



การจดจำท่าทางมนุษย์จากมุมมองสูงโดยใช้ข้อมูลความลึกและสี  
**Top-view Based Human Action Recognition Using Depth  
and Color Information**

สิทธิศักดิ์ เชี่ยวภากร

Sittisuk Seawpakorn

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา  
วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์  
มหาวิทยาลัยสงขลานครินทร์

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

2556

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์

ชื่อวิทยานิพนธ์      การจดจำท่าทางมนุษย์จากมุมสูงโดยใช้ข้อมูลความลึกและสี  
ผู้เขียน                นายสิทธิศักดิ์ เชี่ยวภากร  
สาขาวิชา              วิศวกรรมคอมพิวเตอร์

---

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

คณะกรรมการสอบ

.....  
(ผู้ช่วยศาสตราจารย์ ดร.นิคม สุวรรณวร)

.....ประธานกรรมการ  
(ดร.อนันต์ ชกสุริวงค์)

.....กรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร.นิคม สุวรรณวร)

.....กรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร.ปรัชญสิทธิ์ สมานพิบูรณ์)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้รับวิทยานิพนธ์ฉบับนี้ สำหรับ  
การศึกษา ตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์

.....  
(รองศาสตราจารย์ ดร.ธีระพล ศรีชนะ)

คณบดีบัณฑิตวิทยาลัย

ขอรับรองว่า ผลงานวิจัยนี้มาจากการศึกษาวิจัยของนักศึกษาเอง และได้แสดงความขอบคุณบุคคลที่มีส่วนช่วยเหลือแล้ว

ลงชื่อ.....  
(ผู้ช่วยศาสตราจารย์ ดร.นิคม สุวรรณวร)  
อาจารย์ที่ปรึกษาวิทยานิพนธ์

ลงชื่อ.....  
( นายสิทธิศักดิ์ เชี่ยวภากร )  
นักศึกษา

ข้าพเจ้าขอรับรองว่า ผลงานวิจัยนี้ไม่เคยเป็นส่วนหนึ่งในการอนุมัติปริญญาในระดับใดมาก่อน และ  
ไม่ได้ถูกใช้ในการยื่นขออนุมัติปริญญาในขณะนี้

ลงชื่อ.....

( นายสิทธิศักดิ์ เชี่ยวภากร )

นักศึกษา

ชื่อวิทยานิพนธ์	การจดจำท่าทางมนุษย์จากมุมมอง โดยใช้ข้อมูลความลึกและสี
ผู้เขียน นาย	สิทธิศักดิ์ เชี่ยวภากร
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
ปีการศึกษา	2555

### บทคัดย่อ

วิทยานิพนธ์ฉบับนี้เป็นงานวิจัยเพื่อประยุกต์ใช้เทคนิคด้านการประมวลผลภาพ วิเคราะห์การรู้จำท่าทางของมนุษย์เพื่อใช้สำหรับการเฝ้าระวังเหตุการณ์ที่ผิดปกติในสถานที่สำคัญต่างๆ และในผู้ป่วย ผู้พิการ คนชราที่อาศัยอยู่เพียงลำพัง เมื่อเกิดเหตุการณ์ผิดปกติสามารถให้การช่วยเหลือได้อย่างรวดเร็ว โดยในงานวิจัยนี้จะมุ่งเน้นไปที่เทคนิคการรู้จำท่าทางของมนุษย์ ประกอบด้วยท่าทางพื้นฐาน 5 ท่าทาง ได้แก่ การยืน การเดิน การนั่ง การก้ม และการนอน ระบบจะมีการนำข้อมูลความลึก สี และการเคลื่อนไหวของวัตถุเข้ามาช่วยในการวิเคราะห์ภาพจากมุมมอง ซึ่งภาพจากมุมมองจะช่วยลดความซับซ้อนของร่างกายมนุษย์ด้วย โดยเทคนิคการรู้จำท่าทางของมนุษย์ประกอบด้วย (1) กระบวนการตรวจจับและติดตามมนุษย์ เพื่อให้ได้บริเวณที่สนใจสำหรับการวิเคราะห์ (2) กระบวนการสร้างโครงสร้างเสมือนมนุษย์ เพื่อหาลักษณะเด่นของท่าทางมนุษย์ โดยการรวมความสัมพันธ์ของการประมาณระยะทางของวัตถุจากแผนผังคิซพาริตี ในแกน U และ V ซึ่งจะได้โครงสร้างมนุษย์ใหม่ นำมาคำนวณหามุม และค่าเฉลี่ยความสูงของโครงสร้างมนุษย์ สำหรับใช้เป็นลักษณะเด่นในการรู้จำ (3) กระบวนการรู้จำท่าทางมนุษย์ จะแบ่งเป็น 4 กลุ่มท่าทางคือ 1.การยืนและการเดิน 2.การนั่ง 3.การก้ม 4.การนอน โดยใช้วิธีของ นิวรอนเน็ตเวิร์ค สำหรับการเทรนลักษณะเด่นและทดสอบความถูกต้องของการรู้จำท่าทาง โดยใช้ตัวอย่างสำหรับการทดลอง 400 ข้อมูล ท่าทางละ 100 ข้อมูล ได้ผลการรู้จำท่าทางมนุษย์ให้ความถูกต้องโดยเฉลี่ย 99.3 เปอร์เซ็นต์ โดยที่ระบบสามารถทำงานแบบเรียลไทม์ได้อย่างมีประสิทธิภาพ และรองรับการติดตั้งกล้องบนเพดานในความสูงที่แตกต่างกัน โดยไม่ส่งผลกระทบต่อความถูกต้องของระบบ

<b>Thesis Title</b>	Top-View Based Human Action Recognition Using Depth and Color Information
<b>Author</b>	Mr.Sittisuk Seawpakorn
<b>Major Program</b>	Computer Engineering
<b>Academic Year</b>	2012

## ABSTRACT

This thesis applies image processing techniques for human action recognition in surveillance systems for accident events in habitation, or to observe disabled persons when they slip, alerting the attendant. This thesis recognizes five human actions: standing, walking, sitting, bending, and laying. The system uses color images, depth images, and motion information from a top-view camera. The strength of a top-view camera is that it reduces the complexity of human patterns. The system is established through the following steps. First, motion object extraction is used to obtain position, size, and object area. Secondly, human structure analysis is performed, where the human structure is created from a human structure project via color UV-disparity. The feature is the angle of the human structure. Finally, human action recognition is divided into four groups: 1. standing or walking, 2. sitting, 3. bending, 4. laying. A neural network is applied as the classification method. The dataset consists of 400 images, with 100 images for each action. The recognition rate of human action recognition is 99.3%. The system can recognize actions accurately in real-time and support different camera heights.

## กิตติกรรมประกาศ

สำหรับการดำเนินการวิจัยและจัดทำวิทยานิพนธ์นี้ ผู้วิจัยขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.นิคม สุวรรณวร ประธานกรรมการที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้คำปรึกษา ชี้แนะแนวทางในการทำงาน ทั้งยังให้กำลังใจและการเอาใจใส่กับข้าพเจ้าเป็นอย่างมาก กระตุ้นให้ ข้าพเจ้า ได้มีความมุ่งมั่นในการทำงานให้สำเร็จลุล่วงไปได้เป็นอย่างดีเรื่อยมา รวมถึงการตรวจและ แก้ไขเนื้อหาวิทยานิพนธ์ให้สำเร็จสมบูรณ์

ขอขอบพระคุณ ดร.อนันต์ ชกสูริวงศ์ กรรมการที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้ คำแนะนำในการปรับปรุงวิทยานิพนธ์ให้สมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณคณาจารย์ และบุคลากรทุกท่านในภาควิชาวิศวกรรม คอมพิวเตอร์ทุกท่านที่ให้ความช่วยเหลือในระหว่างการทำวิทยานิพนธ์

ขอขอบพระคุณเพื่อน ๆ และนักศึกษาปริญญาโททุกท่านที่คอยให้คำแนะนำและ คอยให้ความช่วยเหลือด้วยดีตลอดมา

สุดท้ายนี้ ขอกราบขอบพระคุณ บิดา มารดาและญาติพี่น้องทุกคน ซึ่งเป็นผู้มี พระคุณสูงสุดที่ให้กำลังใจและให้การสนับสนุนทุกสิ่งทุกอย่างด้วยดีตลอดมาในชีวิตของข้าพเจ้า

สิทธิศักดิ์ เขียวภากร

## สารบัญ

สารบัญ.....	(8)
รายการรูปภาพ.....	(11)
รายการตาราง.....	(13)
บทที่ 1 บทนำ .....	14
1.1. ความสำคัญและที่มาของงานวิจัย .....	14
1.2. งานวิจัยที่เกี่ยวข้อง .....	15
1.2.1. การสกัดวัตถุที่สนใจ (foreground extraction).....	15
1.2.2. การจดจำท่าทางมนุษย์ (action recognition).....	21
1.3. วัตถุประสงค์ .....	23
1.4. ขอบเขตงานวิจัย .....	23
1.5. ประโยชน์ที่คาดว่าจะได้รับ .....	24
บทที่ 2 ทฤษฎีและหลักการ .....	25
2.1. กระบวนการตรวจจับการเคลื่อนไหว (Motion Detection).....	25
2.1.1. การตรวจจับการเคลื่อนไหวพื้นฐาน (ความแตกต่างระหว่างเฟรม).....	25
2.1.2. การสร้างพื้นหลังจากการผสมของระเบียบวิธีเกาส์เซียนหลายรูปแบบ .....	26
2.2. การติดตามวัตถุที่สนใจ .....	27
2.2.1. การวิเคราะห์กลุ่มของจุดภาพที่เป็นสีขาวดำ (Blob analysis).....	28
2.3. การประมวลผลภาพกับรูปร่างและโครงสร้างของภาพ (Morphological Image Processing) ..	28
2.3.1. เทคนิคของการ Hit และ Miss .....	28
2.3.2. โอเปอเรชันพื้นฐานสำหรับรูปร่างหรือโครงสร้างพื้นฐาน .....	29
2.3.3. การขยายภาพ (Dilation).....	29
2.3.4. การย่อภาพ (Erosion) .....	30
2.3.5. โอเปอเรชัน Closing และ Opening .....	32
2.4. แบบจำลองของกล้อง (Camera model) .....	34



2.5. การแยกวัตถุจากการแผนผังดิสพาริตี.....	38
2.6. โครงข่ายประสาทเทียม (Artificial Neural network).....	39
บทที่ 3 การออกแบบและพัฒนาระบบ .....	42
3.1. กระบวนการตรวจจับและติดตามการเคลื่อนที่ของวัตถุ (Motion Object Extraction) .....	44
3.1.1. การลบภาพพื้นหลัง (Background Subtraction) .....	44
3.1.2. การวิเคราะห์กลุ่มของจุดภาพ (Blob Analysis) .....	47
3.2. กระบวนการจับคู่สีกับวัตถุที่สนใจ (Color Match).....	48
3.2.1. การสอบเทียบภาพ (Calibration Image).....	48
3.3. การวิเคราะห์โครงสร้างมนุษย์ (Human Structure Analysis).....	51
3.3.1. การฉายโครงสร้างมนุษย์ด้วยการระบุสีให้กับการประมาณระยะทางของวัตถุจาก แผนผังดิสพาริตี (Human structure project via color UV-Disparity) .....	51
3.3.2. การสร้างโครงสร้างมนุษย์ (Human Structure) .....	55
3.3.3. การรวมโครงสร้างมนุษย์ (Structure of Human of Combination) .....	58
3.4. การสกัดลักษณะเด่น.....	62
3.4.1. การหามุมจากการระบุสีให้กับการประมาณระยะทางของวัตถุของแผนผังดิสพาริตี ...	63
3.4.2. การหามุมจากการรวมโครงสร้างมนุษย์.....	64
3.4.3. การหามุมและค่าเฉลี่ยความสูงจากการรวมโครงสร้างมนุษย์.....	64
3.5. การรู้จำท่าทางมนุษย์ (Human Recognition) .....	65
3.6. สรุป.....	68
บทที่ 4 ผลการทดลอง.....	69
4.1. การทดลองที่ 1 การรู้จำท่าทางมนุษย์โดยใช้มุมจากโครงสร้างมนุษย์ในภาพการระบุสีให้กับ การประมาณระยะทางของวัตถุจากแผนผังดิสพาริตี.....	69
4.1.1. สมมติฐาน .....	69
4.1.2. ปัจจัยกำหนดในการทดลอง .....	70
4.1.3. ผลการทดลอง .....	70

4.1.4. วิเคราะห์และสรุปผลการทดลอง.....	73
4.2. การทดลองที่ 2 การรู้จำท่าทางมนุษย์โดยใช้มุมจากภาพการรวมโครงสร้างมนุษย์ .....	73
4.2.1. สมมติฐาน .....	74
4.2.2. ปัจจัยกำหนดในการทดลอง .....	74
4.2.3. ผลการทดลอง .....	74
4.2.4. วิเคราะห์และสรุปผลการทดลอง.....	77
4.3. การทดลองที่ 3 การรู้จำท่าทางมนุษย์โดยใช้มุมและค่าเฉลี่ยความสูงจากภาพการรวม โครงสร้างมนุษย์ .....	77
4.3.1. สมมติฐาน .....	78
4.3.2. ปัจจัยกำหนดในการทดลอง .....	78
4.3.3. ผลการทดลอง .....	78
4.3.4. วิเคราะห์และสรุปผลการทดลอง.....	81
4.4. การจำแนกท่าทางมนุษย์แบบเรียลไทม์ (Realtime Action Recognition).....	82
4.5. สรุป.....	85
บทที่ 5 สรุป.....	86
5.1. สรุปผลการวิจัย .....	86
5.2. อภิปรายผล .....	87
5.3. ข้อเสนอแนะ .....	88
บรรณานุกรม .....	89
ภาคผนวก .....	91
ภาคผนวก ก. ผลงานตีพิมพ์เผยแพร่จากวิทยานิพนธ์.....	92
ประวัติผู้เขียน .....	99

## รายการรูปภาพ

ภาพประกอบ 1- 1 ภาพประกอบ 1(a) ภาพสี (b) ภาพประกอบด้วยความลึก (c) สกัควัตถุที่สนใจ (d) ภาพ Grayscale จากมุมสูง (e) ภาพมุมสูงที่ประกอบด้วยความลึก (f) การแมปสีให้กับภาพมุมสูง ...	16
ภาพประกอบ 1- 2 แม่แบบ (templates) สำหรับติดตามบุคคลที่สนใจ .....	16
ภาพประกอบ 1- 3 ผลลัพธ์จากเทคนิค background subtraction .....	17
ภาพประกอบ 1- 4 Log-polar Histogram of Human Posture.....	17
ภาพประกอบ 1- 5 กระบวนการทำ background subtraction (เฉพาะในวงกลม) .....	18
ภาพประกอบ 1- 6 ภาพรวมของระบบ.....	19
ภาพประกอบ 1- 7 (a) bounding box ของกล้องซ้าย (b) ลดขนาดและรวม bounding box (c) ผลลัพธ์สุดท้าย .....	20
ภาพประกอบ 1- 8 การสัควัตถุที่สนใจของ N.Noorit.....	20
ภาพประกอบ 1- 9 โมเดลมนุษย์และคั่นแปร .....	23
ภาพประกอบ 1- 10 ท่าทางพื้นฐาน (a) ก้ม (b) นั่ง (c) นอนราบ (d) ยืน หรือ เดิน .....	23
ภาพประกอบ 2- 1 การแยกประเภทจุดสี ของวิธี Mixture of Gaussians [10] .....	27
ภาพประกอบ 2- 2 ผลจากวิธี Mixture of Gaussians [9] (a) ภาพวิดีโอตัดต่อเฟรมปัจจุบัน (b) ภาพวัตถุเคลื่อนที่ .....	27
ภาพประกอบ 2- 3 ตัวอย่างการทำ blob tracking .....	28
ภาพประกอบ 2- 4 ตัวอย่างข้อมูลภาพ .....	29
ภาพประกอบ 2- 5 ตัวอย่างข้อมูลภาพสำหรับทำโอเปอเรชั่น .....	29
ภาพประกอบ 2- 6 ตัวอย่างการขยายภาพ (ก) ข้อมูลภาพ (ข) template.....	30
ภาพประกอบ 2- 7 ตัวอย่างการลบภาพ (ก) ข้อมูลภาพ (ข) template .....	31
ภาพประกอบ 2- 8 ผลจากการย่อภาพ .....	31
ภาพประกอบ 2- 9 ผลลัพธ์จากการย่อภาพด้วย template ใหม่.....	31
ภาพประกอบ 2- 10 ตัวอย่างการทำโอเปอเรชั่น OPEN (ก) ภาพต้นฉบับ (ข) template.....	32
ภาพประกอบ 2- 11 ตัวอย่างการทำโอเปอเรชั่น CLOSE (ก) ภาพต้นฉบับ (ข) template.....	33
ภาพประกอบ 2- 12 แบบจำลองของกล้องรูเข็ม.....	34
ภาพประกอบ 2- 13 ความแตกต่างระหว่างตำแหน่งแกนของพิกัดจุดบนระนาบกับพิกัดบนภาพทั่วไป.....	35
ภาพประกอบ 2- 14 Calibration Chart.....	38
ภาพประกอบ 2- 15 แกน U และ V สำหรับคิสพาริตี .....	38

ภาพประกอบ 2- 16 การสร้างภาพ UV-disparity (ก) ภาพต้นฉบับ (ข) ภาพ V-Disparity .....	39
ภาพประกอบ 2- 17 โมเดลของข่ายประสาทเทียม .....	40
ภาพประกอบ 2- 18 แสดงสถาปัตยกรรม Feedforward network.....	41
ภาพประกอบ 3- 1 กระบวนการโดยรวมของการวิเคราะห์พฤติกรรมมนุษย์ .....	44
ภาพประกอบ 3- 2 กระบวนการตรวจจับและติดตามการเคลื่อนที่ของมนุษย์ (ก) ภาพต้นฉบับ (ข) ภาพผลลัพธ์ของการตรวจจับวัตถุที่เคลื่อนที่ .....	46
ภาพประกอบ 3- 3 การกำจัดสัญญาณรบกวน (ก) ภาพต้นฉบับ (ข) ภาพผลลัพธ์ของการกำจัด สัญญาณรบกวน.....	47
ภาพประกอบ 3- 4 การวิเคราะห์กลุ่มของจุดภาพ .....	48
ภาพประกอบ 3- 5 การย้ายตำแหน่งของภาพความลึกไปภาพสี (ก) ภาพความลึก (ข) ภาพสี .....	50
ภาพประกอบ 3- 6 (ก) การเลือกบริเวณที่สนใจจากภาพสี (ข) การเลือกบริเวณที่สนใจจากภาพความ ลึก (ค) ผลลัพธ์จาก match กัน .....	51
ภาพประกอบ 3- 7 (ก) ผลลัพธ์ของกระบวนการติดตามเฉพาะบริเวณวัตถุที่เคลื่อนที่.....	53
ภาพประกอบ 3- 8 (ก) ภาพสี (ข) Color V-Disparity (ค) Color U-Disparity .....	55
ภาพประกอบ 3- 9 ตัวอย่างการประมาณความสูงมนุษย์.....	55
ภาพประกอบ 3- 10 โครงสร้างมนุษย์สำหรับ U-disparity: (ก) การยืน (ข) การเดิน .....	56
ภาพประกอบ 3- 11 โครงสร้างมนุษย์สำหรับ V-disparity: (ก) การยืน (ข) การเดิน .....	57
ภาพประกอบ 3- 12 ผลลัพธ์ของโครงสร้างมนุษย์จากภาพกระบวนการระบุสีให้กับการประมาณ ระยะทางของวัตถุจากแผนผังดิสพาริตี (ก) ภาพต้นฉบับ (ข) U-Disparity (ค) V-Disparity .....	58
ภาพประกอบ 3- 13 การคำนวณระยะห่างระหว่างจุดอ้างอิงกับจุดของโครงสร้างมนุษย์ : .....	60
ภาพประกอบ 3- 14 การคำนวณหาทิศทางของโครงสร้างมนุษย์.....	61
ภาพประกอบ 3- 15 โครงสร้างมนุษย์แบบใหม่: (ก) การยืน (ข) การเดิน (ค) การนั่ง .....	62
ภาพประกอบ 3- 16 มุมสำหรับโครงสร้างมนุษย์.....	62
ภาพประกอบ 3- 17 การหามุมจากภาพการระบุสีให้กับการประมาณระยะทางของวัตถุของ แผนผังดิสพาริตี .....	63
ภาพประกอบ 3- 18 โครงสร้างมนุษย์แบบฝึกความสูงของแกน V: (ก) การยืน (ข) การเดิน .....	63
ภาพประกอบ 3- 19 โครงสร้างมนุษย์แบบฝึกความสูงของแกน U: (ก) การยืน (ข) การเดิน .....	64
ภาพประกอบ 4- 1 แผนผังการคำนวณของข่ายงานประสาท .....	82
ภาพประกอบ 4- 2 โปรแกรมการรู้จำท่าทางมนุษย์.....	84

### รายการตาราง

ตารางที่ 1- 1 อัตราค่าเฉลี่ยความสูง-กว้างของท่าทางมนุษย์ .....	17
ตารางที่ 3- 1 ลักษณะเด่นของแต่ละท่าทางโครงสร้างมนุษย์จากภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสปาริที .....	65
ตารางที่ 3- 2 ลักษณะเด่นของแต่ละท่าทาง .....	66
ตารางที่ 3- 3 จำนวนโหนดที่ใช้สำหรับข่ายงานประสาท .....	67
ตารางที่ 4- 1 อัตราความถูกต้องสำหรับการสุ่มจำนวนโหนดของการทดลองที่ 1 .....	70
ตารางที่ 4- 2 แสดงผลลัพธ์จากการใช้ลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสปาริที .....	71
ตารางที่ 4- 3 ความถูกต้องของการฝึกสอนสำหรับลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสปาริที .....	71
ตารางที่ 4- 4 ความถูกต้องของการตรวจสอบสำหรับลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสปาริที .....	72
ตารางที่ 4- 5 ความถูกต้องของการทดสอบสำหรับลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสปาริที .....	72
ตารางที่ 4- 6 อัตราความถูกต้องสำหรับการสุ่มจำนวนโหนดของการทดลองที่ 2 .....	74
ตารางที่ 4- 7 แสดงผลลัพธ์จากการใช้ลักษณะเด่นจากมุมของภาพการรวมโครงสร้างมนุษย์ .....	75
ตารางที่ 4- 8 ความถูกต้องของการฝึกสอนสำหรับลักษณะเด่นจากมุมของ .....	75
ตารางที่ 4- 9 ความถูกต้องของการตรวจสอบสำหรับลักษณะเด่นจากมุมของ .....	76
ตารางที่ 4- 10 ความถูกต้องของการทดสอบสำหรับลักษณะเด่นจากมุมของ .....	76
ตารางที่ 4- 11 อัตราความถูกต้องสำหรับการสุ่มจำนวนโหนดของ .....	78
ตารางที่ 4- 12 แสดงผลลัพธ์จากการใช้ลักษณะเด่นจากมุมและค่าเฉลี่ยความสูงของ .....	79
ตารางที่ 4- 13 ความถูกต้องของการฝึกสอนสำหรับลักษณะเด่นจากมุมและค่าเฉลี่ยความสูงของภาพการรวมโครงสร้างมนุษย์ .....	79
ตารางที่ 4- 14 ความถูกต้องของการตรวจสอบสำหรับลักษณะเด่นจากมุมและค่าเฉลี่ยความสูงของภาพการรวมโครงสร้างมนุษย์ .....	80
ตารางที่ 4- 15 ความถูกต้องของการทดสอบสำหรับลักษณะเด่นจากมุมและค่าเฉลี่ยความสูงของภาพการรวมโครงสร้างมนุษย์ .....	80
ตารางที่ 4- 16 แสดงผลลัพธ์การรู้จำท่าทางมนุษย์ในความสูงที่แตกต่างกัน .....	81

## บทที่ 1

### บทนำ

#### 1.1. ความสำคัญและที่มาของงานวิจัย

ปัจจุบันงานวิจัยเกี่ยวกับการวิเคราะห์พฤติกรรมมนุษย์มีผู้ให้ความสนใจเป็นจำนวนมาก เนื่องจากเป็นงานวิจัยพื้นฐานที่สำคัญสามารถนำไปประยุกต์ใช้สำหรับงานเกี่ยวกับการรักษาความปลอดภัยในสถานที่สำคัญต่างๆ หรือการเฝ้าระวังอุบัติเหตุที่สามารถเกิดขึ้นได้กับบุคคลที่ช่วยเหลือตัวเองไม่ได้เช่น คนชราและคนพิการ โดยการทำงานของระบบจะใช้ภาพจากกล้องวีดีโอนำมาวิเคราะห์ลักษณะพฤติกรรมของมนุษย์ เพื่อจำแนกท่าทางต่างๆ โดยสามารถบ่งบอกเหตุการณ์ผิดปกติที่อาจจะเกิดขึ้นได้อย่างอัตโนมัติ เมื่อเกิดเหตุการณ์ผิดปกติระบบสามารถแจ้งเตือนสำหรับผู้ดูแลสถานที่ให้เข้ามาทำการช่วยเหลืออย่างทันที ซึ่งระบบนี้จะช่วยลดภาระงานของคนที่ทำหน้าที่เฝ้าดูแลรักษาความปลอดภัย และจะช่วยลดความผิดพลาดที่เกิดจากคนที่ต้องทำงานเป็นเวลานานต่อเนื่องกันทั้งวันได้

อย่างไรก็ตามในปัจจุบันยังไม่มีรูปแบบหรือโมเดลเฉพาะที่จะใช้วิเคราะห์พฤติกรรมมนุษย์ที่ยืนยันได้ว่าสามารถจำแนกการกระทำของมนุษย์ได้ 100% ปัญหาส่วนใหญ่ที่เกิดขึ้นมาจากความซับซ้อนของร่างกายมนุษย์เอง ปัญหามุมมองของภาพที่ใช้ในการวิเคราะห์ หรือแม้กระทั่งปัญหาทางธรรมชาติเรื่องการเปลี่ยนแปลงแสงสว่าง ซึ่งจะทำให้เกิดความผิดพลาดในกระบวนการของการตรวจจับวัตถุที่สนใจ

แนวทางที่ถูกนำมาใช้ในการแก้ไขปัญหาคือการจำแนกพฤติกรรมมนุษย์ในปัจจุบัน ซึ่งมีด้วยกันหลายวิธี โดยวิธีที่นิยมนำใช้กันอย่างแพร่หลาย ตัวอย่างเช่น การพิจารณาพฤติกรรมมนุษย์จากรูปร่าง จากสี จากการเคลื่อนไหว จากความร้อน จากเทมเพลตของแต่ละท่าทาง เป็นต้น ซึ่งวิธีที่ได้กล่าวไว้ข้างต้น สามารถจำแนกพฤติกรรมมนุษย์ได้ผลในระดับหนึ่ง

ในส่วนของวิทยานิพนธ์นี้จะเน้นในเรื่องของการจำแนก ท่าทางมนุษย์ ทั้ง 5 ท่าทางพื้นฐาน คือ การยืน การเดิน การนั่ง การก้ม และการนอน โดยได้นำเสนอการใช้ข้อมูลของความลึก สี และการเคลื่อนไหวของมนุษย์จากภาพในมุมมองสูง ซึ่งจากภาพมุมมองสูงจะช่วยลดความซับซ้อนของร่างกายมนุษย์ทำให้ง่ายต่อการวิเคราะห์ โดยระบบจะ สร้างภาพเสมือนในมุมมองสูงของมนุษย์ ซึ่งสามารถนำมาใช้ประมาณ โครงสร้างของท่าทางมนุษย์สำหรับนำไปจำแนกแต่ละท่าทางของมนุษย์ได้อย่างมีประสิทธิภาพในรูปแบบของเรียลไทม์ (realtime)

## 1.2. งานวิจัยที่เกี่ยวข้อง

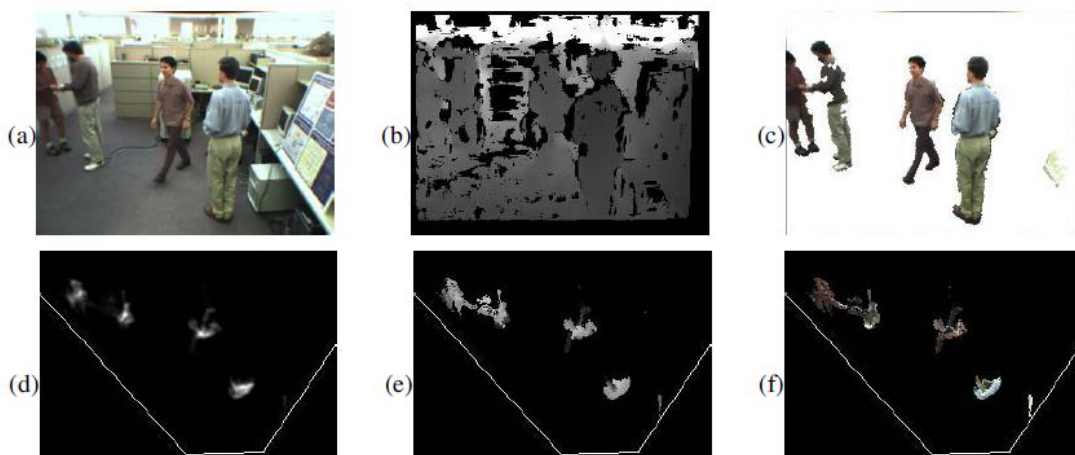
การวิเคราะห์พฤติกรรมมนุษย์เป็นงานวิจัยพื้นฐานสำหรับนำไปประยุกต์ใช้งานทางด้านต่างๆ ได้ ซึ่งมีผู้สนใจที่จะศึกษาและได้นำเสนอรูปแบบการวิเคราะห์พฤติกรรมมนุษย์ในหลายวิธี [1] โดยแต่ละวิธีสามารถสรุปเป็นหัวข้อใหญ่ได้ 2 หัวข้อ คือ การสกัดวัตถุที่สนใจ (foreground extraction) และ การจดจำท่าทางมนุษย์ (action recognition) ในส่วนของการสกัดวัตถุที่สนใจจะเป็นการตรวจจับ (detect) และการติดตาม (tracking) วัตถุที่สนใจ ส่วนการจดจำท่าทางมนุษย์จะเป็นการวิเคราะห์ลักษณะเด่นของมนุษย์เพื่อระบุว่าเป็นการกระทำใด เช่น การยืน การนอน การนั่ง การก้ม การวิ่ง เป็นต้น

### 1.2.1. การสกัดวัตถุที่สนใจ (foreground extraction)

การสกัดวัตถุที่สนใจ เป็นการตรวจจับและติดตามวัตถุที่สนใจจากภาพวิดีโอ สำหรับนำมาใช้ในการประมวลผล โดยวัตถุที่สนใจจะเป็นได้ทั้งสิ่งมีชีวิตและไม่มีชีวิต ซึ่งสามารถเคลื่อนที่และไม่สามารถเคลื่อนที่ได้ วิธีใช้สำหรับการสกัดวัตถุที่สนใจมีจำนวนมากสามารถแบ่งเป็นประเภทได้ตามข้อมูลที่ใช้ในการวิเคราะห์ ได้แก่ การใช้สีสำหรับการวิเคราะห์ โดยการตรวจจับเฉพาะสีที่สนใจวิธีที่นิยมนำมาใช้ได้แก่ CamShift, meanshift, k-mean เป็นต้น การวิเคราะห์จากการเคลื่อนไหว โดยกำหนดให้สิ่งที่มีเคลื่อนไหวเป็นวัตถุที่สนใจ วิธีที่นิยมนำมาใช้ได้แก่ Background Subtraction, Optical Flow เป็นต้น โดยได้มีผู้นำเสนอวิธีการสกัดวัตถุที่สนใจในรูปแบบต่างๆ ดังนี้

Michael Harville [2] ได้นำเสนอวิธีการสกัดวัตถุที่สนใจ โดยจะพิจารณาจากข้อมูลสีและความลึกจากวัตถุที่สนใจ โดยในงานวิจัยของเขาต้องการวิเคราะห์พฤติกรรมมนุษย์จากภาพในมุมมองสูง (plan-view) แต่กล้องถ่ายภาพติดอยู่บริเวณผนังห้อง ดังนั้นจึงจะมี กระบวนการ โพรเจกชันภาพจากด้านหน้าของวัตถุที่สนใจไปยังด้านบนของวัตถุ ก่อนจะนำไปเข้ากระบวนการสกัดวัตถุที่สนใจ

จากภาพประกอบ 1-1 การสกัดวัตถุที่สนใจ ซึ่งในที่นี้คือมนุษย์ ประกอบด้วยอินพุต 2 ส่วน ได้แก่ ข้อมูลสีและความลึกจากภาพต้นฉบับดัง ภาพประกอบ 1-1 (a),(b) จากนั้นใช้เทคนิค Time-Adaptive Per-Pixel Mixture of Gaussians (TAPPMOGs) หาความสัมพันธ์ของสีและความลึกเพื่อสกัดวัตถุที่สนใจออกจากภาพต้นฉบับดัง ภาพประกอบ 1-1 (c) แล้วใช้ฟังก์ชัน ( $O_{raw}$ ) กับความลึก ( $H_{raw}$ ) ของวัตถุที่สนใจสร้างภาพเสมือนมุมมองสูง ดัง ภาพประกอบ 1-1 (d) เป็นภาพมุมมองสูงแบบ Grayscale และภาพประกอบ 1-1 (e) เป็นภาพมุมมองสูงที่ประกอบด้วยความลึกของวัตถุ จากนั้นทำการแมปสี ( $C_{raw}$ ) ให้กับภาพมุมมองสูง ดังภาพประกอบ 1-1 (f)



ภาพประกอบ 1- 1 ภาพประกอบ 1(a) ภาพสี (b) ภาพประกอบด้วยความลึก (c) สกัดวัตถุที่สนใจ (d) ภาพ Grayscale จากมุมมองสูง (e) ภาพมุมมองสูงที่ประกอบด้วยความลึก (f) การแมพสีให้กับภาพมุมมองสูง

Michael Harville [2] ได้นำเสนอวิธีการติดตาม (tracking) วัตถุที่สนใจ ซึ่งเรียกว่า short-term ( $\vec{S}_i$ ) โดย short-term จะเป็นการติดตามมนุษย์ โดยสร้าง template สำหรับภาพมุมมองสูงเพื่อติดตาม template นั้น ประกอบด้วยพารามิเตอร์คือ  $\vec{S}_i = \langle \vec{p}_i, \vec{v}_i, T_{H,i}, T_{C,i} \rangle$ , ซึ่ง  $T_{H,i}$  และ  $T_{C,i}$  คือ ความลึกและสีของ template ตามลำดับ  $\vec{p}_i$  คือ ตำแหน่งจากภาพมุมมองสูง  $\vec{v}_i$  คือ อัตราความเร็วของวัตถุ จากภาพมุมมองสูง ซึ่งแสดงได้ดังภาพประกอบ 1-2



ภาพประกอบ 1- 2 แม่แบบ (templates) สำหรับติดตามบุคคลที่สนใจ

Skulkittiyut Weerachai and Makoto MIZUKAWA [3] ได้นำเสนอวิธีการวิเคราะห์พฤติกรรมมนุษย์โดยใช้ภาพมุมมองสูง (top-view) จากกล้องวิดีโอที่ใช้เลนส์ wide ซึ่งจะได้มุมมองภาพที่กว้างขึ้น โดยใช้เทคนิค background subtraction สำหรับตรวจจับและติดตามวัตถุที่สนใจดังภาพประกอบ 1-3





ภาพประกอบ 1-3 ผลลัพธ์จากเทคนิค background subtraction

จากนั้นสกัดวัตถุที่สนใจ (feature extraction) เพื่อสร้างคุณลักษณะเด่น (feature) สำหรับนำไปจดจำท่าทาง (action recognition) โดยคุณลักษณะเด่นที่เขาใช้ประกอบด้วย 3 ส่วนคือ 1. Position of Human Posture Position เป็นการวิเคราะห์การเปลี่ยนแปลงตำแหน่งของท่าทาง ซึ่งสามารถช่วยระบุการเปลี่ยนพฤติกรรมได้, 2. Height-Width Ratio เป็นอัตราความสูงและความกว้าง ของมนุษย์จากภาพมุมสูงในท่าทางต่างๆ ซึ่งมีส่วนช่วยในการบ่งบอกท่าทางได้ เช่น การนอน จะมี Ratio มากที่สุดดังตาราง 1-1

ตารางที่ 1-1 อัตราค่าเฉลี่ยความสูง-กว้างของท่าทางมนุษย์

Posture	Height-width
Lying	2.750670
Standing	1.386048
Sitting	1.239172

และ 3. DFT Coefficient of Log-polar Histogram เป็นแปลงภาพให้อยู่ใน frequency domain ซึ่งจะช่วยให้ง่ายต่อการวิเคราะห์และจำแนกพฤติกรรมมนุษย์ยิ่งขึ้นดังภาพประกอบ 1-4



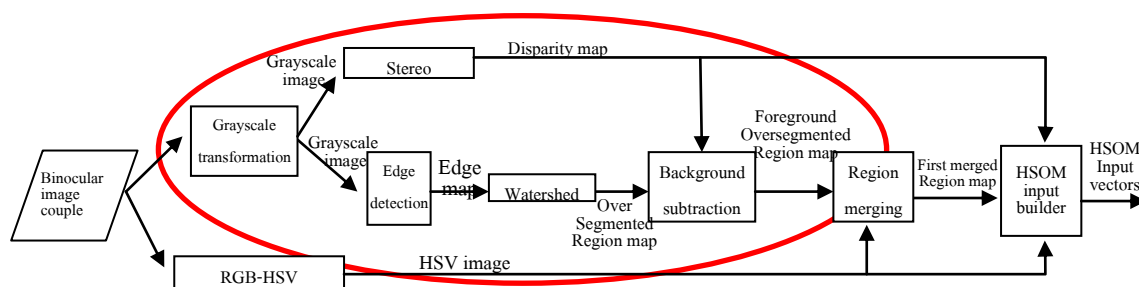
ภาพประกอบ 1-4 Log-polar Histogram of Human Posture

Yong Wang, Tianli Yu, Larry Shi and Zhu Li [4] ได้นำเสนอวิธีการวิเคราะห์ท่าทางมนุษย์ สำหรับใช้ในการควบคุมการเล่นเกมส์ ซึ่งวิธีการสกัดวัตถุที่สนใจจะมีการนำคุณสมบัติของความลึก และจำนวนพิกเซลในบริเวณที่เกี่ยวข้องมาใช้สำหรับการตัดสินใจ โดยเงื่อนไขที่ใช้ในการสกัดวัตถุที่สนใจดังสมการที่ (1.1)

$$C(k) = \begin{cases} \text{hand} & k=0 \text{ and } A(k) < t_h \\ \text{body} & R(k) > T_b, R(i) \leq T_b \text{ when } i < k \\ \text{ignored} & \text{others} \end{cases} \quad (1.1)$$

โดย  $C(k)$  เป็นผลลัพธ์ของการตัดสินใจ โดย  $k$  คือระดับความลึกที่มีมือ และ  $A(k)$  คือจำนวนพิกเซล ถ้า  $k$  คือระดับความลึกลำดับแรกในลิสและมีจำนวนพิกเซลน้อยกว่าค่าเทรชโฮลด์ ( $t_h$ ) สิ่งที่สกัดได้คือมือ แต่ถ้าระดับความลึกแรกของลิสอยู่ในบริเวณที่  $R(k)$  มากกว่าค่าเทรชโฮลด์ ( $T_b$ ) สิ่งที่สกัดได้คือร่างกาย

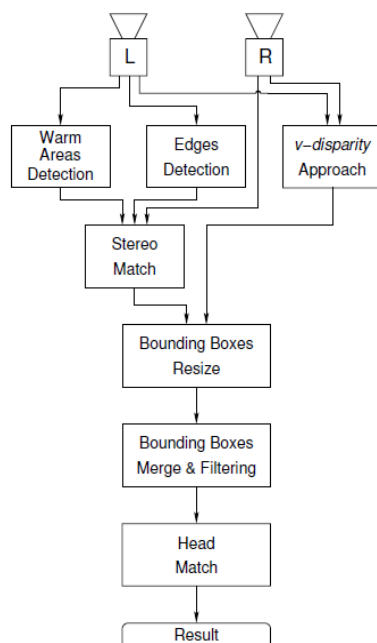
Giovanni Bertolini, Stefano Ramat [5] ได้นำเสนอวิธีการสกัดวัตถุที่สนใจ โดยใช้วิธี background subtraction ซึ่งวิธีนี้จะมีการนำข้อมูลความลึกและขอบ (edge) ที่ผ่านการทำ watershed ซึ่งจะเป็นการ segment วัตถุที่สนใจมาพิจารณาในการทำ background subtraction ด้วย แสดงภาพรวมของระบบในส่วนของการทำ background subtraction ดังภาพประกอบ 1-5



ภาพประกอบ 1- 5 กระบวนการทำ background subtraction (เฉพาะในวงกลม)

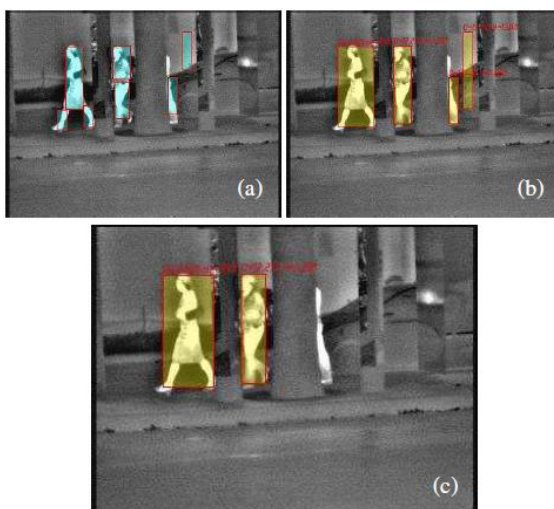
จากภาพประกอบ 1-5 กระบวนการ background subtraction จะอยู่ในวงกลมสีแดง โดยเริ่มจากแปลงภาพให้เป็นภาพขาวเทา (grayscale) จากนั้นหาขอบของวัตถุที่สนใจด้วยวิธี watershed และประมาณความลึกของภาพจากกระบวนการ stereo แล้วนำข้อมูลทั้งสองไปใช้ประมวลผลในกระบวนการ background subtraction เพื่อสกัดวัตถุที่สนใจออกมา

M. Bertozzi, E. Binelli, A. Broggi, M. Del Rose [6] ได้นำเสนอวิธีการสกัดวัตถุที่สนใจ โดยอัลกอริทึมที่ประกอบไปด้วย 3 ความแตกต่าง ได้แก่ warm area detection, edge based detection and v-disparity computation ซึ่งงานวิจัยของเขาต้องการที่จะติดตามคนเดินบนถนน ดังภาพประกอบ 1-6



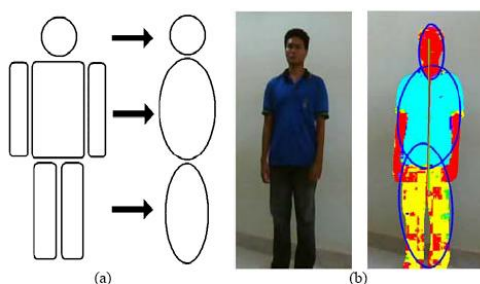
ภาพประกอบ 1- 6 ภาพรวมของระบบ

ภาพประกอบ 1-6 warm areas detection จะเป็นการอาศัยความร้อนของวัตถุเปรียบเทียบกับภาพพื้นหลังเพื่อตรวจจับวัตถุที่สนใจ edges detection จะเป็นการตรวจจับจากขอบของวัตถุโดยใช้ sobel operator และ v-disparity approach จะเป็นการใช้ระยะทางของวัตถุมาพิจารณา ซึ่งทั้งสามวิธีนี้จะมีการสร้าง rectangular bounding box บริเวณจุดที่สนใจทั้งหมด ซึ่งจะสร้างจากการประมาณ bounding box ด้วยระยะทางและตำแหน่งของวัตถุที่สัมพันธ์กัน ซึ่ง v-disparity ประกอบไปด้วยค่าของระยะทางอยู่แล้วแต่ warm area detection, edge based detection จะต้องนำมาผ่าน Stereo Match ก่อน ผลลัพธ์จากกระบวนการที่ผ่านมาจะได้ bounding box หลายๆ ส่วนในบริเวณที่สนใจ จึงต้องนำมาผ่านกระบวนการ resize , merge เพื่อลดขนาดและรวม bounding box ย่อยๆเข้าด้วยกันและกระบวนการ filter เพื่อลบ bounding box ที่เกิดความผิดพลาดจากการตรวจจับ ดังภาพประกอบ 1-7 สุดท้ายจะผ่านกระบวนการ Head Match เพื่อค้นหาหัวของคนจาก bounding box ซึ่งจะทำให้เพิ่มประสิทธิภาพของการตรวจจับคนเดินถนนมากขึ้น



ภาพประกอบ 1- 7 (a) bounding box ของกล้องซ้าย (b) ลดขนาดและรวม bounding box (c) ผลลัพธ์สุดท้าย

Nattapon Noorit, Nikom Suvonvorn, and Montri Karnchanadecha [7] ได้นำเสนอวิธีการสกัดวัตถุที่สนใจ โดยได้ ลดองค์ประกอบที่ซับซ้อน ของร่างกายมนุษย์ ออกไปให้เหลือเพียง 3 องค์ประกอบหลักคือ หัว ลำตัว และขา ซึ่งกระบวนการสร้าง โครงสร้างมนุษย์ขึ้นใหม่ประกอบด้วย 2 ขั้นตอน คือ 1. Motion-texture based Human Detection and Tracking ขั้นตอนนี้จะใช้บริเวณวัตถุเคลื่อนไหว (ภาพพื้นหน้า) ที่ได้จากระบบ การลบภาพพื้นหลัง ซึ่งบริเวณดังกล่าวจะถูกนำไปแยกพื้นที่ออกเป็นส่วน ๆ โดยใช้คุณลักษณะสีที่มีความใกล้เคียงกันและอยู่ติดกันเป็นกลุ่มก่อนจะถูกจัดเป็นพื้นที่วัตถุชิ้นเดียวกัน หลังจากแยกพื้นที่ออกเป็นส่วน ๆ แล้วแต่ละส่วนจะถูกนำไปจับคู่กับโครงสร้างอย่างง่ายของร่างกาย ดังภาพประกอบ 1-8



ภาพประกอบ 1- 8 การสกัดวัตถุที่สนใจของ N.Noorit

(a) การลดความซับซ้อนของมนุษย์ (b) ตัวอย่างโครงร่างใหม่ของโมเดลมนุษย์

Myung-Cheol Roh, Ho-Keun Shin, and Seong-Whan Lee [8] ได้นำเสนอการใช้เทคนิค VMT (Volume Motion Template) ซึ่งจะช่วยให้สำหรับการสร้างเทมเพลตของการเคลื่อนที่แบบสามมิติ เทคนิคนี้จะมีการนำแผนผังดิสพาทิที่เข้ามาใช้ เพื่อจะบ่งบอกข้อมูลความลึกและทิศทาง

เคลื่อนที่ในรูปแบบสามมิติ โดยจะสกัดวัตถุที่สนใจด้วยวิธี background subtraction ซึ่งใช้แผนผังคิสพารีสำหรับการวิเคราะห์ โดยประมาณหาส่วนของวัตถุที่สนใจดังสมการที่ (1.2)

$$O_t(x,y,z) = \begin{cases} 1 & \text{if } S_z(x,y) = 1 \text{ and } Z_t(x,y) = z \\ 0 & \text{otherwise} \end{cases} \quad (1.2)$$

โดยที่  $S_z$  เป็นพิกเซลที่ต้องการ  $Z_t$  เป็นแผนผังคิสพารีที่พิกเซลนั้น แล้วประมาณหาวัตถุที่เคลื่อนที่โดยการนำภาพจากเฟรมที่ต่อเนื่องกัน มาทำการลบกันดังสมการที่ (1.3)

$$\sigma_t(x,y,z) = \left| O_t(x,y,z) - O_{t-1}(x,y,z) \right| \quad (1.3)$$

ซึ่ง VMT คือ เซตของข้อมูลจาก  $\sigma_j$  ที่  $1 < j < t$  ซึ่งจำนวนข้อมูลเหล่านี้จะถูกรวมตัวกันเป็นภาพ อธิบายได้ดังสมการที่ (1.4)

$$V_t(x,y,z) = \begin{cases} I_{\max} & \text{if } \sigma_t(x,y,z) = 1 \\ \max(0, V_{t-1}(x,y,z) - \eta\mu_t) & \text{otherwise} \end{cases} \quad (1.4)$$

ที่  $I_{\max}$  เป็นค่ามากที่สุดของความเข้มสี (intensity) ตัวอย่างเช่น 255 ที่ 8 บิตในภาพขาวเทา  $\eta$  เป็นการลดลงของค่าคงที่  $\mu_t$  เป็นความน่าจะเป็นของวัตถุที่มีการเคลื่อนที่

### 1.2.2. การจดจำท่าทางมนุษย์ (action recognition)

การจดจำท่าทางมนุษย์ เป็นการวิเคราะห์ลักษณะเด่นที่สำคัญ เพื่อที่จะจัดกลุ่มหรือจำแนกท่าทางของมนุษย์ ซึ่งวิธีที่นิยมใช้สำหรับการจัดกลุ่มมีด้วยกันหลายวิธี ตัวอย่างเช่น SVM (Support Vector Machines), HMM (Hidden Markov Models), CRF (Conditional Random Fields), NN(neuron network) เป็นต้น โดยได้มีผู้เสนอวิธีการจดจำท่าทางมนุษย์ที่น่าสนใจไว้ดังนี้

Michael Harville [2] ได้นำเสนอวิธีการจดจำท่าทางมนุษย์ โดยเรียกว่า long-term ( $\vec{L}_i$ ) ซึ่งจะใช้จดจำท่ามนุษย์ระหว่างการติดตาม (tracking) ในภาพมุมมองสูง ซึ่งกำหนดให้  $\vec{L}_i = \langle \vec{L}_{s,i}, \vec{L}_{c,i} \rangle$  โดย long-term จะเป็นการสร้างโมเดลของมนุษย์จากมุมมองสูง ซึ่งจะนำลักษณะเด่นเกี่ยวกับรูปร่าง ( $\vec{L}_{s,i}$ ) ของมนุษย์จากมุมมองสูงและสี ( $\vec{L}_{c,i}$ ) จากภาพมาสร้างโมเดล

Skulkittiyut Weerachai and Prof.Makoto MIZUKAWA [3] นำเสนอวิธีการจดจำท่าทางมนุษย์โดยใช้เทคนิคของ นิวรอนเน็ตเวิร์ก อัลกอริทึม (neuron network algorithm) โดยเขาได้แบ่งท่าทางมนุษย์ออกเป็น 2 กลุ่ม คือ 1. ท่าทางที่หยุดนิ่ง คือ การยืน การนั่ง และการนอน , 2. ท่าทางที่เคลื่อนที่ โดยมี CP (changing of human posture position) เปรียบเทียบกับค่า threshold เพื่อบ่งบอก

การเคลื่อนที่ของวัตถุที่สนใจ อินพุตสำหรับนิเวศน์ตเวริกประกอบด้วย Height-Width Ratio และ DFT Coefficient n Significant เพื่อจำแนกท่าทาง 3 ท่าทาง คือ การยืน การนั่ง และการนอน

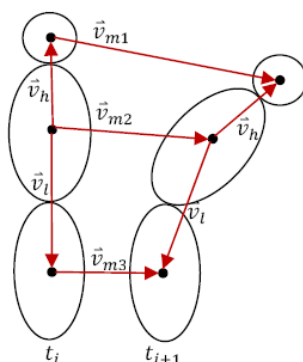
Yong Wang, Tianli Yu, Larry Shi and Zhu Li [4] ได้นำเสนอวิธีการจดจำท่าทางมนุษย์ โดยใช้ Gaussian Hidden Markov Model (GHMM) สำหรับจดจำท่าทางการเล่นเกมส ได้แก่ การหลบ(dodge), การต่อยอย่างรวดเร็ว (jab), การดิ่งศัตรู(hook), การชกหมัดเสยขึ้น (uppercut) ซึ่งลักษณะเด่นที่ใช้คือ ระดับของความลึกที่บ่งบอกถึงร่างกายหรือมือที่ปรากฏในภาพ จะถูกในเรียงจากลำดับการทำงานของ HMM ดังสมการที่ (1.5)

$$G(s) = \arg \max_n \{P(s|\lambda_n)\}, n=0,1,2, \dots, N-1 \quad (1.5)$$

โดยที่  $P(x)$  เป็นความน่าจะเป็นของ  $x$  และ  $\lambda_n$  คือ โมเดลของท่าทาง  $n^{\text{th}}$

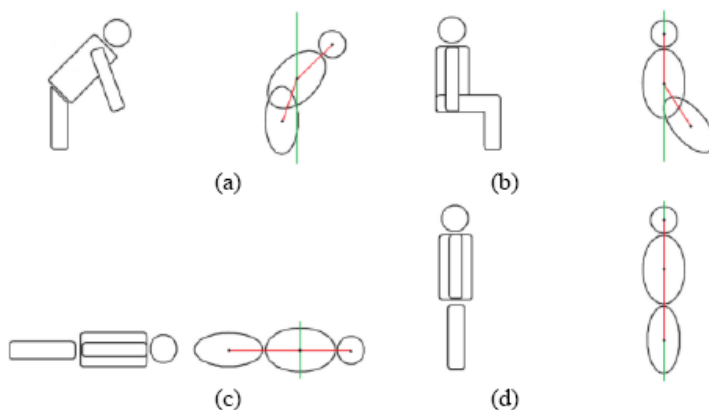
Giovanni Bertolini, Stefano Ramat [5] ได้นำเสนอวิธีการจดจำท่าทางมนุษย์ โดยใช้ hierachial SOM (HSOM) neural network โดยนำข้อมูลของระบบสี HSV, ความลึกของวัตถุในภาพ และวัตถุที่สนใจซึ่งผ่านการทำ foreground extraction มาให้ HSOM พิจารณาเพื่อจดจำท่าทางมนุษย์

Nattapon Noorit, Nikom Suvonvorn, and Montri Karnchanadecha [7] ได้นำเสนอวิธีการจดจำท่าทางมนุษย์ โดยเรียกว่า Parametric Model Definition สามารถอธิบายได้จากภาพประกอบ 1-9 ณ เวลา  $t_i$  ขณะยังไม่มีกรเคลื่อนไหว ประกอบด้วยพารามิเตอร์ที่ใช้พิจารณา 2 กลุ่ม ได้แก่ พารามิเตอร์ภายใน และพารามิเตอร์ภายนอก โดยพารามิเตอร์ภายใน ( $\vec{v}_h, \vec{v}_l$ ) แสดงถึงคุณลักษณะของโครงสร้างมนุษย์ที่อธิบายถึงระยะห่าง ( $\vec{v}_h, \vec{v}_l$ ) และทิศทาง ( $\vec{v}_h, \vec{v}_l$ ) จากองค์ประกอบส่วนลำตัวไปยังศีรษะและขา พารามิเตอร์ภายนอก ( $\vec{v}_{m1}, \vec{v}_{m2}, \vec{v}_{m3}$ ) จะสัมพันธ์กับคุณสมบัติการเคลื่อนไหวจากเฟรมหนึ่งไปสู่เฟรมอื่น ๆ โดยอธิบายถึงลักษณะของความเร็ว ( $\vec{v}_{m1}, \vec{v}_{m2}, \vec{v}_{m3}$ ) และทิศทาง การเคลื่อนไหว ( $\vec{v}_{m1}, \vec{v}_{m2}, \vec{v}_{m3}$ ) ขององค์ประกอบศีรษะ ลำตัว และขา โดยพารามิเตอร์จะสัมพันธ์กับศูนย์กลางของแต่ละองค์ประกอบซึ่งการประมาณค่าโดยใช้ศูนย์กลางเช่นนี้จะช่วยลดสัญญาณรบกวนจากกระบวนการย่อย ๆ เช่น การตรวจหาการเคลื่อนไหว กระบวนการแยกกลุ่มสี กระบวนการติดตามการเคลื่อนที่



ภาพประกอบ 1-9 โมเดลมนุษย์และต้นแปร

เมื่อได้แบบจำลองอย่างง่ายต่อไปจะนิยามและรู้จำท่าทางมนุษย์ ซึ่งสามารถแยกท่าทางพื้นฐานได้ 5 ประเภทได้แก่ การยืน การก้ม การนั่ง การเดิน และการนอนราบ เมื่อพิจารณาท่าทางพื้นฐาน 5 ท่า เทียบเคียงกับแบบจำลองมนุษย์อย่างง่ายจะพบว่าแต่ละท่าทางจะมีลักษณะที่เป็นเอกลักษณ์และสังเกตเห็นความแตกต่างกันอย่างชัดเจน ทำให้เราสามารถที่จะบ่งชี้ความแตกต่างของท่าทางได้อย่างถูกต้อง ดัง ภาพประกอบ 1-10 แสดงท่าทางพื้นฐานและพารามิเตอร์ภายใน เมื่อใช้เกณฑ์ด้านการเคลื่อนที่แบ่งประเภท เราสามารถแบ่งประเภทท่าทางออกเป็น 2 ประเภทได้แก่ ท่าทางสถิต (static action) ซึ่งจะจัดอยู่ในกลุ่มแบบจำลองสถิตและท่าทางพลวัต (dynamic action) ซึ่งจะจัดอยู่ในกลุ่มแบบจำลองพลวัต



ภาพประกอบ 1-10 ท่าทางพื้นฐาน (a) ก้ม (b) นั่ง (c) นอนราบ (d) ยืน หรือ เดิน

### 1.3. วัตถุประสงค์

วิจัยและพัฒนาาระบบเพื่อจำแนกท่าทางของมนุษย์โดยใช้ข้อมูลความลึกและสี

### 1.4. ขอบเขตงานวิจัย

งานวิจัยนี้มุ่งเน้นไปยัง 2 ส่วนได้แก่

1.การศึกษาและพัฒนาระบบการจำแนกพฤติกรรมมนุษย์จากบริเวณที่สนใจ โดยวิเคราะห์จากภาพมุมสูง

2.การศึกษาและพัฒนาข้อมูลความลึกและสี เพื่อจำแนกพฤติกรรมมนุษย์ ทั้ง 5 ท่าทาง พื้นฐาน คือ การยืน , การเดิน , การนั่ง , การก้ม และการนอน ได้อย่างถูกต้อง โดยประมวลผลแบบ Real-Time ได้อย่างมีประสิทธิภาพ

### 1.5. ประโยชน์ที่คาดว่าจะได้รับ

ได้เทคนิคการติดตามการเคลื่อนที่ของวัตถุ พร้อมทั้งข้อมูลพฤติกรรมของวัตถุ ซึ่งสามารถนำไปประยุกต์ใช้กับการค้นหาเหตุการณ์ผิดปกติ ถ้านำไปใช้ร่วมกับระบบกล้องวงจรปิดเพื่อช่วยบรรเทาปัญหาด้านความปลอดภัย จะทำให้สามารถลดโอกาสการเกิดความสูญเสียอันเนื่องมาจากการก่อการร้าย และยังใช้เป็นหลักฐานยืนยันการกระทำผิดได้อีกด้วย หรือจะนำมาใช้สำหรับเฝ้าระวังผู้พิการและคนชราที่ต้องอาศัยอยู่ที่พักอาศัยตามลำพัง เมื่อผู้ดูแลต้องออกไปทำธุระนอกที่พักอาศัย ซึ่งจะช่วยให้ผู้ดูแลรับทราบพฤติกรรมที่ผิดปกติ จากการเกิดอุบัติเหตุและสามารถให้ความช่วยเหลือได้อย่างทันที



## บทที่ 2

### ทฤษฎีและหลักการ

#### 2.1. กระบวนการตรวจจับการเคลื่อนไหว (Motion Detection)

กระบวนการตรวจจับการเคลื่อนไหว คือ ระบบที่ใช้สำหรับการตรวจสอบการเคลื่อนไหวของวัตถุที่ผ่านเข้ามาในกล้องวิดีโอ ซึ่งจะช่วยให้ระบบการเฝ้าระวังมีประสิทธิภาพมากขึ้น เนื่องจากจะช่วยคัดแยกเฉพาะวัตถุที่สนใจสำหรับนำไปวิเคราะห์ และตัดวัตถุที่เป็นพื้นหลังออกไป โดยกระบวนการตรวจจับการเคลื่อนไหวสามารถแบ่งออกเป็น 2 ประเภทใหญ่ๆ คือ 1. การลบพื้นหลัง (Background Subtraction)[9] วิธีนี้เหมาะกับกล้องวิดีโอที่อยู่นิ่ง ซึ่งภาพพื้นหลัง (background) ไม่มีการเปลี่ยนแปลง และ 2. ออปติคัลโฟว์ (Optical Flow) เป็นการประมาณการเปลี่ยนแปลงระหว่างเฟรมรูปภาพ 2 เฟรม โดยสมมติฐานพื้นฐานของการคำนวณเกี่ยวกับการเคลื่อนที่ของส่วนที่เป็นแสงสว่างคือ การป้องกันการสูญเสียค่าความเข้มของพิกเซล ซึ่งก็คือความสดหรือความบริสุทธิ์ของสี ซึ่งตัวสีแต่ละสีที่สะท้อนออกมาต่างกันตามคุณสมบัติของสีนั้นๆ เพื่อเปรียบเทียบกับสีข้างเคียง ซึ่งวิธีนี้เหมาะกับการใช้กล้องวิดีโอที่สามารถเคลื่อนที่ได้

แต่เนื่องจากในงานวิทยานิพนธ์นี้ ได้เลือกใช้กล้องวิดีโอที่อยู่นิ่งติดตั้งบนเพดาน ดังนั้น กระบวนการตรวจจับการเคลื่อนไหวที่เหมาะสมคือ การลบพื้นหลัง ซึ่งมีการนำเสนออยู่หลายเทคนิคแต่จะขอเสนอเทคนิคที่ใช้คือ 1.การตรวจจับการเคลื่อนไหวพื้นฐาน (ความแตกต่างระหว่างเฟรม) และ 2.การสร้างพื้นหลังจากการผสมของระเบียบวิธีเกาส์เซียนหลายรูปแบบ ( Mixture of Gaussian)

##### 2.1.1. การตรวจจับการเคลื่อนไหวพื้นฐาน (ความแตกต่างระหว่างเฟรม)

การตรวจจับการเคลื่อนไหวพื้นฐาน คือ การประมวลผลตรวจหาความแตกต่างของพิกเซลระหว่างภาพใหม่กับภาพเก่าที่มีความแตกต่างกันมากเกินกว่าค่าที่ตั้งไว้หรือไม่ ถ้ามีความแตกต่างกันมากระบบสามารถรับรู้ได้ว่าเกิดการเคลื่อนไหวขึ้น ซึ่งในการตรวจหาการเคลื่อนที่จะใช้วิธีตรวจหาการเคลื่อนที่ของ วัตถุที่สนใจ ซึ่งจะใช้การหาความแตกต่างระหว่างจุดภาพ ( Pixel difference) โดยมีหลักการ คือ เฟรมภาพแต่ละภาพจะถูกนำไปเปรียบเทียบกับภาพอ้างอิงทุกจุดภาพต่อจุดภาพ ถ้าค่าความแตกต่างที่จุดภาพหนึ่งมีค่ามากกว่าค่าขีดเริ่มเปลี่ยน ( threshold) ที่ตั้งไว้จุดภาพนั้นจะถูกตัดสินว่าเป็นส่วนของวัตถุที่มีการเคลื่อนที่ แต่ถ้าค่าความแตกต่างที่จุดภาพนั้นน้อย

กว่าค่าขีดเริ่มเปลี่ยน จุดภาพนั้นก็จะถูกจัดเป็นฉากหลัง ซึ่งสามารถอธิบายในรูปแบบของสมการ (2.1)

$$\left\| \text{frame}_i - \text{frame}_{i-1} \right\| > T \quad (2.1)$$

กำหนดให้

$\text{frame}_i$  คือ ค่า ณ จุดภาพใด ๆ ในภาพที่เวลา  $i$

$\text{frame}_{i-1}$  คือ ค่า ณ จุดภาพใด ๆ ในภาพที่เวลา  $i-1$

$T$  คือ ค่าขีดแบ่งระหว่างการถูกจำแนกเป็นพื้นหลังหรือวัตถุเคลื่อนไหว

### 2.1.2. การสร้างพื้นหลังจากการผสมของระเบียบวิธีเกาส์เซียนหลายรูปแบบ

เทคนิคนี้เป็นการสร้างโมเดลของภาพพื้นหลังที่สามารถปรับค่าให้เป็นปัจจุบันได้ตลอดเวลาให้เหมือนกับพื้นหลังของทุกๆ เฟรมปัจจุบัน ที่มีวัตถุเคลื่อนไหวที่เกิดขึ้นภายในเฟรมนั้นๆ แต่ละจุดสีของเฟรมภาพจะถูกแยกประเภทเป็นจุดสีของภาพพื้นหลัง หรือจุดสีของวัตถุด้วยวิธีเกาส์เซียนดิสทริบิวชันที่มากกว่า 1 ที่มีประสิทธิภาพในการหาโมเดลของภาพพื้นหลังที่มีความซับซ้อน กล่าวคือในสภาพแวดล้อมจริงที่มีการเปลี่ยนแปลงที่ไม่แน่นอนของสภาพแวดล้อม เช่น ความเข้มของแสง เงามาจากต้นไม้ หรือเงาจากอาคารต่างๆ ทำให้ค่าของจุดสีที่ตำแหน่งเดิมของแต่ละลำดับเฟรมภาพเปลี่ยนแปลงตามสภาพอากาศ และสภาพแวดล้อมได้ ทำให้การสร้างโมเดลของภาพพื้นหลังเพียงหนึ่งโมเดลอาจจะไม่เพียงพอกับภาพวิดีโอที่มีความซับซ้อนสามารถอธิบายได้ดังสมการ (2.2), (2.3)

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (2.2)$$

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^2 |\Sigma|^2} e^{-\frac{1}{2} (X_t - \mu)^T \Sigma^{-1} (X_t - \mu)} \quad (2.3)$$

กำหนดให้

$K$  คือ จำนวนของการกระจายของเกาส์เซียน

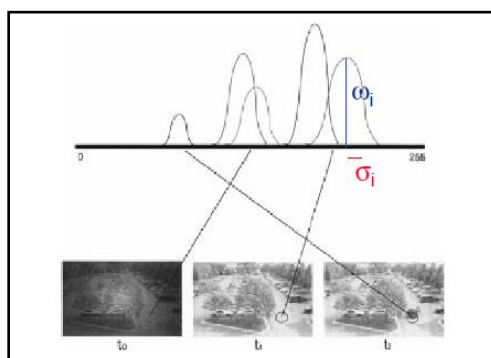
$\omega_{i,t}$  คือ ค่าน้ำหนักของเกาส์เซียนตัวที่  $i^{\text{th}}$  ที่เวลา  $t$

$\mu_{i,t}$  คือ ค่าเฉลี่ยของเกาส์เซียนตัวที่  $i^{\text{th}}$  ที่เวลา  $t$

$\Sigma_{i,t}$  คือ ค่าความแปรปรวนร่วมของเกาส์เซียนตัวที่  $i^{\text{th}}$  ที่เวลา  $t$

$\eta$  คือ Gaussian Probability Density Function

เกณฑ์ในการแยกประเภทจุดสี ว่าเป็นจุดสีของภาพพื้นหลังหรือจุดสีของวัตถุเคลื่อนที่ หาได้จากเมื่อพิจารณาจุดสีค่าแอมพลิจูดสูงสุด จะมีค่าความน่าจะเป็นและมีค่าของส่วนเบี่ยงเบนมาตรฐานดังภาพประกอบ 2-1 ถ้าคิสทริบิวชันที่มีความหนาแน่นมากจะเป็นส่วนของภาพพื้นหลัง ดังสมการ (2.4)

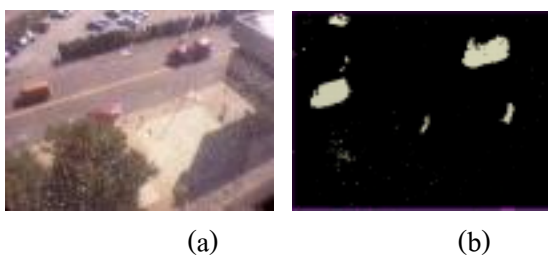


ภาพประกอบ 2- 1 การแยกประเภทจุดสี ของวิธี Mixture of Gaussians [10]

$$\sum_{i=1}^B \omega_i > T \quad (2.4)$$

เมื่อ  $T$  คือ ค่าขีดแบ่ง

จากสมการ (2.4) ถ้าแอมพลิจูดของคิสทริบิวชันมีค่ามากกว่าค่าขีดแบ่งคิสทริบิวชันนั้นจะเป็นส่วนของภาพพื้นหลังและทำการปรับค่าพารามิเตอร์  $(\omega_{i,t}, \mu_{i,t}, \sigma_{i,t})$  ให้เป็นปัจจุบัน แต่ถ้านอกเหนือจากนี้คิสทริบิวชันนั้นจะกลายเป็นส่วนของวัตถุที่เคลื่อนที่ในเฟรมภาพ ตัวอย่างผลจากวิธี Mixture of Gaussians ดังภาพประกอบ 2-2



ภาพประกอบ 2- 2 ผลจากวิธี Mixture of Gaussians [9] (a) ภาพวิดีโอตัดต่อเฟรมปัจจุบัน (b) ภาพวัตถุเคลื่อนที่

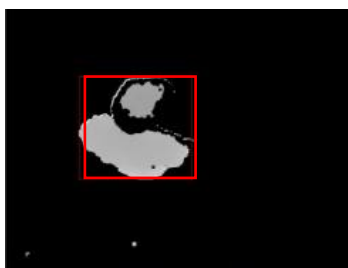
## 2.2. การติดตามวัตถุที่สนใจ

กระบวนการติดตามวัตถุที่สนใจ เป็นการนำผลลัพธ์ที่ได้จากกระบวนการตรวจจับวัตถุที่เคลื่อนที่มาวิเคราะห์ เพื่อจะได้เฉพาะข้อมูลบริเวณหนึ่งที่สนใจ เช่น ตำแหน่ง สี ความลึก ของวัตถุ รูปแบบการติดตามวัตถุที่สนใจมีผู้นำเสนอด้วยกันหลายวิธี เช่น ติดตามจากรูปร่างวัตถุ ได้แก่ วัตถุ

รูปทรงกลม สีเหลือง สามเหลี่ยม เป็นต้น ติดตามจากสี ได้แก่ ลูกบาศก์สีแดง ลูกเทนนิสสีเขียว เป็นต้น อัลกอริทึมที่นิยมใช้สำหรับการติดตามวัตถุคือ mean shift , blob tracking ซึ่งในวิทยานิพนธ์นี้ขอนำเสนอการติดตามวัตถุที่สนใจด้วย blob tracking

### 2.2.1. การวิเคราะห์กลุ่มของจุดภาพที่เป็นสีขาวดำ (Blob analysis)

การวิเคราะห์กลุ่มของจุดภาพที่เป็นสีขาวดำ เป็นการแยกกลุ่มของจุดภาพที่แตกต่างกันด้วยการคำนวณหาตำแหน่งและขนาดของวัตถุ แล้วทำการสร้างบล็อกลีเหลือง เพื่อตัดเฉพาะบริเวณที่สนใจ ซึ่งจากบล็อกลีเหลืองนี้เราจะสามารถรู้ จุดกึ่งกลางของวัตถุ ความกว้าง ความยาวของบล็อกลีเหลืองนำไปใช้ในการวิเคราะห์ต่อไป ดังภาพประกอบ 2-3



ภาพประกอบ 2- 3 ตัวอย่างการทำ blob tracking

### 2.3. การประมวลผลภาพกับรูปร่างและโครงสร้างของภาพ (Morphological Image Processing)

Morphological Image Processing เป็นการประมวลผลภาพโดยการเปลี่ยนแปลงลักษณะรูปร่างหรือโครงสร้างของภาพ โอเปอเรชันพื้นฐานโดยทั่วไปได้แก่ การ Dilation Erosion และ Skeleton โดยการ Dilation คือการขยายภาพโดยมีสัดส่วนเท่ากันทั่วทั้งภาพ (Uniform) การ Erosion คือการย่อภาพ ส่วนการทำ Skeleton เป็นการหาโครงสร้างหลักของวัตถุ นอกจากโอเปอเรชันพื้นฐานดังที่ได้กล่าวข้างต้นแล้วยังมีโอเปอเรชันอื่น ๆ อีกได้แก่การ Opening และ Closing เป็นต้น

#### 2.3.1. เทคนิคของการ Hit และ Miss

โอเปอเรชันพื้นฐานสำหรับการกระทำกับรูปร่างหรือโครงสร้างของภาพ ไม่ว่าจะเป็นการย่อหรือการขยายภาพ จำเป็นที่จะต้องมีการนำเอาเทคนิคการ Hit และ Miss มาใช้ โดยแนวคิดนี้คือการกำหนดให้มีเมตริก (Template) ที่มีขนาดเล็กและเป็นจำนวนคี่ (โดยทั่วไปจะมีค่าเท่ากับ 3x3) ไปเปรียบเทียบกับข้อมูลภาพ โดยการเปรียบเทียบจะทำตลอดทั้งภาพตั้งแต่ต้นภาพจนถึงท้ายภาพ ถ้าข้อมูลของภาพมีลักษณะเหมือนกับเมตริกดังกล่าวเอาพุทที่ได้จะขึ้นอยู่กับพิกเซลที่เป็นศูนย์กลางของเมตริกซึ่งจะถูกกำหนดให้เป็นค่าตามต้องการ (1 หรือ 0) แต่ถ้าข้อมูลในเมตริกไม่เหมือนกับข้อมูลภาพข้อมูลเอาพุทที่ได้จะมีค่าตรงกันข้าม

### 2.3.2. โอเปอเรชันพื้นฐานสำหรับรูปร่างหรือโครงสร้างพื้นฐาน

โอเปอเรชันพื้นฐานสำหรับรูปร่างหรือโครงสร้างพื้นฐานสามารถพิจารณาได้จากตัวอย่างข้อมูลภาพดังภาพประกอบ 2-4

$$\begin{array}{ccccc} 1 & * & 1 & * & 1 \\ * & 1 & * & 1 & * \\ 1 & * & 1 & * & 1 \end{array}$$

ภาพประกอบ 2- 4 ตัวอย่างข้อมูลภาพ

จากข้อมูลภาพดังภาพประกอบ 2-4 แสดงถึงพิกเซลที่ทราบค่าเท่ากับ 1 และพิกเซลที่ไม่ทราบค่าแสดงด้วย \* ซึ่งสามารถแทนด้วยเซตในทางคณิตศาสตร์ได้ดังนี้คือ  $\{(0,0),(0,2), (0,4), (1,1), (1,3), (2,0), (2,2), (2,4)\}$  จากข้อมูลที่ได้สามารถกำหนดให้มีข้อมูลภาพสำหรับการทำโอเปอเรชันได้ดังภาพประกอบ 2-5

$$A = \begin{array}{ccccc} 1 & * & 1 & * & 1 \\ * & 1 & * & 1 & * \\ 1 & * & 1 & * & 1 \end{array} \quad B = \begin{array}{ccccc} * & * & * & 1 & 1 \\ * & * & * & 1 & 1 \\ * & * & * & 1 & 1 \end{array}$$

ภาพประกอบ 2- 5 ตัวอย่างข้อมูลภาพสำหรับทำโอเปอเรชัน

โดยสามารถนำข้อมูลภาพ A และ B มาทำโอเปอเรชันได้ดังนี้

$$A \text{ ยูเนียน } B = \begin{array}{ccccc} 1 & * & 1 & 1 & 1 \\ * & 1 & * & 1 & 1 \\ 1 & * & 1 & 1 & 1 \end{array}$$

$$A \text{ อินเตอร์เซกชัน } B = \begin{array}{ccccc} * & * & * & * & 1 \\ * & * & * & 1 & * \\ * & * & * & * & 1 \end{array}$$

### 2.3.3 การขยายภาพ (Dilation)

การขยายภาพสามารถพิจารณาสำหรับข้อมูลภาพที่เป็นแบบไบนารีโดยการใช้เทคนิคการ Hit และ Miss การขยายภาพจะทำได้โดยกำหนด Template (ซึ่งสามารถสร้างได้จาก \* และ 1 โดยมีจุดเริ่มต้นที่กำหนดโดยวงกลมรอบข้อมูลภาพ ) และนำ Template นี้สแกนไปบนข้อมูลภาพตามลำดับตลอดทั้งภาพซึ่งในขณะที่จุดเริ่ม (Origin) ของ Template ตรงกับตำแหน่งข้อมูลภาพที่พิกเซลมีค่าเท่ากับ 1 นั่นก็จะทำการยูเนียน Template เข้ากับข้อมูลภาพดังตัวอย่างต่อไปนี้





ซึ่งจากผลลัพธ์ของการเปลี่ยน template สรุปได้ว่า Template  $\begin{matrix} 1 & * \\ 1 & 1 \end{matrix}$  สามารถย่อขนาดของภาพได้มากกว่า template  $\begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix}$  เพราะฉะนั้นการเลือก template ที่เหมาะสมกับภาพต้นฉบับเป็นสิ่งที่สำคัญในการย่อภาพหรือการขยายภาพ

### 2.3.5. โอเปอเรชัน Closing และ Opening

#### โอเปอเรชันการ Open

โดยกำหนดให้ OPEN (I, T) เป็นการกระทำ Opening ของภาพ I โดยใช้ Template (T) ซึ่งมีลักษณะดังสมการ(2.5)

$$\text{OPEN (I, T)} = D(E(I)) \quad (2.5)$$

จากสมการ (2.5) จะได้ว่าการทำโอเปอเรชัน OPEN คือการนำข้อมูลภาพ I ผ่านการขยายภาพ(Erosion) แล้วตามด้วยการย่อภาพ (Dilation) โดยใช้ Template ชุดเดียวกันคือ T ซึ่งตัวอย่างการทำโอเปอเรชัน OPEN ดังภาพประกอบ 2-1

0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	4
0	0	0	0	1	2	4	5	
0	0	1	1	2	5	5	4	1
0	0	0	0	0	1	2	3	1
0	0	0	0	0	0	1	3	
0	0	0	0	0	1	2	3	
0	0	0	0	1	2	2	3	

(ก)

(ข)

0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	4
0	0	0	0	1	2	4	4	
0	0	0	0	1	2	4	4	
0	0	0	0	0	1	2	3	
0	0	0	0	0	0	1	3	
0	0	0	0	0	1	2	3	
0	0	0	0	0	1	2	3	

(ค)

ภาพประกอบ 2- 10 ตัวอย่างการทำโอเปอเรชัน OPEN (ก) ภาพต้นฉบับ (ข) template



(ค) ผลลัพธ์การทำโอเปอเรชั่น OPEN

### โอเปอเรชั่นการ Close

กำหนดให้ CLOSE (I, T) เป็นการกระทำแบบ Closing ของภาพ I โดยใช้ Template T ซึ่งมีลักษณะดังสมการ (2.6)

$$\text{CLOSE (I, T)} = \text{E(D(I))} \quad (2.6)$$

จากสมการ (2.6) จะได้ว่าการทำโอเปอเรชั่น CLOSE คือการนำข้อมูลภาพ I ผ่านการย่อภาพ (Dilation) แล้วตามด้วยการขยายภาพ (Erosion) โดยใช้ Template ชุดเดียวกันคือ T ซึ่งตัวอย่างการทำโอเปอเรชั่น CLOSE ดังภาพประกอบ 2-11

0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	1	4	
0	0	0	0	1	2	4	5		
0	0	1	1	2	5	5	4		1
0	0	0	0	0	1	2	3		1
0	0	0	0	0	0	1	3		
0	0	0	0	0	1	2	3		
0	0	0	0	1	2	2	3		
				(ก)					(ข)

0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	4
0	0	0	0	1	2	4	5	
0	0	1	1	2	5	5	4	
0	0	0	0	0	1	2	3	
0	0	0	0	0	1	2	3	
0	0	0	0	0	1	2	3	
0	0	0	0	1	2	2	3	

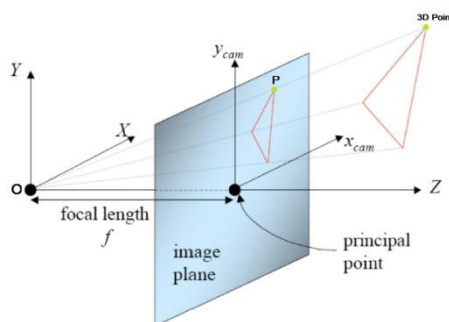
(ค)

ภาพประกอบ 2- 11 ตัวอย่างการทำโอเปอเรชั่น CLOSE (ก) ภาพต้นฉบับ (ข) template

(ค) ผลลัพธ์การทำโอเปอเรชั่น CLOSE

## 2.4. แบบจำลองของกล้อง (Camera model)

แบบจำลองของกล้องที่นิยมใช้กันโดยทั่วไปจะเป็นแบบจำลองกล้องรูเข็ม (Pin-hole Camera Model) โดยประกอบด้วยจุดศูนย์กลางการฉาย O (center of projection), ระนาบรับภาพ (image plane), จุดพิกัดใน 3 มิติ (3D Point) และจุดบนระนาบรับภาพ P ดังภาพประกอบ 2-12 โดยการประมาณพิกัดของจุดบนระนาบรับภาพ P สามารถคำนวณได้จากหลักการของสามเหลี่ยมคล้าย ดังสมการ (2.7) (2.8)



ภาพประกอบ 2- 12 แบบจำลองของกล้องรูเข็ม

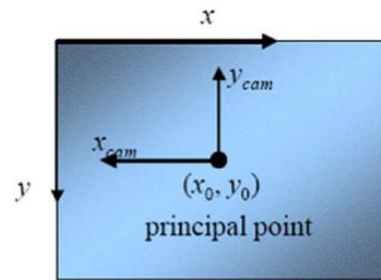
$$x_{cam} = f \frac{x}{z} \quad (2.7)$$

$$y_{cam} = f \frac{y}{z} \quad (2.8)$$

โดย  $x_{cam}$ ,  $y_{cam}$  เป็นพิกัดของจุดบนระนาบภาพในแนวแกน x และแกน y, f คือ ความยาวโฟกัส ซึ่งสามารถเขียนจุดพิกัดใน 3 มิติที่ถูกแปลงเป็น 2 มิติบนระนาบภาพได้ใหม่ในรูปแบบของพิกัดเอกพันธ์ (homogeneous coordinate) ซึ่งการแปลงพิกัดด้วยผลการแปลงเชิงเส้นนี้จะช่วยแก้ปัญหาการที่ระบบพิกัดของโลกภายนอกกับตัวกล้องมักไม่สอดคล้องกันโดยตรงดังสมการ (2.9)

$$\begin{bmatrix} x_{cam} \\ y_{cam} \\ f \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.9)$$

เมื่อได้พิกัดของจุดบนระนาบภาพแล้ว แต่เนื่องจากแกนของภาพทั่วไปกับแกนของพิกัดของจุดบนระนาบภาพมีความแตกต่างกันดังภาพประกอบ 2-13 จึงต้องมีการปรับเพื่อให้แกนทั้ง 2 ภาพตรงกัน ดังสมการ (2.10)



ภาพประกอบ 2- 13 ความแตกต่างระหว่างตำแหน่งแกนของพิกัดจุดบนระนาบกับพิกัดบนภาพ  
ทั่วไป

$$x = -k_x x_{cam} + x_0 \quad (2.10)$$

$$y = -k_y y_{cam} + y_0 \quad (2.11)$$

โดยกำหนดให้  $\alpha_x = -fk_x$  และ  $\alpha_y = -fk_y$  ซึ่งสมการ (2.10), (2.11) สามารถแปลงเป็นพิกัดเอกพจน์ได้ดังสมการ (2.12)

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{f} \begin{bmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{cam} \\ y_{cam} \\ f \end{bmatrix} \quad (2.12)$$

$$\begin{bmatrix} x_{cam} \\ y_{cam} \\ f \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} \quad (2.13)$$

$$\begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} = \begin{bmatrix} R^T & -t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.14)$$

ซึ่งสมการ (2.13) เป็นสมการที่ใช้สำหรับการหาพิกัดของจุดบนระนาบภาพ และสมการ (2.14) เป็นสมการการหาพิกัด 3 มิติของกล้อง โดยการนำตำแหน่ง 3 มิติของวัตถุจริงคูณกับค่าการหมุน (R) และการเคลื่อนที่ของตำแหน่งพิกัด (t) เพื่อความเข้าใจที่ง่ายขึ้นสามารถย่อสมการได้ดังสมการที่ (2.15)(2.16)(2.17)

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R^T & -t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.15)$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} [R^T | -t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.16)$$

$$x = K [R^T | -t] X \quad (2.17)$$

จากสมการ (2.17) K เป็นเมทริกซ์การสอบเทียบกล้อง (camera calibration matrix) เป็นเมทริกซ์ขนาด 3x3 ที่มีความสัมพันธ์โดยตรงกับพารามิเตอร์ภายในทั้ง 4 ตัวของกล้องโดยที่

$$K = \begin{bmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

โดย  $\alpha_x, \alpha_y$  เป็นค่าคงที่ซึ่งเกี่ยวข้องกับคุณสมบัติของเลนส์และความละเอียดของอุปกรณ์รับรูปภาพ,  $x_0, y_0$  เป็นตำแหน่งจุดศูนย์กลางของภาพซึ่งคือจุดกำเนิดของแกนอ้างอิงภาพ และเป็นจุดตัดของเลนส์मुखสำคัญของเลนส์ (optical axis) กับระนาบรับภาพ

$[R^T | -t]$  เป็นพารามิเตอร์ภายนอกมีขนาด 3x4 โดยแบ่งเป็นส่วนของการหมุนและส่วนของการเลื่อนที่ ซึ่งมีองศาอิสระเท่ากับ 6 (องศาอิสระการเลื่อนเท่ากับ 3 และองศาอิสระการหมุนเท่ากับ 3) เมทริกซ์กล้อง จะมีองศาอิสระรวม 11 จะสอดคล้องกับการที่เมทริกซ์กล้องมีสมาชิก 12 ตัวซึ่งมีค่าขึ้นอยู่กับสเกล (up to scale) ดังนั้นองศาจะเหลือเพียง 11 เท่านั้น และ X เป็นพิกัด 3 มิติ

ซึ่งจากสมการนี้สามารถนำมาใช้สำหรับการปรับปรุงและแก้ปัญหาค่าความผิดเพี้ยนของกล้องที่เกิดจากการ misalignment ของตัวเซ็นเซอร์กับศูนย์กลางของเลนส์และความผิดเพี้ยนของตัวเองในขั้นตอนของการผลิตและการประกอบกล้อง ซึ่งสามารถแบ่งความผิดเพี้ยนของเลนส์เป็น 3 ชนิดดังนี้

1. ความผิดเพี้ยนในแนวรัศมี (Radial Distortion) เกิดจากความบกพร่องของเลนส์ทำให้จุดภาพมีการเคลื่อนที่เข้าหรือออกมาข้างนอกจากตำแหน่งที่ถูกต้อง ซึ่งสามารถแสดงให้อยู่ในรูปสมการดังสมการ (2.18) (2.19) (2.20)

$$x_{rd} = a_1 x_d p^2 + a_2 x_d p^4 \quad (2.18)$$

$$y_{rd} = a_1 y_d p^2 + a_2 y_d p^4 \quad (2.19)$$

$$p = \sqrt{X_d^2 + Y_d^2} \quad (2.20)$$

เมื่อ  $x_{rd}, y_{rd}$  คือความผิดเพี้ยนในแนวรัศมีตามแนวแกน x และ y

$a_1, a_2$  คือสัมประสิทธิ์ความผิดเพี้ยน

$X_d, Y_d$  คือพิกัดของจุดภาพบนระนาบภาพในแนวแกน x และ y

2. ความผิดเพี้ยนไม่ตรงศูนย์กลาง (Decentering Distortion) เกิดจากศูนย์กลางทางเดินแสงของเลนส์ไม่ได้อยู่ที่จุดเดียวกัน ดังสมการ (2.21)

$$x_{dd} = 2b_1 X_d Y_d + b_2 (3X_d^2 + Y_d^2) \quad (2.21)$$

$$y_{dd} = b_1 (X_d^2 + 3Y_d^2) + 2b_2 X_d Y_d \quad (2.22)$$

เมื่อ  $x_{dd}, y_{dd}$  คือความผิดเพี้ยนไม่ตรงศูนย์กลางในแนวแกน x และ y

$b_1, b_2$  คือสัมประสิทธิ์ความผิดเพี้ยน

2. ความผิดเพี้ยนจากความหนาของเลนส์ (Thin Prism Distortion) ดังสมการ(2.23)(2.24)

$$x_{pd} = c_1 p^2 \quad (2.23)$$

$$y_{pd} = c_2 p^2 \quad (2.24)$$

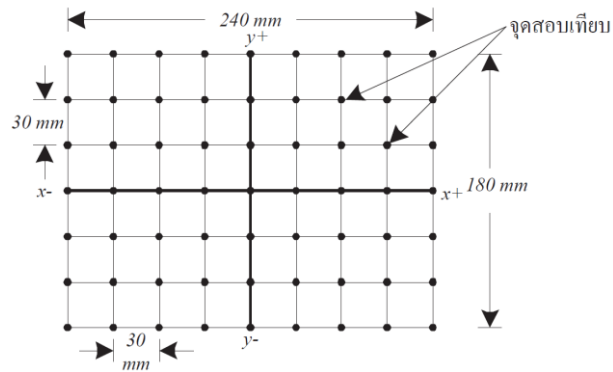
เมื่อ  $x_{pd}, y_{pd}$  คือความผิดเพี้ยนจากความหนาของเลนส์ในแนวแกน x และ y

$c_1, c_2$  คือสัมประสิทธิ์ความผิดเพี้ยน

ดังนั้นจึงนำไปสู่กระบวนการสอบเทียบกล้อง เพื่อจะแก้ไขปัญหาความผิดเพี้ยนของกล้องด้วยการคำนวณภาพใหม่ที่มีความถูกต้องมากขึ้น

### การสอบเทียบกล้อง (Camera Calibration)

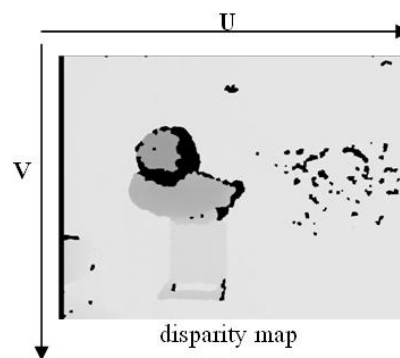
ในการสอบเทียบสิ่งที่ต้องการหาคือพารามิเตอร์ภายในและภายนอกของกล้อง การสอบเทียบทำได้โดยการถ่ายภาพ Calibration Chart ดังภาพประกอบ 2-14 เนื่องจากจุดสอบเทียบทุกจุดบน Calibration Chart มีระยะห่างทั้งแนวตั้งและแนวนอนเท่ากับทุกๆจุด เมื่อใช้กล้องที่ต้องการสอบเทียบถ่ายภาพ Calibration Chart จะได้ระยะห่างระหว่างจุดสอบเทียบทั้งแนวนอนและแนวตั้งใกล้เคียงหรือเท่ากับที่บริเวณจุดศูนย์กลางของภาพ แต่จะต่างกันมากขึ้นเมื่อจุดสอบเทียบอยู่ที่เส้นรอบวงภายนอก ด้วยเหตุนี้จะสมมุติได้ว่าระยะทางระหว่างจุดสอบเทียบที่อยู่ใกล้บริเวณศูนย์กลางของภาพเป็นค่าที่ถูกต้องและนำค่าระยะทางระหว่างจุดสอบเทียบที่นำมาคำนวณหาพิกัดของจุดสอบเทียบอื่นๆได้ ซึ่งจุดพิกัดของจุดสอบเทียบที่คำนวณได้นี้เป็นพิกัดภาพของจุดสอบเทียบบนระนาบภาพที่ไม่มีความผิดพลาด พิกัดของจุดสอบเทียบเหล่านี้จะถูกนำมาใช้เพื่อคำนวณหาค่า element ของเมทริกซ์การหมุน element ของ Translation Vector และความยาวโฟกัส f



ภาพประกอบ 2- 14 Calibration Chart

## 2.5. การแยกวัตถุจากการแผนผังดิสพาริตี

การแยกวัตถุจากแผนผังดิสพาริตีที่มีผู้นำเสนออยู่หลายวิธีด้วยกัน เช่น การใช้งาน Euclidean [11][12] และ V-Disparity [11][13] สำหรับ Euclidean นั้นเป็นการใช้งานโครงสร้างของข้อมูลในพิกัด 3 มิติ ซึ่งไม่สามารถนำมาใช้ประมวลผลภาพในวิทยานิพนธ์นี้ได้ เนื่องจากผลลัพธ์ของความลึกที่ได้จากกล้อง kinect มีลักษณะเป็นแผนผังดิสพาริตี ดังนั้นวิทยานิพนธ์นี้จึงสนใจการใช้งาน V-Disparity เพื่อสร้างภาพประมาณระยะทางของวัตถุ ซึ่งแบ่งออกเป็น 2 แกน คือ แกน U และ แกน V ดังภาพประกอบ 2-15 เพื่อให้ครอบคลุมทิศทางการเคลื่อนที่ของมนุษย์



ภาพประกอบ 2- 15 แกน U และ V สำหรับดิสพาริตี

### การประมาณระยะทางของวัตถุจากแผนผังดิสพาริตี (UV-disparity)

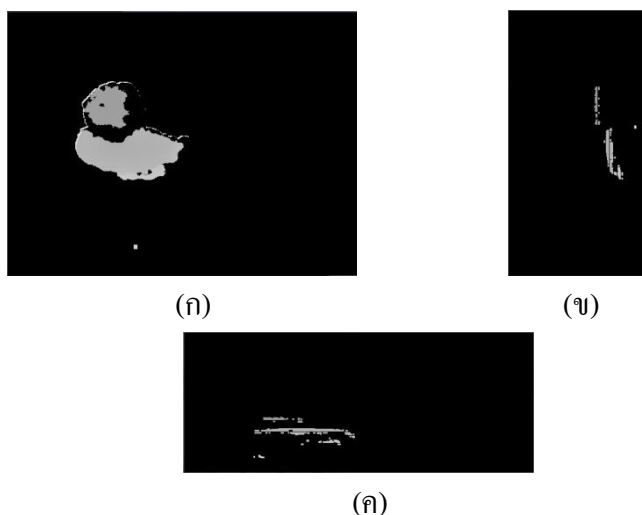
V-disparity คือข้อมูลที่ประมวลผลมาจากแผนผังตามแนวแกน V (แนวตั้ง) โดยให้  $I_{\Delta}$  คือแผนผังดิสพาริตีที่ได้จากการประมวลผลภาพสเตอริโอ H เป็นฟังก์ชันของตัวแปร  $I_{\Delta}$  ดังนั้นจะได้สมการที่ (2.25) ดังนี้

$$H(I_{\Delta})=I_{\Delta_v} \quad (2.25)$$

$I_{\Delta_v}$  เรียกว่า V-disparity โดยที่ H คือการสะสมจำนวนคิสพาริตีที่มีค่าคิสพาริตีเดียวกันที่ปรากฏอยู่บนรูปภาพแถวที่ i เช่นเดียวกับ U-disparity คือข้อมูลที่ประมวลผลมาจากแผนผังคิสพาริตีตามแนวแกน U และ V เป็นฟังก์ชันของ  $I_{\Delta}$  ดังนั้นจะได้สมการที่ (2.26) ดังนี้

$$V(I_{\Delta})=I_{\Delta_u} \quad (2.26)$$

$I_{\Delta_u}$  เรียกว่า U-disparity โดยที่ V คือการสะสมจำนวนคิสพาริตี ที่มีค่าคิสพาริตีเดียวกันที่ปรากฏอยู่บนรูปภาพคอลัมน์ที่ j ซึ่งการใช้งานฟังก์ชัน H และ V เพื่อแปลงแผนผังคิสพาริตีที่ดังภาพประกอบ 2-16

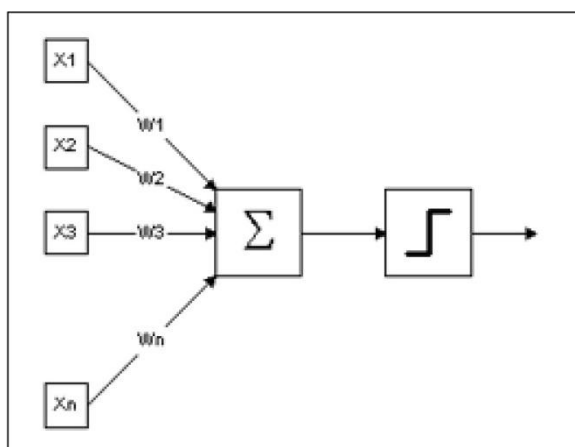


ภาพประกอบ 2- 16 การสร้างภาพ UV-disparity (ก) ภาพต้นฉบับ (ข) ภาพ V-Disparity  
(ค) ภาพ U-Disparity

## 2.6. โครงข่ายประสาทเทียม (Artificial Neural network)

โครงข่ายประสาทเทียม (Artificial Neural network) คือ การสร้างคอมพิวเตอร์ที่มีการจำลองการทำงานของสมองมนุษย์ หรือทำให้คอมพิวเตอร์รู้จักคิดจดจำในแนวเดียวกับโครงข่ายประสาทของมนุษย์ เพื่อช่วยให้คอมพิวเตอร์ฟังภาษามนุษย์ได้เข้าใจ อ่านออก และรู้จำได้ ซึ่งอาจจะเรียกได้ว่าเป็น “สมองกล” โครงสร้างของโครงข่ายประสาทเทียม ประกอบด้วย input units, output units การทำงานของข่ายประสาทเทียม คือเมื่อมี input เข้ามายัง network ก็เอา input มาคูณกับ weight ของแต่ละขา ผลที่ได้จาก input ทุกๆขาของ neuron จะเอามารวมกันแล้วเอามาเทียบกับ threshold ที่กำหนดไว้ ถ้าผลรวมมีค่ามากกว่า threshold แล้ว neuron ก็จะส่ง output ออกไป output

นี่ก็จะถูกส่งไปยัง input ของ neuron อื่นๆ ที่เชื่อมกันใน network ถ้าค่าน้อยกว่า threshold ก็จะไม่เกิด output ดังภาพประกอบ 2-17



ภาพประกอบ 2- 17 โมเดลของข่ายประสาทเทียม

สิ่งสำคัญที่สุดในการทำงาน คือการทราบค่า weight และ threshold สำหรับใช้ในการรู้จำของคอมพิวเตอร์ ซึ่งเป็นค่าที่ไม่แน่นอน แต่สามารถกำหนดให้คอมพิวเตอร์ปรับค่าเหล่านั้นได้โดยการสอนให้มันรู้จัก pattern ของสิ่งที่มันต้องการรู้จำ ซึ่งเรียกว่า back propagation ซึ่งเป็นกระบวนการย้อนกลับของการรู้จำ

การสร้างการเรียนรู้สำหรับ neural network เพื่อให้มีความคิดเสมือนมนุษย์ มีสองวิธี คือ 1. Supervised learning การเรียนแบบมีการสอน คือ เป็นการเรียนแบบที่มีการตรวจคำตอบเพื่อให้วงจรข่ายปรับตัว ชุดข้อมูลที่ใช้สอนวงจรข่ายจะมีคำตอบไว้คอยตรวจสอบว่าวงจรข่ายให้คำตอบที่ถูกต้องหรือไม่ ถ้าตอบไม่ถูก วงจรข่ายจะปรับตัวเพื่อให้ได้คำตอบที่ดีขึ้น เปรียบเทียบกับคนเหมือนเป็นการ

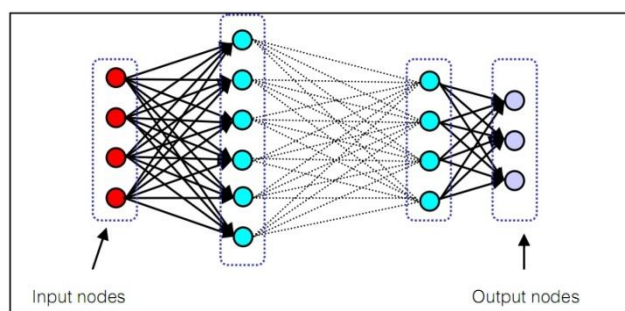
สอนนักเรียน โดยมีครูผู้สอนคอยแนะนำ 2.Unsupervised learning การเรียนแบบมีการสอน คือ เป็นการเรียนแบบไม่มีผู้แนะนำ ไม่มีการตรวจคำตอบว่าถูกหรือผิด วงจรข่ายจะจัดโครงสร้างด้วยตัวเองตามลักษณะของข้อมูล ผลลัพธ์ที่ได้ วงจรข่ายจะสามารถจัดข้อมูลได้ เปรียบเทียบกับคน เช่น การที่เราสามารถแยกแยะพันธุ์พืช พันธุ์สัตว์ตามลักษณะรูปร่างของมันได้เองโดยไม่มีใครสอน

สถาปัตยกรรมโครงข่าย แบ่งเป็น 4 แบบ คือ Feedforward network, Feedback network, Network layer และ Perceptrons ซึ่งในวิทยานิพนธ์นี้สนใจจะใช้สถาปัตยกรรมแบบ feedforward network



### Feedforward network

ข้อมูลที่ประมวลผลในวงจรข่ายจะถูกส่งไปในทิศทางเดียวจาก input nodes ส่งต่อมาเรื่อยๆ จนถึง output nodes โดยไม่มีการย้อนกลับของข้อมูล หรือแม้แต่ nodes ใน layer เดียวกันก็ไม่มี การเชื่อมต่อกัน ดังภาพประกอบ 2-18



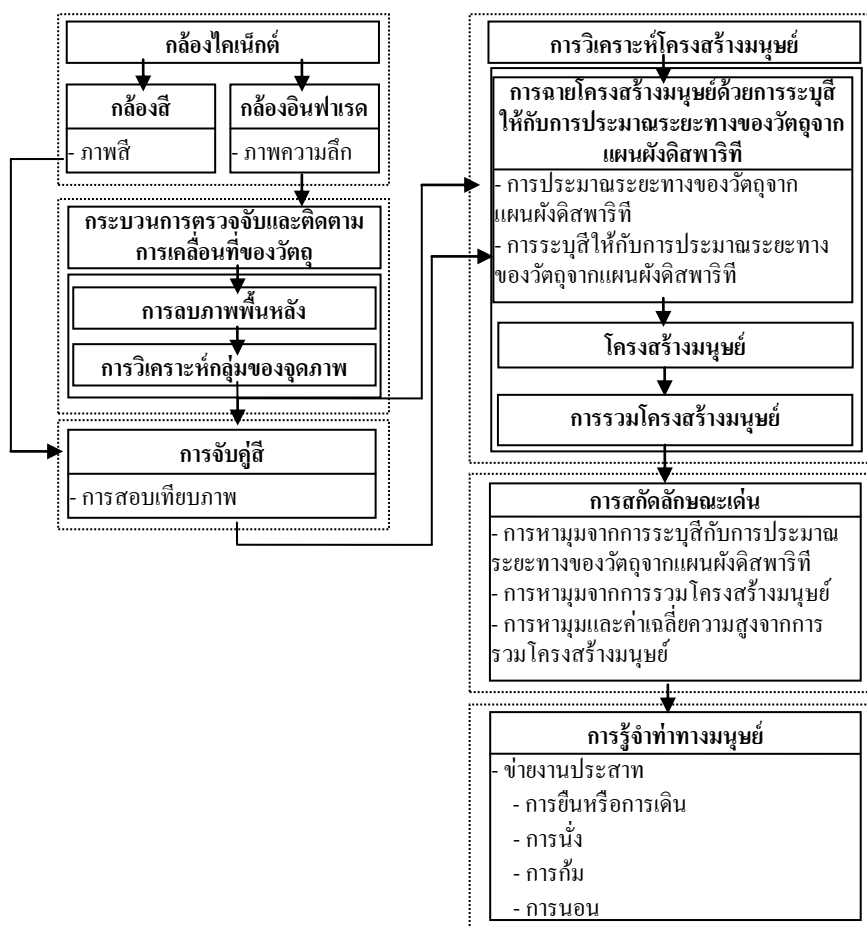
ภาพประกอบ 2- 18 แสดงสถาปัตยกรรม Feedforward network

### บทที่ 3

#### การออกแบบและพัฒนาระบบ

งานวิจัยนี้ได้ประยุกต์ใช้กระบวนการตรวจจับและติดตามวัตถุที่เคลื่อนไหวจากกล้องวิดีโอไคเน็คต์ (kinect camera) เพื่อพัฒนาเทคนิคการจำแนกพฤติกรรมพื้นฐานของมนุษย์ ได้แก่ การยืน การเดิน การนั่ง การก้ม และการนอน ซึ่งสามารถนำไปสู่เทคนิคการตรวจจับพฤติกรรมผิดปกติของมนุษย์ได้ โดยระบบจะใช้ข้อมูลภาพจากมุมมองด้านบน (top-view) จากกล้องวิดีโอไคเน็คต์ ซึ่งประกอบด้วยข้อมูลสี ความลึก และการเคลื่อนไหวของวัตถุ สำหรับวิเคราะห์พฤติกรรมของมนุษย์ กระบวนการโดยรวมแสดงในภาพประกอบ 3-1 เริ่มต้นระบบรับภาพวิดีโอจากกล้องไคเน็คต์ สำหรับใช้เป็นอินพุต ซึ่งประกอบด้วยภาพ 2 แบบ คือ ภาพสี (Color image) จาก RGB Camera และ ภาพความลึก (Depth image) จาก IR Camera ทั้ง 2 ภาพจะถูกนำไปวิเคราะห์ในกระบวนการสกัดลักษณะเด่นของวัตถุที่สนใจ (Feature extraction) โดยภาพความลึกซึ่งมีลักษณะเป็นภาพเกรย์สเกลที่มีการตอบสนองต่อแสงที่มีการเปลี่ยนแปลงน้อยกว่าภาพสีถูกนำไปใช้ในกระบวนการตรวจจับและติดตามการเคลื่อนที่ของวัตถุ (Motion object extraction) ซึ่งเลือกใช้เทคนิคการลบภาพพื้นหลัง (Background Subtraction) เนื่องจากระบบจะติดตั้งกล้องวิดีโอไคเน็คต์บนเพดานแบบไม่มีการเคลื่อนไหว หลังจากได้ผลลัพธ์ของเทคนิคการลบภาพพื้นหลังจะนำมาผ่านกระบวนการวิเคราะห์จุดของกลุ่มภาพ (Blob Analysis) เพื่อที่จะหาตำแหน่งของบริเวณที่มีการเคลื่อนที่ของวัตถุ สำหรับนำไปใช้ในกระบวนการวิเคราะห์โครงสร้างมนุษย์ (Human structure analysis) ซึ่งแบ่งเป็น 3 ขั้นตอน 1.การฉายโครงสร้างมนุษย์ด้วยการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริตี (Human structure project via color UV-Disparity) โดยวัตถุที่สนใจจะถูกนำมาใช้เป็นอินพุตสำหรับการสร้างภาพเสมือนโครงสร้างมนุษย์ด้วยวิธีการประมาณระยะทางของวัตถุจากแผนผังคิสพาริตี (UV-Disparity) ซึ่งผลลัพธ์จะเป็นการประมาณลักษณะ โครงสร้างมนุษย์ตามระดับความลึกที่แตกต่างกัน จะแบ่งเป็นแกน  $U(x)$  และ  $V(y)$  เพื่อใช้แสดงภาพ 2 มิติที่อิงกับค่าความลึก ซึ่งสามารถบ่งบอกถึงลักษณะท่าทางของมนุษย์ได้ แต่จะมีลักษณะเป็นภาพแบบเกรย์สเกลซึ่งจะจำแนกส่วนต่างๆของร่างกายมนุษย์จากระดับความลึก เพื่อเป็นการเพิ่มลักษณะเด่นสำหรับการใช้การจำแนกส่วนต่างๆได้ดียิ่งขึ้นจึงจะนำสีเข้ามาช่วยในการจำแนก โดยนำภาพสีมาผ่านกระบวนการจับคู่สีกับวัตถุที่สนใจ (Color match) เพื่อที่จะแมทสีให้กับส่วนต่างๆของร่างกายมนุษย์ แต่เนื่องจากกล้องสำหรับภาพสีกับกล้องสำหรับภาพความลึกตั้งอยู่ในตำแหน่งที่ต่างกันส่งผลให้ภาพทั้งสองมีตำแหน่งที่คลาดเคลื่อนกันจึงไม่สามารถแมทสีให้กับวัตถุได้โดยตรง ดังนั้นจะต้องใช้วิธีการสอบ

เทียบภาพ (Calibration Image) เพื่อปรับขนาดเซตตำแหน่งที่มีการคลาดเคลื่อนให้ตรงกันก่อนจึงจะนำภาพทั้งสองมาแมทกัน แล้วจึงนำผลลัพธ์การแมทที่ได้ไปใช้ในกระบวนการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริติ ซึ่งจะทำให้ผลลัพธ์มีลักษณะเด่นที่เพิ่มมากขึ้นจากเดิมที่เป็นภาพเกรย์สเกลวิเคราะห์จากระดับความลึกที่แตกต่างกันก็สามารถจำแนกส่วนต่างๆของร่างกายได้ด้วยสีที่ต่างกัน เช่น ผมสีดำ เสื้อสีแดง กางเกงสีน้ำเงิน เป็นต้น 2.การสร้างโครงสร้างมนุษย์ (Human Structure) โดยนำผลลัพธ์ก่อนหน้ามาทำการประมาณเพื่อหาจุดกึ่งกลางของร่างกายมนุษย์ โดยแบ่งเป็น 7 จุด มี 6 ส่วนคือ ศีรษะ 1 ส่วน ลำตัว 3 ส่วน และขา 3 ส่วน ซึ่งจะได้โครงสร้างมนุษย์โดยประมาณสำหรับภาพในแกน U และ V แต่จากโครงสร้างที่ได้จากภาพทั้งสองไม่สามารถนำมาใช้เพื่อบ่งบอกท่าทางของมนุษย์ได้ทั้งสองภาพ คือบางกรณีภาพในแกน U จะใช้บ่งบอกท่าทางได้ถูกต้อง บางกรณีภาพในแกน V จะใช้บ่งบอกท่าทางได้ถูกต้อง ดังนั้นจึงจะไปปรับปรุงในกระบวนการต่อไป 3.การรวมโครงสร้างมนุษย์ (Structure of human of combination) เนื่องจากการนำภาพโครงสร้างมนุษย์ทั้งแกน U และ V มาใช้สำหรับการวิเคราะห์ท่าทางมนุษย์จะทำให้ข้อมูลที่ใช้สำหรับการวิเคราะห์มีจำนวนมาก ซึ่งจะส่งผลทำให้ความถูกต้องในการจำแนกท่าทางลดน้อยลงตามไปด้วย ดังนั้นจึงจะลดความซับซ้อนในส่วนนี้โดยรวมภาพให้เหลือเพียงภาพเดียว โดยอาศัยความสัมพันธ์ของทั้งสองภาพ แล้วนำภาพที่ได้เข้าสู่กระบวนการสกัดลักษณะเด่น (Feature Extraction) ซึ่งแบ่งลักษณะเด่นที่ใช้สำหรับการทดลองออกเป็น 3 แบบ คือ 1. ลักษณะเด่นที่ได้จากการคำนวณมุมของโครงสร้างมนุษย์จากผลลัพธ์ของกระบวนการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริติจำนวน 18 ข้อมูล 2. ลักษณะเด่นที่ได้จากการคำนวณมุมของโครงสร้างมนุษย์จากผลลัพธ์ของการรวมโครงสร้างมนุษย์จำนวน 6 ข้อมูล 3. ลักษณะเด่นที่ได้จากการคำนวณมุมและค่าเฉลี่ยความสูงจากโครงสร้างมนุษย์จากผลลัพธ์ของการรวมโครงสร้างมนุษย์จำนวน 7 ข้อมูล จากนั้นนำลักษณะเด่นทั้ง 3 แบบเข้าสู่กระบวนการรู้จำท่าทางมนุษย์ (Model recognition) เพื่อจำแนกท่าทางมนุษย์ออกเป็น 4 กลุ่ม ได้แก่ 1. การยืนหรือการเดิน 2. การนั่ง 3. การก้ม 4. การนอน โดยเลือกใช้เทคนิคของข่ายงานประสาท (neuron network) สำหรับการเทรนลักษณะเด่น



ภาพประกอบ 3-1 กระบวนการ โดยรวมของการวิเคราะห์พฤติกรรมมนุษย์

### 3.1. กระบวนการตรวจจับและติดตามการเคลื่อนที่ของวัตถุ (Motion Object Extraction)

ในกระบวนการนี้ใช้สำหรับตรวจจับและติดตามวัตถุที่สนใจ ซึ่งในที่นี้คือมนุษย์ การทำงานคือจะตรวจจับเฉพาะวัตถุที่เคลื่อนที่ในภาพ โดยเลือกใช้ข้อมูลภาพความลึกสำหรับนำไปใช้ในการตรวจจับการเคลื่อนที่ของวัตถุที่สนใจ เนื่องจากภาพความลึกเป็นภาพเกรย์สเกลซึ่งจะไม่มีปัญหาในเรื่องของแสงที่จะมีการเปลี่ยนแปลงอยู่ตลอดเวลา ซึ่งจะส่งผลให้การตรวจจับวัตถุที่สนใจมีการคลาดเคลื่อนได้ อัลกอริทึมที่นิยมใช้สำหรับตรวจจับการเคลื่อนที่ของวัตถุมีด้วยกัน 2 วิธีหลัก คือ 1. การลบภาพพื้นหลัง และ 2. ออปติคัล โฟว์ ซึ่งในระบบนี้เลือกใช้อัลกอริทึมการลบภาพพื้นหลัง

#### 3.1.1. การลบภาพพื้นหลัง (Background Subtraction)

อัลกอริทึมการลบภาพพื้นหลัง ซึ่งจะเหมาะกับระบบนี้ เนื่องจากกล้องที่ใช้สำหรับการรับภาพจะอยู่นิ่งกับที่ โดยอัลกอริทึมการลบพื้นหลังจะนำเอาภาพเฟรมหนึ่งๆ มาลบกับพื้นหลังซึ่งจะได้เฉพาะวัตถุที่สนใจ ดังต่อไปนี้

$$O = I - B \quad (3.1)$$

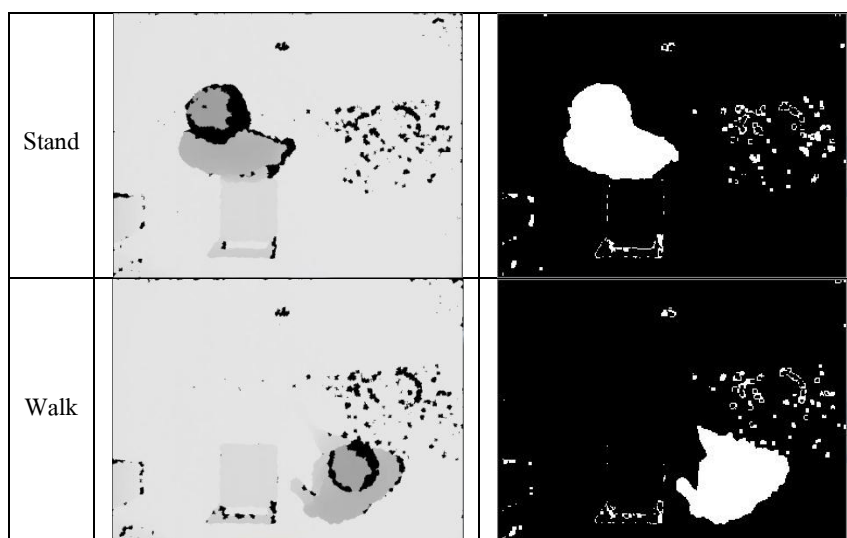
เมื่อ O คือ Object  
I คือ ภาพของเฟรมหนึ่ง  
B คือ ภาพพื้นหลัง

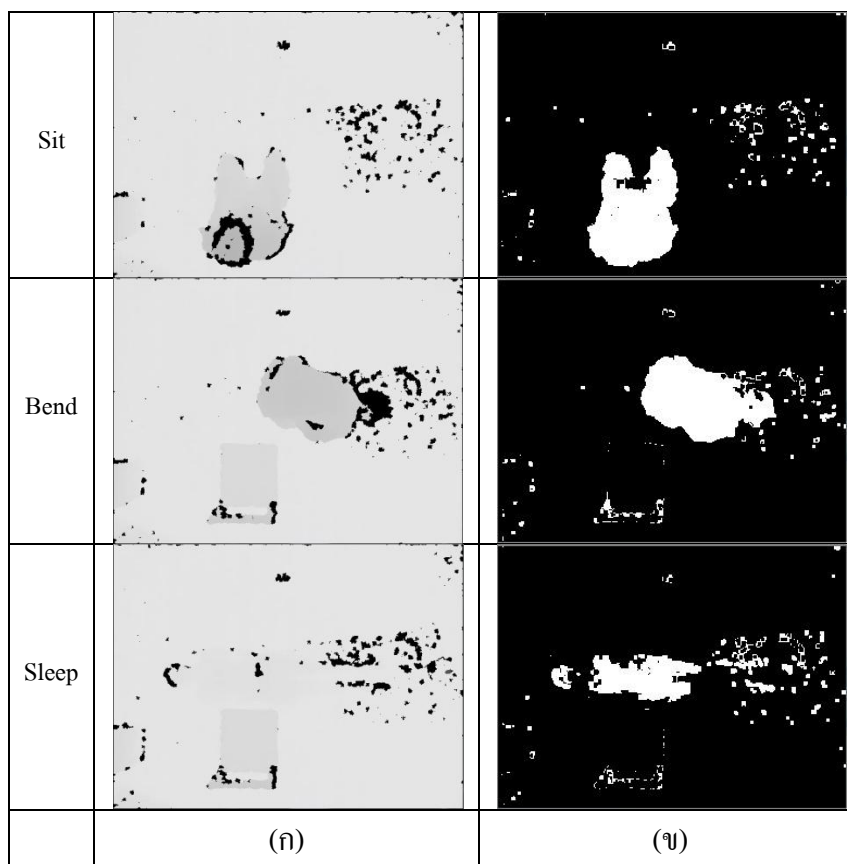
ในกรณีที่ภาพพื้นหลังมีการเปลี่ยนแปลงไปจากภาพแรกจำเป็นที่จะต้องมีการอัปเดตภาพพื้นหลัง ซึ่งในที่นี้จะใช้วิธีของ mixture of gaussian ซึ่งจะใช้ค่าสถิติมาเปรียบเทียบกับค่าความแปรปรวนของภาพพื้นหลัง (background) และภาพด้านหน้า (foreground) ซึ่งนิยามได้ดังต่อไปนี้

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (3.2)$$

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^2 |\Sigma|^2} e^{-\frac{1}{2}(X_t - \mu_t)^T \Sigma^{-1} (X_t - \mu_t)} \quad (3.3)$$

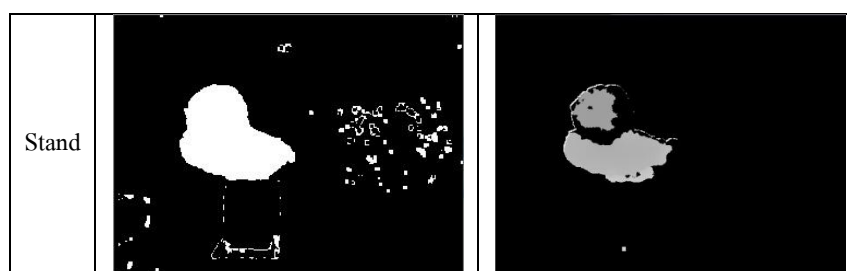
โดย  $P(X_t)$  คือค่าความน่าจะเป็นที่ใช้สำหรับอัปเดตจุด  $X$  ณ เวลาใดๆ คำนวณมาจาก  $\omega_{i,t}$  คูณกับฟังก์ชัน  $\eta$  โดยที่  $\omega_{i,t}$  ซึ่งเป็นค่าน้ำหนักที่ใช้ในการปรับความเหมาะสมของฟังก์ชันความน่าจะเป็นเกาส์เซียน และ  $\eta$  เป็นฟังก์ชันความน่าจะเป็นของเกาส์เซียน ที่คำนวณมาจากค่าเฉลี่ยของพิกเซล ( $\mu$ ) และค่าเบี่ยงเบนมาตรฐาน ( $\Sigma$ ) โดยในงานวิทยานิพนธ์นี้จะกำหนดจำนวนเกาส์เซียน  $K = 5$  ผลลัพธ์ของการตรวจจับและติดตามการเคลื่อนที่ของวัตถุ ดังภาพประกอบ 3-2

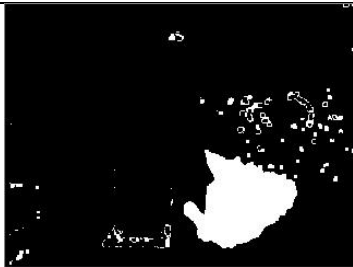
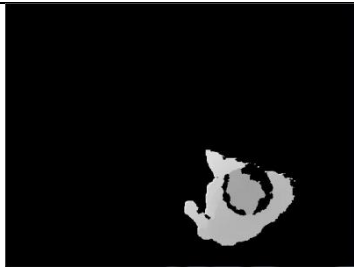
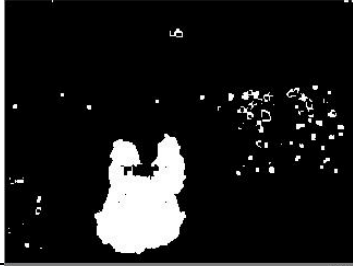
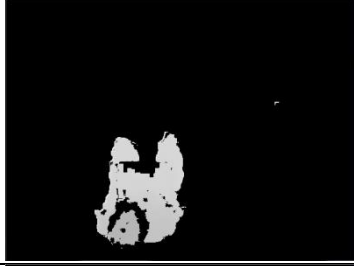
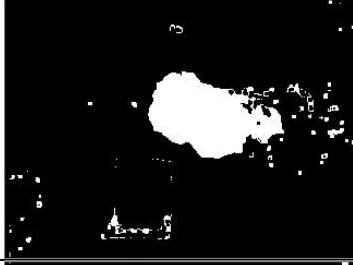

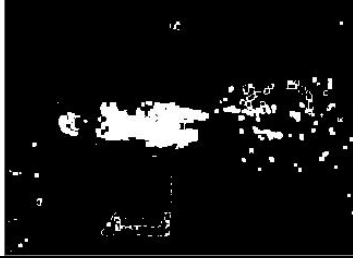
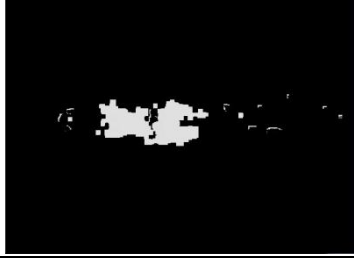




ภาพประกอบ 3-2 กระบวนการตรวจจับและติดตามการเคลื่อนที่ของมนุษย์ (ก) ภาพต้นฉบับ (ข) ภาพผลลัพธ์ของการตรวจจับวัตถุที่เคลื่อนที่

จากภาพประกอบ 3-2 (ก) เป็นภาพต้นฉบับ (ข) เป็นผลลัพธ์การลบภาพพื้นหลัง ซึ่งจากผลลัพธ์ของทุกท่าทางสามารถตรวจจับวัตถุที่มีการเคลื่อนที่ได้อย่างถูกต้อง โดยวัตถุที่มีการเคลื่อนที่จะเป็นจุดสีขาวตามลักษณะท่าทางมนุษย์จากมุมมองด้านบน ส่วนภาพพื้นหลังจะเป็นจุดสีดำ แต่จากการสังเกตผลลัพธ์ที่ได้จะมีจุดสีขาวบางส่วนในภาพที่ไม่ต้องการ ซึ่งเกิดจากความผิดพลาดของภาพความลึกในบริเวณที่เป็นจุดอับที่มีเปลี่ยนแปลงตลอดเวลา โดยจะสมมุติให้เป็นสัญญาณรบกวน (noise) โดยสามารถกำจัดสัญญาณรบกวนเหล่านี้ด้วยเทคนิค open morphology เลือกใช้ kernel แบบสี่เหลี่ยม ทำการวนลูป 2 ครั้ง ซึ่งทำหน้าที่เปิดหรือเซาะพิกเซลที่ไม่ต้องการให้หมดไป ดังภาพประกอบ 3-3



Walk		
Sit		
Bend		
Sleep		
	(ก)	(ข)

ภาพประกอบ 3- 3 การกำจัดสัญญาณรบกวน (ก) ภาพต้นฉบับ (ข) ภาพผลลัพธ์ของการกำจัดสัญญาณรบกวน

จากภาพประกอบ 3-3 (ก) ภาพต้นฉบับ (ข) ผลลัพธ์การกำจัดสัญญาณรบกวน ซึ่งจากผลลัพธ์ของการกำจัดสัญญาณรบกวนสามารถทำได้ดีมาก โดยเปรียบเทียบได้จากภาพต้นฉบับ สัญญาณรบกวนจะถูกลบไปได้จำนวนมาก แต่อาจจะเหลืออยู่บ้าง แต่ก็ไม่ได้ส่งผลต่อการวิเคราะห์ภาพ

### 3.1.2. การวิเคราะห์กลุ่มของจุดภาพ (Blob Analysis)

กระบวนการวิเคราะห์กลุ่มของจุดภาพ เป็นการวิเคราะห์กลุ่มสีของวัตถุหรือเป้าหมายที่เป็นไบนารี โดยจะแยกกลุ่มของจุดภาพที่แตกต่างกันแล้วคำนวณหาตำแหน่งและขนาดของวัตถุ

เพื่อสร้าง blob ของวัตถุที่สนใจ ดังภาพประกอบ 3-4



ภาพประกอบ 3- 4 การวิเคราะห์หากลุ่มของจุดภาพ

### 3.2. กระบวนการจับคู่สีกับวัตถุที่สนใจ (Color Match)

กระบวนการจับคู่สีกับวัตถุที่สนใจ เพื่อที่จะเพิ่มคุณลักษณะเด่นสำหรับการประมวลผล โดยจะนำผลลัพธ์ที่ได้จากกระบวนการติดตามวัตถุ ซึ่งจะได้อำนาจของบริเวณวัตถุที่สนใจสำหรับนำไปเลือกบริเวณที่สนใจ (region of interest) ในภาพสี แต่เนื่องจากตำแหน่งที่ได้จากกระบวนการติดตามเฉพาะวัตถุที่เคลื่อนที่ได้มาจากภาพความลึก ซึ่งเป็นตำแหน่งที่ไม่ตรงกันกับภาพสี ดังนั้นจึงต้องมีการสอบเทียบภาพ (calibration image) เพื่อคำนวณหาความสัมพันธ์ของภาพทั้งสอง

#### 3.2.1. การสอบเทียบภาพ (Calibration Image)

การสอบเทียบภาพเพื่อหาความสัมพันธ์ระหว่างภาพ 2 ภาพ เกิดขึ้นเนื่องจากภาพ 2 ภาพเกิดจากกล้องสองตัวที่ติดตั้งในตำแหน่งที่ต่างกัน ส่งผลให้บริเวณของวัตถุที่เหมือนกันมีตำแหน่งของภาพ  $(x, y)$  ที่ไม่ตรงกัน ซึ่งการสอบเทียบภาพจะแบ่งออกเป็น 3 ขั้นตอนดังนี้

1. แปลงภาพความลึก 2 มิติ เป็นตำแหน่งความลึก 3 มิติ

วิธีการสอบเทียบภาพในขั้นตอนแรกจะใช้ค่าความเข้มของสี (intensity) และตำแหน่งพิกเซลของภาพความลึก สำหรับคำนวณหาตำแหน่งในพิกัด 3 มิติของกล้องความลึก ได้แก่ P3D.x, P3D.y และ P3D.z ซึ่งสามารถอธิบายได้ดังนี้

$$P3D.x = (x_d - cx_d) * \left( \frac{\text{depth}(x_d, y_d)}{fx_d} \right) \quad (3.4)$$

$$P3D.y = (y_d - cy_d) * \left( \frac{\text{depth}(x_d, y_d)}{fy_d} \right) \quad (3.5)$$

$$P3D.z = \text{depth}(x_d, y_d) \quad (3.6)$$



เมื่อ  $x_d, y_d$  เป็นตำแหน่งพิกัด  $x, y$  ของภาพความลึก  $\text{depth}(x_d, y_d)$  เป็นค่าความเข้มสีที่ตำแหน่ง  $x, y$  ในภาพความลึก  $cx_d, cy_d, fx_d, fy_d$  เป็นค่าที่บ่งบอกลักษณะเฉพาะของกล้องความลึก (depth intrinsic parameter) ดังต่อไปนี้

$$fx_d = 5.9421434211923247e+02$$

$$fy_d = 5.9104053696870778e+02$$

$$cx_d = 3.3930780975300314e+02$$

$$cy_d = 2.4273913761751615e+02$$

## 2. แปลงตำแหน่ง 3 มิติของกล้องความลึกเป็นตำแหน่ง 3 มิติของกล้องสี

เมื่อได้ตำแหน่ง 3 มิติของกล้องความลึก จะถูกนำมาแปลงเป็นตำแหน่ง 3 มิติของกล้องสี โดยจะอาศัยความสัมพันธ์เชิง geometry ระหว่างกล้องทั้ง 2 กล้อง ซึ่งจะใช้เมตริกซ์การหมุน (R) และเวกเตอร์การย้ายตำแหน่ง (T) เข้ามาช่วยสำหรับการปรับตำแหน่งภาพของกล้องความลึกให้เป็นค่าที่เหมาะสมสำหรับภาพของกล้องสี สามารถอธิบายเป็นสมการดังนี้

$$\begin{bmatrix} P3D'.x \\ P3D'.y \\ P3D'.z \end{bmatrix} = R \cdot \begin{bmatrix} P3D.x \\ P3D.y \\ P3D.z \end{bmatrix} + T \quad (3.7)$$

เมื่อ  $P3D'.x, P3D'.y, P3D'.z$  เป็นตำแหน่ง 3 มิติของกล้องสี, R เป็นเมตริกซ์การหมุนและ T เป็นเวกเตอร์การย้ายตำแหน่ง ซึ่งเป็นค่าคงที่ดังนี้

$$R = \begin{bmatrix} 9.9984628826577793e-01 & 1.2635359098409581e-03 & -1.7487233004436643e-02 \\ -1.4779096108364480e-03 & 9.9992385683542895e-01 & -1.2251380107679535e-02 \\ 1.7470421412464927e-02 & 1.2275341476520762e-02 & 9.9977202419716948e-01 \end{bmatrix}$$

$$T = \begin{bmatrix} 1.9985242312092553e-02 \\ -7.4423738761617583e-04 \\ -1.0916736334336222e-02 \end{bmatrix}$$

## 3. แปลงตำแหน่ง 3 มิติของกล้องสีเป็นตำแหน่ง 2 มิติในภาพสี

เมื่อได้ตำแหน่ง 3 มิติของกล้องสี จะถูกนำมาหาตำแหน่ง 2 มิติในภาพสี โดยสามารถอธิบายเป็นสมการได้ดังนี้

$$P2D_{rgb}.x = \frac{P3D'.x * fx_{rgb}}{P3D'.z} + cx_{rgb} \quad (3.8)$$

$$P2D_{rgb}.y = \frac{P3D'.y * fy_{rgb}}{P3D'.z} + cy_{rgb} \quad (3.9)$$

เมื่อ  $P2D_{rgb}.x$ ,  $P2D_{rgb}.y$  เป็นตำแหน่ง 2 มิติของภาพสี,  $cx_{rgb}$ ,  $cy_{rgb}$ ,  $fx_{rgb}$ ,  $fy_{rgb}$  ค่าที่บ่งบอกลักษณะเฉพาะของกล้องสี (color intrinsic parameter) ดังต่อไปนี้

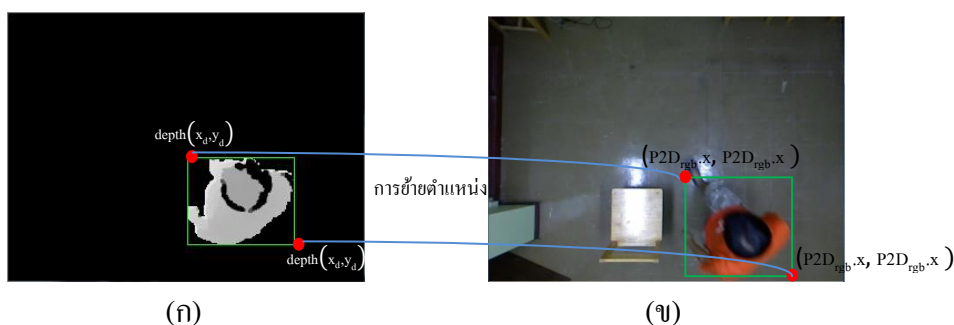
$$fx_{rgb} = 5.2921508098293293e+02$$

$$fy_{rgb} = 5.2556393630057437e+02$$

$$cx_{rgb} = 3.2894272028759258e+02$$

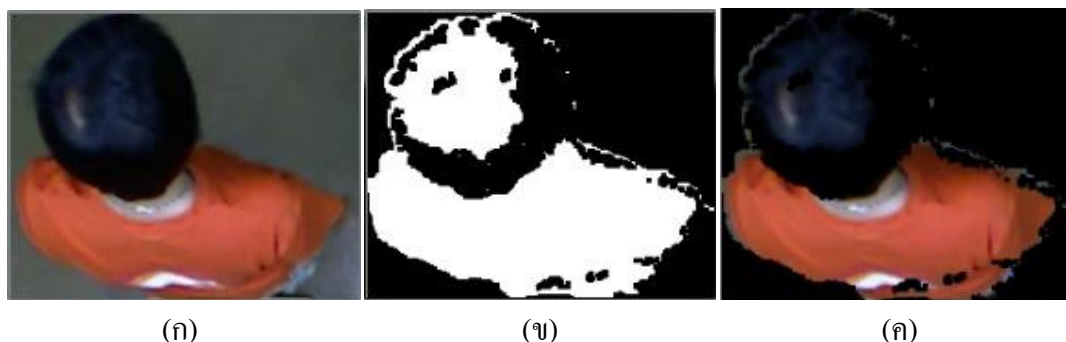
$$cy_{rgb} = 2.6748068171871557e+02$$

กระบวนการสอบเทียบสำหรับในระบบนี้ จะใช้ตำแหน่งเพียง 2 จุด จากผลลัพธ์ของการติดตามเฉพาะบริเวณวัตถุที่เคลื่อนที่ ซึ่งก็คือ จุดซ้ายบนสุดและจุดขวาล่างสุดของกรอบสี่เหลี่ยมสำหรับนำมาทำการสอบเทียบประมาณหาความสัมพันธ์ของตำแหน่งที่ต้องการระหว่างภาพความลึกและภาพสี ดังภาพประกอบ 3-5



ภาพประกอบ 3-5 การย้ายตำแหน่งของภาพความลึกไปภาพสี (ก) ภาพความลึก (ข) ภาพสี

ผลลัพธ์จากกระบวนการสอบเทียบจะถูกนำไปใช้สำหรับการประมาณบริเวณที่สนใจของภาพสี ดังภาพประกอบ 3-6 (ก) แล้วนำมาจับคู่ (match) บริเวณที่เหมือนกันกับผลลัพธ์ของการเลือกบริเวณที่สนใจ (ROI) จากภาพความลึก ซึ่งจะถูกแปลงเป็นภาพขาวดำ ดังภาพประกอบ 3-6 (ข) สุดท้ายทั้งสองภาพจะถูก match กันด้วยการ And operator ซึ่งจะได้ผลลัพธ์เป็นภาพสีเฉพาะลักษณะรูปร่างมนุษย์ ดังภาพประกอบ 3-6 (ค)



ภาพประกอบ 3- 6 (ก) การเลือกบริเวณที่สนใจจากภาพสี (ข) การเลือกบริเวณที่สนใจจากภาพความ  
 ลึก (ค) ผลลัพธ์จาก match กัน

$$\hat{O}_c = O_c \cap O_D \quad (3.10)$$

หรือสามารถอธิบายได้ดังสมการ (3.10) เมื่อ  $\hat{O}_c$  เป็นวัตถุที่มีการระบุสี,  $O_c$  เป็นภาพสีของ  
 วัตถุ ซึ่งนำมา AND กับ  $O_D$  ที่เป็นภาพความลึกที่ถูกแปลงเป็นภาพขาวดำ

### 3.3. การวิเคราะห์โครงสร้างมนุษย์ (Human Structure Analysis)

กระบวนการวิเคราะห์โครงสร้างมนุษย์จะเป็นการสร้างภาพเหมือน โครงสร้างมนุษย์เพื่อนำ ไป  
 ใช้สำหรับการหาลักษณะเด่นของท่าทางมนุษย์ การใช้ภาพในมุมมองด้านบนสามารถช่วยลดความ  
 ซ้ำซ้อนของร่างกายมนุษย์ลงได้ เมื่อนำมาใช้ร่วมกับข้อมูลความลึกก็จะสามารถช่วยบ่งบอกส่วน  
 ต่างๆของร่างกายได้จากความลึกที่ต่างกัน ซึ่งในวิทยานิพนธ์นี้จะสร้างโครงสร้างมนุษย์โดย  
 กำหนดให้โครงสร้างมนุษย์แบ่งเป็น 6 ส่วน ประกอบด้วย ศีรษะ 1 ส่วน ลำตัว 2 ส่วน และขา 3  
 ส่วน การวิเคราะห์โครงสร้างมนุษย์จะเป็น 3 ขั้นตอนใหญ่ ได้แก่ 1.การฉายโครงสร้างมนุษย์ด้วย  
 การระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังดิสพาริตี 2.การสร้างโครงสร้างมนุษย์ 3.  
 การรวมโครงสร้างมนุษย์

#### 3.3.1. การฉายโครงสร้างมนุษย์ด้วยการระบุสีให้กับการประมาณระยะทางของวัตถุจาก แผนผังดิสพาริตี (Human structure project via color UV-Disparity)

การฉายโครงสร้างมนุษย์ด้วยการระบุสีให้กับการประมาณระยะทางของวัตถุจาก  
 แผนผังดิสพาริตี เป็นการสร้างภาพใหม่โดยการฉายภาพให้มีความสัมพันธ์กับค่าความลึก ซึ่งบ่ง  
 บอกถึงรูปทรง ความสูง ความกว้างของมนุษย์ได้ แล้วจึงนำมาจับคู่สีกับส่วนต่างๆของร่างกาย  
 มนุษย์ เพื่อเพิ่มคุณสมบัติสำหรับการจำแนกส่วนต่างๆของร่างกาย ซึ่งสามารถแบ่งได้ 2 ขั้นตอน  
 ดังนี้

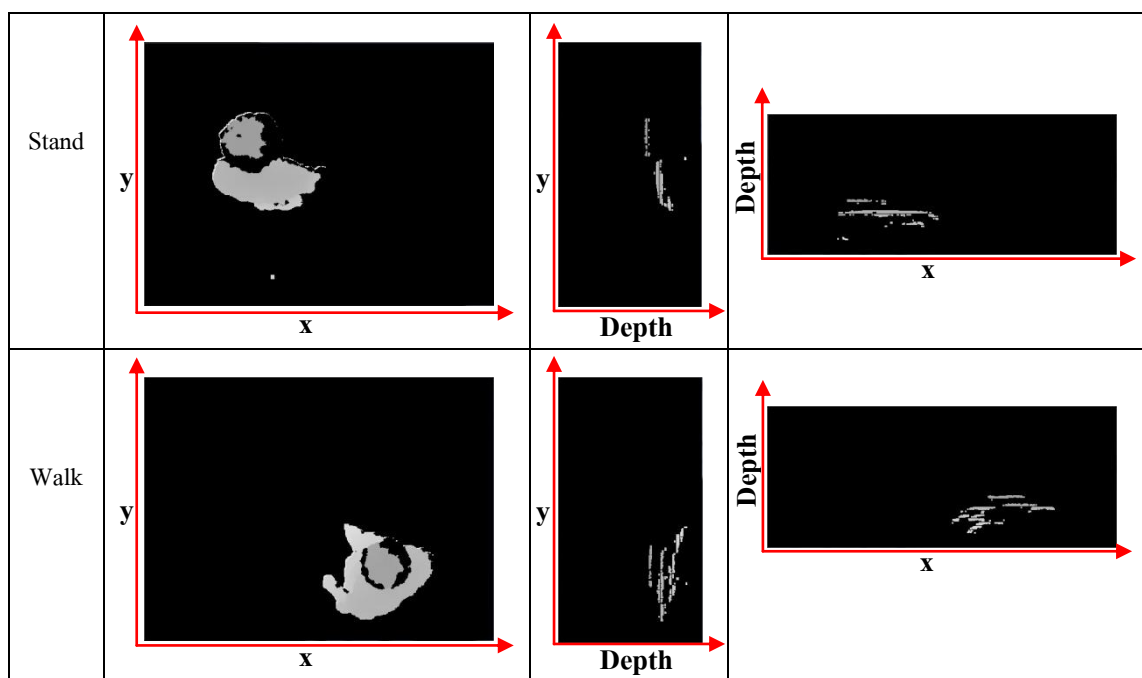
### 3.3.1.1. การประมาณระยะทางของวัตถุจากแผนผังดิสพาริตี (UV-Disparity)

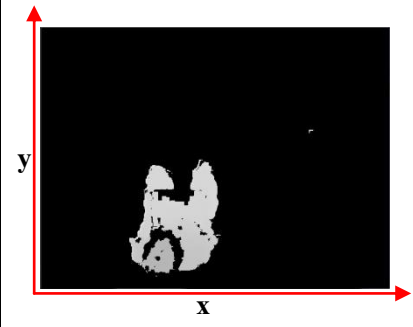
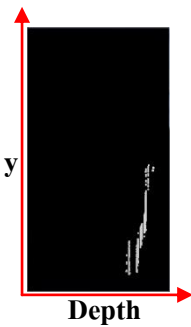
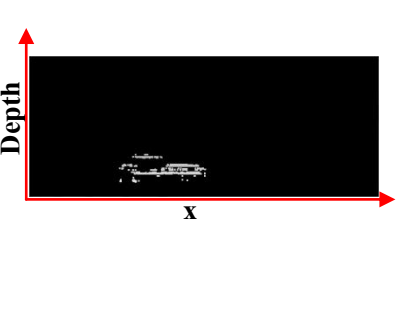
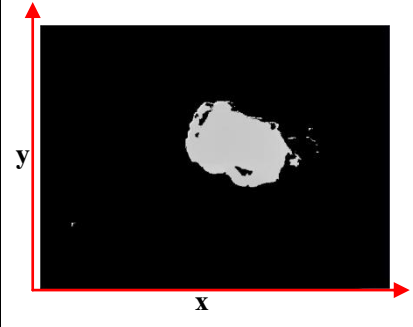


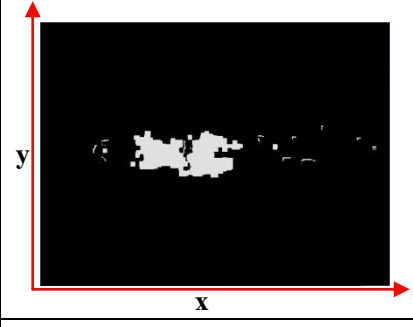

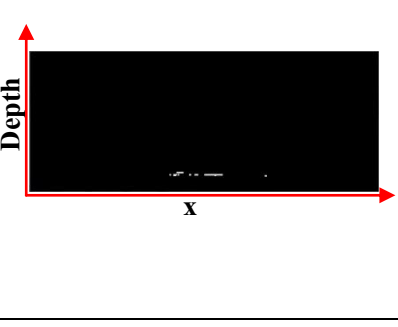
กระบวนการประมาณระยะทางของวัตถุจากแผนผังดิสพาริตี เป็นการแยกวัตถุที่สนใจตามระดับความลึกที่แตกต่าง ซึ่งในวิทยานิพนธ์นี้จะใช้งานร่วมกับมนุษย์จากภาพในมุมมองสูง จึงทำให้สามารถแยกระดับร่างกายของมนุษย์สำหรับใช้ในการวิเคราะห์พฤติกรรมมนุษย์ได้ เช่น จากภาพมุมมองสูง จะพบส่วนของหัว ไหล่ และลำตัว กำหนดว่าเป็นท่าทางการยืน เป็นต้น หลักการทำงานจะถูกแบ่งออกเป็น 2 แขนงคือ V-Disparity และ U-Disparity โดยผลลัพธ์จากการติดตามเฉพาะบริเวณวัตถุที่เคลื่อนที่จะถูกใช้สำหรับประมวลผลในกระบวนการนี้

V-disparity คือข้อมูลที่ประมวลผลมาจากแผนผังตามแนวแกน V (แนวตั้ง) โดยให้  $I_{\Delta}$  คือแผนผังดิสพาริตีที่ได้จากภาพความลึก H เป็นฟังก์ชันของตัวแปร  $I_{\Delta}$  ดังนี้

$$H(I_{\Delta})=I_{\Delta v} \quad (3.11)$$

เมื่อ  $I_{\Delta v}$  เรียกว่า V-disparity โดยที่ H คือการสะสมจำนวนดิสพาริตีที่มีค่าดิสพาริตีเดียวกันที่ปรากฏอยู่บนรูปภาพแถวที่  $i$  ดังภาพประกอบ 3-7 (ข)



Sit			
Bend			
Sleep			
	(ก)	(ข)	(ค)

ภาพประกอบ 3- 7 (ก) ผลลัพธ์ของการติดตามเฉพาะบริเวณวัตถุที่เคลื่อนที่

(ข) V-Disparity (ค) U-Dispairty

U-disparity คือข้อมูลที่ประมวลผลมาจากแผนผังดิสพาริตีตามแนวแกน U และ V เป็นฟังก์ชันของ  $I_{\Delta}$  ดังนี้

$$V(I_{\Delta})=I_{\Delta_u} \quad (3.12)$$

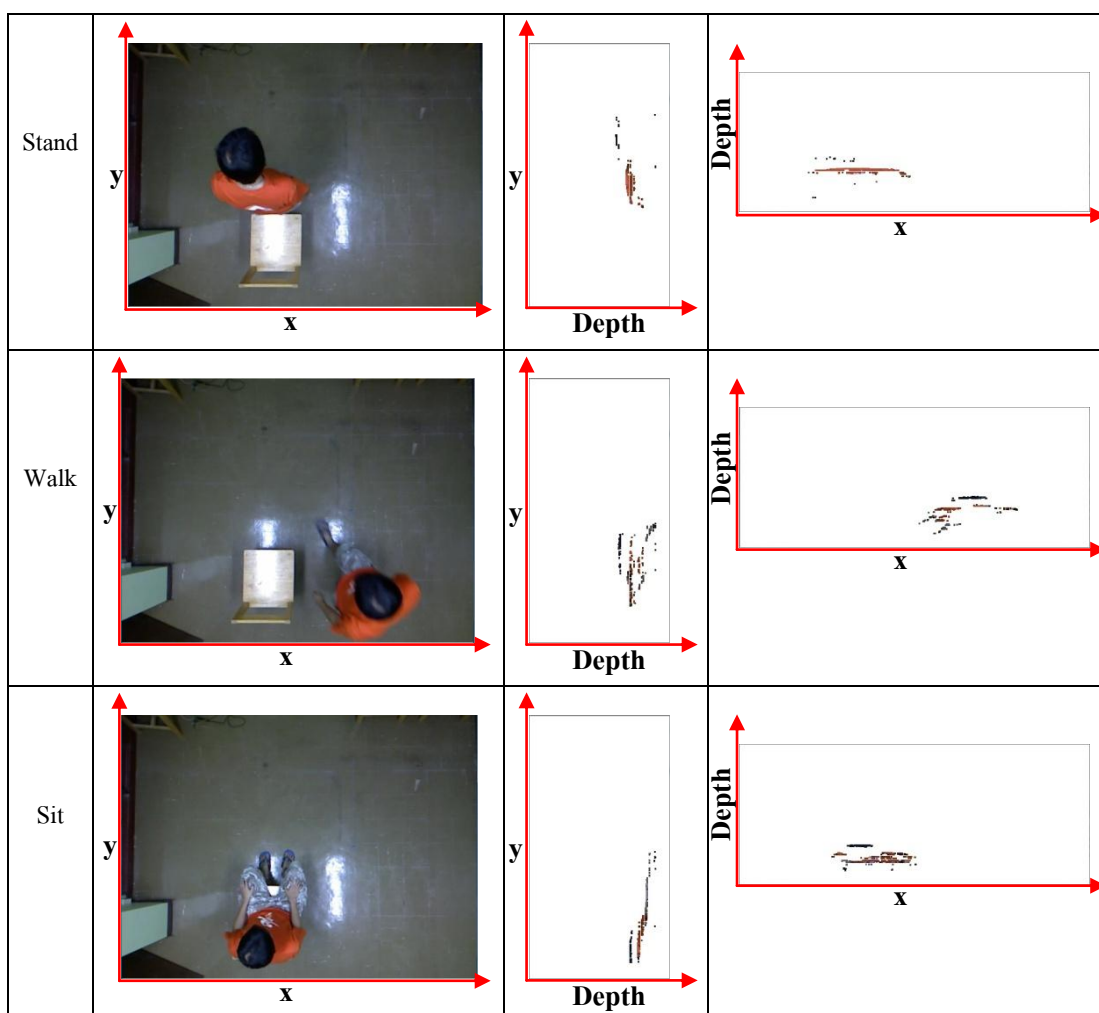
$I_{\Delta_u}$  เรียกว่า U-disparity โดยที่ V คือการสะสมจำนวนดิสพาริตี ที่มีค่าดิสพาริตีเดียวกันที่ปรากฏอยู่บนรูปภาพคอลัมน์ที่ j ซึ่งการใช้งานฟังก์ชัน H และ V เพื่อแปลงแผนผังดิสพาริตีดังกล่าวประกอบ 3-7 (ค)


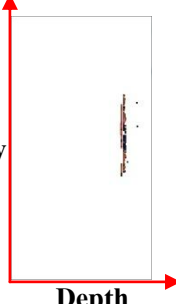
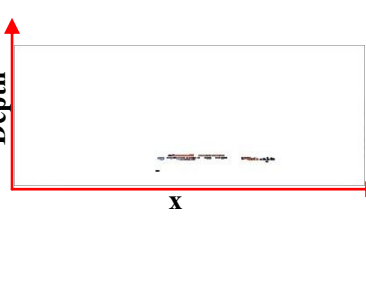


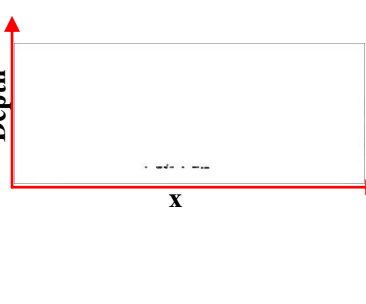
จากผลลัพธ์ของ UV-Disparity จะประกอบด้วย 5 ท่าทางได้แก่ การยืน การเดิน การนั่ง การก้ม และการนอน ซึ่งสามารถสังเกตลักษณะท่าทางต่างๆของมนุษย์ได้จากระดับความลึกที่แตกต่าง

กัน จากมุมมองของ U และ V ที่มีความสัมพันธ์กัน การใช้เฉพาะค่าความลึกสำหรับการวิเคราะห์ท่าทางมนุษย์อาจจะไม่เพียงพอ จึงมีแนวคิดที่เพิ่มลักษณะเด่นให้กับ UV-Disparity โดยการนำสีเข้ามาช่วยในการระบุส่วนต่างๆของร่างกายมนุษย์ เช่น การระบุสีของหัว เสื้อ กางเกง เป็นต้น

### 3.3.1.2. การระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังดิสพาริตี (Color UV-Disparity)

กระบวนการระบุสีสำหรับการประมาณระยะทางของวัตถุจากแผนผังดิสพาริตี จะช่วยให้การแยกแยะส่วนต่างๆของมนุษย์มีความถูกต้องมากขึ้น โดยการนำผลลัพธ์จากกระบวนการจับคู่สีกับวัตถุที่สนใจ (color match) มารวมกับการประมาณระยะทางของวัตถุจากแผนผังดิสพาริตี ซึ่งการรวมกันจะต้องอาศัยตำแหน่ง และค่าความลึกที่เหมือนกัน ดังภาพประกอบ 3-8



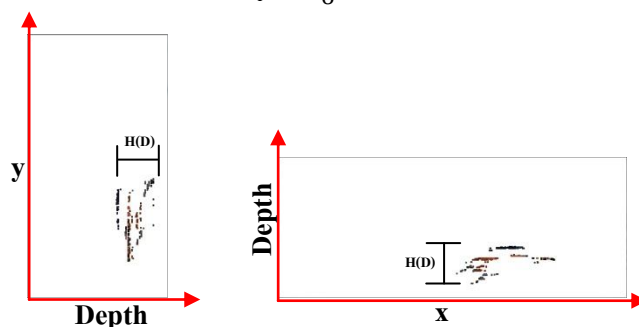
Bend			
Sleep			
	(ก)	(ข)	(ค)

ภาพประกอบ 3- 8 (ก) ภาพสี (ข) Color V-Disparity (ค) Color U-Disparity

### 3.3.2. การสร้างโครงสร้างมนุษย์ (Human Structure)

ผลลัพธ์จากกระบวนการระบุสีสำหรับการประมาณระยะทางของวัตถุจากแผนผังดิสพาริตี จะถูกนำสร้างเป็นโครงสร้างใหม่ซึ่งจะใช้ระบุท่าทางของมนุษย์ โดยเริ่มจากแบ่งข้อมูลออกเป็น 6 ส่วนตามระดับความสูง ดังสมการ (3.13) เมื่อ  $H(D)$  เป็นความสูงของมนุษย์  $i = \{1, 2, 3, 4, 5, 6\}$  ในกรณีนี้เมื่อหารแล้วมีเศษจะนำเศษไปรวมกับ  $part_6$  (ขาคส่วนที่ 3)

$$part_i = \frac{H(D)}{6} \tag{3.13}$$

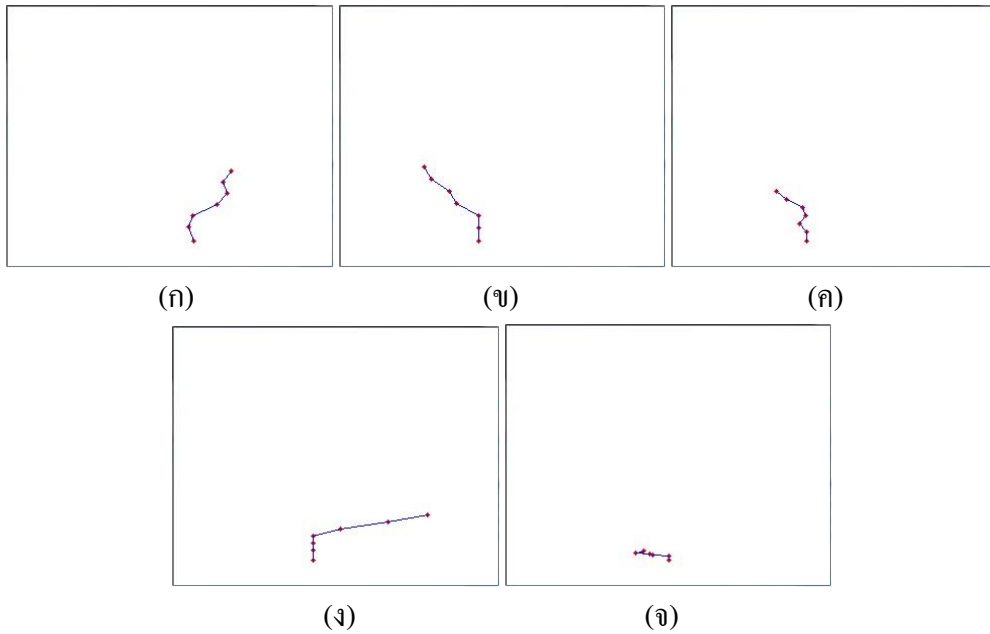


ภาพประกอบ 3- 9 ตัวอย่างการประมาณความสูงมนุษย์

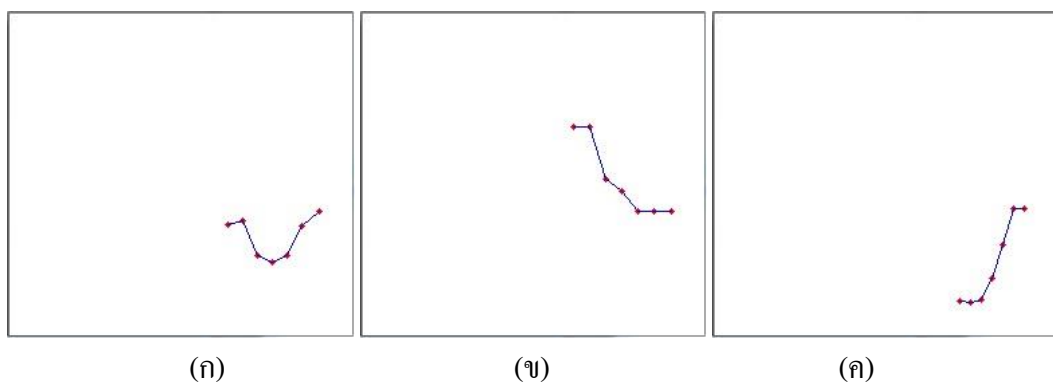
การสร้างโครงสร้างใหม่นั้นจะใช้แกนกลางของข้อมูล โดยการหาจุดศูนย์กลางของแต่ละ  $part_i$  ด้วยการหาค่าเฉลี่ยของจำนวนจุดพิกเซลในแต่ละ  $part_i$  ดังสมการ (3.14)

$$P(x_i, y_i) = \frac{\sum_0^n b(x, y)}{n} \quad (3.14)$$

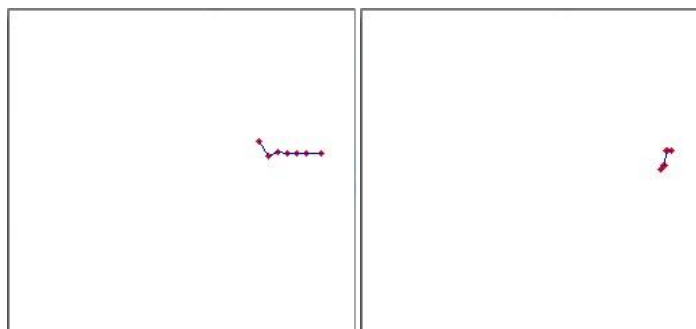
เมื่อ  $P(x_i, y_i)$  เป็นจุดตรงกลางของแต่ละ  $part_i$ ,  $p(x, y)$  เป็นตำแหน่งของจุดพิกเซลที่อยู่ในแต่ละ  $part_i$ ,  $n$  คือจำนวนของจุดพิกเซลที่ปรากฏในแต่ละ  $part_i$  ผลลัพธ์การสร้างโครงสร้างมนุษย์ดังภาพประกอบ 3-10 และ ภาพประกอบ 3-11



ภาพประกอบ 3- 10 โครงสร้างมนุษย์สำหรับ U-disparity: (ก) การยืน (ข) การเดิน  
(ค) การนั่ง (ง) การก้ม (จ) การนอน





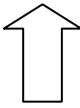



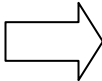





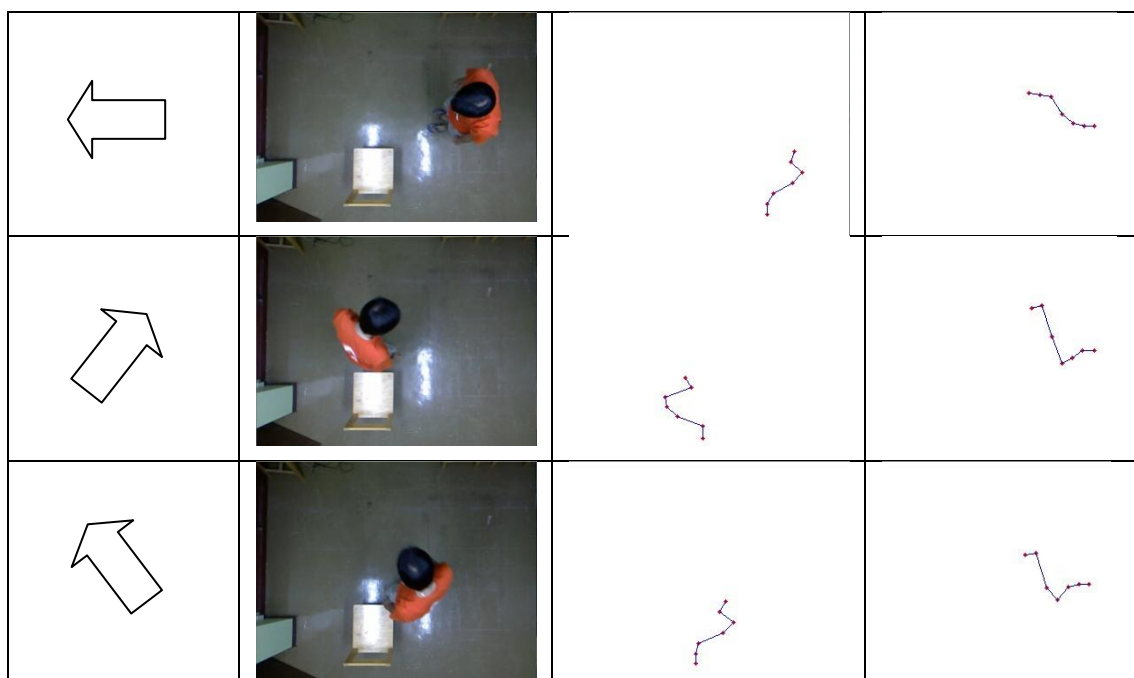
(ง)

(จ)

ภาพประกอบ 3- 11 โครงสร้างมนุษย์สำหรับ V-disparity: (ก) การขึ้น (ข) การเดิน  
(ค) การนั่ง (ง) การก้ม (จ) การนอน

จากภาพประกอบ 3-10,3-11 เป็นผลลัพธ์ของการสร้างโครงสร้างมนุษย์จะประกอบด้วยจุดศูนย์กลาง 7 จุด โดยจุดแรก  $part_0$  จะเป็นจุดอ้างอิงคำนวณค่าเฉลี่ยจากระดับความสูงแรก ส่วน 6 จุดที่เหลือจะคำนวณดังสมการ 3.14 โดยผลลัพธ์สำหรับ 1 เฟรมจะประกอบด้วย 2 โครงสร้างมนุษย์ในแกน  $u$  และแกน  $v$  ซึ่งโครงสร้างมนุษย์ทั้ง 2 แกน อาจจะมีเพียงแค่แกนเดียวที่สามารถบ่งบอกถึงท่าทางมนุษย์ได้ ตัวอย่างเช่น ภาพประกอบ 3-10 (ค) และ ภาพประกอบ 3-11 (ค) ทั้ง 2 ภาพคือผลลัพธ์ที่ได้จากการนั่ง แต่มีเพียงภาพประกอบ 3-11 (ค) เท่านั้นที่บ่งบอกถึงท่าทางการนั่ง ส่วนอีกภาพจะไม่ใช่ สาเหตุมาจากทิศทางของมนุษย์ในภาพจะมีผลต่อการฉายภาพในกระบวนการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริตี ซึ่งทำให้โครงสร้างมนุษย์เกิดความผิดพลาดดังภาพประกอบ 3-12

ทิศทาง	(ก)	(ข)	(ค)
			
			



ภาพประกอบ 3- 12 ผลลัพธ์ของโครงสร้างมนุษย์จากภาพกระบวนการระบุสีให้กับการประมวล  
ระยะทางของวัตถุจากแผนผังคิสพาริตี (ก) ภาพต้นฉบับ (ข) U-Disparity (ค) V-Disparity

จากภาพประกอบ 3-12 เป็นผลลัพธ์ของโครงสร้างมนุษย์ในท่าทางการยืนที่มีทิศทาง  
หันหน้าที่แตกต่างกัน ซึ่งจะส่งผลให้ผลลัพธ์ของโครงสร้างมนุษย์ของทั้งแกน U และ V มีความต่างกัน  
ไปด้วย โดยทิศทางหันของโครงสร้างมนุษย์จะมีผลต่อการฉายภาพทำให้โครงสร้างมนุษย์ที่ได้  
ในแกน U และ V เกิดความผิดพลาดในบางกรณี โครงสร้างมนุษย์ในแกน U จำแนกท่าทางได้  
ถูกต้องแต่โครงสร้างมนุษย์ในแกน V ผิดพลาดหรือโครงสร้างมนุษย์ในแกน U ผิดพลาด แต่  
โครงสร้างมนุษย์ในแกน V จำแนกท่าทางได้ถูกต้อง

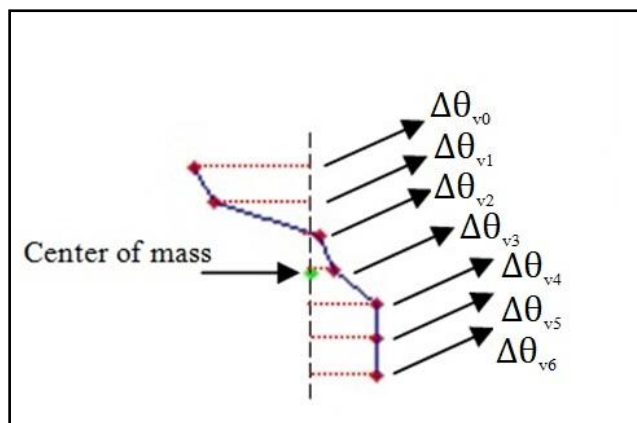
ดังนั้นจึงมีแนวคิดที่จะรวม color u-disparity และ color v-disparity เข้าด้วยกัน เพื่อที่ลด  
ความซับซ้อนที่ของข้อมูลและแก้ปัญหาความผิดพลาดของโครงสร้างมนุษย์ที่เกิดขึ้น

### 3.3.3. การรวมโครงสร้างมนุษย์ (Structure of Human of Combination)

การรวมโครงสร้างมนุษย์เป็นการหาความสัมพันธ์ของโครงสร้างมนุษย์ในแกน U และ V  
ขั้นตอนการรวมจะเริ่มจากการปรับทิศทางของแกนโครงสร้างมนุษย์จากแกน V ให้อยู่ในทิศทาง  
เดียวกับแกน U ก่อน โดยการหมุนภาพ 90 องศา และตามด้วยการกลับภาพดังสมการ (3.15)

$$\hat{I}_V = -R_{90}(I_V) \quad (3.15)$$





(ข)

ภาพประกอบ 3- 13 การคำนวณระยะห่างระหว่างจุดอ้างอิงกับจุดของโครงสร้างมนุษย์ :

(ก) บนแกน U (ข) บนแกน V

จากภาพประกอบ 3-13 เป็นตัวอย่างการคำนวณหาระยะห่างระหว่างจุดอ้างอิงกับจุดของโครงสร้างมนุษย์บนแกน U และ V ก่อนนำระยะห่างมาบวกกันตามสมการ (3.17) โดยผลลัพธ์จะได้ค่า  $\Delta\theta_i$  จำนวน 7 ค่าสำหรับใช้สร้างโครงสร้างมนุษย์ใหม่ โดยความสูงของโครงสร้างมนุษย์แบบใหม่จะเปลี่ยนแปลงตามความสูงของแต่ละคนที่ผ่านเข้ามาในเฟรมภาพ ซึ่งทำให้เกิดความซับซ้อนของข้อมูลได้ ดังนั้นจึงจะนำโครงสร้างใหม่มาผ่านกระบวนการ นอร์มัลไลเซชัน (normalization) โดยการกำหนดสเกล (scale) ความสูงของมนุษย์ให้ไม่เกิน 100 ดังสมการ (3.21)

$$S(h) = \frac{h}{100} \quad (3.20)$$

$$H = S(h) * 100 \quad (3.21)$$

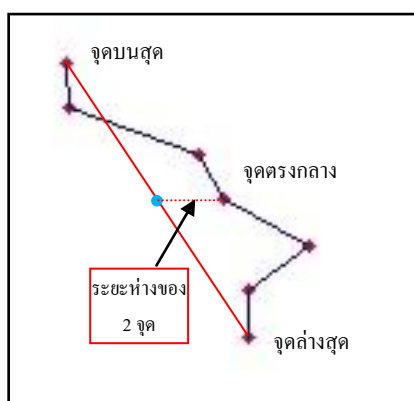
เมื่อ  $S(h)$  เป็นอัตราของความสูงมีค่า 0 - 1 ,  $H$  เป็นค่าความสูงที่กำหนดสเกลไม่เกิน 100

ทิศทางการหันหน้าของโครงสร้างมนุษย์ก็มีผลต่อการรู้จำท่าทาง เพราะฉะนั้นจะต้องลดความซับซ้อนที่อาจจะทำให้เกิดข้อผิดพลาดต่อระบบการรู้จำท่าทาง โดยจะสร้างมาตรฐานทิศทางการหันของโครงสร้างมนุษย์ให้เหมือนกัน จึงกำหนดให้โครงสร้างมนุษย์ทั้งหมดจะต้องมีทิศทางหันไปทางขวาไม่ใช่หันไปทางซ้ายดังภาพประกอบ 3-13 โดยขั้นตอนแรกจะต้องคำนวณหาทิศทางของโครงสร้าง เพื่อใช้สำหรับกลับทิศทางในกรณีที่หันซ้าย การคำนวณหาทิศทางจะใช้ข้อมูลจากโครงสร้างมนุษย์ 3 จุด ได้แก่ จุดบนสุด ( $\Delta\theta_0$ ) จุดตรงกลาง ( $\Delta\theta_3$ ) และจุดล่างสุด ( $\Delta\theta_6$ ) โดยใช้จุดบนสุดและจุดล่างสุดนำมาคำนวณหาความชัน (slope: M) ดังสมการ (3.22) แล้วสร้างเส้นตรงขึ้นดังสมการ (3.23) โดยจะนำจุดบนเส้นตรงที่อยู่แกน x เดียวกับจุดตรงกลางของ

ข้อมูลโครงสร้างมาเปรียบเทียบ ถ้าค่าแกน  $y$  ของจุดบนเส้นตรงมีค่าน้อยกว่าค่าแกน  $y$  ของจุดตรงกลางของข้อมูลโครงสร้างมนุษย์แสดงว่า โครงสร้างมนุษย์มีทิศทางไปทางซ้าย ดังภาพประกอบ 3-14

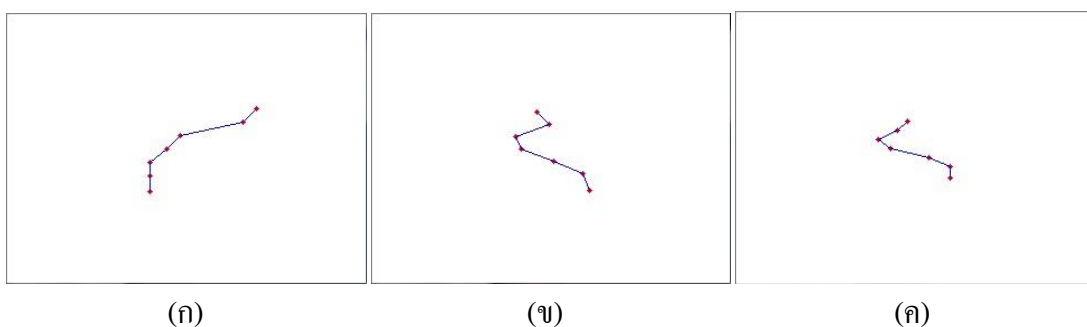
$$M = \frac{\Delta\theta_0(y) - \Delta\theta_6(y)}{\Delta\theta_0(x) - \Delta\theta_6(x)} \quad (3.22)$$

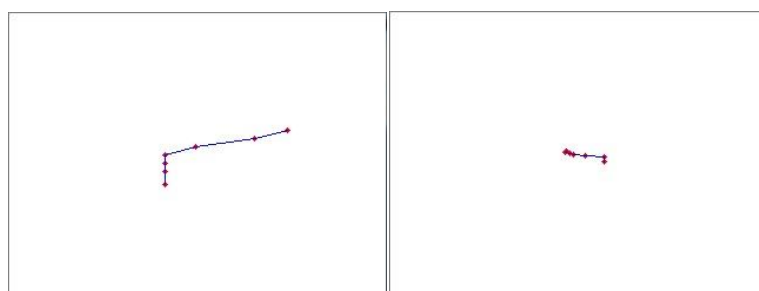
$$Y = MX + C \quad (3.23)$$



ภาพประกอบ 3-14 การคำนวณหาทิศทางของโครงสร้างมนุษย์

จากภาพประกอบ 3-14 เป็นการคำนวณหาทิศทางของโครงสร้างมนุษย์ โดยถ้าค่า  $x$  ที่จุดสี่ฟ้ามี่ค่าน้อยกว่าค่า  $x$  ที่จุด  $\Delta\theta_3$  แสดงว่าโครงสร้างมนุษย์กำลังหันหน้าไปทางซ้าย เมื่อสามารถประมาณหาทิศทางของโครงสร้างมนุษย์ได้ ก็จะสามารกลับทิศทางเพื่อให้โครงสร้างมนุษย์หันหน้าไปทางขวาทางเดียว ดังภาพประกอบ 3-15





(ก)

(ข)

ภาพประกอบ 3- 15 โครงสร้างมนุษย์แบบใหม่: (ก) การขึ้น (ข) การเดิน (ค) การนั่ง  
(ง) การก้ม (จ) การนอน

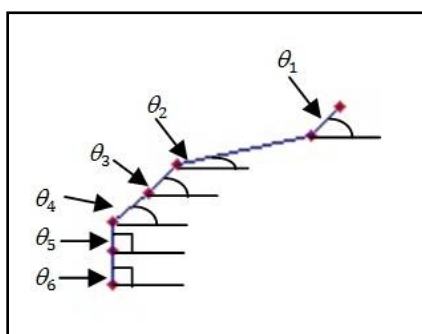
### 3.4. การสกัดลักษณะเด่น

การสกัดลักษณะเด่นจากโครงสร้างมนุษย์ที่ได้เป็นสิ่งสำคัญ ซึ่งจะช่วยให้คอมพิวเตอร์สามารถเรียนรู้และวิเคราะห์พฤติกรรมมนุษย์ได้อย่างเข้าใจเหมือนกับที่มนุษย์เห็น ดังภาพประกอบ 3-15 ถ้ามองด้วยตาก็จะสามารถบ่งบอกลักษณะท่าทางจากโครงสร้างเหล่านี้ได้ไม่ยาก เพื่อให้คอมพิวเตอร์สามารถเข้าใจได้จึงจะใช้มุมของแต่ละจุดที่เรียงต่อกันสำหรับการบ่งบอกลักษณะโครงสร้างมนุษย์

#### การคำนวณมุมสำหรับโครงสร้างมนุษย์

การคำนวณมุมสำหรับเป็นลักษณะเด่นของโครงมนุษย์ จะคำนวณจากจุด 2 จุดที่เรียงติดกัน ดังตัวอย่างภาพประกอบ 3-16 โดยคำนวณได้จากสมการ (3.24) ซึ่งเป็นสมการพื้นฐานสำหรับการหามุมของโครงสร้างมนุษย์

$$\theta_i = \tan^{-1} \left( \frac{p(x_i) - p(x_{i-1})}{p(y_i) - p(y_{i-1})} \right) \quad (3.24)$$

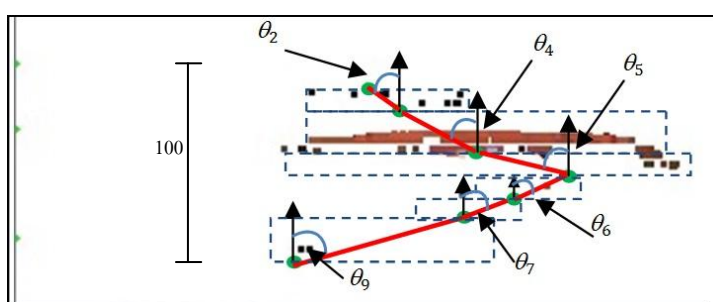


ภาพประกอบ 3- 16 มุมสำหรับโครงสร้างมนุษย์

การหาลักษณะเด่นที่ใช้สำหรับกระบวนการรู้จำท่าทางมนุษย์ แบ่งเป็น 3 แบบ เพื่อทดลองหาลักษณะเด่นที่เหมาะสมและได้ผลลัพธ์ของการรู้จำท่าทางที่ดีที่สุดดังนี้

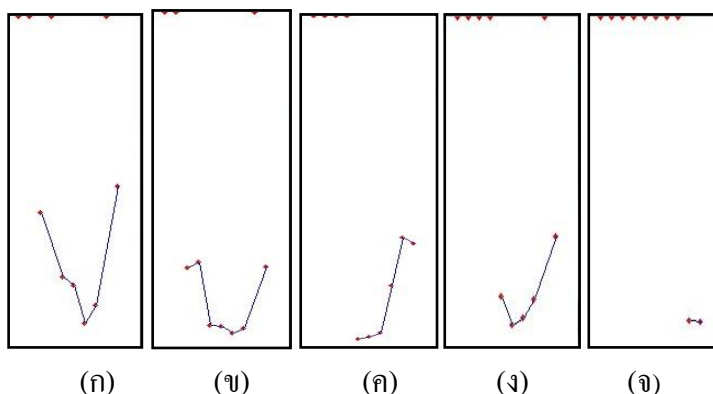
### 3.4.1. การหามุมจากการระบุสีให้กับการประมาณระยะทางของวัตถุของแผนผังคิสพารีที่

การหาลักษณะเด่นวิธีแรกจะหาจากภาพการระบุสีให้กับการประมาณระยะทางของวัตถุของแผนผังคิสพารีที่ ซึ่งโครงสร้างมนุษย์ในแบบนี้จะเป็นแบบฟิกความสูงของโครงสร้างมนุษย์ไว้ที่ 100 แล้วแบ่งจุดกึ่งกลางสำหรับสร้าง โครงสร้างมนุษย์เป็น 10 จุดตามความสูงที่เท่ากัน จากนั้นจึงคำนวณหามุมจากจุดที่เรียงกัน ดังภาพประกอบ 3-17

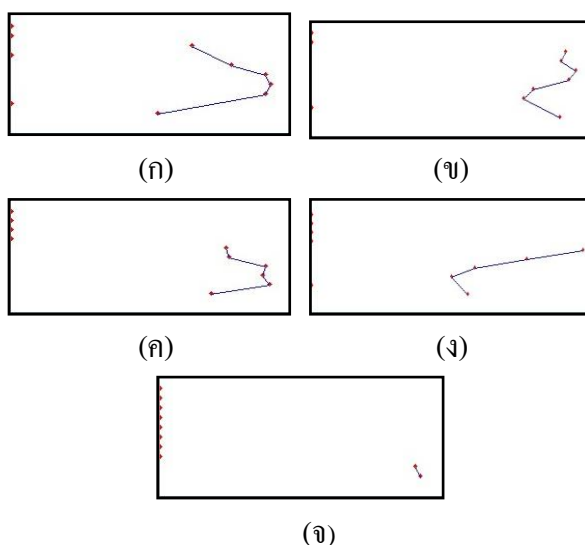


ภาพประกอบ 3- 17 การหามุมจากภาพการระบุสีให้กับการประมาณระยะทางของวัตถุของแผนผังคิสพารีที่

เนื่องจากภาพการระบุสีให้กับการประมาณระยะทางของวัตถุของแผนผังคิสพารีที่จะประกอบด้วย 2 ภาพ คือภาพจากแกน U และ V ซึ่งแต่ละภาพจะได้ลักษณะเด่นของท่าทางมนุษย์จำนวน 9 มุม ดังนั้นจึงมีลักษณะเด่นทั้ง 18 มุม คือ  $F_1 = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}, \theta_{15}, \theta_{16}, \theta_{17}, \theta_{18}]$  ผลลัพธ์ดังภาพประกอบ 3-18, 3-19



ภาพประกอบ 3- 18 โครงสร้างมนุษย์แบบฟิกความสูงของแกน V: (ก) การยืน (ข) การเดิน (ค) การนั่ง (ง) การก้ม (จ) การนอน



ภาพประกอบ 3- 19 โครงสร้างมนุษย์แบบพีกความสูงของแกน U: (ก) การยื่น (ข) การเดิน  
(ค) การนั่ง (ง) การก้ม (จ) การนอน

### 3.4.2. การหามุมจากการรวมโครงสร้างมนุษย์

การหาลักษณะเด่นวิธีที่สองจะหาจากภาพการรวมโครงสร้างมนุษย์ โดยจะทำการหามุมเพื่อบ่งบอกโครงสร้างของท่าทางมนุษย์ ซึ่งจะได้ลักษณะเด่นทั้งหมด 6 มุม คือ  $F_2=[\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6]$

### 3.4.3. การหามุมและค่าเฉลี่ยความสูงจากการรวมโครงสร้างมนุษย์

การหาลักษณะเด่นวิธีที่สามจะเหมือนกับวิธีที่สอง คือ ได้มุมจากภาพการรวมโครงสร้างมนุษย์ทั้ง 6 มุม แต่จะเพิ่มลักษณะเด่นในส่วน of ค่าเฉลี่ยความสูงของโครงสร้างมนุษย์เข้ามา โดยการหาค่าเฉลี่ยความสูงจะเริ่มจากการหาค่าความสูงที่มากที่สุดของโครงสร้างมนุษย์แต่ละคน แล้วนำมาหาค่าเฉลี่ยของความสูงของโครงสร้างมนุษย์แต่ละท่าทาง ดังสมการ 3.25

$$A(H) = \frac{H}{\max(H)} \quad (3.25)$$

เมื่อ  $A(H)$  เป็นค่าเฉลี่ยความสูงของโครงสร้างมนุษย์ที่ทำทางนั้น,  $H$  เป็นความสูงของโครงสร้างมนุษย์ที่ทำทางนั้น และ  $\max(H)$  เป็นค่าความสูงที่มากที่สุดของโครงสร้างมนุษย์คนนั้น โดยค่าเฉลี่ยความสูงของโครงสร้างมนุษย์จะมีค่าระหว่าง 0 – 1 ซึ่งจะได้ลักษณะเด่นทั้งหมด 7 ข้อมูล คือ  $F_3=[\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, H]$



### 3.5. การรู้จำท่าทางมนุษย์ (Human Recognition)

การสร้างโมเดลท่าทางมนุษย์ในระบบนี้จะประกอบไปด้วย 5 ท่าทาง ได้แก่ การยืน การเดิน การนั่ง การก้ม และการนอน ซึ่งเป็นท่าทางพื้นฐานที่นำไปสู่พฤติกรรมต่างๆที่มนุษย์ปฏิบัติในชีวิตประจำวัน โดยสามารถอธิบายลักษณะเด่นของแต่ละท่าทางได้ด้วยลักษณะโครงสร้างมนุษย์ที่ต่างกันดังนี้

ถ้าใช้ลักษณะเด่นที่เป็นมุมจากภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริที ดังภาพประกอบ 3-18, 3-19 ลักษณะท่าทางโครงสร้างมนุษย์ที่ได้จากแกน U และ V บางกรณีสามารถจำแนกกว่าเป็นท่าทางใดท่าทางหนึ่งได้ทั้ง 2 แกน หรือบางกรณีสามารถจำแนกท่าทางได้เพียงแกนใดแกนหนึ่งเท่านั้น โดยลักษณะของแต่ละท่าทางจะเป็นดังภาพประกอบ 3-18, 3-19 คือ

การยืนและการเดินจะมีลักษณะโครงสร้างมนุษย์ใกล้เคียงกัน โดยจะมีลักษณะโค้งในแนวนอน เพราะจากภาพในมุมด้านบนศีรษะ ใหญ่ ลำตัว และขาจะไม่ตรงกันค่อนข้างมีส่วนโค้ง การนั่ง จะนั่งบนเก้าอี้ ซึ่งขาจะเป็นแนวนอนตามเก้าอี้ โดยจากภาพตัวอย่างจะสามารถจำแนกท่าทางได้เฉพาะแกน V

การก้ม ศีรษะและลำตัวจะมีลักษณะโค้งไปด้านบน

การนอน ความสูงของโครงสร้างมนุษย์จะมีระดับต่ำ สามารถจำแนกได้อย่างชัดเจน

ซึ่งจะถูกจัดให้อยู่ในรูปของ  $F_1 = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}, \theta_{15}, \theta_{16}, \theta_{17}, \theta_{18}]$  สามารถแสดงตัวอย่างของลักษณะเด่นของโครงสร้างมนุษย์ดังตาราง 3-1

ตารางที่ 3- 1 ลักษณะเด่นของแต่ละท่าทาง โครงสร้างมนุษย์จากภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริที

ลักษณะเด่น	ท่าทาง			
	การยืน, การเดิน	การนั่ง	การก้ม	การนอน
$\theta_1$	-1.000000	-1.000000	-1.000000	-1.000000
$\theta_2$	-1.000000	-1.000000	-1.000000	-1.000000
$\theta_3$	56.309933	-1.000000	-1.000000	-1.000000
$\theta_4$	82.504142	-1.000000	-1.000000	-1.000000
$\theta_5$	63.434949	-16.699244	-54.462322	-1.000000
$\theta_6$	59.534455	-26.565051	-70.974394	-1.000000
$\theta_7$	-75.256437	-80.837653	-68.198591	-1.000000

$\theta_8$	79.114473	-80.980678	-1.000000	68.198591
$\theta_9$	-85.763605	38.659808	-40.364537	-70.974394
$\theta_{10}$	-1.000000	-1.000000	-1.000000	-1.000000
$\theta_{11}$	-1.000000	-1.000000	-1.000000	-1.000000
$\theta_{12}$	60.945396	-1.000000	-1.000000	-1.000000
$\theta_{13}$	45.000000	-1.000000	-1.000000	-1.000000
$\theta_{14}$	72.121303	16.699244	-81.119341	-1.000000
$\theta_{15}$	75.963757	74.875993	-82.874984	-1.000000
$\theta_{16}$	70.974394	-16.699244	-79.114473	-1.000000
$\theta_{17}$	-1.000000	34.992020	-1.000000	73.141601
$\theta_{18}$	-81.634114	-80.380272	-14.036243	75.963757

ถ้าใช้ลักษณะเด่นที่เป็นมุมจากภาพการรวมโครงสร้างมนุษย์ , มุมและค่าเฉลี่ยความสูงจากภาพการรวมโครงสร้างมนุษย์ จะมีลักษณะท่าทางโครงสร้างมนุษย์ดังภาพประกอบ3-15 คือ

การยืน จะมีลักษณะโครงสร้างเป็นขาตรงทำมุม 90 องศา ส่วนลำตัวกับศีรษะจะโค้ง หรือจะมีลักษณะตรงทั้งโครงสร้าง

การเดิน จะมีลักษณะโครงสร้างในส่วนขาจะทำมุมขึ้นไปข้างหน้า ตามการเดินของเท้า

การนั่ง คือจะนั่งบนเก้าอี้ ซึ่งจะมีโครงสร้างตามเก้าอี้ ส่วนของศีรษะกับลำตัวจะค่อนข้างตรง ส่วนต้นขาจะเป็นแนวนอนและส่วนหน้าแข้งจะเป็นแนวตั้งตามเก้าอี้

การนอน จะมีลักษณะโครงสร้างเป็นแนวนอนอย่างชัดเจน

ซึ่งลักษณะเด่นของทั้ง 2 แบบจะถูกจัดให้อยู่ในรูป  $F_2=[\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6]$  และ  $F_3=[\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6,H]$  สามารถแสดงได้ดังตารางที่ 3-2 โดยลักษณะเด่นทั้งสองแบบจะมีความแตกต่างกันที่การใช้หรือไม่ใช้ค่าเฉลี่ยความสูง

ตารางที่ 3- 2 ลักษณะเด่นของแต่ละท่าทาง

ท่าทาง	ลักษณะเด่น						
	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$	$\theta_6$	H(ความสูง)
การยืน	-45.0	33.690068	-59.036244	-20.854458	-36.869898	-59.420773	0.985915
การเดิน	45.0	16.886791	54.782407	48.576334	90.0	90.0	1
การนั่ง	-84.80557	29.744881	-38.157227	-19.536655	-24.623565	90.0	0.704225
การก้ม	13.240520	10.007980	57.994617	90.0	90.0	90.0	0.577465

การนอน	45.0	-26.565051	-18.434949	-5.710593	-7.125016	90.0	0.126761
--------	------	------------	------------	-----------	-----------	------	----------

การรู้จำพฤติกรรมมนุษย์ในระบบนี้จะใช้เทคนิคของข่ายงานประสาท (neural network) เพื่อจำลองการทำงานของเครือข่ายประสาท เหมือนใน สมองมนุษย์ เพื่อที่จะสร้างเครื่องมือซึ่งมีความสามารถในการเรียนรู้การจดจำรูปแบบ (Pattern Recognition) และการอุปมาความรู้ เช่นเดียวกับความสามารถที่มีในสมองมนุษย์ โดยระบบนี้จะเลือกใช้การเรียนรู้แบบมีการสอน (supervised learning) สถาปัตยกรรมโครงข่ายแบบ feedforward network การทำงานจะถูกแบ่งเป็น 3 layer คือ input node, hidden node และ output node โดยข้อมูลที่ประมวลผลจะอยู่ใน input node แล้วจะถูกส่งต่อมาที่ hidden node เพื่อนำมาคูณกับ weight ที่เหมาะสมแล้วจึงตัดสินใจแบ่งกลุ่มให้กับข้อมูลที่เข้ามา ซึ่งจะถูกระบุไว้ที่ output node

ในวิทยานิพนธ์นี้จะประกอบด้วยลักษณะเด่นที่จะใช้เป็น input ในข่ายงานประสาทจำนวน 3 รูปแบบ คือ  $F_1=[\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}, \theta_{15}, \theta_{16}, \theta_{17}, \theta_{18}]$ ,  $F_2=[\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6]$ ,  $F_3=[\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, H]$  ในส่วนของ hidden node จะสุ่มจำนวน โหนดสำหรับแต่ละ input ที่ต่างกัน ซึ่งจำนวน โหนดที่ทำให้ได้ผลลัพธ์ที่ดีของแต่ละ input เป็นดังนี้

ตารางที่ 3- 3 จำนวนโหนดที่ใช้สำหรับข่ายงานประสาท

ลักษณะเด่น	จำนวน โหนด
$F_1$	5
$F_2$	36
$F_3$	7

ในส่วนของ output จะแบ่งออกเป็น 4 กลุ่ม คือ การยืนหรือการเดิน (0001) การนั่ง (0010) การก้ม (0100) การนอน (1000) สาเหตุที่การยืนกับการเดินถูกกำหนดให้อยู่ในกลุ่มเดียวกันเพราะ โครงสร้างมนุษย์มีลักษณะใกล้เคียงกัน แต่จะจำแนกได้จากการเคลื่อนที่ของ โครงสร้างมนุษย์ในกลุ่มนี้

ข่ายงานประสาทในวิทยานิพนธ์นี้จะนำ tool สำเร็จรูปที่มีอยู่ใน Matlab มาใช้งาน ซึ่งมีชื่อว่า Neural network pattern recognition tool (nprtool) โดยการทำงานของ nprtool สามารถทำได้โดยการใส่ input และ output โดยกำหนดให้ input มี 400 ข้อมูล ข้อมูลที่ใช้สำหรับการรู้จำท่าทางจะถูกแบ่งออกเป็น 3 ส่วนคือ การฝึกสอน (train) การตรวจสอบ (validation) และการทดสอบ (test) ซึ่ง จะกำหนดสัดส่วนเป็น การฝึกสอน 80% การตรวจสอบ 10% การทดสอบ 10%

### 3.6. สรุป

ในงานวิจัยนี้ได้ออกแบบและพัฒนาระบบการวิเคราะห์ท่าทางมนุษย์ โดยนำเสนอการใช้ข้อมูลความลึก สี และการเคลื่อนที่ของวัตถุสำหรับใช้สร้างภาพเสมือน โครงสร้างมนุษย์จากภาพมุมด้านบน ระบบจะรับข้อมูลภาพมาจากกล้องไคน์คอร์ทที่ติดตั้งอยู่บนเพดาน ซึ่งประกอบด้วย ภาพความลึกและภาพสี ขั้นตอนแรกนำภาพความลึกเข้าสู่กระบวนการตรวจจับและติดตามการเคลื่อนที่ของวัตถุ โดยใช้เทคนิคการลบภาพพื้นหลังและการติดตามวัตถุเพื่อจะได้บริเวณกับตำแหน่งของวัตถุที่จะนำมาใช้ในการประมวลภาพ ต่อไปนำภาพสีเข้าสู่กระบวนการจับคู่สี โดยจะนำภาพสีมาแมทช์กับผลลัพธ์ของกระบวนการตรวจจับและติดตามการเคลื่อนที่ของวัตถุเพื่อจะได้วัตถุที่ถูกแมทช์กับสี ต่อไปเป็นกระบวนการวิเคราะห์โครงสร้างมนุษย์สำหรับใช้สร้างภาพเสมือน โครงสร้างมนุษย์ ซึ่งจะแบ่งเป็น 3 ขั้นตอนใหญ่ คือ 1. การฉายโครงสร้างมนุษย์ด้วยการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริติ โดยการนำผลลัพธ์ของกระบวนการตรวจจับและติดตามการเคลื่อนที่ของวัตถุมาผ่านกระบวนการประมาณระยะทางของวัตถุจากแผนผังคิสพาริติ ซึ่งจะได้ภาพลักษณะท่าทางของมนุษย์มา 2 แกนคือ U และ V แล้วจึงนำผลลัพธ์นี้เข้าสู่กระบวนการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริติ โดยจะแมทช์กับผลลัพธ์ของกระบวนการจับคู่สี แล้วจะได้ภาพลักษณะท่าทางมนุษย์ที่ถูกระบุสี 2 ภาพจากแกน U และ V นำไปสร้างภาพโครงสร้างมนุษย์ โดยการประมาณจุดกึ่งกลางของลักษณะท่าทางมนุษย์ แล้วจึงเข้าสู่กระบวนการรวมโครงสร้างมนุษย์ เพื่อลดความซับซ้อนและความผิดพลาดที่เกิดจากภาพโครงสร้างมนุษย์ที่มี 2 แกน แล้วจึงนำมาหาลักษณะเด่นของโครงสร้างมนุษย์ โดยจะแบ่งออกเป็น 3 ลักษณะเด่นสำหรับใช้ทดสอบระบบ คือ 1. ลักษณะเด่นที่ได้จากการคำนวณมุมของโครงสร้างมนุษย์จากผลลัพธ์ของกระบวนการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริติจำนวน 18 ข้อมูล 2. ลักษณะเด่นที่ได้จากการคำนวณมุมของโครงสร้างมนุษย์จากผลลัพธ์ของการรวมโครงสร้างมนุษย์จำนวน 6 ข้อมูล 3. ลักษณะเด่นที่ได้จากการคำนวณมุมและค่าเฉลี่ยความสูงจากโครงสร้างมนุษย์จากผลลัพธ์ของการรวมโครงสร้างมนุษย์จำนวน 7 ข้อมูล จากนั้นเข้าสู่กระบวนการรู้จำท่าทางมนุษย์ โดยใช้เทคนิคข่ายงานประสาทสำหรับจำแนกท่าทางมนุษย์ โดยจะแบ่งเป็น 4 กลุ่ม คือ การยืนหรือการเดิน การนั่ง การก้ม การนอน

## บทที่ 4

### ผลการทดลอง

ในบทนี้ได้ทำการทดสอบการรู้จำท่าทางมนุษย์ ซึ่งประกอบด้วยท่าทางพื้นฐานทั้ง 5 ท่าทาง ได้แก่ การยืน การเดิน การนั่ง การก้ม และการนอน โดยระบบจะรองรับการวิเคราะห์ท่าทางมนุษย์ จากคนเพียงคนเดียวในแต่ละเฟรม และสามารถจำแนกกลุ่มของท่าทางในรูปแบบ real time ได้ อย่างมีประสิทธิภาพ

สำหรับการทดสอบประสิทธิภาพการรู้จำท่าทางมนุษย์จะขึ้นอยู่กับลักษณะเด่นที่ใช้ในการทดสอบ ซึ่งถูกแบ่งออกเป็น 3 ลักษณะเด่น ได้แก่ 1. การใช้มุมจากโครงสร้างมนุษย์ในภาพการระบุ

สีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริตี ซึ่งประกอบด้วย 18 ข้อมูล  $F_1=[\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}, \theta_{15}, \theta_{16}, \theta_{17}, \theta_{18}]$  2. การใช้มุมจาก

โครงสร้างมนุษย์ในภาพการรวมโครงสร้างมนุษย์ ซึ่งประกอบด้วย 6 ข้อมูล

$F_2=[\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6]$  3. การใช้มุมและค่าเฉลี่ยความสูงจากโครงสร้างมนุษย์ในภาพการรวมโครงสร้างมนุษย์ ซึ่งประกอบด้วย 7 ข้อมูล  $F_3=[\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, H]$  ซึ่งทั้ง 3 ลักษณะเด่นจะเป็นข้อมูลของมุมที่ใช้สำหรับการระบุโครงสร้างท่าทางมนุษย์ แต่จะมีจำนวนของข้อมูลที่แตกต่างกัน

#### 4.1. การทดลองที่ 1 การรู้จำท่าทางมนุษย์โดยใช้มุมจากโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริตี

การทดลองที่ 1 ใช้ลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริตี ซึ่งประกอบด้วย 18 ข้อมูลได้มาจากภาพในแกน U และ V อย่างละ 9 ข้อมูล ดังรูปแบบของ  $F_1=[\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}, \theta_{15}, \theta_{16}, \theta_{17}, \theta_{18}]$  ตัวอย่างของลักษณะเด่นนี้สำหรับแต่ละท่าทางมนุษย์ดังตาราง 3-1 และภาพประกอบ 3-18, 3-19

##### 4.1.1. สมมติฐาน

การใช้ลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริตี จะประกอบด้วยโครงสร้างมนุษย์ 2 แบบจากแกน U และ V ซึ่งจะมีความสัมพันธ์กันในเชิงความลึกทำให้ทราบความกว้าง ความยาวของโครงสร้างมนุษย์สำหรับ

นำไปใช้ในการวิเคราะห์ท่าทางมนุษย์แล้วยังมีการพิจารณาความสูงของโครงสร้างมนุษย์ไว้ โดยแบ่งโครงสร้างมนุษย์เป็น 9 ส่วนเท่าๆกัน ทำให้จำแนกแต่ละท่าทางได้จากความสูงที่แตกต่างกันไป

#### 4.1.2. ปัจจัยกำหนดในการทดลอง

1) ฐานข้อมูลวิดีโอที่ติดตั้งบนความสูง 3 เมตร ประกอบด้วยการจำลองท่าทางของมนุษย์จำนวน 4 คน ทำการคัดข้อมูลสำหรับการเทรนจำนวน 400 ข้อมูล แบ่งเป็นการยืนหรือการเดิน 100 ข้อมูล การนั่ง 100 ข้อมูล การก้ม 100 ข้อมูล และการนอน 100 ข้อมูล

2) จำนวนฮิดเดน โหนด (hidden node) ที่ใช้สำหรับการเทรนข้อมูล คือ 5, 10, 15, 20, 25, 30, 35, 40

#### 4.1.3. ผลการทดลอง

ผลการทดลองจากการใช้ลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับ การประมาณระยะทางของวัตถุจากแผนผังดิสพาริตีที่จะได้มาจากกล้องเครื่องมือสำเร็จรูปของเมทแลปคือ Neural Pattern Recognition Tool (nprtool) โดยการนำอินพุตจากฐานข้อมูล และเอาต์พุตที่กำหนดขึ้นได้แก่ การยืนหรือการเดิน (0001), การนั่ง (0010), การก้ม (0100) และการนอน (1000) นำเข้าไปเทรนเพื่อหาค่าน้ำหนัก (weight) ที่เหมาะสมสำหรับนำกลับมาใช้คู่กับอินพุตเพื่อตัดสินใจในการเลือกกลุ่มของเอาต์พุตตามที่เทรนไป โดยที่ nprtool สามารถคำนวณความถูกต้องของผลลัพธ์ได้ โดยจะแบ่งฐานข้อมูลออกเป็น 3 ส่วนคือ การฝึกสอน (training) 80%, การตรวจสอบ (Validation) 10% และการทดสอบ (test) 10% การกำหนดจำนวนฮิดเดน โหนดที่จะใช้สำหรับหาค่าน้ำหนักที่เหมาะสมจึงเป็นสิ่งสำคัญ เพื่อให้ใช้ฮิดเดน โหนดให้น้อยที่สุดสำหรับผลลัพธ์ที่มีประสิทธิภาพ โดยการสุ่มผลลัพธ์ของฮิดเดน โหนด ซึ่งจะเพิ่มขึ้นทีละ 5 โหนด ดังตารางที่ 4-1

ตารางที่ 4- 1 อัตราความถูกต้องสำหรับการสุ่มจำนวนโหนดของการทดลองที่ 1

จำนวนฮิดเดนโหนด	อัตราความถูกต้อง (%)
5	99.0
10	99.0
15	99.3
20	99.5
25	99.5
30	99.5

35	99.8
40	99.8

จากตารางที่ 4-1 สามารถบอกได้ว่าจำนวนโหนดที่เพิ่มขึ้นจะทำให้อัตราความถูกต้องเพิ่มขึ้น แต่จะเพิ่มขึ้นน้อยมาก ดังนั้นจึงเลือกใช้ผลลัพธ์จาก 5 โหนดสำหรับใช้ในการรู้จำท่าทางมีความถูกต้อง 99.0% ซึ่งถือเป็นผลลัพธ์ที่ดี เมื่อเทียบกับจำนวนโหนดที่เพิ่มขึ้นซึ่งจะทำให้ความเร็วในการวิเคราะห์ของระบบเพิ่มขึ้นด้วย แสดงผลลัพธ์อย่างละเอียดของ 5 โหนดดังตารางที่ 4-2

ตารางที่ 4- 2 แสดงผลลัพธ์จากการใช้ลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสปาริตี

ท่าทาง	วิธีการ			
	การฝึกสอน	การตรวจสอบ	การทดสอบ	ผลลัพธ์
การขึ้นและการเดิน	100%	100%	100%	100%
การนั่ง	98.8%	100%	100%	99.0%
การก้ม	98.8%	90.0%	90.0%	97.1%
การนอน	100%	100%	100%	100%
				99.0%

สามารถแสดงผลของความถูกต้องจากการใช้ลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสปาริตีสำหรับการฝึกสอนอย่างละเอียดดังตารางที่ 4-3 สำหรับการตรวจสอบอย่างละเอียดดังตารางที่ 4-4 สำหรับการทดสอบอย่างละเอียดดังตารางที่ 4-5

ตารางที่ 4- 3 ความถูกต้องของการฝึกสอนสำหรับลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสปาริตี

ข้อมูล	จำนวน	วิเคราะห์ท่าทาง				ความถูกต้อง(%)
		การนอน	การก้ม	การนั่ง	การขึ้นและการเดิน	
การขึ้นและการเดิน	81	0	0	0	81	100
การนั่ง	82	0	1	81	0	98.8
การก้ม	82	0	81	1	0	98.8

การนอน	75	75	0	0	0	100
						99.4

จากตาราง 4-3 ความถูกต้องของการฝึกสอนข้อมูลสำหรับทำการยื่นหรือการเดินจำนวน 81 ข้อมูล ฝึกสอนได้ความถูกต้อง 100 % , ทำการนั่งจำนวน 82 ข้อมูล วิเคราะห์เป็นทำการก้ม 1 ข้อมูล ที่เหลือถูกต้อง ฝึกสอนได้ความถูกต้อง 98.8 % , ทำการก้มจำนวน 82 ข้อมูล วิเคราะห์เป็นทำการนั่ง 1 ข้อมูล ที่เหลือถูกต้อง ฝึกสอนได้ความถูกต้อง 98.8 % , ทำการนอนจำนวน 75 ข้อมูล ฝึกสอนได้ความถูกต้อง 100 %

ตารางที่ 4- 4 ความถูกต้องของการตรวจสอบสำหรับลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริที

ข้อมูล	จำนวน	วิเคราะห์ท่าทาง				ความถูกต้อง(%)
		การนอน	การก้ม	การนั่ง	การยื่นและการเดิน	
การยื่นและการเดิน	10	0	0	0	10	100
การนั่ง	10	0	0	10	0	100
การก้ม	10	1	9	0	0	90.0
การนอน	10	10	0	0	0	100
						97.5

จากตาราง 4-4 ความถูกต้องของการตรวจสอบข้อมูลสำหรับทำการยื่นหรือการเดินจำนวน 10 ข้อมูล ตรวจสอบได้ความถูกต้อง 100 % , ทำการนั่งจำนวน 10 ข้อมูล ตรวจสอบได้ความถูกต้อง 100 % , ทำการก้มจำนวน 10 ข้อมูล วิเคราะห์เป็นทำการนอน 1 ข้อมูล ที่เหลือถูกต้อง ตรวจสอบได้ความถูกต้อง 90.0 % , ทำการนอนจำนวน 10 ข้อมูล ตรวจสอบได้ความถูกต้อง 100 %

ตารางที่ 4- 5 ความถูกต้องของการทดสอบสำหรับลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริที

ข้อมูล	จำนวน	วิเคราะห์ท่าทาง				ความถูกต้อง(%)
		การนอน	การก้ม	การนั่ง	การยื่นและการเดิน	
การยื่นและการเดิน	9	0	0	0	9	100
การนั่ง	8	0	0	8	0	100



การก้ม	10	1	9	0	0	90.0
การนอน	13	13	0	0	0	100
						97.5

จากตาราง 4-5 ความถูกต้องของการทดสอบข้อมูลสำหรับท่าการยืนหรือการเดินจำนวน 9 ข้อมูล ทดสอบได้ความถูกต้อง 100 % , ท่าการนั่งจำนวน 8 ข้อมูล ทดสอบได้ความถูกต้อง 100 % , ท่าการก้มจำนวน 10 ข้อมูล วิเคราะห์เป็นท่าการนอน 1 ข้อมูล ที่เหลือถูกต้อง ทดสอบได้ความถูกต้อง 90.0 % , ท่าการนอนจำนวน 13 ข้อมูล ทดสอบได้ความถูกต้อง 100 %

#### 4.1.4. วิเคราะห์และสรุปผลการทดลอง

จากตารางที่ 4-2 แสดงผลลัพธ์ที่ได้จากการทดลองการใช้ลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังทัศนพริทัศน์ ซึ่งสามารถสรุปผลการทดลองได้ว่า ความถูกต้องในการรู้จำท่าการยืน 100 % , การนั่ง 99.0% , การก้ม 97.1% และการนอน 100% ค่าเฉลี่ยความถูกต้องคือ 99.0% ซึ่งถือว่ามีประสิทธิภาพดีมากในการรู้จำท่าทางต่างๆ แต่การเลือกใช้ลักษณะเด่นในการทดลองนี้จะมีจำนวนลักษณะเด่น 18 ข้อมูล ซึ่งถือว่ามีจำนวนมากจะส่งผลทำให้การวิเคราะห์ท่าทางมนุษย์ช้าลง และการพีกระดับความสูงให้กับโครงสร้างมนุษย์ที่ 100 จะทำให้ระบบการวิเคราะห์ท่าทางมนุษย์ไม่สามารถทำงานได้อย่างครอบคลุมในสถานการณ์ต่างๆ เช่น ถ้าระบบมีการเพิ่มหรือลดความสูงให้กับกล้องวิดีโอที่ติดตั้งบนเพดานระบบก็จะเกิดความผิดพลาดได้ ดังนั้นจึงมีแนวคิดที่จะนำภาพการรวมโครงสร้างมนุษย์มาใช้สำหรับการหามุมของโครงสร้างมนุษย์

#### 4.2. การทดลองที่ 2 การรู้จำท่าทางมนุษย์โดยใช้มุมจากภาพการรวมโครงสร้างมนุษย์

การทดลองที่ 2 ใช้ลักษณะเด่นจากมุมของภาพการรวมโครงสร้างมนุษย์ ซึ่งจะช่วยลดจำนวนข้อมูลที่มากเกินไปจนเกิดความจำเป็น โดยจะทำการรวมภาพจากแกน U และ V เข้าด้วยกันแล้วสร้างโครงสร้างมนุษย์ใหม่ที่มีระดับความสูงเปลี่ยนแปลงไปตามความสูงของคนที่เข้ามาในระบบ ซึ่งจะทำให้ระบบมีความยืดหยุ่นต่อการใช้งานมากยิ่งขึ้น จัดอยู่ในรูปสมการ  $F_1=[\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6]$  ซึ่งประกอบด้วย 6 ข้อมูลสำหรับเป็นอินพุตของระบบ ตัวอย่างลักษณะเด่นดังตารางที่ 3-2 และภาพประกอบ 3-15

#### 4.2.1. สมมติฐาน

การใช้ ลักษณะเด่น จากมุมของภาพการรวมโครงสร้างมนุษย์ จะช่วยลดจำนวนของ ลักษณะเด่นลงจากการทดลองที่ 1 ซึ่งจะทำให้ระบบการวิเคราะห์ท่าทางมนุษย์มีความเร็วและความ ถูกต้องมากขึ้น

#### 4.2.2. ปัจจัยกำหนดในการทดลอง

1). ฐานข้อมูลวิดีโอที่ติดตั้งบนความสูง 3 เมตร ประกอบด้วยการจำลองท่าทางของมนุษย์ จำนวน 4 คน ทำการคัดข้อมูลสำหรับใช้ในการเทรนจำนวน 400 ข้อมูล แบ่งเป็น การยืนหรือการ เดิน 100 ข้อมูล การนั่ง 100 ข้อมูล การก้ม 100 ข้อมูล และการนอน 100 ข้อมูล

2).จำนวนฮิดเดนโหนด (hidden node) ที่ใช้สำหรับการเทรนข้อมูล คือ 6, 12, 18, 24, 30, 36, 42, 48

#### 4.2.3. ผลการทดลอง

ผลการทดลองจากการใช้ ลักษณะเด่น จากมุมของภาพการรวมโครงสร้างมนุษย์ จะใช้การ ทดสอบเหมือนกับในการทดลองที่ 1 คือใช้ nprtool สำหรับคำนวณผลลัพธ์ แต่จะต่างกันที่การสุ่ม จำนวนโหนดที่เพิ่มขึ้น โดยจะเพิ่มขึ้นทีละ 6 โหนด ดังตารางที่ 4-6

ตารางที่ 4- 6 อัตราความถูกต้องสำหรับการสุ่มจำนวนโหนดของการทดลองที่ 2

จำนวนฮิดเดนโหนด	อัตราความถูกต้อง (%)
6	72.0
12	75.8
18	80.8
24	86.0
30	88.8
36	91.5
42	88.3
48	87.0

จากตารางที่ 4-6 สามารถบอกได้ว่าจำนวนฮิดเดนโหนดที่เพิ่มขึ้นจะทำให้ผลลัพธ์ที่ได้ดีขึ้น แต่จะมีผลลัพธ์ที่ดีที่สุดอยู่ที่โหนดที่ 36 หลังจากนั้นอัตราความถูกต้องจะลดลง ดังนั้นจึงเลือกใช้ 36 โหนดสำหรับการทดลองนี้ แสดงผลลัพธ์อย่างละเอียดได้ดังตารางที่ 4-7

ตารางที่ 4- 7 แสดงผลลัพธ์จากการใช้ลักษณะเด่นจากมุมของภาพการรวมโครงสร้างมนุษย์

ท่าทาง	วิธีการ			
	การฝึกสอน	การตรวจสอบ	การทดสอบ	ผลลัพธ์
การยืนและการเดิน	92.0%	80.0%	78.6%	88.9%
การนั่ง	100%	88.9%	83.3%	96.9%
การก้ม	90.0%	75.0%	80.0%	88.3%
การนอน	92.5%	100%	77.8%	92.2%
				91.5%

สามารถแสดงผลลัพธ์ความถูกต้องจากใช้ ลักษณะเด่นจากมุมของภาพการรวมโครงสร้างมนุษย์สำหรับการฝึกสอนอย่างละเอียดดังตารางที่ 4-8 สำหรับการตรวจสอบอย่างละเอียดดังตารางที่ 4-9 สำหรับการทดสอบอย่างละเอียดดังตารางที่ 4-10

ตารางที่ 4- 8 ความถูกต้องของการฝึกสอนสำหรับลักษณะเด่นจากมุมของภาพการรวมโครงสร้างมนุษย์

ข้อมูล	จำนวน	วิเคราะห์ท่าทาง				ความถูกต้อง(%)
		การนอน	การก้ม	การนั่ง	การยืนและการเดิน	
การยืนและการเดิน	75	1	3	2	69	92.0
การนั่ง	75	0	0	75	0	100
การก้ม	90	3	81	0	6	90.0
การนอน	80	74	4	0	2	92.5
						93.4

จากตาราง 4-8 ความถูกต้องของการฝึกสอนข้อมูลสำหรับการยืนหรือการเดินจำนวน 75 ข้อมูล วิเคราะห์เป็นท่าการนอน 1 ข้อมูล วิเคราะห์เป็นท่าการก้ม 3 ข้อมูล วิเคราะห์เป็นท่าการนั่ง 2 ข้อมูล ที่เหลือถูกต้อง ฝึกสอนได้ความถูกต้อง 92 % , ท่าการนั่งจำนวน 75 ข้อมูล ฝึกสอนได้ความถูกต้อง 100 % , ท่าการก้มจำนวน 90 ข้อมูล วิเคราะห์เป็นท่าการนอน 3 ข้อมูล วิเคราะห์เป็นท่าการยืนและการเดิน 6 ข้อมูล ที่เหลือถูกต้อง ฝึกสอนได้ความถูกต้อง 90.0 % , ท่าการนอนจำนวน 80 ข้อมูล วิเคราะห์เป็นท่าการก้ม 4 ข้อมูล วิเคราะห์เป็นท่าการยืนและการเดิน 2 ข้อมูล ที่เหลือถูกต้อง ฝึกสอนได้ความถูกต้อง 92.5 %

ตารางที่ 4-9 ความถูกต้องของการตรวจสอบสำหรับลักษณะเด่นจากมุมมองของ  
ภาพการรวมโครงสร้างมนุษย์

ข้อมูล	จำนวน	วิเคราะห์ทำทาง				ความถูกต้อง(%)
		การนอน	การก้ม	การนั่ง	การขึ้นและการเดิน	
การขึ้นและ การเดิน	10	1	0	1	8	80.0
การนั่ง	9	0	1	8	0	88.9
การก้ม	8	1	6	0	1	75.0
การนอน	13	13	0	0	0	100
						87.5

จากตาราง 4-9 ความถูกต้องของการตรวจสอบข้อมูลสำหรับทำการขึ้นหรือการเดินจำนวน 10 ข้อมูล วิเคราะห์เป็นทำการนอน 1 ข้อมูล วิเคราะห์เป็นทำการนั่ง 1 ข้อมูล ที่เหลือถูกต้อง ตรวจสอบได้ความถูกต้อง 80.0 % , ทำการนั่งจำนวน 9 ข้อมูล วิเคราะห์เป็นทำการก้ม 1 ข้อมูล ที่เหลือถูกต้อง ตรวจสอบได้ความถูกต้อง 88.9 % , ทำการก้มจำนวน 8 ข้อมูล วิเคราะห์เป็นทำการนอน 1 ข้อมูล วิเคราะห์เป็นทำการขึ้นและการเดิน 1 ข้อมูล ที่เหลือถูกต้อง ตรวจสอบได้ความถูกต้อง 75.0 % , ทำการนอนจำนวน 13 ข้อมูล ตรวจสอบได้ความถูกต้อง 100 %

ตารางที่ 4-10 ความถูกต้องของการทดสอบสำหรับลักษณะเด่นจากมุมมองของ  
ภาพการรวมโครงสร้างมนุษย์

ข้อมูล	จำนวน	วิเคราะห์ทำทาง				ความถูกต้อง(%)
		การนอน	การก้ม	การนั่ง	การขึ้นและการเดิน	
การขึ้นและ การเดิน	14	0	0	3	11	78.6
การนั่ง	12	0	0	10	2	83.3
การก้ม	5	0	4	0	1	80.0
การนอน	9	7	1	1	0	77.8
						80.0

จากตาราง 4-10 ความถูกต้องของการทดสอบข้อมูลสำหรับทำการขึ้นหรือการเดินจำนวน 14 ข้อมูล วิเคราะห์เป็นทำการนั่ง 3 ข้อมูล ที่เหลือถูกต้อง ทดสอบได้ความถูกต้อง 78.6 % , ทำการ

นั่งจำนวน 12 ข้อมูล วิเคราะห์เป็นทำการขึ้นและการนอน 2 ข้อมูล ที่เหลือถูกต้อง ทดสอบได้ความถูกต้อง 83.3 % , ทำการก้มจำนวน 5 ข้อมูล วิเคราะห์เป็นทำการขึ้นและการนอน 1 ข้อมูล ที่เหลือถูกต้อง ทดสอบได้ความถูกต้อง 80.0 % , ทำการนอนจำนวน 9 ข้อมูล วิเคราะห์เป็นทำการก้ม 1 ข้อมูล วิเคราะห์เป็นทำการนั่ง 1 ข้อมูล ที่เหลือถูกต้อง ทดสอบได้ความถูกต้อง 77.8 %

#### 4.2.4. วิเคราะห์และสรุปผลการทดลอง

จากตารางที่ 4-7 แสดงผลลัพธ์ที่ได้จากการทดลองการใช้ ลักษณะเด่นจากมุมของภาพการรวมโครงสร้างมนุษย์ สามารถสรุปผลการทดลองได้ว่า ความถูกต้องในการรู้จำทำการขึ้น 88.9% , การนั่ง 96.9% , การก้ม 88.3% และการนอน 92.2% ค่าเฉลี่ยความถูกต้องคือ 91.5% ซึ่งมีประสิทธิภาพดี แต่จากการเปลี่ยนมาใช้โครงสร้างมนุษย์แบบใหม่จะทำให้ความถูกต้องของการจำแนกท่าทางมนุษย์ลดลงเมื่อเทียบกับการทดลองที่ 1 ซึ่งมีผลมาจากการใช้เฉพาะ โครงสร้างมนุษย์แบบใหม่ในการเทรน เพราะลักษณะ โครงสร้างมนุษย์ในทำการขึ้น การก้ม และการนอนในบางภาพมีความผิดพลาดหรืออาจจะเกิดความเหมือนกัน จึงทำให้การรู้จำเกิดความผิดพลาดได้ ส่วนทำการนั่งนั้นลักษณะโครงสร้างจะต่างออกไปมากจึงทำให้ความถูกต้องมาก เพื่อแก้ไขปัญหาในส่วนจึงมีแนวคิดที่นำความสูงของแต่ละ โครงสร้างมาช่วยในการวิเคราะห์ ซึ่งน่าจะทำได้ผลลัพธ์ที่ดีขึ้น

#### 4.3. การทดลองที่ 3 การรู้จำท่าทางมนุษย์โดยใช้มุมและค่าเฉลี่ยความสูงจากภาพการรวมโครงสร้างมนุษย์

การทดลองที่ 3 ใช้มุมและค่าเฉลี่ยความสูงจากภาพการรวมโครงสร้างมนุษย์ โดยจะใช้มุมเหมือนกับการทดลองที่ 2 เพื่อแก้ไขปัญหาการเหมือนกันของลักษณะ โครงสร้างมนุษย์ในบางกรณี โดยมีเงื่อนไขดังนี้

$$1 \geq H_{\text{การขึ้น}} > H_{\text{การนั่ง}} > H_{\text{การก้ม}} > H_{\text{การนอน}} > 0$$

โดยกำหนดให้ H เท่ากับค่าเฉลี่ยความสูง จากเงื่อนไขข้างต้นค่าเฉลี่ยความสูงเข้าใกล้ 1 หรือมีค่าเท่ากับ 1 คือทำการขึ้น ค่าเฉลี่ยความสูงเข้าใกล้ 0 คือทำการนอน ส่วนทำการนั่งจะมีค่าเฉลี่ยความสูงมากกว่าการก้ม แต่จะน้อยกว่าการขึ้น ลักษณะเด่นจัดอยู่ในสมการ  $F_3 = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, H]$  ซึ่งจะประกอบด้วย 7 ข้อมูลสำหรับเป็นอินพุตของระบบ ตัวอย่างลักษณะเด่นดังตารางที่ 3-2 และภาพประกอบ 3-15

#### 4.3.1. สมมติฐาน

การใช้มุมและค่าเฉลี่ยความสูงจากภาพการรวมโครงสร้างมนุษย์ จะช่วยแก้ปัญหาความเหมือนกันของลักษณะโครงสร้างมนุษย์ และจากการใช้โครงสร้างมนุษย์แบบใหม่จะทำให้การใช้งานระบบมีความยืดหยุ่นยิ่งขึ้น โดยเมื่อมีการเปลี่ยนแปลงความสูงของกล้องวิดีโอที่ติดตั้งบนเพดานระบบจะยังสามารถทำงานได้อย่างมีประสิทธิภาพ

#### 4.3.2. ปัจจัยกำหนดในการทดลอง

1).ฐานข้อมูลวิดีโอ ประกอบด้วย 3 ระดับความสูงโดยจะติดตั้งกล้องวิดีโอที่ 3 เมตร, 3.3 เมตร และ 2.7 เมตร ทุกวิดีโอจะมีการจำลองพฤติกรรมของมนุษย์จำนวน 4 คน ซึ่งทำการคัดข้อมูลสำหรับการใช้ในการเทรนจำนวน 400 ข้อมูล แบ่งเป็น การยืนหรือการเดิน 100 ข้อมูล การนั่ง 100 ข้อมูล การก้ม 100 ข้อมูล และการนอน 100 ข้อมูล

2).จำนวนฮิดเดนโหนด (hidden node) ที่ใช้สำหรับการเทรนข้อมูล คือ 7, 14, 21, 28, 35, 42, 49

#### 4.3.3. ผลการทดลอง

ผลการทดลองจากการใช้มุมและค่าเฉลี่ยความสูงจากภาพการรวมโครงสร้างมนุษย์ จะใช้การทดสอบเหมือนกับในการทดลองที่ 1 และ 2 คือใช้ nprtool โดยการนำข้อมูลจากฐานข้อมูลวิดีโอที่ความสูง 3 เมตร ไปทำการเทรนใน nprtool ซึ่งจะสุ่มโหนดสำหรับหาค่าน้ำหนักที่จะทำให้ได้ผลลัพธ์ที่เหมาะสมดังตารางที่ 4-11

ตารางที่ 4- 11 อัตราความถูกต้องสำหรับการสุ่มจำนวนโหนดของ  
การทดลองที่ 3(ที่ความสูง 3 เมตร)

จำนวนฮิดเดนโหนด	อัตราความถูกต้อง (%)
7	99.3
14	99.3
21	99.5
28	90.0
35	99.5
42	99.0
49	99.0

จากตารางที่ 4-11 จะสังเกตเห็นจำนวนโหนดที่เพิ่มขึ้น ไม่ได้ทำให้ความถูกต้องเพิ่มขึ้นมากนัก ดังนั้นจึงเลือกใช้จำนวน 7 โหนดสำหรับการทดลองนี้ เนื่องจากมีการใช้จำนวนโหนดที่น้อยที่สุด แต่ได้ความถูกต้องที่ใกล้เคียงกับจำนวนโหนดที่เพิ่มขึ้น แสดงผลลัพธ์อย่างละเอียดดังตารางที่ 4-12

ตารางที่ 4- 12 แสดงผลลัพธ์จากการใช้ลักษณะเด่นจากมุมและค่าเฉลี่ยความสูงของภาพการรวมโครงสร้างมนุษย์

ท่าทาง	วิธีการ			
	การฝึกสอน	การตรวจสอบ	การทดสอบ	ผลลัพธ์
การยืนและการเดิน	98.7%	100%	100%	99.0%
การนั่ง	100%	100%	100%	100%
การก้ม	98.8%	100%	91.7%	98.0%
การนอน	100%	100%	100%	100%
				99.3%

สามารถแสดงผลของความถูกต้องจากการใช้ลักษณะเด่นจากมุมและค่าเฉลี่ยความสูงของภาพการรวมโครงสร้างมนุษย์สำหรับการฝึกสอนอย่างละเอียดดังตารางที่ 4-13 สำหรับการตรวจสอบอย่างละเอียดดังตารางที่ 4-14 สำหรับการทดสอบอย่างละเอียดดังตารางที่ 4-15

ตารางที่ 4- 13 ความถูกต้องของการฝึกสอนสำหรับลักษณะเด่นจากมุมและค่าเฉลี่ยความสูงของภาพการรวมโครงสร้างมนุษย์

ข้อมูล	จำนวน	วิเคราะห์ท่าทาง				ความถูกต้อง(%)
		การนอน	การก้ม	การนั่ง	การยืนและการเดิน	
การยืนและการเดิน	79	0	1	0	78	98.71
การนั่ง	72	0	0	72	0	100
การก้ม	87	0	80	1	6	98.8
การนอน	88	88	0	0	0	100
						99.4

จากตาราง 4-13 ความถูกต้องของการฝึกสอนข้อมูลสำหรับท่าการยืนหรือการเดินจำนวน 79 ข้อมูล วิเคราะห์เป็นท่าการก้ม 3 ข้อมูล ที่เหลือถูกต้อง ฝึกสอนได้ความถูกต้อง 98.71 % , ท่าการ

นั่งจำนวน 72 ข้อมูล ฝึกสอนได้ความถูกต้อง 100 % , ทำการก้มจำนวน 87 ข้อมูล วิเคราะห์เป็นทำการนั่ง 1 ข้อมูล ที่เหลือถูกต้อง ฝึกสอนได้ความถูกต้อง 98.8 % , ทำการนอนจำนวน 88 ข้อมูล ฝึกสอนได้ความถูกต้อง 100 %

ตารางที่ 4- 14 ความถูกต้องของการตรวจสอบสำหรับลักษณะเด่นจากมุมและค่าเฉลี่ยความสูงของภาพการรวมโครงสร้างมนุษย์

ข้อมูล	จำนวน	วิเคราะห์ท่าทาง				ความถูกต้อง(%)
		การนอน	การก้ม	การนั่ง	การยืนและการเดิน	
การยืนและการเดิน	5	5	0	0	0	100
การนั่ง	8	0	8	0	0	100
การก้ม	16	0	0	16	0	100
การนอน	11	0	0	0	11	100
						100

จากตาราง 4-14 ความถูกต้องของการตรวจสอบข้อมูลสำหรับทำการยืนหรือการเดินจำนวน 5 ข้อมูล ตรวจสอบได้ความถูกต้อง 100 % , ทำการนั่งจำนวน 8 ข้อมูล ตรวจสอบได้ความถูกต้อง 100 % , ทำการก้มจำนวน 16 ข้อมูล ตรวจสอบได้ความถูกต้อง 100 % , ทำการนอนจำนวน 11 ข้อมูล ตรวจสอบได้ความถูกต้อง 100 %

ตารางที่ 4- 15 ความถูกต้องของการทดสอบสำหรับลักษณะเด่นจากมุมและค่าเฉลี่ยความสูงของภาพการรวมโครงสร้างมนุษย์

ข้อมูล	จำนวน	วิเคราะห์ท่าทาง				ความถูกต้อง(%)
		การนอน	การก้ม	การนั่ง	การยืนและการเดิน	
การยืนและการเดิน	11	0	0	0	11	100
การนั่ง	10	0	0	10	0	100
การก้ม	12	0	11	1	0	91.7
การนอน	7	7	0	0	0	100
						97.5



จากตาราง 4-15 ความถูกต้องของการทดสอบข้อมูลสำหรับทำการขึ้นหรือการเดินจำนวน 11 ข้อมูล ทดสอบได้ความถูกต้อง 100 % , ทำการนั่งจำนวน 10 ข้อมูล ทดสอบได้ความถูกต้อง 100 % , ทำการก้มจำนวน 12 ข้อมูล วิเคราะห์เป็นทำการนั่ง 1 ข้อมูล ที่เหลือถูกต้อง ทดสอบได้ความถูกต้อง 91.7 % , ทำการนอนจำนวน 7 ข้อมูล ทดสอบได้ความถูกต้อง 100 %

จากสมมติฐานที่ว่าการใช้โครงสร้างมนุษย์แบบใหม่จะทำให้ระบบการรู้จำท่าทางมนุษย์มีความยืดหยุ่นในการใช้งานมากขึ้น แต่ยังสามารถรู้จำท่าทางต่างๆ ได้ถูกต้อง จึงต้องมีการพิสูจน์โดยการนำข้อมูลของฐานข้อมูลวิดีโอที่มีความสูงแตกต่างกัน คือ 3.3 เมตร และ 2.7 เมตร มาทำการทดสอบกับค่าน้ำหนักที่ได้จากการเทรนของฐานข้อมูลวิดีโอที่มีความสูง 3 เมตร โดยเมทแลปก็จะมีฟังก์ชันที่ใช้สำหรับการนำข้อมูลไปทดสอบกับค่าน้ำหนักเดิมอยู่แล้ว ซึ่งจะได้ผลลัพธ์ดังตารางที่ 4-16

ตารางที่ 4- 16 แสดงผลลัพธ์การรู้จำท่าทางมนุษย์ในความสูงที่แตกต่างกัน

ท่าทาง	ความสูงเดิม 3 เมตร	3.3 เมตร	2.7 เมตร
การขึ้นหรือการเดิน	99.0%	100%	97.1%
การนั่ง	100%	80.8%	93.1%
การก้ม	98.0%	78.4%	97.9%
การนอน	100%	99.0%	100%
	99.3%	89.5%	97.0%

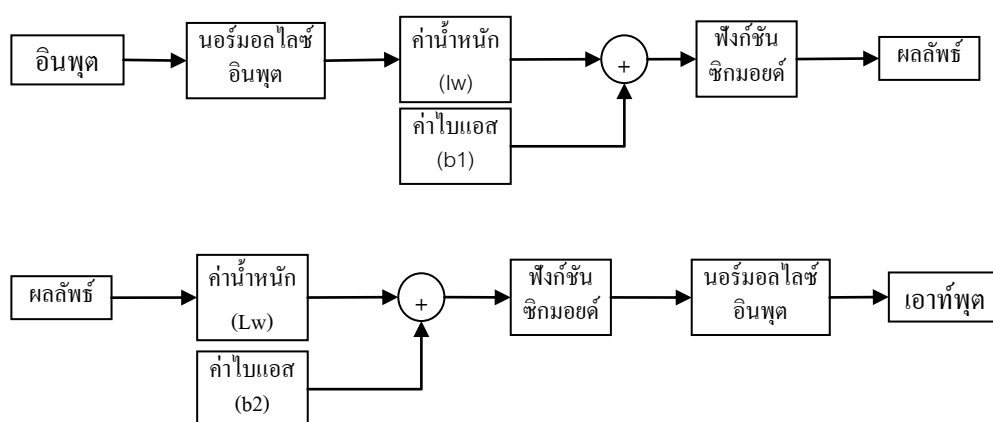
#### 4.3.4. วิเคราะห์และสรุปผลการทดลอง

จากตารางที่ 4-12 แสดงผลลัพธ์ที่ได้จากการทดลองการรู้จำและค่าเฉลี่ยความสูงจากภาพการรวมโครงสร้างมนุษย์ สามารถสรุปผลการทดลองได้ว่า ความถูกต้องในการรู้จำทำการขึ้น 99.0% , การนั่ง 100% , การก้ม 98.0% และการนอน 100% ค่าเฉลี่ยความถูกต้องคือ 99.3% และจากตารางที่ 4-7 แสดงผลลัพธ์การรู้จำท่าทางมนุษย์ในความสูงที่แตกต่างกัน จากการนำข้อมูลที่มีความสูง 3.3 และ 2.7 เมตร นำไปทดสอบกับน้ำหนักการเทรนของข้อมูลที่มีความสูง 3 เมตร ซึ่งความถูกต้องของข้อมูลที่มีความสูง 3.3 เมตรคือ 89.5% และความถูกต้องของข้อมูลที่มีความสูง 2.7 เมตรคือ 97.0% ซึ่งผลลัพธ์ทั้งสองมีความถูกต้องดีมาก แต่จะมีความผิดพลาดในท่าทางของการนั่งและการก้ม เนื่องจากอาจจะมีลักษณะ โครงสร้างมนุษย์บางกรณีที่คล้ายกันและมีค่าเฉลี่ยความสูงที่ใกล้เคียงกัน ดังนั้นจึงสามารถสรุปได้ว่าการเปลี่ยนแปลงของความสูงจากกล้องวิดีโอจะส่งผลกระทบต่อผลลัพธ์ที่ได้เล็กน้อย แต่การทำงานของระบบการรู้จำท่าทางยังคงมีประสิทธิภาพ

การทดลองที่ 3 จะแตกต่างกับการทดลองที่ 2 ในส่วนของการเพิ่มลักษณะเด่นในเรื่องของค่าเฉลี่ยความสูง เพื่อแก้ปัญหาความเหมือนกันในบางกรณีของลักษณะเด่นจากโครงสร้างมนุษย์ที่ทำให้กับการจำแนกท่าทางเกิดความผิดพลาด โดยจากการเพิ่มค่าเฉลี่ยเข้าไปทำให้ความถูกต้องของการวิเคราะห์ท่าทางมนุษย์ดีขึ้นจากเดิมที่ 91.5 % สำหรับจำนวนโหนด 36 โหนด กลายเป็น 99.3 โดยใช้จำนวนโหนดเพียง 7 โหนดเท่านั้น

#### 4.4. การจำแนกท่าทางมนุษย์แบบเรียลไทม์ (Realtime Action Recognition)

การจำแนกท่าทางมนุษย์แบบเรียลไทม์ เป็นโปรแกรมสำหรับการจำแนกท่าทางมนุษย์จากภาพของกล้องวิดีโอ โดยการนำค่าน้ำหนักของการเทรนที่ได้จากลักษณะเด่นของการใช้มุมและค่าเฉลี่ยความสูงจากภาพการรวมโครงสร้างมนุษย์ที่ความสูงของกล้องวิดีโอ 3 เมตร มาทำการคูณกับลักษณะเด่นของโครงสร้างมนุษย์เพื่อตัดสินใจว่าลักษณะเด่นที่ได้จะอยู่ในกลุ่มใด ซึ่งมี 4 กลุ่ม ได้แก่ 1.การยืนหรือการเดิน (0001) 2.การนั่ง (0010) 3.การก้ม (0100) 4.การนอน(1000) ซึ่งค่าน้ำหนักที่ได้จะมาจากเทคนิคข่ายงานประสาทของแมทแลป ดังนั้นการคำนวณเพื่อหาเอาต์พุตจึงต้องมีรูปแบบการคำนวณเหมือนกับที่แมทแลปใช้งาน โดยการคำนวณสามารถอธิบายได้ตามภาพประกอบที่ 4-1



ภาพประกอบ 4- 1 แผนผังการคำนวณของข่ายงานประสาท

ข่ายงานประสาทถูกกำหนดให้มีโหนดจำนวน 7 โหนด อินพุตประกอบด้วย 7 ข้อมูล โดยเริ่มต้นอินพุตจะถูกนอร์มอลไลซ์ เพื่อปรับสเกลให้อินพุตมีค่าไม่เกินช่วงที่กำหนด ซึ่งแต่ละอินพุตจะมีช่วงที่กำหนดแตกต่างกันดังนี้

$$\text{Max} = [-86.423666, -83.659808, -86.185925, -86.185925, -86.423666, -87.797402, 0.085366]$$

$$\text{Min} = [90, 90, 90, 90, 90, 90, 1]$$

แล้วจึงนำอินพุตที่ผ่านการนอร์มอลไลซ์มาทำการคูณกับค่าน้ำหนัก ( $Iw$ ) แล้วนำไปบวกกับค่าไบแอส ( $b1$ )

$$Iw_{7 \times 7} = \begin{bmatrix} 3.4051 & 0.6128 & 1.2979 & 0.9576 & -0.8075 & 0.7877 & -0.3638 \\ 2.7399 & 1.2689 & -0.5430 & 0.3911 & 0.2768 & 0.5842 & 0.8820 \\ 0.2134 & -0.2746 & -0.5624 & 0.5457 & -0.4883 & -0.7735 & -0.6778 \\ -0.1982 & -1.0931 & -0.2071 & 0.2136 & -2.6158 & -0.9145 & 5.2909 \\ 0.0463 & -0.9731 & -0.0035 & 0.1949 & -0.0839 & -0.7643 & 4.7005 \\ 0.1159 & -0.6164 & 0.4011 & 0.5815 & 0.2390 & -0.9062 & -3.8817 \\ 0.5261 & 0.2015 & 0.7563 & 0.3957 & 0.5402 & 1.4445 & 4.5972 \end{bmatrix}$$

$$b1_{7 \times 1} = \begin{bmatrix} 0.4352 \\ -1.6251 \\ -2.3912 \\ 0.7018 \\ -1.8908 \\ 2.4743 \\ 0.4950 \end{bmatrix}$$

จากนั้นนำค่าที่ได้ผ่านฟังก์ชันซิกมอยด์ ดังสมการที่ (4.1) โดยจะทำการเปรียบเทียบกับค่าเทรซโฮลด์ เพื่อคำนวณหาค่าที่ลู่อื่น 0 หรือ 1

$$S = \left( \frac{2}{1 + e^{(-2n)}} \right) - 1 \quad (4.1)$$

จากนั้นนำผลลัพธ์ที่ได้มาคูณกับค่าน้ำหนัก ( $Lw$ ) แล้วนำไปบวกค่าไบแอส ( $b2$ ) ก่อนจะนำไปเข้าสมการซิกมอยด์อีกครั้ง


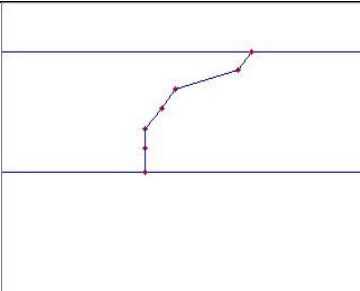

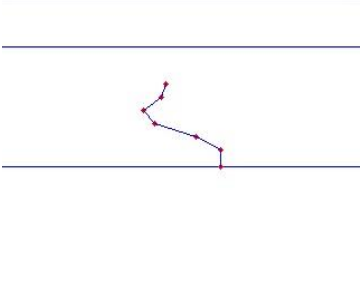
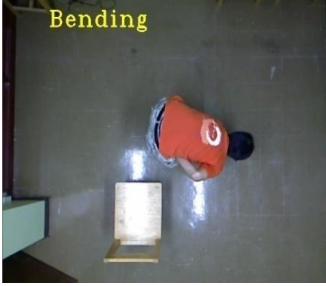
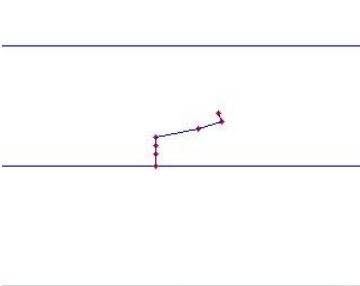

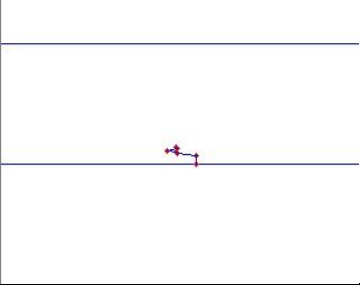
$$Lw_{4 \times 7} = \begin{bmatrix} 0.3733 & 0.2180 & 2.7418 & -1.5900 & -0.1946 & 1.0185 & -3.1416 \\ 1.3579 & -1.4462 & 1.8536 & -3.0415 & 0.2456 & -0.5160 & 3.7968 \\ -2.3482 & 1.7209 & 1.7469 & 4.1149 & -2.2962 & 2.4795 & 0.4105 \\ -0.4476 & 1.6294 & 2.0981 & 0.5866 & 2.3689 & -3.4532 & 0.4389 \end{bmatrix}$$

$$b2_{4 \times 1} = \begin{bmatrix} -0.3770 \\ -1.7770 \\ -1.2100 \\ -0.4028 \end{bmatrix}$$

สุดท้ายนำผลลัพธ์ไปนอร์มอลไลซ์อีกครั้ง เพื่อปรับเป็นค่าจำนวนเต็ม 0 หรือ 1 เพื่อจำแนกกลุ่มตามที่ได้กำหนดไว้

## โปรแกรมจำแนกท่าทางมนุษย์

การสร้างโปรแกรมจำแนกท่าทางมนุษย์ จะสร้างจาก Microsoft visual studio 2008 โดยมีการใช้ไลบรารี OpenCV 2.1 เพื่อประมวลผลภาพสำหรับการคำนวณหาลักษณะเด่นของโครงสร้างมนุษย์ การทำงานของโปรแกรมจะแสดงดังภาพประกอบ 4-2

กลุ่มท่าทาง	การรู้จำท่าทางมนุษย์	โครงสร้างมนุษย์	ลักษณะเด่น
กลุ่มการขึ้นหรือการเดิน			$F = [51.340192, 15.945396, 53.130102, 48.576334, 90.0, 90.0, 1.0]$
กลุ่มการนั่ง			$F = [70016894, 34.508523, -47.726311, -16.565051, 26.565051, 90.0, 0.684932]$
กลุ่มการก้ม			$F = [-66.801410, 15.945396, 10.437475, 90.0, 90.0, 90.0, 0.438356]$
กลุ่มการนอน			$F = [-45.0, 12.528808, -7.125016, -45.0, -6.709837, 90.0, 0.136986]$

ภาพประกอบ 4- 2 โปรแกรมการรู้จำท่าทางมนุษย์

จากภาพประกอบที่ 4-2 เป็นการรู้จำท่าทางมนุษย์ในท่าทางต่างๆ โดยโปรแกรมจะนำภาพมาประมวลผล เพื่อสร้างโครงสร้างมนุษย์แล้วทำการคำนวณลักษณะเด่นของโครงสร้างมนุษย์ จากนั้นตัดสินใจเพื่อจำแนกกลุ่มท่าทางของมนุษย์ โดยแสดงบนภาพในเวลานั้น

การทำงานของโปรแกรมรู้จำท่าทางมนุษย์สามารถทำงานได้อย่างรวดเร็ว โดยประมาณความเร็วได้ 22 เฟรมต่อวินาที

#### 4.5. สรุป

การทดลองการรู้จำท่าทางมนุษย์ จะใช้ข่ายงานประสาทสำหรับการแทนลักษณะเด่นของโครงสร้างมนุษย์ โดยลักษณะเด่นของโครงสร้างมนุษย์ที่จะนำมาใช้สำหรับการทดลองจะแบ่งเป็น 3 รูปแบบ คือ 1. ลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังดิสพาริที มีอินพุตจำนวน 18 ข้อมูลและใช้อัดเดนโหนดสำหรับการทดลองจำนวน 5 โหนด ซึ่งได้ความถูกต้อง 99.0% แต่การใช้ลักษณะเด่นรูปแบบนี้จะมีปัญหาในเรื่องการพิกัดลักษณะโครงสร้างมนุษย์และจำนวนของข้อมูลที่มาก ซึ่งจะส่งผลให้ระบบไม่มีความยืดหยุ่นในการใช้งานและเกิดซ้ำในการประมวลผล 2. ลักษณะเด่น จากมุมของภาพการรวมโครงสร้างมนุษย์ มีอินพุตจำนวน 6 ข้อมูลและใช้อัดเดนโหนดสำหรับการทดลองจำนวน 36 โหนด ซึ่งได้ความถูกต้อง 91.5% จะเห็นว่าความถูกต้องของการรู้จำท่าทางมนุษย์ลดลงจากการทดลองที่ 1 เนื่องจากลักษณะเด่นในแบบนี้จะมีบางกรณีที่มีความคล้ายกันทำให้การรู้จำท่าทางมีความผิดพลาด ความถูกต้องจึงลดลง แต่การนำไปใช้งานจะมีความยืดหยุ่นมากขึ้นเนื่องจากความสูงโครงสร้างมนุษย์จะเปลี่ยนแปลงตามความสูงของคนที่มา 3. ลักษณะเด่น จากมุมและค่าเฉลี่ยความสูงของภาพการรวมโครงสร้างมนุษย์ มีอินพุตจำนวน 7 ข้อมูลและใช้อัดเดนโหนดสำหรับการทดลองจำนวน 7 โหนด ซึ่งได้ความถูกต้อง 99.3% ลักษณะรูปแบบนี้จะต่างจากของการทดลองที่ 2 คือมีการเพิ่มความสูงของโครงสร้างมนุษย์เข้ามา เพื่อแก้ไขปัญหาค่าความคล้ายกันของโครงสร้าง เนื่องจากแต่ละท่าทางจะมีความสูงที่ต่างกัน

ก้าน้ำหนักจากการแทนของลักษณะเด่นในการทดลองที่ 3 ซึ่งใช้ข้อมูลจากกล้องวิดีโอที่ติดตั้งบนเพดานที่ความสูง 3 เมตร จะถูกนำไปทดสอบกับลักษณะเด่นแบบเดียวกันแต่ความสูงของการติดตั้งกล้องวิดีโอที่ต่างกัน คือ 2.7 เมตร และ 3.3 เมตร จากการทดสอบฐานข้อมูลที่มีความสูง 2.7 เมตร ได้ความถูกต้อง 97.0 % และฐานข้อมูลที่มีความสูง 3.3 เมตร ได้ความถูกต้อง 89.5% ซึ่งเป็นความถูกต้องที่น่าพอใจ เนื่องจากผลของการทดสอบจะต้องลดลงจากผลของการแทนอยู่แล้ว

สุดท้ายแล้วจะทำการสร้างโปรแกรมสำหรับการรู้จำท่าทางมนุษย์ โดยจะนำก้าน้ำหนักที่ได้จากการแทนเข้ามาประมวลผลกับลักษณะเด่นโครงสร้างมนุษย์ที่ได้ โดยการรู้จำท่าทางมนุษย์ของโปรแกรมที่สร้างขึ้นสามารถทำงานได้อย่างรวดเร็วมีประสิทธิภาพโดยวัดความเร็วได้ที่ 22 เฟรมต่อวินาที

## บทที่ 5

### สรุปผลการวิจัยและข้อเสนอแนะ

ปัจจุบันการรักษาความปลอดภัยหรือการเฝ้าระวังเหตุการณ์ผิดปกติต่างๆเป็นสิ่งสำคัญ การแก้ปัญหาโดยการให้คนทำหน้าที่เฝ้าระวังเพื่อรักษาความปลอดภัย ณ สถานที่ที่อาจจะเกิดเหตุการณ์ผิดปกติ หรือการติดตั้งกล้องวงจรปิดแล้วให้คนเป็นผู้เฝ้าดูเหตุการณ์ต่างๆที่ผิดปกติด้วยตัวเองผ่านทางจอมอนิเตอร์อาจจะได้ผลเล็กน้อย เนื่องจากการให้คนเฝ้าระวังเป็นเวลานานๆจะส่งผลทำให้ร่างกายมีการเหนื่อยล้า อ่อนเพลีย และนำไปสู่ความผิดพลาดในการทำงานได้ ดังนั้นการนำเอาระบบการประมวลผลภาพเข้ามาช่วยสำหรับวิเคราะห์เหตุการณ์ผิดปกติที่เกิดแบบอัตโนมัติจะช่วยให้ลดภาระของคน และทำให้ได้ผลลัพธ์ที่ดีกว่า

#### 5.1. สรุปผลการวิจัย

ในงานวิจัยนี้ได้วิจัยและพัฒนาวิธีการวิเคราะห์ท่าทางมนุษย์ เพื่อประยุกต์ใช้ในการเฝ้าระวังเหตุการณ์ผิดปกติต่างๆ เช่น การเฝ้าระวังผู้ป่วยที่อยู่ตามลำพัง เมื่อเกิดการหกล้มซึ่งอาจจะทำให้ได้รับบาดเจ็บ ผู้ดูแลสามารถรู้และให้การช่วยได้ทันเวลา การเฝ้าระวังความปลอดภัยในสถานที่สำคัญต่างๆ พฤติกรรมมนุษย์สามารถจำแนกได้หลายท่าทาง ซึ่งในระบบนี้จะเน้นไปที่การวิเคราะห์เฉพาะท่าทางพื้นฐานซึ่งนำไปสู่การทำท่าทางอื่นได้แก่ การยืน การเดิน การนั่ง การก้ม และการนอน ระบบนี้จะนำเสนอการรับภาพวิดีโอจากมุมสูง เพื่อจะช่วยลดความซับซ้อนของคนลงได้ และใช้คุณสมบัติของสี ความลึกและการเคลื่อนที่ของวัตถุเข้ามาช่วยสำหรับการวิเคราะห์ โดยกระบวนการในการวิเคราะห์ท่าทางมนุษย์จะประมวลผลตามขั้นตอน คือ ขั้นตอนแรกนำภาพความลึกเข้าสู่กระบวนการตรวจจับและติดตามการเคลื่อนที่ของวัตถุ โดยใช้เทคนิคการลบภาพพื้นหลังและการติดตามวัตถุเพื่อจะได้บริเวณกับตำแหน่งของวัตถุที่จะนำมาใช้ในการประมวลผลภาพต่อไป นำภาพสีเข้าสู่กระบวนการจับคู่สี โดยจะนำภาพสีมาแมทช์กับผลลัพธ์ของกระบวนการตรวจจับและติดตามการเคลื่อนที่ของวัตถุเพื่อจะได้วัตถุที่ถูกแมทช์กับสี ต่อไปเป็นกระบวนการวิเคราะห์โครงสร้างมนุษย์สำหรับใช้สร้างภาพเสมือนโครงสร้างมนุษย์ ซึ่งจะแบ่งเป็น 3 ขั้นตอนใหญ่ คือ 1. การฉายโครงสร้างมนุษย์ด้วยการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริติ โดยการนำผลลัพธ์ของกระบวนการตรวจจับและติดตามการเคลื่อนที่ของวัตถุมาผ่านกระบวนการประมาณระยะทางของวัตถุจากแผนผังคิสพาริติ ซึ่งจะได้ภาพลักษณะท่าทางของมนุษย์มา 2 แขน คือ U และ V แล้วจึงนำผลลัพธ์นี้เข้ากระบวนการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพาริติ โดยจะแมทช์กับผลลัพธ์ของกระบวนการจับคู่สี แล้วจะได้ภาพลักษณะท่าทาง

มนุษย์ที่ถูกระบุสี 2 ภาพจากแกน U และ V นำไปสร้างภาพโครงสร้างมนุษย์ โดยการประมาณจุดกึ่งกลางของลักษณะท่าทางมนุษย์ แล้วจึงเข้าสู่กระบวนการรวมโครงสร้างมนุษย์ เพื่อลดความซับซ้อนและความผิดพลาดที่เกิดจากภาพโครงสร้างมนุษย์ที่มี 2 แกน แล้วจึงนำมาหาลักษณะเด่นของโครงสร้างมนุษย์ โดยจะแบ่งออกเป็น 3 ลักษณะเด่นสำหรับใช้ทดสอบระบบ คือ 1. ลักษณะเด่นที่ได้จากการคำนวณมุมของโครงสร้างมนุษย์จากผลลัพธ์ของกระบวนการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพริที่จำนวน 18 ข้อมูล 2. ลักษณะเด่นที่ได้จากการคำนวณมุมของโครงสร้างมนุษย์จากผลลัพธ์ของการรวมโครงสร้างมนุษย์จำนวน 6 ข้อมูล 3. ลักษณะเด่นที่ได้จากการคำนวณและค่าเฉลี่ยความสูงจากโครงสร้างมนุษย์จากผลลัพธ์ของการรวมโครงสร้างมนุษย์จำนวน 7 ข้อมูล จากนั้นเข้าสู่กระบวนการรู้จำท่าทางมนุษย์ โดยใช้เทคนิคข่ายงานประสาทสำหรับจำแนกท่าทางมนุษย์ โดยจะแบ่งเป็น 4 กลุ่ม คือ การยืนหรือการเดิน การนั่ง การก้ม การนอน

## 5.2. อภิปรายผล

ลักษณะเด่นของโครงสร้างมนุษย์ที่จะนำมาใช้สำหรับการทดลองจะแบ่งเป็น 3 รูปแบบ คือ 1. ลักษณะเด่นจากมุมของโครงสร้างมนุษย์ในภาพการระบุสีให้กับการประมาณระยะทางของวัตถุจากแผนผังคิสพริที่มีอินพุตจำนวน 18 ข้อมูลและใช้อัดเดนโหนดสำหรับการทดลองจำนวน 5 โหนด ซึ่งได้ความถูกต้อง 99.0% แต่การใช้ลักษณะเด่นรูปแบบนี้จะมีปัญหาในเรื่องการฝึกลักษณะโครงสร้างมนุษย์และจำนวนของข้อมูลที่มาก ซึ่งจะส่งผลให้ระบบไม่มีความยืดหยุ่นในการใช้งานและเกิดช้าในการประมวลผล 2. ลักษณะเด่น จากมุมของภาพการรวมโครงสร้างมนุษย์มีอินพุตจำนวน 6 ข้อมูลและใช้อัดเดนโหนดสำหรับการทดลองจำนวน 36 โหนด ซึ่งได้ความถูกต้อง 91.5% จะเห็นว่าความถูกต้องของการรู้จำท่าทางมนุษย์ลดลงจากการทดลองที่ 1 เนื่องจากลักษณะเด่นในแบบนี้จะมีบางกรณีที่มีความคล้ายกันทำให้การรู้จำท่าทางมีความผิดพลาด ความถูกต้องจึงลดลง แต่การนำไปใช้งานจะมีความยืดหยุ่นมากขึ้นเนื่องจากความสูงโครงสร้างมนุษย์จะเปลี่ยนแปลงตามความสูงของคนที่เข้ามา 3. ลักษณะเด่น จากมุมและค่าเฉลี่ยความสูงของภาพการรวมโครงสร้างมนุษย์มีอินพุตจำนวน 7 ข้อมูลและใช้อัดเดนโหนดสำหรับการทดลองจำนวน 7 โหนด ซึ่งได้ความถูกต้อง 99.3% ลักษณะรูปแบบนี้จะต่างจากของการทดลองที่ 2 คือมีการเพิ่มความสูงของโครงสร้างมนุษย์เข้ามา เพื่อแก้ไขปัญหาค่าความคล้ายกันของโครงสร้าง เนื่องจากแต่ละท่าทางจะมีความสูงที่ต่างกัน

การนำค่าน้ำหนักจากการเทรนของลักษณะเด่นในการทดลองที่ 3 ซึ่งใช้ฐานข้อมูลจากกล้องวิดีโอที่ติดตั้งบนเพดานที่ความสูง 3 เมตร จะถูกนำไปทดสอบกับลักษณะเด่นแบบเดียวกันแต่ความ

สูงของการติดตั้งกล้องวิดีโอที่ต่างกัน คือ 2.7 เมตร และ 3.3 เมตร จากการทดสอบฐานข้อมูลที่สูง 2.7 เมตร ได้รับความถูกต้อง 97.0 % และฐานข้อมูลที่สูง 3.3 เมตร ได้รับความถูกต้อง 89.5% ซึ่งเป็นความถูกต้องที่น่าพอใจ เนื่องจากผลของการทดสอบจะต้องลดลงจากผลของการเรนเดอร์แล้ว

### 5.3. ข้อเสนอแนะ

งานวิจัยนี้สามารถนำไปประยุกต์ใช้ในระบบการเฝ้าระวังผู้ป่วยหรือผู้สูงอายุที่ต้องอาศัยอยู่เพียงลำพัง ซึ่งมีโอกาสที่อาจจะเกิดอุบัติเหตุการล้มในระหว่างการใช้ชีวิตประจำวันตามปกติ ซึ่งระบบจะช่วยให้ผู้ดูแลรับรู้ถึงเหตุการณ์ที่เกิดขึ้นและสามารถให้การช่วยเหลือได้อย่างทันที จากการทดสอบของกระบวนการจากวิทยานิพนธ์ชิ้นนี้จะให้ผลลัพธ์ของการวิเคราะห์ท่าทางมนุษย์ที่แม่นยำและรองรับความสูงของการติดตั้งกล้องที่แตกต่างกันอย่างดี แต่เนื่องจากการติดตั้งกล้องวิดีโอบนเพดานจะทำให้มุมมองของภาพที่ได้แคบลง ซึ่งจะส่งผลต่อการทำงานของระบบโดยรวมคือ ระบบจะทำงานได้ ณ บริเวณใดบริเวณหนึ่งเท่านั้น เพื่อแก้ปัญหาดังกล่าวจึงเป็นประเด็นที่สามารถนำไปพัฒนาต่อในอนาคต เพื่อเพิ่มบริเวณการทำงานของระบบให้มากขึ้น โดยอาจจะเพิ่มจำนวนกล้องในการประมวลผลเป็น 2 กล้อง แล้วทำให้กล้องทั้งสองสามารถทำงานร่วมกันได้ หรือการหาเลนส์ช่วงถ่ายที่กว้าง (lens wide) มาติดตั้งที่กล้อง ซึ่งทำให้ภาพที่ได้มีบริเวณที่กว้างขึ้น



### บรรณานุกรม

- [1] D.Weinland, R.Ronfard, E.Boyer, “A Survey of Vision-Based Methods for Action Representation, Segmentation and Recognition”, *inria-00459653*, version 1-24 Feb 2010.
- [2] M. Harville, “Stereo person tracking with short and long term plan-view appearance models of shape and color”, *In Proceeding IEEE Conference on Advanced Video and Signal Based Surveillance*, 2005, pp 522–527.
- [3] Weerachai S., Mizukawa M., “Human behavior recognition via top-view vision for intelligent space”, *In Proceeding 2010 International Conference On Control Automation And Systems(ICCAS)*, pp 1687 – 1690.
- [4] Y. Wang, T. Yu, L. Shi, and Z. Li, “Using human body gestures as inputs for gaming via depth analysis,” in *Multimedia and Expo, 2008 IEEE International Conference on*, 2008, pp. 993 –996.
- [5] Giovanni Bertolini, Stefano Ramat, “Identification and Recognition of Object in Color Stereo Image Using a HierachialSOM”, *In Proceeding of the 2007 IEEE Fourth Canadian Conference on Computer and Robot Vision(CRV’07)*.
- [6] M. Bertozzi, E. Binelli, A. Broggi, M. Del Rose, “Stereo Vision-based approaches for Pedestrian Detection”, *In Proceeding of the 2005 IEEE Computer Society Conference on CVPR’05*, PP 1063-6919/05.
- [7] Nattapon Noorit, Nikom Suvonvorn, and Montri Karnchanadecha. “Model-based Human Action Recognition”, *Proc. SPIE*, Vol. 7546, 75460P (2010).
- [8] M.-C. Roh, H.-K. Shin, and S.-W. Lee, “View-Independent Human Action Recognition Based on a Stereo Camera,” in *Chinese Conference on Pattern Recognition, 2009. CCPR 2009*, 2009, pp. 1 –5.
- [9] M. Piccardi, “Background subtraction techniques: a review,” in *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, 2004, vol. 4, pp. 3099 – 3104 vol.4.
- [10] I. Pavlidis, V. Morellas, P. Tsiamyrtzis and S. Harp, “Urban surveillance systems: from the laboratory to the commercial world,” in *Proceedings of the IEEE*, vol. 89, no. 10, 2001, pp. 1478 -1497.

- [11] A. Sappa, R. Herrero, F. Dornaika, D. Geronimo, and A. Lopez, "Road Approximation in Euclidean and  $v$ -Disparity Space: A Comparative Study," in *Computer Aided Systems Theory U EUROCAST 2007*, 2007, pp. 1105-1112.
- [12] A. Sappa, D. Geronimo, F. Dornaika, and A. Lopez, "On-board camera extrinsic parameter estimation," *Electronics Letters*, vol. 42, pp. 745-747, 2006.
- [13] R. Labayrade, D. Aubert, and J. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation", *Intelligent Vehicle Symposium*, 2002. IEEE, vol.2, pp. 646-651, 2002.

**ภาคผนวก**

**ภาคผนวก ก.****ผลงานตีพิมพ์เผยแพร่จากวิทยานิพนธ์**

1. S. Seawpakorn and N. Suvonvorn, "Top-view based human action recognition using depth and color information," in *Proceedings of the IASTED International Conference on Advances in Computer Science and Engineering, ACSE 2012*, 2012, pp. 376–380.



## Top-view Based Human Action Recognition Using Depth and Color Information

Sittisuk Seawpakorn<sup>1</sup>, Nikom Suwornvorn<sup>2</sup>

Department of Computer Engineering, Faculty of Engineering, Prince of Songkla University  
Hatyai, Songkhla, Thailand, 90112

E-mail: [saint.memory@gmail.com](mailto:saint.memory@gmail.com)<sup>1</sup>, [kom@coe.psu.ac.th](mailto:kom@coe.psu.ac.th)<sup>2</sup>

### ABSTRACT

Human action recognition is an important step for human behavior analysis, which can apply to many applications, such as, surveillance system, and medical analysis tools. In this paper, we propose an alternative technique for human action recognition using color and depth image from top-view camera. The specific representation, called the Color UV-disparity, is proposed to solve the universal unsolved problem. The four basic actions are concerned: standing or walking, sitting, bending and laying. The neural network is applied as classification method. Our method can recognize actions accurately in real-time.

### KEY WORDS

Action recognition, Color UV-disparity, RGB-D image

### 1. Introduction

Action recognition is necessary knowledge for human behavior understanding that can apply in many applications, such as, surveillance system for public safety, assistance systems for disable person, and patient monitoring in hospital. Actually, many researches are done in this field [1], amount of methods are proposed and analysed based on many properties, such as, shape, color, motion, model, and etc. However, the systems generally failed when the behavior become complex, due to many factors, such as, angle of camera, illumination change, and etc. In our research, we believe that any complex behaviors can be decomposed into sequence of small atomic actions: standing, walking, sitting, blending, and laying. If the actions are well identified, the complex behavior can be better modeled.

In this paper, we propose an alternative technique for human action recognition using color and depth image from top-view camera. The four basic actions are concerned: standing/walking, sitting, bending and laying. The important idea of our method is that depth information from the top-view camera provides the vital information to the action classification. For example, ideally, the two levels of depth (head, shoulder) could be observed for standing/walking actions, and in contrast only one small depth level with large color area should identify the laying action. Based on the idea, we proposed a specific representation, called the Color UV-disparity, as an intermediate transformed image space in order to recognize the four basic actions. Then neural network is applied in classification process.

### 2. Related work

The previous research works on human action recognition [1] can be grouped into two main categories: body modeling and image modeling. In body modeling, the methods try to recover the human body structure from the images and identified actions. Bertozzi et al.[2] proposed a system for pedestrian detection combining three different methods: warm area detection, edge based detection and v-disparity computation. W. Wang et al.[3] propose an articulated model of human using the body's trunk and limbs contour angles. The angles representing the pose of the skeleton model and the length-width ratio of the human are used as feature vector. The performance of method depends on the accuracy of coordinates of human body articulation. In contrast for image modeling, actions are recognized directly from specific image features. Giovanni Bertolini et al.[4] introduced feature clustering method extracted from HSV color space and depth, using a hierarchical self-organizing map(HSOM). Liang Wang et al.[5] proposed combines kernel principal component analysis (KPCA) based feature extraction and using factorial conditional random field (FCRF) for motion modeling.

Some works are nearly related to our method. Weerachai S[6] proposed the human action recognition which focuses on a top-view single camera. Human posture features are described based on DFT coefficients of log-polar histogram. Neural network is used for classification of 4 actions: standing, sitting, lying and moving. Michael Harville[7] proposes the method for recognizing human action from plan-view template. Shape and color feature is used for describing the long-term model. Nittapon Noorit et al.[8] introduce the method for action classification using simple human model with color features. Five basic actions can be classified using neural network.

### 3. Proposed method

This section we introduce our propose system for human action recognition from the top view, shown in figure 1. Image sequence is acquired from Kinect camera, as color and depth images. Firstly, depth images are analyzed for detecting moving object using background subtraction technique. Then, the detected object is projected into the UV-disparity representation. At the same time, the segmentation of ROI corresponding to concerning object in the color image is performed via calibrated geometric transformation between cameras.

The process will introduce the color properties into the UV-disparity. Finally, the feature representing actions, such as, standing/walking, sitting, bending, and laying, are extracted from our representation, and then input to the neural network for further action classification.

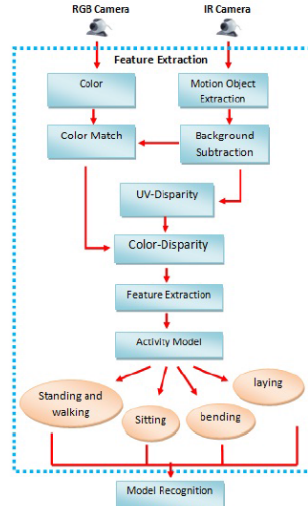


Figure 1. System overview

### 3.1 Object tracking from depth image

In our system, the motion detection is used to segment moving object in depth image using background subtraction technique. We found that performing motion detection on depth image is more reliable than on color image. Note that changing of illumination has strong impact to intensity of color image but not to the depth information. In our system, we apply the mixture of Gaussian method for background modeling. The background is modeled by calculating the  $P(X_t)$  probability as the following.

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (1)$$

Where  $\omega_{i,t}$  is weighted Gaussian,  $K$  is number of Gaussian (fixed to 5), and  $\eta$  is probability of Gaussian function defined by.

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)} \quad (2)$$

The result of background subtraction is shown in figure 2. We can notice that motion object can be well segmented in most actions. However the laying action gets less accurate than others due to its depth information that does not enough distinguish to the floor.

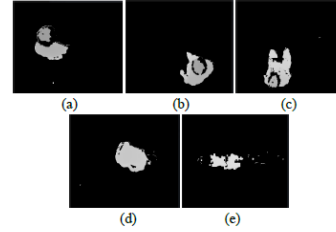


Figure 2. the result of motion detection: (a) standing, (b) walking, (c) sitting, (d) bending, (e) laying.

The blob tracking is used to track motion object in depth image. Minimum bounding box is determined around moving object for globally tracking the direction of moving object.

### 3.2 UV-disparity

UV-disparity is a representation which is obtained from the projected depth image over x-axis, called “u-disparity”, and y-axis, called “v-disparity”. This representation provides an intermediate way of disparity analysis of moving object on the x and y spaces separately. V-disparity [10] is estimated from disparity map on y-axis by the following.

$$H(I_\Delta) = I_{\Delta v} \quad (3)$$

Where  $I_\Delta$  is disparity map,  $I_{\Delta v}$  is v-disparity, and  $H$  is a horizontal accumulate operator of similar intensity on each row of disparity map. Figure 3 shows results of v-disparity of moving object in five actions.

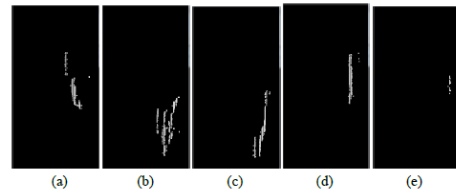


Figure 3. the V-disparity: (a) standing, (b) walking, (c) sitting, (d) bending, (e) laying.

In the same manner, u-disparity is estimated from disparity map on x-axis defined by following equation.

$$V(I_\Delta) = I_{\Delta u} \quad (4)$$

Where  $I_{\Delta u}$  is u-disparity, and  $V$  is a vertical accumulate operator of similar intensity on each column of disparity map. Figure 4 shows results of u-disparity for moving object of five actions. From the results, we can notice that the uv-disparity can discriminate the structure of human in difference actions by its depth information from the top-view. However, it cannot clearly define physically the parts of body which gives more details on the specific actions, for example, head, body, or leg.

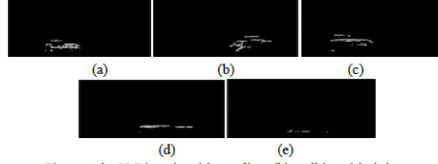


Figure 4. the U-Disparity: (a) standing, (b) walking, (c) sitting, (d) bending, (e) laying

### 3.3 Color UV-disparity

The color uv-disparity is proposed for identifying parts of human structure in the uv-disparity. Colors can represent the parts of body depending on hair or cloth colors. This step can be done by segmenting the moving object in color image corresponding to object found in the depth image. The camera calibration allows us to transform depth pixels to color pixels geometrically. Firstly, the depth pixels will be transformed to 3D points  $P3D.x$ ,  $P3D.y$ ,  $P3D.z$  using the following equations.

$$P3D.x = (x_d - cx_d) * \left( \frac{depth(x_d, y_d)}{fx_d} \right) \quad (5)$$

$$P3D.y = (y_d - cy_d) * \left( \frac{depth(x_d, y_d)}{fy_d} \right) \quad (6)$$

$$P3D.z = depth(x_d, y_d) \quad (7)$$

where  $fx_d, fy_d, cx_d, cy_d$  are intrinsic of IR camera

$$fx_d = 5.9421434211923247e+02$$

$$fy_d = 5.9104053696870778e+02$$

$$cx_d = 3.3930780975300314e+02$$

$$cy_d = 2.4273913761751615e+02$$

The 3D points are then converted to 2D points in color image.

$$\begin{bmatrix} P3D'.x \\ P3D'.y \\ P3D'.z \end{bmatrix} = R \cdot \begin{bmatrix} P3D.x \\ P3D.y \\ P3D.z \end{bmatrix} + T \quad (8)$$

$$P2D_{rgb}.x = \frac{P3D'.x + fx_{rgb}}{P3D'.z} + cx_{rgb} \quad (9)$$

$$P2D_{rgb}.y = \frac{P3D'.y + fy_{rgb}}{P3D'.z} + cy_{rgb} \quad (10)$$

where R and T are the rotation and translation matrix respectively.

$$R = \begin{bmatrix} 9.9984628826577793e-01 & 1.2635359098409581e-03 & -1.7487233004436643e-02 \\ -1.4779096108364480e-03 & 9.9992385683542895e-01 & -1.2251380107679535e-02 \\ 1.7470421412464927e-02 & 1.2275341476520762e-02 & 9.9977202419716948e-01 \end{bmatrix}$$

$$T = \begin{bmatrix} 1.9985242312092553e-02 \\ -7.4423738761617583e-04 \\ -1.0916736334336222e-02 \end{bmatrix}$$

$$fx_{rgb} = 5.2921508098293293e+02$$

$$fy_{rgb} = 5.2556393630057437e+02$$

$$cx_{rgb} = 3.2894272028759258e+02$$

$$cy_{rgb} = 2.6748068171871557e+02$$

Finally, the transformed depth pixels, figure 5(b), and color pixels, figure 5(a), is matched with a simple *and* operator, which produces result shown in figure 5(c).



Figure 5. (a) color image (b) depth image (c) color object

By using the same definition of uv-disparity in 3.2, but apply to color moving object; we obtain finally the color uv-disparity. Figures 6 and 7 show results of moving objects in different actions.

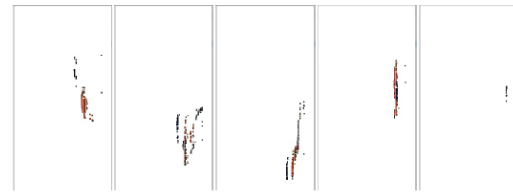


Figure 6. the color V-disparity: (a) standing, (b) walking, (c) sitting, (d) bending, (e) laying.

We can notice that the color provides us a superior quality of actions representation via human structure in the uv-disparity. So, the head and body parts can be clearly identified, although the leg can be located hardly.

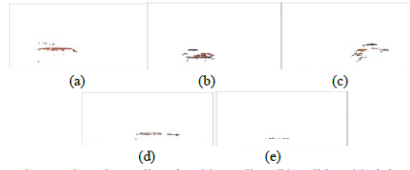


Figure 7. the color U-disparity: (a) standing, (b) walking, (c) sitting, (d) bending, (e) laying.

### 3.4 Feature extraction

Our features are estimated from the color UV-disparity. Firstly, the disparity levels in the U and V disparity are quantized into 10 levels with respect to color (human structures). In each level of U or V disparity, the center of mass are computed, which represents the center of mass of body at the specific disparity level in the x and y axes respectively. Then, the features were defined by angles between the two successive centers of mass in the adjacent disparity levels. The angles are calculated with reference to horizontal axe for v-disparity, and reference to vertical axe for u-disparity, shown in figure 8.



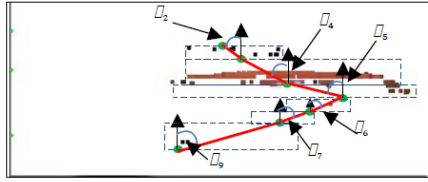


Figure 8. Angle feature definition.

Ideally or physically, the angle becomes zero when standing or walking, or positive (but not close to 90 degrees) when bending (this angle reflex relatively to bending's angle), or near to 90 degrees when laying action. By following this procedure, we got 18 angles from u (9 angles) and v (9 angles) disparity. Figure 9 and 10 shows the example of angle features.

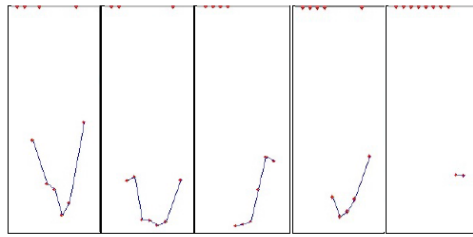


Figure 9 the angle features of color V-disparity: (a) standing, (b) walking, (c) sitting, (d) bending, (e) laying.

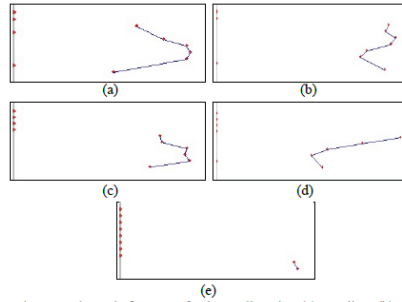


Figure 10 the angle features of color U-disparity: (a) standing, (b) walking, (c) sitting, (d) bending, (e) laying.

Then, the feature vector is defined as the 18 angles for 1 frame. Note that for one action, the feature is defined as 18x100 of 2D matrix.

### 3.5 Action classification

In our recognition system for human action classification, we apply the feed forward neurons network with multi-layer perception network using back-propagation algorithm and sigmoid activate function. We fix 18 nodes for input layer, corresponding to our feature angles, and 4 nodes for four basic actions:

standing/walking, sitting, bending, and laying. The number of hidden nodes is predefined to 5,10,15,20,25,30,35 and 40 nodes, the optimum number is determined during the experimentation.

## 4. Experimental results

The human action recognition is acquired from Kinect camera installed on the top of a room. Color and depth image is record at the same time via RGB camera and IR camera, shown in figures 11 and 12. The datasets is done using our adaptive version of CL NUI Platform. Four persons with different clothes act randomly and continuously with respect to testing actions: 100 standing, 100 sitting, 100 bending, and 100 laying actions, shown in figure 10 and 11. Testing system is run on Intel processor Dual Core, 2.4 GHz, 1 GB memory. The human action was divided into 4 groups: standing/walking(0001), sitting(0010), bending (0100), and laying (1000). The features used to recognize in neurons network is consisted 18 angles of the color UV-disparity. For each action, the training set is 400 images. This dataset is split into 3 groups: training 80%, validation 10% and testing 10%.

In the first experimentation, we would like to determine the optimal number of hidden layer nodes for this classification problem. We define 8 neural networks with different hidden nodes, such as, 5,10,15,20,25,30,35, and 40. The table 1 shows correction rate with respect to the number of hidden nodes.

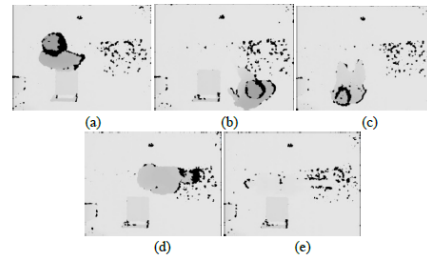


Figure 11.(a) standing, (b) walking, (c) sitting, (d) bending, (e) laying.

Nodes	Percentage of the correction rate in average
5	99.0
10	99.0
15	99.3
20	99.5
25	99.5
30	99.5
35	99.8
40	99.8

Table1: the correction rate with random nodes

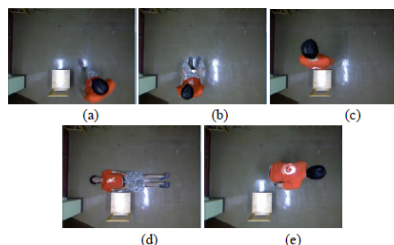


Figure 12. (a) standing, (b) walking, (c) sitting, (d) bending, (e) laying.

We can note that the correction rate increase when the number increase. However, with more than 10 nodes the increasing rate is very small (up to 0.8% for 38 additional nodes). Then, we conclude that the neural network with only 5 nodes is enough for solving the problem. The table 2 shows the recognition rate of four actions with 5 hidden nodes.

Actions	Our method				Weerachai [6]
	Training	Validation	Test	Final	
Standing and Walking	100%	100%	100%	100%	84.0%
Sitting	98.8%	100%	100%	99.0%	84.0%
Bending	98.8%	90.0%	90.0%	97.1%	-
Laying	100%	100%	100%	100%	86.0%
				99.0%	83.3%

Table 2. Actions recognition rate.

We found that the standing/walking and laying actions can be recognized perfectly due to the distinguished disparity levels. However, the system provides a little bit less powerful on bending action at 97.1%. Globally, our system produces a good recognition rate in average at 99.0%

We compare the performance of our technique with the method of Weerachai S.[6]. As shown in table 2, we can notice that in very action our method can perform better results. This can be explained intuitively that depth information is representative and valuable for action classification from the top-view of human structure. The classification using only color image with primitive properties, such as, height-width ratio, will limit totally the classification of ambiguity actions (bending/sitting actions). However, analyzing image from top-view has an important limitation on the field of view. More cooperative cameras should improve the better results.

## 5. Conclusion

In this paper, we present a novel method of human action recognition for top-view images using color and depth information. The color UV-disparity is specially proposed for describing the moving actions. Features extracted from the color uv-disparity are feed to neural

network. The system provides a good recognition rate for at 99.0%.

## Acknowledgement

This work was supported by the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission

## References

- [1] D.Weinland, R.Ronfard, E.Boyer, "A Survey of Vision-Based Methods for Action Representation, Segmentation and Recognition", *inria-00459653*, version 1-24 Feb 2010.
- [2] M. Bertozzi, E. Binelli, A. Broggi, M. Del Rose, "Stereo Vision-based approaches for Pedestrian Detection", *In Proceeding of the 2005 IEEE Computer Society Conference on CVPR'05*, PP 1063-6919/05.
- [3] Weihua Wang, Zhijing Liu, "Real-time human behavior recognition based on articulated model", *In Proceeding 2010 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, Vol 4, pp 254 – 258.
- [4] Giovanni Bertolini, Stefano Ramat, "Identification and Recognition of Object in Color Stereo Image Using a HierachialSOM", *In Proceeding of the 2007 IEEE Fourth Canadian Conference on Computer and Robot Vision(CRV'07)*.
- [5] Liang Wang, Suter D., "Recognizing Human Activities from Silhouettes: Motion Subspace and Factorial Discriminative Graphical Model", *In Proceeding 2007 IEEE Conference on Computer Vision and Pattern Recognition CVPR '07*, pp 1-8.
- [6] Weerachai S., Mizukawa M., "Human behavior recognition via top-view vision for intelligent space", *In Proceeding 2010 International Conference On Control Automation And Systems(ICCAS)*, pp 1687 – 1690.
- [7] M. Harville, "Stereo person tracking with short and long term plan-view appearance models of shape and color", *In Proceeding IEEE Conference on Advanced Video and Signal Based Surveillance*, 2005, pp 522–527.
- [8] Nattapon Noorit, Nikom Suvonvorn, and Montri Kamchanadecha. "Model-based Human Action Recognition", *Proc. SPIE*, Vol. 7546, 75460P (2010).
- [9] Piccardi M., "Background subtraction techniques: A review", *Conference Proceedings – IEEE International Conference on Systems, Man and Cybernetics*, 2004, 4, pp. 3099-3104.
- [10] R. Labayrade, D. Aubert, and J. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation", *Intelligent Vehicle Symposium*, 2002. IEEE, vol.2, pp. 646-651.

## ประวัติผู้เขียน

ชื่อ สกุล	นายสิทธิศักดิ์ เชี่ยวภากร		
รหัสประจำตัวนักศึกษา	5310120080		
วุฒิการศึกษา			
	วุฒิ	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
	วิศวกรรมศาสตรบัณฑิต (วิศวกรรมคอมพิวเตอร์)	มหาวิทยาลัยสงขลานครินทร์	2552

### ทุนการศึกษา(ที่ได้รับระหว่างการศึกษา)

ทุนบัณฑิตศึกษาภายใต้โครงการมหาวิทยาลัยแห่งชาติมหาวิทยาลัยสงขลานครินทร์

### การตีพิมพ์เผยแพร่ผลงาน

1. S. Seawpakorn and N. Suvonvorn, "Top-view based human action recognition using depth and color information," in *Proceedings of the IASTED International Conference on Advances in Computer Science and Engineering, ACSE 2012*, 2012, pp. 376–380.