การวิเคราะห์ภาพวิดีโอโดยใช้ความลึกสำหรับการจดจำท่าทางของมนุษย์

Side-View Based Human Action Recognition Using Stereo Vision

พงศธร ชวลิตสิทธิกุล

Pongsatorn Chawalitsittikul

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

ชื่อวิทยานิพนธ์ การวิเคราะห์ภาพวิดีโอโดยใช้ความลึกสำหรับการจดจำท่าทางของมนุษย์
ผู้เขียน นายพงศธร ชวลิตสิทธิกุล
สาขาวิชา วิศวกรรมคอมพิวเตอร์

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

……..
(ผู้ช่วยศาสตราจารย์ดร.นิคม สุวรรณวร)

คณะกรรมการสอบ

……..ประธานกรรมการ
(ดร.อนันท์ ชกสุริวงค์)

……..กรรมการ
(ผู้ช่วยศาสตราจารย์ดร.นิคม สุวรรณวร)

……..ใ..กรรมการ
(ผู้ช่วยศาสตราจารย์ดร.ปรัญชลีย์ สมานพิบูรณ์)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้นับวิทยานิพนธ์ฉบับนี้สำหรับการศึกษา ตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์

……..…….

(รองศาสตราจารย์ ดร.ธีระพล ศรีชนะ)
คณบดีบัณฑิตวิทยาลัย

ขอรับรองว่า ผลงานวิจัยนี้มาจากการศึกษาวิจัยของนักศึกษาเอง และได้แสดงความขอบคุณบุคคลที่มีส่วนช่วยเหลือแล้ว

ลงชื่อ……..
(ผู้ช่วยศาสตราจารย์ดร.นิคม สุวรรณวร)
อาจารย์ที่ปรึกษาวิทยานิพนธ์

ลงชื่อ……..
(นายพงศธร ชวลิตสิทธิกุล)
นักศึกษา

ข้าพเจ้าขอรับรองว่า ผลงานวิจัยนี้ ไม่เคยเป็นส่วนหนึ่งในการอนุมัติปริญญาในระดับใดมาก่อน และไม่ได้ถูกใช้ในการยื่นขออนุมัติปริญญาในขณะนี้

ลงชื่อ……..

(นายพงศธร ชวลิตสิทธิกุล)

นักศึกษา

ชื่อวิทยานิพนธ์ การวิเคราะห์ภาพวิดีโอโดยใช้ความลึกสำหรับการจดจำท่าทางของมนุษย์
ผู้เขียน นายพงศธร ชวลิตสิทธิกุล
สาขาวิชา วิศวกรรมคอมพิวเตอร์
ปีการศึกษา 2555

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้เป็นงานวิจัยเพื่อประยุกต์ใช้เทคนิคด้านการประมวลผลภาพวิเคราะห์โมเดลมนุษย์สำหรับการรู้จำท่าทางของมนุษย์ซึ่งมุ่งเน้นไปที่การเฝ้าระวังเหตุการณ์ที่ผิดปกติในสถานที่สำคัญต่างๆและในผู้ป่วย ผู้พิการ คนชราที่อาศัยอยู่เพียงลำพัง เมื่อเกิดเหตุการณ์ผิดปกติสามารถให้การช่วยเหลือได้อย่างรวดเร็ว ซึ่งประกอบด้วยกระบวนการสร้างโครงสร้างมนุษย์ขึ้นใหม่ให้อยู่ในรูปแบบจำลองโครงสร้างมนุษย์อย่างง่ายที่มีเพียง 3 องค์ประกอบสำคัญของร่างกาย คือ ศีรษะ ลำตัวและขาจากนั้นจึงติดตามการเคลื่อนไหวของแต่ละองค์ประกอบอย่างเป็นอิสระต่อกัน ระบบจะมีการนำข้อมูลความลึก สี และการเคลื่อนไหวของวัตถุเข้ามาช่วยในการวิเคราะห์ภาพพื้นฐาน ผลของการติดตามจะได้ข้อมูลของความสัมพันธ์ของมุม และความแตกต่างของความลึก ของลำตัวถึงศีรษะ และลำตัวถึงขา โดยข้อมูลเหล่านี้จะถูกนำไปใช้ในกระบวนการรู้จำท่าทางพื้นฐานทั้งหมด 5 ท่าทางได้แก่ การยืน การเดิน การนั่ง การก้ม และการนอน โดยใช้กระบวนการรู้จำของนิวรอนเน็ตเวิร์ค และSVM(Support Vector Machine) สำหรับการเทรนลักษณะเด่นและทดสอบความถูกต้องของการรู้จำท่าทาง โดยใช้ตัวอย่างสำหรับการทดลอง 800 ข้อมูล ท่าทางละ 200 ข้อมูลซึ่งผลการรู้จำท่าทางพื้นฐานให้ความถูกต้องโดยเฉลี่ยมากที่สุด 96.5% และจากการนำท่าทางพื้นฐานเหล่านี้มาผสมผสานกันจะสามารถใช้อธิบายถึงกิจกรรมมนุษย์ที่มีความซับซ้อนมากขึ้นได้โดยที่ระบบสามารถทำงานแบบเรียลไทม์ได้อย่างมีประสิทธิภาพ และมีความคงทนต่อมุมที่เปลี่ยนไประดับหนึ่งโดยไม่ทำให้การวิเคราะห์คลาดเคลื่อนมากนัก

Thesis Title Side-View based Human Action Recognition using Stereo Vision

Author Mr. Pongsatorn Chawalitsittikul

Major Program Computer Engineering

Academic Year 2012

ABSTRACT

This thesis is to research practical techniques by using image processing, analysis, modeling human action recognition, which focuses on monitoring the unusual landmark, and in patients with impaired seniors. Abnormal events can be of assistance quickly. The process consists of creating a new human to human in a simple model with only three key elements of the body is the head, torso and legs and then track the movement of each element independently of one another. It has a depth of color and movement of objects to help in the image analysis. The results of the investigation of the relationship and the difference in depth of the body, head, and legs. This information will be used in the recognition stance underlying all five postures such as standing, walking, sitting, bending and lying to the recognition of the neuron network and SVM (Support Vector Machine) for training. Features and testing the accuracy of human action recognition for samples were 800 data signal 200 for which the recognition, gesture based for accuracy by the 96.5% and the attitude underlying these combinations can be used to describe human activities are complex. The system can run in real-time effectively and robust to the change to a certain extent without causing much error analysis.

กิตติกรรมประกาศ

สำหรับการดำเนินการวิจัยและจัดทำวิทยานิพนธ์นี้ ผู้วิจัยขอขอบพระคุณผู้ช่วยศาสตราจารย์ ดร.นิคม สุวรรณวร ประธานกรรมการที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้คำปรึกษา ชี้แนะแนวทางในการทำงาน ทั้งยังให้กำลังใจและการเอาใจใส่กับข้าพเจ้าเป็นอย่างมาก กระตุ้นให้ข้าพเจ้าได้มีความมุ่งมั่นในการทำงานให้สำเร็จลุล่วงไปได้เป็นอย่างดีเรื่อยมา รวมถึงการตรวจและแก้ไขเนื้อหาวิทยานิพนธ์ให้สำเร็จสมบูรณ์

ขอขอบพระคุณรองศาสตราจารย์ ดร.อนันท์ ชกสุริวงค์ กรรมการที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้คำแนะนำในการปรับปรุงวิทยานิพนธ์ให้สมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณคณาจารย์ และบุคคลากรทุกท่านในภาควิชาวิศวกรรมคอมพิวเตอร์ทุกท่านที่ให้ความช่วยเหลือในระหว่างการทำวิทยานิพนธ์

ขอขอบพระคุณเพื่อน ๆ และนักศึกษาปริญญาโททุกท่านที่คอยให้คำแนะนำและคอยให้ความช่วยเหลือด้วยดีตลอดมา

สุดท้ายนี้ ขอกราบขอบพระคุณ บิดา มารดาและญาติพี่น้องทุกคน ซึ่งเป็นผู้มีพระคุณสูงสุดที่ให้กำลังในและให้การสนับสนุนทุกสิ่งทุกอย่างด้วยดีตลอดมาในชีวิตของข้าพเจ้า

พงศธร ชวลิตสิทธิกุล

สารบัญ

สารบัญรูปภาพ

สารบัญตาราง

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของโครงงาน

ปัจจุบันการเกิดเหตุการณ์ที่ไม่พึงประสงค์สามารถเกิดได้ตลอดเวลา ไม่ว่าจะเป็นเหตุการณ์ที่เกี่ยวกับความปลอดภัยในชีวิต และทรัพย์สิน จากข่าวตามสื่อต่างๆมากมาย เช่น การปล้น การลอบสังหาร การลอบวางระเบิด เป็นต้น หรือเหตุการณ์ที่เกี่ยวกับการเฝ้าระวังมนุษย์เพื่อวิเคราะห์พฤติกรรมความเป็นอยู่ เช่น การดูแลผู้สูงอายุ การสังเกตการณ์ทำงานของพนักงานในโรงงาน เป็นต้น เหตุการณ์ที่กล่าวมาถือเป็นเรื่องที่ควรให้ความสำคัญเป็นอย่างมาก เนื่องจากมีผลกระทบต่อการดำรงชีวิตของมนุษย์

อย่างไรก็ตามเรื่องของการเฝ้าระวังพฤติกรรมมนุษย์ในปัจจุบันยังไม่สามารถทำได้ดีพออันเนื่องมาจากระบบรักษาความปลอดภัยที่มีอยู่ไม่สามารถสังเกตได้ตลอดเวลา มาจากหลายสาเหตุ เช่น บุคคลที่นำมาเฝ้าเกิดการบกพร่องในหน้าที่ไม่ว่าจะเป็นกรณีใดๆก็ตาม หรือแม้กระทั่งเป็นผู้กระทำผิดเองเป็นต้น เรื่องของการเฝ้าระวังมนุษย์เพื่อวิเคราะห์พฤติกรรมความเป็นอยู่ก็เช่นกันซึ่งถ้าใช้มนุษย์ในการสังเกตก็ไม่สามารถสังเกตได้ตลอดเวลา และอาจจะดูว่าเป็นการรุกล้ำสิทธิ ซึ่งจะทำให้ผู้ที่ถูกสังเกตการณ์เกิดความอึดอัดได้

ดังนั้นในปัจจุบันจึงมีการติดตั้งกล้องวงจรปิดอย่างแพร่หลายในองค์กรไม่ว่าจะเป็นองค์กรขนาดใหญ่หรือขนาดเล็ก ห้างสรรสินค้า ร้านค้า รวมไปถึงบ้านเรือน ก็ยังมีการติดตั้งกล้องวงจรปิดเพื่อใช้สำหรับการสังเกตการณ์ และใช้ดูวิดีโอย้อนหลังเมื่อเกิดเหตุการณ์ที่สนใจขึ้น อย่างไรก็ตามถึงแม้จะมีการติดตั้งกล้องวงจรปิดแล้วก็ไม่สามารถระงับเหตุการณ์ที่เกิดขึ้นได้ทันท่วงที เช่น มีเหตุการณ์คนร้ายวางระเบิดในห้างสรรพสินค้ากว่าที่ระบบรักษาความปลอดภัยของห้างจะรู้ว่ามีการวางระเบิดก็อาจจะสายไประเบิดอาจจะทำงาน กล้องวงจรปิดก็ทำได้เพียงบันทึกข้อมูล เพื่อใช้สำหรับการค้นหาตัวผู้กระทำผิดภายหลัง ซึ่งมันก็สายไปสำหรับความสูญเสียที่เกิดขึ้นต่อชีวิตและทรัพย์สินรวมไปถึงสภาพจิตใจของผู้คนที่อยู่ในเหตุการณ์ซึ่งไม่อาจจะประเมินค่าได้ หรือเรื่องของการดูแลผู้สูงอายุถ้าหากผู้สูงอายุเกิดการล้มกล้องวงจรปิดก็ไม่สามารถบอกได้ว่าเกิดการล้ม ณ เวลานั้นได้ทันท่วงที ดังนั้นเรื่องของการสังเกตพฤติกรรมมนุษย์จึงมีความสำคัญอย่างมากถ้าหากต้องการจะแก้ปัญหาเหตุการณ์เหล่านี้

งานวิจัยชิ้นนี้จึงเน้นเรื่องของการเฝ้าระวังมนุษย์ในเรื่องของการสังเกตท่าทางอันนำไปสู่การวิเคราะห์พฤติกรรมตลอด 24 ชั่วโมงโดยใช้การประมวลผลภาพเข้ามาช่วยสำหรับการวิเคราะห์ เพื่อใช้สำหรับการแจ้งเตือนได้อย่างทันท่วงที [7] โดยใช้การตรวจจับจากวัตถุที่เคลื่อนไหวในเฟรม

วิดีโอ และสีเสื้อผ้าของผู้ที่สวมใส่เพื่อได้มาซึ่งคุณลักษณะของมนุษย์โดยอ้างอิงจากโมเดลอันนำไปสู่การวิเคราะห์ท่าทางต่าง เช่น เดิน นั่ง ยืน นอน ก้ม เป็นต้น โดยการตรวจจับ (Detection) และติดตาม (Tracking) มีความสำคัญต่อความถูกต้องของการวิเคราะห์ท่าทางมนุษย์ และเพิ่มประสิทธิภาพสำหรับความผิดพลาดที่เกิดจากมุมกล้องจึงเป็นที่มาของการใช้กล้องมากกว่าหนึ่งตัวสำหรับการวิเคราะห์ [8] เพื่อให้มุมมองที่แตกต่างในกรณีของกล้องตัวเดียวถ้าหันหน้าเข้าหากล้อง เช่น การก้ม จะไม่สามารถวิเคราะห์ได้ว่าบุคคลคนนั้นก้ม สำหรับมุมมองของภาพในกรณีที่ได้จากกล้องหลายตัวจะทำให้สามารถได้มุมมองที่หลากหลายมากขึ้นถ้าหากก้มหน้าใส่กล้องตัวใดตัวหนึ่ง กล้องตัวอื่นที่เหลือก็สามารถวิเคราะห์ได้ ทำให้สามารถวิเคราะห์ท่าทางได้มีประสิทธิภาพมากกว่ากล้องตัวเดียว

1.2 วัตถุประสงค์ของโครงงาน

1.2.1 เพื่อวิจัยการวิเคราะห์ท่าทางของมนุษย์จากโมเดลโดยใช้ข้อมูลการเคลื่อนไหว และข้อมูลความลึก

1.2.2 เพื่อพัฒนาเทคนิคการสร้างโมเดลมนุษย์ในรูปแบบสามมิติ (3D-Model) จากหลายมุมมอง (Multi View)

1.2.3 เพื่อพัฒนาเทคนิคการรู้จำ(Recognition) ท่าทางของมนุษย์

1.3 ประโยชน์ที่คาดว่าจะได้รับ

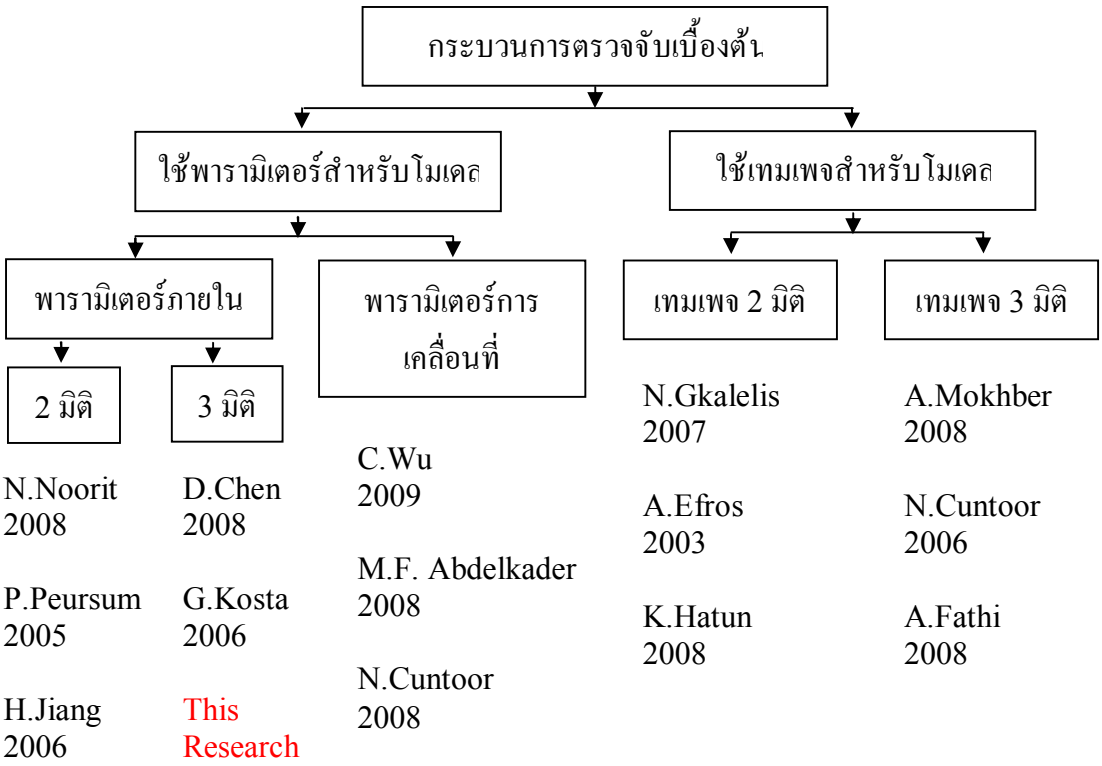
1.3.1 สามารถเฝ้าระวัง และวิเคราะห์รูปแบบการทำกิจกรรมของมนุษย์โดยที่มนุษย์สามารถดำเนินชีวิตได้อย่างอิสระ

1.3.2 สามารถใช้ประโยชน์สูงสุดจากกล้องของระบบรักษาความปลอดภัย

1.3.3 ได้ความรู้สำหรับการประยุกต์ใช้กล้องสำหรับการประมวลผลภาพ (Image Processing) ที่ได้จากกล้องหลายมุมมอง(Multi View)

1.4 งานวิจัยที่เกี่ยวข้อง

การวิเคราะห์ท่าทางมนุษย์โดยใช้การประมวลผลภาพจากเฟรมวิดีโอถือเป็นเรื่องที่น่าสนใจและได้ประโยชน์อย่างมาก เนื่องจากการใช้ชีวิตในยามที่ไม่สามารถช่วยเหลือตัวเองได้เต็มประสิทธิภาพทำให้ต้องมีคนมาคอยดูแล รวมไปถึงการวิเคราะห์ท่าทางสำหรับสังเกตพฤติกรรมที่น่าสงสัย จึงทำให้มีงานวิจัยที่เกี่ยวกับการรู้จำท่าทางมนุษย์จากการใช้การประมวลผลภาพเกิดขึ้นมากมาย ดังภาพประกอบที่ 1-1



ภาพประกอบ 1-1 แผนภาพแสดงประเภทของงานวิจัยที่เกี่ยวข้องกับมนุษย์

จากงานวิจัยที่เกี่ยวข้องกับเรื่องของการวิเคราะห์ท่าทางมนุษย์สามารถแบ่งกลุ่มของงานวิจัยออกเป็น 2 กลุ่มหลักๆ คือ กลุ่มที่ใช้คุณลักษณะของมนุษย์จากของจริง (Parametric Object Model) และกลุ่มที่ใช้คุณลักษณะมนุษย์จากเทมเพจ(Implicit Object Model) [12]

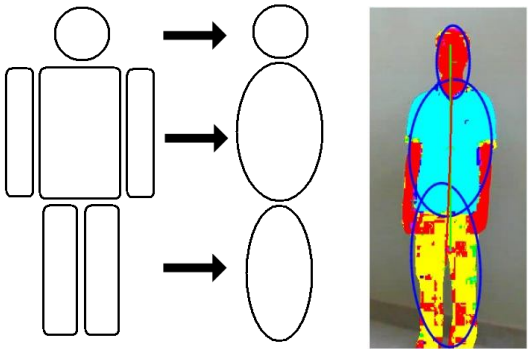
กลุ่มแรกสามารถแบ่งได้อีกเป็นสองกลุ่มคือกลุ่มที่ใช้คุณลักษณะทางโครงสร้างภายในร่างกาย (Internal Model) ซึ่งมีเป็นแบบ 2D และ 3D โดยงานวิจัยชิ้นนี้จัดอยู่ในกลุ่มของInternal Model ที่เป็นลักษณะของโมเดลสามมิติ (3D) และอีกกลุ่มคือกลุ่มที่เป็นลักษณะของการดูการเคลื่อนโดยเทียบกับบริเวณแล้วนำไปวิเคราะห์(Trajectory Model)

กลุ่มที่สองสามารถแบ่งได้อีกเป็นสองกลุ่มเช่นกันคือกลุ่มที่มีลักษณะเป็นเทมเพจสองมิติ (2D Template) และเทมเพจสามมิติ (3D Template) ซึ่งไม่เกี่ยวข้องกับงานวิจัยนี้มากนัก

1.4.1 การวิเคราะห์หาโมเดลมนุษย์ในรูปแบบสองมิติ(2D Human Modeling)

Nattapon Noorit, Nikom Suvonvorn, และ Montri Karnchanadecha [7] ได้เสนอเทคนิคการค้นหากระบวนการสร้างโครงสร้างมนุษย์ขึ้นใหม่ให้อยู่ในรูปแบบจำลองโครงสร้างมนุษย์อย่างง่ายที่มีเพียง 3 องค์ประกอบสำคัญของร่างกาย คือ ศีรษะ, ลำตัวและขา ดังภาพประกอบที่ 1-2a โดยใช้การตรวจจับการเคลื่อนไหว (Motion Detection) สำหรับตรวจจับการเคลื่อนไหวที่คาดว่าจะเป็น

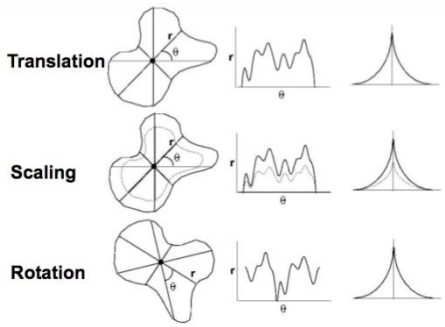
มนุษย์ จากนั้นใช้ฮิสโทแกรมของสีจากบริเวณทั้ง3 ส่วน สำหรับการติดตาม (Tracking) โดยการตั้งสมมุติฐานว่าลำตัวจะมีบริเวณขนาดของสีใหญ่ที่สุดเมื่อเทียบกับ ศีรษะ และขา จึงเริ่มจากการหาบริเวณที่ใหญ่ที่สุด 2 บริเวณ จากนั้นหาบริเวณที่คาดว่าจะเป็นลำตัวสำหรับเป็นจุดอ้างอิง โดยที่ศีรษะจะตั้งอยู่ในตำแหน่ง45° ถึง 135° ในด้านบน และ ขา จะอยู่ในตำแหน่ง -45° ถึง -135° ในด้านล่าง ดังภาพประกอบที่ 1-2b ถ้าหากผลการหาความสัมพันธ์ตรงตามเงื่อนไข พื้นที่บริเวณ ศีรษะ ลำตัวและขาจะถูกทำเครื่องหมายไว้เพื่อใช้ในกระบวนการติดตามต่อไป โดยงานวิจัยชิ้นนี้จัดอยู่ในกลุ่มของ Parametric Object Model -> Internal Model -> 2D



(a) (b)

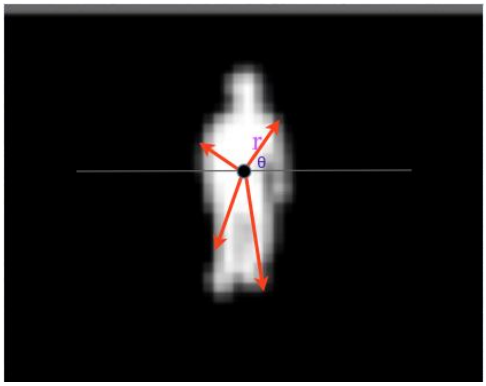
ภาพประกอบ 1-2 แบบจำลองมนุษย์อย่างง่าย

N. Gkalelis, N. Nikolaidis, และ I. Pitas [8] ได้เสนอเทคนิคการตรวจจับ(Detection) และติดตาม(Tracking) มนุษย์โดยใช้Discrete Fourier Transform (DFT) เนื่องจากคุณสมบัติที่ได้จากการแปลง DFT สามารถที่ระบุว่าวัตถุนั้นเป็นวัตถุเดียวกันโดยไม่คำนึงเรื่องของตำแหน่งของมุมไม่ว่าวัตถุนั้นจะหมุนไปเท่าไหร่เมื่อทำการแปลงDFT ก็จะได้รูปกราฟเหมือนกันเสมอและถ้าหากวัตถุที่มีรูปร่างเหมือนกันแต่ขนาดต่างกันรูปแบบของกราฟก็ยังคงเหมือนเดิมต่างกับเพียงความสูงของกราฟโดยวัตถุที่ใหญ่กราฟจะมีขนาดสูงกว่าวัตถุที่เล็กกว่าดังภาพประกอบที่ 1-3



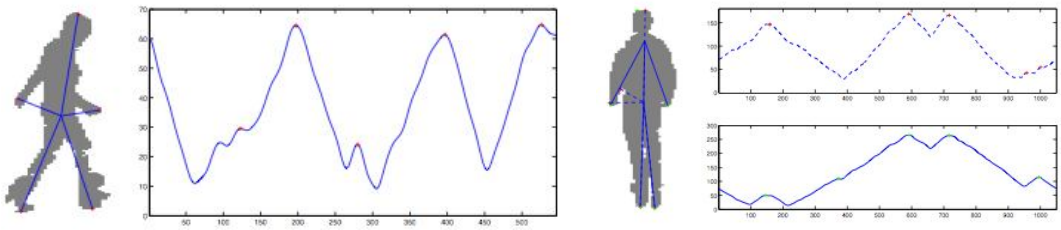
ภาพประกอบ 1-3 คุณสมบัติของ Discrete Fourier Transform (DFT)

ในงานวิจัยของ N. Gkalelis, N. Nikolaidis, และ I. Pitas จะถือว่าทุกอย่างที่เคลื่อนไหวมีโอกาสเป็นไปได้ทั้งหมดที่จะเป็นมนุษย์ โดยจะนำมาหาว่าสิ่งที่เคลื่อนไหวใดตรงกันกับโมเดลมนุษย์ที่อยู่ในรูปของแบบของที่ DFT ที่มีอยู่ในฐานข้อมูล โดยการแปลงทำได้จากการนำพิกเซลรอบๆวัตถุที่เคลื่อนไหวมาหารัศมี และมุมโดยเทียบกับจุดอ้างอิงตรงกลางวัตถุจะได้กราฟที่อยู่ในรูปของรัศมีเทียบกับมุม ดังภาพประกอบที่ 1-4 จากนั้นนำไปเทียบว่าใกล้วัตถุใดคือมนุษย์โดยงานวิจัยชิ้นนี้จัดอยู่ในกลุ่มของImplicit Object Model -> 2D Template แต่ก็มีบางงานวิจัยจะสมมุติให้วัตถุที่เคลื่อนไหวในเฟรมวิดีโอทั้งหมดที่ได้จากการตรวจจับการเคลื่อนไหว(Motion Detection) เป็นมนุษย์ [5][6] เช่น งานวิจัยของ C. Canton-Ferrer และ M. Ahmad and Seong-Whan Lee



ภาพประกอบ 1-4 การแปลง Discrete Fourier Transform (DFT)

Daniel Chen, Pi-chi Chou, Clinton Fookes และ Sridha Sridharan [9] ได้เสนอเทคนิคการตรวจจับ (Detection) และติดตาม (Tracking) มนุษย์จากสมมุติฐานที่ว่ามนุษย์จากมุมสองมิติจะมีลักษณะเป็นเหมือนรูปดาวที่มีลักษณะเป็น5 แฉก โดยเริ่มจากการหาระยะห่างจากเส้นรอบรูปโดยเทียบกับจุดศูนย์กลางของวัตถุที่เคลื่อนไหวมาทำการพอร์ตกราฟ หลังจากนั้นทำการเลือกจุดที่มีระยะห่างจากจุดศูนย์กลางมากที่สุด5 จุด จากจุดที่มีความชันเท่ากับศูนย์ เพื่อนำไปวิเคราะห์ความเป็นไปได้ที่จะเป็นมนุษย์ ถ้าวัตถุที่ตรวจพบเป็นมนุษย์จะพบว่าระยะของจุดที่เป็นศีรษะ และขาทั้งสองข้างจะมีระยะใกล้เคียงกัน แต่จะยาวกว่ามือทั้งสองข้างที่มีระยะใกล้เคียงกัน ดังภาพประกอบที่ 1-5 (a) เนื่องจากพบว่าการที่กำหนดจุดอ้างอิงอยู่บริเวณตรงกลางของวัตถุจะสามารถตรวจจับได้ดีเมื่อบุคคลนั้นหันด้านข้างแต่จะมีข้อผิดพลาดถ้าบุคคลนั้นหันหน้าเข้าหากล้องจึงได้มีวิธีแก้ไขโดยการย้ายจุดอ้างอิงไปอยู่บริเวณส่วนบนของลำตัวทำให้ประสิทธิภาพสำหรับการตรวจจับดีขึ้น ดังภาพประกอบที่ 1-5 (b)

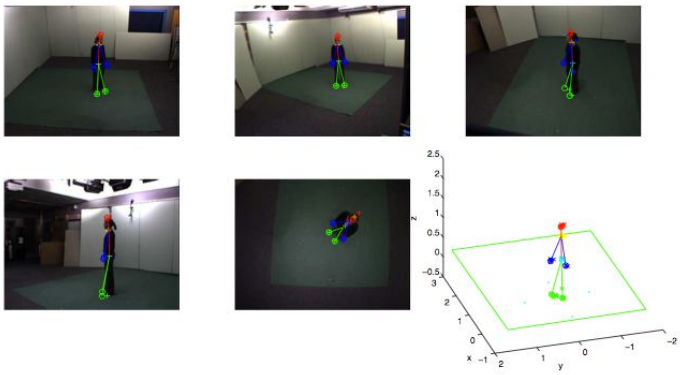


(a) (b)

ภาพประกอบ 1-5 ตัวอย่าง Star Skeleton

1.4.2 การวิเคราะห์หาโมเดลมนุษย์ในรูปแบบสามมิติ (3D Human Modeling)

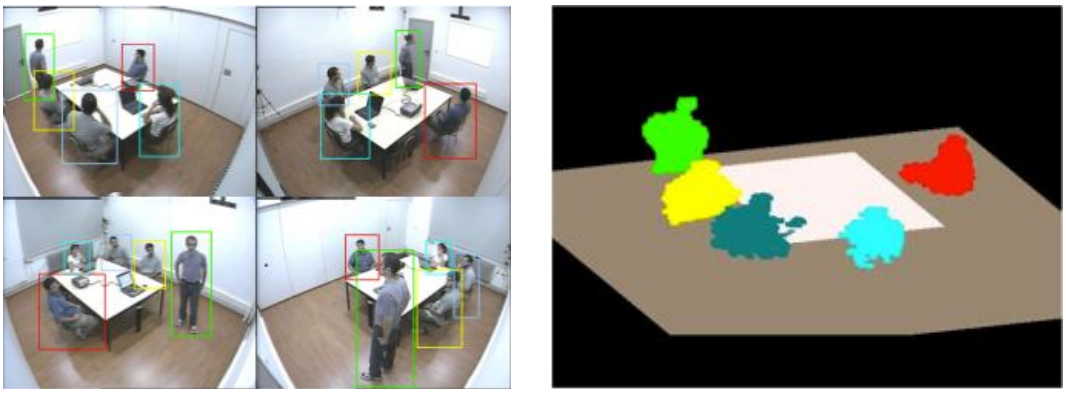
Daniel Chen, Pi-chi Chou, Clinton Fookes และ Sridha Sridharan [9] ได้เสนอเทคนิคการสร้างโมเดลมนุษย์สามมิติจากภาพสองมิติ (3D Reconstruction) โดยการใช้พารามิเตอร์จากจุดที่ได้จากอัลกอริทึม Star Skeleton ที่มาจากหลายมุมมองซึ่งจะได้จุดทั้งหมด5 จุดต่อหนึ่งกล้อง คือ ศีรษะ มือ และขาทั้งสองข้าง โดยจะทำการแยกสีของจุดที่คาดว่าจะเป็นจุดเดียวกัน เช่น มือที่มีระยะห่างน้อยกว่าศีรษะและขา เป็นสีน้ำเงิน ขาที่ยาวกว่ามือและมีระยะใกล้เคียงกันสองจุให้เป็นสีเขียว และศีรษะคือจุดที่มีระยะไม่ใกล้เคียงกับทั้งสี่จุดให้เป็นสีแดงจากนั้นนำค่าที่คาดว่าเป็นจุดเดียวกันจากมุมมองที่ต่างกันโดยกล้องทั้งหมดผ่านการทำให้เป็นมาตรฐานสำหรับการหาจุดที่เป็นตำแหน่งเดียวกันในแต่ละมุมมอง เพื่อนำมาทำการคำนวณจุดในสามมิติ ดังภาพประกอบที่ 1-6 โดยความแม่นยำจะแปรผันตรงกับจำนวนของมุมมองโดยงานวิจัยชิ้นนี้ จัดอยู่ในกลุ่มของ Parametric Object Model -> Internal Model -> 3D



ภาพประกอบ 1-6 ผลลัพธ์จากการทำโมเดลมนุษย์ที่เป็นสามมิติ

C. Canton-Ferrer [5] ได้เสนอเทคนิคการสร้างโมเดลมนุษย์สามมิติ จากการนำข้อมูลจากหลายมุมมอง (Data Fusion) เป็นหลักโดยเรื่องของโครงสร้างมนุษย์ให้ความสนใจรองลงมา ดังภาพประกอบที่ 1-7 (a) ที่ผ่านการคาลิเบตสำหรับการหาจุดที่เป็นตำแหน่งเดียวกันในมุมมองที่

ต่างกัน โดยนำเฉพาะส่วนที่มีการเคลื่อนไหวที่สำคัญ(Foreground) ซึ่งได้จากการใช้อัลกอริทึมของ Stauffer-Grimson [10] เป็นการตรวจจับการเคลื่อนไหวโดยการนำภาพมาลบกับพื้นหลัง (Background Subtraction) เพื่อนำมาสร้างภาพใหม่ที่เป็นสามมิติเพื่อระบุตำแหน่งของบุคคลภายในห้อง ดังภาพประกอบที่ 1-7 (b)

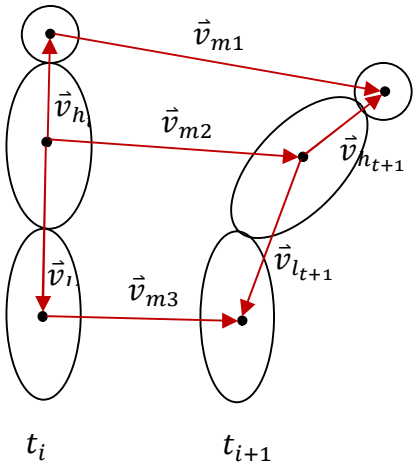


(a) (b)

ภาพประกอบ 1-7 ผลลัพธ์จากการนำภาพจากหลายมุมมองมาทำเป็นภาพสามมิติเพื่อระบุตำแหน่ง

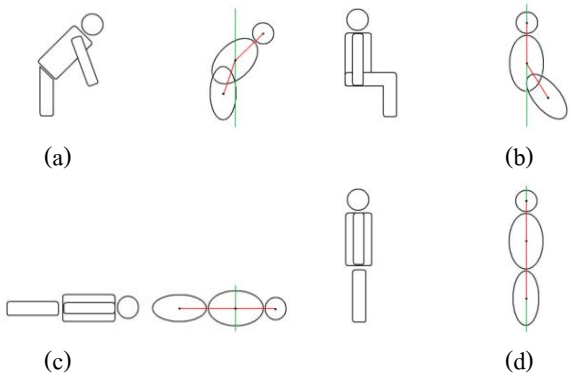
1.4.3 การวิเคราะห์ท่าทาง (Action Analysis)

Nattapon Noorit, Nikom Suvonvorn, และ Montri Karnchanadecha [7] วิเคราะห์ท่าทางโดยใช้พารามิเตอร์ที่ได้จากโมเดลมนุษย์จากสามจุดคือ ศีรษะ ลำตัว และขา ในขณะที่ยังไม่มีการเคลื่อนไหว ณ เวลา t_i จะมีพารามิเตอร์ที่เกิดขึ้นสองตัวคือ $\vec{V}_h$ และ $\vec{V}_l$ โดยในขณะที่มีการเคลื่อนไหว ณ เวลา $t_i + 1$ จะมีพารามิเตอร์เกิดขึ้นเพิ่มอีกสามตัวคือ$\vec{V}_{m1}$ $\vec{V}_{m2}$ และ $\vec{V}_{m3}$ ดังภาพประกอบที่ 1-8



ภาพประกอบ 1-8 ต้นแบบมนุษย์อย่างง่ายและพารามิเตอร์ทั้งหมด ณ ขณะเวลา t_i และ t_{i+1}

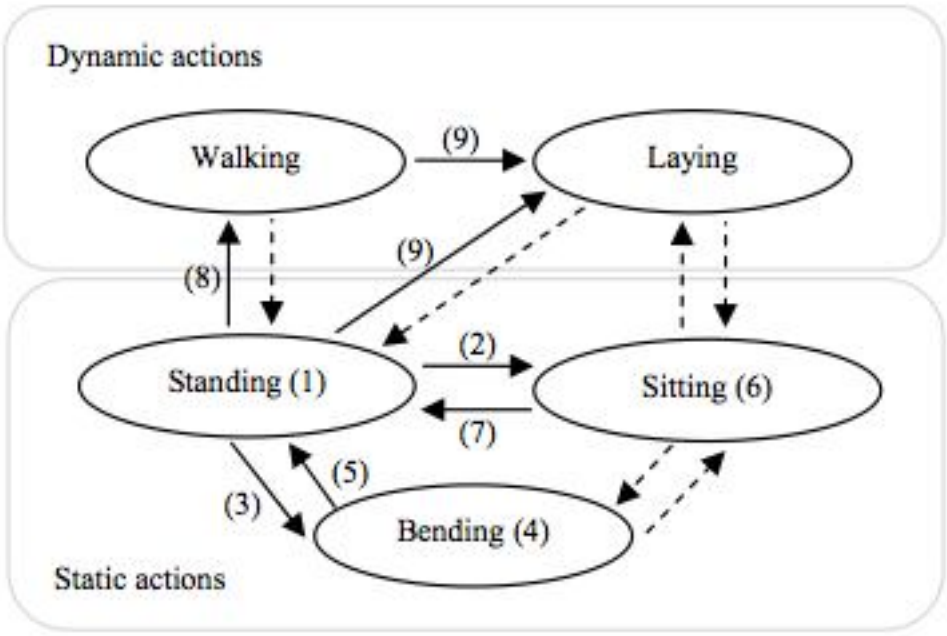
สำหรับการวิเคราะห์ท่าทางพื้นฐานมีทั้งหมด5 ท่าทาง ได้แก่ การยืน การก้ม การนั่ง การเดิน และการนอนราบ ดังภาพประกอบที่ 1-9 โดยแบ่งท่าทางออกเป็น 2 ลักษณะคือ ท่าทางสถิต (Static Action) ซึ่งจะจัดอยู่ในกลุ่มแบบจำลองสถิต และท่าทางพลวัต (Dynamic Action) ซึ่งจะจัดอยู่ในกลุ่มแบบจำลองพลวัตเช่น $\frac{\partial \vec{v}^{s}_{m1}}{\partial t} = \frac{\partial \vec{v}^{s}_{m2}}{\partial t} = \frac{\partial \vec{v}^{s}_{m3}}{\partial t} = 0$ เป็นสมการที่แสดงว่าไม่มีการเคลื่อนที่ของบุคคลแปลความหมายได้ว่าบุคคลนั้นยืนอยู่โดยไม่มีการเคลื่อนที่



ภาพประกอบ 1-9 ท่าทางพื้นฐาน (a) ก้ม (b) นั่ง (c) นอนราบ (d) ยืน หรือ เดิน

สำหรับการวิเคราะห์ท่าทางจะมีลำดับของการเปลี่ยนแปลงตามความเป็นไปได้ที่สามารถเกิดขึ้นได้ โดยแสดงต้นแบบการแสดงท่าทาง ลูกศรเส้นทึบ และหมายเลขกำกับแสดงถึงคุณลักษณะที่ได้กำหนดไว้ตามลำดับสมการข้างต้น ลูกศรเส้นประแสดงการเปลี่ยนสถานะที่ไม่ได้กำหนดคุณลักษณะในงานวิจัยนี้ ดังภาพประกอบที่ 1-10 ซึ่งเงื่อนไขการเปลี่ยนสถานะเป็นไปตามสมการดังต่อไปนี้

(1) $(\vec{v}^{\theta}_{h} = 0° \mp 30°, \vec{v}^{\theta}_{l} = 180° \mp 30°, \frac{\partial \vec{v}^{s}_{m1}}{\partial t} = \frac{\partial \vec{v}^{s}_{m2}}{\partial t} = \frac{\partial \vec{v}^{s}_{m3}}{\partial t} = 0 \mp 3)$
(2) $(\frac{\partial \vec{v}^{\theta}_{l}}{\partial t} < 0)$, **(3)** $(\frac{\partial \vec{v}^{\theta}_{h}}{\partial t} > 0)$, **(4)** $(\vec{v}^{\theta}_{l} = 180° \mp 30°, \frac{\partial \vec{v}^{s}_{m2}}{\partial t} = \frac{\partial \vec{v}^{s}_{m3}}{\partial t} = 0 \mp 3)$,
(5) $(\frac{\partial \vec{v}^{\theta}_{h}}{\partial t} < 0)$, **(6)** $(\vec{v}^{\theta}_{h} = 0° \mp 30°, \frac{\partial \vec{v}^{s}_{m3}}{\partial t} = 0 \mp 3)$, **(7)** $(\frac{\partial \vec{v}^{\theta}_{l}}{\partial t} > 0)$,
(8) $(\frac{\partial \vec{v}^{s}_{m1}}{\partial t} \cong \frac{\partial \vec{v}^{s}_{m2}}{\partial t} \cong \frac{\partial \vec{v}^{s}_{m3}}{\partial t} > 3)$, **(9)** $(\frac{\partial^2 \vec{v}^{\theta}_{m1}}{\partial t^2} > \frac{\partial^2 \vec{v}^{\theta}_{m2}}{\partial t^2} \geq \frac{\partial^2 \vec{v}^{\theta}_{m3}}{\partial t^2} > 3)$

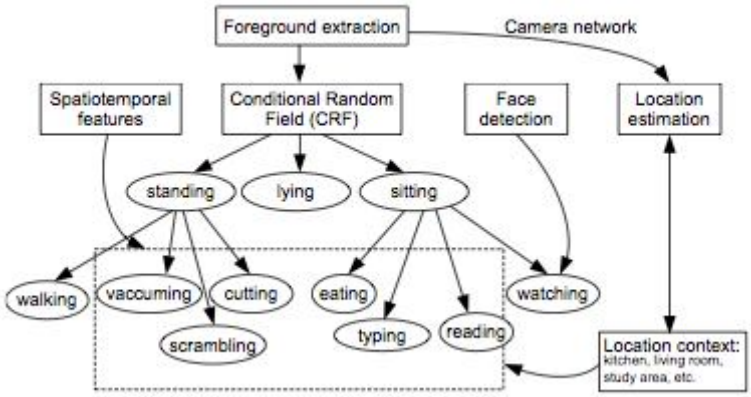


ภาพประกอบ 1-10 แผนภาพแสดงการสถานะจำกัดการเคลื่อนไหว

C. Wu, A. H Khalili, และ H. Aghajan [11] วิเคราะห์ท่าทางโดยการใช้ข้อมูลจากกล้องมากกว่าหนึ่งตัวซึ่งมีมุมมองแตกต่างกันสำหรับมองไปยังสภาพแวดล้อมต่างๆ ดังภาพประกอบที่ 1-11 วิเคราะห์จากการแยกข้อมูลของกล้องแต่ละตัวเพื่อหาเหตุการณ์โดยนำข้อมูลจากกล้องต่างๆมารวมกันสำหรับการวิเคราะห์ท่าทาง

ภาพประกอบ 1-11 ตำแหน่งการติดตั้งกล้อง และตัวอย่างภาพจากสองมุมมอง

สำหรับการวิเคราะห์ท่าทางมนุษย์โดยใช้การตรวจจับการเคลื่อนไหว (Motion Detection) โดยนำเฉพาะส่วนสำคัญที่มีการเคลื่อนไหว (Foreground) จากนั้นทำการวิเคราะห์สถานที่ที่เกิดการเคลื่อนไหวโดยอ้างอิงตามตำแหน่งของกล้องแต่ละตัว เช่น ถ้ากล้องตัวที่5 สามารถตรวจจับการเคลื่อนไหวได้แสดงว่ามนุษย์อยู่ในห้องนั่งเล่น เป็นต้น



ภาพประกอบ 1-12 แผนภาพแสดงการสถานะจำกัดการเคลื่อนไหว

การวิเคราะห์ท่าทางจะใช้ข้อมูลจากหลายกล้องมาหาความสัมพันธ์ที่น่าจะเป็นไปได้สำหรับเหตุการณ์ที่มีโอกาสเกิดขึ้น ดังภาพประกอบที่ 1-12 ถ้าหากบุคคลกำลังนั่ง และอยู่ในห้องเรียน มีโอกาสที่จะ กินข้าว, พิมพ์ดีด และอ่านหนังสือ ดังตารางที่ 1-1 แต่ถ้าไม่อยู่ในห้องเรียนในขณะเดียวกันกล้องตัวที่ 4 มีการตรวจจับใบหน้าได้ก็แสดงว่าบุคคลนั้นกำลังนั่งดูTV เป็นต้น สำหรับวิธีวิเคราะห์ท่าทางหลังจากการจำแนกในเบื้องต้นใช้วิธี Bag-of-features (BoF) จากภาพที่ได้รับจากกล้องเพื่อเก็บข้อมูลสำหรับเป็นอินพุตให้ Support Vector Machine (SVM) ที่เป็นวิธีการสำหรับการวิเคราะห์เหตุการณ์จากการสอนระบบจากข้อมูลที่ส่งเข้ามา โดยงานวิจัยชิ้นนี้จัดอยู่ในกลุ่มของ Parametric Object Model -> Trajectory Model

ตารางที่ 1-1 แสดงความสัมพันธ์ระหว่างพื้นที่กับเหตุการณ์

ตำแหน่ง	**พฤติกรรม**
ห้องครัว	ตัด, กวน, ดูดฝุ่น, อื่นๆ
ห้องกินข้าว	กิน, ดูดฝุ่น, อื่นๆ
ห้องนั่งเล่น	ดู TV, อ่านหนังสือ, ดูดฝุ่น, อื่นๆ
ห้องอ่านหนังสือ	เขียนหนังสือ, อ่านหนังสือ, ดูดฝุ่น, อื่นๆ

1.4.4 สรุปงานวิจัยที่เกี่ยวข้อง

ในงานวิจัยของ Nattapon Noorit, Nikom Suvonvorn, และ Montri Karnchanadecha [7] สามารถระบุท่าได้อย่างแม่นยำ และรวดเร็วในงบประมาณที่ต่ำเนื่องจากใช้กล้องเพียงตัวเดียวซึ่งการใช้กล้องเพียงตัวเดียวนี้เองทำให้เกิดปัญหาการระบุท่าทางในมุมมองที่น้อยกว่าทำให้ไม่สามารถระบุท่าทางในขณะที่มนุษย์หันหน้าเข้าหากล้องได้N. Gkalelis N. Nikolaidis และ I. Pitas [8] , C. Canton-Ferrer [5] , C. Wu A. H Khalili และ H. Aghajan [11] สามารถระบุท่าได้ทั้งมุมมองหน้าตรง และมุมด้านข้าง โดยใช้ภาพในมุมมองที่มากกว่า 1 มุมมองจึงทำให้การประมวลผลทำได้ช้างบประมาณที่ค่อนข้างสูงและการวิเคราะห์การเป็นไปได้ที่จะเป็นมนุษย์ยังทำได้ไม่ดีDaniel Chen, Pi-chi Chou, Clinton Fookes และ Sridha Sridharan [9] สามารถระบุท่าทางได้ทุกมุมมองโดยใช้ภาพในมุมมองที่มาก และมีการวิเคราะห์ความเป็นไปได้ที่เป็นมนุษย์สิ่งที่สนใจ เนื่องจากการใช้มุมมองที่มากทำให้การประมวลผลทำได้ช้ามาก และงบประมาณก็สูงตามจำนวนมุมมองที่ใช้ ซึ่งการติดตั้งระบบก็ทำได้ยากเช่นกัน ตารางที่ 1-2 เป็นการเปรียบเทียบงานวิจัยที่เกี่ยวข้อง

ตารางที่ 1-2 เปรียบเทียบงานวิจัยทีเกี่ยวข้อง

ชื่อผู้ทำวิจัยที่เกี่ยวข้อง	วัตถุหันข้างให้กล้อง	วัตถุหันหน้าให้กล้อง	ความเร็วในการประมวลผล	การตรวจสอบการเป็นมนุษย์	งบประมาณต่ำ
N.Noorit [7]	/	X	/	/	/
N. Gkalelis [8]	/	/	X	X	X
D.Chen [9]	/	/	X	/	X
C. Canton [5]	/	/	X	X	X
C. Wu [11]	/	/	X	X	X
งานวิจัยนี้	/	/	/	/	/

จากงานวิจัยของ Nattapon Noorit, Nikom Suvonvorn, และ Montri Karnchanadecha [7] ที่สามารถระบุท่าได้อย่างแม่นยำ และรวดเร็ว ในงบประมาณที่ต่ำ แต่ไม่สามารถระบุท่าทางจาก

มุมมองด้านหน้าได้เนื่องจากใช้เพียงมุมมองเดียวในการประมวลผล แต่จากงานวิจัยอื่นแก้ปัญหานี้โดยการใช้ภาพจากมุมมองที่มากกว่า 1 มุมมองมาทำการแก้ปัญหานี้สำหรับในงายวิจัยนี้ใช้ภาพจากแค่ 2 มุม โดยการใช้พารามิเตอร์มาประมวลผลไม่ได้ใช้ทั้งภาพเหมือนในบางงานวิจัยที่กล่าวมาทำให้ลดการประมวลผลไปได้มาก และงบประมาณที่ต่ำกว่าจากการใช้เพียง2 กล้อง

1.5 ขอบเขตของการวิจัย

1.5.1 สามารถวิเคราะห์ได้มีทั้งหมด5 ท่าทาง คือ การเดิน การนอน การนั่ง การยืน การก้ม

1.5.2 สามารถค้นหา และระบุได้ว่าสิ่งที่ตรวจพบเป็นมนุษย์จริงๆโดยจะต้องมีความแม่นยำมากกว่า 90% ซึ่งการระบุลักษณะอ้างอิงจากโมเดลมนุษย์อันนำไปสู่การใช้วิเคราะห์หาท่าทางมนุษย์

1.5.3 สามารถใช้งานได้ในสภาพแวดล้อมจริง

1.6 ขั้นตอนและวิธีการดำเนินงานวิจัย

การดำเนินงานของโครงงานนี้จะแบ่งออกเป็น 4 ช่วงหลักๆดังนี้คือ

1.6.1 พัฒนาเทคนิคการตรวจจับ และติดตาม มนุษย์ให้ได้โดยอ้างอิงจากโมเดลที่ได้ออกแบบไว้ไม่ว่าจะเคลื่อนไปที่ใดโดยสามารถใช้งานได้ในหลายสภาวะแวดล้อมต่างๆโดยจะต้องมีความแม่นยำมากกว่า90% ของสำหรับกล้องแต่ละตัว

1.6.2 พัฒนาเทคนิคการสร้างโมเดลสามมิติ(3D-Model)โดยนำคุณลักษณะที่มาจากกล้องแต่ละตัวในมุมมองที่ต่างกัน(Multi-View) สิ่งที่ได้จะต้องเป็นโมเดลมนุษย์ที่มองจากมุมด้านข้างเสมอและได้พารามิเตอร์เพื่อเข้าสู่กระบวนการวิเคราะห์ท่าทาง

1.6.3 พัฒนาเทคนิคการวิเคราะห์ท่าทางจากของเดิมที่มีอยู่แล้ว โดยรับพารามิเตอร์สำหรับการวิเคราะห์จากโมเดลสามมิติ(3D-Model)

1.6.4 รวมระบบที่ได้พัฒนาร่วมกับซอฟแวร์ระบบรักษาความปลอดภัยโดยใช้กล้องวงจรปิด

บทที่ 2
ทฤษฎีและหลักการ

2.1 การตรวจจับการเคลื่อนไหว

2.1.1 การหาค่าเฉลี่ยอย่างต่อเนื่อง (Running average)

พฤติกรรมบุคคลเกิดขึ้นเมื่อเกิดการเคลื่อนไหว ดังนั้นขั้นตอนแรกจึงจำเป็นต้องมีการตรวจจับวัตถุเคลื่อนไหวต่าง ๆ ที่เกิดขึ้นเพื่อใช้เป็นเงื่อนไขเริ่มต้นในการติดตามพฤติกรรมของบุคคล กระบวนการตรวจจับความเคลื่อนไหวสามารถทำได้โดยใช้เทคนิคการลบพื้นหลัง (Background Subtraction) ซึ่งอธิบายได้โดยสมการดังนี้

$$F_{i+1}(x,y) = \begin{cases} 0 & \text{if } |img_{i+1}(x,y) - B_i(x,y)| > Th \\ 1 & \text{else} \end{cases} \quad (2.1)$$

$$B_{i+1}(x,y) = \begin{cases} \alpha\, img_{i+1}(x,y) + (1-\alpha)B_i(x,y) & \text{if } F_{i+1}(x,y) = 1 \\ B_i(x,y) & \text{else} \end{cases} \quad (2.2)$$

กำหนดให้

$img_i(x,y)$ เป็นค่าของจุดภาพ ณ ตำแหน่งพิกัด(x,y) ของลำดับภาพที่i

$B_{i+1}(x,y)$ และ $B_i(x,y)$ เป็นค่าของจุดภาพ ณ ตำแหน่งพิกัด(x,y) ที่ถูกกำหนดเป็นพื้นหลังของลำดับ ภาพที่ i+1 และ i ตามลำดับ

$F_{i+1}(x,y)$ เป็นค่าของจุดภาพ ณ ตำแหน่งพิกัด(x,y) ที่ถูกกำหนดเป็นพื้นหน้าของลำดับภาพที่ i+1

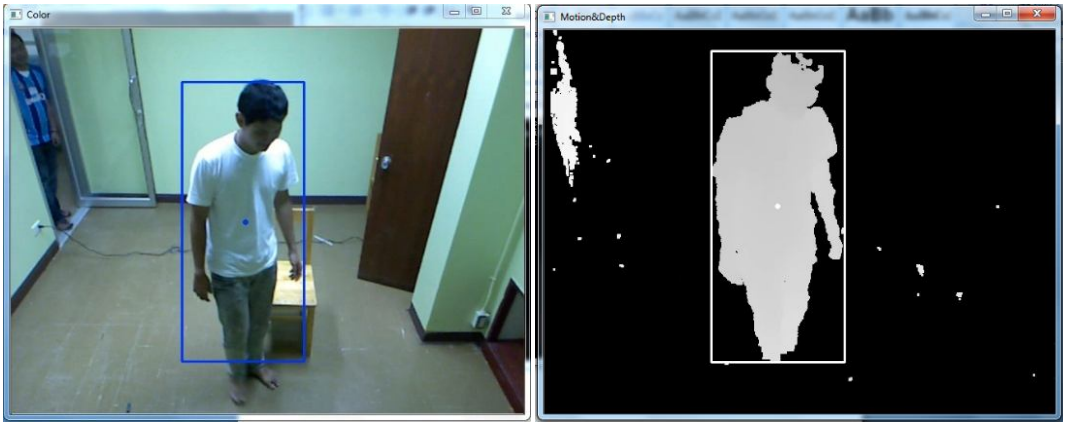
α เป็นตัวแปรซึ่งใช้กำหนดค่าน้ำหนักในการปรับปรุงค่าจุดภาพที่เป็นพื้นหลัง

จากสมการ (2.1) แสดงให้เห็นถึงการคำนวณบริเวณที่เกิดการเปลี่ยนแปลงล่าสุด โดยใช้ค่าภาพที่เฟรมล่าสุดลบด้วยค่าภาพพื้นหลังลำดับก่อนหน้า ถ้าผลการคำนวณค่าความแตกต่างสัมบูรณ์ที่จุดภาพใดมีค่ามากกว่าขีดแบ่ง Th ค่าจุดภาพ $F_{i+1}(x,y)$ จะถูกกำหนดค่าเป็น 0 ซึ่งหมายถึงจุดภาพเหล่านี้เกิดการเปลี่ยนแปลงค่าสีอย่างมาก และอาจเป็นไปได้ว่าอาจเป็นบริเวณที่เกิดการเคลื่อนไหวของวัตถุใด ๆ หลังจากขั้นตอนนี้ภาพ F_{i+1} จะถูกปรับปรุงโดยเทคนิคการดำเนินการปิดและการดำเนินการเปิดเพื่อลบจุดภาพเดี่ยว ๆ หรือกลุ่มจุดภาพที่มีขนาดเล็กมาก ทำให้ F_{i+1} เหลือเพียงวัตถุเคลื่อนไหวที่มีขนาดใหญ่เท่านั้นซึ่งบริเวณเหล่านี้จะถูกคาดหมายว่าเป็นวัตถุเคลื่อนไหว

ซึ่งอาจเป็นมนุษย์ โดยบริเวณที่ถูกตรวจพบว่าเป็นวัตถุเคลื่อนไหว (พื้นหน้า) จะไม่ถูกนำไปปรับปรุงค่าพื้นหลังตามเงื่อนไขในสมการ(2.2)

จากสมการ (2.2) พิกัดจุดภาพที่ไม่ถูกตรวจพบเป็นวัตถุเคลื่อนไหว($F_i(x,y)$ มีค่าเป็น 1) จะเป็นเงื่อนไขในการพิจารณาการปรับปรุงภาพพื้นหลังB_{i+1} โดย ค่า $B_{i+1}(x,y)$ จะถูกปรับปรุงโดยอาศัยค่าจากสองส่วน ได้แก่ ค่าพื้นหลังเดิม $B_i(x,y)$ และค่าจากภาพเฟรมล่าสุด $img_{i+1}(x,y)$ โดยใช้ค่า α เป็นตัวแปรกำหนดน้ำหนักอัตราส่วนของการปรับปรุงค่า$B_{i+1}(x,y)$ ยิ่งค่า α มีค่ามาก ค่า $B_{i+1}(x,y)$ ก็จะมีค่าลู่เข้าใกล้ค่า$img_{i+1}(x,y)$ มาก ภายในงานวิจัยนี้เลือกใช้ค่าα เท่ากับ 0.05

อย่างไรก็ตาม วิธีการนี้ก็ยังคงใช้ค่าขีดแบ่งแบบคงที่ไม่ได้กำหนดกระบวนการคำนวณค่าขีดแบ่งอย่างชัดเจน และวิธีการนี้ยังคงมีพื้นฐานอยู่บนการอ้างอิงค่าเพียงค่าเดียวซึ่งทำให้ไม่สามารถจัดการกับค่าพื้นหลังที่มีการกระจายหลายรูปแบบได้



(a) (b)

ภาพประกอบ 2-1 ตัวอย่างผลการลบพื้นหลังโดยการหาค่าเฉลี่ยอย่างต่อเนื่อง (a) เฟรมปัจจุบัน (b) ภาพวัตถุเคลื่อนไหว

2.1.2 การสร้างพื้นหลังจากการผสมของระเบียบวิธีเกาส์เซียนหลายรูปแบบ(Mixture of Gaussian)

เทคนิคนี้เป็นการสร้างโมเดลของภาพพื้นหลังที่สามารถปรับค่าให้เป็นปัจจุบันได้ตลอดเวลาให้เหมือนกับพื้นหลังของทุกๆเฟรมปัจจุบัน ที่มีวัตถุเคลื่อนที่เกิดขึ้นภายในเฟรมนั้นๆ แต่ละจุดสีของเฟรมภาพจะถูกแยกประเภทเป็นจุดสีของภาพพื้นหลัง หรือจุดสีของวัตถุด้วยวิธีเกาส์เซียนดิสทริบิวชันที่มากกว่า 1 ที่มีประสิทธิภาพในการหาโมเดลของภาพพื้นหลังที่มีความซับซ้อน กล่าวคือในสภาพแวดล้อมจริงที่มีการเปลี่ยนแปลงที่ไม่แน่นอนของสภาพแวดล้อม เช่น ความเข้ม

ของแสง เงาจากต้นไม้ หรือเงาจากอาคารต่างๆ ทำให้ค่าของจุดสีที่ตำแหน่งเดิมของแต่ละลำดับเฟรมภาพเปลี่ยนแปลงตามสภาพอากาศ และสภาพแวดล้อมได้ ทำให้การสร้างโมเดลของภาพพื้นหลังเพียงหนึ่งโมเดลอาจจะไม่เพียงพอกับภาพวิดีโอที่มีความซับซ้อน

ตัวอย่างที่น่าสนใจอย่างยิ่ง คือ จากงานวิจัยของ(C.Stauffer,1999) สร้างโมเดลภาพพื้นหลังที่มีหลายค่าเพื่อแก้ปัญหาการของจุดสีของภาพพื้นหลังที่มีหลายค่า ทำให้การแบ่งประเภทจุดสีถูกต้องมากยิ่งขึ้น ซึ่งความน่าจะเป็นของการพิจารณาค่าจุดสี สามารถอธิบายได้ดังสมการ(2.3)

$$P(x_t) = \sum_{i=1}^{K} \omega_{i,t}\eta(x_t - \mu_{i,t}\sum_{i,t} \quad) \tag{2.3}$$

กำหนดให้

K คือ จำนวนของการกระจายของเกาส์เซียน

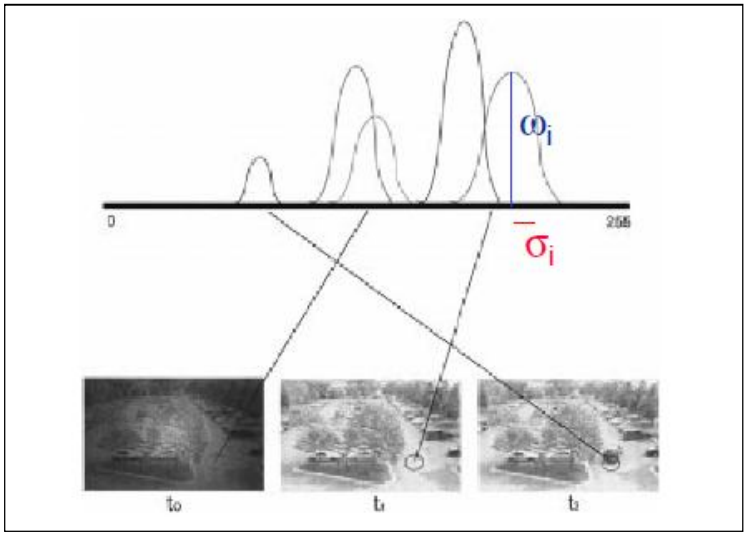
$\omega_{i,t}$ คือ ค่าน้ำหนักของเกาส์เซียนตัวที่th ที่เวลา t

$\mu_{i,t}$ คือ ค่าเฉลี่ยของเกาส์เซียนตัวที่ ith ที่เวลา t

$\sum_{i,t}$ คือ ค่าความแปรปรวนร่วมของเกาส์เซียนตัวที่ ith ที่เวลา t

η คือ Gaussian Probability Density Function

เกณฑ์ในการแยกประเภทจุดสี ว่าเป็นจุดสีของภาพพื้นหลังหรือจุดสีของวัตถุเคลื่อนที่ หาได้จากเมื่อพิจารณาจุดสีค่าแอมพลิจูดสูงสุดจะมีค่าของส่วนเบี่ยงเบนมาตรฐานเป็น σ ดังภาพประกอบ 2-2 ถ้าดิสทริบิวชันที่มีความหนาแน่นมากจะเป็นส่วนของภาพพื้นหลัง ดังสมการ(2.4)



ภาพประกอบ 2-2 การแยกประเภทจุดสี ของวิธี Mixture of Gaussians

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

เมื่อ T คือ ค่าขีดแบ่ง

จากสมการ(2.4) ถ้าแอมพลิจูดของดิสทริบิวชันมีค่ามากกว่าค่าขีดแบ่งดิสทริบิวชันนั้นจะเป็นส่วนของภาพพื้นหลังและทำการปรับค่าพารามิเตอร์$(\omega_{i,t}, \mu_{\mathrm{i,t}}, \sigma_{\mathrm{i,t}})$ ให้เป็นปัจจุบัน แต่ถ้านอกเหนือจากนี้ดิสทริบิวชันนั้นจะกลายเป็นส่วนของวัตถุที่เคลื่อนที่ในเฟรมภาพตัวอย่างผลจากวิธี Mixture of Gaussians ดังภาพประกอบ 2-3

(a) (b)

ภาพประกอบ 2-3 ผลจากวิธี Mixture of Gaussians (a) ภาพวิดีโอตัดต่อเฟรมปัจจุบัน (b) ภาพวัตถุเคลื่อนที่

2.2 การติดตามสีโดยใช้ Mean-Shift

Mean-shift algorithm [3] [4] เป็นวิธีการหยาบๆของการหาความแน่นอนเฉพาะที่ในการจัดวางข้อมูลซึ่งเป็นกระบวนการง่ายๆในการจำแนกอย่างต่อเนื่อง ในคำอธิบาย เป็นสิ่งที่จำเป็นอย่างยิ่งในการนำมาประยุกต์hill climbing ในการจำแนกความหนาแน่นของข้อมูล อย่างไรก็ตาม สำหรับความไม่ต่อเนื่องของการวางข้อมูล ถือเป็นปัญหาเล็กน้อยที่ไม่สำคัญมาก ได้มีผู้อธิบายว่า การนำมาใช้เพื่อความเข้าใจทางสถิติ นั่นคือMean-shift ไม่ยอมรับข้อมูลที่มาจากภายนอกนั่นหมายถึงเป็นการไม่ยอมรับจุดที่เป็นข้อมูลซึ่งอยู่ไกลจากจุดสูงสุดของข้อมูล วิธีการคือ ประมวลเฉพาะจุดนั้นภายในlocal window ของข้อมูลและจากนั้นก็จะนำออกจากwindow นั้นเอง

สำหรับโครงงานนี้ใช้Cam-Shift สำหรับการติดตาม (tracking) โดย Cam-Shift จะต่างจาก Mean-shift ก็ตรงที่ Cam-Shift สามารถปรับเปลี่ยนขนาดของกรอบที่ติดตามวัตถุได้เท่านั่นเอง

มีขั้นตอนของอัลกอริทึมดังนี้

ขั้นตอนที่1 เลือกสิ่งที่เราสนใจที่จะติดตามที่มีความเป็นไปได้ที่จะจำแนกได้

ขั้นตอนที่2 ตีกรอบเริ่มต้นของสิ่งที่ต้องการจะติดตาม

ขั้นตอนที่3 คำนวณสีที่เป็นไปได้สำหรับการจำแนก(Histogram)

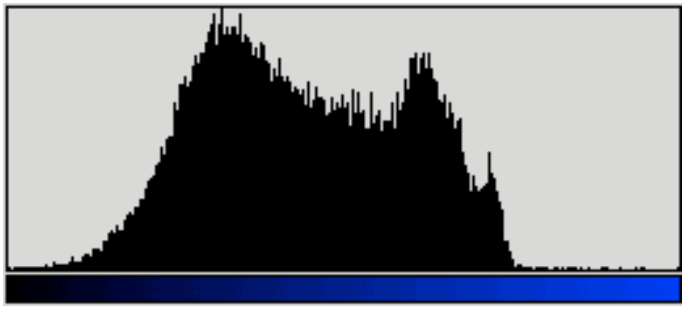
ขั้นตอนที่4 หาค่า center of mass ของค่าที่เป็นไปได้ของรูปและเก็บค่า zero moment (distribution area) และ center of mass location.

ขั้นตอนที่5 เลื่อน frame จุดกึ่งกลางของกรอบคือจุดที่หาได้จากขั้นตอนที่4 และขนาดของกรอบคือค่าที่หาได้จากฟังก์ชั่นหาzero moment จากนั้นกลับไปทำขั้นตอนที่3

2.2.1 Histogram

การหา Histogram จะหาจากค่า ROI (Region of Interest) ที่ได้จากขั้นตอนที่1 หาโดยใช้ Hue channel ใน HSV color model สำหรับการหาโดยการหาจะนำค่าที่หาได้เก็บเอาไว้เพื่อที่จะลดการคำนวณ ความซับซ้อน และสามารถจัดกลุ่มที่เหมือนให้อยู่ด้วยกันได้ และสามารถกำหนดขนาดได้ ดังภาพประกอบที่ 2-4 ในกรอบสีเหลืองคือวัตถุที่เราสนใจจะติดตาม(tracking) และภาพประกอบที่ 2-5 คือ histogram ของวัตถุที่จะติดตาม

ภาพประกอบ 2-4 Original Image



ภาพประกอบ 2-5 Histogram

2.2.2 Histogram Back-Projection

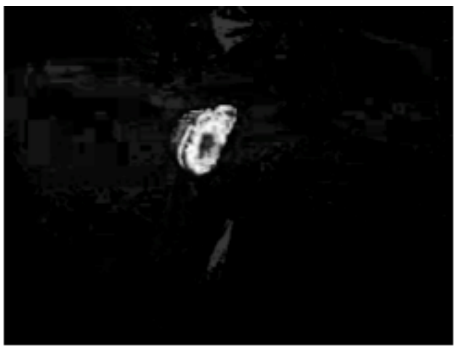
Back-Projection โดยเริ่มแรกจะเป็นการคำนวณความสัมพันธ์ระหว่างค่าในจุด(pixel) ของรูปทั้งหมดนำมาเทียบหาค่าที่ตรงกันกับค่าในHistogram ที่เก็บไว้ โดยการคำนวณสามารถคำนวณได้จากสมการดังนี้

$$q_u = \sum_{i=0}^{n} \rho[c(x_i) - u] \tag{2.5}$$

$$\left\{ p_u = \min\left(\frac{255}{\max(q)} q_u\right), 255 \right\}_{u=1\ldots m} \tag{2.6}$$

เป้าหมาย คือ ความเป็นไปได้ทั้งหมดของเฟรมว่าค่าในแต่ละจุด(Pixel) ของรูปที่มีค่าตรงกับค่าใน Histogram ดังภาพประกอบที่ 2-7 โดยส่วนที่เป็นสีขาวคือผลจากการที่ค่าในจุด (pixel) ตรงกับค่าของ histogram ของวัตถุที่สนใจ โดยวัตถุที่สนใจในรูปคือหมวกที่โดนตีกรอบ ดังภาพประกอบที่ 2-6

ภาพประกอบ 2-6 Original Image



ภาพประกอบ 2-7 Back-Projection Image

2.2.3 Mass Centre Calculation

Center of mass คือจุดศูนย์กลางของกรอบที่ตีอยู่รอบๆสิ่งที่กำลังติดตาม โดยสามารถคำนวณหาจุด center of mass สามารถคำนวณได้ดังต่อไปนี้

คำนวณหาค่าจุดศูนย์กลาง

$$M_{00} = \sum_x \sum_y I(x, y)$$

คำนวณหาค่าจุดศูนย์กลางของx และ y เริ่มต้น

$$M_{10} = \sum_x \sum_y xI(x, y)$$

$$M_{01} = \sum_x \sum_y yI(x, y)$$

คำนวณหาค่าเฉลี่ย

$$x_c = \frac{M_{10}}{M_{00}} \; ; \; y_c = \frac{M_{01}}{M_{00}}$$

โดยกำหนดให้I(X, Y) คือ intensity ของ pixel นั้น

2.3 การประมวลผลภาพแบบสเตอริโอ

การประมวลผลภาพแบบสเตอริโอ (Stereo Vision) เป็นการประมวลผลภาพที่ทำให้ได้ข้อมูลความลึกของวัตถุในภาพ ซึ่งเกิดจากการนำภาพสองภาพจากกล้องที่ถ่ายจากมุมมองที่อยู่ในระนาบเดียวกัน โดยกล้องจะต้องมีระยะห่างค่าหนึ่งมาประมวลผลภาพ โดยจับคู่เปรียบเทียบ (Stereo Matching) ตำแหน่งของภาพที่เป็นจุดเดียวกันระหว่างภาพซ้าย และภาพขวา โดยเรียกระยะห่างระหว่างจุดทั้งสองจุดนี้ ว่าดิสพาริที(Disparity) ซึ่งกระบวนการจับคู่นี้ สามารถนำไปคำนวณหาความลึกของวัตถุได้แสดงตัวอย่างดังภาพประกอบที่ 2-8

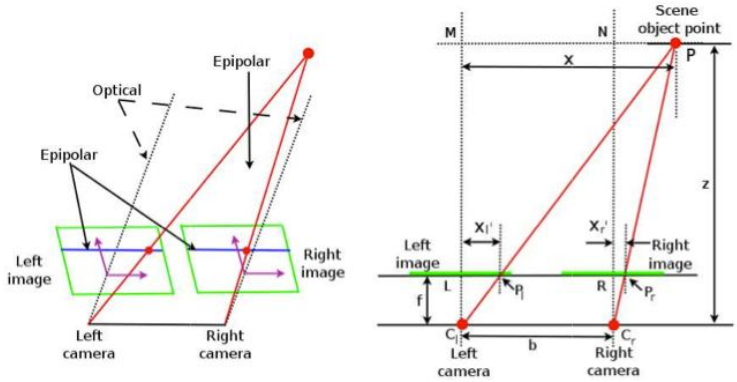
Left Image Right Image

Disparity Image

ภาพประกอบ 2-8 ตัวอย่างภาพจากกล้องซ้ายๆ กล้องขวา และภาพที่ผ่านการจับคู่จากทั้งสองกล้อง

2.3.1 ลักษณะทางเลขาคณิต (Epipolar Geometry)

เป็นการคำนวณหาระยะของจุดที่สนใจเทียบกับกล้องที่อยู่ในระนาบเดียวกันโดยใช้ลักษณะทางเรขาคณิตเข้ามาช่วยในการคำนวณ โดยการวางตำแหน่งของกล้องในแบบจำลองจะถูกติดตั้งในลักษณะที่อยู่ในระนาบเดียวกันในทางแกนX โดยระยะห่างจากจุดศูนย์กลางของกล้องทั้งสองคือจุด b (Base Line) ซึ่งจุดที่เราสนใจเมื่อมาอยู่บนภาพทั้งสองระยะห่างของทั้งสองจุดเรียกว่าดิสพาริทีและระนาบสามเหลี่ยมเมื่อเทียบจากจุดบนภาพซ้ายไปยังจุดที่เราสนใจจากภายนอกผ่านจุดบนภาพขวา ดังภาพประกอบที่ 2-9a เส้นสีแดงเรียกว่า Epipolar Plane และเส้นที่เกิดการตัดกันของระนาบภาพกับ Epipolar Plane เรียกว่า Epipolar Line ซึ่งหมายถึงเส้นที่กล้องขวามองเห็นทิศทางการพุ่งไปยังจุดที่สนใจของกล้องซ้ายจากตำแหน่งของกล้องขวานั่นคือEpipolar Line ในกล้องขวา ซึ่งกล้องซ้ายก็มีลักษณะวิธีการเดียวกัน ดังภาพประกอบที่ 2-9a เส้นสีน้ำเงิน



(a) (b)

ภาพประกอบ 2-9 ลักษณะทางเลขาคณิตที่สัมพันธ์กันระหว่างกล้องทั้งสอง

ดังภาพประกอบที่ 2-9b สามารถหาระยะจากจุดที่สนใจเทียบกับกล้องได้จากการใช้คุณสมบัติของสามเหลี่ยมคล้าย โดยที่จุด $\mathbf{P_l}$ และ $\mathbf{P_r}$ คือจุดที่สนใจซึ่งอยู่บนระนาบภาพซ้าย และระนาบภาพขวาตามลำดับ และ f คือ ระยะโฟกัสของกล้องทั้งสองในที่นี้คือ $\mathbf{C_l}$ และ $\mathbf{C_r}$ ซึ่งก็คือตำแหน่งศูนย์กลางของเลนส์กล้องซ้าย และกล้องขวา ซึ่งX คือระยะจากจุดศูนย์กลางเลนส์กล้องซ้าย ($\mathbf{C_l}$) เทียบกับจุด P ซึ่งก็คือจุดที่เราต้องการจะหาระยะ Z นั่นเอง โดยที่สมการหาค่า Z หาได้จากการเปรียบเทียบสามเหลี่ยมคล้ายดังสมการที่(2.7), (2.8) และ (2.9) ตามลำดับ

จากสามเหลี่ยมคล้าย $\mathbf{PMC_l}$ และ $\mathbf{P_lLC_l}$ จะได้สมการดังนี้

$$\frac{\mathbf{X}}{\mathbf{Z}} = \frac{\mathbf{X_l'}}{\mathbf{f}} \qquad (2.7)$$

จากสามเหลี่ยมคล้าย $\mathbf{PNC_r}$ และ $\mathbf{P_rLC_r}$ จะได้สมการดังนี้

$$\frac{\mathbf{X-b}}{\mathbf{Z}} = \frac{\mathbf{X'_r}}{\mathbf{f}} \qquad (2.8)$$

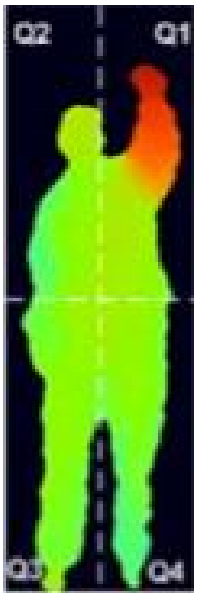
แทนสมการ(2.7) ในสมการ (2.8)

$$\mathbf{Z} = \frac{\mathbf{bf}}{\mathbf{X'_l - X'_r}} \qquad (2.9)$$

จากสมการที่ (2.9) แสดงให้เห็นว่าสามารถหาระยะห่างระหว่างจุดที่สนใจกับจุดศูนย์กลางเลนส์ของกล้องได้ จากสมการ โดยใช้ที่ค่า $\mathbf{X'_l - X'_r}$ ซึ่งก็คือค่า ดิสพาริที นั่นเอง

2.4 3D reconstruction

สำหรับงานวิจัยนี้เรื่องของการโมเดลภาพสามมิติจะไม่ให้ความสำคัญเรื่องของความแม่นยำของค่าพารามิเตอร์ใน(x, y, z) มากนัก แต่จะให้ความสำคัญเรื่องความสัมพันธ์ของข้อมูลความลึกที่เปลี่ยนไปของจุดที่สนใจมากกว่า จากตัวอย่างเป็นวิธีการวิเคราะห์ท่าทางโดยแบ่งใช้พื้นที่ในภาพเข้ามาผนวกกับข้อมูลความลึก โดยที่ส่วนที่เป็นสีแดงจะอยู่ห่างจากกล้องมากกว่าส่วนที่เป็นสีเขียว [13] ซึ่งการวิเคราะห์จะใช้แค่ว่าถ้าส่วนที่เป็นสีแดงอยู่ในบริเวณภาพ Q1 ซึ่งก็คือมีส่วนที่เข้าใกล้ยื่นเข้าใกล้กล้องมากกว่าส่วนอื่นก็สามารถวิเคราะห์ได้ว่าเป็นการยกมือซ้าย ดังภาพประกอบที่ 2-10



ภาพประกอบ 2-10 ตัวอย่างการวิเคราะห์ท่าทางการยกมือโดยใช้ภาพสามมิติ

2.5 Morphological

Morphological Image Processing เป็นการประมวลผลภาพโดยการเปลี่ยนแปลงลักษณะรูปร่างหรือโครงสร้างของภาพ โดยมี structuring element เป็น input เพื่อที่จะประมวลผลโดยใช้ operator ต่างๆมากระทำ เช่น การนำมาintersection, union เป็นต้น [1]

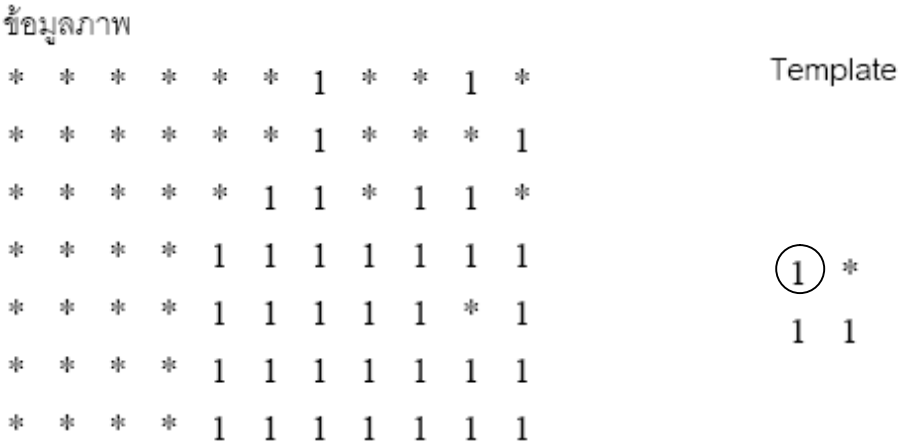
โดยปกติ structuring element เป็นเมกตริกซ์ ขนาด 3 คูณ 3 มีจุดกำเนิดที่ center โดยกระบวนของอัลกอริทึมนี้คือ กลุ่ม pixel ของ structuring element จะถูกนำไปเปรียบเทียบกับกลุ่ม pixel ของภาพ ไปเรื่อย ๆ จนทั่วทั้งภาพต้นแบบ หาก2 กลุ่ม pixel ใดที่ทำการเปรียบเทียบสามารถจับคู่กันได้ลงตัว ภายใต้เงื่อนไขของoperator ที่เอามาใช้ จะทำให้เกิดการเปลี่ยนแปลงของกลุ่มpixel นั้น ๆ จะทำให้เกิดการเปลี่ยนโครงสร้างของภาพใหม่ เรียกเทคนิคนี้ ว่า การhit และ Miss

โดยโอเปอเตอร์ต่างๆจะถูกรวมไว้อยู่ในโอเปอร์เรชั่นพื้นฐาน ได้แก่ Dilation, Erosion และ Skeleton โดยการ Dilation คือการขยายภาพโดยมีสัดส่วนเท่ากันทั่วทั้งภาพ(Uniform) การ Erosion คือการย่อภาพ ส่วนการทำ Skeleton เป็นการหาโครงสร้างหลักของวัตถุซึ่งจะกล่าวโดยละเอียดต่อจากนี้ นอกจากโอเปอเรชั่นพื้นฐานดังที่ได้กล่าข้างต้นแล้วยังมีโอเปอเรชั่นอื่นๆได้แก่ การ Opening และ closing เป็นต้น

การทำ Morphological สามารถนำไปประยุกต์ใช้ได้กับ ภาพ Grey-level ได้เช่น นำไปลด noise ของภาพ หรือ จะปรับความสว่างของภาพ เป็นต้น อย่างไรก็ตาม การทำ Morphological operations เป็นเพียงส่วนหนึ่งของ Image processing ยังมีกระบวนการอื่น ๆ อีกมากมายที่สามารถนำมาใช้ปรับแต่งโครงสร้างภาพให้มีความเหมาะสมตามความต้องการของผู้ใช้ แต่ในโครงงานนี้ใช้เพียง การกัดกร่อนภาพ (Erosion) และการพอกภาพ (Dilation)

2.5.1 การกัดกร่อนภาพ (Erosion)

การกัดกร่อนภาพเป็นลักษณะของการลบข้อมูลภาพบริเวณขอบของภาพการกัดกร่อนภาพสามารถทำได้มีลักษณะคล้ายกับการพอกภาพโดยการสร้าง Template ขึ้นแล้วนำ Template ไปสแกนตามข้อมูลภาพสำหรับทุกตำแหน่งที่เลื่อน Template ไปบนภาพก็จะมีการเปรียบเทียบกับข้อมูลภาพถ้าข้อมูลภาพมีความเหมือนกับ Template จะทำการกำหนดค่าข้อมูลภาพในตำแหน่งที่ตรงกับจุดเริ่มต้น(Origin) ของ Template ถูกกำหนดให้มีค่าเท่ากับ1



ภาพประกอบ 2-11 ตัวอย่างข้อมูลภาพ และTemplate การทำErosion

ผลที่ได้จะมีเพียง 3 ตำแหน่งเท่านั้นที่มีค่าเหมือนกับTemplate

ภาพประกอบ 2-12 ตัวอย่างผลลัพธ์การทำErosion

ผลที่ได้ดังภาพประกอบที่ 2-12 ข้อมูลภาพที่ผ่านการทำโอเปอเรชั่นกับTemplate แล้วพบว่ามีข้อมูลของภาพเพียง 3 ตำแหน่งเท่านั้นที่เหมือนกับTemplate ถ้ามีการเปลี่ยน Template เป็น 1 1

ผลที่ได้มีลักษณะดังนี้คือ 1 1

```
* * * * * * * * * * *
* * * * * * * * * * *
* * * * * * * * * * *
* * * * * 1 * * 1 * *
* * * * 1 1 1 1 1 * *
* * * * 1 1 1 1 * * *
* * * * 1 1 1 1 1 1 *
```

ภาพประกอบ 2-13 ผลลัพธ์เมื่อทำการเปลี่ยนแปลงTemplate

ผลที่ได้ดังภาพประกอบที่ 2-13 จะเห็นว่าเป็นการย่อขนาดภาพแต่ย่อได้น้อยกว่าเมื่อใช้ Template แรก จะเห็นได้ว่าการเลือก Template เป็นสิ่งสำคัญอย่างหนึ่งในการย่อและขยายภาพ

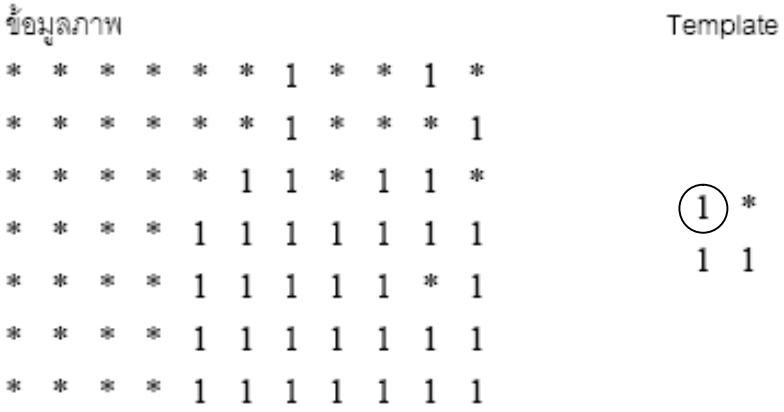
ผลของการทำErosion ย่อพื้นหน้า ขยายพื้นหลังในภาพแบบGray level สามารถกำจัด “salt noise” (จุดเล็ก ๆ สีขาวบนภาพ) ออกไปได้ลดความสว่างของภาพ แยกวัตถุออกเป็นส่วน ๆ (เห็นได้ชัดในภาพแบบ Gray level) โดยลดขอบของวัตถุในภาพลง

(a) (b)

ภาพประกอบ 2-14 ภาพก่อนทำและหลังทำErosionจากตัวอย่างจะเห็นว่าหลังจากการทำErosion ทำให้“salt noise” จุดเล็ก ๆ สีขาวในภาพหายไป

2.5.2 การพอกภาพ (Dilation)

การขยายภาพในที่นี้จะพิจารณาสำหรับข้อมูลภาพที่เป็นแบบไบนารี โดยการใช้เทคนิคการ Hit และ Miss ตามที่ได้กล่าวไว้การพอกภาพจะทำได้โดยกำหนดTemplate (ซึ่งสามารถสร้างได้จาก * และ 1 โดยมีจุดเริ่มต้นที่กำหนดโดยวงกลม) และนำ Template นี้สแกนไปบนข้อมูลภาพตามลำดับตลอดทั้งภาพซึ่งในขณะที่จุดเริ่ม(Origin) ของ Template ตรงกับตำแหน่ง ข้อมูลภาพที่พิกเซลมีค่าเท่ากับ 1 นั้นก็จะทำการยูเนียนTemplate นี้เข้ากับข้อมูลภาพดังตัวอย่าง



ภาพประกอบ 2-15 ตัวอย่างข้อมูลภาพ และTemplate การทำDilation

ข้อมูลแถวแรกของภาพเป็นดังนี้

* * * * * * 1 * * 1 *

เมื่อทำการยูเนียนกับTemplate ณ.ตำแหน่งข้อมูลภาพที่พิกเซลเท่ากับ1 ในแถวแรก

* * * * * * 1 * * 1 *

* * * * * * 1 1 * * *

และเมื่อยูเนียนกับTemplate เข้ากับพิกเซลที่มีค่าเท่ากับ1 ณ. ตำแหน่งพิกเซลที่สองในแถวแรก

* * * * * * 1 * * 1 *

* * * * * * 1 1 * 1 1

และเมื่อทำการยูเนียนทั้งภาพจะได้ภาพสุดท้ายดังนี้

*	*	*	*	*	*	1	*	*	1	*	*
*	*	*	*	*	*	1	1	*	1	1	*
*	*	*	*	*	1	1	1	1	1	1	1
*	*	*	*	1	1	1	1	1	1	1	1
*	*	*	*	1	1	1	1	1	1	1	1
*	*	*	*	1	1	1	1	1	1	1	1
*	*	*	*	1	1	1	1	1	1	1	1
*	*	*	*	1	1	1	1	1	1	1	1

ภาพประกอบ 2-16 ตัวอย่างผลลัพธ์การทำ Dilation

ผลของการทำ Dilation ขยายพื้นหน้า ย่อพื้นหลังในภาพแบบ Gray level สามารถกำจัด “pepper noise” (จุดดำเล็ก ๆ บนภาพ) ออกไปได้เพิ่มความสว่างของภาพทำให้เส้นขอบเรียบขึ้น

(a) (b)

ภาพประกอบ 2-17 ภาพก่อนทำและหลังทำ Dilation จากตัวอย่างจะเห็นว่า“pepper noise” (จุดดำเล็ก ๆ บนภาพ) หายไปและเพิ่มความสว่างของภาพ

2.6 Neural Network

โครงข่ายประสาทเทียม (Artificial neural network) คือ แบบจำลองทางคณิตศาสตร์สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (connectionist) แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาข่ายงานไฟฟ้าชีวภาพ(bioelectric network) ในสมอง ซึ่งประกอบด้วยเซลล์ประสาท หรือ นิวรอน (neurons) และ จุดประสานประสาท (synapses) ข่ายงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาท จนเป็นเครือข่ายที่ทำงานร่วมกัน ซึ่งรูปแบบการคำนวณนั้นค่อนข้างซับซ้อนโดยใช้ข้อมูลต่างๆที่มีอยู่ให้ออกมาในรูปแบบของโครงสร้างวิธีการคำนวณ หรือที่เรียกว่าความรู้ ประสบการณ์ เพื่อที่จะนำรูปแบบที่ได้จากการเรียนรู้นี้ไปใช้ในการวิเคราะห์ตีความหรือคาดคะเนความหมายของข้อมูลที่อยู่ในลักษณะเดียวกัน ซึ่งวิธีการดังกล่าวจะเป็นการเลียนแบบวิธีการทำงานของสมองมนุษย์ หรืออาจจะกล่าวได้ว่าวิธีการดังกล่าวเป็นการจำลองการทำงานของสมองนั่นเองและความรู้ที่ได้อาจจะเกิดขึ้นได้จากกระบวนการเรียนรู้

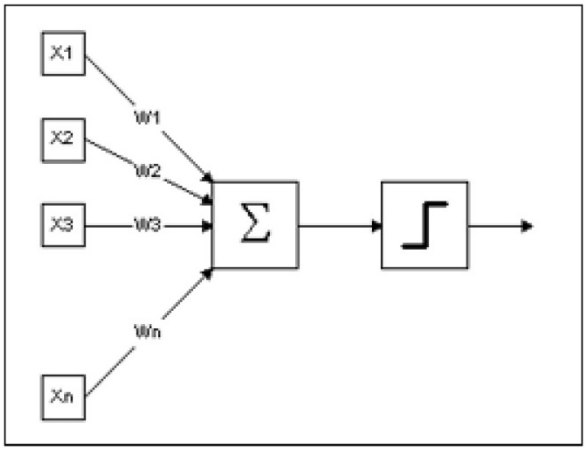
โดยโครงข่ายงานประสาทเทียมจะรวมกลุ่มแบบขนานของหน่วยประมวลผลย่อยๆ ซึ่งการเชื่อมต่อในโครงสร้าง ทำให้เกิดความรู้ ประสบการณ์ ความฉลาดของข่ายงาน ซึ่งนำรูปแบบที่ได้จากการเรียนรู้นี้ไปใช้ในการวิเคราะห์ตีความหรือความหมายของข้อมูลที่อยู่ในลักษณะคล้ายกัน ซึ่งวิธีการดังกล่าวจะเป็นการเลียนแบบวิธีการทำงานของสมองมนุษย์ ใน 2 ลักษณะดังนี้

1. ความรู้ ประสบการณ์ หรือความฉลาดนั้นเกิดจากกระบวนการเรียนรู้(learning process)

2. ความรู้ถูกเก็บที่ การเชื่อมต่อระหว่างเซลล์ประสาท หรือ นิวรอน (neurons) และ จุดประสานประสาท (synapses) เรียกว่า synaptic weights

โครงสร้างของโครงข่ายประสาทเทียม ประกอบด้วย Input units, output units การทำงานของข่ายประสาทเทียม คือเมื่อมี input เข้ามายัง network ก็เอา input มาคูณกับ weight ของแต่ละขา ผลที่ได้จาก input ทุกๆขาของ neuron จะเอามารวมกันแล้วเอามาเทียบกับthreshold ที่กำหนดไว้ถ้าผลรวมมีค่ามากกว่า threshold แล้ว neuron ก็จะส่ง output ออกไป output นี้ก็จะถูกส่งไปยังinput ของ neuron อื่นๆ ที่เชื่อมกันใน network ถ้าค่าน้อยกว่าthreshold ก็จะไม่เกิด output

แนวคิดของแบบจำลองคณิตศาสตร์ดังกล่าวแสดงให้เห็นว่าเริ่มต้นที่การมีสัญญาณนาเข้า (input) คือ X จำนวนหนึ่ง (เช่น n สัญญาณ) เข้ามารวมกันอยู่ในที่ๆ หนึ่ง (ซึ่งแสดงด้วยเครื่องหมาย sum) แล้วก่อนที่จะยิงสัญญาณนั้นออกไปก็อาจจะมีการแปลงสัญญาณซึ่งการแปลงสัญญาณนี้อาจจกระทำผ่านฟังก์ชั่นบางอย่าง แล้วจึงจะออกมาเป็นสัญญาณส่งออก (output) คือ Y ที่จะส่งไปยังเป็น input ของเซลล์สมองตัวต่อไปต่อมาเมื่อได้แบบจำลองเซลล์สมองแล้ว ก็อาจจะพิจารณาได้ว่าในบรรดาข้อมูลนำเข้าทั้งหลายนั้น (X) ข้อมูลแต่ละเรื่องอาจจะมีความสำคัญมากน้อยต่างกัน จึงมีการกำหนดค่าน้ำหนักให้กับแต่ละข้อมูลดังภาพประกอบ 2-18



ภาพประกอบที่ 2-18 โมเดลของข่ายประสาทเทียม

ผลลัพธ์ที่ได้ออกมาดังกล่าวหากจะส่งต่อต้องมาแปลงสัญญาณก่อนที่จะส่งต่อออกไป การแปลงสัญญาณนี้มีเหตุผลสองประการคือ หนึ่งเพื่อที่จะสะท้อนการตัดสินใจบางอย่างเช่น การสั่งการให้ทำหรือ ไม่ทำและสอง เพื่อที่จะจัดระเบียบของ Input ที่เซลล์อื่นจะได้รับต่อไป

ในการแปลงสัญญาณ F(x) ให้เป็น 0 หรือ 1 นั้น เราจะแปลงผ่านฟังก์ชั่น (threshold function) หากค่าของ F(x) มีค่าเกินค่า ๆหนึ่งเช่น จะให้ค่าเป็น 1 แต่หากไม่เป็นเช่นนั้นก็จะให้ค่าเป็น 0

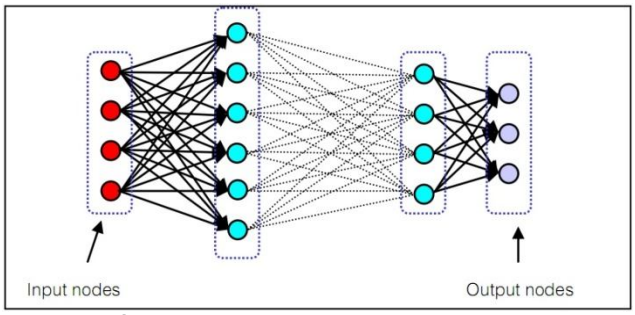
สิ่งสำคัญที่สุดในการทำงาน คือการทราบค่าWeight และ threshold สำหรับใช้ในการรู้จำของคอมพิวเตอร์ ซึ่งเป็นค่าที่ไม่แน่นอน แต่สามารถกำหนดให้คอมพิวเตอร์ปรับค่าเหล่านั้นได้โดยการสอนให้มันรู้จัก Pattern ของสิ่งที่มันต้องการรู้จำ ซึ่งเรียกว่า back propagation ซึ่งเป็นกระบวนการย้อนกลับของการรู้จำ การสร้างการเรียนสำหรับ neural network เพื่อให้มีความคิดเสมือนมนุษย์ มีสองวิธี คือ

1. Supervised learning การเรียนแบบมีการสอน คือ เป็นการเรียนแบบที่มีการตรวจคำตอบเพื่อให้วงจรข่ายปรับตัว ชุดข้อมูลที่ใช้สอนวงจรข่ายจะมีคำตอบไว้คอยตรวจดูว่าวงจรข่ายให้คำตอบที่ถูกหรือไม่ ถ้าตอบไม่ถูก วงจรข่ายจะปรับตัวเพื่อให้ได้คำตอบที่ดีขึ้น

2. Unsupervised learning การเรียนแบบมีการสอน คือ เป็นการเรียนแบบไม่มีการตรวจคำตอบว่าถูกหรือผิด วงจรข่ายจะจัดโครงสร้างด้วยตัวเองตามลักษณะของข้อมูล ผลลัพธ์ที่ได้ วงจรข่ายจะสามารถจัดข้อมูลได้

โครงข่ายประสาทเทียมสามารถแบ่งออกเป็น 4 แบบ คือ Feedforward network, Feedback network, Network layer และ Perceptrons

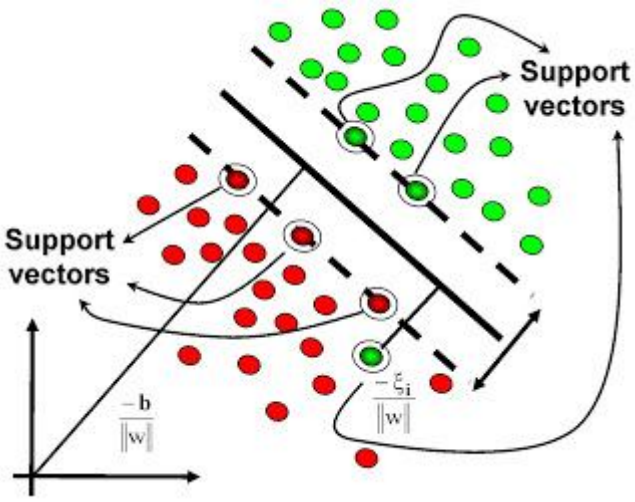
โดยข้อมูลจะถูกประมวลในแต่ละโหนดและจะถูกส่งต่อไปอีกโหนดในทิศทางเดียวกันจาก input nodes จนถึง output nodes โดยไม่มีการย้อนกลับของข้อมูล ดังภาพประกอบ 2-19



ภาพประกอบที่ 2-19 แสดงสถาปัตยกรรม Feedforward network

2.7 SVM (Support Vector Machine)

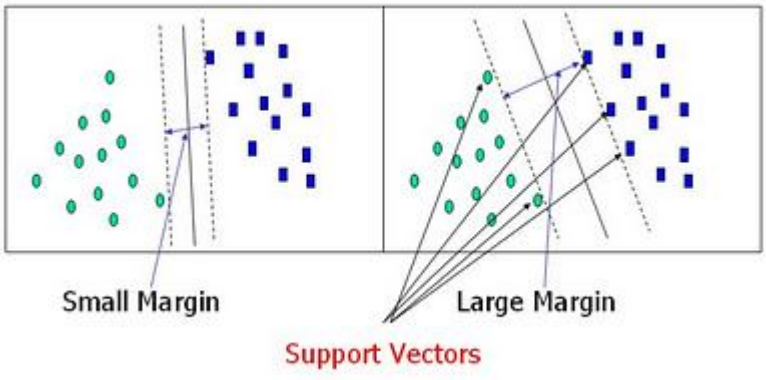
SVM เป็นอัลกอริทึมในการคัดแยกที่มีการนำมาใช้กันอย่างกว้างขวางในด้านการประมวลผลเป็นภาพดิจิตอล หลักการของ SVM คือการให้อินพุทที่ใช้ฝึกเป็นเวคเตอร์ในสเปซ N มิติ เช่นถ้าในกรณีของ 2 มิติ และ 3 มิติ จะเป็นจุดที่อยู่ในระนาบ xy และสเปซ xyz ตามลำดับ จากนั้นทำการสร้างไฮเปอร์เพลน (Hyperplane) ที่จะแยกกลุ่มของเวคเตอร์อินพุทออกเป็นประเภทต่างๆ ในกรณีที่เป็น 2 มิติ และ 3 มิติ ไฮเปอร์เพลน คือเส้นตรงและระนาบตามลำดับข้อเด่นของ SVM จะทำการเก็บแมพ (Map) เวคเตอร์ในสเปซอินพุทให้เข้าสู่ Feature Space โดยใช้ฟังก์ชั่นหรือเรียกว่าเคอร์นัล (kernel) ชนิดต่างๆ เช่น โพลีโนเมียล (Polynomial) เรเดียล (Radial) เป็นต้น ใน Feature Space ดังกล่าวเวคเตอร์อินพุท สามารถแยกประเภทได้โดยไฮเปอร์เพลน



ภาพประกอบที่ 2-20 ตัวอย่าง SVM ใน 2 มิติ

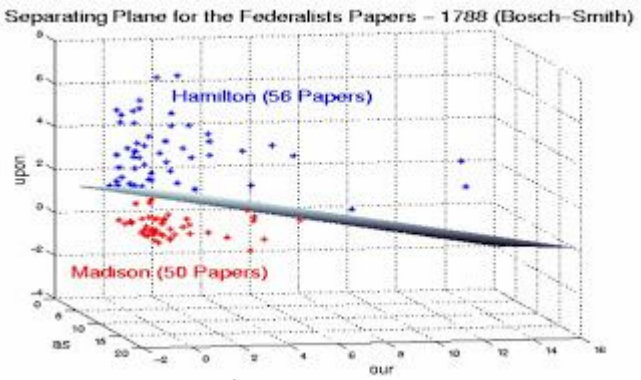
เครือข่ายปัญญาประดิษฐ์ กล่าวคือ SVM ที่ใช้ฟังก์ชั่นซิกมอยด์ในการแมพเทียบเท่ากับเครือข่ายปัญญาประดิษฐ์แบบ Feedforward ที่มี 2 ชั้นมีข้อแตกต่างจากเครือข่ายปัญญาประดิษฐ์ก็

คือ การแก้สมการหาค่าน้ำหนักใช้ในการแก้สมการ Quadratic ที่มีข้อบังคับเชิงเส้น (Linear Constrained) แทนที่จะเป็นการหาค่าต่ำสุด (minimization) อย่างในกรณีของเครือข่ายปัญญาประดิษฐ์



ภาพประกอบที่ 2-21 ตัวอย่างค่าMargin

สมมติว่าเราต้องการคัดแยกอินพุทออกเป็น 2 กลุ่ม โดยใช้ไฮเปอร์เพลน ที่เป็นเส้นตรง จะเห็นว่ามีเส้นตรงจำนวนมากที่สามารถคัดแยกได้ แต่เส้นตรงเส้นไหนที่ดีที่สุด (Optimal Line) ภาพประกอบที่ 2-21 แสดงตัวอย่างของ 2 เส้นตรง เราจะนิยาม Margin เป็นผลรวมระยะห่างของเส้นตรงที่เป็นไฮเปอร์เพลน (ภาพประกอบที่ 2-21) ถึงเส้นตรงที่ผ่านอินพุทที่ ใกล้ที่สุดและขนานกับไฮเปอร์เพลน ของทั้งสองกลุ่ม(ภาพประกอบที่ 2-21) ระยะดังกล่าวอาจมองเป็นเวคเตอร์และมีชื่อว่า ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) อัลกอริทึม SVM จะเลือกไฮเปอร์เพลนที่ให้ค่า Margin มีค่าสูงสุด ดังภาพประกอบที่ 2-21 กรณีของ 3 มิติ จะเป็นทำนองเดียวกันอัลกอริทึมSVM ใน 3 มิติ

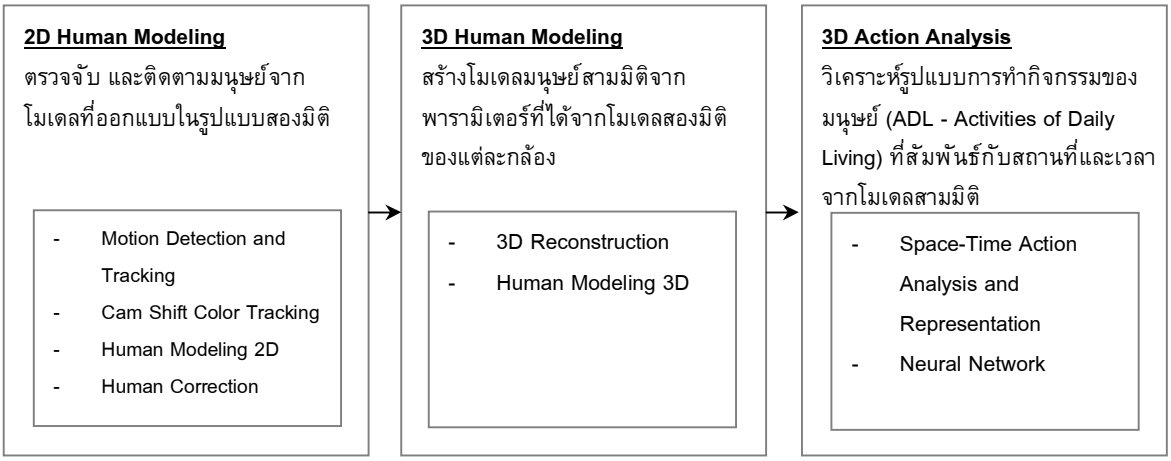


ภาพประกอบที่ 2-22 ตัวอย่างSVM ใน 3 มิติ

บทที่ 3
การออกแบบและพัฒนาระบบ

3.1 แนวคิดเบื้องต้นงานวิจัย

หลักการและทฤษฎีของระบบเบื้องต้น จะประกอบด้วย 3 ส่วนสำคัญคือ ขั้นตอนการวิเคราะห์หาโมเดลมนุษย์ในรูปแบบสองมิติ (2D Human Modeling) ขั้นตอนการวิเคราะห์หาโมเดลมนุษย์ในรูปแบบสามมิติจากโมเดลสองมิติ (3D Human Modeling) และการวิเคราะห์พฤติกรรมของมนุษย์สัมพันธ์กับเวลาและสถานที่(ADL Action Analysis) ดังไดอะแกรมต่อไปนี้



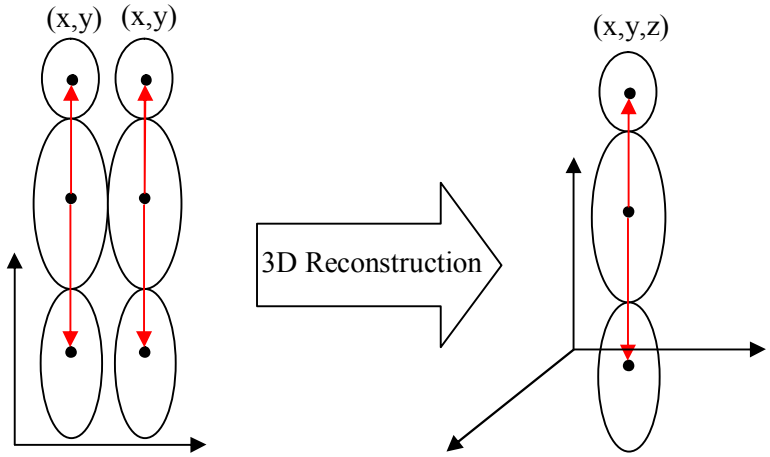
การทำงานวิจัยในแต่ละส่วนเป็นดังนี้

ส่วนที่ 1 การวิเคราะห์การกระทำพื้นฐาน(2D Human Modeling)

ซึ่งจำเป็นต้องการมีวิเคราะห์โดยทั่วไปดังนี้ การวิเคราะห์วัตถุเคลื่อนไหวและติดตาม ทำหน้าที่แยกแยะวัตถุที่กำลังเคลื่อนไหวภายในภาพวีดีโอ โดยจะทำการตรวจสอบความน่าจะเป็นมนุษย์ของวัตถุที่เคลื่อนไหวเบื้องต้นจากการนำภาพที่มีความลึกมาวิเคราะห์วัตถุที่เคลื่อนไหวถ้าอยู่ในระนาบเดียวกันแสดงว่ามีความเป็นไปได้ที่วัตถุนั้นจะเป็นมนุษย์สูงและเพิ่มความถูกต้องของวัตถุที่เราสนใจโดยการใช้ภาพที่มีความลึก และภาพที่ผ่านกระบวนการตรวจจับการเคลื่อนไหว (Motion Detection) มาทำรวมกันเพื่อให้ได้ซึ่งวัตถุที่มีความถูกต้องมากที่สุด พร้อมทั้งดึงลักษณะเฉพาะของแต่ละวัตถุออกมาโดยรับภาพของแต่ละเฟรมจากกล้องวีดีโอเข้ามาประมวลผลและทำการติดตามพฤติกรรมของวัตถุนั้นจากกล้องแต่ละตัว จากนั้นทำการสร้างโมเดลของมนุษย์จากวัตถุเคลื่อนไหวในรูปแบบสองมิติ (2D)

ส่วนที่ 2 การโมเดลกิจกรรมและรู้จำ(3D Human Modeling)

เป็นการนำพารามิเตอร์ที่ได้จากการวิเคราะห์ของโมเดลมนุษย์ในสองมิติ (2D) ของแต่ละกล้องเพื่อสร้างโมเดลมนุษย์ในรูปแบบสามมิติ (3D) [5] [6] ดังภาพประกอบที่ 3-1 เพื่อใช้สำหรับการวิเคราะห์หาท่าทางสำหรับส่วนถัดไป



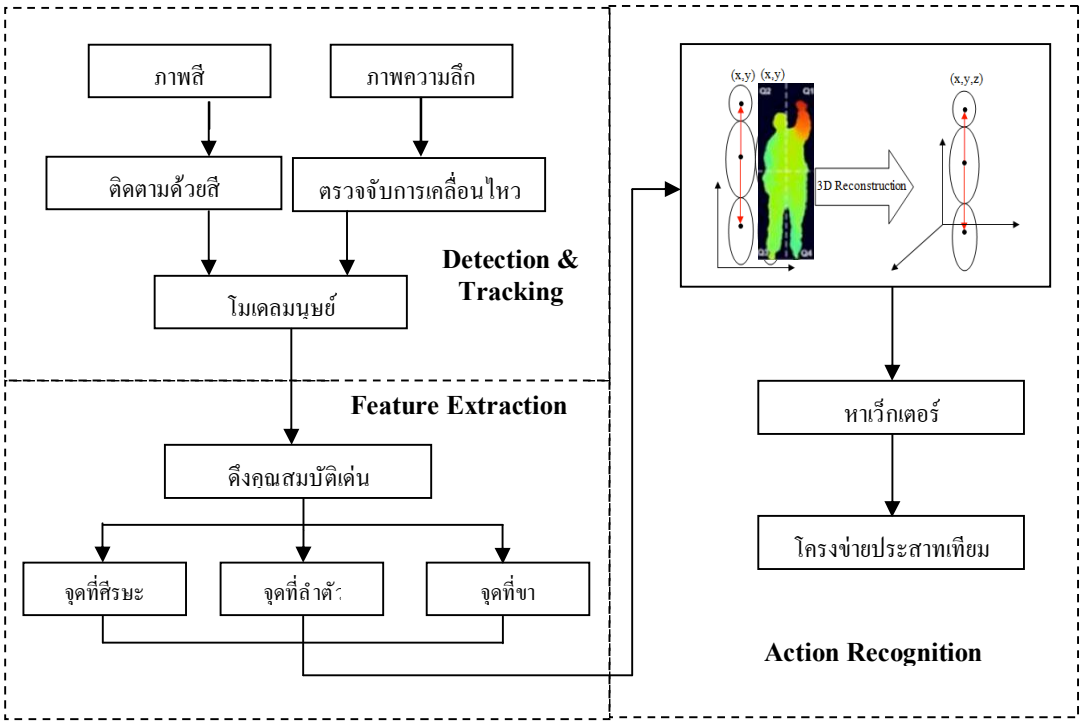
ภาพประกอบ 3-1 3D Reconstruction

ส่วนที่ 3 วิเคราะห์รูปแบบการทำกิจกรรมของมนุษย์ (3D Action Analysis)

เป็นการวิเคราะห์ความสัมพันธ์เชิงสถานที่และเวลาของมนุษย์ โดยจะต้องมีรูปแบบการนำเสนอข้อมูลที่เหมาะสมเป็นการวิเคราะห์ภาพวิดีโอจากโมเดลสามมิติ(3D) เพื่อให้ทราบว่าเกิดการกระทำพื้นฐานอะไรบ้าง อันได้แก่ การยื่น การเดิน การนั่ง การนอนและการก้ม เป็นต้น

3.2 การพัฒนาระบบ

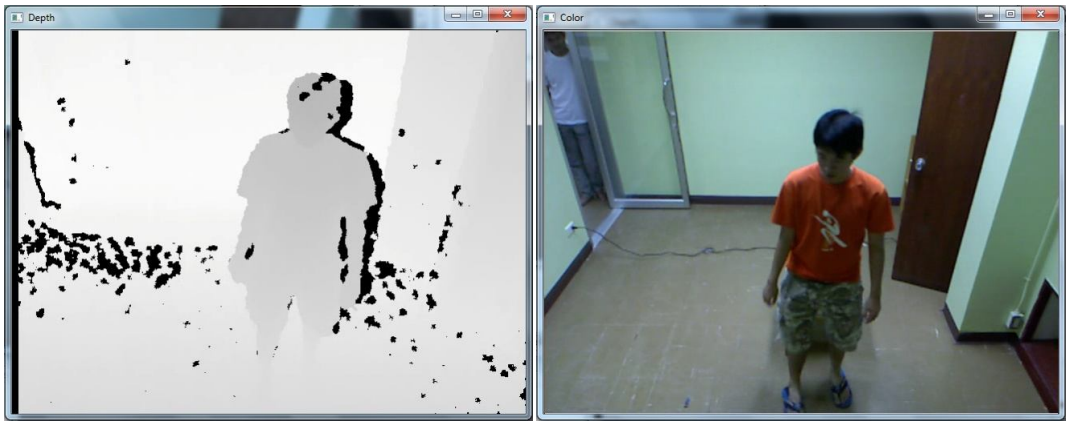
งานวิจัยนี้ได้ประยุกต์ใช้กระบวนตรวจจับการเคลื่อนไหวจากภาพความลึก และใช้สีจากภาพสีสำหรับการติดตามเพื่อใช้สำหรับการโมเดล โดยมนุษย์ภายในงานวิจัยนี้ได้ถูกอธิบายอยู่ในรูปแบบจำลองโครงสร้างมนุษย์อย่างง่ายโดยทำการลดความซับซ้อนขององค์ประกอบในร่างกายมนุษย์ให้เหลือเพียง3 องค์ประกอบโดยโมเดลที่ได้จะอยู่ในสองมิติ(2D) จากนั้นนำโมเดลที่ได้และภาพความลึกไปทำการสร้างใหม่ในรูปของสามมิติ (3D) จะทำให้ได้คุณลักษณะของมนุษย์ซึ่งจะถูกนำไปใช้ในกระบวนการรู้จำท่าทางมนุษย์โดยใช้กระบวนการของโครงข่ายประสาทเทียม (Neural Network) ผลจากกระบวนการนี้จะทำให้ทราบว่ามนุษย์ได้กระทำท่าทางใด ดังภาพประกอบที่ 3-2



ภาพประกอบ 3-2 ภาพรวมของระบบ

3.2.1 การเตรียมภาพสำหรับการวิเคราะห์

สำหรับงานวิจัยนี้ ภาพจากกล้องวิดีโอที่ใช้สำหรับการวิเคราะห์มีสองแบบ1) ภาพความลึกที่ให้ผลลัพธ์ออกมาในรูปแบบภาพที่เป็นDisparity ขนาด 8 บิต และ 2) ภาพสีที่ให้ผลลัพธ์ออกมาในรูปแบบ RGB ขนาด 8 บิท



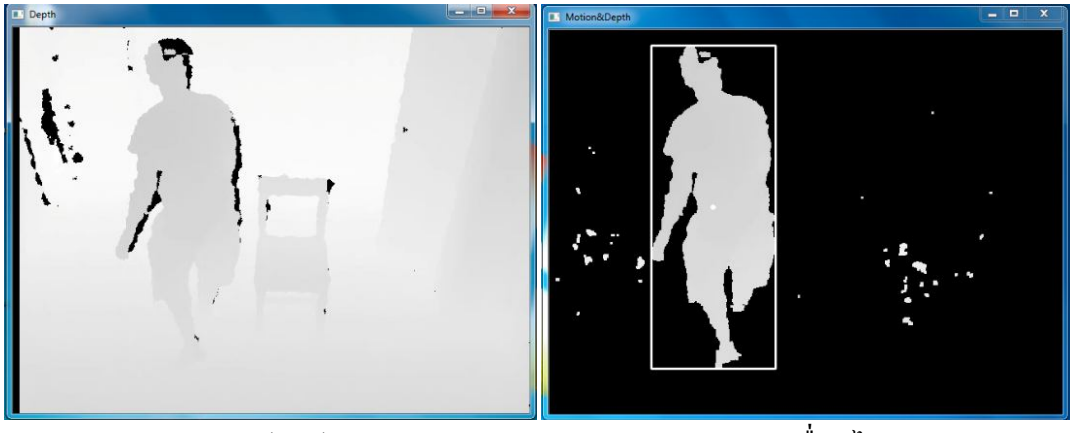
a) ภาพความลึก b) ภาพสี

ภาพประกอบ 3-3 ภาพเปรียบเทียบความลึกกับภาพสี

ระยะห่างของวัตถุจะแปรผันตามค่าสีที่เป็นภาพ Gray กล่าวคือถ้าวัตถุอยู่ไกลจากกล้องมาก ค่าของสีก็จะเข้าใกล้ 255 และวัตถุอยู่ใกล้ค่าสีก็จะเข้าใกล้0 แต่ไม่ใช่ 0 เนื่องจากค่า 0 สำหรับภาพความลึกคือไม่สามารถหาระยะได้ ดังภาพประกอบที่ 3-3a และภาพประกอบที่ 3-3b เป็นภาพสีที่ใช้สำหรับงานวิจัยนี้

3.2.2 การตรวจจับการเคลื่อนไหว

สำหรับการตรวจจับการเคลื่อนไหวเป็นกระบวนการที่มีความสำคัญอย่างมากเนื่องจากตั้งสมุติฐานว่าวัตถุที่เคลื่อนไหวในเฟรมวิดีโอเป็นมนุษย์ทั้งหมด ถ้าการตรวจจับการเคลื่อนไหวมีความแม่นยำสูงจะส่งผลให้กระบวนการอื่นมีความแม่นยำขึ้นอย่างมาก สำหรับการตรวจจับการเคลื่อนไหวในงานวิจัยนี้ใช้การตรวจจับการเคลื่อนไหวที่เรียกว่า Background Subtraction การสร้างพื้นหลังจากวิธีเกาส์เซียนหลายรูปแบบ (Mixture of Gaussian)



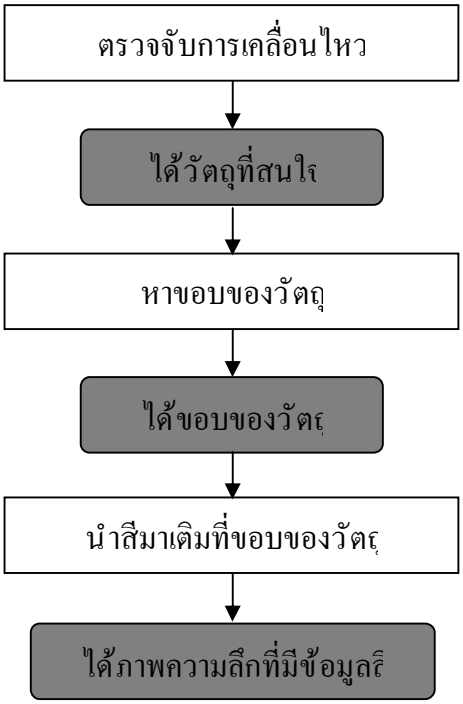
a) ภาพเฟรมปัจจุบัน b) ภาพการเคลื่อนไหว

ภาพประกอบ 3-4 ภาพตัวอย่างการตรวจจับการเคลื่อนไหวจากภาพความลึก

สำหรับการตรวจจับการเคลื่อนไหวโดยใช้ภาพสีในกรณีแสงเกิดการเปลี่ยนแปลงมากๆ ในทันทีจะทำให้เกิดความผิดพลาดได้พื้นหลังมาเป็นวัตถุที่สนใจ ข้อเสียอีกอย่างหนึ่งคือเมื่อสีของวัตถุที่เคลื่อนที่ตรงกับสีของพื้นหลังจะทำให้วัตถุนั้นกลายเป็นพื้นหลังซึ่งจากข้อผิดพลาดดังกล่าวทำให้ในงานวิจัยนี้ใช้ภาพความลึกเพียงอย่างเดียวสำหรับการตรวจจับการเคลื่อนไหวเนื่องจากแสงและสีไม่มีผลต่อภาพความลึกทำให้ไม่เกิดปัญหาดังกล่าว และโอกาสที่จะเกิดข้อผิดพลาดน้อยมาก เนื่องจากถ้าวัตถุที่มีระยะห่างเท่ากันจะไม่มาทับซ้อนกันจึงทำให้การใช้ภาพความลึกมีความแม่นยำสูงกว่าภาพสีมาก ดังภาพประกอบที่ 3-4

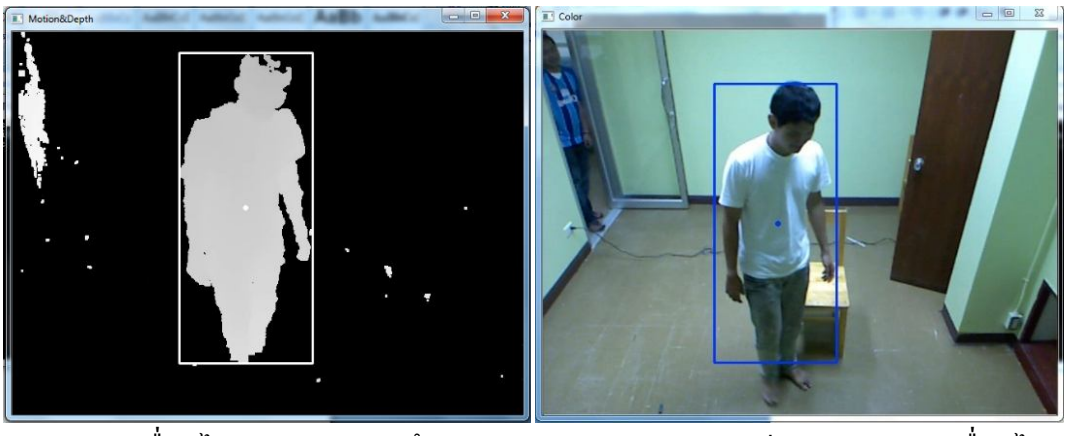
3.2.3 การผสมข้อมูลภาพความลึก และข้อมูลภาพสี

งานวิจัยนี้ ได้ทำการนำภาพที่มีข้อมูลของความลึก(Depth Image) และภาพที่มีข้อมูลของสี (Color Image) มาทำการผสมผสานกัน เพื่อเพิ่มความแม่นยำสำหรับการติดตามพิกัดที่สนใจซึ่งในที่นี้คือ ศีรษะ ลำตัว และขา อันนำไปสู่การโมเดลมนุษย์อย่างง่าย ดังภาพประกอบที่ 3-5 เริ่มจากการตรวจจับการเคลื่อนไหว โดยกระบวนการลบพื้นหลังเพื่อให้ได้วัตถุที่สนใจจากนั้นหาขอบของวัตถุโดยจากกระบวนการกัดกร่อน (Erosion) และนำมาเทียบสีจากภาพสีไปยังขอบของวัตถุ



ภาพประกอบ 3-6 กระบวนการโดยรวมของการผสมข้อมูลสี และความลึก

สำหรับการเทียบพิกัดจุดจากสองกล้องที่ตั้งขนานกันถ้าเทียบกันจุดต่อจุดจะทำให้เกิดความผิดพลาดของการเหลื่อมล้ำของพิกัดจึงทำให้ต้องมีการปรับให้อยู่ในมาตรฐานเดียวกัน (Calibration) สำหรับการเทียบพิกัดที่ตรงกัน ดังภาพประกอบที่ 3-6 เนื่องจากกล้องที่ใช้เป็นกล้องที่ใช้กันแพร่หลายในท้องตลาดจึงได้มีงานวิจัยสำหรับการปรับให้อยู่ในมาตรฐานเดียวกัน (Calibration) มากมาย ดังนั้นกระบวนการปรับให้อยู่ในมาตรฐานเดียวกัน (Calibration) ของกล้องสำหรับการเทียบพิกัดจึงอ้างอิงจากงานวิจัยอื่น สำหรับในงานวิจัยนี้ที่มุ่งเน้นไปที่การวิเคราะห์ท่าทางเป็นหลัก



a) ภาพการเคลื่อนไหวจากภาพความลึก b) ภาพจากการเทียบจุดจากภาพเคลื่อนไหว

ภาพประกอบ 3-7 ภาพตัวอย่างแสดงการเทียบพิกัดจากภาพความลึกไปภาพสี

สำหรับขั้นตอนแรกคือการบิดเบือนภาพสี และภาพความลึกโดยใช้ค่าสัมประสิทธิ์การบิดเบือนจากพารามิเตอร์ภายในของกล้องทั้งสอง(Intrinsic Parameter) แต่ละพิกเซลของภาพความลึกสามารถฉายไปยังพื้นที่สามมิติจากสมการต่อไปนี้

สมการสำหรับการฉายภาพความลึกไปยังพิกัดโลก

$$P3D.x = \left(x_d - cx_d\right) * \left(\frac{depth\left(x_d, y_d\right)}{fx_d}\right) \quad (3.1)$$

$$P3D.y = \left(y_d - cy_d\right) * \left(\frac{depth\left(x_d, y_d\right)}{fy_d}\right) \quad (3.2)$$

$$P3D.z = depth\left(x_d, y_d\right) \quad (3.3)$$

โดยค่าที่ใช้สำหรับการคำนวณเป็นค่าที่ได้มาจากพารามิเตอร์ภายใน(Intrinsic Parameter) ของกล้องความลึกจากการปรับให้อยู่ในมาตรฐานเดียวกัน(Calibration) โดยใช้ตารางหมากลุก

fx_d = 5.9421434211923247e+02

fy_d = 5.9104053696870778e+02

cx_d = 3.3930780975300314e+02

cy_d = 2.4273913761751615e+02

ปรับภาพความลึกสำหรับการเทียบไปยังภาพสีโดย การRotation และ Translation

$$\begin{bmatrix} P3D'.x \\ P3D'.y \\ P3D'.z \end{bmatrix} = R. \begin{bmatrix} P3D.x \\ P3D.y \\ P3D.z \end{bmatrix} + T \quad (3.4)$$

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

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

สมการสำหรับเทียบพิกัดจากภาพความลึกไปยังภาพสี

$$P2D_{rgb}.x = \frac{P3D'.x * fx_{rgb}}{P3D'.z} + cx_{rgb} \quad (3.5)$$

$$P2D_{rgb}.y = \frac{P3D'.y * fy_{rgb}}{P3D'.z} + cy_{rgb} \quad (3.6)$$

โดยค่าที่ใช้สำหรับการคำนวณเป็นค่าที่ได้มาจากพารามิเตอร์ภายใน (Intrinsic Parameter) ของกล้องสีจากการคาลิเบตโดยใช้ตารางหมากลุก

fx_rgb = 5.2921508098293293e+02

fy_rgb = 5.2556393630057437e+02

cx_rgb = 3.2894272028759258e+02

cy_rgb = 2.6748068171871557e+02

เนื่องจากการเทียบพิกัดจุดไปจากภาพความลึกไปยังภาพสีมีการคำนวณที่ซับซ้อน ถ้าหากทำการเทียบทั้งบริเวณที่เป็นวัตถุที่สนใจการประมวลผลทำได้ล่าช้าจึงทำการดึงสีจากภาพสีเฉพาะบริเวณที่สนใจซึ่งในที่นี้คือสีบริเวณขอบของวัตถุ ดังสมการที่(3.7) โดยอ้างอิงมาจากขอบของวัตถุที่เคลื่อนไหวในภาพความลึกที่ได้จากกระบวนการตรวจจับการเคลื่อนไหว (Motion Detection) ผลลัพธ์ดังภาพประกอบที่3-7

สมการสำหรับการรวมภาพความลึกและภาพสี

$$I_{DC} = E(I_D) \cup (I_C \cap (O_D - E(O_D))) \quad (3.7)$$

I_D คือ ภาพความลึก

I_C คือ ภาพสี

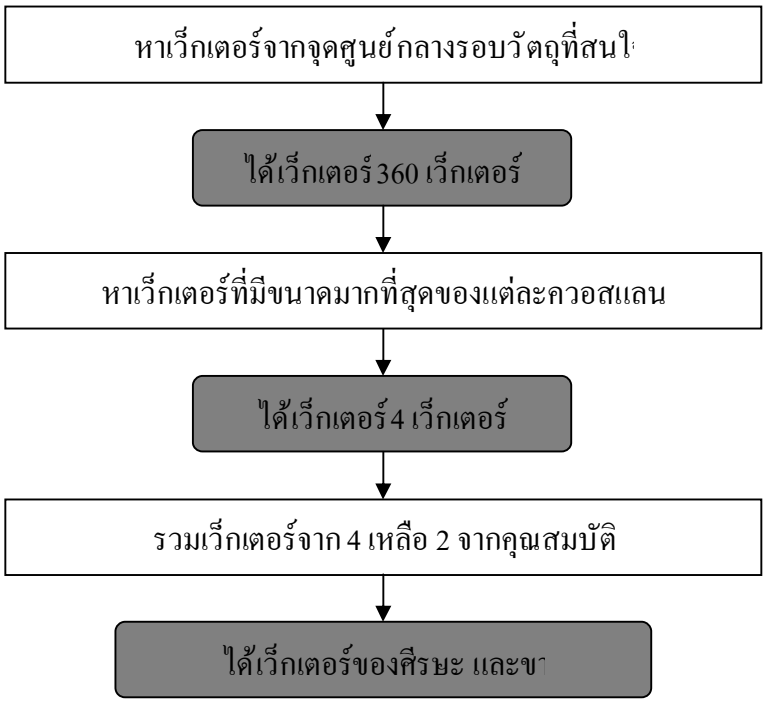
E คือ โอเปอร์เรชั่นของการกัดกร่อน (Erosion)

a) ตัวอย่างภาพภาพสี b) ผลลัพธ์การดึงสีจากภาพสีมายังภาพความลึก

ภาพประกอบ 3-7 ภาพตัวอย่างแสดงการรวมภาพความลึกและภาพสี

3.2.4 โมเดลโครงสร้างมนุษย์อย่างง่าย

เนื่องจากการวิเคราะห์ท่าทางของงานวิจัยนี้ ใช้โมเดลมนุษย์อย่างง่ายที่ประกอบด้วย3 ส่วน คือ ศีรษะ ลำตัว และขา โดยให้ลำตัวเป็นจุดศูนย์กลาง จากนั้นหาเว็กเตอร์จากจุดบนศีรษะ และขา กระบวนการโดยรวมของการสร้างโมเดลมนุษย์จะแบ่งออกเป็น 3 ขั้นตอน ดังภาพประกอบที่3-8



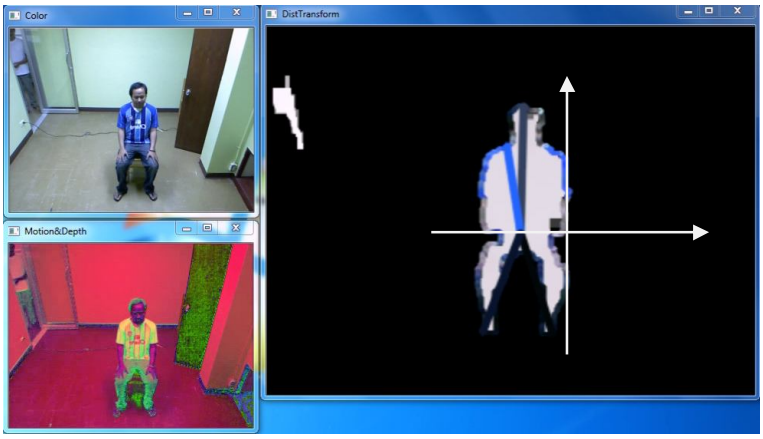
ภาพประกอบ 3-8 กระบวนการโดยรวมของการสร้างโมเดลมนุษย์อย่างง่าย

<u>ขั้นตอนที่1</u> หาเว็กเตอร์จากจุดบนขอบของวัตถุที่สนใจที่ได้จากกระบวนก่อนหน้าเทียบไปยังจุดศูนย์กลางที่คำนวณจากสมการ (3.8) จำนวน 360 เว็กเตอร์รอบจุดศูนย์กลางมวล

$$\bar{x}, \bar{y} = \frac{\sum_0^N x_i}{N}, \frac{\sum_0^N y_i}{N} \quad (3.8)$$

<u>ขั้นตอนที่2</u> จากสมมุติฐานที่ว่าจุดจากขอบที่อยู่ไกลจากจุดศูนย์กลางมีโอกาสเป็น ศีรษะ และขา มากกว่าจุดที่อยู่ใกล้ จึงทำการแบ่งวัตถุที่สนใจออกเป็น 4 ส่วน โดยอ้างอิงจากจุดศูนย์กลางมวลที่คำนวณจากสมการ (3.8) จากนั้นทำการหาเว็กเตอร์ที่มีขนาดมากที่สุดในแต่ละส่วน ผลลัพธ์จะได้เว็กเตอร์ที่สนในจำนวน 4 เว็กเตอร์ ที่มีโอกาสเป็นศีรษะ และขาของมนุษย์ ดังภาพประกอบที่3-9

$$\Delta_O = \{\Delta_i | i = 1 \dots 4\} \quad (3.9)$$



ภาพประกอบ 3-9 ภาพตัวอย่างการหาโมเดลโครงสร้างมนุษย์

<u>ขั้นตอนที่3</u> สำหรับในขั้นตอนนี้เป็นการรวมเว็กเตอร์จาก 4 ให้เหลือ 2 เว็กเตอร์ เริ่มจากการจับคู่เว็กเตอร์โดยอ้างอิงจากมุมที่กระทำกันซึ่งเว็กเตอร์ที่มีค่าความต่างของมุมน้อยที่สุดในแต่ละคู่จะจับคู่กับ และอีกสองเว็กเตอร์ที่เหลือก็จับคู่กันเอง จากนั้นใช้ความน่าจะเป็นเข้ามาช่วยสำหรับการรวมเว็กเตอร์ โดยเว็กเตอร์ที่มีขนาด(D_v) มากจะมีความน่าจะเป็นที่จะเป็นศีรษะ หรือ ขา มากกว่าเว็กเตอร์ที่มีขนาดน้อยกว่า ดังสมการที่3.11

$$D_{v_i} = \sqrt{(\Delta_{x_i} - \bar{x})^2 + (\Delta_{y_i} - \bar{y})^2} \ | \ i = 1 \dots 4 \quad (3.10)$$

$$\omega_{v_i} = \left[\sum_{j=1}^{2} \left(\frac{D_{v_1}}{D_{v_j}} \right)^{\frac{2}{m-1}} \right]^{-1} \ | \ i = 1 \dots 4 \quad (3.11)$$

เนื่องจากใช้ขนาดของเว็กเตอร์เพียงอย่างเดียวมีความน่าเชื่อถือไม่เพียงพอสำหรับการรวมเว็กเตอร์จึงเพิ่มความน่าเชื่อถือโดยการนำสีบริเวณที่เว็กเตอร์ชี้ไปมาทำการหาความน่าจะเป็นสำหรับการรวมเว็กเตอร์ โดยเทียบจากสีอ้างอิงหากสีบริเวณที่เว็กเตอร์ชี้ไปมีความใกล้เคียงกับสีอ้างอิง (D_c) มากความน่าจะเป็นก็จะมากแต่หากมีความต่างกันความน่าจะเป็นก็จะน้อยลงตามลำดับ ดังสมการที่ 3.13

$$D_{c_i} = \sqrt{(\Delta_{c_i} - Cr_t)^2} \mid i = 1 \dots 4 \tag{3.12}$$

$$\omega_{c_i} = \left[\sum_{j=1}^{2} \left(\frac{D_{c_1}}{D_{c_j}} \right)^{\frac{2}{m-1}} \right]^{-1} \mid i = 1 \dots 4 \tag{3.13}$$

สมการที่ 3.14 ใช้สำหรับการปรับปรุงค่าสีอ้างอิงโดยอัตราการปรับปรุงขึ้นอยู่กับค่า α_c

$$Cr_{t+1} = \alpha_c Cr_t - (1 - \alpha_c)(\omega_1 C_1 + \omega_2 C_2) \tag{3.14}$$

จากสมการที่ 3.15 ใช้สำหรับการปรับค่าความน่าจะเป็นของการเลือกเชื่อระหว่างความน่าจะเป็นที่ได้จากขนาดของเว็กเตอร์ ω_v กับ ความน่าจะเป็นที่ได้จากสีเทียบกับสีอ้างอิงω_c โดยขึ้นอยู่กับค่า α_v

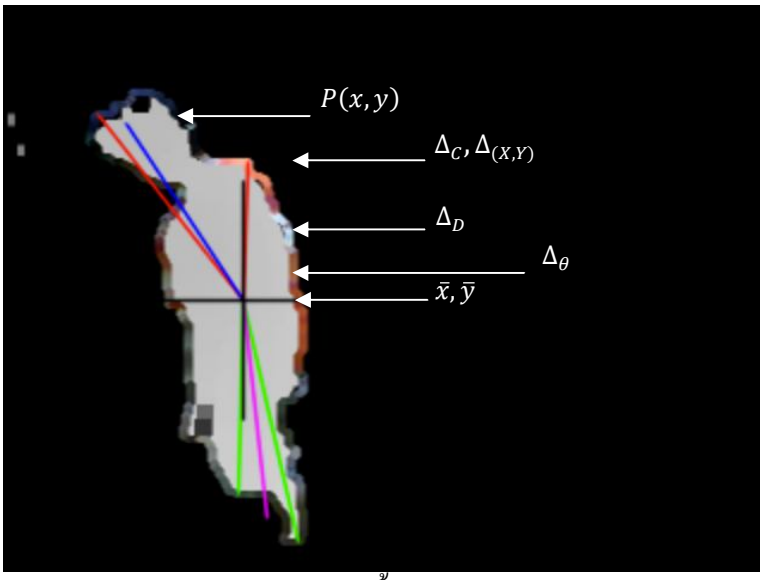
$$\omega_1 = \alpha_v \omega_v + (1 - \alpha_v)\omega_c \tag{3.15}$$

$$\omega_2 = 1 - \omega_1 \tag{3.16}$$

นำความน่าจะเป็นที่ได้จากสมการ 3.15 และ 3.16 มาทำการคูณกับพิกัดจุดที่เว็กเตอร์ชี้ไปทั้งสองเว็กเตอร์ โดยเป็นไปตามสมการที่ 3.17 ผลลัพธ์จะได้พิกัดใหม่จากสองพิกัดที่จับคู่กัน จากนั้นนำจุดที่ได้มาทำการหาเว็กเตอร์ไปยังจุดศูนย์กลางมวลผลลัพธ์จะเหลือเว็กเตอร์ที่ชี้ไปยัง ศีรษะ และขา ของมนุษย์เพื่อเข้าสู่กระบวนการต่อไป

$$P(x, y) = \omega_1 P_1(\Delta_{x_1}, \Delta_{y_1}) + \omega_2 P_2(\Delta_{x_2}, \Delta_{y_2}) \tag{3.17}$$

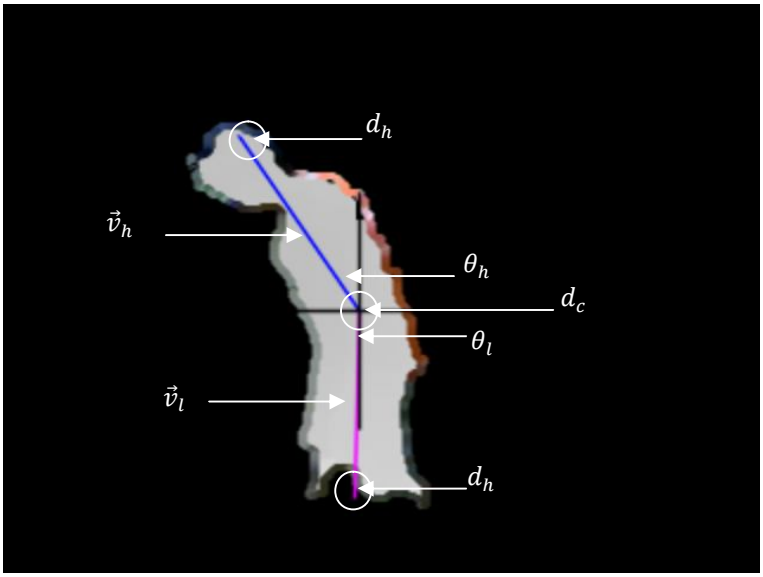
ผลลัพธ์ในขั้นตอนนี้ทำให้ได้โมเดลมนุษย์อย่างง่ายโดยการนำเอาเว็กเตอร์สีแดงรวมเป็นเว็กเตอร์สีน้ำเงิน และเว็กเตอร์สีเขียวรวมเป็นเว็กเตอร์สีชมพู ดังภาพประกอบที่ 3-10 โดยเว็กเตอร์สีน้ำเงินอธิบายความสัมพันธ์ระหว่างลำตัว และศีรษะ ส่วนเว็กเตอร์สีชมพูอธิบายความสัมพันธ์ระหว่างลำตัว และขา ซึ่งสามารถนำไปดึงคุณลักษณะบางอย่างที่สามารถนำไปสู่การวิเคราะห์ท่าของมนุษย์



ภาพประกอบ 3-10 โมเดลมนุษย์อย่างง่ายเส้นสีน้ำเงิน และสีชมพูคือผลลัพธ์จากการรวมเส้นสีแดงและสีเขียวตามลำดับ

3.2.5 การดึงลักษณะเด่น

สำหรับการดึงลักษณะเด่นจากโมเดลมนุษย์มีความสำคัญอย่างมากที่จะช่วยให้ระบบสามารถวิเคราะห์ท่าทางได้อย่างถูกต้อง และแม่นยำ สำหรับการคำนวณพารามิเตอร์อันนำไปเข้าสู่กระบวนรู้จำเพื่อวิเคราะห์ท่าทางสามารถคำนวณได้จากแบบจำลองมนุษย์อย่างง่ายดังภาพประกอบ 3-11 โดยพิจารณาความสัมพันธ์ของแต่ละองค์ประกอบภายในที่สัมพันธ์กันของโครงสร้างมนุษย์ ประกอบด้วย 2 เว็กเตอร์ คือ เว็กเตอร์สีน้ำเงิน$\vec{v}_h$ อธิบายความสัมพันธ์ระหว่างลำตัว และศีรษะ ส่วนเว็กเตอร์สีชมพู $\vec{v}_l$ อธิบายความสัมพันธ์ระหว่างลำตัว และขา



ภาพประกอบ 3-11 พารามิเตอร์ $[\theta_h, \theta_l]$ และ $[D_h, D_l]$ สำหรับการนำไปวิเคราะห์ท่าทาง

สำหรับพารามิเตอร์ที่สำคัญสำหรับการนำไปวิเคราะห์ท่าทางแบ่งเป็น 2 ประเภท คือ พารามิเตอร์ที่ได้มากจากค่ามุม $[\theta_h, \theta_l]$ ของเว็กเตอร์ที่สัมพันธ์กันของลำตัวกับศีรษะ และลำตัวกับขา ดังสมการที่ (3.18) สามารถวิเคราะห์ท่าทางได้แม่นยำในมุมมองด้านข้าง เนื่องจากค่ามุมของความสัมพันธ์ดังกล่าวสามารถอธิบายได้ถึงท่าทางที่เกิดขึ้นในขณะนั้น

$$[\theta_h, \theta_l] = \cos^{-1}(\frac{P_x}{R}) \tag{3.18}$$

และพารามิเตอร์ที่ได้มาจากค่าความต่างของความลึก $[D_h, D_l]$ ระหว่างลำตัวกับศีรษะ และลำตัวกับขาดังสมการที่ (3.19) สามารถวิเคราะห์ท่าทางได้แม่นยำในมุมมองด้านหน้าเนื่องจากค่าความสัมพันธ์ของความแตกต่างของความลึกดังกล่าวสามารถอธิบายได้ถึงท่าทางที่เกิดขึ้นในขณะนั้น

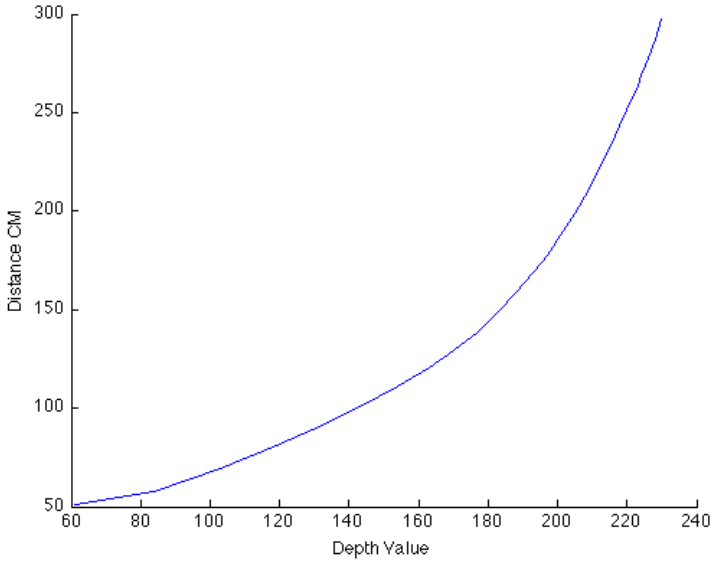
$$[D_h, D_l] = \sqrt{(d_{(h,l)} - d_c)^2} \tag{3.19}$$

เนื่องจากมุมกล้องมีผลต่อความคลาดเคลื่อนของพารามิเตอร์ที่ได้มาจากค่าความต่างของความลึก $[D_h, D_l]$ จึงจำเป็นต้องใช้การหมุน(Rotation Matrix) เพื่อปรับมุมกล้องให้เสมือนกับว่าขนานกับวัตถุมากที่สุดเพื่อปรับความลึกให้อยู่ในเกณฑ์เดียวกันโดยไม่ขึ้นกับมุมกล้องซึ่งจะปรับเฉพาะแกน y เพราะว่ามีผลต่อมุมกล้องโดยตรงดังสมการที่ 3.20

$$\begin{matrix} x \\ y \\ [\mathrm{D_h}, \mathrm{D_l}] \end{matrix} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \\ [\mathrm{D_h}, \mathrm{D_l}] \end{bmatrix} \quad (3.20)$$

เนื่องจากข้อมูลความลึกที่ได้มาจากกล้องนั้นมีขนาด 8 บิท โดยสามารถแบ่งเป็นค่าความลึกได้เป็น 256 ระดับ แต่เนื่องจากระดับที่ได้ออกมาไม่มีหน่วยที่สามารถบอกระยะจริงได้จึงต้องทำการปรับให้อยู่ในมาตรฐานเดียวกัน (Calibration) ของระดับค่าความลึกให้เป็นเซนติเมตร โดยระดับความลึกที่ได้จากกล้องเมื่อนำไปวาดกราฟปรากฏว่ากราฟไม่เป็นเส้นตรงดังภาพประกอบที่ 3-12

$$7.229247947975168e^{-7}x^4 - 3.434034079954699e^{-4}x^3 + 0.062936247222164x^2 \\ -4.506183418260807x + 1.596646613741274e^2$$

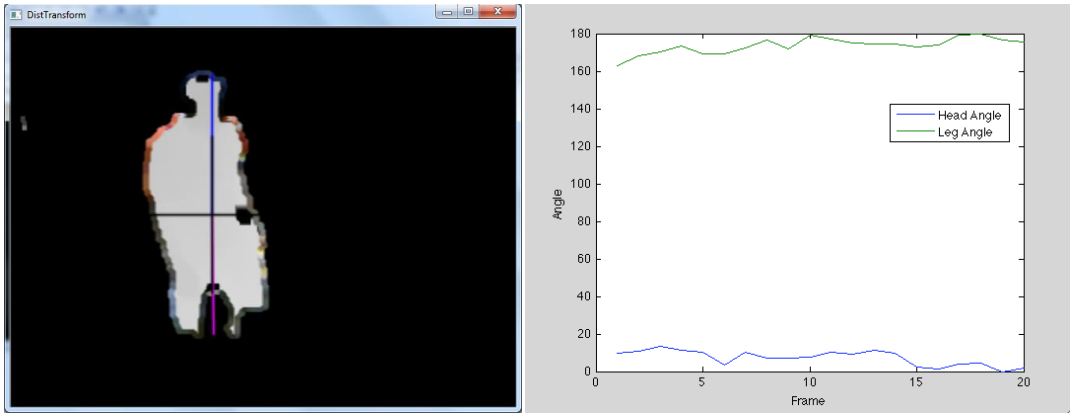


ภาพประกอบ 3-12 กราฟความสัมพันธ์ระหว่างระยะทางจริงกับค่าความลึกจากกล้อง

แต่เนื่องจากกราฟไม่เป็นเส้นตรงจึงต้องใช้สมการถดถอยแบบไม่เป็นเส้นตรง(Polynomial Regression) เข้ามาช่วยสำหรับการหาสมการความสัมพันธ์ระหว่างระยะจริงกับค่าความลึกจากกล้องโดยการวัดระยะจริงสำหรับการนำไปหาสมการถดถอยแบบไม่เป็นเส้นตรงเริ่มวัดตั้งแต่ 0 เซนติเมตรเนื่องจากกล้องเริ่มวัดระดับความลึกตั้งแต่ 50 เซตติเมตรขึ้นไป และวัดไปจนถึง 300 เซนติเมตร โดยแต่ละระดับห่างกัน 10 เซนติเมตร รวมทั้งหมด 27 ระดับ

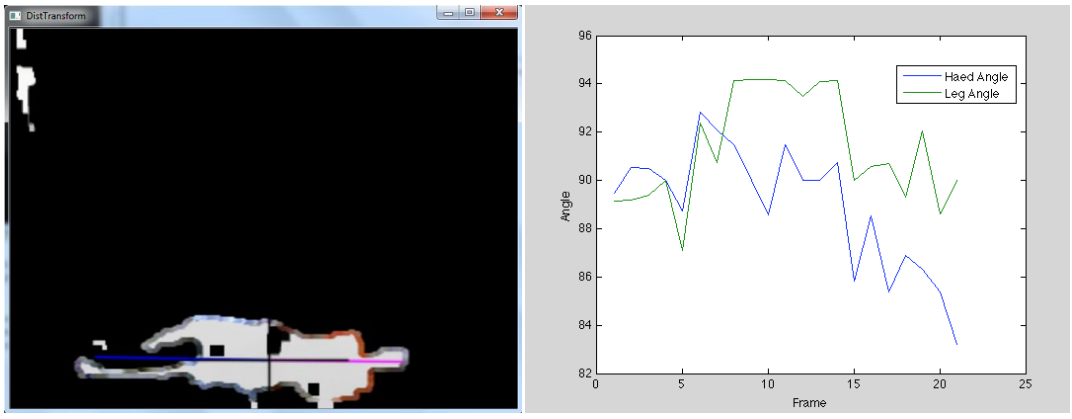
ตัวอย่างพารามิเตอร์ของการวิเคราะห์ท่าทาง

a) ดังภาพประกอบที่ 3-13 เป็นตัวอย่างพารามิเตอร์ที่บ่งบอกว่าเป็นการยืนหรือการเดิน โดยพิจารณาจากมุมของลำตัวถึงศีรษะใกล้เคียง0 องศา และมุมของลำตัวถึงขาใกล้เคียง180 องศา



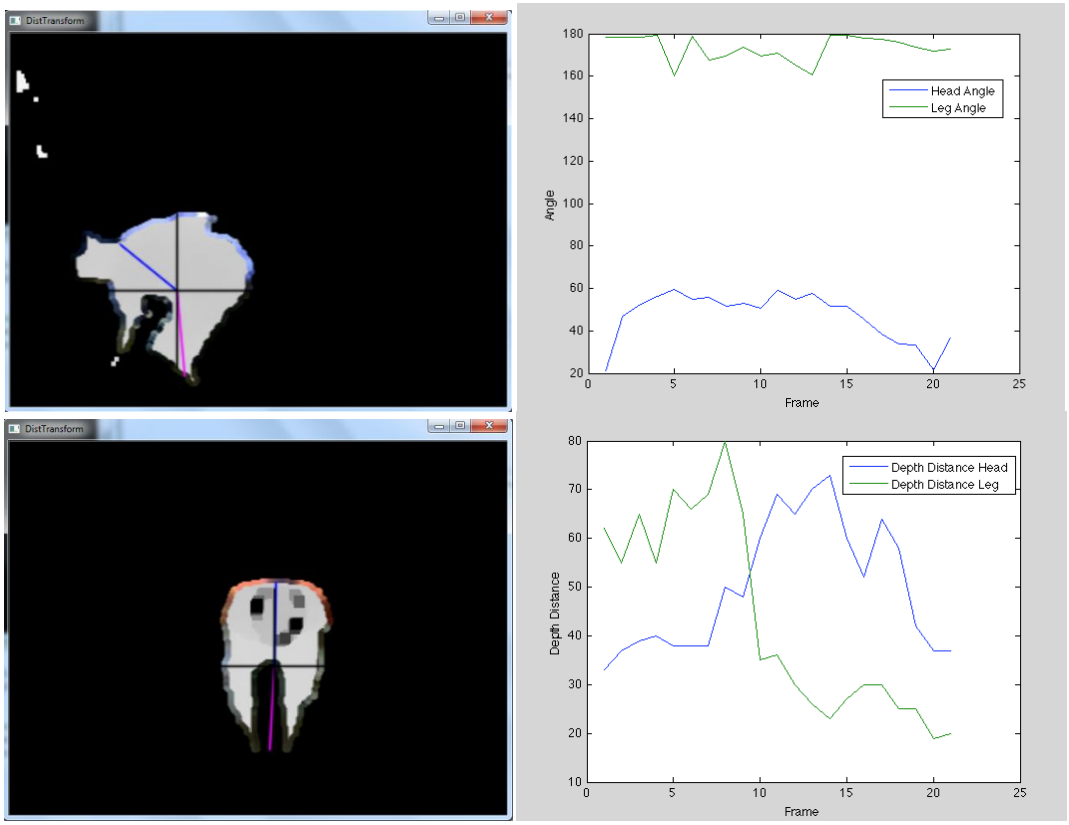
ภาพประกอบ 3-13 ตัวอย่างพารามิเตอร์ของการยืน และการเดิน

b) ดังภาพประกอบที่ 3-14 เป็นตัวอย่างพารามิเตอร์ที่บ่งบอกว่าเป็นการนอน โดยพิจารณาจากมุมของลำตัวถึงศีรษะใกล้เคียง90 องศา และมุมของลำตัวถึงขาใกล้เคียง90 องศา



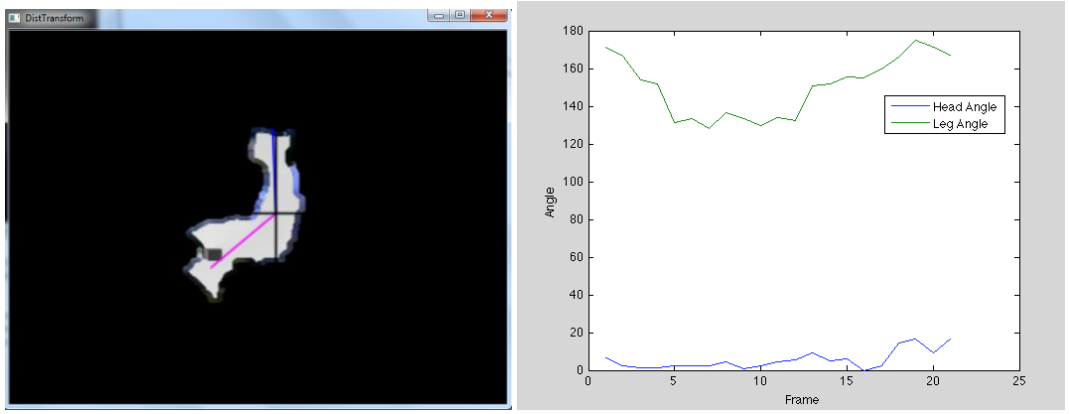
ภาพประกอบ 3-14 ตัวอย่างพารามิเตอร์ของการนอน

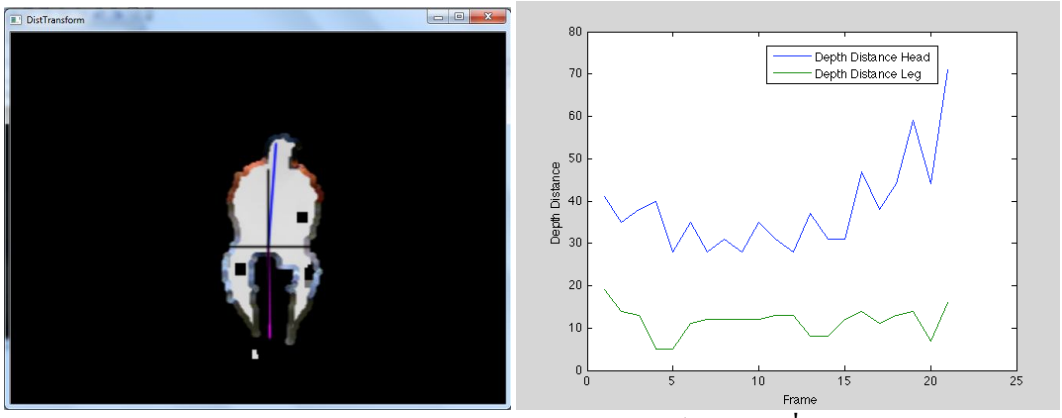
c) ภาพประกอบที่ 3-15 เป็นตัวอย่างพารามิเตอร์ที่บ่งบอกว่าเป็นการก้ม โดยพิจารณาจากมุมของลำตัวถึงศีรษะใกล้เคียง45 - 60 องศา และมุมของลำตัวถึงขาใกล้เคียง180 องศา และอีกพารามิเตอร์หนึ่งที่สามารถบ่งบอกว่าเป็นการก้มคือค่าความต่างของความลึกระหว่างลำตัวถึงศีรษะใกล้เคียง70 ซ.ม. และค่าความต่างของความลึกระหว่างลำตัวถึงขาใกล้เคียง 20 ซ.ม.



ภาพประกอบ 3-15 ตัวอย่างพารามิเตอร์ของการก้ม

d) ภาพประกอบที่ 3-16 เป็นตัวอย่างพารามิเตอร์ที่บ่งบอกว่าเป็นการก้ม โดยพิจารณาจากมุมของลำตัวถึงศีรษะใกล้เคียง120 - 140 องศา และมุมของลำตัวถึงขาใกล้เคียง0 องศา และอีกพารามิเตอร์หนึ่งที่สามารถบ่งบอกว่าเป็นการก้มคือค่าความต่างของความลึกระหว่างลำตัวถึงศีรษะใกล้เคียง40 ซ.ม. และค่าความต่างของความลึกระหว่างลำตัวถึงขาใกล้เคียง 20 ซ.ม.





ภาพประกอบ 3-16 ตัวอย่างพารามิเตอร์ของการนั่ง

e) จากตัวอย่างเป็นการนำพารามิเตอร์มาวิเคราะห์ความเป็นไปได้ที่มีโอกาสเกิดเป็นท่าทางจากสมมุติฐานที่ตั้งไว้ โดยมุมมองที่ใช้เป็นมุมมอง 60 องศา จากนั้นนำพารามิเตอร์ที่ได้เข้าสู่กระบวนการรู้จำท่าทางขั้นตอนต่อไป

3.2.6 การรู้จำท่าทางมนุษย์

สำหรับการรู้จำท่าทางมนุษย์ในงานวิจัยนี้ ใช้อัลกอริทึมที่มีความสามารถในการเรียนรู้จากการสอน ในงานวิจัยนี้ใช้สองเทคนิคสำหรับการรู้จำ คือ เทคนิคโครงข่ายปราสาทเทียม(Neural network) โดยอัลกอริทึมสำหรับการเทรนเป็นแบบการรู้จำแบบย้อนกลับ (back propagation algorithm) และใช้ซิกมอยด์ฟังก์ชั่นสำหรับเป็นแอคติเวทฟังก์ชั่น ซึ่งมีอินพุตทั้งหมด 4 อินพุต คือ ค่ามุม $[\theta_h, \theta_l]$ ของเว็กเตอร์ที่สัมพันธ์กันของลำตัวกับศีรษะ และลำตัวกับขา และค่าความต่างของความลึก $[D_h, D_l]$ ระหว่างลำตัวกับศีรษะ และลำตัวกับขา โดยมีเลเยอร์ซ่อน 1 เลเยอร์ และเอาท์พุทมีทั้งหมด4 เอาท์พุท นั่นก็คือท่าทาง การยืนหรือการเดิน, การนั่ง, การนอน และการก้ม และอีกเทคนิคที่ใช้นำมาเปรียบเทียบคือ SVM (Support Vector Machine) ซึ่งมีทั้งหมด4 อินพุต และ 4 เอาท์พุท เช่นเดียวกันกับเทคนิคโครงข่ายปราสาทเทียม

3.2.7 สรุป

สำหรับในงานวิจัยนี้ ได้ออกแบบและพัฒนาระบบวิเคราะห์ท่าทางมนุษย์ โดยรับภาพจากมุมมองด้านข้าง โดยใช้ค่าความลึกเข้ามาช่วยแก้ปัญหาในส่วนของข้อผิดพลาดที่เกิดจากมุมมองในกรณีที่เกิดท่าทางในมุมมองหน้าตรงกับกล้องซึ่งจะทำให้เกิดข้อผิดพลาดในส่วนของการดึงลักษณะเด่น แต่เมื่อนำความลึกเข้ามาช่วยทำให้สามารถดึงลักษณะเด่นอันนำไปสู่การวิเคราะห์ท่าทางได้อย่างแม่นยำมากยิ่งขึ้น โดยคุณลักษณะเด่นได้มาจากโมเดลมนุษย์อย่างง่ายที่มีความสัมพันธ์กันของลำตัวกับศีรษะ และลำตัวกับขาจากนั้นเข้าสู่กระบวนการวิเคราะห์ท่าทางโดยใช้กระบวนการ

สำหรับการรู้จำจากการสอน Neural Network และ SVM เพื่อเปรียบเทียบผลลัพธ์สำหรับรู้จำท่าทาง ได้แก่ การยืนหรือการเดิน การนั่ง การก้ม และการนอนโดยไม่ขึ้นอยู่กับมุมมองของกล้อง

บทที่ 4

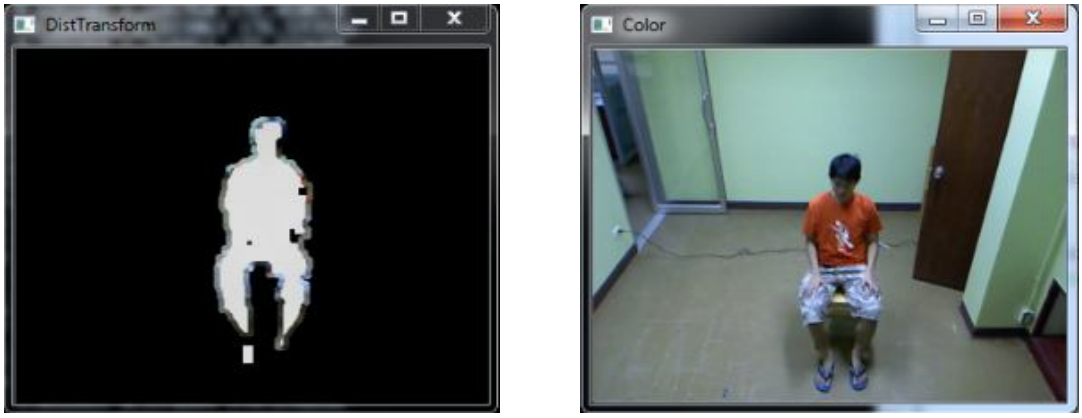
ผลการทดลอง

ผลการทดสอบวิเคราะห์ท่าทางมนุษย์ ซึ่งประกอบด้วยท่าทางพื้นฐานทั้ง 5 ท่าทาง ได้แก่ การยืน การเดิน การนั่ง การนอน และ การก้ม โดยใช้กล้องจำนวน2 ตัว คือกล้องที่เป็นภาพสี ความละเอียดขนาด 640x480 จุด และกล้องที่เป็นภาพความลึก ขนาด 640x480 จุด โดยระดับของความลึกที่ได้อยู่ที่256 ระดับ โดยการทดลองจะเปลี่ยนมุมกล้อง 3 มุมมองคือ 30, 45, และ60 องศา โดยใช้วัตถุทดลองเป็นบุคคลทดสอบสี่คนที่ใส่เสื้อและกางเกงสีแตกต่างกันซึ่งแต่ละบุคคลจะแสดงท่าทางครบทั้ง5 ท่าทางคือ การยืน การเดิน การนั่ง การนอน และการก้ม โดยระบบจะรองรับการวิเคราะห์ท่าทางเพียงคนเดียว

4.1 การทดสอบการรวมภาพความลึกและภาพสี

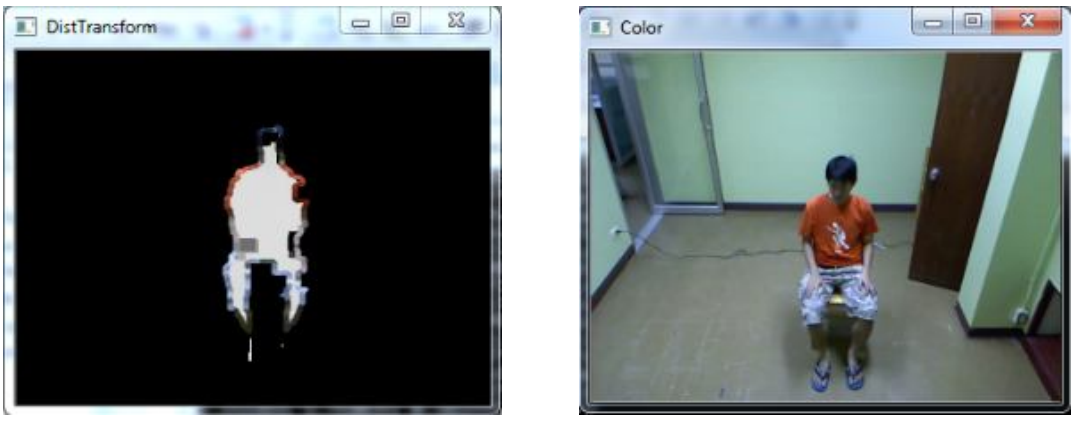
สำหรับการทดสอบการรวมภาพสีและภาพความลึกได้มีการใช้กระบวนการที่เรียกว่า“การกัดกร่อน” ซึ่งใช้ทรัพยากรเครื่องมากถ้าหากใช้ในปริมาณรอบของการทำซ้ำมากเกินไป จะทำให้ระบบไม่สามารถประมวลผลได้ในระยะเวลาที่ทันท่วงทีจึงได้มีการทดสอบการรวมภาพสีและภาพความลึกเพื่อหาจำนวนรอบที่เหมาะสมที่สุดของการใช้กระบวนการกัดกร่อน

a) ภาพประกอบที่ 4-1 แสดงให้เห็นผลลัพธ์ที่มีความผิดพลาดสูงเนื่องจากภาพสี และภาพความลึกมีระยะซูมที่แตกต่างกัน กล่าวคือวัตถุในภาพความลึกมีขนาดใหญ่กว่าวัตถุในภาพสีทำให้การกัดกร่อนแค่ 1 รอบยังไม่ทำให้ถึงสีของวัตถุที่สนใจซึ่งทำให้เกิดความผิดพลาดสูง



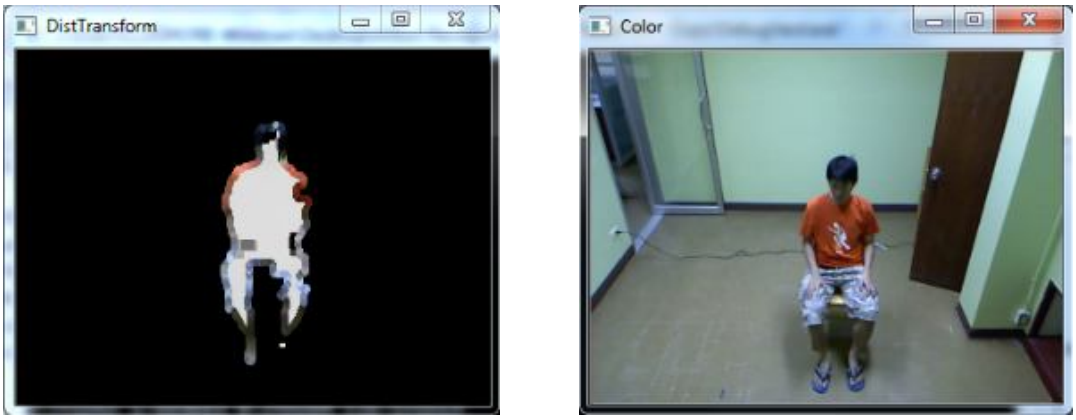
ภาพประกอบ 4-1 ตัวอย่างผลลัพธ์ของการทำกระบวนการกัดกร่อน 1 รอบ

b) ภาพประกอบที่ 4-2 ผลลัพธ์ของการกัดกร่อนจำนวน2 รอบได้ผลลัพธ์ของสีที่ครบถ้วน เนื่องจากสามารถดึงค่าสีจากบริเวณขอบของวัตถุได้อย่างแม่นยำโดยที่การประมวลผลยังสามารถประมวลผลได้เร็ว



ภาพประกอบ 4-2 ตัวอย่างผลลัพธ์ของการทำกระบวนการกัดกร่อน2 รอบ

c) ภาพประกอบที่ 4-3 ผลลัพธ์ของการกัดกร่อนจำนวน รอบได้ผลลัพธ์ของสีที่ครบถ้วน ดีกว่าการใช้กระบวนการกัดกร่อน2 รอบเพียงเล็กน้อยแต่แต่ใช้ทรัพยากรการประมวลผลที่มากกว่าพอสมควร



ภาพประกอบ 4-3 ตัวอย่างผลลัพธ์ของการทำกระบวนการกัดกร่อน3 รอบ

จากการทดสอบการรวมภาพความลึกและภาพสีการใช้กระบวนการกัดกร่อนเพียง1 รอบไม่สามารถนำไปใช้ในกระบวนการติดตามสำหรับการวิเคราะห์ได้ เพราะมีความผิดพลาดสูงในการติดตามสีของวัตถุที่สนใจ และการใช้กระบวนการกัดกร่อนจำนวน3 รอบผลลัพธ์ออกมาเป็นที่น่าพอใจแต่เมื่อเปรียบเทียบกับการกัดกร่อนจำนวน2 รอบผลลัพธ์มีความแตกต่างกันน้อยมากงานวิจัยนี้จึงเลือกการใช้การกัดกร่อนจำนวน2 รอบเนื่องจากผลลัพธ์ออกมามีความถูกต้องที่เหมาะสมกับทรัพยากรที่ใช้สำหรับการประมวลผล

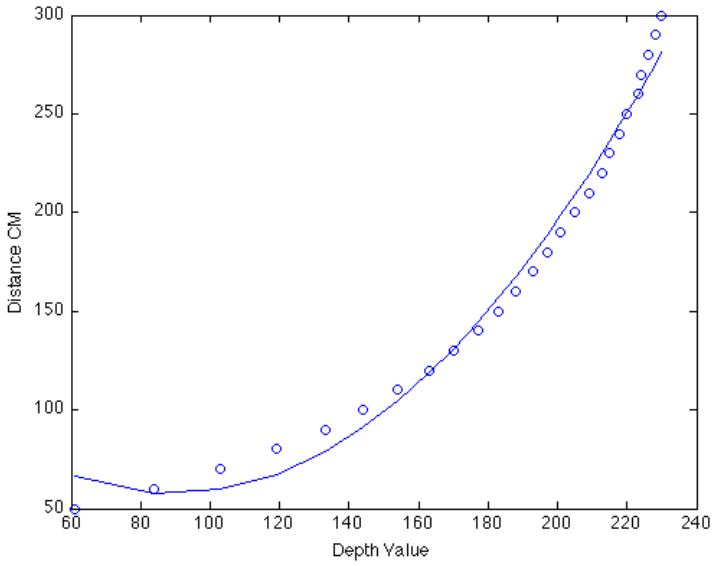
4.2 การทดสอบความสัมพันธ์ระหว่างระยะทางจริงกับค่าจากกล้อง

สำหรับการทดสอบหาความสัมพันธ์ระหว่างระยะจริงกับค่าจากกล้องที่เหมาะสมระหว่างความถูกต้องกับทรัพยากรในการประมวลผลมีความสำคัญเนื่องจากสมการถดถอยแบบไม่เป็นเส้นตรงอยู่ในรูปของเลขยกกำลังซึ่งถ้าหากยกกำลังมากๆจะทำให้การประมวลผลช้าซึ่งจะส่งผลประทบต่อการประมวลผลทั้งระบบ

a) ภาพประกอบที่ 4-4 ผลลัพธ์ของสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง2 กราฟวงกลมคือความสัมพันธ์ของระยะจริงกับค่าจากกล้องที่ได้จากการวัด และกราฟเส้นคือกราฟที่ได้จากการหาสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง2 จากผลลัพธ์ดังกล่าวอธิบายได้ว่าสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง 2 มีความคลาดเคลื่อนจากกราฟที่ได้จากกว่าวัดจริงอยู่มากพอสมควร

b)

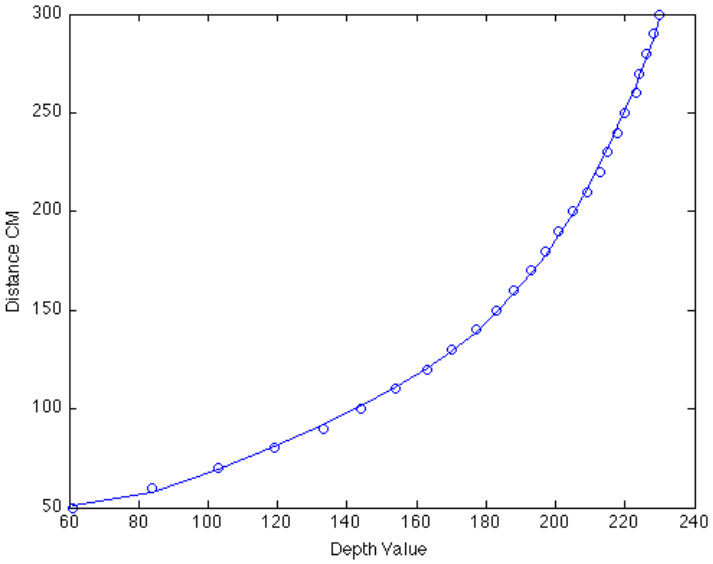
$$0.011318764249569x^2 - 2.023866610995049x + 1.480636900519317e^2$$



ภาพประกอบ 4-4 ตัวอย่างผลลัพธ์ของสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง2

c) ภาพประกอบที่ 4-5 ผลลัพธ์ของสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง2 กราฟวงกลมคือความสัมพันธ์ของระยะจริงกับค่าจากกล้องที่ได้จากการวัด และกราฟเส้นคือกราฟที่ได้จากการหาสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง4 จากผลลัพธ์ดังกล่าวอธิบายได้ว่าสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง 4 มีความถูกต้องมากกว่าสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง2 โดยที่ผลลัพธ์มีความคลาดเคลื่อนกับค่าที่วัดจริงน้อย

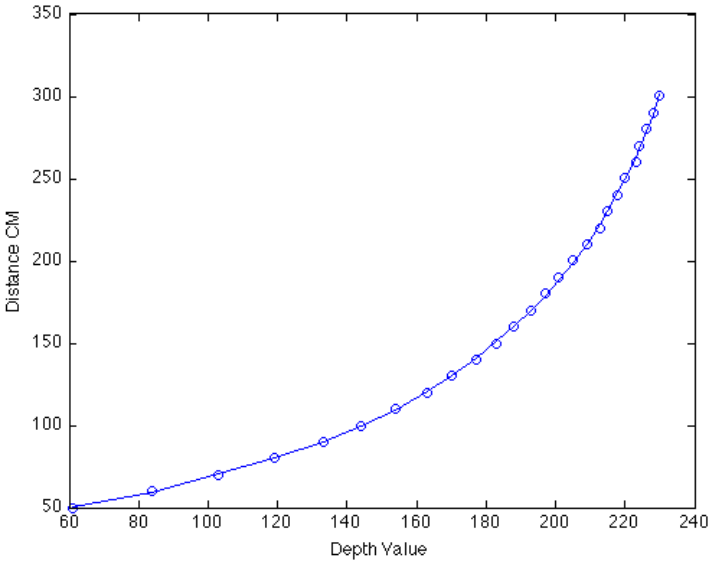
$$7.229247947975168e^{-7}x^4 - 3.434034079954699e^{-4}x^3 + 0.062936247222164x^2 \\ -4.506183418260807x + 1.596646613741274e^2$$



ภาพประกอบ 4-5 ตัวอย่างผลลัพธ์ของสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง4

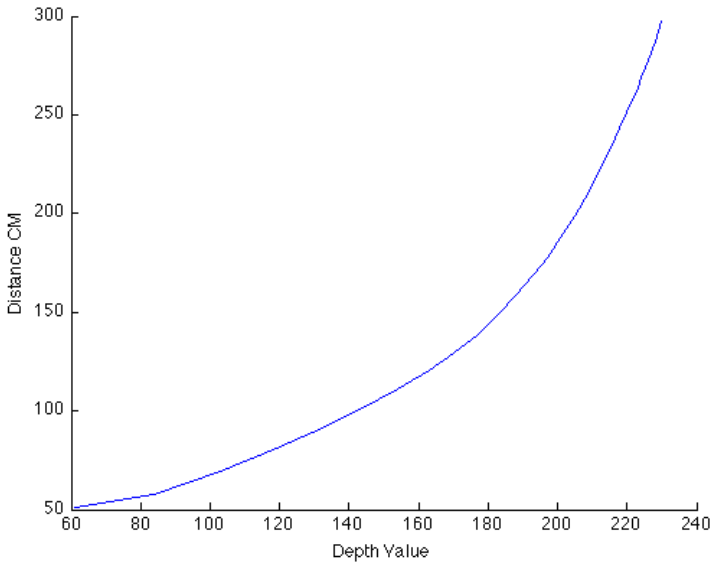
d) ภาพประกอบที่ 4-6 ผลลัพธ์ของสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง6 กราฟวงกลมคือความสัมพันธ์ของระยะจริงกับค่าจากกล้องที่ได้จากการวัด และกราฟเส้นคือกราฟที่ได้จากการหาสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง6 จากผลลัพธ์ดังกล่าวอธิบายได้ว่าสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง 6 มีความถูกต้องมากกว่าสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง2 และกำลัง4 โดยที่ผลลัพธ์มีความคลาดเคลื่อนกับค่าที่วัดจริงน้อยมาก

$$1.251217487491828e^{-10}x^6 - 1.037037939785234e^{-7}x^5 + 3.501481072103725e^{-5}x^4 - 0.006100639363181x^3 + 0.577595387872622x^2 - 27.634268166355310x + 5.675994039761157e^2$$



ภาพประกอบ 4-6 ตัวอย่างผลลัพธ์ของสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง 6

จากผลการทดลองดังกล่าวสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง 2 มีความคลาดเคลื่อนมากจึงไม่สามารถใช้งานได้ในงานวิจัยนี้ และสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง 6 มีความถูกต้องมากกว่าสมการถดถอยแบบไม่เป็นเส้นตรงกำลัง 4 ก็จริงแต่เนื่องจากเป็นถึงกำลัง 6 แต่มีความถูกต้องเมื่อเทียบกับสมการกำลัง 4 ผลลัพธ์ถูกต้องมากกว่าเพียงเล็กน้อยเมื่อเทียบกับทรัพยากรการประมวลที่เพิ่มขึ้นงานวิจัยนี้จึงเลือกใช้สมการถดถอยแบบไม่เป็นเส้นตรงกำลัง 4 เนื่องจากมีความเหมาะสมในเรื่องของความถูกต้องเมื่อเทียบกับทรัพยากรสำหรับการประมวลผล



ภาพประกอบ 4-7 กราฟความสัมพันธ์ระหว่างระยะทางจริงกับค่าความลึกจากกล้อง

ภาพประกอบที่ 4-7 กราฟแสดงความสัมพันธ์ระหว่างระยะทางจริงกับค่าความลึกจากกล้อง แสดงให้เห็นถึงความละเอียดของระยะทางจริงจะลดน้อยลงเมื่ออยู่ไกลจากกล้องซึ่งจะส่งผลให้ความถูกต้องของพารามิเตอร์สำหรับการวิเคราะห์ท่าทางมีความคลาดเคลื่อนเล็กน้อย

4.3 การทดสอบความคลาดเคลื่อนของมุมกล้อง

การทดสอบมุมกล้องในงานวิจัยนี้ใช้มุมกล้องที่60 องศา เป็นหลักสำหรับการนำพารามิเตอร์ไปทำการสอนระบบสำหรับการวิเคราะห์ท่าทาง เนื่องจากมุมกล้องที่ 60 องศา เป็นมุมกล้องที่สามารถมองเห็นสภาวะแวดล้อมที่กว้างโดยไม่ทำให้พารามิเตอร์คลาดเคลื่อน ดังนั้นในงานวิจัยนี้จึงใช้มุมกล้องที่ 60 องศาเป็นหลัก สำหรับกรณีมุมกล้องที่ไม่ใช่ 60 องศา จึงจำเป็นต้องทำการ Rotation เพื่อปรับพารามิเตอร์ให้เสมือนอยู่ในมุมมองที่ 60 องศา เนื่องจากถ้าหากปรับมุมกล้องเพียงอย่างเดียวมุมมองการมองเห็นพื้นที่จะเปลี่ยนไปจึงต้องปรับความสูงของกล้องเพื่อให้กล้องสามารถมองเห็นพื้นที่ที่มุมมองที่ 60 องศา สามารถมองเห็นจึงเทียบจากพื้นที่ที่มุม 60 องศา และสามารถวิคราะห์โดยที่มุมกล้องปรับเป็นไปตามตาราง 4-1 ข้อมูลที่นำมาทดสอบจำนวน 400 ชุด ข้อมูลของแต่ละมุมกล้องผลการทดลองคือค่าเฉลี่ยความผิดพลาดของแต่ละพารามิเตอร์ความต่างของความลึกระหว่างลำตัวและศีรษะ กับลำตัวและขา(D_h, D_l)

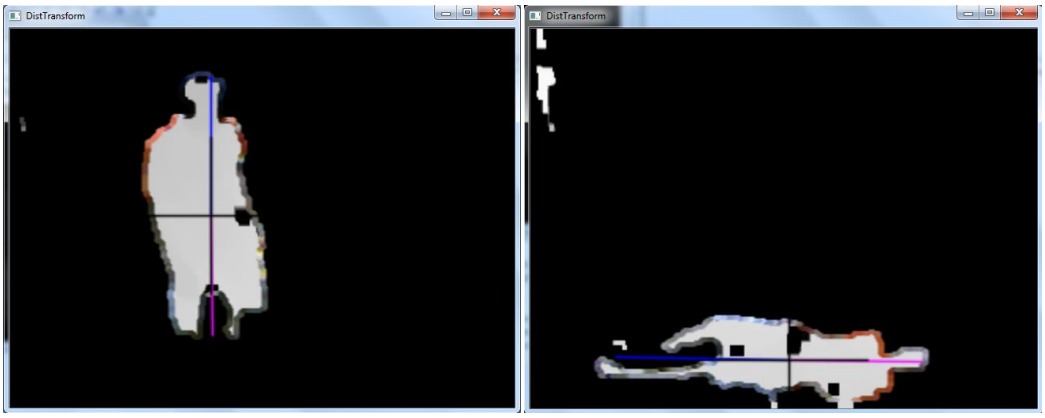
ตารางที่ 4-1 ผลการทดลองการเปรียบเทียบของแต่ละมุมกล้อง

มุมกล้อง(องศา)	ความคลาดเคลื่อน	ระยะเริ่มการทำงาน
0	15.40%	70cm
45	13.93%	50cm
75	10.44%	30cm
90	98.8%	0

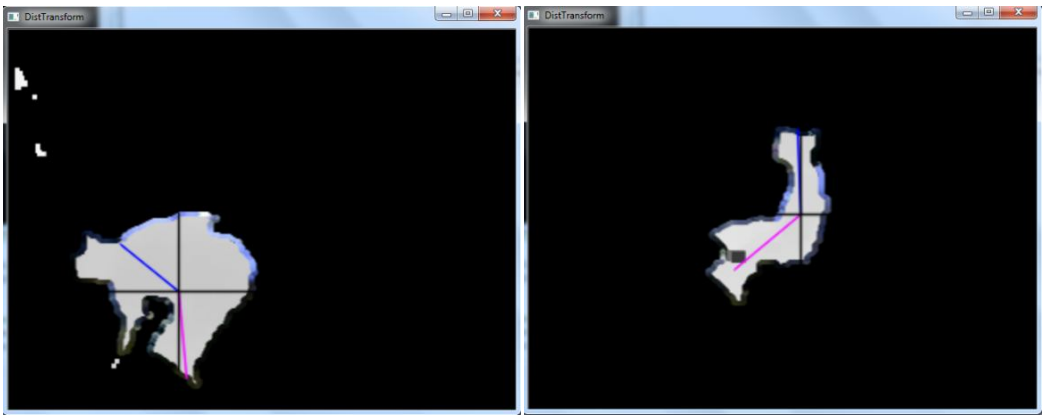
จากผลการทดลองมุมกล้องที่องศาต่ำกว่า60 องศา ความคลาดเคลื่อนของพารามิเตอร์มีค่าน้อยแต่ระยะการทำงานจะลดลงเนื่องจากมุมกล้องที่ต่ำกว่า60 องศา การวางกล้องจะวางในตำแหน่งที่ต่ำกว่าเพื่อสามารถทำให้มองเห็นตัวบุคคลได้เต็มตัว และที่มุมกล้องที่มากกว่า60 องศา ระยะการทำงานครอบคลุมมากกว่าแต่เนื่องจากงานวิจัยนี้ใช้มุมมองด้านข้างของตัวบุคคลเพื่อใช้สำหรับการวิเคราะห์ท่าทางจึงทำให้มุมมองที่มีอาศาที่สูงกว่า60 องศา การมองเห็นด้านข้างของตัวบุคคลจึงมีความคลาดเคลื่อนสูงตามไปด้วย

4.4 การทดสอบการวิเคราะห์ท่าทางมนุษย์

สำหรับการวิเคราะห์ท่าทางการยืน การเดิน การนั่ง การนอน และการก้ม สามารถวิเคราะห์ได้จากค่ามุมของเว็กเตอร์ $[\theta_h, \theta_l]$ เส้นสีน้ำเงิน และเส้นสีชมพู ซึ่งอธิบายถึงความสัมพันธ์กันของลำตัวกับศีรษะ และลำตัวกับขาแสดงให้เห็นถึงองค์ประกอบสำคัญแต่ละส่วนของแบบจำลองของพารามิเตอร์ภายใน คือ $\vec{v}_h, \vec{v}_l$ ตามลำดับ โดยใช้ผลจากกระบวนการติดตามการเปลี่ยนแปลงในแต่ละเฟรม ดังภาพประกอบที่ 4-8



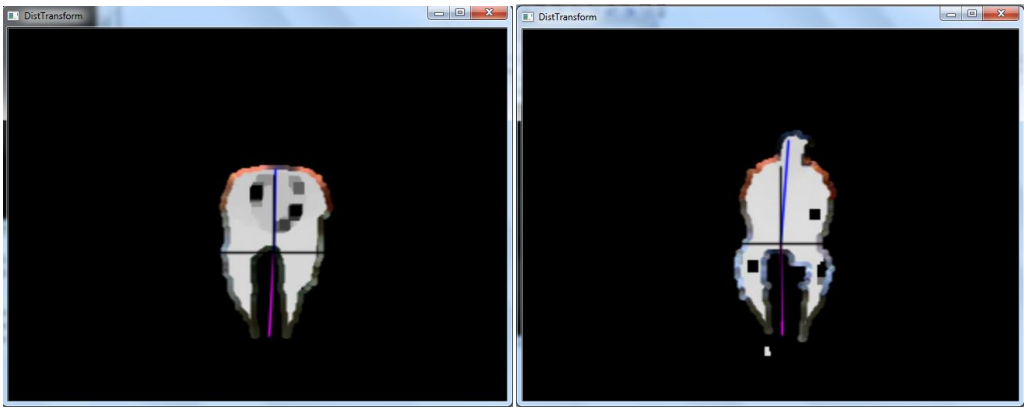
(a) (b)



(c) (d)

ภาพประกอบ 4-8 ตัวอย่างท่าทางที่พิจารณาท่าทางโดยใช้ค่ามุมของเว็กเตอร์(a) การเดิน (b) การนอน (c) การก้ม (d) การนั่ง

แต่เนื่องจากมีบางกรณีที่ค่ามุมของเว็กเตอร์ $[\theta_h, \theta_l]$ เพียงอย่างเดียวไม่สามารถวิเคราะห์ได้ในมุมมองด้านหน้าจึงใช้พารามิเตอร์ที่ได้มาจากค่าความต่างของความลึก $[D_h, D_l]$ ระหว่างลำตัวกับศีรษะ และลำตัวกับขา มาช่วยสำหรับการวิเคราะห์ท่าทางได้แม่นยำในมุมมองด้านหน้า เนื่องจากค่าความสัมพันธ์ของความแตกต่างของความลึกดังกล่าวสามารถอธิบายได้ถึงท่าทางที่เกิดขึ้นในขณะนั้นดังภาพประกอบที่ 4-9



(a) (b)

ภาพประกอบ 4-9 ตัวอย่างท่าทางที่พิจารณาท่าทางโดยใช้ความลึก(a) การก้ม (b) การนั่ง

สำหรับการทดลองจะแบ่งการทดลองออกเป็น 2 การทดลองหลัก ได้แก่ การทดลองที่ 1 วิเคราะห์โดยใช้โครงข่ายประสาทเทียม (Neural network) และการทดลองที่ 2 วิเคราะห์โดยใช้ SVM (Support Vector Machine)

ตัวอย่างของท่าทางสำหรับการวิเคราะห์ท่าทาง

ภาพประกอบ 4-9 ตัวอย่างการแสดงท่าทางการยืน และการเดิน

ภาพประกอบ 4-10 ตัวอย่างการแสดงท่าทางการนอน

ภาพประกอบ 4-11 ตัวอย่างการแสดงท่าทางการนั่ง

ภาพประกอบ 4-12 ตัวอย่างการแสดงท่าทางการก้ม

4.4.1 การทดลองที่ 1 วิเคราะห์โดยใช้โครงข่ายประสาทเทียม (Neural network)

4.4.1.1 สมมุติฐาน

การใช้คุณสมบัติเด่นจากโมเดลมนุษย์สำหรับการคำนวณพารามิเตอร์อันนำไปเข้าสู่กระบวนรู้จำเพื่อวิเคราะห์ท่าทางสามารถคำนวณได้จากแบบจำลองมนุษย์อย่างง่าย โดยพิจารณาความสัมพันธ์ของแต่ละองค์ประกอบภายในที่สัมพันธ์กันของโครงสร้างมนุษย์ ประกอบด้วย 2 เว็กเตอร์ คือ เว็กเตอร์สีน้ำเงิน $\vec{v}_h$ อธิบายความสัมพันธ์ระหว่างลำตัว และศีรษะ ส่วนเว็กเตอร์สีชมพู $\vec{v}_l$ อธิบายความสัมพันธ์ระหว่างลำตัว และขาโดยค่ามุมของเว็กเตอร์ $[\theta_h, \theta_l]$ และค่าความต่างของความลึก $[D_h, D_l]$ สามารถอธิบายถึงท่าทางของมนุษย์โดยการใช้โครงข่ายประสาทเทียม (Neural Network) สำหรับรู้จำท่าทางดังตัวอย่างพารามิเตอร์สำหรับการรู้จำในบทที่ 3

4.4.1.2 ปัจจัยกำหนดในการทดลอง

1) ใช้ OpenCv library สำหรับการประมวลผลภาพ และกล้อง Kinect Camera สำหรับทำวิดีโอข้อมูลที่ให้ภาพสี ความละเอียดขนาด 640x480 จุด และภาพความลึก ขนาด 640x480 จุด โดยระดับของความลึกที่ได้อยู่ที่ 256 ระดับ คอมพิวเตอร์สำหรับการประมวลผล CPU Dual Core 2.4 Ghz, แรม 1 Gb

2) ภาพวีดิโอข้อมูลประกอบด้วยการจำลองท่าทางของมนุษย์จำนวน 4 คน ทำการคัดข้อมูลสำหรับใช้ในการเทรนโครงข่ายประสาทเทียม (Neural Network) โดยใช้แมทแลป (Mat Lab) ฟังก์ชั่น Neural Pattern Recognition Tool เป็นเครื่องมือสำหรับการเทรน

4.4.1.3 ผลการทดลอง

ผลการทดลองจากการใช้โครงข่ายประสาทเทียม (Neural Network) ได้มีการกำหนดอินพุตทั้งหมด 4 อินพุต คือ ค่ามุม $[\theta_h, \theta_l]$ ของเว็กเตอร์ที่สัมพันธ์กันของลำตัวกับศีรษะ และลำตัวกับขา และค่าความต่างของความลึก $[D_h, D_l]$ ระหว่างลำตัวกับศีรษะ และลำตัวกับขา โดยมีเลเยอร์ซ่อน 1 เลเยอร์ และเอ้าพุตมีทั้งหมด 4 เอ้าพุต นั่นก็คือท่าทาง การยืนหรือการเดิน (0001), การนั่ง (0010), การนอน (0100) และการก้ม (1000)

<u>การทดลองย่อยที่ 1 ทดลองที่มุมมอง 60 องศา</u>

สำหรับการเทรนจะแบ่งเป็นสองการทำลองคือที่จำนวน 400 ชุดข้อมูล แบ่งเป็น การยืนหรือการเดิน 100 ข้อมูล การนั่ง 100 ข้อมูล การก้ม 100 ข้อมูล และการนอน 100 ข้อมูล จำนวนโหนด (node) ที่ใช้สำหรับการเทรนข้อมูล คือ 20, 25, 30, 35, 40

ตารางที่ 4-2 อัตราความถูกต้องสำหรับการสุ่มจำนวนโหนดที่ชุดข้อมูล400

โหนด	การรู้จำท่าทาง(เฉลี่ย)			
	Training	Validation	Testing	Total
20	95.0%	95.0%	96.7%	95.2%
25	95.0%	98.3%	96.7%	95.8%
30	94.6%	93.3%	95.0%	94.5%
35	95.7%	93.3%	93.3%	95.0%
40	93.9%	96.7%	98.3%	95.0%

และจำนวน 800 ชุดข้อมูลแบ่งเป็น การยืนหรือการเดิน 200 ข้อมูล การนั่ง200 ข้อมูล การก้ม 200 ข้อมูล และการนอน 200 ข้อมูล จำนวนโหนด (node) ที่ใช้สำหรับการเทรนข้อมูล คือ 20, 25, 30, 35, 40

ตารางที่ 4-3 อัตราความถูกต้องสำหรับการสุ่มจำนวนโหนดที่ชุดข้อมูล800

โหนด	การรู้จำท่าทาง(เฉลี่ย)			
	Training	Validation	Testing	Total
20	95.4%	95.8%	96.3%	95.1%
25	95.9%	95.0%	97.5%	96.0%
30	96.4%	96.7%	96.7%	96.5%
35	96.4%	94.2%	94.2%	95.8%
40	94.5%	95.8%	93.3%	94.5%

จากตารางที่ 4-1 และ4-2 อธิบายถึงความเหมาะสมของจำนวนโหนด กับข้อมูลที่ใช้สำหรับเทรนการเทรนเมื่อจำนวนข้อมูลมีอัตราส่วนมากกว่าจำนวนโหนด การเพิ่มโหนดจะทำให้ความถูกต้องมากขึ้นแต่เมื่อเพิ่มถึงจุดหนึ่งที่จำนวนข้อมูลมีความเหมาะสมกับจำนวนโหนดจะได้ผลลัพธ์ที่มีความถูกต้องมากที่สุด หลังจากนั้นความถูกต้องจะลงลดเนื่องจากจำนวนโหนดมีอัตราส่วนมากกว่าจำนวนข้อมูลและความถูกต้องจะมากขึ้นโดยขึ้นอยู่กับปริมาณข้อมูลที่มากขึ้นอีกด้วยโดยจำนวนโหนดที่มีความเหมาะสมมากที่สุดคือ 30โหนด ต่อ จำนวนข้อมูล800 ชุดข้อมูล ดังตารางที่ 4-3

ตารางที่ 4-4 อัตราความถูกต้องของแต่ละท่าทางที่มากที่สุด

ท่าทาง	การรู้จำท่าทาง(เฉลี่ย)
การยืน หรือการเดิน	95.8%
การนั่ง	96.5%
การนอน	96.6%
การก้ม	97.0%

จากตารางที่ 4-4 แสดงผลการทดลอง การยืนหรือการเดิน การนอน การก้มและการนั่ง โดยในลำดับที่ต่ำที่สุดมีความถูกต้อง95.80 % ที่ท่าการยืนและการเดิน และสูงที่สุดมีความถูกต้อง 97.0% ที่ท่าการก้ม โดยความถูกต้องของการรู้จำท่าทางพื้นฐานโดยเฉลี่ยอยู่ที่96.5%

ตารางที่ 4-5 อัตราความถูกต้องของแต่ละท่าทาง และความผิดพลาด

	การยืน และการเดิน	การนั่ง	การนอน	การก้ม
การยืน หรือการเดิน	184	5	1	2
การนั่ง	7	194	0	0
การนอน	3	1	199	3
การก้ม	6	0	0	195

จากตารางที่ 4-5 อธิบายถึงความผิดพลาดของการวิเคราะห์ท่าทางสำหรับของแต่ละท่าทาง จำนวนท่าทางละ200 ครั้ง โดยความผิดพลาดที่น้อยที่สุดคือท่านอนมีความผิดพลาดโดยวิเคราะห์ว่าเป็นท่าการยืน หรือการเดิน 1 ครั้ง และท่าทางที่มีความผิดพลาดมากที่สุดคือท่าการยืน หรือการเดิน วิเคราะห์ผิดพลาดไปที่ท่า การนั่ง7 ครั้ง,การนอน 3 ครั้ง และการก้ม6 ครั้งเนื่องจาก การยืน หรือเดิน เป็นท่าทางหลักสำหรับก่อนที่จะเปลี่ยนไปท่าทางอื่นๆจึงทำให้ข้อมูลที่นำไปเทรนอาจเกิดความทับซ้อนของข้อมูลในช่วงของการเปลี่ยนท่าจึงทำให้ค่าความผิดพลาดมีค่าสูงกว่าท่าอื่นๆ

ตารางที่ 4-6 อัตราความถูกต้องของแต่ละมุมมอง

มุมกล้อง(องศา)	การรู้จำท่าทาง(เฉลี่ย)
0	87.56%
45	90.73%
75	93.38%
90	1.33%

จากตารางที่ 4-6 จากผลการทดลองมุมกล้องที่เข้าใกล้ 60 องศา ความคลาดเคลื่อนจะมีค่าน้อยลงเรื่อยๆ และจะไม่สามารถวิเคราะห์ท่าทางได้เมื่อถึงมุมมองที่ไม่สามารถเห็นร่างกายมนุษย์จากด้านข้างได้ครบทุกส่วนดังมุมมองที่90 องศา จากตารางที่ 4-6 ผลการทดลอง การยืนและการเดิน การนอน การก้มและการนั่ง โดยการปรับมุมมององศาของกล้องที่แตกต่างออกไป มุมมองที่มีความแม่นยำต่ำสุดคือ90 องศา มีความแม่นยำ1.33% และมุมมองที่มีความแม่นยำสูงที่สุดคือ 75 องศา มีความแม่นยำ93.38%

4.4.1.4 วิเคราะห์และสรุปผลการทดลอง

จากผลลัพธ์ที่ได้จากการทดลองสามารถสรุปผลการทดลองได้ว่าความถูกต้องขึ้นอยู่กับปริมาณของจำนวนของข้อมูล และจำนวนโหนดของโครงข่ายประสาทเทียม (Neural Network) และมีความเหมาะสมกันระหว่างจำนวนของข้อมูล และโหนด ซึ่งความถูกต้องของแต่ละมุมมองที่มีความคลาดเคลื่อนกันเกิดจากหลายสาเหตุ คือ การถ่ายวิดีโอข้อมูลที่ไม่ได้มาจากการถ่ายในครั้งเดียวกันเนื่องจากจำเป็นต้องใช้กล้องตัวเดียวกันสำหรับการทดลองเพราะพารามิเตอร์ของการคาลิเบตของกล้องแต่ละตัวไม่เหมือนกันอาจทำให้พารามิเตอร์อื่นได้รับผลกระทบ

4.4.2 การทดลองที่ 2 วิเคราะห์โดยใช้ SVM (Support Vector Machine) สำหรับการวิเคราะห์ท่าทาง

4.4.2.1 สมมติฐาน

การใช้คุณสมบัติเด่นจากโมเดลมนุษย์สำหรับการคำนวณพารามิเตอร์อันนำไปเข้าสู่กระบวนรู้จำเพื่อวิเคราะห์ท่าทางสามารถคำนวณได้จากแบบจำลองมนุษย์อย่างง่าย โดยพิจารณาความสัมพันธ์ของแต่ละองค์ประกอบภายในที่สัมพันธ์กันของโครงสร้างมนุษย์ ประกอบด้วย 2 เว็กเตอร์ คือ เว็กเตอร์สีน้ำเงิน $\vec{v}_h$ อธิบายความสัมพันธ์ระหว่างลำตัว และศีรษะ ส่วนเว็กเตอร์สีชมพู $\vec{v}_l$ อธิบายความสัมพันธ์ระหว่างลำตัว และขาโดยค่ามุมของเว็กเตอร์ $[\theta_h, \theta_l]$ และค่าความต่างของ

ความลึก $[D_h, D_l]$ สามารถอธิบายถึงท่าทางของมนุษย์โดยการใช้ SVM (Support Vector Machine) สำหรับรู้จำท่าทางดังตัวอย่างพารามิเตอร์สำหรับการรู้จำในบทที่3

4.4.2.2 ปัจจัยกำหนดในการทดลอง

1) ใช้ OpenCv library สำหรับการประมวลผลภาพ และกล้อง Kinect Camera สำหรับทำวิดีโอข้อมูลที่ให้ภาพสี ความละเอียดขนาด 640x480 จุด และภาพความลึก ขนาด 640x480 จุด โดยระดับของความลึกที่ได้อยู่ที่256 ระดับ คอมพิวเตอร์สำหรับการประมวลผล CPU Dual Core 2.4 Ghz, แรม 1 Gb

2) ภาพวีดิโอข้อมูลประกอบด้วยการจำลองท่าทางของมนุษย์จำนวน4 คน ทำการคัดข้อมูลสำหรับใช้ในการเทรน SVM (Support Vector Machine) โดยใช้ฟังก์ชั่นจากOpenCv สำหรับการเทรน และทดสอบ

4.4.2.3 ผลการทดลอง

ผลการทดลองจากการใช้ SVM (Support Vector Machine) ได้มีการกำหนดอินพุตทั้งหมด 4 อินพุต คือ ค่ามุม $[\theta_h, \theta_l]$ ของเว็กเตอร์ที่สัมพันธ์กันของลำตัวกับศีรษะ และลำตัวกับขา และค่าความต่างของความลึก $[D_h, D_l]$ ระหว่างลำตัวกับศีรษะ และลำตัวกับขา และเอ้าพุตมีทั้งหมด4เอ้าพุต นั่นก็คือท่าทาง การยืนหรือการเดิน(คลาส 1), การนั่ง(คลาส 2), การนอน (คลาส 3) และการก้ม (คลาส 4)

<u>การทดลองย่อยที่ 1 ทดลองที่มุมมอง 60 องศา</u>

สำหรับการเทรนจะแบ่งเป็นสองการทำลองคือที่จำนวน 400 ชุดข้อมูล แบ่งเป็น การยืนหรือการเดิน 100 ข้อมูล การนั่ง100 ข้อมูลการก้ม 100 ข้อมูล และการนอน 100 ข้อมูลจำนวนรอบของการเทรนที่ใช้สำหรับการเทรนข้อมูล คือ10000, 100000, 1000000, 10000000, 10000000

ตารางที่ 4-7 อัตราความถูกต้องสำหรับการสุ่มจำนวนโหนดที่ชุดข้อมูล400

จำนวนรอบการเทรน	การรู้จำท่าทาง(เฉลี่ย)
10000	90.8%
100000	91.8%
1000000	90.6%
10000000	91.0%

และจำนวน 800 ชุดข้อมูลแบ่งเป็น การยืนหรือการเดิน 200 ข้อมูล การนั่ง 200 ข้อมูล การก้ม 200 ข้อมูล และการนอน 200 ข้อมูล จำนวนรอบของการเทรนที่ใช้สำหรับการเทรนข้อมูล คือ 10000, 100000, 1000000, 10000000, 10000000

ตารางที่ 4-8 อัตราความถูกต้องสำหรับการสุ่มจำนวนโหนดที่ชุดข้อมูล800

จำนวนรอบการเทรน	การรู้จำท่าทาง(เฉลี่ย)
10000	90.8%
100000	91.5%
1000000	92.6%
10000000	91.4%

จากตารางที่ 4-7 และ4-8 อธิบายถึงความเหมาะสมของจำนวนรอบของการเทรน กับข้อมูลที่ใช้สำหรับเทรนการเทรนเมื่อจำนวนข้อมูลมีอัตราส่วนมากกว่าจำนวนรอบ การเพิ่มจำนวนรอบจะทำให้ความถูกต้องมากขึ้นแต่เมื่อเพิ่มถึงจุดหนึ่งที่จำนวนข้อมูลมีความเหมาะสมกับจำนวนรอบจะได้ผลลัพธ์ที่มีความถูกต้องมากที่สุด หลังจากนั้นความถูกต้องจะมีการเปลี่ยนแปลงน้อยมากหรือไม่มีเลยเนื่องจากไม่มีการเปลี่ยนแปลงของกลุ่มข้อมูลที่เข้าใกล้เส้นซับพอร์ตเว็กเตอร์ และความถูกต้องจะมากขึ้นโดยขึ้นอยู่กับปริมาณข้อมูลที่มากขึ้นอีกด้วยโดยจำนวนรอบที่มีความเหมาะสมมากที่สุดคือ 100000 รอบ ต่อ จำนวนข้อมูล800 ชุดข้อมูล ดังตารางที่4-9

ตารางที่ 4-9 อัตราความถูกต้องของแต่ละท่าทางที่มากที่สุด

ท่าทาง	การรู้จำท่าทาง(เฉลี่ย)
การยืน หรือการเดิน	90.5%
การนั่ง	97.0%
การนอน	97.0%
การก้ม	87.5%

จากตารางที่ 4-9 แสดงผลการทดลอง การยืนหรือการเดิน การนอน การก้มและการนั่ง โดยในลำดับที่ต่ำที่สุดมีความถูกต้อง87.50 % ที่ท่าการก้ม และสูงที่สุดมีความถูกต้อง97.0% ที่ท่าการนั่ง และการนอน โดยความถูกต้องของการรู้จำท่าทางพื้นฐานโดยเฉลี่ยอยู่ที่92.6%

ตารางที่ 4-10 อัตราความถูกต้องของแต่ละท่าทาง และความผิดพลาด

	การยืน หรือ การเดิน	การนั่ง	การนอน	การก้ม
การยืน หรือ การเดิน	178	6	2	7
การนั่ง	8	186	2	4
การนอน	5	6	194	5
การก้ม	9	2	2	184

จากตารางที่ 4-10 อธิบายถึงความผิดพลาดของการวิเคราะห์ท่าทางสำหรับของแต่ละท่าทาง จำนวนท่าทางละ200 ครั้ง โดยความผิดพลาดที่น้อยที่สุดคือท่านอนมีความผิดพลาดโดยวิเคราะห์ว่าเป็นท่าการยืน หรือการเดิน 2 ครั้ง,การนั่ง 2 ครั้งและการก้ม 2 ครั้ง และท่าทางที่มีความผิดพลาดมากที่สุดคือท่าการยืน หรือการเดิน วิเคราะห์ผิดพลาดไปที่ท่า การนั่ง8 ครั้ง,การนอน 5 ครั้ง และการก้ม9 ครั้ง เนื่องจาก การยืน หรือเดิน เป็นท่าทางหลักสำหรับก่อนที่จะเปลี่ยนไปท่าทางอื่นๆจึงทำให้ข้อมูลที่นำไปเทรนอาจเกิดความทับซ้อนของข้อมูลในช่วงของการเปลี่ยนท่าจึงทำให้ค่าความผิดพลาดมีค่าสูงกว่าท่าอื่นๆ

ตารางที่ 4-11 อัตราความถูกต้องของแต่ละมุมมอง

มุมกล้อง(องศา)	การรู้จำท่าทาง(เฉลี่ย)
0	84.26%
45	87.55%
75	89.14%
90	1.31%

จากตารางที่ 4-11 จากผลการทดลองมุมกล้องที่เข้าใกล้ 60 องศา ความคลาดเคลื่อนจะมีค่าน้อยลงเรื่อยๆ และจะไม่สามารถวิเคราะห์ท่าทางได้เมื่อถึงมุมมองที่ไม่สามารถเห็นร่างกายมนุษย์

จากด้านข้างได้ครบทุกส่วนดังมุมมองที่90 องศา จากตารางที่ 4-11 ผลการทดลอง การยืนและการเดิน การนอน การก้มและการนั่ง โดยการปรับมุมมององศาของกล้องที่แตกต่างออกไป มุมมองที่มีความแม่นยำต่ำสุดคือ90 องศา มีความแม่นยำ1.31% และมุมมองที่มีความแม่นยำสูงที่สุดคือ 75 องศา มีความแม่นยำ89.14%

4.4.2.4 วิเคราะห์และสรุปผลการทดลอง

จากผลลัพธ์ที่ได้จากการทดลองสามารถสรุปผลการทดลองได้ว่าความถูกต้องขึ้นอยู่กับปริมาณของจำนวนของข้อมูล และจำนวนรอบของการเทรน และมีความเหมาะสมกันระหว่างจำนวนของข้อมูล และจำนวนรอบ และเมื่อถึงจุดๆหนึ่งความถูกต้องจะมีการเปลี่ยนแปลงน้อยมากถึงไม่มีเลยนั่นก็ไม่มีการเปลี่ยนแปลงของตัวซับพอร์ตเว็กเตอร์ซึ่งความถูกต้องของแต่ละมุมมองที่มีความคลาดเคลื่อนกันเกิดจากหลายสาเหตุ คือ การถ่ายวิดีโอข้อมูลที่ไม่ได้มาจากการถ่ายในครั้งเดียวกันเนื่องจากจำเป็นต้องใช้กล้องตัวเดียวกันสำหรับการทดลองเพราะพารามิเตอร์ของกล้องปรับให้อยู่ในมาตรฐานเดียวกัน (Calibrate) ของกล้องแต่ละตัวไม่เหมือนกันอาจทำให้พารามิเตอร์อื่นได้รับผลกระทบ และอีกหนึ่งปัจจัยหลักที่ทำให้เกิดข้อผิดพลาดคือระดับความละเอียดที่ได้จากภาพความลึกเมื่อวัตถุอยู่ห่างจากกล้องมากๆความละเอียดจะลดน้อยลงซึ่งทำให้พารามิเตอร์ที่ได้มีความคลาดเคลื่อนทำให้ส่งผลต่อการวิเคราะห์ท่าทางที่คลาดเคลื่อนด้วยเช่นกัน

4.5 การวิเคราะห์ท่าทางมนุษย์แบบเรียลไทม์ (Realtime Implementation)

สำหรับการวิเคราะห์ท่าทางสามารถทำได้แบบเรียลไทม์ด้วยความเร็วในการประมวลผลที่ความเร็ว 22 เฟรม/วินาที สำหรับการวิเคราะห์ด้วยโครงข่ายประสาทเทียม และวิเคราะห์โดยใช้ SVM (Support Vector Machine) ประมวลผลที่ความเร็ว 23 เฟรม/วินาที โดยใช้ภาพที่มีความละเอียด 320x240 พิกเซล

โดยการนำค่าน้ำหนักที่ได้จากการเทรนในโครงข่ายประสาทเทียมโดยใช้โปรแกรม Matlab มาทำการคูณกับพารามิเตอร์ที่ได้จากกระบวนการดึงคุณสมบัติจะได้เป็นค่าซึ่งมีทั้งหมด4 กลุ่ม ได้แก่ การยืนหรือการเดิน(0001), การนั่ง(0010), การก้ม(0100) และ การนอน (1000) สำหรับกระบวนการวิเคราะห์โดยใช้ SVM ใช้ library ของ opencv สำหรับการทำเรียลไทม์โดยใช้ข้อมูลเทรนชุดเดียวกับโครงข่ายประสาทเทียมและแบ่งเป็น 4 กลุ่มเช่น

4.6 สรุป

สำหรับการประเมินประสิทธิภาพจะแบ่งการทดลองออกเป็น 2 การทดลองหลัก ได้แก่ การทดลองที่ 1 วิเคราะห์โดยใช้โครงข่ายประสาทเทียม (Neural network) และการทดลองที่ 2 วิเคราะห์โดยใช้ SVM (Support Vector Machine) โดยใช้ลักษณะเด่นเป็นอินพุตสำหรับการเทรนคือ $[\theta_h, \theta_l]$ และ $[D_h, D_l]$ และเอ้าพุตทั้งหมด4 เอ้าพุตตามจำนวนท่าทางคือ การยืนหรือการเดิน การนั่ง การนอน และการก้ม

การทดลองที่ 1 ความถูกต้องขึ้นอยู่กับความเหมาะสมของอัตราส่วนระหว่างจำนวนข้อมูลที่เทรน และจำนวนโหนดของการเทรน และจะแปรผันตามปริมาณจำนวนของข้อมูลที่นำมาเทรน การทดลองที่ 2 ความถูกต้องขึ้นอยู่กับจำนวนปริมาณของข้อมูลที่นำมาเทรน และความครอบคลุมของข้อมูล โดยทั้งสองวิธีที่ใช้ในการทดสอบสิ่งสำคัญที่ทำให้ได้ความถูกต้องคือลักษณะเด่นที่เทรนมีความเป็นลักษณะเด่นเฉพาะที่มีความทับซ้อนกันน้อยมากซึ่งความถูกต้องของแต่ละมุมมองที่มีความคลาดเคลื่อนกันเกิดจากหลายสาเหตุ คือ การถ่ายวิดีโอข้อมูลที่ไม่ได้มาจากการถ่ายในครั้งเดียวกันเนื่องจากจำเป็นต้องใช้กล้องตัวเดียวกันสำหรับการทดลองเพราะพารามิเตอร์ของการคาลิเบตของกล้องแต่ละตัวไม่เหมือนกันอาจทำให้พารามิเตอร์อื่นได้รับผลกระทบ

โดยผลลัพธ์จากการทดลองที่ 1 มีความแม่นยำมากกว่าการทดลองที่ 2 เนื่องจากการใช้โครงข่ายประสาทเทียม (Neural Network) ผลจากการเทรนจะได้เป็นค่าน้ำหนักสำหรับการนำไปเข้าสู่กระบวนวิเคราะห์ซึ่งมากความยืดหยุ่นมากกว่า SVM ที่ใช้ตัวซับพอร์ตเว็กเตอร์สำหรับการแบ่งคลาสทำให้ความยืดหยุ่นของข้อมูลอินพุตที่คลุมเครือน้อยกว่าวิธีโครงข่ายประสาทเทียมซึ่งทำให้ผลลัพธ์มีความถูกต้องน้อยลง

จากผลการทดลองในบทนี้กล่าวถึงการใช้ภาพสี และภาพความลึกนำมาดึงลักษณะเด่นจากนั้นวิเคราะห์ท่าทางโดยใช้วิธีการสอนคอมพิวเตอร์ให้มีความเข้าพารามิเตอร์เพื่อจำแนกท่าทาง(Machine Learning) สำหรับการทดสอบระบบมีความยืดหยุ่นในเรื่องของมุมกล้องซึ่งผลก็คือมีความอิสระในเรื่องของการติดตั้ง

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

ปัจจุบันการเกิดเหตุการณ์ที่ไม่พึงประสงค์สามารถเกิดได้ตลอดเวลา ไม่ว่าจะเป็นเหตุการณ์ที่เกี่ยวกับความปลอดภัยในชีวิต และทรัพย์สิน จากข่าวตามสื่อต่างๆมากมาย เช่น การปล้น การลอบสังหาร การลอบวางระเบิด เป็นต้น หรือเหตุการณ์ที่เกี่ยวกับการเฝ้าระวังมนุษย์เพื่อวิเคราะห์พฤติกรรมความเป็นอยู่ เช่น การดูแลผู้สูงอายุ การสังเกตการณ์ทำงานของพนักงานในโรงงาน เป็นต้น เหตุการณ์ที่กล่าวมาถือเป็นเรื่องที่ควรให้ความสำคัญเป็นอย่างมาก เนื่องจากมีผลกระทบต่อการดำรงชีวิตของมนุษย์อย่างไรก็ตามเรื่องของการเฝ้าระวังพฤติกรรมมนุษย์ในปัจจุบันยังไม่สามารถทำได้ดีพออันเนื่องมาจากระบบรักษาความปลอดภัยที่มีอยู่ไม่สามารถสังเกตได้ตลอดเวลา มาจากหลายสาเหตุ เช่น บุคคลที่นำมาเฝ้าเกิดการบกพร่องในหน้าที่ไม่ว่าจะเป็นกรณีใดๆก็ตาม หรือแม้กระทั่งเป็นผู้กระทำผิดเอง เป็นต้นดังนั้นการนำเอาระบบการประมวลภาพเข้ามาช่วยสำหรับวิเคราะห์เหตุการณ์ผิดปกติที่เกิดแบบอัตโนมัติจะให้ช่วยลดภาระ และทำให้ได้ผลลัพธ์ที่ดีกว่า

5.1 สรุปผลการวิจัย

สำหรับงานวิจัยนี้ได้วิจัยและพัฒนาวิธีการวิเคราะห์ท่าทางพื้นฐานได้แก่ การยืน การเดิน การนั่ง การก้ม และการนอน ซึ่งสามารถนำไปประยุกต์จำแนกพฤติกรรมเพื่อใช้ในการเฝ้าระวังเหตุการณ์ผิดปกติ โดยใช้กล้องที่มีคุณสมบัติสี และความลึก จากมุมมองด้านข้างโดยใช้ค่าความลึกเข้ามาช่วยแก้ปัญหาในส่วนของข้อผิดพลาดที่เกิดจากมุมมองในกรณีที่เกิดท่าทางในมุมมองหน้าตรงกับกล้องซึ่งจะทำให้เกิดข้อผิดพลาดในส่วนของการดึงลักษณะเด่น แต่เมื่อนำความลึกเข้ามาช่วยทำให้สามารถดึงลักษณะเด่นอันนำไปสู่การวิเคราะห์ท่าทางได้อย่างแม่นยำมากยิ่งขึ้น โดยคุณลักษณะเด่นได้มาจากโมเดลมนุษย์อย่างง่ายที่มีความสัมพันธ์กันของลำตัวกับศีรษะ และลำตัวกับขา จากนั้นเข้าสู่กระบวนการวิเคราะห์ท่าทางโดยใช้กระบวนการสำหรับการรู้จำจากการสอน Neural Network และ SVM เพื่อเปรียบเทียบผลลัพธ์สำหรับรู้จำท่าทาง ได้แก่การยืนหรือการเดิน การนั่ง การก้ม และการนอน

ความผิดพลาดที่เกิดจากข้อจำกัดของกระบวนการลบพื้นหลังแยกเป็น 2 กรณี ดังนี้ (1) กรณีวัตถุอยู่ห่างจากกล้องทำให้ค่าความละเอียดของภาพความลึกมีค่าน้อยลงทำให้ค่าพื้นหลังกับวัตถุถ้าอยู่ติดกันมากๆจะไม่สามารถแยกได้จึงถูกจำแนกเป็นพื้นหลังไปด้วยทำให้บริเวณของวัตถุบางส่วนหายไปซึ่งถ้าหากเป็นพื้นที่เพียงเล็กน้อยกระบวนการติดตามโดยใช้ขั้นตอนวิธีการย้ายเข้าสู่ค่ากลางโดยการปรับตัวอย่างต่อเนื่องจะยังคงให้ผลการติดตามที่ไม่เกิดความผิดพลาดมากนัก

เนื่องจากกระบวนการนี้ ใช้พื้นฐานทางสถิติโดยการหาตำแหน่งค่ากลาง ถ้าหากจุดข้อมูลหายไปเพียงเล็กน้อยเมื่อคำควณหาค่ากลางย่อมได้ค่าที่มีความผิดพลาดเพียงน้อยเท่านั้น แต่ถ้าหากพื้นที่บริเวณวัตถุมีค่าความลึกเหมือนกับพื้นหลังเป็นบริเวณกว้างจะทำให้ผลการหาตำแหน่งของวัตถุผิดไปอย่างมาก ซึ่งท้ายที่สุดจะส่งผลไปยังการจำแนกท่าทางที่ผิดพลาด ดังนั้นการพัฒนากระบวนการลบพื้นหลังให้มีประสิทธิภาพดีขึ้นย่อมมีความสำคัญกับการพัฒนากระบวนรู้จำท่าทางและรู้จำกิจกรรมต่อไป

5.2 อภิปรายผล

สำหรับการประเมินของระบบได้แบ่งการทดสอบเป็น 2 การทดลอง การทดลองที่ 1 วิเคราะห์โดยใช้โครงข่ายประสาทเทียม (Neural network) และการทดลองที่ 2 วิเคราะห์โดยใช้ SVM (Support Vector Machine)

การทดลองที่ 1 ใช้เทคนิคโครงข่ายประสาทเทียม (Neural network) ผลการทดลองที่มีความแม่นยำมากที่สุด96.5% โดยในลำดับที่ต่ำที่สุดมีความถูกต้อง 95.80 % ที่ท่าการยืนและการเดิน และสูงที่สุดมีความถูกต้อง97.0% ที่ท่าการก้ม โดยที่จำนวนโหนดและจำนวนข้อมูลที่เทรนมีผลต่อความแม่นยำซึ่งในที่นี้คือ30 โหนด ที่ข้อมูล800 ชุด ต่อ 1 มุมมอง สรุปคือถ้าอัตราส่วนของจำนวนชุดข้อมูลและจำนวนโหนดของการเทรนที่ไม่เหมาะสมผลลัพธ์ก็จะมีความคลาดเคลื่อน

การทดลองที่ 2 ใช้เทคนิค SVM (Support Vector Machine) ผลการทดลองที่มีความแม่นยำมากที่สุด 91.5% โดยในลำดับที่ต่ำที่สุดมีความถูกต้อง87.50 % ที่ท่าการก้ม และสูงที่สุดมีความถูกต้อง97.0% ที่ท่าการนั่ง และการนอนที่ชุดข้อมูลเดียวกันคือ800 ชุด ต่อ 1 มุมมอง

สำหรับความถูกต้องของมุมมองพบว่ามุมมองที่มีความใกล้เคียงกับมุมมองหลัก(มุมมองที่ 60 องศา) มากที่สุดจะมีความแม่นยำถูกต้องมากกว่ามุมมองที่ห่างออกไปเนื่องจากค่าพารามิเตอร์หลังผ่านกระบวนการโรเตชั่น (rotation) แล้วมีความผิดพลาดไม่แม่นยำ100% และทำให้กระทบต่อความค่าพารามิเตอร์ที่ใช้สำหรับการวิเคราะห์ซึ่งจะกระทบมากขึ้นเมื่อวัตถุอยู่ห่างจากกล้องมากขึ้น เนื่องจากค่าความละเอียดของกล้องที่เป็นภาพความลึกจะละเอียดน้อยลงเมื่อวัตถุอยู่ห่างจากกล้องมากขึ้น

5.3 ข้อเสนอแนะ

สำหรับงานวิจัยนี้มุ่งเน้นไปที่การวิเคราะห์ท่าทางซึ่งเป็นท่าทางพื้นฐานที่สามารถนำไปประยุกต์เพื่อวิเคราะห์เป็นพฤติกรรม ที่สามารถนำไปใช้เป็นระบบเฝ้าระวังต่างๆ เช่น ดูแลผู้สูงอายุ ผู้ป่วยหรือไม่ว่าจะเป็นเหตุการณ์ที่เกี่ยวกับความปลอดภัยในชีวิต และทรัพย์สิน ซึ่งสามารถแจ้ง

เตือนได้อย่างทันท่วงที จากการทดสอบระบบประสิทธิภาพสูง และการติดตั้งกล้องทำได้ง่ายเป็นระบบที่ไม่ซับซ้อน แต่เนื่องจากกล้องที่ใช้เป็นกล้องการประยุกต์ใช้กล้องที่ทำงานเฉพาะด้านสำหรับการเล่นเกมส์ (Kinect Camera) ที่ใช้งานในที่ไม่มีแสงแดด หรือมีความสว่างน้องกว่าอินฟาเรดเท่านั้นเนื่องจากตัวกล้องใช้อินฟาเรดสำหรับการสร้างภาพความลึกโดยบางมุมที่อินฟาเรดส่องไปไม่ถึงทำให้ส่วนนั้นเป็นจุดบอดที่ไม่สามารถวิเคราะห์ได้ จึงทำให้งานวิจัยนี้มีข้อจำกัดในเรื่องของสภาพแวดล้อมที่ปิด และขนาดพื้นที่ของการวิเคราะห์ถูกจำกัดโดยกล้องในเรื่องของความคลาดเคลื่อนเมื่อวัตถุอยู่ห่างจากกล้องมากๆทำให้ความละเอียดลดน้อยลงซึ่งจะส่งผลโดยตรงกับพารามิเตอร์ความลึกที่นำไปเป็นอินพุตของการวิเคราะห์ท่าทาง สำหรับการพัฒนาต่อควรใช้กล้องที่เป็น Stereo Vision จากการนำภาพสีมาสร้างเป็นภาพความลึกซึ่งจะทำให้ผลลัพธ์มีความแม่นยำมากขึ้น สำหรับความผิดพลาดที่เกิดจากวัตถุอยู่ห่างจากกล้องมากส่งผลกระทบให้ค่าความละเอียดของภาพความลึกลดลงจนทำให้พารามิเตอร์คลาดเคลื่อนสามารถแก้ไขเพิ่มความละเอียดของภาพจากปกติใช้เป็นภาพความละเอียดขนาดเพียง 8 บิท ซึ่งกล้องkinect camera สามารถให้ภาพที่มีความละเอียดได้สูงสุดที่ 16 บิท

เอกสารอ้างอิง

[1] รศ.ดร.มนตรี กาญจนะเดชะ, หนังสือการประมวลผลภาพ, last modified: unknown, Access date: 24 November 2008

[2] ผศ.ดร.สุเทพ มาดารัศมี, http://cpe.kmutt.ac.th/previousproject/2005/2/index.htm, last modified: unknown, Access date: 30 January 2009.

[3] Gary Bradski and Adrian Kaebler, Learning OpenCV, last modified: September 2008, Access date: 31 January 2009.

[4] John G. Allen, Richard Y. D. Xu and Jesse S. Jin, "Object Tracking Using CamShift Algorithm and Multiple Quantized", in Proc. 2003 Pan-Sydney Area Workshop on Visual Information Processing (VIP2003), Sydney, Australia. CRPIT.

[5] C. Canton-Ferrer, J.R.Casas, M.Pardas, M.E.Sargin and A.M.Tekalp, "3D Human Action Recognition in Multiple View Scenarios," in ICIP (2006).

[6] M. Ahmad and Seong-Whan Lee, "Human action recognition using multi-view image sequences," in Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on, 2006, 523-528, 10.1109/FGR.2006.65.

[7] Nattapon Noorit, Nikom Suvonvorn, and Montri Karnchanadecha, "Model-based human action recognition," in (presented at the Second International Conference on Digital Image Processing, Singapore, Singapore, 2010), 75460P-75460P-6, http://spie.org/x648.html?product_id=853223.

[8] N. Gkalelis, N. Nikolaidis, and I. Pitas, "View indepedent human movement recognition from multi-view video exploiting a circular invariant posture representation," in Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on, 2009, 394-397, 10.1109/ICME.2009.5202517.

[9] Daniel Chen, Pi-chi Chou, Clinton Fookes and Sridha Sridharan, "Multi-view human pose estimation using modified five-point skeleton model," in ICSP (2007).

[10] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in Proc. IEEE Int. Conf. on Computer Vision and Pattern Rec., 1999, pp. 252–259.

[11] C. Wu, A. H Khalili, and H. Aghajan, "Multiview Activity Recognition in Smart Homes with Spatio Temporal Features", in Proceedings of the 4th ACM/IEEE ICDSC, pp 142–149.

[12] Ana Paula Brandão Lopes et al., "Action Recognition in Videos: from Motion Capture Labs to the Web," 1006.3506 (June 17, 2010), http://arxiv.org/abs/1006.3506.

[13] Wanqing Li, Zhengyou Zhang, and Zicheng Liu, “Action recognition based on a bag of 3D points,” in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, 2010, 9-14.

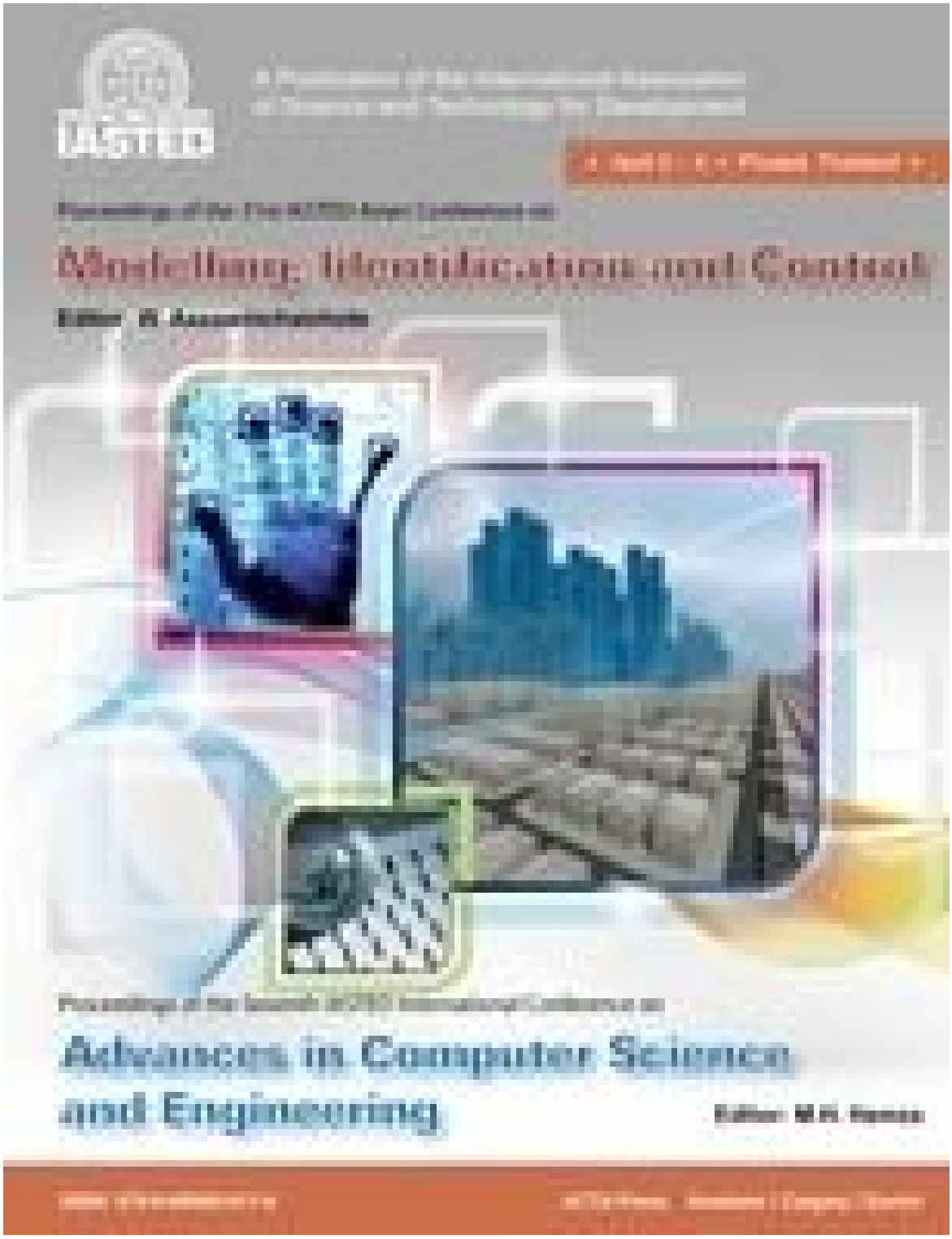
[14] R. Verma and A. Dev, “Vision based hand gesture recognition using finite state machines and fuzzy logic,” in Ultra Modern Telecommunications & Workshops, 2009. ICUMT '09. International Conference on, 2009, 1-6.

[15] Pengyu Hong, Thomas S. Huang, and Matthew Turk, “Gesture Modeling and Recognition Using Finite State Machines,” in Automatic Face and Gesture Recognition, IEEE International Conference on, vol. 0 (Los Alamitos, CA, USA: IEEE Computer Society, 2000), 410.

ภาคผนวก ก.

ผลงานตีพิมพ์เผยแพร่จากวิทยานิพนธ์

1. P.Chawalitsittikul and N. Suvonvorn, "*PROFILE-BASED HUMAN ACTION RECOGNITION USING DEPTH INFORMATION*," in *Proceedings of the IASTED International Conference on Advances in Computer Science and Engineering, ACSE 2012*, 2012, pp. 376–380.


IASTED
Modelling, Identification and Control
Proceedings of the Fourth IASTED International Conference on
Advances in Computer Science and Engineering

PROFILE-BASED HUMAN ACTION RECOGNITION USING DEPTH INFORMATION

Pongsatorn Chawalitsittikul and Nikom Suvonvorn*
Department of Computer Engineering, Faculty of Engineering, Prince of Songkla University,
Hatyai, Songkhla, 90112
E-mail: armclashblack@hotmail.com, kom@coe.psu.ac.th*

ABSTRACT

The recognition of human actions is an important step for human behaviour understanding via video processing. In this paper, we propose an accurate method of model-based action recognition using color and depth information. The parametric model of human is extracted from image sequences using mixture of motion, color and depth information. Extracted model features are then classified into five basic actions using artificial neural network. The experimentation result shows that our method gives high recognition rate more than 90%.

KEY WORDS

Action recognition, Human modeling, Multi-views, RGB-D image, Video surveillance

1. Introduction

In the near future, the automatic video processing for human behaviour recognition will play the important role in many applications, such as, the unusual situation detection, elderly people monitoring, activities analysis in factory, and etc. In our research, we assume that the complex behaviour is the composition of the atomic actions: standing, walking, sitting, blending, and laying. Thus, the specific order of these actions can construct the complex behaviour. If the actions can be correctly recognized, the behaviour can then be better modeled.

In this paper, we propose a method for human action recognition from side view using motion, color image, and depth information. Recently, many researches proposed various solutions that can be divided into two main approaches [1]: parametric object modeling, and implicit object modeling. Hierarchically, the parametric model [2] [3] would study either internal model or trajectory model. For the implicit model [4][5], it concerns for both 2D and 3D image template analysis. For parametric model, N.Noorit [2] proposed a simple human structure consisting head, body and leg, together with motion detection and cam-shift color tracking for classifying the five basic actions. C.Wu [3] uses the trajectory model for recognizing activities in a home environment. Multiple cameras are needed to cover the whole observation area. They focus on spatiotemporal features with three different fusion methods. In the implicit modeling, N.Gkalelis [4] proposed a method for view independent human movement representation and recognition. The binary masks of multi-view posture image are vectorized and concatenated, and then using circular shift invariance for view correspondence. Fuzzy vector quantization and linear discriminant analysis are used for classification. D.Chen [5] proposed the 3D human pose utilizing a modified five-point skeleton model. In this paper, we describe our proposed technique in section 2, then the experimentation result and conclusion is presented in section 3 and 4 respectively.

2. Proposed method

In this section, we describe how to determine the parametric model of human. We attempt to reconstruct the 3D human pose from 2D human pose using motion, color, and depth information. The 3D parameters of our reconstructed human model are then used for recognizing actions by using neural network. The specific methods on each stage, such as, camera calibration, motion detection and tracking, skeleton algorithm, feature extraction, and action recognition are introduced. The figure 1 shows the overview of our method.

2.1 Moving object tracking

We found that in the depth image the moving object can be identified easier than moving object in the color image. Motion detection using color images is sensible to the sudden changing of light, and is inacceptable especially when the color of object is the same with background that cause catastrophic errors. In our case, we then firstly separate moving object from background by using depth image applying background subtraction technique. Since light and color don't affect to depth information, objects obtained from the process is then quit robust and more accurate than using color images. However, we find the small noises that can be removed by

morphological opening and closing filters. The filtered moving object as region of interest (ROI) is then tracked.

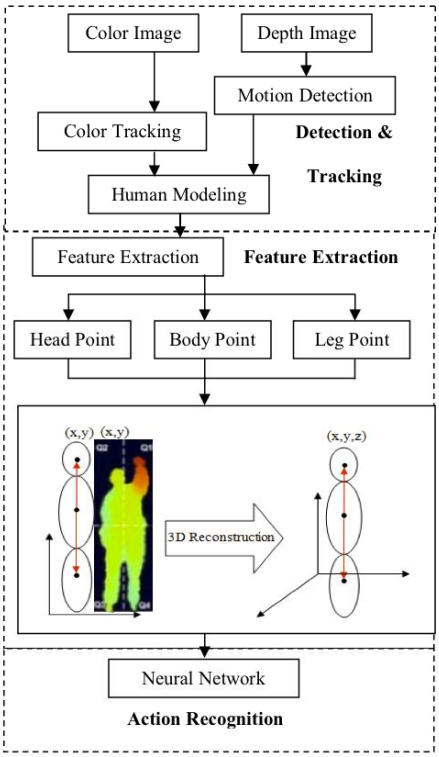


Figure1. Our proposed system

2.2Mixture of Color and Depth

Moving object with depth image gives a robust object. However, in order to track object in the time sequence, or to reconstruct the human body with many parts, we found that using only depth information for these tasks is quit sensible comparing to color information that represents the texture of objects, note that depth information is a gray scale image. To improve our human model, we then mix the color and depth of object, try to take the advantage of color to the moving object obtained by depth image. Notice that in the Kinect the depth and color camera are parallel. In theory, only translation matrix can transform the color image to overlay with depth image. However, in practice this is not applicable. Additionally, we need the use the calibration matrix obtained from classical calibration method. It consists of two matrices: rotation and translation:

Rotation Matrix

[9.9977321644139494e-01,1.7292658422779497e-03,-2.1225581878346968e-02;
-2.0032487074002391e-03, 9.9991486643051353e-01,-1.2893676196675344e-02;
2.1201478274968936e-02,1.2933272242365573e-02, 9.9969156632836553e-01]

Translation Matrix

[2.1354778990792557e-02,2.5073334719943473e-03,-1.2922411623995907e-02]

Then, the depth image was superimposed with color image which is transformed by calibration matrices (rotation and translation matrix). We construct a specific object space as mixture of color and depth information, called "color-depth object", depicted in figure 2. This object O_{DC} can be described by the following equation.

$$O_{DC} = E(O_D) \cup (I_C \cap (O_D - E(O_D))) \quad (1)$$

Where O_D is an interested object in depth image, I_C is the color image, and E is an erosion operator. This representation gives us the advantages for human modeling and tracking in real time.

Figure2.A mixture of color and depth object: left is I_C, and right is O_{DC}.

2.3Adaptive Human modeling

In this stage, we build the human model for the future recognition stage. The color-depth object obtained from the last stage will be divided into three parts of human structure (corresponding to three color regions of interest) referred to the center of mass, such as, head, body and legs, where the body is the central of human structure. Then, the vectors representing the relative links between body-head and body-leg are established. In order to obtain robust vectors, the skeleton algorithm is applied by three steps. Firstly, finding vectors defining the radius of the color-depth object, where the origin is fixed at the center of mass.

$$\bar{x}, \bar{y} = \frac{\sum_0^N x_i}{N}, \frac{\sum_0^N y_i}{N} \quad (2)$$

Secondly, all vectors are separated into four quadrants, and select the vectors $\Delta_O(X, Y, \theta, D, C)$ with maximum magnitude in each quadrant.

$$\Delta_O = \{\Delta_i | i = 1 \dots 4\} \quad (3)$$

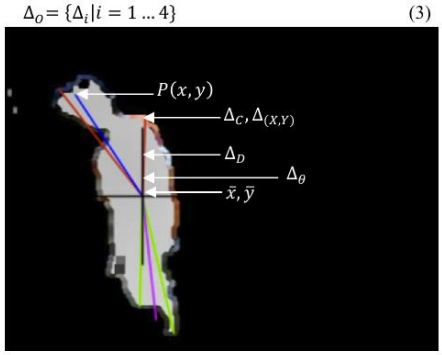


Figure3.Our human model: blue and purple lines show the fused vectors.

The last step, color of the maximum vector is determined by color of region at the end of that vector. We introduce the color property to vector in order to increase the accuracy of object (vectors) tracking, as shown in figure 3. From skeleton algorithm, we have four maximum vectors, in order to simplify our human model but having enough parameters for defining the basic actions in our recognition system, we found that these can be reduced to only two vectors, representing head and leg. The vector fusion is done by using its properties (magnitudes D_{v_i} and colors D_{c_i}).

$$D_{v_i} = \sqrt{(\Delta_{x_i} - \bar{x})^2 + (\Delta_{y_i} - \bar{y})^2} \ | \ i = 1 \dots 4 \quad (4)$$

$$D_{c_i} = \sqrt{(\Delta_{c_i} - Cr_t)^2} \ | \ i = 1 \dots 4 \quad (5)$$

The intermediate weights ω_{v_i} and ω_{c_i} are determined for the further vector fusion, which is described by fuzzilization of vector ratio as the following.

$$\omega_{v_i} = \left[\sum_{j=1}^{2} \left(\frac{D_{v_1}}{D_{v_j}} \right)^{\frac{2}{m-1}} \right]^{-1} \ | \ i = 1 \dots 4 \quad (6)$$

$$\omega_{c_i} = \left[\sum_{j=1}^{2} \left(\frac{D_{c_1}}{D_{c_j}} \right)^{\frac{2}{m-1}} \right]^{-1} \ | \ i = 1 \dots 4 \quad (7)$$

Then, the learning weights ω_1 and ω_2 are established for objectively updating vector proportion by a learning rate α_v.

$$\omega_1 = \alpha_v \omega_v + (1 - \alpha_v)\omega_c \quad (8)$$

$$\omega_2 = 1 - \omega_1 \quad (9)$$

Finally, the vectors are fused for both magnitude (eq. 10) and color (eq.11), where α_c is learning rate of color update.

$$P(x,y) = \omega_1 P_1(\Delta_{x_1}, \Delta_{y_1}) + \omega_2 P_2(\Delta_{x_2}, \Delta_{y_2}) \quad (10)$$

$$Cr_{t+1} = \alpha_c Cr_t - (1 - \alpha_c)(\omega_1 C_1 + \omega_2 C_2) \quad (11)$$

Where $P(x,y)$ is 3D point at the end of vector and Cr_{t+1} is color of vector at time t+1. Figure 3 shows human model representing by our method.

2.4Human Features

The parameters of model are considered as internal parameters representing the characteristic of human structure, which describes by the directions $[\theta_h, \theta_l]$ and distances $[D_h, D_l]$ between the object to the camera from body component to the head and legs components respectively in figure 4.

$$R = \sqrt{(P_x^{\,2} + P_y^{\,2})} \quad (12).$$

$$[\theta_h, \theta_l] = \cos^{-1}\left(\frac{P_x}{R}\right) \quad (13).$$

$$[D_h, D_l] = \sqrt{(d_{(h,l)} - d_c)^2} \quad (14).$$

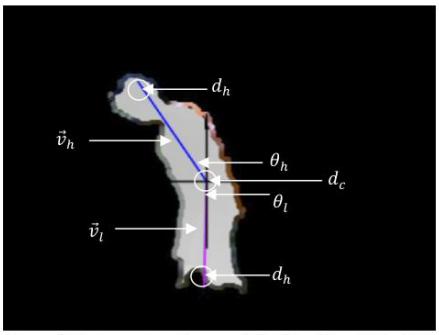


Figure4.Feature extraction result is $[\theta_h, \theta_l]$ and $[D_h, D_l]$ for action recognition.

The feature vector that will be used for action recognition is defined as $[\theta_h, \theta_l, D_h, D_l]$.

2.5. Action Recognition

The multilayer perception of neural network is applied for action recognition. The network is trained by back-propagation algorithm with one hidden layer and sigmoid is used as activate function.We consider various nodes in hidden layer in order to find the best recognition rate. In

the case of using neural network, fixing the number of nodes of input and output layer is necessary. In our case, 4 nodes of input layer is used because of directions $[\theta_h, \theta_l]$ (the angle ratio between head and legs with respect to the body) and distance D_h, D_l (the distance of depth data between head and leg with respect body). We use four actions for output layer has 4 nodes

3. Experimental Result

Our system is implemented using OpenCV and Kinect library. The experimentation is performed using the datasets established under indoor environment, no change of light condition with the following test. Four persons with different clothes act randomly and continuously with respect to testing actions: 100 standing, 100 sitting, 100 bending, and 100 laying actions. Testing system is run on Intel processor Dual Core, 2.4 Ghz, 1 GB memory. The video sequence using Microsoft Kinect Camera is analyzed with image resolutions 640x480 pixels at 15 fps for both color and depth images. Around 1,600 features are extracted using 70% for training, 15% for validation and 15% for testing. Table 1 shows the result of recognition rate while varying the number of hidden nodes. We can notice that at 20 nodes the system gives an optimal result at 95.8%.

Table1. Neural network with difference hidden nodes

Nodes	Recognition rate (average)			
	Training	Validation	Testing	Total
10	95.0%	95.0%	96.7%	95.2%
20	95.0%	98.3%	96.7%	95.8%
40	94.6%	93.3%	95.0%	94.5%
60	95.7%	93.3%	93.3%	95.0%
80	93.9%	96.7%	98.3%	95.0%

Then, we use neural network with 20 nodes for evaluating our action recognition system, classified into 4 basic actions, such as,standing &walking, bending, sitting and laying. From the table 2, we found that our system performs a good result up to 98.0% for laying and standing/walking actions. Sitting and bending actions get less accurate result at 93.0% and 94.1% respectively due to its ambiguities between these two actions.

Table2. Correction rate of our recognition system.

Actions	Our method	N.Noorit [2]	Chi-Hung[12] $(0°, -45°, 45°)$
Standing	98.0%	99.41 %	-
Walking	98.0%	80.65 %	92.4%,89.54%,88.89%
Bending	94.1%	94.35 %	95.4%, 90.38%, 89.02%
Sitting	93.0%	89.26 %	97.6%, 91.38%, 90.41%
Laying	98.0%	100 %	-

We compare the performance of our method to the methods of N.Noorit, and Chi-Hung, the result is shown in the table2. N.Noorit [2] can classify the human basic actions based on color image analysis for the five basic actions. We can notice that our method can perform better for walking and sitting, and a little bit lower for laying and standing. The result is quite reasonable because the depth information doesn't distinguish much to standing and laying actions from the profiled view of human structure. And, using only color image cannot well define the body structure with respect to different camera's angles.

For the comparison with Chi-Hung method, only walking, sitting, and bending are tested. We found that our method can perform totally better in the walking action and better in some angles for bending and sitting. We notice that Chi-Hung method gives better accuracy for the$0°$ degree, and reduce at $-45°$ and $45°$ degrees.

In summary, our method able to solve problems and improve the accuracy of action recognition by using "Mixture of Color and Depth", since light and color don't affect to depth information, objects obtained from the process is then quit robust and more accurate than using color images, and can analyze both the front and side view, with 95.8% accuracy. Moreover, our method can run in real time at 22 fps.

Conclusion

In this paper, we have presented an efficient technique for recognizing human actions using color and depth information from the side view of human. The human parametric model is introduced, its features can discriminate four basic actions via neural network. The system gives encourage recognition rate at 95.8% in average.

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] Ana Paula Brandão Lopes et al., "Action Recognition in Videos: from Motion Capture Labs to the Web," *1006.3506* (June 17, 2010),http://arxiv.org/abs/1006.3506.

[2] Nattapon Noorit, Nikom Suvonvorn, and Montri Karnchanadecha, "Model-based human action recognition," *in the Second International Conference on Digital Image Processing, Singapore*, Singapore, 2010, 75460P-75460P-6,

[3] C.Wu, A.H Khalili, and H.Aghajan, "Multiview Activity Recognition in Smart Homes with Spatio

Temporal Features", in *Proceedings of the 4th ACM/IEEE ICDSC*, pp 142–149.

[4] N.Gkalelis, N.Nikolaidis, and I.Pitas, "View indepedent human movement recognition from multi-view video exploiting a circular invariant posture representation," in *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on, 2009, 394-397, 10.1109/ICME.2009.5202517.*

[5] Daniel Chen, Pi-chi Chou, Clinton Fookes, and SridhaSridharan, "Multi-view human pose estimation using modified five-point skeleton model," in *ICSP* (2007).representation," in *Multimedia and Expo, 2009.ICME 2009.IEEE International Conference on*, 2009, 394-397, 10.1109/ICME.2009.5202517.

[6] C.Canton-Ferrer, J.R.Casas, M.Pardas, M.E.Sargin, and A.M.Tekalp, "3D Human Action Recognition in Multiple View Scenarios," in *ICIP* (2006).

[7] M. Ahmad and Seong-Whan Lee, "Human action recognition using multi-view image sequences," in *Automatic Face and Gesture Recognition, 2006. FGR 2006.7th International Conference on*, 2006, 523-528, 10.1109/FGR.2006.65.

[8] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Int. Conf. on ComputerVision and Pattern Rec.*, 1999, pp. 252–259.

[9] Ana Paula Brandão Lopes et al., "Action Recognition in Videos: from Motion Capture Labs to the Web," *1006.3506* (June 17, 2010), http://arxiv.org/abs/1006.3506.

[10] Wanqing Li, Zhengyou Zhang, and Zicheng Liu, "Action recognition based on a bag of 3D points," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, 2010,9-14.

[11] Gary Bradski and Adrian Kaebler, Learning OpenCV : Computer Vision with the OpenCV Library, O'Reilly, Inc, 2008.

[12] Chi-Hung Chuang, Jun-Wei Hsieh, Luo-Wei Tsai, and Kuo-Chin Fan, "Human Action Recognition Using Star Templates and Delaunay Triangulation," in *Intelligent Information Hiding and Multimedia Signal Processing, 2008. IIHMSP '08 International Conference on*, 2008, pp. 179 – 182.

ประวัติผู้เขียน

ชื่อ สกุล นายพงศธร ชวลิตสิทธิกุล

รหัสประจำตัวนักศึกษา 5310120016

วุฒิการศึกษา

วุฒิ	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
วิศวกรรมศาสตรบัณฑิต (วิศวกรรมคอมพิวเตอร์)	มหาวิทยาลัยสงขลานครินทร์	2552

ทุนการศึกษา

1. ทุนโครงการพัฒนามหาวิทยาลัยวิจัยแห่งชาติ(NRU)
2. ทุนอุดหนุนการวิจัยเพื่อวิทยานิพนธ์

การตีพิมพ์เผยแพร่ผลงาน

1. P. Chawalitsittikul and N. Suvonvorn, "*PROFILE-BASED HUMAN ACTION RECOGNITION USING DEPTH INFORMATION*," in *Proceedings of the IASTED International Conference on Advances in Computer Science and Engineering, ACSE 2012*, 2012, pp. 376–380.