *ICCV*, 2005.

[12] Weinland, D., Ronfard, R. and Boyer, E., Motion history volumes for free viewpoint action recognition. *IEEE International Workshop on modeling People and Human Interaction*, 2005. Available from: http://perception.inrialpes.fr/Publications/2005/WRB05 [2011, April 20]

[13] Ke, Y., Sukthankar, R. and Hebert, M., Event Detection in Crowded Videos. *ICCV*, 2007.

[14] Kuno, Y., Watanabe, T., Shimosakoda, Y. and Nakagawa, S. Automated Detection of Human for Visual Surveillance System. *ICPR,* 1996.

[15] Dedeoğlu, Y., Töreyin, B.U., Güdükbay, U. and Çetin, A. E. Silhouette-Based Method for Object Classification and Human Action Recognition in Video. HCI/ECCV, 2006.

[16] Lv, F. and Nevatia, R., Single View Human Action Recognition using Key Pose Matching and Viterbi Path Searching, *In Proceedings of CVPR*, 2007.

[17] Niebles, J.C. and Li, F., A Hierarchical Model of Shape and Appearance for Human Action Classification, *In Proceedings of CVPR*, 2007.

[18] Yang, W., Wang, Y. and Mori, G., Human Action Recognition from a Single Clip per Action. *ICCV,* 2009.

[19] Castellanos, R., Kalva, H., Marques, O. and Furht, B., Event Detection in Video Using Motion Analysis. *ARTEMIS*, 2010.

[20] National Institute of Standards and Technology (NIST). TRECVID 2008 Evaluation for Surveillance Event Detection [Online]. Available: http://www.nist.gov/speech/tests/trecvid/2008/, [2011, April 20].

[21] Yao, B. and Zhu, S. Learning Deformable Action Template from Cluttered Videos. *ICCV*, 2009.

[22] Guo, P., Miao, Z. and Cheng, H. Mask Based Human Action Detection in Crowded Videos. *ICIP,* 2010.

[23] Valstar, M., Pantic, M. and Patras, I., Motion history for facial action detection in video. *In: Proc. IEEE Int. Conf. SMC*, vol. 1, pp. 635–640, 2004.

[24] Pantic, M., Patras, I. and Valstar, M.F., Learning spatio-temporal models of facial expressions. *In: Proc. Int. Conf. on Measuring Behavior*, pp. 7–10, September 2005.

[25] Valstar, M., Patras, I. and Pantic, M., Facial action recognition using temporal templates. *In: Proc. IEEE Workshop on Robot and Human Interactive Communication*, pp. 253–258, 2004.

[26] Ahad Md.A.R., Ogata T., Tan J.K., Kim H. and Ishikawa S., A complex motion recognition technique employing directional motion templates. *Int. J. Innov. Comput. Inf.* Control **4**(8), pp.1943–1954 , 2008.

[27] Kellokumpu, V., Zhao, G. and Pietikäinen, M., Texture based description of movements for activity analysis. *In: Proc. Conf. Computer Vision Theory and Applications (VISAPP'08)*, vol. 2, pp. 368–374, Portugal, 2008.

ภาคผนวก

ภาคผนวก ก

จำนวนเฟรมของแต่ละท่าทางในหนึ่งชุดภาพวีดิทัศน์

ตารางที่ ก-1 แสดงจำนวนเฟรมของแต่ละท่าทางในหนึ่งชุดภาพวีดิทัศน์ ที่ฉากหลังคงที่

ท่าทาง	ทิศการเคลื่อนที่	ลำดับชุดข้อมูล	ลำดับผู้แสดงท่าทาง										
			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
เตะ	ขวา	1	15	14	18	14	17	17	20	15	17	15	15
		2	15	14	18	12	17	19	19	16	18	14	17
		3	14	15	16	12	17	18	19	17	16	14	19
	ซ้าย	1	15	20	16	14	16	17	18	18	18	16	15
		2	17	19	17	14	17	18	23	17	18	15	15
		3	15	17	17	14	19	17	18	16	18	16	15
ถีบ	ขวา	1	22	20	19	22	19	18	20	18	23	23	13
		2	22	18	15	19	20	20	19	20	23	19	14
		3	17	18	14	21	17	18	18	20	21	19	13
	ซ้าย	1	23	21	21	19	19	18	17	17	25	17	13
		2	17	20	19	18	22	17	16	17	18	20	14
		3	21	19	17	19	18	18	16	20	20	18	15
ชก	ขวา	1	14	11	10	13	10	9	14	12	12	10	7
		2	14	12	9	12	12	9	13	14	13	12	9
		3	16	12	12	11	10	13	13	11	12	10	7
	ซ้าย	1	12	10	11	11	12	10	12	12	10	11	9
		2	13	10	10	9	11	10	14	11	14	11	9
		3	12	12	14	12	12	13	14	12	13	10	7
ผลัก	ขวา	1	12	15	12	13	12	12	14	16	13	12	11
		2	12	10	14	13	11	11	10	13	10	10	10
		3	11	12	11	13	12	10	10	9	11	12	10
	ซ้าย	1	9	15	10	11	11	10	14	13	12	11	11
		2	9	12	12	11	16	9	9	17	12	13	12
		3	10	12	11	12	10	10	10	10	17	11	11

การเตะ ค่าเฉลี่ย: 16.48 SD: 1.99 การชก ค่าเฉลี่ย: 11.37 SD: 1.86

การถีบ ค่าเฉลี่ย: 18.65 SD: 2.59 การผลัก ค่าเฉลี่ย: 11.66 SD: 1.88

ตารางที่ ก-2 แสดงจำนวนเฟรมของแต่ละท่าทางในหนึ่งชุดภาพวีดิทัศน์ ที่ฉากหลังที่มีความหลากหลาย

ท่าทาง	ทิศการเคลื่อนที่	ลำดับชุดข้อมูล	ลำดับผู้แสดงท่าทาง										
			1	2	3	4	5	6	7	8	9	10	11
เตะ	ขวา	1	14	18	14	13	14	14	16	13	16	15	13
		2	14	17	13	12	14	14	15	12	15	14	13
		3	14	16	13	12	14	14	16	14	13	15	14
	ซ้าย	1	14	15	12	12	14	16	15	12	13	15	13
		2	14	14	13	11	13	15	15	13	15	16	14
		3	13	13	13	12	12	15	14	15	15	17	13
ถีบ	ขวา	1	23	19	17	18	15	18	20	17	16	17	14
		2	16	17	19	16	17	16	17	15	16	16	14
		3	17	19	16	17	17	17	16	16	17	18	14
	ซ้าย	1	19	17	15	17	20	17	16	18	17	19	13
		2	16	16	15	13	20	17	16	16	17	18	15
		3	16	16	16	14	19	17	16	17	18	20	14
ชก	ขวา	1	11	11	14	10	11	10	11	10	11	13	12
		2	11	15	13	10	11	8	13	10	11	12	10
		3	10	14	14	10	10	11	11	11	11	13	10
	ซ้าย	1	9	15	12	8	11	10	10	12	12	12	11
		2	9	15	13	12	10	10	9	13	12	13	11
		3	11	12	11	10	11	12	8	11	12	13	12
ผลัก	ขวา	1	10	13	13	12	12	10	12	10	11	12	10
		2	10	16	13	9	11	10	12	11	9	12	11
		3	13	13	14	13	11	12	13	15	13	14	10
	ซ้าย	1	10	15	10	10	13	10	11	14	11	13	12
		2	10	16	13	9	11	10	10	11	12	13	11
		3	13	12	13	9	13	12	11	17	12	12	12

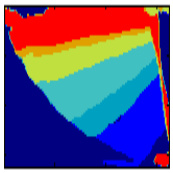
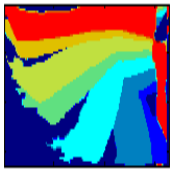
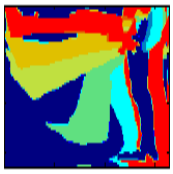
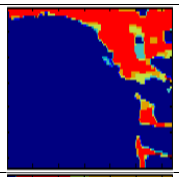
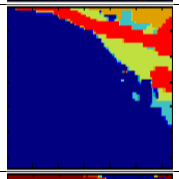
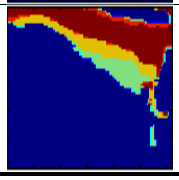
การเตะ	ค่าเฉลี่ย:	14.00	SD:	1.40	การชก	ค่าเฉลี่ย:	11.28	SD:	1.60	
การถีบ	ค่าเฉลี่ย:	16.77	SD:	1.81	การผลัก	ค่าเฉลี่ย:	11.83	SD:	1.76	

ภาคผนวก ข

ตัวอย่างการแสดงภาพ MHI ที่ให้แสดงผลการแยกด้วย Confusion Matrix

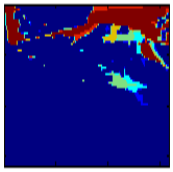
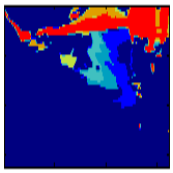
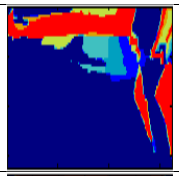
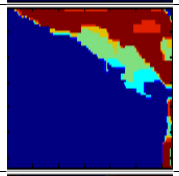
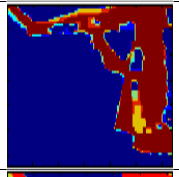
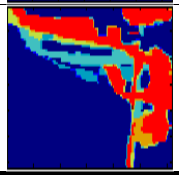
ตารางที่ ข-1 ตัวอย่างภาพ Motion History Image (MHI) ที่แสดงผลการแยกด้วย Confusion Matrix

		Actual Class: Yes	Actual Class: No
Predicted Class	Yes	TP	FP
	No	FN	TN

	ท่าทาง	MHI	หมายเหตุ
True Positive	เตะ		- ผลจากการทดสอบกับท่าทางการถีบหลายๆ ชุด ยังสามารถแยกได้ว่าเป็นการเตะ เพราะพื้นที่ลำดับและวิถีชัดเจน
			- ผลจากการทดสอบกับท่าทางการถีบหลายๆ ชุด ยังสามารถแยกได้ว่าเป็นการเตะ เพราะพื้นที่ลำดับและวิถีชัดเจน
			- ผลจากการทดสอบกับท่าทางการถีบหลายๆ ชุด ยังสามารถแยกได้ว่าเป็นการเตะ เพราะพื้นที่ลำดับและวิถีชัดเจน
	ชก		- ผลจากการทดสอบกับท่าทางการผลักหลายๆ ชุด ยังสามารถแยกได้ว่าเป็นการชก
			- ผลจากการทดสอบกับท่าทางการผลักหลายๆ ชุด ยังสามารถแยกได้ว่าเป็นการชก
			- ผลจากการทดสอบกับท่าทางการผลักหลายๆ ชุด ยังสามารถแยกได้ว่าเป็นการชก

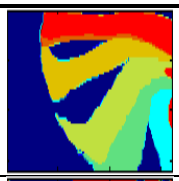
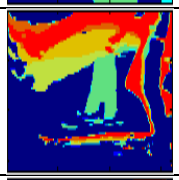
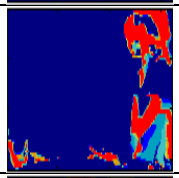
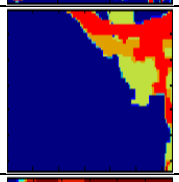
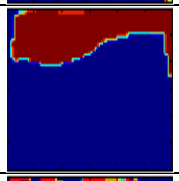
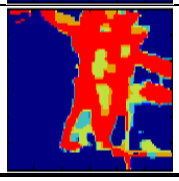
ตารางที่ ข-1 ตัวอย่างภาพ Motion History Image (MHI) ที่แสดงผลการแยกด้วย Confusion Matrix

		Actual Class: Yes	Actual Class: No
Predicted Class	Yes	TP	FP
	No	FN	TN

	ท่าทาง	MHI	หมายเหตุ
False Positive	ถีบ		- พื้นที่ของขาที่เคลื่อนที่ถูกสกัดออกมาไม่ค่อยดี เนื่องจากสีของขาและฉากหลังมีความใกล้เคียงกัน
			- พื้นที่ของขามีขนาดเล็กการสกัดออกจากฉากหลังจึงได้น้อยและไม่สมบูรณ์ - ท่าทางที่ได้คล้ายการเตะมาก แต่ถ้าแบ่งส่วนการเคลื่อนที่เพิ่มจะสามารถช่วยให้แยกได้ดี
			- ท่าทางมีความใกล้เคียงการเตะมาก การแบ่งส่วนการเคลื่อนที่เพิ่มจะสามารถช่วยให้แยกได้ดี
	ผลัก		- ท่าทางใกล้เคียงการชกมาก และรายละเอียดการเคลื่อนที่ไม่ชัดเจน
			- ท่าทางใกล้เคียงการชกมาก และลำดับการเคลื่อนที่ไม่ชัดเจน
			- ท่าทางใกล้เคียงการชกมาก ไม่มีลักษณะพิเศษอื่นที่ช่วยแยกท่าทางทั้งสองได้ชัดเจน

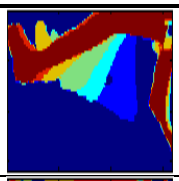
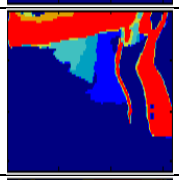
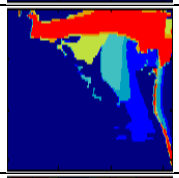
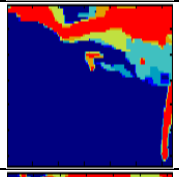
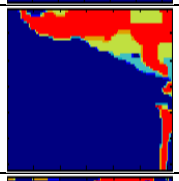
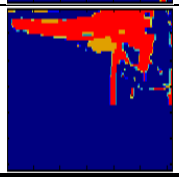
ตารางที่ ข-1 ตัวอย่างภาพ Motion History Image (MHI) ที่แสดงผลการแยกด้วย Confusion Matrix

		Actual Class: Yes	Actual Class: No
Predicted Class	Yes	TP	FP
	No	FN	TN

	ท่าทาง	MHI	หมายเหตุ
False Negative	เตะ		- ท่าทางที่แสดงไม่เต็มพื้นที่ ทำให้ขนาดของพื้นที่ไม่เท่ากับภาพการเตะอื่นๆ จึงทายผลผิด
			- ท่าทางใกล้เคียงการถีบมาก การแบ่งส่วนจะช่วยให้แยกได้ชัดเจนขึ้น
			- บริเวณการเคลื่อนที่ถูกตัดมาผิดตำแหน่ง ทำให้ทายผิด
	ชก		- ท่าทางที่ตัดมาไม่สมบูรณ์ อาจเพราะเลือกเฟรมที่ตำแหน่งกึ่งกลางการเคลื่อนที่ไม่ค่อยดี
			- บริเวณที่ตัดมาผิดตำแหน่ง
			- มีการเคลื่อนที่อื่นปะปนที่ฉากหลัง ทำให้ท่าทางที่แสดงโดย MHI ผิดเพี้ยนไป

ตารางที่ ข-1 ตัวอย่างภาพ Motion History Image (MHI) ที่แสดงผลการแยกด้วย Confusion Matrix

		Actual Class: Yes	Actual Class: No
Predicted Class	Yes	TP	FP
	No	FN	TN

	ท่าทาง	MHI	หมายเหตุ
True Negative	ถีบ		- ผลจากการทดสอบกับท่าทางการเตะหลายๆ ชุด ยังสามารถแยกได้ว่าเป็นการถีบ
			- ผลจากการทดสอบกับท่าทางการเตะหลายๆ ชุด ยังสามารถแยกได้ว่าเป็นการถีบ
			- ผลจากการทดสอบกับท่าทางการเตะหลายๆ ชุด ยังสามารถแยกได้ว่าเป็นการถีบ
	ผลัก		- ผลจากการทดสอบกับท่าทางการชกหลายๆ ชุด ยังสามารถแยกได้ว่าเป็นการผลัก
			- ผลจากการทดสอบกับท่าทางการชกหลายๆ ชุด ยังสามารถแยกได้ว่าเป็นการผลัก
			- ผลจากการทดสอบกับท่าทางการชกหลายๆ ชุด ยังสามารถแยกได้ว่าเป็นการผลัก

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวภาวินี จรูญพันธุ์ เกิดวันที่ 21 มกราคม พ.ศ. 2530 ภูมิลำเนาอยู่ที่ จ.ชัยภูมิ สำเร็จการศึกษาระดับปริญญาวิศวกรรมศาสตรบัณฑิต สาขาอิเล็กทรอนิกส์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ในปีการศึกษา 2552 หลังจากนั้นได้เข้ามาศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2552