



การแยกแยะท่าทางของมนุษย์โดยใช้กล้อง Kinect

นายธรรมศาสตร์ วิสุทธารมณั์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมซอฟต์แวร์
คณะเทคโนโลยีสารสนเทศ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
ปีการศึกษา 2557

การแยกแยะท่าทางของมนุษย์โดยใช้กล้อง Kinect

นายธรรมศาสตร์ วิสุทธารมณ วศ.บ. (วิศวกรรมคอมพิวเตอร์)

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมซอฟต์แวร์
คณะเทคโนโลยีสารสนเทศ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
ปีการศึกษา 2557

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการวิทยานิพนธ์
(ผศ. ดร.ชาคริตา นุกุลกิจ)

..... กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์
(ผศ. ดร.พรชัย มงคลนาม)

..... กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(Assoc. Prof. Dr. Jonathan Hoyin Chan)

..... กรรมการ
(ดร.วิฑิตา จงสุขชัยสิทธิ์)

..... กรรมการ
(ดร.นล เปรมย์เจียร)

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

หัวข้อวิทยานิพนธ์	การแยกแยะท่าทางของมนุษย์โดยใช้กล้อง Kinect
หน่วยกิต	12
ผู้เขียน	นาย ธรรมศาสตร์ วิสุทธารมณ
อาจารย์ที่ปรึกษา	ผศ. ดร.พรชัย มงคลนาม รศ. ดร.โจนาธาน โฮยีน ซาน
หลักสูตร	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมซอฟต์แวร์
คณะ	เทคโนโลยีสารสนเทศ
ปีการศึกษา	2557

บทคัดย่อ

งานวิจัยชิ้นนี้นำเสนอหลักการในการแยกแยะท่าทางที่ถูกแสดงออกโดยมนุษย์ โดยใช้ข้อมูลโครงสร้างร่างกายในรูปแบบ 3 มิติที่วิเคราะห์ได้จากกล้อง Kinect ร่วมกับหลักการในการทำเหมืองข้อมูลเพื่อใช้ในการแยกแยะท่าทาง โดยเลือกอัลกอริทึมสำหรับการเรียนรู้ชุดข้อมูล 4 อัลกอริทึม ประกอบด้วย Neural Network, Naïve Bayes, Logistic Regression และ Decision Tree เพื่อเปรียบเทียบหาอัลกอริทึมที่สามารถเรียนรู้ และแยกแยะข้อมูลท่าทางได้ดีที่สุด โดยชุดข้อมูลที่ถูกใช้เพื่อการเรียนรู้มีด้วยกัน 3 รูปแบบได้แก่ ชุดข้อมูลดิบจากโครงสร้างร่างกายที่วิเคราะห์ได้จากกล้อง Kinect ชุดข้อมูลที่ผ่านการคัดกรองโดยหลักการ Data Selection และชุดข้อมูลที่ผ่านการปรับเปลี่ยนชุดข้อมูลโดยหลักการ Data Transformation โดยโมเดลที่ใช้ในการแยกแยะท่าทางมีด้วยกัน 2 รูปแบบ ประกอบด้วย Single-stage Classifier Model ที่เป็นการใช้โมเดลเพียงอันเดียวในการวิเคราะห์ท่าทาง และ Multiple-stage Classifier Model ที่เป็นโมเดลที่ประกอบด้วยหลายโมเดลย่อยที่ทำงานร่วมกันเพื่อวิเคราะห์ข้อมูลท่าทาง เพื่อเปรียบเทียบศักยภาพระหว่าง 2 หลักการในวิเคราะห์ข้อมูลท่าทางจากกล้อง Kinect

คำสำคัญ : การแยกแยะท่าทาง / กล้อง Kinect / การทำเหมืองข้อมูล / การลดความซับซ้อนในชุดข้อมูล / การตรวจจับโครงสร้างร่างกายมนุษย์

Thesis Title	Human Posture Classification Using Kinect
Thesis Credits	12
Candidate	Mr. Thammarsat Visutharrom
Thesis Advisors	Asst. Prof. Dr. Pornchai Mongkolnam Assoc. Prof. Dr. Jonathan Hoyin Chan
Program	Master of Science
Field of Study	Software Engineering
Faculty	Information Technology
Academic Year	2014

Abstract

In this research, we proposed a method for a real-time human posture classification using 3D skeleton data streaming from Kinect camera. The data mining techniques were used to classify the postures. We selected 4 learning algorithms namely: neural network, naïve Bayes, logistic regression, and decision tree, for learning our data to compare and find the appropriate learning algorithm. Three types of training sets were used in our experiment, including Kinect's raw skeletal training set, skeleton with data selection training set, and with data transformation training set. Two models were used in our experiment. The first model was single-stage classifier model that used only one model to classify all postures, and the second model was multiple-stage classifier model that partitioned model into multiple models working together. We then compared the capability of posture classifications between both models.

Keywords: Postural classification / Kinect camera / Data mining / Data complexity reduction / Human skeleton tracking

กิตติกรรมประกาศ

ข้าพเจ้าขอกราบขอบพระคุณทุนสนับสนุนวิจัยจากทางโครงการส่งเสริมการวิจัยในอุดมศึกษา และการพัฒนามหาวิทยาลัยวิจัย (NRU) ของกระทรวงศึกษาธิการ และเงินทุนสนับสนุนจากคณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

วิทยานิพนธ์ฉบับนี้จะไม่สามารถประสบความสำเร็จลุล่วงไปได้ด้วยดีหากไม่ได้รับความอนุเคราะห์จาก ผศ. ดร.พรชัย มงคลนาม และ รศ. ดร.โจนาธาน โฮยิน ชาน ที่เสียสละเวลาอันมีค่ามาเพื่อให้คำปรึกษา และคำแนะนำแก่ผู้วิจัยเพื่อใช้ในการแก้ไขปัญหาต่าง ๆ ที่เกิดขึ้นในงานวิจัยชิ้นนี้จนสำเร็จไปได้ด้วยดี รวมถึงการเสียสละเวลาอันมีค่าเพื่อช่วยตรวจสอบ และแก้ไขจุดผิดพลาดของงานวิจัยให้ถูกต้องและมีความสมบูรณ์ ข้าพเจ้าจึงขอกราบขอบพระคุณในความอนุเคราะห์จากทั้งสองท่านเป็นอย่างสูงไว้ ณ โอกาสนี้

วิทยานิพนธ์ฉบับนี้จะไม่มีความสมบูรณ์หากไม่ได้รับความกรุณาจาก ผศ. ดร.ชาคริตา นุกุลกิจ ประธานกรรมการสอบวิทยานิพนธ์ ดร.วิฑิตา จงสุขชัยสิทธิ์ และ ดร.นล เปรมชัยเรือง กรรมการสอบวิทยานิพนธ์ ที่ช่วยให้คำแนะนำต่าง ๆ เพื่อทำให้วิทยานิพนธ์ฉบับนี้มีความสมบูรณ์มากยิ่งขึ้น ข้าพเจ้าจึงขอกราบขอบพระคุณในความอนุเคราะห์จากทางคณะกรรมการทุกท่านเป็นอย่างสูงไว้ ณ โอกาสนี้

สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณ บิดา มารดา และครอบครัวที่คอยสนับสนุนและให้กำลังใจในระหว่างการทำงานวิจัยชิ้นนี้ ซึ่งผู้วิจัยมีความหวังเป็นอย่างยิ่งที่จะนำประสบการณ์และความรู้ที่ได้รับจากการทำงานวิจัยครั้งนี้ไปประยุกต์ใช้เพื่อก่อให้เกิดประโยชน์แก่สังคมและประเทศชาติต่อไป โดยประโยชน์ที่เกิดจากงานวิจัยฉบับนี้จะไม่สามารถเกิดขึ้นได้หากไม่ได้รับความกรุณาจากบุคคลทุกท่านที่ได้กล่าวมาข้างต้น

สารบัญ

หน้า

บทคัดย่อภาษาไทย	๗
บทคัดย่อภาษาอังกฤษ	๘
กิตติกรรมประกาศ	๙
สารบัญ	๑๐
รายการตาราง	๑๑
รายการรูปประกอบ	๑๒

บทที่

1. บทนำ	1
1.1 ความสำคัญและที่มาของงานวิจัย	1
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ประโยชน์และผลที่ได้รับจากงานวิจัย	3
1.4 ขอบเขตงานวิจัย	3
1.5 ขั้นตอนการดำเนินงาน	4
1.6 โครงร่างของวิทยานิพนธ์	4
2. ทฤษฎี เทคโนโลยีและงานวิจัยที่เกี่ยวข้อง	5
2.1 Human Posture Analysis	5
2.1.1 Marker Motion Capture	5
2.1.2 Markerless Motion Capture	6
2.2 กล้อง Kinect	7
2.2.1 Kinect Library Function	8
2.2.2 สถาปัตยกรรม Kinect	11
2.3 Data Mining	12
2.3.1 ขั้นตอนในการทำเหมืองข้อมูล	13
2.3.2 Weka Data Mining Tool	13
2.4 งานวิจัยที่มีความเกี่ยวข้อง	14

3. วิธีการดำเนินงานวิจัย	24
3.1 ภาพรวมของงานวิจัย	25
3.2 Data Preprocessing	26
3.2.1 Training Set Detail	26
3.3 Data Selection and Data Transformation	29
3.3.1 Data Selection	30
3.3.2 Data Transformation	30
3.4 Model Learning	36
3.4.1 Classifier Algorithm	39
3.5 ภาพรวมของระบบ	40
4. ผลการวิจัย	46
4.1 รายละเอียดการทดลอง	46
4.1.1 Testing Set Detail	46
4.2 ผลลัพธ์จากการคัดเลือกและจัดลำดับข้อมูลโดยหลักการ Data Selection	48
4.3 ผลการทดลอง	50
5. สรุปผลการวิจัย	54
5.1 การอภิปรายสรุปผลการทดลอง	54
5.2 ข้อจำกัดของงานวิจัย	56
5.3 ข้อเสนอแนะ	57
5.4 แนวทางการพัฒนาต่อ	58
เอกสารอ้างอิง	61
ภาคผนวก	68
ก หนังสืออนุมติจริยธรรมการวิจัย	68
ข ผลงานที่ได้รับการตีพิมพ์ : Postural Classification using Kinect	73
ค ผลงานที่ได้รับการตีพิมพ์ : Multiple-stage Classification of Human Poses while Watching Television	80
ประวัติผู้วิจัย	88

รายการตาราง

ตาราง	หน้า
3.1 จำนวนชุดข้อมูลในแต่ละท่าทางของชุด Training Set	27
3.2 ความกว้างของไหล่และความสูงของร่างกายผู้ร่วมการทดลองในส่วนของ Training Set	27
4.1 จำนวนชุดข้อมูลในแต่ละท่าทางของชุด Testing Set	47
4.2 ความกว้างของไหล่และความสูงของร่างกายผู้ร่วมการทดลองในส่วนของ Testing Set	47
4.3 ผลลัพธ์ในการจัดลำดับของชุดข้อมูลโดยหลักการ Gain Ratio Attribute Evaluator	48
4.4 ผลลัพธ์ในการจัดลำดับของชุดข้อมูลโดยหลักการ Info Gain Attribute Evaluator	49
4.5 ค่าความแม่นยำจากการทดสอบชุดข้อมูลที่ไม่ได้ผ่านการทำ Data Selection และ Data Transformation	50
4.6 ค่าความแม่นยำจากการทดสอบชุดข้อมูลที่ได้ผ่านการทำ Data Selection โดยหลัก Gain Ratio Attribute Evaluator	51
4.7 ค่าความแม่นยำจากการทดสอบชุดข้อมูลที่ได้ผ่านการทำ Data Selection โดยหลัก Info Gain Attribute Evaluator	52
4.8 ค่าความแม่นยำจากการทดสอบชุดข้อมูลที่ได้ผ่านการทำ Data Transformation	53

รายการรูปประกอบ

รูป	หน้า
2.1 การทำสัญลักษณ์บนจุดต่าง ๆ บนร่างกาย	5
2.2 องค์ประกอบหลักของกล้อง Kinect	7
2.3 การเชื่อมต่อเพื่อส่งถ่ายข้อมูลระหว่างกล้อง Kinect และ โปรแกรมที่ถูกพัฒนาขึ้น	8
2.4 การวิเคราะห์จุดต่าง ๆ บนร่างกายโดย Kinect for Windows SDK Library Function	9
2.5 แกน X, Y และ Z ของกล้อง Kinect	10
2.6 โครงสร้างของสถาปัตยกรรมการส่งถ่ายข้อมูลของกล้อง Kinect	12
2.7 การตรวจจับร่างกายมนุษย์จากภาพสามมิติ	16
2.8 การแบ่งแยกท่าทางของสองบุคคลที่ร่างกายสัมผัสกัน	16
2.9 ตัวอย่างพฤติกรรมของเด็กที่บันทึกโดยใช้กล้อง Kinect	17
2.10 ระบบการลองเสื้อผ้าเสมือนจริง	18
2.11 ระบบการควบคุมหุ่นยนต์โดยใช้กล้อง Kinect ในการตรวจจับท่าทางของผู้ควบคุม	18
2.12 ระบบสภาพแวดล้อมเสมือนจริงเพื่อใช้ในการฝึกซ้อมเดิน	19
2.13 ระบบเพื่อฝึกสอนการเดินบัลเลต์	20
2.14 โครงร่างมนุษย์จากมุมมองในแต่ละกล้องและการประกอบโครงร่าง	20
2.15 การติดตามท่าทางขณะนั่งทำงาน	22
2.16 ท่าทางการตีกลองของภาพสองและสามมิติที่บันทึกได้จากกล้อง Kinect	23
2.17 การวิเคราะห์การเคลื่อนไหวบนใบหน้าเพื่อใช้ควบคุมแบบจำลองใบหน้าสามมิติ	24
3.1 ภาพรวมของขั้นตอนการทดลองในวิจัย	25
3.2 รายละเอียดท่าทางที่งานวิจัยให้ความสนใจ	26
3.3 ความผิดพลาดที่เกิดจากการวิเคราะห์ข้อมูลของกล้อง Kinect	28
3.4 อัตราส่วนระหว่างความกว้างและความสูงของร่างกายมนุษย์ในแต่ละท่าทาง	31
3.5 ขนาดของมุมมองของเข้าท่อนบนระหว่างทำยืนและทำนั่ง	32
3.6 ระยะห่างระหว่างเอวมายังพื้นที่แตกต่างกันในแต่ละท่าทาง	33
3.7 ลักษณะท่าทางของหลังในแต่ละแบบ	34
3.8 ผลต่างในแต่ละค่าตัวแปรที่ใช้ในการวิเคราะห์ท่าทางการยกมือ	35
3.9 Single-stage Classifier Model	36
3.10 Multiple-stage Classifier Model	36
3.11 โครงสร้างการทำงานของการทำงานของการวิเคราะห์ข้อมูลแบบ Multiple-stage Classifier Model	37

รูป	หน้า
3.12 รูปแบบชุดข้อมูลแบบต่าง ๆ ที่ใช้ในการทดลอง	38
3.13 หน้าจอการทำงานของโปรแกรมการจัดการชุดข้อมูลในหน้าจอการลบข้อมูล	40
3.14 หน้าจอการทำงานของโปรแกรมการจัดการชุดข้อมูลในหน้าของการดึงข้อมูลที่ถูกลบกลับสู่ชุดข้อมูลอีกครั้ง	41
3.15 หน้าจอการทำงานของโปรแกรมการจัดการชุดข้อมูลในหน้าสร้างไฟล์ CSV	41
3.16 การออกแบบ Class Diagram ในส่วนของโปรแกรมที่ใช้ในการจัดการชุดข้อมูล	42
3.17 หน้าจอการทำงานของโปรแกรมที่ใช้ในการบันทึกและวิเคราะห์ทำทางตามเวลาจริงในโหมดการบันทึกข้อมูล	43
3.18 หน้าจอการทำงานของโปรแกรมที่ใช้ในการบันทึกและวิเคราะห์ทำทางตามเวลาจริงในโหมดการวิเคราะห์ข้อมูลทำทาง	44
3.19 การออกแบบ Class Diagram ในส่วนของโปรแกรมบันทึกและวิเคราะห์ข้อมูลทำทาง	45
4.1 ช่วงระยะทางที่ใช้ในการทดลอง	46
5.1 ปัญหาจากการอยู่คนเดียวของผู้สูงอายุในไทยโดยจำแนกตามลักษณะการอยู่อาศัยในครัวเรือน	59

บทที่ 1 บทนำ

บทนี้ได้กล่าวถึงความเป็นมา วัตถุประสงค์ของงานวิจัย ขอบเขตของงานวิจัย ขั้นตอนการดำเนินงานวิจัย ประโยชน์ที่ได้รับและโครงสร้างของงานวิจัยดังนี้

1.1 ความสำคัญและที่มาของงานวิจัย

งานวิจัยด้านการวิเคราะห์ท่าทางของมนุษย์ (Human Postural Analysis) เป็นอีกสาขาวิจัยที่ได้รับการพัฒนาขึ้นอย่างต่อเนื่อง ซึ่งงานวิจัยด้านการวิเคราะห์ท่าทางของมนุษย์นั้นสามารถนำไปประยุกต์กับงานวิจัยได้หลากหลายสาขาเช่น การติดตามพฤติกรรมในกลุ่มผู้ป่วยหรือผู้สูงอายุจากการวิเคราะห์ท่าทางที่แสดงออกของสาขาวิทยาศาสตร์การแพทย์ (Medical Sciences) หรือการตรวจจับการเคลื่อนไหวของท่าทางนักกีฬาในระหว่างฝึกซ้อมของสาขาวิทยาศาสตร์การกีฬา (Sport Sciences) เป็นต้น หลักการที่ใช้วิเคราะห์ท่าทางสามารถแบ่งออกได้เป็นสองหลักการได้แก่ การตรวจจับท่าทางโดยทำสัญลักษณ์บนจุดต่าง ๆ ของร่างกาย (Marker Motion Capture) และการตรวจจับท่าทางโดยปราศจากการทำเครื่องหมายบนร่างกาย (Markerless Motion Capture) ซึ่งการตรวจจับท่าทางโดยทำสัญลักษณ์บนจุดต่าง ๆ ของร่างกายนั้นถูกพัฒนาขึ้นเพื่อตอบสนองการตรวจจับท่าทางที่ต้องการความละเอียดสูง เช่น การทำภาพเคลื่อนไหว (Animation) หรือ การศึกษาท่าทางการเคลื่อนไหวของนักกีฬา เป็นต้น แม้ว่าความละเอียดในการตรวจจับท่าทางการเคลื่อนไหวของหลักการการตรวจจับท่าทางโดยทำสัญลักษณ์บนจุดต่าง ๆ ของร่างกายนั้นจะมีความละเอียดสูงแต่ก็อาจไม่ใช่หลักการที่เหมาะสมกับทุกงานวิจัยเสมอไป เช่น ในงานวิจัยด้านการติดตามพฤติกรรมโดยการวิเคราะห์จากท่าทางการแสดงออกของมนุษย์อาจไม่เหมาะสมที่จะนำวิธีดังกล่าวมาประยุกต์ใช้ ด้วยสาเหตุที่การติดอุปกรณ์หรือเครื่องหมายไว้บนร่างกายของผู้ถูกตรวจจับตลอดเวลาอาจเป็นการขัดขวางการใช้ชีวิตประจำวันของบุคคลทั่วไป การตรวจจับท่าทางโดยปราศจากการทำเครื่องหมายบนร่างกายจึงเป็นอีกแนวทางหนึ่งที่สามารถช่วยให้การใช้ชีวิตประจำวันของผู้ถูกตรวจจับมีความสะดวกมากยิ่งขึ้น

ปัจจุบันเทคโนโลยีมีส่วนสำคัญในการช่วยให้งานวิจัยด้านการวิเคราะห์ท่าทางของมนุษย์มีความก้าวหน้ามากยิ่งขึ้น กล้อง Kinect เป็นอีกอุปกรณ์หนึ่งที่ได้รับความสนใจจากงานวิจัยหลากหลายสาขา กล้อง Kinect นั้นถูกพัฒนาขึ้นมาเพื่อเป็นอุปกรณ์เสริมของเครื่องเล่นเกม XBOX 360 โดยเป็นอุปกรณ์ควบคุมการเล่นเก้ด้วยท่าทาง และคำสั่งเสียงของผู้เล่น ซึ่งถูกพัฒนาขึ้นโดยบริษัท Microsoft มีองค์ประกอบในรับข้อมูลสามช่องทางด้วยกันได้แก่ กล้องรับภาพสองมิติ (RGB Image Camera) เซ็นเซอร์รับภาพเชิงลึก (Depth Image Sensor) และชุดไมโครโฟนที่ใช้ในการรับคำสั่งจากผู้ใช้งาน (Multi-array Microphone) โดยจุดเด่นของกล้อง Kinect คือมี Library Function ที่ใช้ในการแยก

โครงสร้างของร่างกายมนุษย์ออกจากพื้นหลังที่มีความซับซ้อน และนอกจากนี้ยังสามารถช่วยในการสกัดส่วนต่าง ๆ ของร่างกายเพื่อสร้างเป็นโครงร่างเสมือน (Skeleton) จึงเป็นสาเหตุให้กล้อง Kinect กลายเป็นอุปกรณ์ที่ได้รับความนิยมจากหลากหลายงานวิจัยในปัจจุบัน

เช่นเดียวกับอีกหลากหลายงานวิจัยที่ให้ความสนใจเลือกใช้กล้อง Kinect งานวิจัยชิ้นนี้เป็นอีกงานวิจัยที่เลือกใช้ข้อมูลที่วิเคราะห์ได้จากกล้อง Kinect มาประยุกต์กับหลักการการทำเหมืองข้อมูล (Data Mining) เพื่อใช้ในการแยกแยะท่าทาง โดยอัลกอริทึมที่ถูกเลือกใช้ในการเรียนรู้เพื่อสร้างโมเดลในการวิเคราะห์ท่าทางของงานวิจัยชิ้นนี้ประกอบด้วย Neural Network, Naïve Bayes, Logistic Regression และ Decision Tree โดยงานวิจัยชิ้นนี้จะนำเสนอหลักการในการลดมิติความซับซ้อนของชุดข้อมูลดิบที่สามารถวิเคราะห์ได้จากกล้อง Kinect ซึ่งจะส่งผลให้การเรียนรู้มีความผิดพลาดน้อยลง โดยหลักการที่ใช้ในการลดความซับซ้อนของชุดข้อมูลในงานวิจัยนี้นำเสนอมีด้วยกัน 2 หลักประกอบด้วย Data Selection และ Data Transformation เพื่อเป็นการเพิ่มประสิทธิภาพในการวิเคราะห์ท่าทางโดยใช้ข้อมูลจากกล้อง Kinect ให้มีความถูกต้องมากยิ่งขึ้น และนอกจากนี้ในงานวิจัยชิ้นนี้ยังได้พัฒนาระบบต้นแบบที่ใช้ในการวิเคราะห์ท่าทางที่สามารถบันทึกท่าทางต่าง ๆ ที่แสดงโดยผู้ถูกตรวจจับลงในฐานข้อมูลเพื่อใช้ในการศึกษาย้อนหลังซึ่งสามารถนำไปต่อยอดใช้เป็นระบบในการติดตามพฤติกรรมของมนุษย์ได้ต่อไป

1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อศึกษาหลักการประยุกต์ใช้ข้อมูลในรูปแบบสามมิติในการวิเคราะห์และแยกแยะท่าทางของมนุษย์ โดยข้อมูลที่ใช้ในการวิเคราะห์ท่าทางเป็นข้อมูลจากกล้อง Kinect ซึ่งเป็นอุปกรณ์บันทึกภาพสามมิติ
2. เพื่อศึกษาการนำหลักการการทำเหมืองข้อมูลมาประยุกต์ในการวิเคราะห์และแยกแยะท่าทางของมนุษย์
3. เพื่อศึกษาวิธีการในการลดความซับซ้อนของข้อมูลเพื่อใช้ในการเพิ่มศักยภาพในการแยกแยะท่าทางของมนุษย์ให้ดียิ่งขึ้น
4. ออกแบบและพัฒนาระบบต้นแบบที่สามารถใช้วิเคราะห์ท่าทาง และสามารถจัดเก็บข้อมูลที่วิเคราะห์ได้จากกล้อง Kinect เพื่อเป็นระบบต้นแบบสำหรับการติดตามพฤติกรรมและกิจกรรมต่าง ๆ ส่วนบุคคลต่อไป

1.3 ประโยชน์และผลที่ได้รับจากงานวิจัย

1. ระบบที่ถูกพัฒนาขึ้นมาสามารถนำไปใช้เป็นระบบที่ใช้ในการวิเคราะห์ท่าทางสำหรับการติดตามพฤติกรรมของผู้สูงอายุซึ่งเป็นกลุ่มบุคคลที่มีความจำเป็นที่จะต้องติดตามพฤติกรรม โดยจะเป็นประโยชน์ต่อการแพทย์เพื่อใช้ในการเพิ่มรายละเอียดในการเรียนรู้พฤติกรรมของผู้สูงอายุให้มากยิ่งขึ้น
2. ชุดข้อมูลท่าทางที่ได้จากกล้อง Kinect ของงานวิจัยชิ้นนี้สามารถนำไปประยุกต์ร่วมกับงานวิจัยด้านการวิเคราะห์การแสดงออกของอารมณ์บนใบหน้ามนุษย์ ซึ่งจะสามารถช่วยให้การตีความของการแสดงออกทางอารมณ์มีความชัดเจนมากยิ่งขึ้น และสามารถนำไปพัฒนาต่อยอดในงานวิจัยด้านการวิเคราะห์ความชื่นชอบในการรับชมรายการโทรทัศน์โดยวิเคราะห์จากการแสดงออกทางอารมณ์เป็นต้น
3. หลักการที่นำเสนอไว้ในงานวิจัยชิ้นนี้เป็นองค์ความรู้ที่สามารถช่วยลดทอนและคัดกรองปริมาณชุดข้อมูลลง เพื่อช่วยลดความซับซ้อนภายในชุดข้อมูลที่วิเคราะห์จากกล้อง Kinect ซึ่งยังคงเป็นชุดข้อมูลดิบ ทำให้การเรียนรู้เพื่อค้นหารูปแบบในการวิเคราะห์ข้อมูลมีความถูกต้องและแม่นยำมากยิ่งขึ้น และนอกจากนี้การลดทอนปริมาณชุดข้อมูลยังสามารถทำให้เวลาในการเรียนรู้ของอัลกอริทึมลดลงได้อีกด้วย

1.4 ขอบเขตงานวิจัย

1. Library Function ที่งานวิจัยนี้เลือกใช้คือ Microsoft Kinect for Windows SDK ซึ่งสามารถวิเคราะห์ข้อมูลโครงสร้างของร่างกายได้ทั้งหมด 20 จุด โดยเป็นเครื่องมือในการพัฒนาโปรแกรมตรวจจับท่าทางที่ทำงานร่วมกับกล้อง Kinect ซึ่งเป็น Library Function รุ่น Microsoft Kinect SDK 1.8
2. ระยะที่ใช้ในการทดสอบและบันทึกข้อมูลอยู่ระหว่าง 1.8 ถึง 3.0 เมตร
3. กล้อง Kinect ที่ใช้ในการทดลองเป็นกล้องรุ่น XBOX 360
4. คอมพิวเตอร์ที่สามารถใช้ได้กับกล้อง Kinect จะต้องมีความ RAM ไม่น้อยกว่า 2 Gigabyte และจะต้องมี Chipset ที่ให้ความเร็วไม่น้อยกว่า 2.0 Gigahertz

1.5 ขั้นตอนการดำเนินงาน

1. ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง
2. ศึกษาการทำงานของกล้อง Kinect เพื่อใช้ในการบันทึกข้อมูลและวิเคราะห์ท่าทาง
3. ออกแบบขั้นตอนและรายละเอียดท่าทางทั้งหมดที่จะใช้ในการทดลอง
4. ศึกษาทฤษฎีหลักการการทำเหมืองข้อมูล เพื่อนำมาประยุกต์กับงานวิจัยและเลือกโมเดลที่ใช้ในการแยกแยะท่าทาง
5. ออกแบบและพัฒนาระบบที่ใช้ในการวิเคราะห์ท่าทางซึ่งทำงานโดยการเชื่อมต่อกับกล้อง Kinect
6. นำเสนอแนวทางของงานวิจัยเพื่อขออนุมัติด้านจริยธรรมการวิจัย
7. บันทึกข้อมูลท่าทางของผู้ร่วมการทดลองเพื่อเป็นข้อมูลต้นแบบในการวิเคราะห์ข้อมูล
8. ทำการวิเคราะห์ข้อมูลที่ได้จากการบันทึกข้อมูลท่าทางต่าง ๆ ของผู้ร่วมการทดลอง
9. ศึกษาวิธีการลดทอนปริมาณข้อมูลเพื่อลดความซับซ้อนของชุดข้อมูลและเพิ่มความถูกต้องในการวิเคราะห์ท่าทาง
10. นำเสนอวิธีการในการลดความซับซ้อนของข้อมูลและวิธีการแยกแยะท่าทางต่าง ๆ
11. ออกแบบวิธีการทดสอบความถูกต้องในการแยกแยะท่าทาง
12. ทดสอบและบันทึกผลความถูกต้องในการวิเคราะห์ท่าทาง
13. สรุปและนำเสนอผลลัพธ์ที่ได้จากการทดลอง

1.6 โครงร่างของวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้มีส่วนประกอบดังต่อไปนี้

- บทที่ 1 บทนำ** ในบทนี้ได้กล่าวถึงความเป็นมา วัตถุประสงค์ของงานวิจัย ขอบเขตของงานวิจัย ขั้นตอนการดำเนินงานวิจัย ประโยชน์ที่ได้รับและโครงสร้างของงานวิจัยขึ้นนี้
- บทที่ 2 ทฤษฎี เทคโนโลยีและงานวิจัยที่เกี่ยวข้อง** ในบทนี้จะกล่าวถึงรายละเอียด ความรู้พื้นฐานของเทคโนโลยีและงานวิจัยที่มีความเกี่ยวข้องกับงานวิจัยขึ้นนี้
- บทที่ 3 วิธีการดำเนินงานวิจัย** ในบทนี้จะกล่าวถึงขั้นตอนของการวิเคราะห์ท่าทาง วิธีการลดทอนข้อมูลเพื่อเพิ่มความถูกต้องในการวิเคราะห์ท่าทาง รวมถึงการออกแบบระบบที่ใช้ในการวิเคราะห์ท่าทาง
- บทที่ 4 ผลการวิจัย** ในบทนี้จะกล่าวถึงความแม่นยำในการวิเคราะห์ท่าทางจากการทดสอบความถูกต้องของวิธีการวิเคราะห์ท่าทางที่ได้นำเสนอไว้
- บทที่ 5 สรุปผลการวิจัย** ในบทนี้จะกล่าวถึงการอภิปรายผลสรุป ข้อจำกัด ข้อเสนอแนะ และแนวทางในการนำงานวิจัยขึ้นนี้ไปประยุกต์ใช้และพัฒนาต่อยอด

บทที่ 2 ทฤษฎี เทคโนโลยีและงานวิจัยที่เกี่ยวข้อง

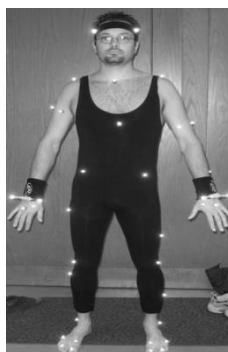
บทนี้จะกล่าวถึงรายละเอียด ความรู้พื้นฐานของเทคโนโลยีและงานวิจัยที่มีความเกี่ยวข้องกับงานวิจัย
ชิ้นนี้

2.1 Human Posture Analysis

การตรวจจับท่าทางการเคลื่อนไหวของร่างกายมนุษย์ (Human Postural Analysis) เป็นการตรวจจับท่าทางของมนุษย์ในอิริยาบถต่าง ๆ เพื่อทำการตีความแต่ละท่าทาง ซึ่งสามารถวิเคราะห์ได้จากท่าทางของอวัยวะในส่วนต่าง ๆ ของร่างกาย เช่น แขน ขา หรือ หัวเข่า เป็นต้น โดยหลักการในการตรวจจับการเคลื่อนไหวเพื่อใช้ในการวิเคราะห์ท่าทางสามารถแบ่งได้เป็นสองหลักการด้วยกันดังต่อไปนี้

2.1.1 Marker Motion Capture

การตรวจจับการเคลื่อนไหวโดยการทำสัญลักษณ์บนจุดต่าง ๆ ของร่างกาย [1-4] (Marker Motion Capture) เป็นการตรวจจับส่วนต่าง ๆ ของร่างกายโดยผู้ถูกตรวจจับจะต้องสวมใส่ชุด หรือ เครื่องแต่งกายที่มีการทำเครื่องหมายในส่วนต่าง ๆ เอาไว้ โดยเครื่องหมายที่ถูกกำหนดไว้บนร่างกายจะเป็นตัวบ่งบอกถึงอวัยวะส่วนต่าง ๆ เพื่อใช้ในการแบ่งแยกอวัยวะแต่ละส่วนที่ต้องการตรวจจับดังแสดงในรูปที่ 2.1 โดยวิธีสัณฐานต่าง ๆ บนร่างกายนั้นสามารถทำได้ทั้งจากการบันทึกภาพผ่านกล้องแล้วนำข้อมูลที่บันทึกได้มาทำการวิเคราะห์หาสัญลักษณ์ที่ทำไว้บนร่างกาย หรือ ในอีกกรณีคือการใช้อุปกรณ์ที่เป็นอุปกรณ์บอกพิกัดซึ่งสามารถช่วยให้รู้พิกัดของส่วนต่าง ๆ บนร่างกายได้ทันทีในระหว่างที่ทำการบันทึกข้อมูลการเคลื่อนไหว



รูปที่ 2.1 การทำสัญลักษณ์บนจุดต่าง ๆ ของร่างกาย [5]

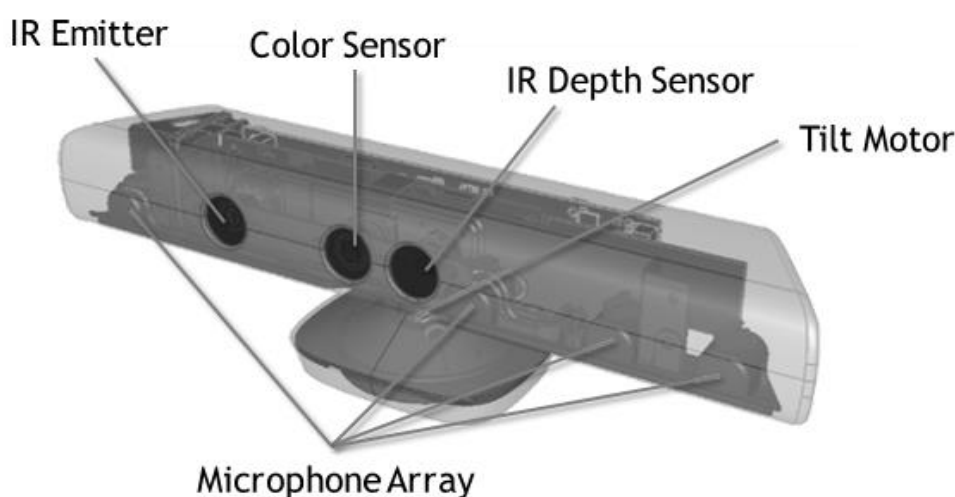
2.1.2 Markerless Motion Capture

การตรวจจับท่าทางโดยปราศจากการทำเครื่องหมายบนร่างกาย (Markerless Motion Capture) เป็นอีกหลักการของการวิเคราะห์การแสดงท่าทางของมนุษย์ หลักการนี้จะไม่มีการทำสัญลักษณ์หรือเครื่องหมายไว้บนจุดต่าง ๆ ของร่างกายของผู้ถูกตรวจจับ โดยในหลักการนี้จะบันทึกภาพการเคลื่อนไหวมาวิเคราะห์ เพื่อทำการแยกแยะส่วนที่คาดการณ์ว่าจะเป็นร่างกายมนุษย์ออกจากพื้นหลัง จากนั้นจึงนำข้อมูลส่วนที่เป็นร่างกายมนุษย์จากภาพมาวิเคราะห์ท่าทาง ในหลักการนี้สามารถแยกย่อยวิธีการในการวิเคราะห์ท่าทางได้ 2 แบบคือ การวิเคราะห์ข้อมูลจากภาพสองมิติ (2 Dimension Image Processing) และการวิเคราะห์ข้อมูลจากภาพสามมิติ (3 Dimension Image Processing) สำหรับการวิเคราะห์ข้อมูลจากภาพสองมิตินั้นเป็นหลักการที่ได้รับความนิยมมาอย่างยาวนาน [6-8] โดยจะใช้รายละเอียดจุดสี (Pixel Color) ของภาพมาทำการวิเคราะห์ท่าทาง แต่อย่างไรก็ตามพบว่าการใช้ข้อมูลจากภาพสองมิตินั้นมักมีปัญหาในเรื่องการแยกแยะส่วนของร่างกายมนุษย์ออกจากพื้นหลังที่มีความซับซ้อน ซึ่งส่งผลกระทบต่อความถูกต้องในการวิเคราะห์ท่าทางตามไปด้วย นอกจากนี้แสงก็ยังเป็นอีกปัจจัยที่ส่งผลกระทบต่อคุณภาพในการวิเคราะห์ข้อมูลท่าทางอีกเช่นกัน ในทางกลับกันการวิเคราะห์ข้อมูลจากภาพสามมิติสามารถช่วยลดปัญหาดังกล่าวของวิธีการวิเคราะห์ข้อมูลจากภาพสองมิติลงได้ ด้วยสาเหตุที่ชุดข้อมูลในภาพสามมิตินั้นมีส่วนของข้อมูลเชิงลึก (Depth Data) ซึ่งช่วยในการแบ่งแยกระหว่างวัตถุแต่ละชิ้นให้มีความชัดเจนมากยิ่งขึ้น นอกจากนี้อุปกรณ์ที่ใช้ในการบันทึกภาพเชิงลึกนั้นเป็นอุปกรณ์ประเภทเซ็นเซอร์ (Sensor) จึงทำให้แสงไม่ได้เป็นปัจจัยสำคัญในการวิเคราะห์ข้อมูล ดังนั้นการวิเคราะห์ข้อมูลจากภาพสามมิติจึงเป็นอีกหนึ่งแนวทางที่หลายงานวิจัยในปัจจุบันให้ความสนใจ [9-11] เช่นเดียวกับหลายงานวิจัยในปัจจุบันงานวิจัยนี้ผู้วิจัยได้เลือกใช้หลักการในการตรวจจับท่าทางโดยปราศจากการทำเครื่องหมายบนร่างกาย โดยใช้วิธีการวิเคราะห์ข้อมูลจากภาพสามมิติในการวิเคราะห์ท่าทางของมนุษย์

อุปกรณ์เป็นหนึ่งในปัจจัยหลักที่ส่งผลกระทบต่อความถูกต้องในการวิเคราะห์ท่าทาง ซึ่งพบว่าในปัจจุบันมีหลายงานวิจัยให้ความสนใจเลือกกล้อง Kinect เพื่อใช้ในงานวิจัย โดยระบบของกล้อง Kinect มีความสามารถในการบันทึกได้ทั้งข้อมูลภาพสองมิติ (RGB Image) และ ภาพสามมิติ (Depth Image) โดยจุดเด่นของกล้อง Kinect คือมี Library Function ที่ช่วยในการวิเคราะห์และทำการแยกแยะส่วนที่เป็นร่างกายมนุษย์ออกจากพื้นหลังที่มีความซับซ้อนได้เป็นอย่างดี และ Library Function ของกล้อง Kinect นั้นยังสามารถช่วยในการวิเคราะห์เพื่อค้นหาจุดสำคัญต่าง ๆ ของร่างกายได้อีกด้วย จากความสามารถของกล้อง Kinect จึงทำให้ปัจจุบันกล้อง Kinect กลายเป็นอุปกรณ์ที่ได้รับความนิยมจากหลายสาขางานวิจัย

2.2 กล้อง Kinect

กล้อง Kinect [12-13] ถูกพัฒนาขึ้นเพื่อใช้เป็นอุปกรณ์ต่อพ่วงเสริมในการเพิ่มความสามารถให้กับเครื่องเล่นเกม XBOX 360 ซึ่งเป็นผลิตภัณฑ์ของบริษัท Microsoft โดยกล้อง Kinect นั้นทำให้เครื่องเล่นเกม XBOX 360 นั้นมีความสามารถในการควบคุมการเล่นเกมโดยสามารถสั่งการผ่านทางท่าทางเคลื่อนไหวและเสียงคำสั่งของผู้เล่น ระบบกล้อง Kinect ได้ถูกพัฒนาภายใต้แนวคิด “คุณคือผู้ควบคุม” (You Are the Controller) โดยการพัฒนาของบริษัท PrimeSense ซึ่งเป็นผู้พัฒนางานด้าน Natural User Interface (NUI) [14] กล้อง Kinect นั้นมีองค์ประกอบหลัก 4 ส่วนด้วยกันดังแสดงในรูปที่ 2.2



รูปที่ 2.2 องค์ประกอบหลักของกล้อง Kinect [15]

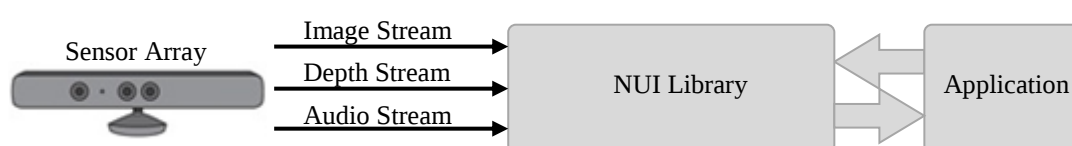
1. **Multi-array Microphone** เป็นไมโครโฟนที่ใช้ในการรับเสียงคำสั่งจากผู้ใช้งาน โดยจะถูกจัดวางเรียงยาวตามแนวนอนของตัวกล้อง Kinect เป็นจำนวน 4 ตัวโดยรับสัญญาณในรูปแบบอนาล็อกและทำการเปลี่ยนรูปแบบสัญญาณเป็นดิจิทัล (ADC: Analog-to-digital Converter) แบบ 24 บิต (Bit) เพื่อนำไปใช้ในการวิเคราะห์ชุดคำสั่งบนระบบทำงานของเครื่องเล่นเกม XBOX 360
2. **Depth Sensor** ถูกแบ่งออกเป็นสองส่วนด้วยกันได้แก่ IR Emitter และ IR Depth Sensor ซึ่งอยู่ด้านขวาและซ้ายของ RGB Camera ตามลำดับ โดย IR Emitter จะทำหน้าที่ในการยิงลำแสงไปยังวัตถุที่อยู่หน้ากล้อง Kinect และ IR Depth Sensor นั้นจะเป็นเซ็นเซอร์ที่ใช้ในการรับลำแสงที่สะท้อนกลับจากวัตถุที่อยู่หน้ากล้อง Kinect และนำไปคำนวณระยะห่างของวัตถุแต่ละชิ้นที่อยู่หน้ากล้อง Kinect โดยอุปกรณ์ทั้งสองชิ้นนี้จะทำให้กล้อง Kinect มีคุณสมบัติที่สามารถตรวจจับ

วัตถุหน้ากล้องได้ทั้งในพื้นที่สว่างและพื้นที่มืด โดยจะทำงานร่วมกันเพื่อวิเคราะห์และสร้างภาพเชิงลึกเพื่อใช้ในการวิเคราะห์และค้นหาร่างกายมนุษย์ต่อไป

3. **RGB Camera** เป็นกล้องที่ใช้ในการบันทึกภาพสองมิติ อยู่ระหว่าง IR Emitter และ IR Depth Sensor โดยภาพที่ได้จากการบันทึกของกล้อง Kinect จะมีความละเอียดได้สูงสุดอยู่ที่ 1280x960 พิกเซล (Pixel)
4. **Motorized Tilt** เป็นมอเตอร์ที่ถูกบรรจุไว้ที่ฐานของกล้องเพื่อใช้ในการปรับมุมมองกล้องให้สามารถก้มและเงย เพื่อให้มีมุมมองที่เหมาะสมในการใช้งานในสถานการณ์ และสถานที่แตกต่างกันออกไป โดยสามารถปรับให้กล้องมีมุมเงยและมุมก้มได้ 27 องศา

2.2.1 Kinect Library Function

Kinect Library Function คือตัวกลาง (Middleware) ที่ทำหน้าที่ในการติดต่อเพื่อดึงข้อมูลจากกล้อง Kinect ไปใช้ในโปรแกรมที่ทำการพัฒนาขึ้นดังแสดงในรูปที่ 2.3 ภาษาทางโปรแกรมที่สามารถใช้งานได้กับ Kinect Library Function นั้นมีด้วยกันสามภาษาคือ ภาษา C++ ภาษา C# และ ภาษา Java โดย Kinect Library Function นั้นเป็น Library ชนิด Natural User Interface Library กล้อง Kinect มี Library หลากหลายตัวที่สามารถใช้ในการสนับสนุนการติดต่อเพื่อดึงข้อมูลจากกล้อง Kinect โดย Library ที่ได้รับความนิยมด้วยกันสองตัวได้แก่ Kinect for Windows SDK Library Function และ OpenNI Library Function โดยงานวิจัยชิ้นนี้เลือกใช้ Kinect for Windows SDK Library Function ซึ่งเป็น Library ที่ถูกพัฒนาขึ้นโดยตรงจากบริษัท Microsoft



รูปที่ 2.3 การเชื่อมต่อเพื่อส่งถ่ายมูลระหว่างกล้อง Kinect และ โปรแกรมที่ถูกพัฒนาขึ้น [16]

การตรวจจับการเคลื่อนไหวของร่างกายมนุษย์นั้นจำเป็นต้องมี Library เพื่อช่วยในการวิเคราะห์ท่าทาง และค้นหาส่วนที่เป็นร่างกายมนุษย์ ซึ่ง Kinect for windows SDK นั้นสามารถวิเคราะห์ส่วนต่าง ๆ บนร่างกายได้สองโหมดซึ่งประกอบด้วย [17] Seat Mode และ Default Mode โดย Default Mode สามารถค้นหาจุดต่าง ๆ บนร่างกายได้ทั้ง 20 จุดดังแสดงในรูปที่ 2.4.1 และ Seat Mode สามารถค้นหาจุดต่าง ๆ บนร่างกายได้ทั้งหมด 10 จุด เฉพาะส่วนของแขนและเหนือไหล่ขึ้นไปดังแสดงในรูปที่ 2.4.2 ซึ่ง Seat Mode นั้นมีความสามารถในการวิเคราะห์ท่าทางในการนั่งได้ดีแต่อย่างไรก็ตาม

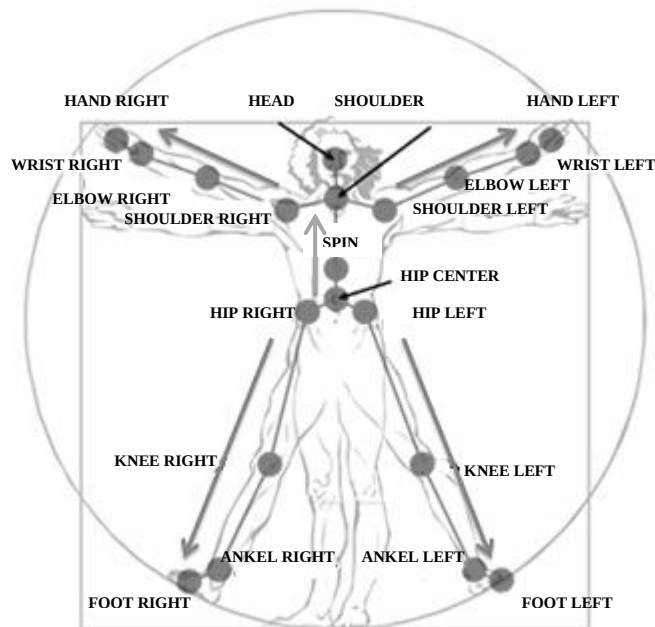
รายละเอียดของข้อมูลส่วนที่ต่ำกว่าไหล่ลงมาจะไม่สามารถนำมาใช้ได้ ดังนั้นในงานวิจัยชิ้นนี้จึงเลือกใช้ Default Mode เพื่อวิเคราะห์การแสดงท่าทาง โดยรายละเอียดของจุดต่าง ๆ บนร่างกายที่สามารถวิเคราะห์ได้จากกล้อง Kinect [18] นั้นได้ถูกแสดงไว้ในรูปที่ 2.4.3



รูปที่ 2.4.1 การวิเคราะห์จุดต่าง ๆ บนร่างกายใน
Default Mode [17]



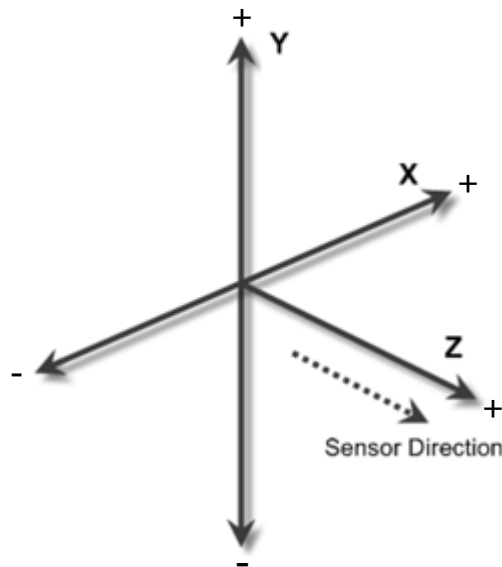
รูปที่ 2.4.2 การวิเคราะห์จุดต่าง ๆ บนร่างกายใน
Seat Mode [17]



รูปที่ 2.4.3 รายละเอียดของจุดต่าง ๆ บนร่างกายที่ถูกวิเคราะห์โดยกล้อง Kinect [18]

รูปที่ 2.4 การวิเคราะห์จุดต่าง ๆ บนร่างกายโดย Kinect for Windows SDK Library Function

ข้อมูลตำแหน่งของจุดต่าง ๆ บนร่างกายที่ได้จาก Kinect for Windows SDK Library Function ของกล้อง Kinect นั้นจะอยู่ในพิกัดสามมิติ ซึ่งประกอบด้วยพิกัดในแกน X, Y และ Z โดยแกน X จะเป็นแกนในแนวนอน แกน Y จะเป็นแกนในแนวตั้ง ซึ่งตัดผ่านจุดศูนย์กลางซึ่งเป็นจุดกำเนิด ($X = Y = Z = 0$) และแกน Z เป็นแกนที่ใช้ในการบ่งบอกถึงระยะห่างระหว่างกล้อง Kinect ไปยังวัตถุ ซึ่งจุดกำเนิดนั้นจะเป็นตำแหน่งที่ใช้ในการอ้างอิงตำแหน่งของวัตถุต่าง ๆ ที่อยู่ตรงหน้ากล้อง Kinect ค่าในแนวแกน X จะเป็นค่าที่บอกระยะห่างจากจุดกำเนิดในแนวนอน ซึ่งค่าในแนวแกน X จะมีค่าเป็นบวกก็ต่อเมื่อค่าในจุดนั้นอยู่ด้านขวาของแกน Y และค่าในแกน Y นั้นจะมีค่าเป็นบวกก็ต่อเมื่อค่าอยู่เหนือแกน X ขึ้นไป ส่วนในแนวแกน Z จะมีค่าเป็นบวกเสมอซึ่งหากจุดของวัตถุนั้นออกห่างจากกล้องมากขึ้นจะทำให้ค่าในแนวแกน Z นั้นค่อย ๆ มีค่าที่เพิ่มขึ้นตามไปด้วย ซึ่งทิศทางของแกน X Y และ Z ถูกแสดงไว้ในรูปที่ 2.5



รูปที่ 2.5 แกน X Y และ Z ของกล้อง Kinect [19]

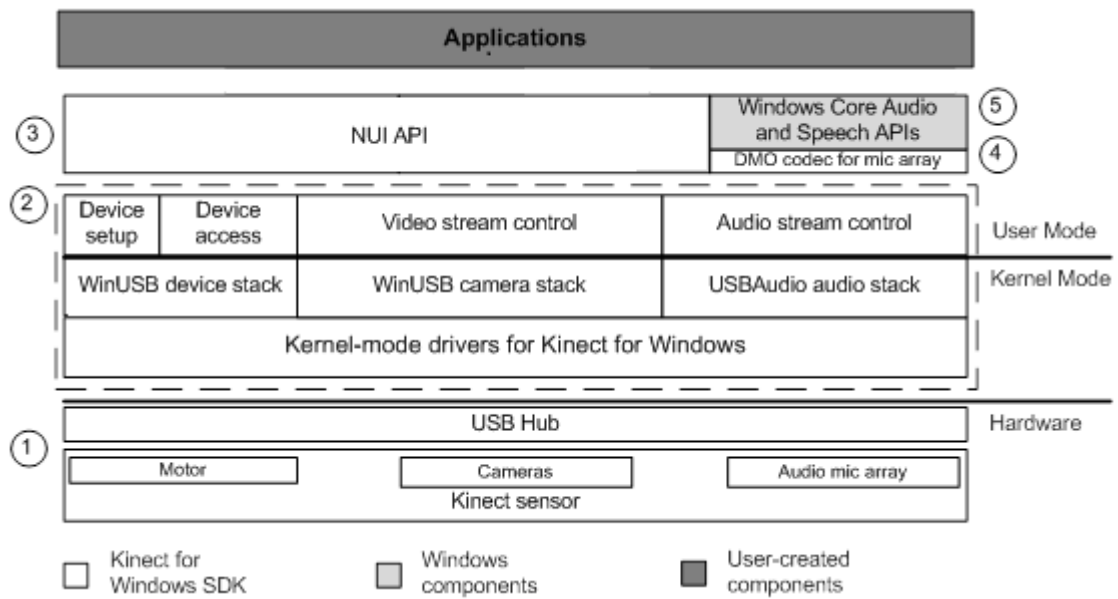
ระยะห่างของกล้องไปยังร่างกายมนุษย์ที่เหมาะสมในการทำงานของ Kinect for Windows SDK Library Function คือช่วงระยะระหว่าง 1.8-3.0 เมตร แม้ว่า Kinect for Windows SDK Library Function นั้นจะมีความสามารถในการแยกแยะส่วนที่เป็นร่างกายคนออกจากพื้นหลังที่มีความซับซ้อนและสามารถวิเคราะห์หาจุดต่าง ๆ บนร่างกายได้ ยังคงพบว่า Kinect for Windows SDK Library Function มีจุดบกพร่องอยู่เช่นกันดังนี้

1. Kinect for Windows SDK Library Function ไม่สามารถจะทำการวิเคราะห์และคาดคะเนส่วนของร่างกายที่ถูกบดบังด้วยวัตถุอื่นได้อย่างสมบูรณ์ ซึ่งข้อมูลที่ได้จากการวิเคราะห์นั้นอาจเป็นข้อมูลที่เกิดจากการคาดเดาที่อาจถูกต้องหรือผิดพลาดก็ได้
2. การเคลื่อนไหวที่รวดเร็ว หรือ การทำท่าทางที่มีความซับซ้อนส่งผลให้ Library Kinect for Windows SDK ไม่สามารถตรวจจับท่าทาง หรือ อาจเกิดการคาดเดาจุดต่าง ๆ ที่ผิดเพี้ยนขึ้นได้
3. ในกรณีที่ขา หรือ แขนชิดติดกัน Library Kinect for Windows SDK อาจมีการวิเคราะห์จุดต่าง ๆ ที่ผิดพลาดมากกว่าเมื่อเทียบกับการแยกแขนและขาออกจากกัน เพื่อให้เกิดช่องว่างระหว่างแขนและลำตัว และขาทั้งสองข้าง เพราะ Kinect for Windows SDK Library Function จะวิเคราะห์ว่าแขนและลำตัวเป็นชิ้นส่วนของร่างกายส่วนเดียวกัน

2.2.2 สถาปัตยกรรม Kinect

รูปแบบของสถาปัตยกรรมในการส่งผ่านข้อมูลของกล้อง Kinect และโปรแกรมที่ถูกพัฒนาขึ้นมีส่วนประกอบ 5 ชั้น ดังต่อไปนี้ (รูปที่ 2.6)

1. **Kinect Hardware** คือส่วนที่เป็นวงจรอิเล็กทรอนิกส์ของกล้อง Kinect ที่ใช้ในการตรวจจับภาพและรับสัญญาณเสียง ซึ่งมีการส่งข้อมูลผ่านสายสัญญาณชนิด USB
2. **Kinect Driver** เป็นตัวขับเคลื่อน (Driver) ซึ่งทำหน้าที่เป็นตัวสั่งการให้กล้องสามารถทำงานได้ตามที่ต้องการ และเป็นตัวช่วยให้คอมพิวเตอร์สามารถเข้าถึงทรัพยากรของกล้อง Kinect ได้ โดยตัวขับเคลื่อนจะมาพร้อมกับ Kinect for Windows SDK Library Function โดยมีองค์ประกอบย่อย 3 ส่วนดังต่อไปนี้
 - 2.1 **Kinect Microphone Array** เป็นส่วนที่ทำให้สามารถควบคุมการใช้งานอุปกรณ์ไมโครโฟนของกล้อง Kinect ได้
 - 2.2 **Audio and Video Streaming** เป็นส่วนที่ใช้ในการดึงข้อมูลเสียง และข้อมูลภาพจากกล้อง Kinect
 - 2.3 **Device Enumeration Functions** เป็นส่วนช่วยให้โปรแกรมที่พัฒนาขึ้นสามารถเข้าถึงกล้อง Kinect ได้มากกว่าหนึ่งเครื่อง
3. **Audio and Video Components** เป็นส่วนที่ใช้ในการติดต่อเพื่อตรวจจับโครงสร้างของร่างกายมนุษย์ทั้งในรูปแบบข้อมูลภาพสองมิติ และข้อมูลภาพเชิงลึก รวมถึงการเข้าใช้งานชุดข้อมูลเสียง
4. **DirectX Media Object (DMO)** เป็นส่วนที่ใช้ในการจัดการคุณภาพข้อมูลเสียงของกล้อง Kinect
5. **Windows 7 Standard APIs** เป็นส่วนที่จัดการเรื่องเสียงและการสั่งการโดยเสียงเพื่อให้สามารถใช้งานไบนารีระบบปฏิบัติการ Windows 7 และ Windows 8



รูปที่ 2.6 โครงสร้างของสถาปัตยกรรมการส่งถ่ายข้อมูลของกล้อง Kinect [16]

2.3 Data Mining

การทำเหมืองข้อมูล (Data Mining) [20-21] คือกระบวนการในการวิเคราะห์ข้อมูลขนาดใหญ่ที่มีความซับซ้อนยากต่อการวิเคราะห์ และการทำความเข้าใจ โดยจะมีการค้นหารูปแบบ (Pattern) และความสัมพันธ์ (Relationship) ในชุดข้อมูลเพื่อให้สามารถจัดการ และวิเคราะห์ชุดข้อมูลได้ง่ายยิ่งขึ้น ซึ่งในปัจจุบันมีหลายสาขาวิจัยนำหลักการด้านการทำเหมืองข้อมูลไปประยุกต์เพื่อช่วยในการวิเคราะห์แยกแยะ และหาความสัมพันธ์ทางข้อมูล เช่น การวิเคราะห์การแสดงออกทางอารมณ์บนหน้า (Facial Emotional Expression Recognition) และการแยกแยะท่าทางของมนุษย์ (Human Postural Recognition) เป็นต้น โดยวิวัฒนาการของการทำเหมืองข้อมูลถูกแบ่งเป็น 4 ช่วงด้วยกันซึ่งได้แก่

- ปี ค.ศ.1960 Data Collection เป็นช่วงที่มีการพัฒนาระบบจัดเก็บข้อมูลที่มีความเหมาะสมบนอุปกรณ์ที่น่าเชื่อถือและมีความสามารถในการป้องกันการสูญหายได้เป็นอย่างดี
- ปี ค.ศ.1980 Data Access เป็นช่วงที่เริ่มมีการนำข้อมูลที่จัดเก็บไว้มาทำการสร้างความสัมพันธ์ต่อกันเพื่อประโยชน์ในการวิเคราะห์ข้อมูล
- ปี ค.ศ. 1990 Data Warehouse & Decision Support คือช่วงที่มีการรวบรวมข้อมูลที่แยกออกจากกัน มาเก็บรวมกันในฐานะข้อมูลที่มีขนาดใหญ่เพื่อให้เห็นภาพรวมของข้อมูล และช่วยสนับสนุนการตัดสินใจต่าง ๆ
- ปี ค.ศ. 2000 Data Mining เป็นช่วงที่ปริมาณข้อมูลที่ใช้ในการวิเคราะห์มีขนาดใหญ่ขึ้น และมีความซับซ้อนส่งผลให้ยากต่อวิเคราะห์ข้อมูล จึงมีการนำข้อมูลมาวิเคราะห์และทำการประมวลผล

โดยการสร้างแบบจำลอง (Model) และหาความสัมพันธ์ทางสถิติ เพื่อช่วยให้เข้าใจชุดข้อมูลได้ง่ายยิ่งขึ้น

2.3.1 ขั้นตอนในการทำเหมืองข้อมูล

ขั้นตอนในการทำเหมืองข้อมูลประกอบด้วย 7 ขั้นตอนต่อไปนี้

1. **Data Cleaning** เป็นขั้นตอนในการวิเคราะห์ข้อมูลเพื่อคัดเลือกว่าเป็นข้อมูลที่ผิดพลาด (Error Data) ออกจากชุดข้อมูลดิบเพื่อลดความผิดพลาดที่จะเกิดขึ้นในการวิเคราะห์ข้อมูล
2. **Data Integration** เป็นขั้นตอนในการรวบรวมข้อมูลในกรณีที่ชุดข้อมูลนั้นมีมากกว่าหนึ่งชุดและแยกกันอยู่ให้เป็นข้อมูลชุดเดียวกัน
3. **Data Selection** เป็นขั้นตอนในการประเมินความสำคัญของแต่ละคุณลักษณะ (Attribute) ของชุดข้อมูลเพื่อแยกแยะข้อมูลที่มีความสำคัญและไม่มีความสำคัญในการวิเคราะห์ข้อมูลออกจากกัน
4. **Data Transformation** เป็นขั้นตอนในการแปลงชุดข้อมูลดิบ (Raw Data) เพื่อลดความซับซ้อนในชุดข้อมูลซึ่งส่งผลให้ชุดข้อมูลมีปริมาณที่น้อยลง และเพิ่มความสัมพันธ์ให้แก่ชุดข้อมูลส่งผลให้การเรียนรู้เพื่อค้นหาความสัมพันธ์ของชุดข้อมูลมีความถูกต้องมากยิ่งขึ้น
5. **Data Mining** เป็นขั้นตอนการเรียนรู้ (Learning) ชุดข้อมูลเพื่อค้นหาความรูปแบบความสัมพันธ์ภายในชุดข้อมูล
6. **Pattern Evaluation** เป็นขั้นตอนการประเมินความสามารถของรูปแบบที่ได้จากการวิเคราะห์ข้อมูลในขั้นตอนการทำเหมืองข้อมูล
7. **Knowledge Representation** เป็นขั้นตอนการนำเสนอองค์ความรู้ที่ค้นพบจากชุดข้อมูล โดยใช้เทคนิคนำเสนอในรูปแบบแผนภาพเพื่อช่วยให้สามารถเข้าใจถึงรูปแบบที่ได้จากการวิเคราะห์ข้อมูลมากยิ่งขึ้น

2.3.2 Weka Data Mining Tool

โปรแกรม Weka [22-23] (Waikato Environment for Knowledge Analysis) เป็นซอฟต์แวร์ฟรีที่รวบรวมอัลกอริทึมเพื่อสนับสนุนการเรียนรู้ของเครื่อง (Machine Learning) และในการทำเหมืองข้อมูล ซึ่งถูกพัฒนาขึ้นโดยมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ โดยผู้ใช้งานสามารถสั่งการผ่านหน้า GUI (Graphic User Interface) และใช้ชุดคำสั่ง (Command Line) ในการประมวลผลชุดข้อมูล โปรแกรม Weka เป็นโปรแกรมที่พัฒนาขึ้นบนพื้นฐานของภาษาจาวา (Java) ซึ่งทำให้ระบบสามารถทำงานได้บนหลากหลายระบบปฏิบัติการ โดยโปรแกรม Weka มีชุดเครื่องมือที่มีฟังก์ชันสำหรับการทำงานร่วมกับข้อมูล 5 ส่วนด้วยกันซึ่งประกอบด้วย

1. **Pre-Processing** เป็นส่วนที่ใช้ในการเลือกไฟล์ข้อมูลที่ต้องการใช้ในการวิเคราะห์ และเป็นส่วนที่ทำการสรุปรายละเอียดพื้นฐานของชุดข้อมูลที่จะทำการวิเคราะห์
2. **Classify** เป็นส่วนที่ใช้ในการวิเคราะห์ข้อมูลด้วยวิธีการจำแนกข้อมูล (Classification) และทำนายข้อมูล (Prediction) ซึ่งมีหลากหลายอัลกอริทึมที่ใช้ในการวิเคราะห์ข้อมูล เช่น Neural Network, Decision Tree และ Naive Bayes เป็นต้น
3. **Cluster** เป็นส่วนที่ใช้ในการวิเคราะห์ข้อมูลด้วยวิธีการจัดกลุ่มชุดข้อมูล (Clustering) โดยมีการจัดกลุ่มข้อมูลตามความสัมพันธ์ของข้อมูลที่มีลักษณะคล้ายกันเข้าไว้ด้วยกัน
4. **Select Attributes** เป็นส่วนที่ใช้ในการวิเคราะห์หากกลุ่มของข้อมูลที่มีความสำคัญและจำเป็นต่อการวิเคราะห์ข้อมูลเพื่อช่วยเพิ่มความถูกต้องในการวิเคราะห์ข้อมูลให้มากยิ่งขึ้น นอกจากนี้ยังช่วยลดปริมาณข้อมูลที่ไม่มีความเกี่ยวข้องต่อการวิเคราะห์ข้อมูลลงได้อีกด้วย
5. **Visualize** เป็นส่วนที่ใช้นำเสนอความสัมพันธ์ของชุดข้อมูลในรูปแบบแผนภาพเพื่อให้เห็นความสัมพันธ์ในแต่ละส่วนของชุดข้อมูล

2.4 งานวิจัยที่มีความเกี่ยวข้อง

ในช่วง ค.ศ. 2006 งานวิจัยของ M. C. Roh และ S. W. Lee [24] ได้นำเสนอวิธีลดความผิดพลาดในการค้นหาและแยกแยะส่วนที่เป็นร่างกายมนุษย์ออกจากพื้นหลังของภาพสองมิติที่มีคุณภาพต่ำ ซึ่งเป็นหนึ่งในปัญหาหลักของการวิเคราะห์เพื่อค้นหาร่างกายมนุษย์จากภาพสองมิติ โดยในงานวิจัยชิ้นนี้ได้เสนอหลักการในการตัดส่วนเกินของพื้นหลังที่ติดมาจากการค้นหาร่างกายมนุษย์ โดยได้ทำการทดสอบผลลัพธ์โดยการเปรียบเทียบกับหลักการค้นหาส่วนที่เป็นร่างกายมนุษย์แบบ CDP (Continuous Dynamic Programming) ซึ่งใช้ชุดข้อมูลในการทดสอบเป็นวิดีโอบันทึกการแข่งขันกีฬาพบว่าหลักการที่งานวิจัยนี้เสนอสามารถทำการค้นหาร่างกายมนุษย์ได้สมบูรณ์กว่าหลักการในแบบ CDP

งานวิจัยของ Y. M. Lianga และคณะ [25] ได้นำเสนอวิธีการตีความหมายของท่าทางของมนุษย์ โดยเสนอกรอบการทำงาน (Framework) ในการวิเคราะห์ความหมายของท่าทาง ซึ่งโครงสร้างของกรอบการทำงานของงานวิจัยชิ้นนี้ประกอบด้วยสองส่วนด้วยกันได้แก่ ส่วนของชุดท่าทางที่ใช้ในการตีความของท่าทาง (Key Post) และส่วนของการวิเคราะห์ความหมายท่าทาง โดยจะนำข้อมูลท่าทางที่สามารถวิเคราะห์ได้มาทำการตีความหมาย ซึ่งในงานวิจัยชิ้นนี้ได้ทำการพัฒนาชุดท่าทางขึ้นมาทั้งหมด 61 ท่าทางเพื่อใช้ในการตีความของการแสดงท่าทาง

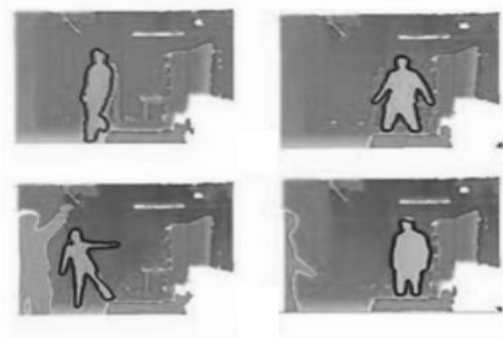
C. C. Chen และคณะ [26] ได้นำเสนอวิธีในการค้นหาส่วนของร่างกายมนุษย์และสัปดาห์ต่าง ๆ ของร่างกาย (Body Part Segmentation) จากภาพสองมิติโดยใช้เทคนิค Gaussian Mixture Model ในการแยกแยะและค้นหาส่วนต่าง ๆ ของร่างกายด้วยหลักการ Depth-first Search จากการทดลองกับภาพทำขึ้นจำนวน 140 ภาพ ซึ่งเป็นภาพที่มีพื้นหลังไม่ซับซ้อนพบว่าด้วยหลักการที่งานวิจัยนี้ได้นำเสนอไว้สามารถแยกแยะแต่ละส่วนของร่างกายได้เป็นอย่างดี

ในช่วง ค.ศ. 2007 งานวิจัยของ C. Wu และ H. Aghajan [27] นำเสนอวิธีการวิเคราะห์ท่าทางของมนุษย์โดยใช้กล้องบันทึกภาพแบบสามมิติ โดยระบบที่ถูกพัฒนาขึ้นนั้นสามารถสนับสนุนการใช้กล้องบันทึกสามมิติหลายตัวในการบันทึกข้อมูล โดยข้อมูลจากแต่ละมุมมองของกล้องสามมิติแต่ละตัวนั้นสามารถช่วยประกอบโครงสร้างร่างกายของมนุษย์ให้มีความถูกต้องมากยิ่งขึ้น นอกจากนี้งานวิจัยชิ้นนี้ยังได้นำเสนอกรอบการทำงาน (Framework) ซึ่งได้กล่าวถึงการกำหนดตัวแปรสำคัญที่มีผลกระทบต่อวิธีการวิเคราะห์ท่าทางในระบบสามมิติซึ่งประกอบด้วย พื้นที่ในการบันทึกข้อมูล เวลา และ Feature ที่ใช้ในการวิเคราะห์ข้อมูลท่าทางของมนุษย์ที่สามารถบันทึกได้ในรูปแบบสามมิติ

ในช่วงปี ค.ศ. 2010 เป็นปีที่กล้อง Kinect รุ่น Kinect for XBOX ถูกจัดวางจำหน่ายสู่ท้องตลาดซึ่งเป็นอุปกรณ์ที่ถูกพัฒนาโดยบริษัท Microsoft โดยมี Library ที่ช่วยสนับสนุนการวิเคราะห์การเคลื่อนไหวของร่างกายมนุษย์ โดยกล้อง Kinect รุ่นแรกถูกพัฒนาขึ้นมาเพื่อเป็นอุปกรณ์เสริมที่ใช้ในการควบคุมการเล่นเกมของเครื่องเล่นเกม XBOX แต่อย่างไรก็ตามด้วยความสามารถในการวิเคราะห์ร่างกายมนุษย์ในระบบสามมิติที่มีความถูกต้องสูงทำให้ภายหลังจากปี ค.ศ. 2010 มีงานวิจัยจำนวนมากให้ความสนใจเลือกใช้กล้อง Kinect เป็นอุปกรณ์ในการตรวจจับท่าทาง

ในช่วงปี ค.ศ. 2011 งานวิจัยของ J. Shotton และคณะ [28] ซึ่งเป็นทีมนักวิจัยของบริษัท Microsoft ได้นำเสนอวิธีการค้นหาส่วนที่เป็นร่างกายมนุษย์ และการแยกแยะส่วนต่าง ๆ ของร่างกายจากภาพสามมิติ โดยหลักการที่นำเสนอสามารถช่วยให้วิเคราะห์ข้อมูลภาพสามมิติได้อย่างรวดเร็วด้วยปริมาณ 200 เฟรมต่อวินาที

งานวิจัยของ L. Xia และคณะ [29] ได้นำเสนอหลักการในการตรวจจับร่างกายของมนุษย์โดยใช้ข้อมูลจากภาพเชิงลึกของกล้อง Kinect ดังแสดงในรูปที่ 2.7 ซึ่งพบว่าข้อดีของกล้อง Kinect คืออุปกรณ์ที่ใช้ในการบันทึกภาพสามมิตินั้นเป็นเซ็นเซอร์อินฟราเรด ทำให้แสงไม่มีผลต่อการวิเคราะห์ข้อมูล จึงทำให้สามารถใช้กล้อง Kinect ได้ทั้งในที่มืดและที่สว่าง นอกจากนี้ยังพบว่าเสื้อผ้าที่สวมใส่ รวมถึงพื้นหลังที่มีความซับซ้อน ไม่มีผลต่อการวิเคราะห์ข้อมูล งานวิจัยนี้จึงแสดงให้เห็นว่าข้อมูลภาพเชิงลึกนั้นสามารถนำมาทดแทนการวิเคราะห์ข้อมูลจากภาพสองมิติได้เป็นอย่างดี



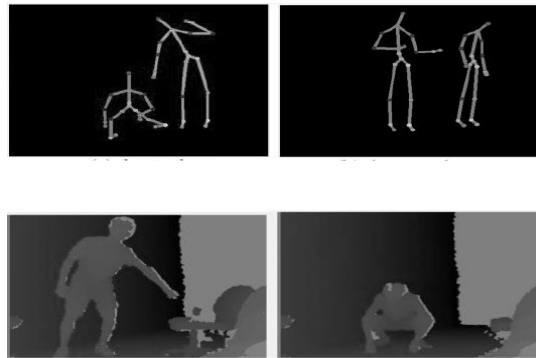
รูปที่ 2.7 การตรวจจับร่างกายมนุษย์จากภาพสามมิติ [29]

งานวิจัยของ Y. Liu และคณะ [30] ได้นำเสนอวิธีการแบ่งแยกบุคคลสองคนที่มีส่วนของร่างกายสัมผัสกัน เช่น การตรวจจับท่าทางของกลุ่มเต้นแต่ละคน โดยวิธีที่ใช้ในงานวิจัยนี้คือการแบ่งแยกชิ้นส่วนในรูปภาพ (Image Segmentation) ซึ่งมีการแบ่งชิ้นส่วนของรูปภาพจากหลายมุมมอง และแต่ละมุมมองจะถูกนำไปประมาณหาพื้นผิว (Surface Estimation) เพื่อสร้างแบบจำลองหุ่นสามมิติดังแสดงในรูปที่ 2.8 ซึ่งผลลัพธ์แสดงถึงวิธีการที่มีความน่าเชื่อถือในการตรวจสอบการเคลื่อนไหวของคนสองคนที่มีร่างกายสัมผัสกันโดยใช้กล้อง Kinect แต่อย่างไรก็ตามยังพบความผิดพลาดในกรณีที่มีการเคลื่อนไหวที่รวดเร็ว และในกรณีที่ส่วนของร่างกายทั้งสองคนมีความใกล้ชิดกันมากเกินไปเช่นการกอด



รูปที่ 2.8 การแบ่งแยกท่าทางของสองบุคคลที่ร่างกายสัมผัสกัน [30]

งานวิจัยของ X. Yu และคณะ [31] ได้ศึกษาและนำเสนอวิธีการวิเคราะห์และประเมินการแสดงออกทางอารมณ์ด้านลบในเด็ก โดยการตรวจจับการเคลื่อนไหวของท่าทางโดยใช้กล้อง Kinect ในการตรวจจับ ดังแสดงไว้ในรูปที่ 2.9 โดยทำการวิเคราะห์อารมณ์ด้านลบจากท่าทางที่แสดงออกของเด็ก ซึ่งจะนำข้อมูลของข้อต่อบนโครงร่างที่ตรวจจับได้จากกล้อง Kinect ไปทำการวิเคราะห์โดยใช้ อัลกอริทึมการวิเคราะห์พฤติกรรม ในงานวิจัยชิ้นนี้ได้ทำการทดลองพัฒนาระบบต้นแบบมาเปรียบเทียบกับระบบบันทึกการเคลื่อนไหว และการบันทึกพฤติกรรมที่เด็กแสดงออกมาจริง ผลสรุปแสดงให้เห็นถึงความสามารถของกล้อง Kinect ว่าเป็นอุปกรณ์ที่สามารถวิเคราะห์โครงสร้างของร่างกายได้อย่างละเอียดส่งผลให้การวิเคราะห์ท่าทางมีความถูกต้องและมีความแม่นยำ จึงทำให้การวิเคราะห์พฤติกรรม และการแสดงออกทางอารมณ์ด้านลบของเด็กมีความถูกต้องสูงตามไปด้วย



รูปที่ 2.9 ตัวอย่างพฤติกรรมของเด็กที่บันทึกโดยใช้กล้อง Kinect [31]

งานวิจัยของ S. Hauswiesner และคณะ [32] ได้นำเสนอระบบในการทดลองเสื้อผ้าเสมือนจริง ซึ่งใช้กล้อง Kinect ในการตรวจจับร่างกาย และท่าทางของผู้ใช้งาน เพื่อนำไปทำการสร้างแบบจำลองสามมิติที่สามารถจำลองร่างกายของผู้ใช้งานที่กำลังใส่เสื้อผ้าที่เลือกไว้ โดยแบบจำลองจะแสดงท่าทางที่เหมือนกับท่าทางที่ผู้ใช้งานกำลังแสดงอยู่ตามเวลาจริงดังแสดงในรูปที่ 2.10 โดยผลลัพธ์แสดงให้เห็นว่ากล้อง Kinect มีความสามารถในการแยกแยะร่างกายมนุษย์ออกจากพื้นหลังได้เป็นอย่างดี ซึ่งเมื่อนำระบบนี้ไปทดสอบกับผู้ใช้งานจริง พบว่าผู้ใช้งานมีความพอใจกับระบบที่ถูกพัฒนาขึ้น



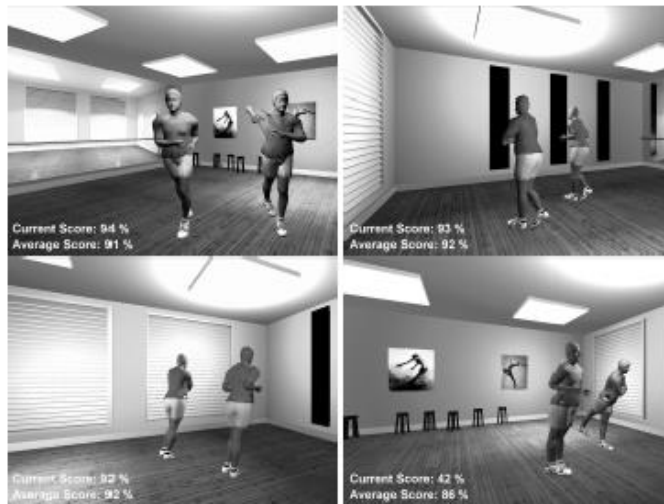
รูปที่ 2.10 ระบบการลองเสื้อผ้าเสมือนจริง [32]

งานวิจัยของ R. Igorevich และคณะ [33] ได้ทำการพัฒนาระบบที่ใช้ควบคุมการเคลื่อนไหวของหุ่นยนต์โดยท่าทางของมนุษย์ ซึ่งผู้ใช้งานสามารถควบคุมให้หุ่นยนต์แสดงท่าทางแบบเดียวกับผู้ควบคุมได้ดังแสดงในรูปที่ 2.11 โดยใช้กล้อง Kinect เป็นอุปกรณ์ในการตรวจจับท่าทางของผู้ใช้งาน โดยทำงานร่วมกับ OpenNI Library Function ในการวิเคราะห์โครงสร้างของร่างกายผู้ควบคุม ข้อมูลที่ใช้ในการควบคุมการเคลื่อนไหวของหุ่นยนต์มาจากขนาดของมุมระหว่างข้อต่อในส่วนต่าง ๆ ของร่างกายผู้ควบคุม ซึ่งข้อมูลเหล่านี้ได้มาจากการวิเคราะห์ข้อมูลของกล้อง Kinect โดยในงานวิจัยนี้ได้ทำการทดลองตรวจสอบความถูกต้องโดยให้ผู้ควบคุมแสดงท่าทางต่าง ๆ เพื่อควบคุมหุ่นยนต์ ผลลัพธ์จากการทดลองพบว่า ข้อมูลจากกล้อง Kinect ที่ทำงานร่วมกับ OpenNI Library Function นั้นสามารถใช้ในการควบคุมการแสดงท่าทางของหุ่นยนต์ได้เป็นอย่างดีซึ่งท่าทางที่หุ่นยนต์แสดงออกมาเหมือนกับท่าทางของผู้ควบคุม แต่อย่างไรก็ตามระบบที่ถูกพัฒนาขึ้นมายังเป็นระบบต้นแบบที่ใช้ในการควบคุมหุ่นยนต์จึงทำให้การแสดงท่าทางของหุ่นยนต์นั้นจะแสดงท่าทางที่ช้ากว่าท่าทางจริงของผู้ควบคุมประมาณ 200 มิลลิวินาที



รูปที่ 2.11 ระบบการควบคุมหุ่นยนต์โดยใช้กล้อง Kinect ในการตรวจจับท่าทางของผู้ควบคุม [33]

งานวิจัยของ D. Alexiadis และคณะ [34] ได้ทำการพัฒนาระบบที่ใช้ในการฝึกซ้อมการเต้น ซึ่งผู้ฝึกซ้อมสามารถฝึกซ้อมการเต้นตามนักเต้นต้นแบบ โดยระบบจะทำการจำลองสภาพแวดล้อมที่เสมือนจริงในรูปแบบสามมิติดังแสดงในรูปที่ 2.12 ซึ่งงานวิจัยชิ้นนี้ใช้กล้อง Kinect ร่วมกับ OpenNI Library Function ในการวิเคราะห์จุดต่าง ๆ ของร่างกาย เพื่อนำไปประเมินความถูกต้องของท่าทางการเต้นของผู้ฝึกซ้อม โดยเปรียบเทียบจุดต่าง ๆ ที่วิเคราะห์ได้จากกล้อง Kinect กับข้อมูลของผู้เต้นต้นแบบ โดยผลการศึกษาพบว่าระบบที่ถูกพัฒนาขึ้นมานั้นสามารถทำงานได้ตามเวลาจริง และข้อมูลจุดต่าง ๆ บนร่างกายที่วิเคราะห์ได้จากกล้อง Kinect นั้นสามารถนำไปใช้การวิเคราะห์ความถูกต้องของท่าเต้นได้จริง



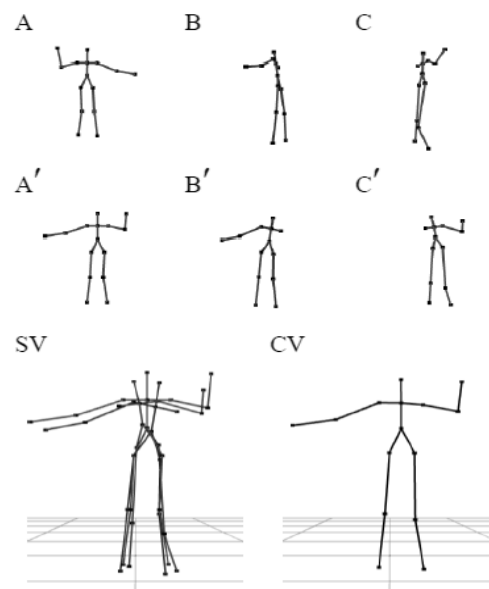
รูปที่ 2.12 ระบบสภาพแวดล้อมเสมือนจริงเพื่อใช้ในการฝึกซ้อมเต้น [34]

ในช่วงปี ค.ศ.2012 งานวิจัยของ Z. Marquardt และคณะ [35] ได้ทำการพัฒนาระบบฝึกซ้อมการเต้น ซึ่งระบบนี้มีชื่อว่า Super Mirror ถูกใช้ในสอนการเต้นบัลเลต์ ผู้ฝึกซ้อมสามารถฝึกซ้อมกับระบบเพื่อดูความถูกต้องของท่าทางการเต้นของผู้ฝึกซ้อม โดยระบบจะทำหน้าที่เป็นเหมือนกระจกที่สะท้อนท่าทางของผู้ฝึกซ้อมให้ผู้ฝึกซ้อมได้เห็น นอกจากนี้ระบบที่ถูกพัฒนาขึ้นสามารถทำการบันทึกข้อมูลการเต้นของผู้ฝึกซ้อม เพื่อใช้ในการติดตามท่าทางการเต้นของผู้ฝึกซ้อมย้อนหลัง ช่วยให้ผู้ฝึกซ้อมสามารถศึกษาการเต้นที่ผิดพลาดของตนเองได้ โดยในงานวิจัยนี้มีการทดสอบระบบโดยการแสดงท่าเต้นดังแสดงในรูปที่ 2.13 โดยกลุ่มนักเต้นบัลเลต์เพื่อทดสอบความถูกต้องในการวิเคราะห์ท่าทางของระบบที่ถูกพัฒนาขึ้น โดยผลลัพธ์จากการทดสอบพบว่าระบบที่ถูกพัฒนาขึ้นมาสามารถตรวจจับท่าทางการเต้นของผู้ทดสอบในระดับพื้นฐาน และระดับกลางได้อย่างถูกต้อง



รูปที่ 2.13 ระบบฝึกสอนการเต้นบัลเลต์ [35]

งานวิจัยของ นราวุฒิ พัฒโนทัย และคณะ [36] ได้พัฒนาระบบที่ใช้ในการตรวจจับการเคลื่อนไหวของมนุษย์โดยใช้กล้อง Kinect หลายตัวเพื่อเพิ่มมุมมองในการตรวจจับท่าทางการเคลื่อนไหวของมนุษย์ ดังแสดงในรูปที่ 2.14 โดยนำเสนอหลักการในการนำข้อมูลจากกล้องทุกตัวมาทำการประกอบเป็นโครงร่างเพียงโครงร่างเดียว ระบบที่ถูกพัฒนาขึ้นมาในงานวิจัยชิ้นนี้สามารถเพิ่มปริมาณของกล้อง Kinect ได้อย่างไม่จำกัดในทางทฤษฎี โดยจุดในการวางกล้อง Kinect นั้นจะต้องมีองศาระหว่างกล้องแต่ละตัวเท่า ๆ กัน โดยจุดเด่นของงานวิจัยชิ้นนี้คือระบบที่ถูกพัฒนาขึ้นสามารถทำให้การวิเคราะห์โครงสร้างของร่างกายนั้นมีความถูกต้องมากกว่าการใช้กล้อง Kinect ตัวเดียวซึ่งมักจะวิเคราะห์ผิดในกรณีที่บางส่วนของร่างกายถูกบดบัง



รูปที่ 2.14 โครงร่างมนุษย์จากมุมมองแต่ละกล้องและการประกอบโครงร่าง [36]

งานวิจัยของ O. Patsadu และคณะ [37] นำเสนอหลักการในการแยกแยะการแสดงท่าทางพื้นฐานของมนุษย์ซึ่งประกอบด้วย ท่านั่ง ท่านอนราบ และ ท่ายืน โดยในงานวิจัยชิ้นนี้ใช้ Kinect for Windows SDK Library Function ซึ่งสามารถให้ข้อมูลจุดต่าง ๆ ของร่างกายได้ทั้งหมด 20 จุด โดยงานวิจัยชิ้นนี้ประยุกต์หลักการการทำเหมืองข้อมูลมาช่วยวิเคราะห์ท่าทางโดยเลือก 4 อัลกอริทึมที่ใช้ในการเรียนรู้ชุดข้อมูลเพื่อค้นหารูปแบบที่ใช้ในการแยกแยะท่าทางซึ่ง 4 อัลกอริทึมที่เลือกใช้ประกอบด้วย Neural Network, Naïve Bayes, Decision Tree และ Support Vector Machine ชุดข้อมูลที่ใช้ในการเรียนรู้ของแต่ละอัลกอริทึมคือจุดต่าง ๆ บนร่างกายทั้งหมด 20 จุดที่วิเคราะห์ได้จากกล้อง Kinect ซึ่งพบว่าหลักการการทำเหมืองข้อมูลเป็นหลักการที่ช่วยแยกแยะท่าทางพื้นฐานได้เป็นอย่างดี

งานวิจัยของ M. Kepski และ B. Kwolek [38] นำเสนอระบบที่ใช้ในการตรวจจับการล้มโดยใช้หลายอุปกรณ์เพื่อเพิ่มความถูกต้องในการวิเคราะห์การล้มให้มีความละเอียดมากยิ่งขึ้น ซึ่งอุปกรณ์ที่ใช้ในการตรวจจับของงานวิจัยนี้ประกอบด้วยกล้อง Kinect, Accelerometer และ Gyroscope โดยระบบที่ถูกพัฒนาขึ้นมาทำงานอยู่บน PandaBoard ซึ่งเป็นอุปกรณ์ที่มีราคาถูกและทำงานบนระบบปฏิบัติการ Android โดยหลักการที่ใช้ในการช่วยวิเคราะห์การล้มคือหลักการ Fuzzy Inference ท่าทางที่งานวิจัยให้ความสนใจสามารถแบ่งได้สองส่วนคือ ท่าที่เป็นการล้มและท่าทางปกติโดยในงานวิจัยนี้ได้ทำการทดสอบระบบกับอาสาสมัครจำนวน 5 คน ซึ่งเป็นการทดสอบโดยให้อาสาสมัครแสดงท่าทางปกติและท่าทางการล้มเพื่อทดสอบความถูกต้อง ผลลัพธ์พบว่าระบบที่ถูกพัฒนาขึ้นนั้นสามารถตรวจจับท่าทางต่าง ๆ และแจ้งการล้มที่เกิดขึ้นได้เป็นอย่างดี

งานวิจัยของ A. Ball และคณะ [39] นำเสนอหลักการในการจับกลุ่มมนุษย์ โดยดูจากรูปแบบการเคลื่อนที่ ซึ่งใช้ข้อมูลจากจุดต่าง ๆ บนร่างกายที่สามารถวิเคราะห์ได้จากกล้อง Kinect ในงานวิจัยนี้ได้ใช้ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของค่าของมุมที่เกิดจากขาส่วนล่างและขาส่วนบนเพื่อใช้ในการวิเคราะห์ความเป็นไปได้ในการตรวจจับกลุ่มของมนุษย์ ซึ่งงานวิจัยนี้แสดงให้เห็นว่าชุดข้อมูลจุดต่าง ๆ บนร่างกายที่สามารถวิเคราะห์ได้จากกล้อง Kinect เป็นข้อมูลที่มีความเสถียรสามารถนำมาใช้ในการคำนวณเพื่อช่วยในการวิเคราะห์ลักษณะการเคลื่อนไหวของร่างกายได้

งานวิจัยของ Y. Liu และคณะ [40] ได้นำเสนอหลักการในการแยกแยะท่ายืนรูปแบบต่าง ๆ ซึ่งประกอบด้วย การยืนหันหน้าเข้าหากล้อง Kinect และ การยืนหันซ้ายและขวา ในมุม 30 และ 60 องศา เพื่อทดสอบความถูกต้องในการแยกแยะท่าทางการยืนในมุมมองต่าง ๆ โดยใช้หลักการการทำเหมืองข้อมูลในการแยกแยะข้อมูลท่าทางการยืน ซึ่งอัลกอริทึมที่ใช้ในการเรียนรู้เพื่อค้นหารูปแบบในการแยกแยะท่าทางที่งานวิจัยนี้เลือกใช้คือ Support Vector Machines ข้อมูลที่ใช้ในการเรียนรู้ของงานวิจัยชิ้นนี้คือมุมการเอียงตัวจากข้อมูลจุดไหล่ซ้ายและขวาที่สามารถวิเคราะห์ได้จากกล้อง Kinect ซึ่งพบว่า

ระบบสามารถแยกแยะท่าขึ้นในมุมมองต่าง ๆ ได้เป็นอย่างดีโดยมีความถูกต้องในการวิเคราะห์แต่ละท่าทางมากกว่า 85 เปอร์เซนต์

ในช่วงปี ค.ศ.2013 งานวิจัยของสุทธิพงษ์ แก่นจันทร์ และคณะ [41] ได้พัฒนาระบบที่ใช้ในการตรวจจับการเคลื่อนไหวของมนุษย์โดยใช้กล้อง Kinect หลายตัวเพื่อเพิ่มมุมมองในการตรวจจับท่าทางการเคลื่อนไหวของมนุษย์ โดยนำงานวิจัยของนราวุฒิ พัฒโนทัย และคณะ [38] มาทำการพัฒนาต่อ ยอดเพื่อให้สมบูรณ์ให้มากยิ่งขึ้น โดยงานวิจัยของนราวุฒิ พัฒโนทัย และคณะ นั้นมีข้อบกพร่องที่มีความจำเป็นที่จะต้องจัดวางกล้องให้มีมุมของกล้องแต่ละตัวห่างเท่า ๆ กัน โดย สุทธิพงษ์ แก่นจันทร์ และคณะ ได้พัฒนาระบบรวบรวมชุดข้อมูลจากกล้องหลายตัวเพื่อประกอบเป็นหนึ่งโครงสร้างร่างกายที่สามารถตั้งกล้องได้อย่างอิสระ โดยจะเลือกกล้องหลักที่จะให้ข้อมูลดีที่สุดเป็นกล้องอ้างอิงข้อมูล โดยใช้หลักการ Brute Force ในการคำนวณหาระยะของกล้องแต่ละตัวเพื่อหาทิศทางความสัมพันธ์ และหาระยะของกล้องตัวอื่น ๆ เพื่อใช้ในการลดขั้นตอนการจัดวางกล้อง Kinect แต่ละตัว และทำให้การใช้งานระบบมีความสะดวกมากยิ่งขึ้น

A. Uribe-Quevedo และคณะ [42] ได้นำเสนอระบบที่ใช้ในการติดตามท่านั่งทำงาน โดยวิเคราะห์มุมการเอียงของไหล่ และศีรษะดังแสดงในรูปที่ 2.15 เพื่อใช้ในการตรวจจับพฤติกรรมการนั่งของคนทำงาน โดยใช้ในการตรวจสอบสุขภาพและคุณภาพชีวิตของแต่ละบุคคล โดยข้อมูลที่ใช้ในการวิเคราะห์พฤติกรรมของงานวิจัยชิ้นนี้มาจากข้อมูลของกล้อง Kinect ซึ่งพบว่าข้อมูลจากกล้อง Kinect นั้นสามารถนำมาวิเคราะห์ท่าทางการเคลื่อนไหวในระยะใกล้ได้เป็นอย่างดี



รูปที่ 2.15 การติดตามท่าทางขณะนั่งทำงาน [42]

งานวิจัยของ K. Kumada และคณะ [43] ได้นำเสนอระบบที่ใช้ในการตรวจจับท่าทางในการตีกอล์ฟ เพื่อใช้ในฝึกซ้อมท่าทางในตีกอล์ฟให้ถูกต้อง โดยผู้ฝึกซ้อมสามารถนำท่าทางการตีกอล์ฟที่ถูกบันทึกไว้ในระบบมาทำการดูย้อนหลังเพื่อศึกษาความผิดพลาดของผู้ฝึกซ้อมได้ ซึ่งผู้ฝึกซ้อมสามารถเลือกดูได้ทั้งข้อมูลจากภาพสองมิติ และข้อมูลจากภาพสามมิติดังแสดงในรูปที่ 2.16 นอกจากนี้ระบบที่ถูกพัฒนาขึ้นยังสามารถแยกแยะแต่ละท่าการตีของนักกอล์ฟได้อีกด้วย โดยงานวิจัยชิ้นนี้ได้ทดสอบความถูกต้องในการตรวจจับท่าทางของการตีกอล์ฟโดยนักเล่นกอล์ฟจำนวน 20 คน ซึ่งจะแสดงท่าตีกอล์ฟหน้ากล้อง Kinect ซึ่งพบว่าระบบสามารถบันทึก และแยกแยะท่าทางของการตีกอล์ฟได้เป็นอย่างดี



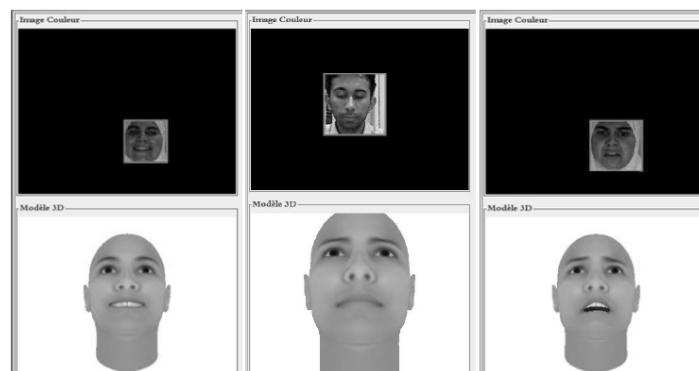
รูปที่ 2.16 ท่าทางการตีกอล์ฟของภาพสองและสามมิติที่บันทึกได้จากกล้อง Kinect [43]

งานวิจัยของ X. Dai และคณะ [44] นำเสนอระบบที่ใช้ในการตรวจจับและแบ่งแยกระหว่างการล้มและ ท่าทางปกติ (ADLs: Activities of Daily Living) โดยงานวิจัยชิ้นนี้ใช้หลักการ Machine Learning ในการค้นหารูปแบบเพื่อใช้ในการแยกแยะระหว่างการล้มและท่าทางปกติ ซึ่งอัลกอริทึมที่ใช้วิเคราะห์ข้อมูลการล้มของงานวิจัยชิ้นนี้คือ Hidden Markov Models โดยข้อมูลที่ใช้ในการเรียนรู้เป็นข้อมูลจากกล้อง Kinect ซึ่งจากการทดสอบระบบสามารถแยกแยะระหว่างการล้มและท่าทางปกติได้เป็นอย่างดี แสดงให้เห็นว่าเห็นข้อมูลที่สามารถวิเคราะห์ได้จากกล้อง Kinect มีความละเอียดซึ่งส่งผลให้การแยกแยะท่าทางของมนุษย์มีความถูกต้องสูงตามไปด้วย

นอกจากกล้อง Kinect จะเป็นอุปกรณ์ที่มีความสามารถในการวิเคราะห์ท่าทางของมนุษย์ได้เป็นอย่างดีแล้ว ยังพบว่ามียานงานวิจัยให้ความสนใจเลือกกล้อง Kinect เพื่อใช้ในการตรวจสอบการแสดงออกบนหน้ามนุษย์ เช่นงานวิจัยของ J. Qiu และคณะ [45] ได้นำเสนอรูปแบบการวิเคราะห์การเคลื่อนไหวบนใบหน้า โดยงานวิจัยชิ้นนี้ได้ใช้ข้อมูลภาพเชิงลึกจากกล้อง Kinect ซึ่งทำให้การแยกแยะใบหน้าของมนุษย์ออกจากพื้นหลังมีความถูกต้องและละเอียดมากยิ่งขึ้น ซึ่งผลลัพธ์ของงานวิจัยชิ้นนี้แสดงให้เห็นว่าเมื่อเปรียบเทียบคุณภาพของการวิเคราะห์โครงร่างบนใบหน้าโดยใช้ข้อมูลเชิงลึกจากภาพสามมิติจะสามารถช่วยเพิ่มความถูกต้องให้กับการวิเคราะห์โครงร่างบนหน้ามากกว่าการวิเคราะห์จากภาพ

สองมิติเพียงอย่างเดียว ซึ่งในกรณีของภาพสองมิตินั้นมักเกิดการวิเคราะห์ที่ผิดพลาดเมื่อเริ่มมีการเคลื่อนไหวหรือหันไปหน้า

งานวิจัยของ B. Seddik และคณะ [46] ได้นำเสนอวิธีการตรวจจับจุดต่าง ๆ ที่มีการเคลื่อนไหวบนใบหน้าเพื่อใช้ในการศึกษาความถูกต้องของการนำจุดต่าง ๆ บนใบหน้าไปใช้ในการควบคุมการแสดงออกบนใบหน้าของแบบจำลองใบหน้าสามมิติ ซึ่งในงานวิจัยชิ้นนี้เลือกใช้ข้อมูลจากภาพเชิงลึกช่วยในการวิเคราะห์จุดต่าง ๆ บนใบหน้าจากกล้อง Kinect โดยผลสรุปของงานวิจัยนี้พบว่าข้อมูลภาพเชิงลึกที่ได้จากกล้อง Kinect สามารถช่วยวิเคราะห์จุดต่าง ๆ บนใบหน้าเพื่อใช้ในการควบคุมใบหน้าจำลอง 3 มิติได้อย่างถูกต้องและมีประสิทธิภาพตามรูปที่ 2.17



รูปที่ 2.17 การวิเคราะห์การเคลื่อนไหวบนใบหน้าเพื่อใช้ควบคุมแบบจำลองใบหน้าสามมิติ [46]

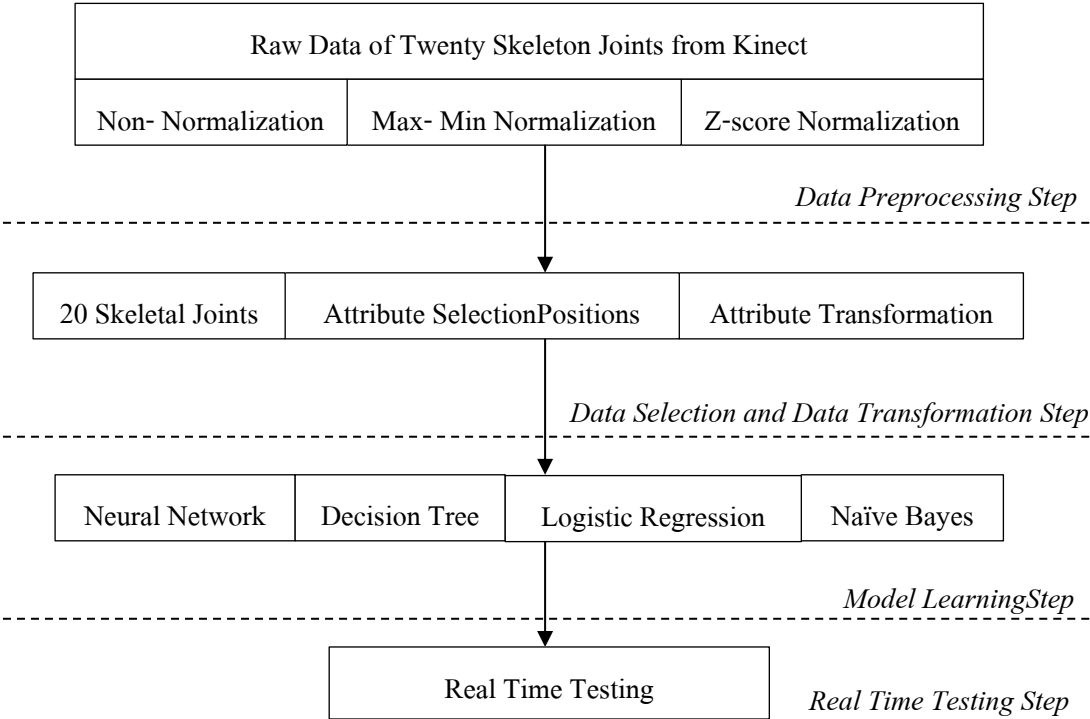
แม้ว่าการแยกแยะท่าทางโดยหลักการวิเคราะห์ภาพสองมิตินั้นเป็นหลักการที่ได้รับการยอมรับในอดีต แต่การแยกแยะร่างกายออกจากพื้นหลังที่มีความซับซ้อนก็ยังคงเป็นปัญหาหลักในการวิเคราะห์ข้อมูล และนอกจากนี้แสงยังเป็นอีกปัจจัยที่มักส่งผลกระทบต่อการใช้ภาพสองมิติ ซึ่งการวิเคราะห์ท่าทางจากภาพสามมิติเป็นอีกหนึ่งแนวทางหนึ่งที่จะช่วยให้การแยกแยะท่าทางมีความถูกต้องมากยิ่งขึ้น อย่างไรก็ตามในอดีตอุปกรณ์ที่ใช้ในการบันทึกภาพสามมิตินั้นมีราคาค่อนข้างสูง จึงทำให้ในอดีตมีงานวิจัยจำนวนไม่มากนักที่สามารถพัฒนาระบบที่ใช้งานร่วมกับอุปกรณ์บันทึกภาพสามมิติได้ จนกระทั่งปี ค.ศ. 2010 กล้อง Kinect ได้ถูกวางจำหน่ายโดยบริษัท Microsoft ซึ่งมีคุณสมบัติในการแยกแยะร่างกายมนุษย์จากพื้นหลังที่มีความซับซ้อน และยังสามารถสกัดส่วนต่าง ๆ ของร่างกายได้อย่างละเอียด จึงทำให้หลายงานวิจัยในปัจจุบันให้ความสนใจเลือกใช้กล้อง Kinect ในงานวิจัย และจากประโยชน์ของกล้อง Kinect ที่ได้กล่าวมาจึงเป็นสาเหตุให้ผู้วิจัยเลือกใช้กล้อง Kinect ในการตรวจจับท่าทางการเคลื่อนไหวของมนุษย์โดยประยุกต์หลักการท่าเหมือนข้อมูลเพื่อช่วยในการวิเคราะห์และแยกแยะท่าทางของมนุษย์

บทที่ 3 วิธีการดำเนินงานวิจัย

บทนี้จะกล่าวถึงขั้นตอนของการวิเคราะห์ท่าทาง วิธีการลดทอนข้อมูลเพื่อเพิ่มความถูกต้องในการวิเคราะห์ท่าทาง รวมถึงการออกแบบระบบที่ใช้ในการวิเคราะห์ท่าทาง

3.1 ภาพรวมของงานวิจัย

ขั้นตอนในการทดลองของงานวิจัยขั้นนี้สามารถแบ่งได้เป็น 4 ขั้นตอนด้วยกัน ได้แก่ ขั้นตอนการเตรียมข้อมูลเพื่อใช้ในการทดลอง (Data Preprocessing) ขั้นตอนการคัดเลือกและแปลงชุดข้อมูลดิบเพื่อเพิ่มประสิทธิภาพในการวิเคราะห์ข้อมูล (Data Selection and Data Transformation) ขั้นตอนการเรียนรู้ชุดข้อมูลเพื่อทำการวิเคราะห์ และสร้างโมเดลที่ใช้ในการวิเคราะห์ข้อมูล (Model Learning) และขั้นตอนการทดสอบระบบที่ถูกพัฒนาขึ้นตามเวลาจริง (Real Time Testing) โดยภาพรวมของแต่ละขั้นตอนนั้นได้แสดงไว้ในรูปที่ 3.1 และรายละเอียดในแต่ละขั้นตอนแสดงไว้ในหัวข้อที่ 3.2-3.4 ตามลำดับ



รูปที่ 3.1 ภาพรวมของขั้นตอนการทดลอง

3.2 Data Preprocessing

ในขั้นตอนนี้จะเริ่มจากการบันทึกท่าทางต่าง ๆ ที่งานวิจัยให้ความสนใจโดยใช้กล้อง Kinect ในการบันทึกข้อมูล ซึ่งท่าทางที่งานวิจัยนี้ให้ความสนใจมีด้วยกันทั้งหมด 18 ท่าทาง โดยสามารถแยกเป็นท่าหลักด้วยกัน 4 ท่าประกอบด้วย ท่ายืน ท่านั่ง ท่านั่งที่พื้น และท่านอน โดยในส่วนของท่ายืน ท่านั่ง และท่านั่งที่พื้นสามารถแยกแยะรายละเอียดท่าทางเพิ่มเติมได้ซึ่งประกอบด้วยลักษณะท่าทางของหลัง และท่าทางการยกมือ โดยในส่วนของท่านอนนั้นในงานวิจัยชิ้นนี้ให้ความสนใจเพียงภาพรวมเท่านั้น ซึ่งรายละเอียดของท่าทางทั้งหมดได้ถูกแสดงไว้ในรูปที่ 3.2

Class Membership							
Stand							
	Normal stand	Stand both hands up	Stand left hand up	Stand right hand up	Stand lean forward		
Sit							
	Sit straight back	Sit lean forward	Sit lean backward	Sit both hands up	Sit left hand up	Sit right hand up	
Sit on floor							
	Sit on floor straight back	Sit on floor lean forward	Sit on floor lean backward	Sit on floor both hands up	Sit on floor left hand up	Sit on floor right hand up	
Lie down							
	Lie down in various manners						

รูปที่ 3.2 รายละเอียดท่าทางภายในงานวิจัย

3.2.1 Training Set Detail

ชุดข้อมูลที่ใช้ในการเรียนรู้ของแต่ละอัลกอริทึมมาจากข้อมูลที่ถูกบันทึกจากผู้เข้าร่วมการทดลองจำนวน 10 คนมีอายุระหว่าง 26-75 ปี มีความสูงระหว่าง 150-180 เซนติเมตร โดยแบ่งเป็น เพศหญิงจำนวน 5 คนและเพศชายจำนวน 5 คน โดยทุกคนจะถูกขอให้แสดงท่าทางต่าง ๆ ที่เห็นได้ชัดเจนหน้ากล้อง Kinect โดยระยะที่ใช้ในการบันทึกข้อมูลท่าทางการเคลื่อนไหวของผู้ร่วมการทดลองอยู่ระหว่าง

1.8-3.0 เมตร โดยท่าทางทั้ง 18 ท่าจะถูกแสดงโดยผู้เข้าร่วมทดลองทุกคน โดยแต่ละคนจะถูกบันทึกท่าทางแต่ละท่าเป็นจำนวน 5 คลิป ซึ่งข้อมูลที่บันทึกได้ในทุกท่าทางมีทั้งหมด 900 คลิปโดยจำนวนเฟรมภาพที่สามารถบันทึกได้จากผู้ร่วมทำการทดลองทุกคนในทุกคลิปมีทั้งหมด 140,048 เฟรม โดยรายละเอียดของจำนวนเฟรมในแต่ละท่าทางได้ถูกแสดงไว้ในตารางที่ 3.1 และข้อมูลความกว้างของไหล่และความสูงของร่างกายผู้ร่วมการทดลองได้ถูกแสดงไว้ในตาราง 3.2

ตารางที่ 3.1 จำนวนชุดข้อมูลในแต่ละท่าทางของชุด Training Set

Posture	Training set (Frames)
Normal stand	8,153
Stand both hands up	7,915
Stand left hand up	8,165
Stand right hand up	7,521
Stand lean forward	7,848
Sit straight back	7,629
Sit lean forward	7,566
Sit lean backward	6,713
Sit both hands up	7,283
Sit left hand up	7,659
Sit right hand up	7,016
Sit on floor straight back	7,268
Sit on floor lean forward	8,513
Sit on floor lean backward	8,003
Sit on floor both hands up	8,620
Sit on floor left hand up	8,626
Sit on floor right hand up	7,292
Lie down in various manners	8,258

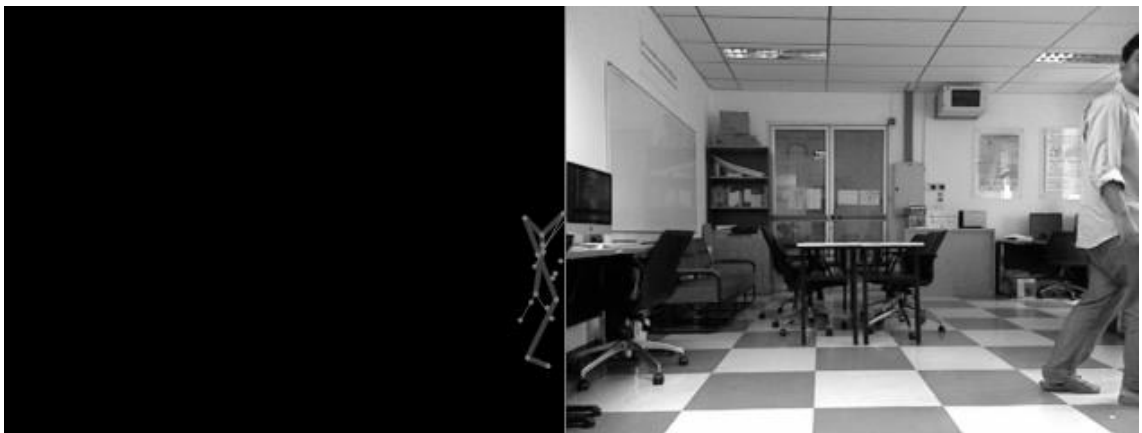
ตารางที่ 3.2 ความกว้างของไหล่และความสูงของร่างกายผู้ร่วมการทดลองในส่วนของ Training Set

	<i>T-1</i>	<i>T-2</i>	<i>T-3</i>	<i>T-4</i>	<i>T-5</i>	<i>T-6</i>	<i>T-7</i>	<i>T-8</i>	<i>T-9</i>	<i>T-10</i>
	<i>Male</i>	<i>Male</i>	<i>Male</i>	<i>Male</i>	<i>Male</i>	<i>Female</i>	<i>Female</i>	<i>Female</i>	<i>Female</i>	<i>Female</i>
Height	165 cm.	150 cm.	175 cm.	180 cm.	158 cm.	170 cm.	160 cm.	157 cm.	166 cm.	169 cm.
Shoulder Width	18 in.	18 in.	20 in.	23 in.	21 in.	18 in.	14 in.	16 in.	16 in.	17 in.

โดยหลังจากบันทึกข้อมูลท่าทางต่าง ๆ ของผู้ร่วมการทดลองทั้งหมดเสร็จแล้ว ผู้วิจัยจะนำข้อมูลดิบที่ได้จากกล้อง Kinect มาทำการคัดเลือกข้อมูลที่มีความผิดพลาดซึ่งเกิดจากการวิเคราะห์ท่าทางที่ผิดพลาดของกล้อง Kinect ออกจากชุดข้อมูลดิบ ซึ่งความผิดพลาดเหล่านี้มักเกิดจากช่วงที่ผู้ร่วมการทดลองออกนอกกรอบที่ได้กำหนดไว้ดังแสดงในรูปที่ 3.3.1 หรือ ในช่วงระหว่างที่กำลังเดินออกจากฉากรับภาพ (Frame) ของกล้อง Kinect ดังแสดงในรูปที่ 3.3.2 ซึ่งหากข้อมูลเหล่านี้ยังคงอยู่ในชุดข้อมูลดิบ และถูกนำไปเรียนรู้เพื่อใช้สร้างโมเดลที่ใช้ในการวิเคราะห์ข้อมูล อาจส่งผลให้การเรียนรู้มีความผิดพลาดมากยิ่งขึ้น



รูปที่ 3.3.1 การวิเคราะห์ท่าทางที่ผิดพลาดของกล้อง Kinect ซึ่งเกิดจากช่วงที่ผู้ร่วมการทดลองออกนอกกรอบที่ได้กำหนดไว้



รูปที่ 3.3.2 การวิเคราะห์ท่าทางที่ผิดพลาดของกล้อง Kinect ซึ่งเกิดในช่วงที่ผู้ร่วมการทดลองกำลังเดินออกจากฉากรับภาพ

รูปที่ 3.3 ความผิดพลาดที่เกิดจากการวิเคราะห์ข้อมูลของกล้อง Kinect

หลังจากทำการคัดกรองชุดข้อมูลดิบที่บันทึกได้จากกล้อง Kinect เสร็จแล้ว ชุดข้อมูลดิบจะถูกนำมาทำการ Normalization เพื่อปรับชุดข้อมูลของผู้ร่วมการทดลองแต่ละคนที่ได้ถูกบันทึกไว้ให้มีความแตกต่างที่น้อยลง ซึ่งในงานวิจัยชิ้นนี้เลือกใช้หลักการ Normalization แบบ Max-Min Normalization และ Z-score Normalization ในการคำนวณเพื่อลดความแตกต่างของชุดข้อมูลโดยวิธีการในการคำนวณนั้นได้แสดงไว้ในสมการที่ 3.1 และ 3.2 ตามลำดับ

$$J' = \frac{J - Min}{Max - Min} \quad (3.1)$$

$$J' = \frac{J - \bar{J}}{\sigma} \quad (3.2)$$

โดยค่าตัวแปร J' นั้นเป็นค่าผลลัพธ์ที่ได้จากการทำ Normalization ของทั้งในแบบ Max-Min Normalization และ Z-score Normalization ค่าตัวแปร J นั้นเป็นข้อมูลในแต่ละแถว (Row) ของชุดข้อมูลดิบ โดยค่าตัวแปร Max และ Min ในสมการที่ 3.1 นั้นเป็นค่าที่มากที่สุดและน้อยสุดในแต่ละหลัก (Column) ของชุดข้อมูลดิบ ในส่วนค่าตัวแปร $\bar{J}$ และ σ ในสมการที่ 3.2 นั้นเป็นค่าเฉลี่ยรวม และ Standard Deviation ในแต่ละหลักของชุดข้อมูลดิบตามลำดับ สาเหตุที่งานวิจัยชิ้นนี้นำข้อมูลดิบมาทำ Normalization เพื่อต้องการทราบว่าในการทำ Normalization นั้นเป็นปัจจัยหนึ่งส่งผลให้การเรียนรู้ของชุดข้อมูลมีความถูกต้องมากขึ้นหรือไม่

3.3 Data Selection and Data Transformation

ในขั้นตอนนี้เป็นขั้นตอนในการเตรียมข้อมูลก่อนนำไปทำการเรียนรู้ และสร้างโมเดลที่ใช้ในการวิเคราะห์ท่าทาง โดยการนำชุดข้อมูลดิบทั้ง 3 แบบ ซึ่งประกอบด้วย ชุดข้อมูลดิบที่ไม่ผ่านการ Normalization (N-NML) และชุดข้อมูลดิบที่ถูก Normalization ในแบบ Max-Min Normalization (MM-NML) และ Z-score Normalization (Z-NML) มาทำการคัดเลือกและแปลงข้อมูลเพื่อให้ได้ชุดข้อมูลที่สามารถช่วยให้การเรียนรู้และค้นหารูปแบบในการวิเคราะห์ข้อมูลมีความถูกต้องมากยิ่งขึ้น โดยหลักการที่จะประยุกต์ใช้ในขั้นตอนนี้มี 2 หลักการคือ Data Selection และ Data Transformation เพื่อเปรียบเทียบความแตกต่างในการรู้เรียนรู้ชุดข้อมูลที่ไม่ได้ผ่านทำ Data Selection และ Data Transformation (20 Skeletal Joints) ว่ามีความแตกต่างกันหรือไม่เพื่อหาวิธีที่เหมาะสมที่สุดในการสร้างโมเดลเพื่อใช้ในการวิเคราะห์ท่าทาง

3.3.1 Data Selection

Data Selection เป็นการจัดลำดับความสำคัญของชุดข้อมูลเพื่อใช้เพิ่มประสิทธิภาพในการเรียนรู้และค้นหารูปแบบในการวิเคราะห์ข้อมูล โดยส่วนของข้อมูลที่มีความสำคัญต่อการเรียนรู้มากจะถูกจัดลำดับไว้ในลำดับต้นและส่วนที่ไม่มีความสำคัญในการเรียนรู้จะถูกจัดลำดับไว้ในลำดับท้าย โดยในงานวิจัยชิ้นนี้เลือกตัวคัดเลือกข้อมูลเพื่อนำมาใช้ในการจัดลำดับชุดข้อมูล 2 ตัวด้วยกันคือ Info-Gain Ranker Evaluation และ Gain Ranker Evaluation เพื่อใช้ในการคัดเลือก และลำดับความสำคัญของชุดข้อมูล โดยตัวคัดเลือกข้อมูลทั้งสองตัวนั้นเป็น Function ที่สามารถเรียกใช้ได้จากโปรแกรม Weka โดยในงานวิจัยชิ้นนี้ได้ทดลองเลือกใช้ข้อมูลใน 10, 20, 30, 40 และ 50 อันแรกเพื่อทดสอบความถูกต้องในการวิเคราะห์ท่าทาง โดยการจัดลำดับชุดข้อมูลได้แสดงไว้ในตารางที่ 4.1 และ 4.2 ของบทที่ 4

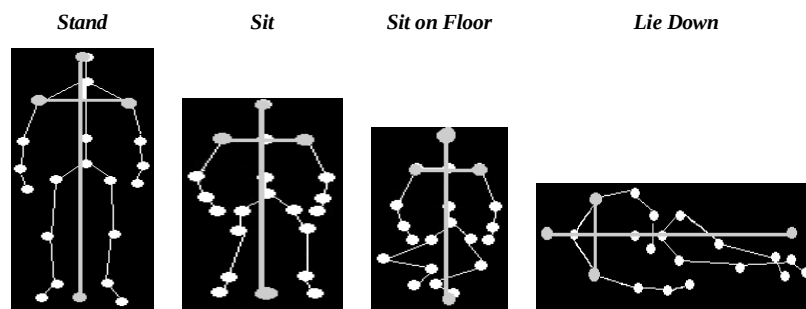
3.3.2 Data Transformation

Data Transformation เป็นหลักการในการนำชุดข้อมูลมาทำการเปลี่ยนแปลงเพื่อลดความซับซ้อนภายในชุดข้อมูล และยังส่งผลให้มีการเพิ่มความสัมพันธ์ภายในชุดข้อมูล ซึ่งจะทำให้การวิเคราะห์ชุดข้อมูลมีความถูกต้องมากขึ้น โดยข้อมูลตั้งต้นที่ใช้ในการเปลี่ยนแปลงข้อมูลของงานวิจัยชิ้นนี้คือข้อมูลพิกัดส่วนต่าง ๆ ของร่างกายที่วิเคราะห์ได้จากกล้อง Kinect โดยในงานวิจัยชิ้นนี้ได้สังเกตหาความแตกต่างของท่าทางของอวัยวะในส่วนต่าง ๆ ของร่างกายจากทั้ง 18 ท่าทางที่ได้กำหนดไว้เพื่อนำมาแปลงให้อยู่ในรูปสมการเพื่อใช้ในการสร้างชุดข้อมูลชุดใหม่ โดยในชุดข้อมูลนี้มี Attribute ที่ใช้ในการเรียนรู้ 9 Attributes ด้วยกันซึ่งได้แก่ มุมการงอของเข่าซ้ายและขวา (Angle of Knee Left and Right) อัตราส่วนระหว่างความกว้างและความสูงของร่างกาย (Aspect Ratio of Width and Height) ระยะห่างระหว่างเอวไปยังระนาบพื้นห้อง (Distance form Hip to Room Floor) ลักษณะการงอของหลัง (Back Status) และท่าทางการยกมือ (Hand Up) ซึ่งแต่ละ Attribute นั้นจะช่วยในการวิเคราะห์ท่าทางในส่วนต่าง ๆ ของร่างกาย โดยวิธีการคำนวณค่าของแต่ละ Attribute ได้แสดงไว้ในสมการที่ 3.3–3.14

Aspect Ratio of Width and Height

Aspect Ratio of Width and Height เป็น Attribute ที่ใช้ในการแยกแยะท่าทางพื้นฐานซึ่งประกอบด้วย ท่ายืน ท่านั่ง ท่านั่งที่พื้น และท่านอน โดยแต่ละท่าทางนั้นจะมีอัตราส่วนระหว่างความกว้าง และความสูง ที่แตกต่างกันออกไปดังแสดงในรูปที่ 3.4 โดยพบว่าค่าอัตราส่วนของท่านอนนั้นค่อนข้างมีความแตกต่างจากท่าทางพื้นฐานอื่น ๆ อย่างชัดเจน จึงทำให้งานวิจัยชิ้นนี้เลือกอัตราส่วนระหว่างความกว้างและความสูงของร่างกายเป็นตัวช่วยในการวิเคราะห์ท่าทาง ซึ่งข้อมูลที่ใช้ในการคำนวณหา

อัตราส่วนมาจากค่าพิกัดในแนวแกน X และค่าพิกัดในแนวแกน Y จาก 5 จุดบนร่างกายซึ่งประกอบด้วย จุดข้อมูลในส่วนหัว ไหล่ซ้าย ไหล่ขวา เท้าซ้ายและขวา โดยนำจุดทั้งหมดที่ได้กล่าวไว้มาเปรียบเทียบหาค่ามากที่สุดและน้อยที่สุดในแต่ละแกน (Max_X , Min_X , Max_Y , Min_Y) โดยค่า $Width$ นั้นเป็นค่าคำนวณได้จากผลต่างระหว่างค่าในแนวแกน X ที่มีค่ามากที่สุดและน้อยสุด และ $Height$ นั้นเป็นค่าที่คำนวณได้จากผลต่างระหว่างค่าในแนวแกน Y ที่มีค่ามากที่สุด และมีค่าน้อยสุดจากจุดทั้ง 5 จุด และค่า $Aspect Ratio$ เกิดจากอัตราส่วนระหว่างค่า $Width$ และ $Height$ ดังแสดงในสมการที่ 3.3–3.5



รูปที่ 3.4 อัตราส่วนระหว่างความกว้างและความสูงของร่างกายมนุษย์ในแต่ละท่าทาง

$$Width = Max_X - Min_X \quad (3.3)$$

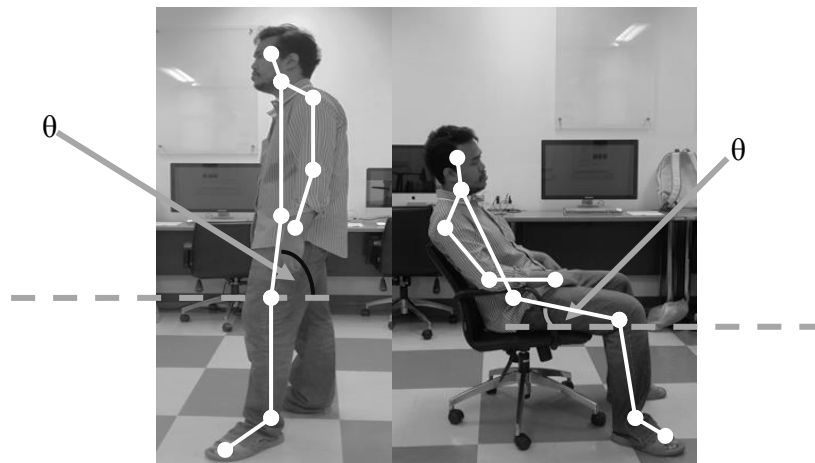
$$Height = Max_Y - Min_Y \quad (3.4)$$

$$Aspect Ratio = \frac{Width}{Height} \quad (3.5)$$

อย่างไรก็ตามแม้ว่า Aspect Ratio of Width and Height นั้นจะเป็น Attribute ที่สามารถช่วยในการวิเคราะห์ท่าทางพื้นฐานได้ แต่ก็ยังพบว่าหลายครั้งมีข้อมูลอัตราส่วนของท่ายืน ท่านั่ง และท่าก้มที่พื้นที่มีค่าคล้ายคลึงกัน ซึ่งอาจจะทำให้การแยกแยะท่าทางมีความผิดพลาดเกิดขึ้นได้ ดังนั้นในงานวิจัยชิ้นนี้จึงได้เพิ่ม Attribute ที่จะช่วยให้การวิเคราะห์ท่าทางพื้นฐานมีความถูกต้องมากยิ่งขึ้น ซึ่งประกอบด้วย การวิเคราะห์มุมการงอของเข่า และระยะทางระหว่างเอวไปยังพื้นห้อง

Angle of Knee Left and Right

Angle of Knee Left and Right เป็น Attribute ที่ใช้ในการแยกแยะท่าทางการนั่งออกจากท่าทางการยืน โดยปกติกรณีที่ร่างกายมนุษย์กำลังอยู่ในท่านั่ง ส่วนของขาที่นอนบนจะมีการงอตัวลงทำให้เกิดมุมของหัวเข่า ในทางกลับกันเมื่อร่างกายมนุษย์กำลังอยู่ในท่ายืนส่วนของขาที่นอนบนนั้นจะกางออกซึ่งทำให้เกิดมุมที่กว้างขึ้นกว่าตอนที่ร่างกายกำลังอยู่ในท่านั่งดังแสดงในรูปที่ 3.5 การงอของขาที่นอนบนจึงเป็นหนึ่งในลักษณะที่สามารถช่วยในการแยกแยะท่าทางระหว่างท่านั่ง และท่ายืนได้ ในงานวิจัยชิ้นนี้จึงเลือกคุณลักษณะนี้เพื่อใช้ในการแยกแยะท่าทางระหว่างท่านั่ง และท่านั่ง โดยวิธีการคำนวณการงอของขาที่นอนบนนั้นได้ถูกแสดงไว้ในสมการที่ 3.6–3.8 โดยค่าตัวแปร A ในสมการที่ 3.6 เป็นค่าระยะห่างระหว่างจุดของหัวเข่า และจุดของเอว ส่วนค่าของตัวแปร B ของสมการที่ 3.7 นั้นเป็นค่าผลต่างระหว่างค่าพิกัดในแนวแกน Y ของจุดหัวเข่าและเอว โดยมุมงอสามารถหาได้จากค่า $\arcsin$ ของอัตราส่วนของค่าตัวแปร A และ B ดังแสดงในสมการที่ 3.8



รูปที่ 3.5 ขนาดของมุมงอของเข่าที่นอนบนระหว่างท่านั่งและท่านั่ง

$$A = \sqrt{(X_{hip} - X_{knee})^2 + (Y_{hip} - Y_{knee})^2 + (Z_{hip} - Z_{knee})^2} \quad (3.6)$$

$$B = |Y_{hip} - Y_{knee}| \quad (3.7)$$

$$\theta = \sin^{-1} \left(\frac{A}{B} \right) \quad (3.8)$$

Distance form Hip to Room Floor

Distance form Hip to Room Floor เป็น Attribute ที่ช่วยในการแยกแยะระหว่างท่าทางนั่งที่พื้น และ นั่งบนเก้าอี้ ซึ่งพบว่าข้อมูลของ Attribute ของ Angle of Knee Left and Right และ Aspect Ratio of Width and Height นั้นไม่สามารถแยกแยะท่าทางทั้งสองแบบได้ ซึ่งความแตกต่างระหว่างสองท่าทางนี้ นั้นคือการนั่งติดพื้นนั้นจะมีระยะห่างระหว่างเอวมายังพื้นที่น้อยกว่าการนั่งบนเก้าอี้หรือโซฟา ดังแสดงในรูป 3.6 ในงานวิจัยชิ้นนี้จึงเลือกระยะห่างระหว่างเอวมายังพื้นเป็นตัวช่วยในการวิเคราะห์ และแยกแยะระหว่างท่าทางนั่งบนเก้าอี้และท่าทางที่พื้น โดยวิธีการคำนวณหาค่าระยะห่างได้ถูกแสดงไว้ในสมการที่ 3.9 ซึ่งเป็นสมการที่ใช้ในการวัดระยะห่างระหว่างจุดใด ๆ ไปยังระนาบ โดยเปรียบเทียบห้องเป็นระนาบ และใช้จุดของเอวในการวัดระยะห่างระหว่างจุดไปยังพื้นระนาบ โดยค่าตัวแปร A , B , C และ D นั้นเป็นค่าสัมประสิทธิ์ระนาบซึ่งเป็นค่าที่ถูกประมาณการด้วยกล้อง Kinect และตัวแปร X , Y และ Z นั้นคือค่าพิกัดของจุดเอวที่ถูกวิเคราะห์ได้จากกล้อง Kinect



รูปที่ 3.6 ระยะห่างระหว่างเอวมายังพื้นที่แตกต่างกันในแต่ละท่าทาง

$$Distance = \frac{AX_{hip\ center} + BY_{hip\ center} + CZ_{hip\ center} + D}{\sqrt{A^2 + B^2 + C^2}} \quad (3.9)$$

Back Status

Back Status เป็น Attribute ที่ใช้ในการวิเคราะห์ลักษณะท่าทางของหลังซึ่งในงานวิจัยชิ้นนี้แบ่งแยกลักษณะท่าทางของหลังเป็น 3 รูปแบบด้วยกันได้แก่ หลังที่มีลักษณะงอ หลังที่มีลักษณะตรง และหลังที่มีลักษณะเอนไปด้านหลัง ดังแสดงไปในรูปที่ 3.7 โดยงานวิจัยชิ้นนี้เลือกใช้การวิเคราะห์ลักษณะของหลังจากค่าผลต่างของค่าในระนาบแกน Z ระหว่างจุดศูนย์กลางของไหล่และจุดศูนย์กลางของเอว ซึ่งข้อมูลทั้งสองจุดนั้นสามารถวิเคราะห์ได้จากกล้อง Kinect โดยวิธีการในการคำนวณได้แสดงไว้ในสมการที่ 3.10 โดยการวิเคราะห์ท่าทางของหลังในงานวิจัยชิ้นนี้ให้ความสนใจที่จะวิเคราะห์ในส่วนของท่านั่ง ท่านั่ง และท่านั่งที่พื้น



รูปที่ 3.7 ลักษณะท่าทางของหลังในแต่ละแบบ

$$\text{Back Status} = Z_{\text{shoulder center}} - Z_{\text{hip}} \quad (3.10)$$

Hand Up

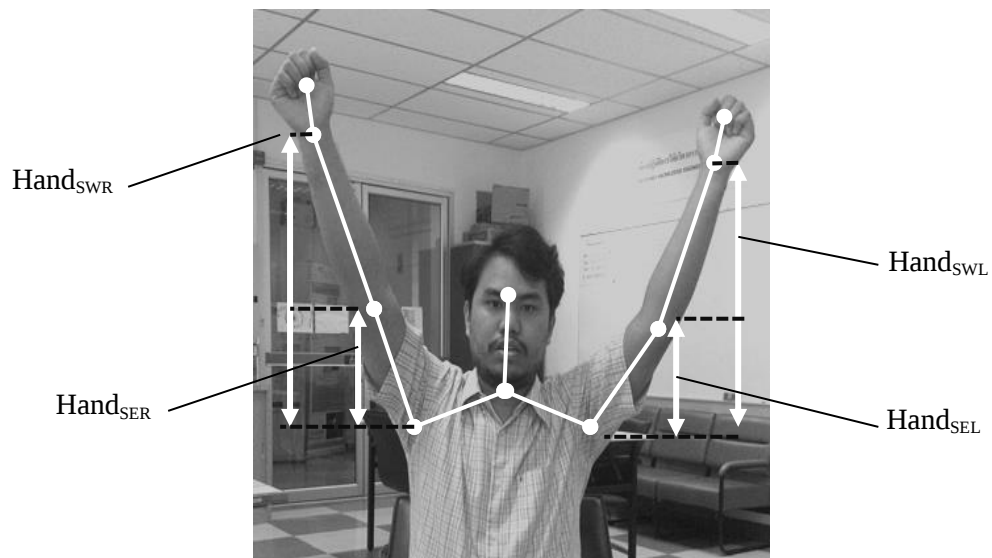
Hand Up คือ Attribute ที่ใช้ในการวิเคราะห์เพื่อแยกแยะท่าทางการยกมือ ซึ่งในงานวิจัยชิ้นนี้แบ่งท่าทางการยกมือเป็นสามรูปแบบด้วยกันได้แก่ การยกมือข้างซ้าย การยกมือข้างขวา และการยกมือทั้งสองข้าง โดยงานวิจัยชิ้นนี้เลือก 4 Attribute เพื่อใช้ในการวิเคราะห์ท่าทางการยกมือ ซึ่งประกอบด้วยผลต่างระหว่างค่าในแนวแกน Y ของไหล่และข้อศอกทั้งด้านซ้ายและด้านขวา และผลต่างระหว่างค่าในแนวแกน Y ของไหล่และข้อมือทั้งด้านซ้ายและด้านขวาดังแสดงในรูปที่ 3.8 ซึ่งวิธีการคำนวณค่าทั้ง 4 นั้นได้ถูกแสดงไว้สมการที่ 3.11–3.14

$$Hand_{SWL} = Y_{shoulder\ left} - Y_{wrist\ left} \quad (3.11)$$

$$Hand_{SWR} = Y_{shoulder\ right} - Y_{wrist\ right} \quad (3.12)$$

$$Hand_{SEL} = Y_{sholder\ left} - Y_{elbow\ left} \quad (3.13)$$

$$Hand_{SER} = Y_{sholder\ right} - Y_{elbow\ right} \quad (3.14)$$



รูปที่ 3.8 ผลต่างในแต่ละค่าตัวแปรที่ใช้ในการวิเคราะห์ท่าทางการยกมือ

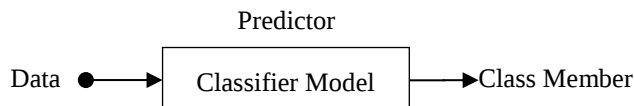
ค่าตัวแปร $Hand_{SWL}$ และ $Hand_{SWR}$ ในสมการที่ 3.11 และ 3.12 นั้นคือผลต่างระหว่างไหล่และข้อมือซ้าย และขวาตามลำดับ ส่วนค่าตัวแปร $Hand_{SEL}$ และ $Hand_{SER}$ ในสมการที่ 3.13 และ 3.14 ก็คือผลต่างระหว่างไหล่และข้อศอกซ้ายและขวาตามลำดับ

3.4 Model Learning

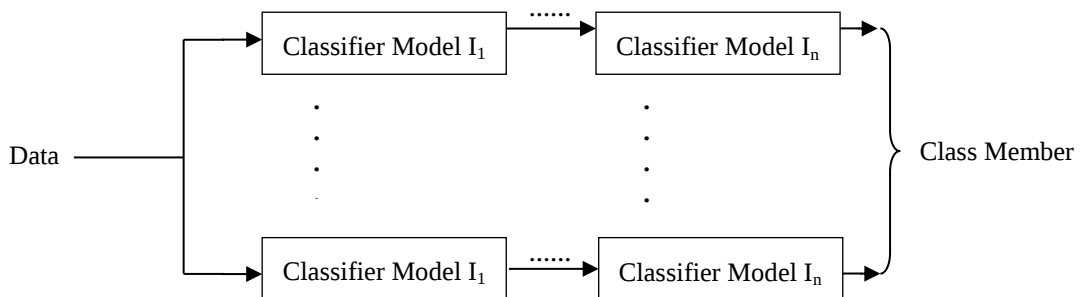
หลังจากนำชุดข้อมูลไปทำการคัดเลือกและทำการแปลงข้อมูลแล้ว ในงานวิจัยชิ้นนี้จะนำชุดข้อมูลแต่ละแบบมาทำการเรียนรู้ เพื่อนำไปสร้างโมเดลที่ใช้ในการวิเคราะห์ข้อมูลซึ่งมี 2 รูปแบบด้วยกันเพื่อใช้ในการเปรียบเทียบความสามารถในการวิเคราะห์ข้อมูลซึ่งประกอบด้วย

Single-stage Classifier Model เป็นโมเดลเชิงเดี่ยวที่เกิดจากการเรียนรู้ชุดข้อมูลทั้งหมดภายในชุดข้อมูลเดียว ซึ่งเป็นการสร้างโมเดลที่สามารถวิเคราะห์ชุดข้อมูลทั้งหมดได้ภายในโมเดลเดียวดังแสดงในรูป 3.9 ซึ่งการใช้โมเดลเดี่ยวในการวิเคราะห์ข้อมูลนั้นจะทำให้ง่ายต่อการพัฒนาระบบ แต่อย่างไรก็ตามการรวมชุดข้อมูลเข้าไว้ด้วยกันเป็นชุดเดียวนั้นทำให้ข้อมูลมีขนาดใหญ่ จึงส่งผลให้การเรียนรู้ของแต่ละอัลกอริทึมกินเวลานาน และนอกจากนี้ยังอาจส่งผลให้การเรียนรู้ในชุดข้อมูลมีความผิดพลาดมากยิ่งขึ้นตามไปด้วย

Multiple-stage Classifier Model เป็นโมเดลแบบเชิงซ้อนที่เกิดจากการเรียนรู้ชุดข้อมูลที่ถูกแบ่งออกเป็นหลายชุด ซึ่งทุกชุดข้อมูลที่ได้รับการเรียนรู้จะสามารถสร้างโมเดลได้หนึ่งโมเดล โดยทุกโมเดลที่ถูกสร้างขึ้นจะทำงานร่วมกันเพื่อทำการวิเคราะห์ชุดข้อมูลดังแสดงในรูปที่ 3.10 ซึ่งการแบ่งแยกชุดข้อมูลนั้นจะลดความซับซ้อนภายในชุดข้อมูล ส่งผลให้การวิเคราะห์ข้อมูลอาจมีความแม่นยำมากยิ่งขึ้น และนอกจากนี้ยังช่วยลดเวลาในการเรียนรู้ของชุดข้อมูลลงได้อีกด้วย

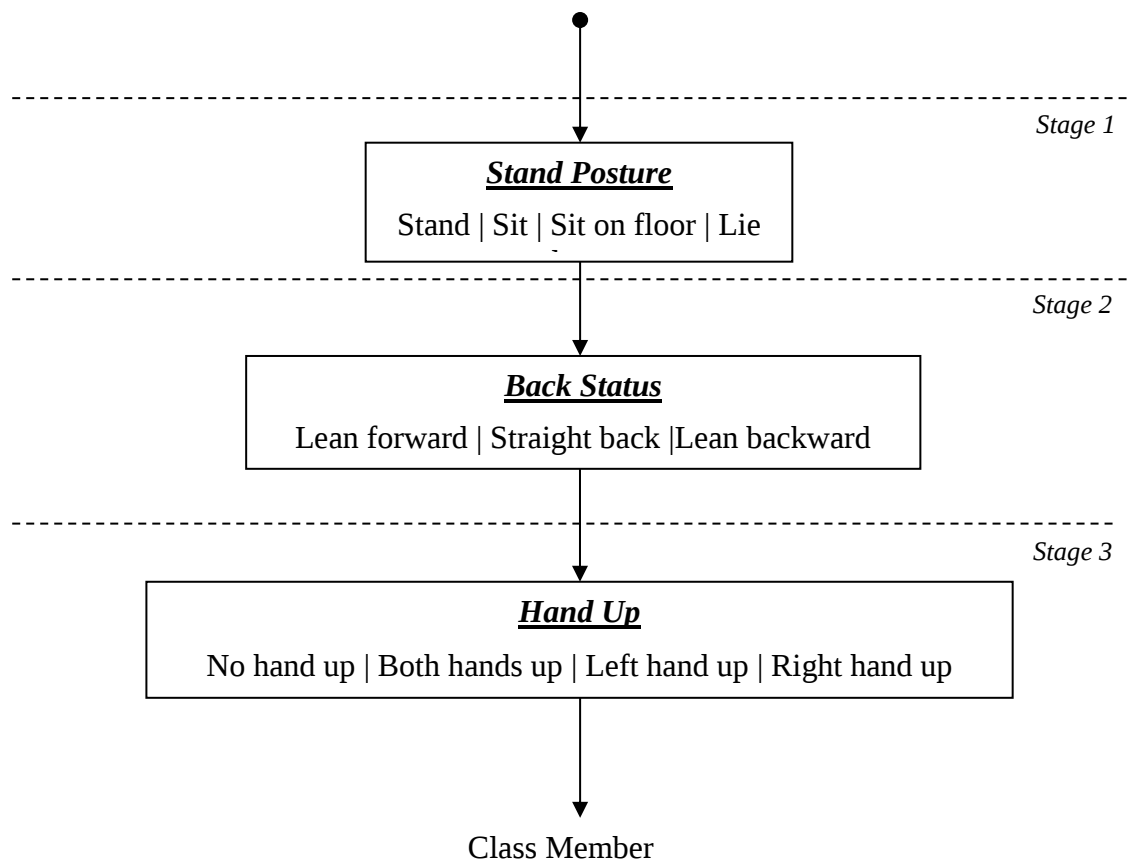


รูปที่ 3.9 Single-stage Classifier Model



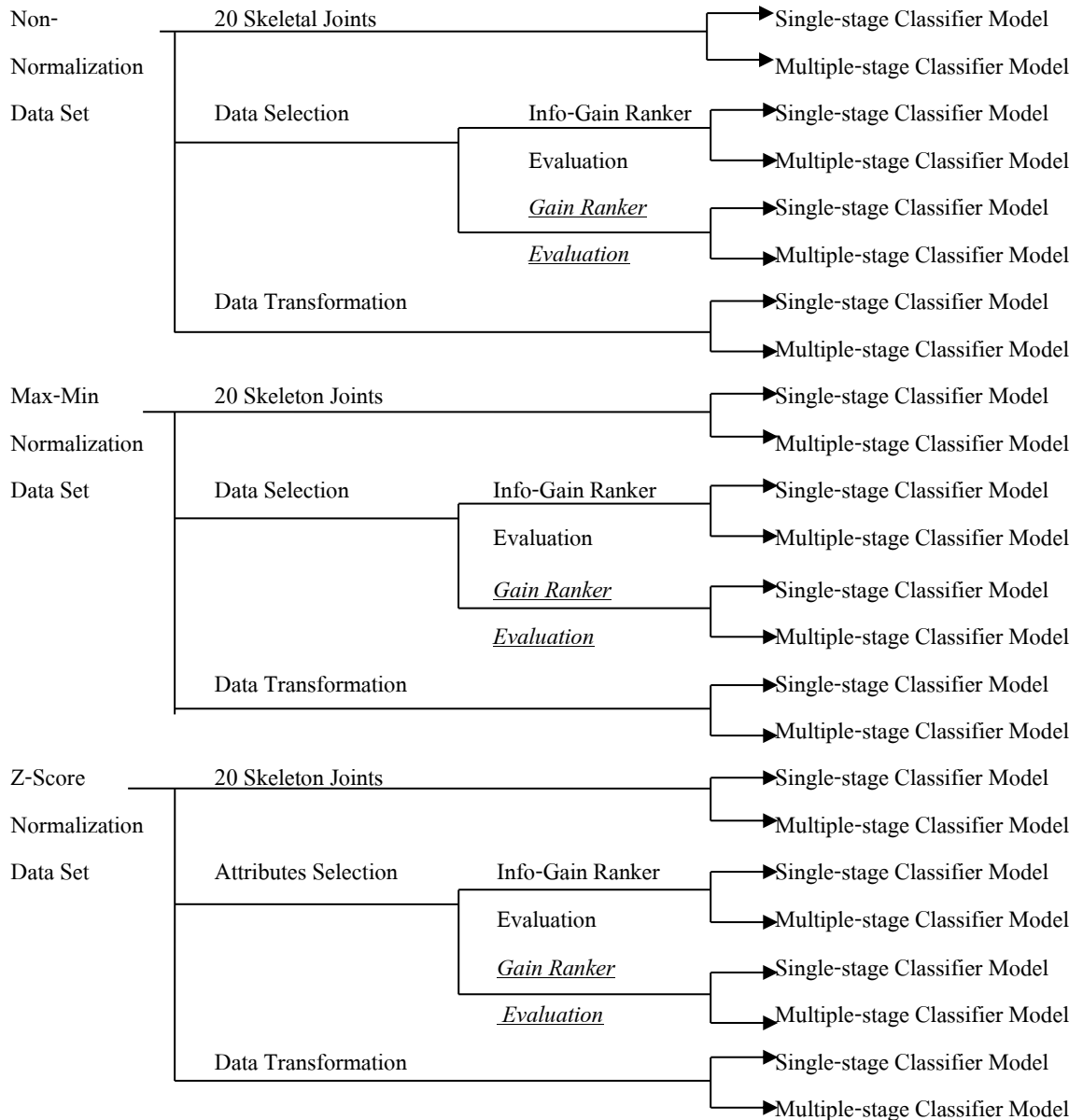
รูปที่ 3.10 Multiple-stage Classifier Model

โดยในส่วนของการแยกชุดข้อมูลของงานวิจัยชิ้นนี้ ได้มีการแบ่งชุดข้อมูลออกเป็น 3 ชุดด้วยกันซึ่งประกอบด้วย ชุดข้อมูลของท่าทางพื้นฐาน ชุดข้อมูลท่าทางการยกมือ และชุดข้อมูลของลักษณะการงอหลัง โดยแต่ละชุดข้อมูลที่ใช้ในการเรียนรู้จะได้ 1 โมเดล ซึ่งในการวิเคราะห์ข้อมูลตามเวลาจริงระบบจะการเริ่มวิเคราะห์จากท่าทางพื้นฐาน ซึ่งประกอบด้วย ท่ายืน ท่านั่ง ท่านั่งที่พื้น และท่านอนเป็นอย่างแรก และในขั้นถัดไปจะเป็นการวิเคราะห์รูปแบบการงอของหลังซึ่งจะสามารถวิเคราะห์ได้ 3 รูปแบบด้วยกัน ซึ่งประกอบด้วย ท่างอหลัง ท่าเอนหลัง และท่าหลังตรง ในขั้นสุดท้ายจะเป็นการวิเคราะห์ท่าทางการยกมือ ซึ่งสามารถวิเคราะห์ได้ 4 รูปแบบด้วยกันได้แก่ ท่ายกมือซ้าย ท่ายกมือขวา ท่ายกมือทั้งสองข้าง และท่าทางปกติไม่ได้ทำการยกมือ โดยรูปแบบการทำงานได้ถูกแสดงไว้ในรูปที่ 3.11



รูปที่ 3.11 โครงสร้างการทำงานของ การวิเคราะห์ข้อมูลแบบ Multiple-stage Classifier Model

เมื่อจัดการชุดข้อมูลที่จะใช้ในการทดลองตามขั้นตอนต่าง ๆ ที่ได้กล่าวไว้ข้างต้น จะทำให้จำนวนชุดข้อมูลที่ใช้ในการทดลองเพื่อเปรียบเทียบความอัตราความแม่นยำในการทดลองของงานวิจัยชิ้นนี้มีทั้งหมด 24 แบบด้วยกันดังแสดงในรูปที่ 3.12



รูปที่ 3.12 รูปแบบชุดข้อมูลแบบต่าง ๆ ที่ใช้ในการทดลอง

3.4.1 Classifier Algorithm

Classifier Algorithm คืออัลกอริทึมที่ใช้ในการเรียนรู้เพื่อสร้างโมเดลที่ใช้ในการวิเคราะห์ข้อมูล ซึ่งงานวิจัยชิ้นนี้เลือกอัลกอริทึมในการเรียนรู้ทั้งหมด 4 อัลกอริทึมซึ่งประกอบด้วย

Decision Tree

Decision Tree (DT) เป็นอัลกอริทึมที่ทำการเรียนรู้ชุดข้อมูลโดยการตัดสินใจทดสอบหาคุณลักษณะที่มีความสำคัญมากที่สุดเพื่อใช้ในการจำแนกชุดข้อมูล ซึ่งผลลัพธ์ที่ได้จากการเรียนรู้ของอัลกอริทึมแบบ Decision Tree นั้นจะอยู่ในรูปแบบของโครงสร้างแผนภาพต้นไม้ (Tree Structure) ข้อมูลที่ถูกวิเคราะห์จากอัลกอริทึมของ Decision Tree นั้นจะเริ่มถูกวิเคราะห์จากรากของแผนภาพต้นไม้ (Root Node) ซึ่งมีความสามารถจำแนกที่หลากหลายผ่านไปยังแขนงต่าง ๆ (Branch) ไปจนถึงใบของแผนภาพต้นไม้ (Leaf Nodes) ซึ่งจะเป็นส่วนที่ใช้ในการจำแนกประเภทของชุดข้อมูลที่ถูกวิเคราะห์ โดยงานวิจัยชิ้นนี้เลือกใช้อัลกอริทึม Decision Tree แบบ J48 ในการวิเคราะห์ข้อมูล

Naïve Bayes

Naïve Bayes (NB) เป็นหนึ่งในอัลกอริทึมพื้นฐานที่ได้รับความนิยมในการจำแนกประเภทในชุดข้อมูล ซึ่งอัลกอริทึม Naïve Bayes เป็นอัลกอริทึมที่สามารถเรียนรู้ชุดข้อมูลได้อย่างรวดเร็วเมื่อเทียบกับอัลกอริทึมอื่น ๆ โดยอัลกอริทึมนี้จะประยุกต์ใช้ทฤษฎีของ Bayes ในการวิเคราะห์ชุดข้อมูล โดยการทำนายประเภทของชุดข้อมูลจากความน่าจะเป็นของชุดข้อมูลที่มีความเป็นไปได้มากที่สุด

Neural Network

Neural Network (NN) เป็นการวิเคราะห์ข้อมูลในรูปแบบเครือข่ายประสาทเทียมซึ่งเป็นการจำลองการทำงานของเครือข่ายประสาทเทียมของสมองมนุษย์ โดยการวิเคราะห์ข้อมูลของอัลกอริทึม Neural Network นั้นจะทำการวิเคราะห์ข้อมูลแบบลำดับชั้น (Layer) โดยทุกลำดับชั้นจะมีการวิเคราะห์หาค่าน้ำหนักของชุดข้อมูล (Weight) เปรียบเทียบกับค่า Threshold เพื่อหาเส้นทางไปยังลำดับชั้นถัดไป โดยในงานวิจัยชิ้นนี้ใช้ Neural Network แบบ Multilayer Perceptron (MLP) โดยใช้อัลกอริทึมแบบ Back-propagation ในการเรียนรู้ชุดข้อมูล

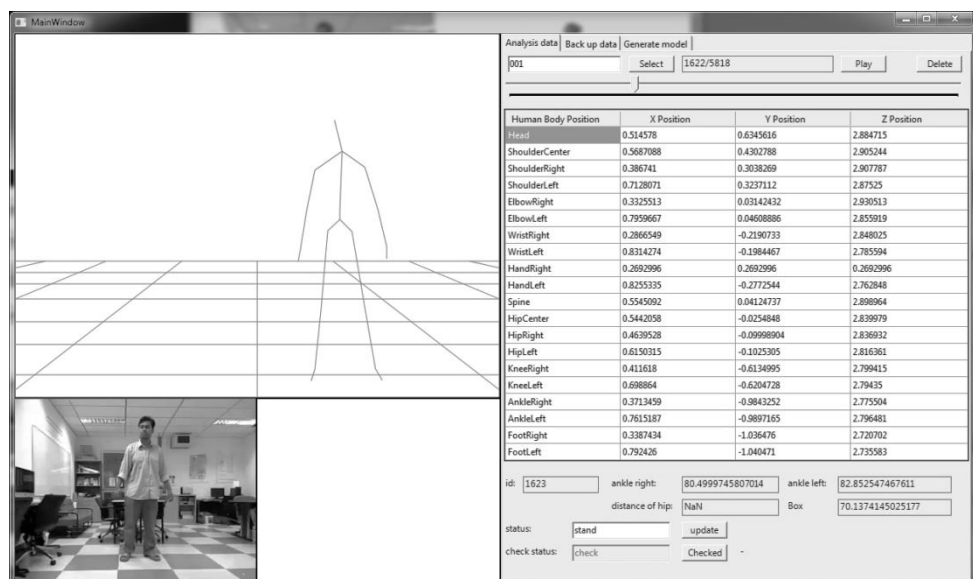
Logistic Regression

Logistic Regression (LGT) เป็นอัลกอริทึมหนึ่งที่ได้รับคามนิยมในการแยกแยะประเภทของชุดข้อมูล โดยใช้หลักการเรียนรู้ทางสถิติแบบเชิงเส้นในการวิเคราะห์เพื่อหารูปแบบของชุดข้อมูลโดย

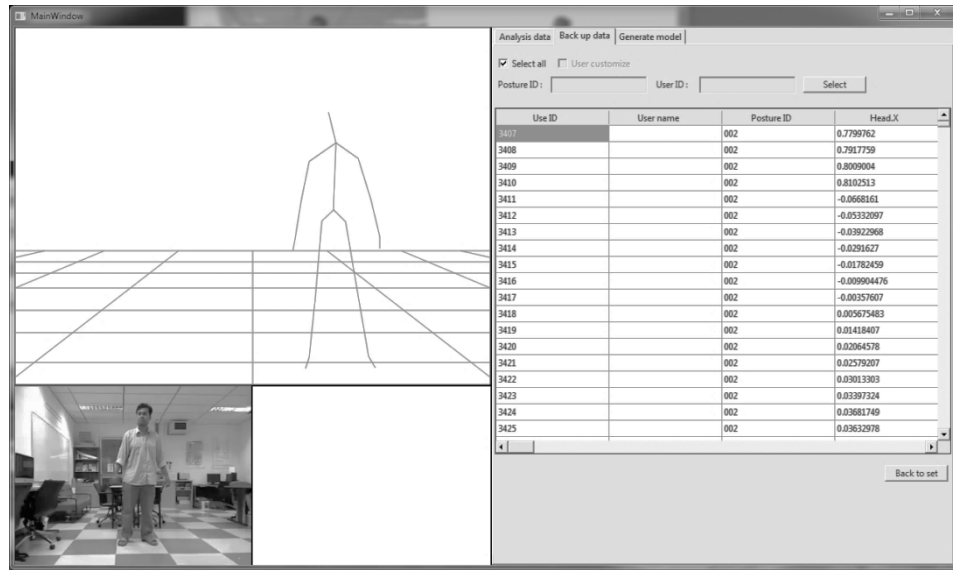
การเปรียบเทียบความเป็นไปได้ในแต่ละประเภทภายในชุดข้อมูล โดยในงานวิจัยชิ้นนี้เลือกอัลกอริทึมแบบ Multinomial Logistic Regression ในการเรียนรู้ชุดข้อมูล

3.5 ภาพรวมของระบบ

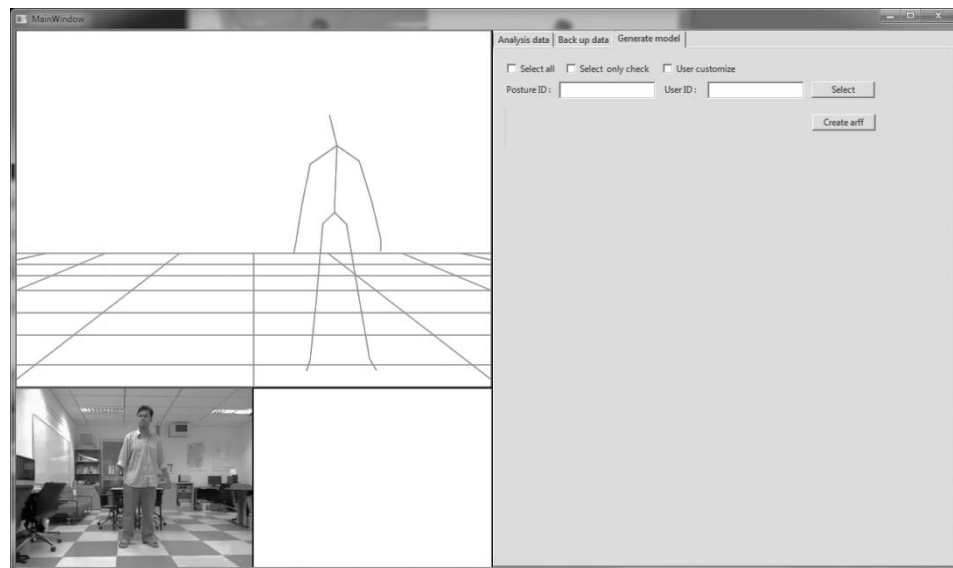
ระบบที่ถูกพัฒนาขึ้นมาเพื่อใช้ในการวิเคราะห์ท่าทางของงานวิจัยชิ้นนี้แบ่งออกเป็น 2 ส่วนด้วยกัน ประกอบด้วย ส่วนที่หนึ่งโปรแกรมที่ใช้ในการจัดการชุดข้อมูลเพื่อใช้ในการเรียนรู้และส่วนที่สองคือโปรแกรมที่ใช้ในการบันทึก และวิเคราะห์ท่าทางตามเวลาจริง โดยในส่วนแรกนั้นสามารถแบ่งย่อยออกเป็น 3 ส่วนด้วยกันซึ่งประกอบด้วย ส่วนที่หนึ่งเป็นส่วนในการจัดการลบข้อมูลที่มีความผิดพลาดจากการชุดข้อมูล โดยข้อมูลที่ได้ออกมาจากกล้อง Kinect อาจมีความผิดพลาดซึ่งข้อมูลเหล่านี้ควรทำการแยกออกก่อนนำชุดข้อมูลไปทำการเรียนรู้ เพื่อไม่ให้ส่งผลต่อการเรียนรู้ที่ผิดพลาด ซึ่งหน้าจอในการทำงานถูกแสดงในรูปที่ 3.13 ส่วนที่สองเป็นส่วนในการจัดการดึงข้อมูลที่ถูกลบกลับเข้ามาสู่ชุดข้อมูล โดยในส่วนนี้ถูกพัฒนาขึ้นมาช่วยในการนำข้อมูลกลับสู่ชุดข้อมูลในกรณีผู้ใช้งานทำการลบข้อมูลที่ผิดพลาด และต้องการนำข้อมูลนั้นกลับมาใช้งานอีกครั้ง ซึ่งหน้าจอการทำงานในส่วนนี้ได้ถูกแสดงไว้ในรูปที่ 3.14 และส่วนสุดท้ายคือส่วนที่ใช้ในการสร้างไฟล์ CSV ซึ่งเป็นไฟล์ที่นำไปใช้ในการเรียนรู้ชุดข้อมูลเพื่อสร้างเป็นโมเดลในโปรแกรม Weka ซึ่งมีหน้าจอการทำงานดังแสดงในรูปที่ 3.15 และ โครงสร้างของ Class Diagram ของโปรแกรมที่ใช้ในการจัดการชุดข้อมูลได้ถูกแสดงไว้ในรูปที่ 3.16



รูปที่ 3.13 หน้าจอการทำงานของโปรแกรมการจัดการชุดข้อมูลในหน้าจอการลบข้อมูล

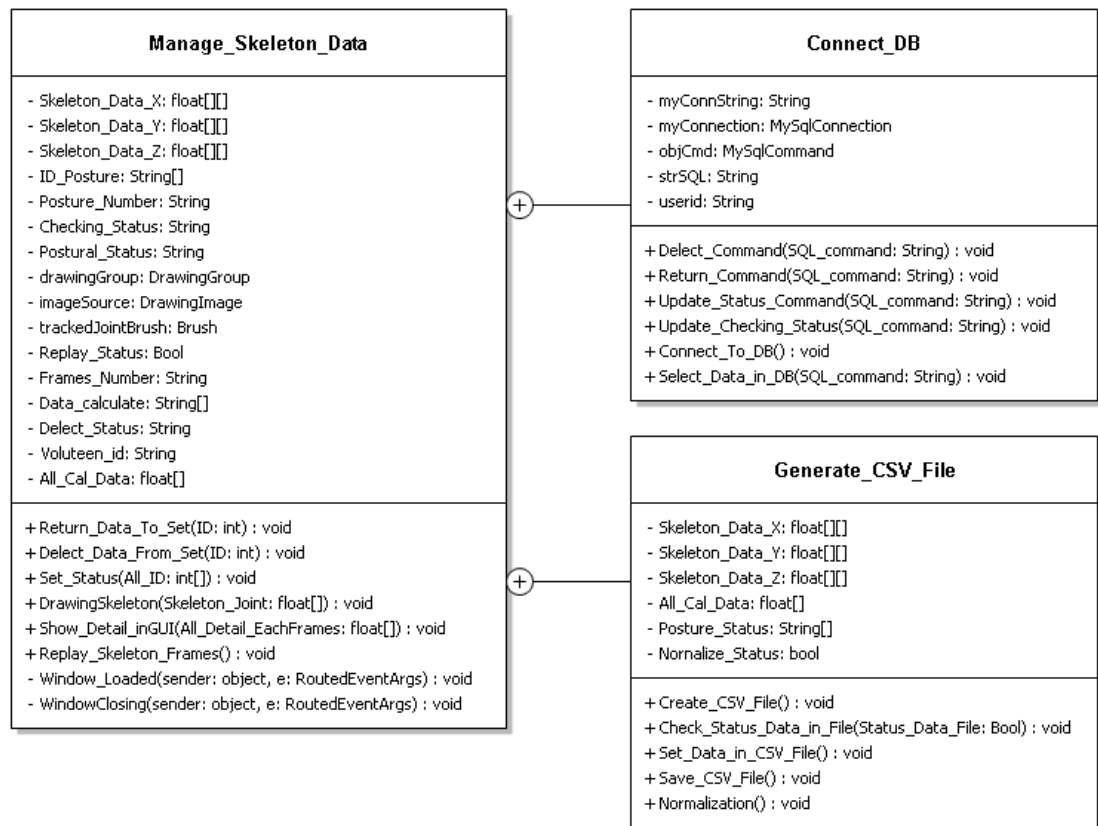


รูปที่ 3.14 หน้าจอการทำงานของโปรแกรมการจัดการชุดข้อมูลในหน้าของการดึงข้อมูลที่ถูกลบกลับ
สู่ชุดข้อมูลอีกครั้ง



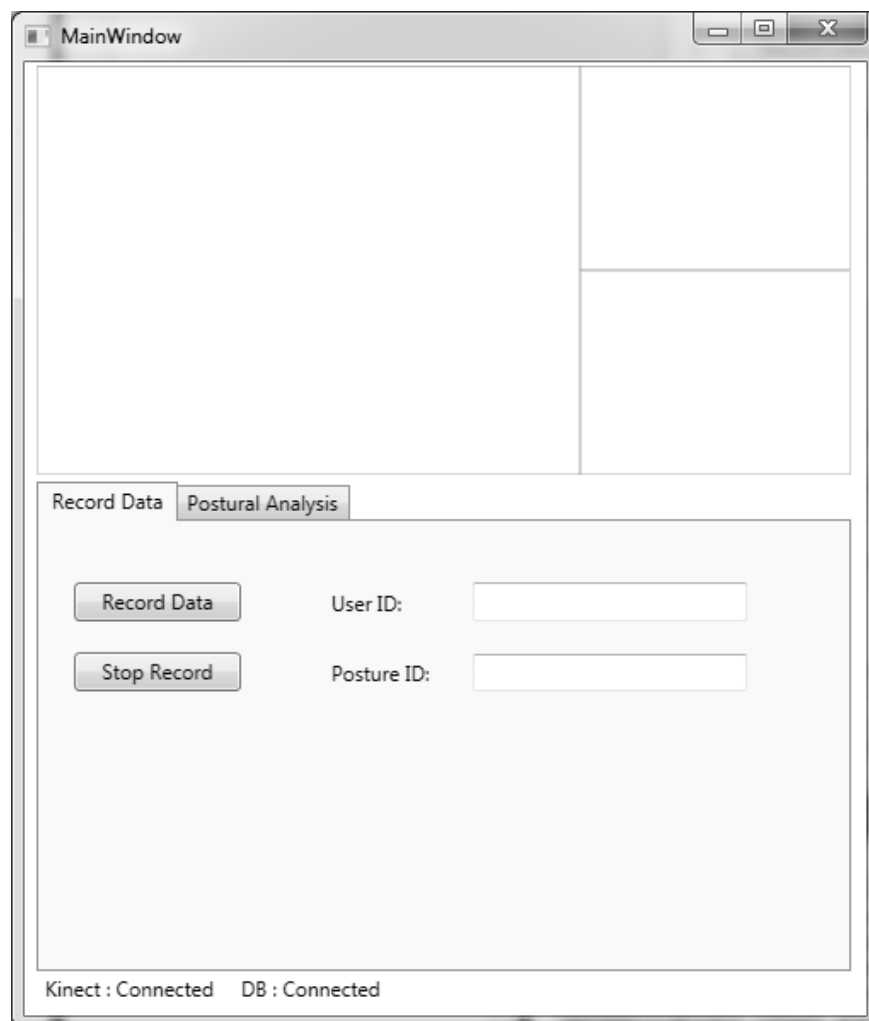
รูปที่ 3.15 หน้าจอการทำงานของโปรแกรมการจัดการชุดข้อมูลในหน้าสร้างไฟล์ CSV

Class Diagram

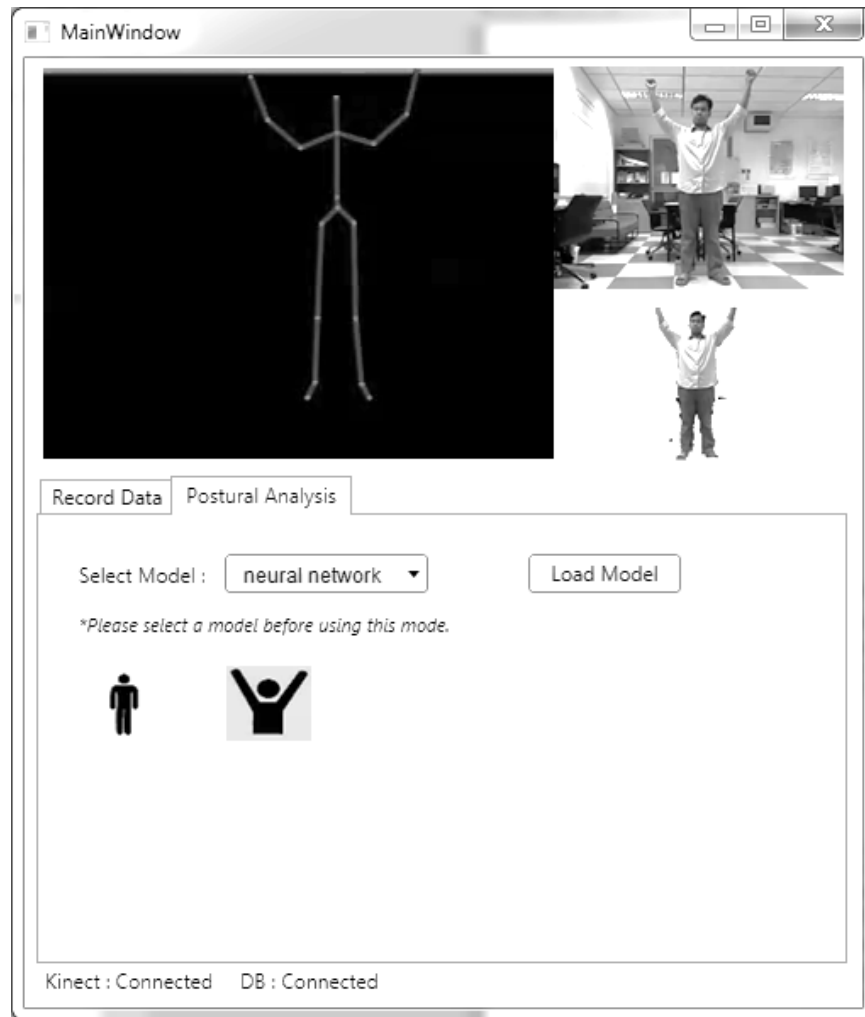


รูปที่ 3.16 การออกแบบ Class Diagram ในส่วนของโปรแกรมที่ใช้ในการจัดการชุดข้อมูล

โดยในส่วนของโปรแกรมที่ใช้ในการจัดการชุดข้อมูลเพื่อใช้ในการเรียนรู้สามารถบ่งชี้แยกได้เป็น 2 ส่วนด้วยกันได้แก่ ส่วนที่ใช้ในการวิเคราะห์ข้อมูลท่าทางตามเวลาจริงและส่วนที่ใช้ในการบันทึกข้อมูลสำหรับการเรียนรู้เพื่อสร้างโมเดล โดยหน้าจอการทำงานนั้นได้ถูกแสดงไว้ในรูปที่ 3.17 และ 3.18 ตามลำดับ โดยโครงสร้างของ Class Diagram ของโปรแกรมที่ใช้ในการบันทึกและวิเคราะห์ข้อมูลได้ถูกแสดงไว้ในรูปที่ 3.19

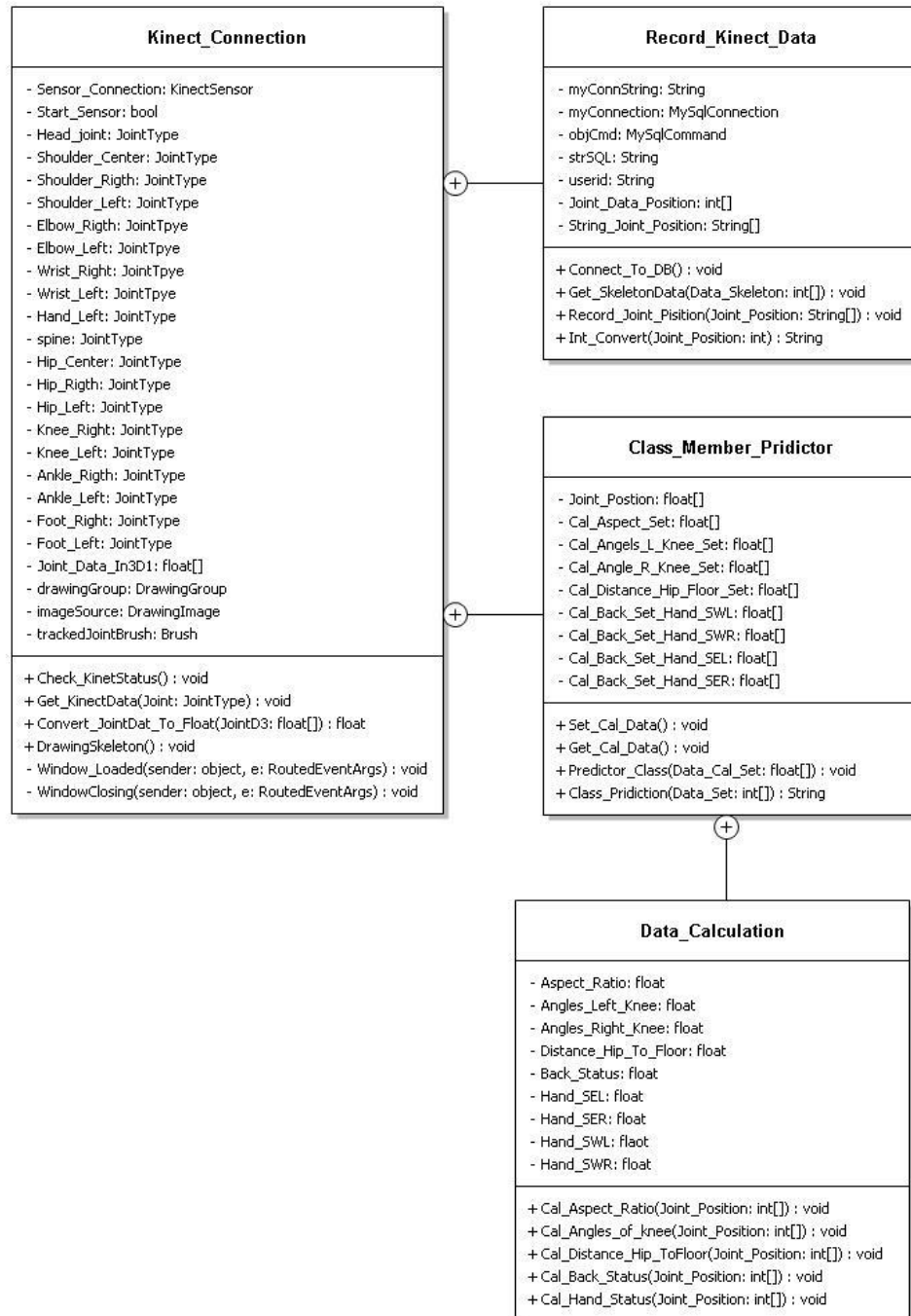


รูปที่ 3.17 หน้าจอการทำงานของโปรแกรมที่ใช้ในการบันทึกและวิเคราะห์ท่าทางตามเวลาจริงใน
โมเดลการบันทึกข้อมูล



รูปที่ 3.18 หน้าจอการทำงานของโปรแกรมที่ใช้ในการบันทึกและวิเคราะห์ท่าทางตามเวลาจริงใน
โหมดการวิเคราะห์ข้อมูลท่าทาง

Class Diagram



รูปที่ 3.19 การออกแบบ Class Diagram ในส่วนของโปรแกรมบันทึกและวิเคราะห์ข้อมูลท่าทาง

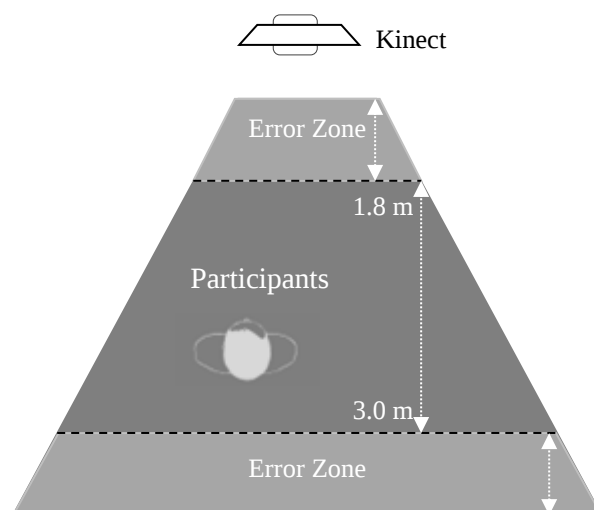
บทที่ 4 ผลการวิจัย

ในบทนี้จะกล่าวถึงความแม่นยำในการวิเคราะห์ท่าทางจากการทดสอบความถูกต้องของวิธีการวิเคราะห์ท่าทางที่ได้นำเสนอไว้

4.1 รายละเอียดการทดลอง

4.1.1 Testing Set Details

ในการทดลองของงานวิจัยชิ้นนี้ได้ทำการทดสอบระบบที่ถูกพัฒนาขึ้นกับผู้ร่วมการทดลองจำนวน 10 คน ซึ่งเป็นคนละกลุ่มกับกลุ่มที่ใช้ในการบันทึกข้อมูลที่ใช้ในการทำการเรียนรู้ โดยแบ่งเป็นเพศชายจำนวน 4 คนและเพศหญิงจำนวน 6 คน กลุ่มผู้ทดสอบนั้นมีอายุระหว่าง 45-70 ปี มีความสูงระหว่าง 155-175 เซนติเมตร ซึ่งมีรูปร่างที่แตกต่างกันออกไป ระยะการทดสอบหน้ากล้อง Kinect ในช่วงระหว่าง 1.8-3.0 เมตร ดังแสดงในรูปที่ 4.1 โดยใช้ Kinect for Windows SDK Library Function ในการวิเคราะห์โครงสร้างของร่างกาย ผู้ร่วมการทดลองทุกคนจะถูกขอให้แสดงท่าทางต่าง ๆ ตามที่ได้กำหนดไว้หน้ากล้อง Kinect โดยผู้ร่วมการทดลองสามารถแสดงท่าทางตามที่ได้กำหนดไว้อย่างอิสระ และผู้ร่วมการทดลองทุกคนจะต้องแสดงแต่ละท่าทางหน้ากล้อง Kinect เป็นจำนวน 5 คลิป ซึ่งข้อมูลที่บันทึกได้จากทุกท่าทางมีทั้งหมด 900 คลิปมีจำนวนเฟรมทั้งหมด 84,690 เฟรมโดยรายละเอียดของจำนวนเฟรมในแต่ละท่าทางได้ถูกแสดงไว้ในตารางที่ 4.1 และรายละเอียดของสัดส่วนของความกว้างของไหล่และความสูงของร่างกายผู้ร่วมการทดลองได้ถูกแสดงไว้ในตารางที่ 4.2



รูปที่ 4.1 ช่วงระยะทางที่ใช้ในการทดลอง

ตารางที่ 4.1 จำนวนชุดข้อมูลในแต่ละท่าทางของชุด Testing set

Posture	Testing Set (Frames)
Normal stand	4,367
Stand both hands up	4,985
Stand left hand up	4,890
Stand right hand up	4,785
Stand lean forward	4,360
Sit straight back	4,109
Sit lean forward	4,958
Sit lean backward	4,583
Sit both hands up	4,639
Sit left hand up	5,380
Sit right hand up	4,269
Sit on floor straight back	4,602
Sit on floor lean forward	5,079
Sit on floor lean backward	4,489
Sit on floor both hands up	4,753
Sit on floor left hand up	4,789
Sit on floor right hand up	4,673
Lie down in various manners	4,980

ตารางที่ 4.2 ความกว้างของไหล่และความสูงของร่างกายผู้ร่วมการทดลองในส่วนของ Testing Set

	<i>T-11</i>	<i>T-12</i>	<i>T-13</i>	<i>T-14</i>	<i>T-15</i>	<i>T-16</i>	<i>T-17</i>	<i>T-18</i>	<i>T-19</i>	<i>T-20</i>
	<i>Male</i>	<i>Male</i>	<i>Male</i>	<i>Male</i>	<i>Female</i>	<i>Female</i>	<i>Female</i>	<i>Female</i>	<i>Female</i>	<i>Female</i>
<i>Height</i>	168 cm.	175 cm.	169 cm.	170 cm.	155 cm.	165 cm.	154 cm.	168 cm.	169 cm.	170 cm.
<i>Shoulder Width</i>	23 in.	20 in.	18 in.	19 in.	16 in.	17 in.	16 in.	17 in.	14 in.	18 in.

4.2 ผลลัพธ์จากการคัดเลือกและจัดลำดับข้อมูลโดยหลักการ Data Selection

ในหัวข้อนี้จะนำเสนอลำดับของข้อมูลที่ถูกคัดเลือกและจัดลำดับโดยหลักการ Data Selection ซึ่งในตารางที่ 4.3 แสดงการจัดลำดับชุดข้อมูลโดยหลักการของ Gain Ratio Attribute Evaluator และในตารางที่ 4.4 นั้นได้แสดงการจัดลำดับชุดข้อมูลโดยหลักการของ Info-Gain Attribute Evaluator

ตารางที่ 4.3 ผลลัพธ์ในการจัดลำดับของชุดข้อมูลโดยหลักการ Gain Ratio Attribute Evaluator

Gain Ratio Attribute Evaluator					
Top 10 Ranker		Top 20 Ranker		Top 30 Ranker	
0.3842	Elbow left y	0.3401	Hip right y	0.2015	Spine z
0.3814	Hand right y	0.3384	Hip center y	0.196	Hip right z
0.378	Wrist right y	0.3327	Hip left y	0.1904	Hip left z
0.3775	Head y	0.2279	Knee right y	0.1892	Shoulder center z
0.3749	Elbow right y	0.2183	Knee left y	0.1858	Foot left z
0.3642	Wrist left y	0.212	Ankle left z	0.1833	Foot right z
0.3616	Shoulder center y	0.2119	Knee left z	0.1711	Shoulder right z
0.359	Shoulder left y	0.2059	Ankle right z	0.171	Knee left x
0.3552	Shoulder right y	0.2052	Knee right z	0.1675	Head z
0.3471	Spine y	0.2038	Hip center z	0.1673	Ankle right y
Top 40 Ranker		Top 50 Ranker		Top 60 Ranker	
0.1634	Shoulder left z	0.1296	Foot left y	0.1108	Wrist left z
0.1593	Knee right x	0.1283	Elbow right z	0.1076	Foot right x
0.1582	Ankle left y	0.1267	Foot left x	0.0987	Elbow left x
0.154	Ankle left x	0.1264	Shoulder right x	0.0959	Elbow right x
0.1471	Hip center x	0.1242	Elbow left z	0.0909	Wrist right x
0.1446	Spine x	0.1239	Shoulder left x	0.0879	Hand right x
0.1427	Hip left x	0.1205	Wrist right z	0.0866	Wrist left x
0.1413	Foot right y	0.1204	Shoulder center x	0.0822	Hand left z
0.1395	Ankle right x	0.1179	Hand right z	0.0822	Hand left y
0.1384	Hip right x	0.1151	Head x	0.0822	Hand left x

ตารางที่ 4.4 ผลลัพธ์ในการจัดลำดับของชุดข้อมูล โดยหลักการ Info Gain Attribute Evaluator

Info Gain Attribute Evaluator					
Top 10 Ranker		Top 20 Ranker		Top 30 Ranker	
2.4095	Shoulder right y	2.2474	Hip left y	1.4642	Knee left y
2.4048	Shoulder left y	2.2457	Hip center y	1.3847	Hip right z
2.3795	Head y	2.2141	Wrist left y	1.3752	Hip left z
2.3736	Shoulder center y	1.6251	Ankle left z	1.3436	Foot left z
2.3186	Spine y	1.6019	Knee left z	1.3191	Foot right z
2.2956	Hip right y	1.6003	Knee right y	1.3137	Knee left x
2.2818	Elbow right y	1.5731	Ankle right z	1.2884	Shoulder center z
2.2803	Elbow left y	1.5699	Knee right z	1.1916	Shoulder right z
2.27	Wrist right y	1.4684	Spine z	1.1765	Knee right x
2.2677	Hand right y	1.4666	Hip center z	1.1696	Ankle left x
Top 40 Ranker		Top 50 Ranker		Top 60 Ranker	
1.1343	Head z	0.8796	Shoulder right x	0.7145	Hand right z
1.1237	Shoulder left z	0.8625	Foot right y	0.6766	Wrist left z
1.0736	Ankle right y	0.8511	Shoulder left x	0.6433	Elbow left x
1.0587	Hip center x	0.8368	Elbow right z	0.6286	Elbow right x
1.0484	Spine x	0.8197	Shoulder center x	0.5733	Wrist right x
1.0424	Ankle left y	0.8045	Foot left y	0.5431	Hand right x
1.034	Ankle right x	0.7923	Elbow left z	0.5368	Wrist left x
1.0232	Hip left x	0.775	Head x	0.489	Hand left z
0.9805	Hip right x	0.7596	Wrist right z	0.489	Hand left y
0.8977	Foot left x	0.7405	Foot right x	0.489	Hand left x

จากการคัดเลือกและจัดลำดับของทั้ง 2 หลักการ พบว่ามีการจัดลำดับที่มีลักษณะคล้ายกัน ซึ่งค่าพิกัดในแนวแกน Y ถูกจัดลำดับไว้ในลำดับต้นเกือบทั้งหมด และค่าในแนวแกน Z ถูกจัดลำดับความสำคัญในลำดับถัดมา โดยที่ค่าในแนวแกน X เกือบทั้งหมดถูกจัดลำดับไว้ในกลุ่มท้ายสุด ซึ่งทำให้เห็นได้ว่าข้อมูลในแนวแกน Y นั้นถูกให้ความสำคัญมากที่สุดในการเรียนรู้ ในทางตรงกันข้ามสำหรับข้อมูลในแนวแกน X ซึ่งเป็นค่าในการบอกพิกัดในแนวนอนของกล้อง Kinect นั้นเป็นชุดข้อมูลที่มีความสำคัญน้อยที่สุดต่อการเรียนรู้ในชุดข้อมูล

4.3 ผลการทดลอง

อัตราความแม่นยำ (Accuracy) ในการวิเคราะห์ท่าทางของงานวิจัยชิ้นนี้ถูกแบ่งออกเป็น 3 ส่วนด้วยกัน ซึ่งประกอบด้วยส่วนของการทดสอบความแม่นยำในการวิเคราะห์ท่าทางของชุดข้อมูลที่ไม่ได้ผ่านการทำ Data Selection และ Data Transformation ส่วนของการทดสอบความแม่นยำในการวิเคราะห์ท่าทางของชุดข้อมูลที่ได้ผ่านการทำ Data Selection และส่วนของการทดสอบความแม่นยำในการวิเคราะห์ท่าทางของชุดข้อมูลที่ได้ผ่านการทำ Data Transformation ซึ่งรายละเอียดของอัตราความแม่นยำในการวิเคราะห์ท่าทางของแต่ละชุดข้อมูลถูกแสดงไว้ในตารางที่ 4.5-4.8 ตามลำดับ โดยรายละเอียดข้อมูลที่ถูกแสดงไว้ในแต่ละตารางนั้นคือความถูกต้องในการวิเคราะห์ท่าทางซึ่งอยู่ในหน่วยเปอร์เซ็นต์

ตารางที่ 4.5 ค่าความแม่นยำจากการทดสอบชุดข้อมูลที่ไม่ได้ผ่านการทำ Data Selection และ Data Transformation

All 20 Joints									
Single-stage Classifier					Multiple-stage Classifier				
	NN	NB	LGT	DT		NN	NB	LGT	DT
N-NML	24.38	18.29	22.40	20.19	N-NML	25.39	18.17	19.97	19.65
MM-NML	26.78	19.89	24.42	21.89	MM-NML	27.13	20.89	23.45	22.56
Z-NML	26.56	19.73	23.89	21.78	Z-NML	25.78	20.17	22.56	22.09

จากการทดลองนำชุดข้อมูลที่ไม่ได้ผ่านการทำ Data Selection และ Data Transformation ไปทำการเรียนรู้โดยอัลกอริทึมต่าง ๆ ที่ได้เลือกไว้สำหรับการสร้างโมเดลเพื่อนำไปวิเคราะห์ท่าทางของกลุ่มผู้ทดสอบ พบว่าให้ค่าความแม่นยำในการวิเคราะห์ท่าทางอยู่ในเกณฑ์ที่ต่ำมาก ซึ่งความแม่นยำที่สามารถวิเคราะห์ได้นั้นอยู่ช่วง 18-27 เปอร์เซ็นต์เท่านั้น โดยการทำ Normalization กับชุดข้อมูลช่วยเพิ่มความแม่นยำในการวิเคราะห์ข้อมูลได้เพียงเล็กน้อย โดยอัลกอริทึมที่สามารถทำการเรียนรู้และสร้างโมเดลเพื่อใช้ในการวิเคราะห์ข้อมูลได้ดีที่สุดสำหรับข้อมูลชุดนี้คืออัลกอริทึม Neural Network ที่เรียนรู้ชุดข้อมูลที่ได้ผ่านการทำ Max-Min Normalization ในแบบ Multiple-stage Classifier

ตารางที่ 4.6 ค่าความแม่นยำจากการทดสอบชุดข้อมูลที่ทำ Data Selection โดยหลัก Gain Ratio Attribute Evaluator

Data Selection									
Single-stage Classifier					Multiple-stage Classifier				
Gain 10									
	NN	NB	LGT	DT		NN	NB	LGT	DT
N-NML	17.07	14.73	19.40	19.91	N-NML	18.89	15.87	19.35	20.72
MM-NML	23.09	20.05	22.89	21.45	MM-NML	24.87	20.32	23.67	21.67
Z-NML	22.34	19.45	21.45	20.89	Z-NML	23.51	19.08	22.56	20.89
Gain 20									
	NN	NB	LGT	DT		NN	NB	LGT	DT
N-NML	24.45	19.89	23.21	21.71	N-NML	29.89	22.31	27.90	24.01
MM-NML	28.67	22.56	28.21	24.90	MM-NML	36.40	24.72	33.34	27.83
Z-NML	28.56	22.56	27.01	24.87	Z-NML	36.56	24.61	34.21	27.93
Gain 30									
	NN	NB	LGT	DT		NN	NB	LGT	DT
N-NML	31.89	24.50	27.45	25.61	N-NML	39.05	27.34	32.23	29.45
MM-NML	38.68	29.04	29.50	33.64	MM-NML	49.89	35.90	46.89	41.89
Z-NML	38.41	28.31	28.95	33.21	Z-NML	48.51	33.49	46.90	41.69
Gain 40									
	NN	NB	LGT	DT		NN	NB	LGT	DT
N-NML	26.15	23.04	26.89	26.90	N-NML	31.51	26.14	27.50	23.45
MM-NML	29.80	25.43	29.36	31.87	MM-NML	33.60	28.56	31.64	27.94
Z-NML	29.75	26.83	28.46	30.65	Z-NML	32.45	27.60	30.99	26.45
Gain 50									
	NN	NB	LGT	DT		NN	NB	LGT	DT
N-NML	25.76	22.51	22.98	24.01	N-NML	27.38	24.37	20.13	20.14
MM-NML	28.89	24.78	24.04	26.74	MM-NML	29.80	26.75	24.58	24.30
Z-NML	29.08	24.53	23.89	25.67	Z-NML	28.90	25.67	24.32	22.10

ตารางที่ 4.7 ค่าความแม่นยำจากการทดสอบชุดข้อมูลที่ผ่านมาการทำ Data Selection โดยหลัก Info Gain Attribute Evaluator

Data Selection									
Single-stage Classifier					Multiple-stage Classifier				
Info-Gain 10									
	NN	NB	LGT	DT		NN	NB	LGT	DT
N-NML	18.45	13.71	16.48	15.96	N-NML	21.50	15.57	21.45	17.98
MM-NML	27.87	23.45	25.34	24.83	MM-NML	29.45	22.45	26.50	21.45
Z-NML	27.94	24.67	25.96	24.45	Z-NML	29.32	23.56	25.76	21.89
Info-Gain 20									
	NN	NB	LGT	DT		NN	NB	LGT	DT
N-NML	27.46	26.34	28.94	26.89	N-NML	35.54	26.30	29.38	28.75
MM-NML	33.67	28.90	31.90	28.90	MM-NML	38.54	29.09	35.56	31.45
Z-NML	32.67	27.56	31.85	27.57	Z-NML	38.90	29.50	34.56	32.45
Info-Gain 30									
	NN	NB	LGT	DT		NN	NB	LGT	DT
N-NML	33.87	28.98	32.51	30.31	N-NML	38.59	28.49	38.90	37.23
MM-NML	37.86	30.45	38.57	34.67	MM-NML	46.90	35.90	42.90	41.56
Z-NML	37.56	31.04	37.60	34.56	Z-NML	46.53	35.37	41.68	41.69
Info-Gain 40									
	NN	NB	LGT	DT		NN	NB	LGT	DT
N-NML	29.40	27.40	28.99	24.78	N-NML	32.67	24.56	26.89	23.00
MM-NML	34.49	25.89	33.05	31.09	MM-NML	34.90	29.09	32.90	27.89
Z-NML	33.56	26.89	32.90	31.90	Z-NML	35.87	28.45	33.89	26.45
Info-Gain 50									
	NN	NB	LGT	DT		NN	NB	LGT	DT
N-NML	22.41	19.53	23.54	21.09	N-NML	24.90	18.40	22.41	21.87
MM-NML	25.45	21.60	25.70	25.67	MM-NML	26.90	24.39	26.47	24.50
Z-NML	26.45	23.67	26.89	24.56	Z-NML	27.90	24.49	26.78	24.90

จากการทดลองพบว่าชุดข้อมูลที่ผ่านการทำ Data Selection ในทั้ง 2 วิธีนั้นให้ค่าความแม่นยำในการวิเคราะห์ทำทางที่ใกล้เคียงกัน ซึ่งชุดข้อมูลที่ให้ค่าความแม่นยำการวิเคราะห์ทำทางที่ดีที่สุดจากทั้ง 2 วิธีคือชุดข้อมูลที่เลือกข้อมูลใน 30 อันดับแรก ซึ่งให้ค่าความแม่นยำอยู่ในช่วงระหว่าง 24-49 เปอร์เซนต์โดยการทำ Normalization กับข้อมูลชุดนี้ช่วยเพิ่มความแม่นยำในการวิเคราะห์ข้อมูลให้แก่ชุดข้อมูลได้ประมาณ 3-10 เปอร์เซนต์ แม้ว่าค่าความแม่นยำที่ได้นี้จะดีกว่าชุดข้อมูลที่ไม่ได้ผ่านการทำ Data Selection และ Data Transformation แต่ก็ยังให้ค่าความแม่นยำในการวิเคราะห์ข้อมูลที่ค่อนข้างต่ำ และยังไม่เหมาะสมที่จะนำมาใช้ในการวิเคราะห์ข้อมูล

ตารางที่ 4.8 ค่าความแม่นยำจากการทดสอบชุดข้อมูลที่ผ่านการทำ Data Transformation

Data Transformation									
Single-stage Classifier					Multiple-stage Classifier				
	NN	NB	LGT	DT		NN	NB	LGT	DT
N-NML	68.43	58.34	64.56	59.50	N-NML	89.77	79.30	87.71	86.79
MM-NML	75.45	64.90	73.45	67.90	MM-NML	91.90	82.40	89.08	89.01
Z-NML	75.60	65.89	76.90	68.90	Z-NML	90.89	83.45	90.80	91.08

จากการทดลองพบว่าชุดข้อมูลที่ผ่านการทำ Data Transformation นั้นให้ค่าความแม่นยำในการวิเคราะห์ทำทางที่ดีที่สุดเมื่อเทียบกับชุดข้อมูลของอีกสองวิธี โดยในแบบ Single-stage Classifier นั้นมีค่าความแม่นยำอยู่ในช่วง 64-76 เปอร์เซนต์ แต่อย่างไรก็ตามแม้ว่าชุดข้อมูลในแบบ Single-stage Classifier นั้นจะมีความสามารถในการวิเคราะห์ทำทางที่ดีขึ้น แต่ก็ยังคงมีค่าความแม่นยำที่ไม่มีความน่าเชื่อถือที่จะนำมาใช้ในการวิเคราะห์ข้อมูล ซึ่งในทางตรงกันข้ามพบว่าชุดข้อมูลในแบบ Multiple-stage Classifier นั้นสามารถใช้ในการเรียนรู้เพื่อสร้างโมเดลที่วิเคราะห์ทำทางให้มีความถูกต้องตรงตามทำทางจริงของผู้ทดสอบได้ดีที่สุดในทุกอัลกอริทึม โดยให้ค่าความแม่นยำในการวิเคราะห์ทำทางอยู่ในช่วง 79-91 เปอร์เซนต์ สำหรับวิธีนี้(Multiple-stage Classifier) การทำ Normalization นั้นให้ค่าความแม่นยำที่ใกล้เคียงกับชุดข้อมูลที่ไม่ได้ทำการ Normalization โดยอัลกอริทึมที่สามารถเรียนรู้เพื่อสร้างโมเดลสำหรับใช้ในการวิเคราะห์ข้อมูลได้ดีที่สุดคือ Neural Network ที่ผ่านการทำ Max-Min Normalization ในแบบ Multiple-stage Classifier โดยให้ค่าความแม่นยำจากการวิเคราะห์ข้อมูลอยู่ที่ 91.90 เปอร์เซนต์

บทที่ 5 สรุปผลการวิจัย

ในบทนี้จะกล่าวถึงการอภิปรายผลสรุป ข้อจำกัด ข้อเสนอแนะ และแนวทางในการนำงานวิจัยชิ้นนี้ไปประยุกต์ใช้และพัฒนาต่อยอด

5.1 การอภิปรายสรุปผลการทดลอง

ในงานวิจัยชิ้นนี้ได้นำเสนอวิธีการวิเคราะห์ทำทาง โดยใช้ข้อมูล 3 มิติจากกล้อง Kinect มาประยุกต์ใช้กับการทำเหมืองข้อมูลเพื่อใช้ในการแยกแยะทำทาง โดยได้นำเสนอวิธีลดความซับซ้อนภายในชุดข้อมูล 3 มิติที่ได้จากกล้อง Kinect เพื่อเพิ่มประสิทธิภาพในการเรียนรู้ชุดข้อมูล ซึ่งวิธีที่งานวิจัยนี้เลือกใช้เพื่อลดความซับซ้อนของชุดข้อมูลมีด้วยกัน 2 วิธีคือ Data Selection และ Data Transformation ชุดอัลกอริทึมที่ใช้ในการเรียนรู้จากชุดข้อมูล เพื่อใช้ในการสร้างโมเดลสำหรับการวิเคราะห์ทำทางที่งานวิจัยนี้ได้เลือกใช้มีทั้งหมด 4 อัลกอริทึมคือ Neural Network, Naïve Bayes, Logistic Regression และ Decision Tree ข้อมูลที่ใช้ในการเรียนรู้ในงานวิจัยชิ้นนี้ถูกบันทึกจากผู้เข้าร่วมการทดลองจำนวน 10 คน และมีการทดสอบความถูกต้องกับผู้ทดสอบ 10 คน ซึ่งเป็นคนละกลุ่มกับกลุ่มคนที่ใช้ในการบันทึกข้อมูล

อัตราความแม่นยำจากการทดลองถูกแบ่งออกเป็น 3 ส่วนด้วยกันประกอบด้วย ความแม่นยำในการวิเคราะห์ทำทางโดยใช้ชุดข้อมูลที่ไม่ได้ผ่านการทำ Data Selection และ Data Transformation ความแม่นยำในการวิเคราะห์ทำทางโดยใช้ชุดข้อมูลที่ได้ผ่านการทำ Data Selection และความแม่นยำในการวิเคราะห์ทำทางของชุดข้อมูลที่ได้ผ่านการทำ Data Transformation ซึ่งพบว่าชุดข้อมูลที่ไม่ได้ผ่านการทำ Data Selection และ Data Transformation ให้ค่าความแม่นยำในการวิเคราะห์ทำทางที่ต่ำมากทั้งข้อมูลในแบบ Single-stage Classifier Model และ Multiple-stage Classifier Model และการทำ Normalization ยังคงไม่สามารถช่วยให้การเรียนรู้ในชุดข้อมูลมีความถูกต้องเพิ่มขึ้น โดยความแม่นยำในการวิเคราะห์ข้อมูลที่ดีที่สุดของชุดข้อมูลมีค่าเพียง 27.13 เปอร์เซ็นต์เท่านั้น สาเหตุที่ทำให้โมเดลที่ได้จากแต่ละอัลกอริทึมมีความสามารถในการวิเคราะห์ข้อมูลได้ต่ำนั้นเกิดมาจากชุดข้อมูลยังคงมีข้อมูลที่ไม่จำเป็นต่อการเรียนรู้อยู่ค่อนข้างมาก จึงทำให้การเรียนรู้ชุดข้อมูลเกิดความผิดพลาด โดยข้อมูลที่ไม่จำเป็นต่อการเรียนรู้ คือข้อมูลในแนวแกน X ซึ่งเป็นส่วนบอกถึงระยะในแนวนอน ซึ่งไม่สามารถช่วยในการบ่งบอกถึงทำทางต่าง ๆ ได้ โดยจะเห็นได้จากการคัดเลือกชุดข้อมูลของหลักการ Data Selection นั้นจะจัดลำดับความสำคัญของค่าในแนวแกน X ไว้ในกลุ่มท้ายสุดเกือบทั้งหมด และในทางกลับกันพบว่าค่าพิกัดในแนวแกน Y นั้นเป็นข้อมูลที่มีความสำคัญต่อการเรียนรู้มากที่สุด ซึ่งจะ

เห็นได้จากการจัดลำดับความสำคัญของชุดข้อมูลในหลักการ Gain Ratio Attribute Evaluator และ Info Gain Attribute Evaluator ได้จัดลำดับค่า Y ไว้ในกลุ่มแรกทั้งคู่ดังแสดงในตารางที่ 4.1 และ 4.2

จากการเรียนรู้ของชุดข้อมูลที่ผ่านมาการทำ Data Selection พบว่าสามารถวิเคราะห์ทำทางได้แม่นยำกว่าชุดข้อมูลที่ไม่ได้ผ่านการทำ Data selection และ Data transformation โดยโมเดลที่สามารถวิเคราะห์ข้อมูลได้ดีที่สุดในหลักการนี้ เป็นโมเดลที่ได้จากการเรียนรู้ชุดข้อมูล 30 อันดับแรก ซึ่งพบว่าชุดข้อมูล 30 อันดับแรกนั้นมีค่าในแนวแกน X อยู่ค่อนข้างน้อยจึงทำให้การเรียนรู้มีความถูกต้องมากยิ่งขึ้น แต่อย่างไรก็ตามผลลัพธ์ที่ได้จากทดสอบข้อมูลนั้นก็ยังคงมีความผิดพลาดค่อนข้างสูง โดยความแม่นยำในการวิเคราะห์ทำทางที่ดีที่สุดของหลักการนี้มีค่าเพียง 49.89 เปอร์เซนต์เท่านั้น สาเหตุที่ชุดข้อมูลนี้ยังทำให้โมเดลมีความสามารถในการเรียนรู้ค่อนข้างต่ำนั้นอาจเกิดมาจากชุดข้อมูลที่ใช้ในการเรียนรู้ นั้นยังมีปริมาณที่ไม่มากพอต่อการเรียนรู้ชุดข้อมูล จึงมีความเป็นไปได้ที่หากมีการเพิ่มปริมาณชุดข้อมูลให้มีความละเอียดมากยิ่งขึ้นอาจส่งผลให้โมเดลที่ได้จากการเรียนรู้สามารถวิเคราะห์ข้อมูลได้ถูกต้องและแม่นยำมากยิ่งขึ้น

ในส่วน of ชุดข้อมูลที่ผ่านมาการทำ Data Transformation นั้นเป็นกลุ่มที่มีค่าความแม่นยำในการวิเคราะห์ทำทางที่ดีที่สุดจากข้อมูลทั้งสามแบบ โดยโมเดลในแบบ Single-stage Classifier Model ให้ค่าความแม่นยำในการวิเคราะห์ทำทางอยู่ช่วง 64-76 เปอร์เซนต์ ซึ่งให้ค่าความแม่นยำที่ดีกว่าชุดข้อมูลทั้งสองแบบที่ได้กล่าวมาก่อนหน้านี้ แต่ก็ยังคงไม่มีความน่าเชื่อถือมากพอที่จะนำมาใช้ในการวิเคราะห์ทำทาง แต่โมเดลในแบบ Multiple-stage Classifier Model เป็นโมเดลที่สามารถวิเคราะห์ทำทางได้ดีที่สุด โดยให้ค่าความแม่นยำในการวิเคราะห์ทำทางอยู่ในช่วง 79-91 เปอร์เซนต์ จึงทำให้สามารถสรุปได้ว่า การนำชุดข้อมูลดิบที่สามารถวิเคราะห์ได้จากกล้อง Kinect มาทำการเปลี่ยนแปลงเพื่อสร้างชุดข้อมูลใหม่สามารถทำให้การเรียนรู้มีความแม่นยำมากที่สุด ซึ่งปัจจัยหลักที่ทำให้ชุดข้อมูลนี้สามารถทำให้การเรียนรู้ได้อย่างถูกต้องคือคุณลักษณะ (Feature) ที่จะนำมาใช้ในการเรียนรู้ หากคุณลักษณะที่นำมาใช้เป็นลักษณะเด่นและเกิดขึ้นเสมอในทุกครั้งที่มีการแสดงทำทางนั้น ๆ ก็จะสามารถทำให้การแยกแยะข้อมูลของแต่ละอัลกอริทึมมีความชัดเจนมากยิ่งขึ้น นอกจากนี้การแปลงชุดข้อมูลยังช่วยลดความซับซ้อนของชุดข้อมูลลงได้อีกด้วย และ เมื่อเปรียบเทียบระหว่างโมเดลในแบบ Single-stage Classifier Model ซึ่งเป็นโมเดลที่เกิดจากการเรียนรู้ชุดข้อมูลทั้งหมดไว้ในชุดเดียวกับ Multiple-stage Classifier Model ซึ่งเป็นโมเดลที่เกิดจากการเรียนรู้ชุดข้อมูลที่มีการแยกข้อมูลออกเป็นชุดตามลักษณะของชุดข้อมูลพบว่าการแยกแยะชุดข้อมูลออกจากกันตามความสัมพันธ์ของข้อมูลสามารถทำให้ชุดข้อมูลมีความซับซ้อนน้อยลง ส่งผลให้การเรียนรู้ในแต่ละอัลกอริทึมมีความถูกต้องมากยิ่งขึ้นตามไปด้วย โดยอัลกอริทึมที่สามารถเรียนรู้เพื่อสร้างโมเดลสำหรับใช้ในการวิเคราะห์ข้อมูลได้ดีที่สุดคือ Neural Network ที่ผ่านการทำ Max-Min Normalization ในแบบ Multiple-stage Classifier โดยให้ค่าผลลัพธ์จากการวิเคราะห์ข้อมูลอยู่ที่ 91.90 เปอร์เซนต์

จากการทดลองพบว่ามีปัจจัยภายนอกที่ส่งผลต่อความแม่นยำของการวิเคราะห์ข้อมูลนั้นมีด้วยกัน 3 ปัจจัย โดยปัจจัยที่หนึ่งคือมุมมองของกล้อง Kinect ที่ใช้ในการบันทึกภาพท่าทางของผู้ร่วมการทดลอง ซึ่งหากกล้อง Kinect สามารถวิเคราะห์โครงสร้างของร่างกายผู้เข้าร่วมการทดลองได้อย่างชัดเจนจะส่งผลให้การวิเคราะห์และทำนายท่าทางมีความแม่นยำตามไปด้วย ปัจจัยที่สองคือสภาพแวดล้อมของบริเวณที่ทำการใช้งานระบบ ซึ่งหากในระหว่างการใช้งานระบบมีวัตถุมาบดบังร่างกายของผู้ร่วมการทดลองอาจส่งผลให้การวิเคราะห์มีความผิดพลาดได้ และปัจจัยที่สามคือความเร็วในการเปลี่ยนแปลงท่าทาง หากมีการเปลี่ยนแปลงท่าทางที่รวดเร็วอาจส่งผลให้กล้อง Kinect ไม่สามารถตรวจจับท่าทางได้ซึ่งจะส่งผลทำให้การวิเคราะห์ท่าทางมีความผิดพลาดเกิดขึ้นตามไปด้วย

5.2 ข้อจำกัดของงานวิจัย

งานวิจัยชิ้นนี้มีข้อจำกัดของระบบที่ใช้ในการวิเคราะห์ข้อมูลท่าทาง ดังต่อไปนี้

ข้อจำกัดด้านระยะการทำงานของกล้อง Kinect ซึ่งกล้อง Kinect นั้นมีความสามารถในการค้นหา และวิเคราะห์ท่าทางของผู้ถูกตรวจจับได้ในระยะทาง 1.8-3.0 เมตรเท่านั้น โดยหากผู้ถูกตรวจจับยืนอยู่นอกระยะการทำงานของกล้อง Kinect ก็จะทำให้ความแม่นยำในการตรวจจับร่างกายมนุษย์น้อยลง

ข้อจำกัดในการวิเคราะห์ท่าทางที่มีความซับซ้อนของ Library Function ในระหว่างการทดลองพบว่า Library Function ของกล้อง Kinect นั้นไม่สามารถที่จะวิเคราะห์ท่าทางที่มีความซับซ้อน หรือท่าทางที่มีอวัยวะซ้อนทับกันได้อย่างแม่นยำ เช่นการกอดอก หรือการนั่งไขว่ห้าง จะพบว่าหลายครั้งที่ Library Function มีการวิเคราะห์ท่าทางเหล่านั้นออกมาผิดพลาด

ข้อจำกัดด้านจำนวนบุคคลที่กล้อง Kinect สามารถตรวจจับได้ กล้อง Kinect รุ่น XBOX 360 นั้นมีความสามารถในการตรวจจับร่างกายมนุษย์ได้เพียงสองคนเท่านั้น ซึ่งในกรณีที่หน้ากล้อง Kinect มีจำนวนผู้ใช้งานครบ 2 คนแล้ว แต่มีบุคคลที่สามเดินผ่านหน้ากล้อง อาจจะทำให้กล้อง Kinect เปลี่ยนผู้ถูกตรวจจับเป็นบุคคลที่สามที่เดินผ่านหน้ากล้องได้

ข้อจำกัดด้านสภาพแวดล้อมที่ส่งผลต่อการวิเคราะห์ข้อมูลของ Library Function โดยพบว่าเมื่อร่างกายถูกบดบังด้วยวัตถุต่าง ๆ อาจทำให้การวิเคราะห์ท่าทางในส่วนของอวัยวะที่ถูกบดบังมีความผิดพลาดได้ นอกจากนี้โซฟายังเป็นหนึ่งในปัจจัยที่ก่อให้เกิดความผิดพลาดในการวิเคราะห์ข้อมูลซึ่งเมื่อผู้ถูกตรวจจับนั่งลงบนโซฟา พบว่าหลายครั้งที่กล้อง Kinect ยังไม่สามารถแยกแยะระหว่างร่างกายมนุษย์ออกพื้นผิวของโซฟาได้เป็นอย่างดี ส่งผลให้การวิเคราะห์ท่าทางอาจมีความผิดพลาดได้

ข้อจำกัดในการวิเคราะห์ท่าทางการนอนของ Library Function สำหรับท่าทางการนอนแม้ว่าจะเป็นท่าทางพื้นฐานแต่ก็ยังพบว่าการประมาณค่าจุดต่าง ๆ ของร่างกายผิดพลาดอยู่ เช่นข้อมูลพิกัดในส่วนแขนหรือขาเป็นต้นโดยอาจมีความผิดพลาดจากท่าทางจริง ซึ่งทำให้การวิเคราะห์ท่าทางการนอนนั้นดูได้เพียงภาพรวมไม่สามารถวิเคราะห์ท่าทางของการนอนอย่างละเอียดได้

5.3 ข้อเสนอแนะ

การสร้างรูปแบบเพื่อช่วยในการตัดสินใจ (Rule-Based) อาจเป็นอีกวิธีที่สามารถช่วยในการวิเคราะห์ข้อมูลให้มีความแม่นยำมากยิ่งขึ้น โดยนำ Attributes ของแต่ละ Class Member ที่สามารถแยกแยะได้อย่างชัดเจนมาทำการสร้างรูปแบบสำหรับการวิเคราะห์ข้อมูล และนำส่วนของข้อมูลที่มีความซับซ้อนจนไม่สามารถแยกแยะด้วยสายตาไปทำการเรียนรู้ด้วยอัลกอริทึมเพื่อทำการสร้างโมเดล ซึ่งจะช่วยประหยัดเวลาในการเรียนรู้ในชุดข้อมูล จึงเป็นอีกหนึ่งวิธีที่ควรนำมาประยุกต์และปรับปรุงระบบเพื่อพัฒนาระบบให้สามารถทำงานได้ถูกต้องและแม่นยำมากยิ่งขึ้น

แม้ว่าตัวจัดลำดับ (Ranker) ของหลักการ Data selection ในงานวิจัยชิ้นนี้ได้เรียงลำดับความสำคัญชุดข้อมูลที่ได้จากกล้อง Kinect ให้ค่าพิกัดในแนวแกน Y ถูกให้ความสำคัญที่สุด และจัดลำดับแนวแกน X ให้มีความสำคัญน้อยที่สุด แต่อย่างไรก็ตามในอนาคตควรมีการเลือกตัวจัดลำดับที่มีความหลากหลายมากยิ่งขึ้น เพื่อทดสอบถึงความแตกต่างในการจัดลำดับของชุดข้อมูล เช่น SVM Attribute Evaluator หรือ Relief Evaluator เป็นต้น รวมไปถึงการทดลองตัดบางแกนของชุดข้อมูลออกไปเพื่อทดสอบว่ามีความแตกต่างในการเรียนรู้ของชุดข้อมูลหรือไม่เช่น ทดสอบตัดแกน Z หรือแกน X ออกเป็นต้น

การใช้กล้อง Kinect ตัวเดียวอาจทำให้ชุดข้อมูลท่าทางที่ได้นั้นไม่ครบถ้วน ซึ่งมักเกิดในกรณีที่ผู้ถูกตรวจจับไม่ได้หันหน้าเข้าหากล้อง หรือในกรณีที่อาจมีวัตถุที่บดบังร่างกายของผู้ถูกตรวจจับซึ่งจะทำให้การวิเคราะห์ท่าทางมีความผิดพลาด ดังนั้นมุมมองของกล้องจึงเป็นอีกปัจจัยที่สำคัญต่อการวิเคราะห์ท่าทาง การเพิ่มมุมมองให้มีความหลากหลายจึงเป็นวิธีการหนึ่งในการแก้ปัญหาเรื่องมุมมองของกล้อง Kinect ดังนั้นการประยุกต์ใช้กล้องหลายตัวจะสามารถช่วยเพิ่มมุมมองซึ่งส่งผลให้ข้อมูลที่ได้นั้นมีความละเอียดมากยิ่งขึ้น เมื่อนำข้อมูลของแต่ละกล้องมารวมกันจะทำให้โครงสร้างของร่างกายมีความสมบูรณ์ส่งผลให้การวิเคราะห์ท่าทางมีความแม่นยำ และมีความน่าเชื่อถือมากยิ่งขึ้น

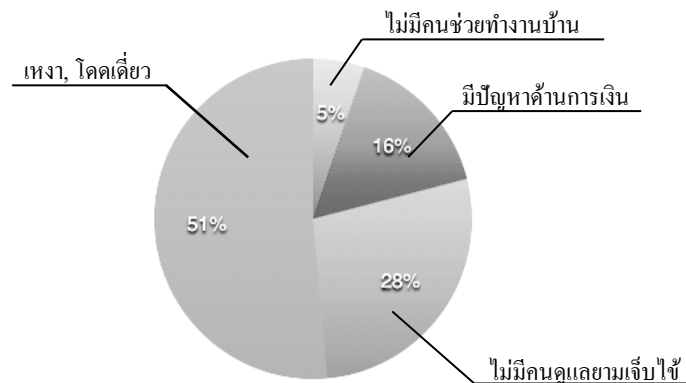
ปัจจัยของระยะการทำงานของกล้อง Kinect ยังคงเป็นข้อจำกัดในการวิเคราะห์ท่าทางซึ่งระยะทำการของกล้อง Kinect ที่สามารถทำงานได้อย่างมีประสิทธิภาพนั้นอยู่ในช่วง 1.8–3.0 เมตรเท่านั้น ดังนั้นการนำเทคโนโลยีอื่น ๆ มาประยุกต์เพื่อใช้ในการวิเคราะห์ท่าทางที่อยู่นอกเหนือจากระยะการทำงานของ

กล้อง Kinect จึงเป็นอีกแนวทางที่จะทำให้ระบบที่ใช้ในการวิเคราะห์ท่าทางตามเวลาจริงมีความยืดหยุ่นในการวิเคราะห์ท่าทางมากยิ่งขึ้น

นอกจากนี้ปัจจัยเรื่องจำนวนคนที่สามารถตรวจจับได้จากกล้องเป็นอีกปัจจัยที่ควรได้รับการแก้ไข โดยในปลายปี ค.ศ. 2014 บริษัท Microsoft ได้พัฒนากล้อง Kinect รุ่น Kinect for Windows V2 ซึ่งสามารถตรวจจับบุคคลที่อยู่หน้ากล้องได้ทั้งหมด 6 คน จึงเป็นอีกอุปกรณ์ที่จะช่วยเพิ่มความสามารถของระบบวิเคราะห์ท่าทางให้มีประสิทธิภาพมากยิ่งขึ้น และนอกจากกล้อง Kinect ที่มีความสามารถในการตรวจจับท่าทางของมนุษย์ได้แล้ว ยังมีกล้อง Wavi Xtion ซึ่งเป็นอุปกรณ์ที่ใช้ในการควบคุมเครื่องเล่นเกมของบริษัท ASUS ซึ่งเป็นอุปกรณ์อีกชิ้นหนึ่งที่มีประสิทธิภาพในการตรวจจับท่าทางของมนุษย์ จึงควรศึกษาเพื่อเปรียบเทียบความสามารถในการทำงานของทั้งสองอุปกรณ์ เพื่อหาอุปกรณ์ที่เหมาะสมที่สุดเพื่อใช้ในการตรวจจับท่าทาง

5.4 แนวทางการพัฒนาต่อ

งานวิจัยชิ้นนี้ได้นำข้อมูล 3 มิติประยุกต์ร่วมกับการทำเหมืองข้อมูลเพื่อใช้ในการวิเคราะห์ท่าทางที่ถูกแสดงออกโดยมนุษย์ ซึ่งหลักการในการวิเคราะห์ท่าทางที่งานวิจัยนี้ได้นำเสนอสามารถใช้ในการวิเคราะห์ท่าทางได้เป็นอย่างดี โดยในอนาคตผู้วิจัยได้วางแผนที่นำงานวิจัยชิ้นนี้ไปประยุกต์เพื่อใช้ในการตีความหมายพฤติกรรม และการแสดงออกทางอารมณ์ในระหว่างการรับชมโทรทัศน์ของผู้สูงอายุ เพื่อใช้ในการนำเสนอสื่อประสมที่เหมาะสมให้แก่ผู้สูงอายุแต่ละคน ซึ่งกลุ่มผู้สูงอายุนั้นเป็นกลุ่มที่ควรได้รับการดูแลจากบุคคลภายในครอบครัว แต่ในปัจจุบันผู้สูงอายุได้รับการดูแลน้อยลง และต้องอยู่ตามลำพัง เนื่องจากที่บุคคลภายในครอบครัวมีความจำเป็นต้องออกไปทำงานนอกบ้านมากยิ่งขึ้น โดยจากข้อมูลในปี พ.ศ. 2550 พบว่ามีผู้สูงอายุในประเทศไทยที่อาศัยอยู่คนเดียวตามลำพังในครัวเรือนคิดเป็นร้อยละ 7.7 และมากกว่าครึ่งหนึ่ง (ร้อยละ 56.7) ของผู้สูงอายุที่อยู่คนเดียวตามลำพังสามารถดำรงชีวิตอยู่ได้โดยไม่มีปัญหา ที่เหลือร้อยละ 43.3 มีปัญหา โดยปัญหาที่พบมากที่สุดคือรู้สึกเหงา (ร้อยละ 51.2) [47] ดังที่แสดงไว้ในรูปที่ 5.1



รูปที่ 5.1 ปัญหาจากการอยู่คนเดียวของผู้สูงอายุในไทยโดยจำแนกตามลักษณะการอยู่อาศัยในครัวเรือน

จากปัญหาที่ได้กล่าวมาทำให้ผู้วิจัยสนใจที่จะหาวิธีการในการช่วยลดความเหงา และอาการซึมเศร้าที่เกิดขึ้นกับผู้สูงอายุ เนื่องจากกิจกรรมหลักของผู้สูงอายุคือการรับชมโทรทัศน์ซึ่งมีหลายงานวิจัยพบว่าผู้สูงอายุยอมที่จะเสียเวลาวันละหลาย ๆ ชั่วโมงไปกับการรับชมโทรทัศน์ [48, 49] ทำให้ผู้วิจัยเล็งเห็นว่าหากการรับชมโทรทัศน์เป็นกิจกรรมที่ผู้สูงอายุชื่นชอบการพัฒนาแบบโทรทัศน์ให้มีความชาญฉลาดที่สามารถวิเคราะห์พฤติกรรม และอารมณ์ของผู้สูงอายุในระหว่างการรับชมโทรทัศน์เพื่อใช้ในการนำเสนอสื่อประสมที่เหมาะสมแก่ผู้สูงอายุอาจเป็นอีกหนึ่งแนวทางที่สามารถช่วยให้ผู้สูงอายุรู้สึกมีความสุขและคลายเหงาได้ โดยระบบที่พัฒนาขึ้นจะประกอบด้วยสองส่วนการทำงานได้แก่ส่วนที่ใช้ในการวิเคราะห์พฤติกรรม และอารมณ์ของผู้สูงอายุในระหว่างการรับชมโทรทัศน์ และส่วนที่ใช้ในการนำเสนอสื่อประสมแก่ผู้สูงอายุ

ในส่วนของการวิเคราะห์พฤติกรรมและอารมณ์ของผู้สูงอายุในระหว่างการรับชมโทรทัศน์นั้นผู้วิจัยได้วางแนวทางที่จะนำหลักการในการวิเคราะห์ทำทางที่ได้นำเสนอในงานวิจัยชิ้นนี้ไปประยุกต์กับการวิเคราะห์การแสดงออกทางอารมณ์บนใบหน้า เพื่อใช้ในการวิเคราะห์อารมณ์ในระหว่างการรับชมโทรทัศน์ของผู้สูงอายุ ซึ่งการนำหลักการทั้งสองมาทำงานร่วมกันจะสามารถช่วยให้การวิเคราะห์พฤติกรรม และอารมณ์ของผู้สูงอายุในแต่ละช่วงเวลามีความชัดเจนมากยิ่งขึ้น และในส่วนของการนำเสนอสื่อประสมแก่ผู้สูงอายุนั้นผู้วิจัยจะนำงานวิจัยของ ของ จุฑารัตน์ ชุมเกษียร พรชัยมงคลนาม และ โจนธาน โสอิน ชาน ที่ได้นำเสนอระบบโทรทัศน์ต้นแบบที่ใช้ในการเสนอสื่อประสมสำหรับผู้สูงอายุ [50] มาทำการประยุกต์และปรับปรุงให้มีความเหมาะสมกับระบบที่พัฒนามากยิ่งขึ้น เพื่อใช้ในการนำเสนอสื่อต่าง ๆ แก่ผู้สูงอายุ เช่น รายการโทรทัศน์ต่าง ๆ ที่ผู้สูงอายุชื่นชอบรวมถึงข้อมูลภาพ และเสียงที่ส่งมาจากผู้คนที่ครอบครัวเป็นต้น

ระบบที่จะพัฒนานี้จะใช้กล้อง Kinect ในการวิเคราะห์ท่าทาง และการแสดงออกบนใบหน้า และในส่วนระบบโทรทัศน์ที่ใช้ในการนำเสนอสื่อ่นั้นจะถูกพัฒนาในระบบ Android Box ซึ่งเป็นอุปกรณ์ที่ใช้ในการรับสัญญาณโทรทัศน์ผ่านทางเครือข่ายอินเทอร์เน็ต โดยทั้งสองส่วนจะทำงานร่วมกันผ่านระบบเครือข่าย

โดยผลประโยชน์ที่ผู้วิจัยคาดหวังจากการพัฒนาระบบที่ได้กล่าวมานี้คือ สามารถนำหลักการวิเคราะห์ท่าทางที่ได้เสนอไว้ในงานวิจัยชิ้นนี้ไปประยุกต์ใช้กับงานวิจัยอื่น ๆ ให้เกิดประโยชน์มากที่สุด และพัฒนาระบบที่ช่วยเพิ่มคุณภาพชีวิตให้แก่ผู้สูงอายุให้ดียิ่งขึ้น

เอกสารอ้างอิง

1. Yonemoto, S., Matsumoto, A., Arita, D. and Taniguchi, R. I., 1999, "A real-time motion capture system with multiple camera fusion," **Image Analysis and Processing, International Conference on**, September 27-29, Venice, Italy, pp. 600-605.
2. Chen, H., Qian, G. and James, J., 2005, "An Autonomous Dance Scoring System Using Marker-based Motion Capture," **Multimedia Signal Processing, 2005 IEEE 7th Workshop on**, October 30 - November 2, Shanghai, China, pp. 1-4.
3. Canton-Ferrer, C., Casas, J. R. and Pardas, M., 2009, "Towards a low cost multi-camera marker based human motion capture system," **Image Processing (ICIP), 2009 16th IEEE International Conference on**, November 7-10, Cairo, Egypt, pp. 2581-2584.
4. Wong, C., Zhang, Z., Kwasnicki, R., Liu, J. and Yang G. Z., 2012, "Motion Reconstruction From Sparse Accelerometer Data Using PLSR," **Wearable and Implantable Body Sensor Networks (BSN), 2012 Ninth International Conference on**, May 9-12, London, England, pp. 178-183.
5. **Carnegie Mellon Graphics Lab: Motion Capture and File Formats** [Online], Available: <http://mocap.cs.cmu.edu/info.php> [2014, October 10].
6. Xian-Jie, Q., Wang, Z. Q., Shi-Hong, X. and Jin-Tao, L., 2005, "Estimating Articulated Human Pose from Video Using Shape Context," **Signal Processing and Information Technology, 2005. Proceedings of the Fifth IEEE International Symposium on**, December 21, Athens, Greece, pp. 583-588.
7. Lee, C. S. and Elgammal, A., 2006, "Gait tracking and recognition using person-dependent dynamic shape model," **Automatic Face and Gesture Recognition FGR 2006. 7th International Conference on**, April 2-6, Southampton, England, pp. 553-559.

8. Chuang, C. H., Jun-Wei, H., Tsai, L. W. and Fan, K. C., 2008, "Human Action Recognition Using Star Templates and Delaunay Triangulation," **Intelligent Information Hiding and Multimedia Signal Processing, 2008. IHHMSP '08 International Conference on**, August 15-17, Harbin, China, pp. 179-182.
9. Heickal, H., Zhang, T., and Hasanuzzaman, M., 2013, "Real-time 3D full body motion gesture recognition," **Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on**, December 12-14, Shenzhen, China, pp. 798-803.
10. Park, C. B., Roh, M. C., and Lee, S. W., 2008, "Real-Time 3D Pointing Gesture Recognition in Mobile Space," **Automatic Face & Gesture Recognition, 2008th IEEE International Conference on**, September 17-19, Amsterdam, Netherlands, pp. 1-6.
11. Hu, G. and Gao, Q., 2012, "A 3D gesture recognition framework based on hierarchical visual attention and perceptual organization models," **Pattern Recognition (ICPR), 2012 21st International Conference on**, November 11-15, Tsukuba, Japan, pp. 1411-1414.
12. Webb, J. and Ashley, J., 2012, **Beginning Kinect programming with the Microsoft Kinect SDK**, Apress©, New York, pp. 1-321.
13. Jana, A., 2012, **Kinect for Windows SDK Programming Guide**, Packt Publishing Ltd., Birmingham, pp. 1-330.
14. Microsoft Research, **NUI: Natural User Interface** [Online], Available: <http://research.microsoft.com/en-us/collaboration/focus/nui/> [2014, October 10].
15. Microsoft Developer Network, **Kinect for Windows Sensor Components and Specifications** [Online], Available: <http://msdn.microsoft.com/en-us/library/jj131033.aspx> [2014, October 10].
16. Microsoft Developer Network, **Kinect for Windows Architecture** [Online], Available: <http://msdn.microsoft.com/en-us/library/jj131023.aspx> [2014, October 10].

17. Microsoft Developer Network, **Tracking Modes (Seated and Default)** [Online], Available: <http://msdn.microsoft.com/en-us/library/hh973077.aspx> [2014, October 10].

18. Microsoft Developer Network, **Tracking Users with Kinect Skeletal Tracking** [Online], Available: <http://msdn.microsoft.com/en-us/library/jj131025.aspx> [2014, October 10].

19. Microsoft Developer Network, **Kinect Sensor** [Online], Available: <http://msdn.microsoft.com/en-us/library/hh438998.aspx> [2014, October 10].

20. Han, J. and Kamber, M., 2006, **Data mining concepts and techniques**, 2nd ed., Kaufmann Publishers, San Francisco, pp. 285-378.

21. Hastie, T., Tibshirani, R. and Friedman, J., 2008, **The Elements of Statistical Learning Data Mining, Inference, and Prediction**, 2nd ed., Springer, pp. 1-42, 119-603.

22. The University of Waikato, **Weka 3-Data Mining with Open Source** [Online], Available: <http://www.cs.waikato.ac.nz/ml/weka/> [2014, October 10].

23. The University of Waikato, **Weka Manual for Version 3-7-10**, [E-Book], pp. 9-211.

24. Roh, M. C. and Lee, S. W., 2006, "Gesture Detection in Low-Quality Video," **Pattern Recognition, 2006 18th International Conference on**, August 20-24, Hong Kong, China, pp. 791-794.

25. Lianga, Y. M., Shih, S. W., Shih, A. C. C. and Liaoa, H. Y. M., 2006, "Understanding Human Behavior Using a Language Modeling Approach," **Intelligent Information Hiding and Multimedia Signal Processing, 2006. IHH-MSP '06. International Conference on**, Pasadena, California, USA, pp. 331-334.

26. Chen, C. C., Hsieh, J. W., Hsu, Y. T. and Huang, C. Y., 2006, "Segmentation of Human Body Parts Using Deformable Triangulation," **Pattern Recognition, 2006. ICPR 18th International Conference on**, Hong Kong, Cahina, pp. 355-358 .

27. Wu, C. and Aghajan, H., 2007, "Model-based Human Posture Estimation for Gesture Analysis in an Opportunistic Fusion Smart Camera Network," **Advanced Video and Signal Based Surveillance AVSS 2007, IEEE Conference on**, September 5-7, London, England, pp. 453-458.
28. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A. and Blake, A., 2011, "Real-Time Human Pose Recognition in Parts from Single Depth Images," **Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on**, June 20-25, Providence, Rhode Island, USA, pp. 1297-1304.
29. Xia, L., Chen, C. and Aggarwal, J. K., 2011, "Human detection using depth information by Kinect," **Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on**, June 20-25, Colorado Springs, USA, pp. 15-22.
30. Liu, Y., Stoll, C., Gall, J., Seidel, H. and Theobalt, C., 2011, "Markerless Motion Capture of Interacting Characters Using Multi-view Image Segmentation," **Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on**, June 20-25, Providence, Rhode Island, USA, pp. 1249-1256.
31. Yu, X., Wu, L., Liu, Q. and Zhou, H., 2011, "Children Tantrum Behaviour Analysis Based on Kinect Sensor," **Intelligent Visual Surveillance (IVS), 2011 Third Chinese Conference on**, December 1-2, Beijing, China, pp. 49-52.
32. Hauswiesner, S., Straka, M. and Reitmayr, G., 2011, "Free Viewpoint Virtual Try-On With Commodity Depth Cameras," **Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry**, December 11-12, Hong Kong, China, pp. 23-30.
33. Igorevich, R. R., Ismoilovich, E. P. and Min, D., 2011, "Behavioral Synchronization of Human and Humanoid Robot," **Ubiquitous Robots and Ambient Intelligence (URAI), 2011 8th International Conference on**, November 23-26, Incheon, Korea, pp. 655-660.

34. Alexiadis, D., Kelly, P., Daras, P., O'Connor, N. E., Boubekeur, T. and Moussa, M. B., 2011, "Evaluating a Dancer's Performance using Kinect-based Skeleton Tracking," **Proceedings of the 19th ACM international conference on Multimedia**, November 28-December 1, Arizona, USA, pp. 659-662.
35. Marquardt, Z., Beira, J., Em, N., Paiva, I. and Kox, S., 2012, "super mirror: a kinect interface for ballet dancers," **Proceeding of CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)**, New York, USA, pp. 1619-1624.
36. นราวุฒิ พัฒโนทัย, พรชัย มงคลนาม และ บัณฑิต วรรณภา, 2555, "การประกอบรวมโครงร่าง มนุษย์จากการตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect หลายตัว," **การประชุมวิชาการระดับประเทศทางด้านเทคโนโลยีสารสนเทศ (NCIT), ครั้งที่ 4**, 26-27 เมษายน 2555, อ. ชะอำ, จ. เพชรบุรี, หน้า 369-374.
37. Patsadu, O., Nukoolkit, C. and Watanapa, B., 2012, "Human gesture recognition using Kinect camera," **Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference on**, May 30–June 1, Bangkok, Thailand, pp. 28-32.
38. Kepski, M. and Kwolek, B., 2012, "Fall detection on embedded platform using kinect and wireless accelerometer," **Proceedings of the 13th international conference on Computers Helping People with Special Needs - Volume Part II (ICHP'12)**, Part II. Springer-Verlag, Berlin, Heidelberg, pp. 407-414.
39. Ball, A., Rye, D., Ramos, F. and Velonaki, M., 2012, "Unsupervised clustering of people from 'skeleton' data," **Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on**, March 5-8, Massachusetts, USA, pp. 225-226.
40. Liu, Y., Zhang, Z., Li, A. and Wang, M., 2013, "View independent human posture identification using Kinect," **Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on**, October 16-18, Chongqing, China, pp. 1590-1593.

41. สุทธิพงษ์ แก่นจันทร์, พรชัย มงคลนาม, บัณฑิต วรรณภา และ ศศิภา เสถียรพงษ์, 2556, "การติดตั้งกล้อง Kinect หลายตัวเพื่อประกอบโครงร่างมนุษย์ โดยปราศจากการใช้เครื่องมือวัดตำแหน่งและทิศทางการวางกล้อง," **การประชุมวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ (NCCIT), ครั้งที่ 10, พฤษภาคม 9-10 2556, คณะเทคโนโลยีสารสนเทศ, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, หน้า 369-374.**
42. Uribe-Quevedo, A., Perez-Gutierrez, B. and Guerrero-Rincon, C., 2013, "Seated Tracking for Correcting Computer Work Postures," **Biomedical Engineering Conference (SBEC), 2013 29th Southern**, Miami, Florida, USA, pp. 169-170.
43. Kumada, K., Usui, Y. and Kondo, K., 2013, "Golf Swing Tracking and Evaluation Using Kinect Sensor and Particle Filter," **Intelligent Signal Processing and Communications Systems (ISPACS), 2013 International Symposium on**, November 12-15, Naha, Japan, pp. 698-703.
44. Dai, X., Wu, M., Davidson, B., Mahoor, M. and Zhang, J., 2013, "Image-based Fall Detection with Human Posture Sequence Modeling," **Healthcare Informatics (ICHI), 2013 IEEE International Conference on**, September 9-11, Philadelphia, Pennsylvania, USA, pp. 376-381.
45. Qiu, J., Jieyu, Z. and Yuanyuan, Z., 2012, "Facial feature extraction with a depth AAM algorithm," **Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on**, May 29-31, Sichuan, China, pp. 1792-1796.
46. Seddik, B., Maâmatou, H., Gazzah, S. and Chateau, T., 2013, "Unsupervised facial expressions recognition and avatar reconstruction from kinect," **Systems, Signals & Devices (SSD), 2013 10th International Multi-Conference on**, March 18-21, Hammamet, Tunisia, pp. 1-6.
47. Thailand's National Statistical Office, **Information Service (2014)** [Online], Available: http://service.nso.go.th/nso/nsopublish/citizen/news/news_older.jsp [2014, October 10].

48. Center for Media Literacy, **Lifeline or Leisure?: TV's Role in the Lives of the Elderly** [Online], Available: <http://www.medialit.org/reading-room/lifeline-or-leisure-tvs-role-lives-elderly> [2014, October 10].

49. Oestlund, B., Jonsson, B. and Waller, P., 2010, "Watching Television in Later Life: A deeper," **Scandinavian Journal of Caring Sciences**, 24(2):233-43, Epub 20 December 2009.

50. Choomkasean, J., Mongkolnam, P. and Chan, J. H., 2009, "Multimedia Delivery for Elderly People: A Conceptual Model," **The 5th International Conference on Advances in Information Technology (IAIT 2012)**, December 6-7, Bangkok, Thailand, pp. 58-69.

ภาคผนวก ก
หนังสืออนุมัติจริยธรรมการวิจัย

แบบเสนอขออนุมัติจริยธรรมการวิจัยจากคณะกรรมการจริยธรรมการวิจัย

ชื่อ: นาย ธรรมศาสตร์ วิชาธรรมณ์
 ศศ.ดร. พรชัย มงคลนาม
 ศศ.ดร. โจนนาธาน โอฮิน ชาน
 หัวข้อวิจัย: การจำแนกท่าทางขณะรับชมโทรทัศน์โดยใช้กล้อง Kinect
 วันที่: 7 กรกฎาคม 2557

สรุปย่อโครงการวิจัย:

ในงานวิจัยนี้มีจุดประสงค์เพื่อประยุกต์เทคโนโลยีสารสนเทศในการช่วยติดตามพฤติกรรมของผู้สูงอายุในระหว่างการรับชมโทรทัศน์โดยใช้กล้อง Kinect ในการวิเคราะห์การเคลื่อนไหวของผู้สูงอายุ กล้อง Kinect จะทำหน้าที่ในการค้นหาท่าทางกายมนุษย์ และ ทำการวิเคราะห์จุดต่างๆบนร่างกาย ซึ่งจุดบนร่างกายที่ผ่านการวิเคราะห์จากกล้อง Kinect นั้นจะอยู่ในรูปแบบข้อมูลพิกัดสามมิติ ซึ่งประกอบด้วย ข้อมูลแกน X, Y และ Z โดยในงานวิจัยชิ้นนี้จะนำเอาข้อมูลที่สามารถบันทึกได้จากกล้อง Kinect ไปทำการวิเคราะห์ท่าทางของผู้สูงอายุในระหว่างการรับชมโทรทัศน์ ซึ่งในการทดลองของงานวิจัยชิ้นนี้จะเลือกใช้กลุ่มตัวอย่างเป็นกลุ่มผู้สูงอายุจำนวน 10 คน ซึ่งเป็นบุคคลที่มีอายุตั้งแต่ 60 ปีบริบูรณ์ โดยประกอบด้วยเพศชาย 5 คน และ เพศหญิง 5 คน ซึ่งมีความสูง และ รูปร่างแตกต่างกันออกไปเพื่อให้เกิดความหลากหลายทางข้อมูล ในการทดลองจะทำการบันทึกข้อมูลพฤติกรรมการแสดงออกของท่าทางจริงของผู้สูงอายุในระหว่างการรับชมโทรทัศน์ โดยท่าทางที่งานวิจัยให้ความสนใจเป็นท่าทางพื้นฐานในการรับชมโทรทัศน์ เช่น การนั่ง การยืน การนอน การกอดอก และ การยกมือเป็นต้น โดยผู้เข้าร่วมงานวิจัยทุกคนจะต้องทำการรับชมโทรทัศน์เป็นเวลา 5 นาที เป็นจำนวน 5 ครั้ง ในแต่ละครั้งสามารถแสดงท่าทางได้อย่างอิสระ และ ไม่มีการลำดับท่าทางในระหว่างการรับชมโทรทัศน์ โดยผู้เข้าร่วมทำการทดลองจะต้องทำท่าทางต่าง ๆ ตามที่งานวิจัยได้กำหนดไว้ให้ครบทุกท่าทาง ซึ่งข้อมูลที่ได้จากการบันทึกท่าทางของผู้เข้าร่วมงานวิจัยทั้ง 10 คนจะถูกนำมาประยุกต์ใช้กับหลักการการทำเหมืองข้อมูล (Data mining) โดยในงานวิจัยนี้เลือกแบบจำลอง (model) เพื่อทดลองทำการเปรียบเทียบผลลัพธ์ของแบบจำลองที่เหมาะสมกับงานวิจัย ซึ่งประกอบด้วย Neural Networks, Support Vector Machine, Naïve Bayes, Decision-tree, Logistic regression และ Random forest โดยแบบจำลองที่ให้ผลลัพธ์ที่ดีที่สุดจะถูกเลือกมาใช้ในการพัฒนาระบบและแยกท่าทางในระหว่างรับชมโทรทัศน์ โดยระบบที่ถูกพัฒนาเสร็จสมบูรณ์จะถูกนำไปติดตั้งตามบ้านผู้สูงอายุเพื่อใช้ในการช่วยติดตามพฤติกรรมของผู้สูงอายุในระหว่างรับชมโทรทัศน์

ประเด็นเชิงจริยธรรมและการดำเนินการ

- การยินยอมเข้าร่วมในการวิจัย
 - กลุ่มตัวอย่างทั้งหมด 10 คน จะได้รับหนังสือขอความยินยอมเข้าร่วมในงานวิจัย และ ได้รับการบอกกล่าวรวมถึงอธิบายว่าการเข้าร่วมในงานวิจัยนี้เป็นไปโดยความสมัครใจ และ หากผู้เข้าร่วมงานวิจัย ไม่ลงนามในหนังสือยินยอมเข้าร่วมในงานวิจัย ข้อมูลที่เก็บรวบรวมได้จะไม่นำมาเป็นส่วนหนึ่งของการวิเคราะห์ข้อมูล
- สิทธิส่วนบุคคล/ การรักษาข้อมูลเป็นความลับ
 - ชื่อนามสกุลจริงของผู้เข้าร่วมงานวิจัยทุกคนจะถูกใช้แทนด้วยนามสมมติ ซึ่งจะไม่สามารถระบุถึงตัวตนของบุคคลที่เข้าร่วมการทดลองได้
- การปกป้องผู้เข้าร่วมงานวิจัยจากความเสี่ยงต่อการได้รับผลกระทบเชิงลบจากการทำวิจัย
 - กรณีผู้เข้าร่วมงานวิจัยได้รับอุบัติเหตุจากการทดลองผู้เข้าร่วมงานวิจัยจะได้รับค่าใช้จ่ายในการรักษาพยาบาลจากผู้วิจัย



หนังสืออนุมัติจริยธรรมการวิจัย



ชื่อโครงการวิจัย การจำแนกท่าทางขณะรับชมโทรทัศน์โดยใช้กล้องคินค
 ชื่อผู้วิจัย นาย ธรรมศาสตร์ วิสุทธารมณ
 ที่อยู่ติดต่อได้ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี เบอร์โทรศัพท์ 087-470-8106
 อีเมลล์ Thammasat@gmail.com

โครงการวิจัยดังกล่าวนี้ได้รับการพิจารณาตรวจสอบ และได้รับการอนุมัติจริยธรรมการวิจัยที่ถือปฏิบัติตามกฎเกณฑ์และ
 มาตรฐานของคณะกรรมการจริยธรรมการวิจัยคณะศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรีแล้ว

คณะกรรมการจริยธรรมการวิจัย

ลงนาม	<u></u>
ชื่อ	รองศาสตราจารย์ ดร.วิฑริศ วัฒน ทอดค์
วันที่	<u>24 ก.ค. 2557</u>
ลงนาม	<u></u>
ชื่อ	ผู้ช่วยศาสตราจารย์ ดร.ปัทมวรรณ ชิลลิ
วันที่	<u>24 ก.ค. 2557</u>
ลงนาม	<u></u>
ชื่อ	ดร.ภาสนันท์ อัครักษ์
วันที่	<u>24 ก.ค. 2557</u>

หนังสือยินยอมเข้าร่วมในการวิจัย

ชื่อโครงการวิจัย การจำแนกท่าทางขณะรับชมโทรทัศน์โดยใช้กล้องคินเนค

ชื่อผู้วิจัย นาย ธรรมศาสตร์ วิชาธรรมณ์

ที่อยู่ติดต่อได้ คณะเทคโนโลยีสารสนเทศมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรีเบอร์โทรศัพท์ 087-470-8106

อีเมล Thanmarsat@gmail.com

ในงานวิจัยนี้มุ่งเน้นการ ประยุกต์ระบบเทคโนโลยีสารสนเทศในการติดตามพฤติกรรมของผู้สูงอายุในระหว่างการรับชมโทรทัศน์ ซึ่งกลุ่มเป้าหมายคือบุคคลที่มีอายุตั้งแต่ 60 ปีบริบูรณ์ ซึ่งในงานวิจัยนี้ผู้วิจัยจะทำการจำลองสถานการณ์การรับชมโทรทัศน์ โดยต้องการให้ผู้เข้าร่วมงานวิจัยทำการแสดงกิจกรรมต่าง ๆ ในระหว่างการรับชมโทรทัศน์ จำนวน 5 ครั้ง ในแต่ละครั้งใช้เวลาประมาณ 5 นาที โดยใช้กล้อง Kinect ในการบันทึกพฤติกรรมของผู้เข้าร่วมงานวิจัย โดยผู้เข้าร่วมงานวิจัยสามารถแสดงท่าทางต่างๆตามที่ได้กำหนดไว้อย่างอิสระ ซึ่งผู้เข้าร่วมงานวิจัยจะต้องแสดงทุกท่าทางให้ครบตามที่งานวิจัยได้กำหนดไว้ โดยท่าทางที่งานวิจัยให้ความสนใจจะเป็นเพียงท่าทางพื้นฐานซึ่งจะไม่ส่งผลเสียต่อร่างกายของผู้เข้าร่วมงานวิจัย อย่างไรก็ตาม ในกรณีที่ผู้เข้าร่วมงานวิจัยได้รับการบาดเจ็บจากการร่วมทำการทดลอง ทางผู้วิจัยจะเป็นผู้ออกค่าใช้จ่ายในการรักษาพยาบาลให้แก่ผู้เข้าร่วมงานวิจัย

ในกรณีที่ผู้เข้าร่วมงานวิจัยมีความต้องการถอนตัวจากการทดลอง ผู้เข้าร่วมงานวิจัยสามารถถอนตัวจากการทดลองได้ตลอดเวลา และ ข้อมูลที่ได้ทำการบันทึกไว้แล้วของผู้เข้าร่วมงานวิจัยที่ได้ถอนตัวจากการทดลอง ผู้วิจัยจะไม่นำข้อมูลเหล่านั้นมาใช้ต่อในงานวิจัย

ทั้งนี้ข้อมูลส่วนบุคคลอันได้แก่ใบหน้า ชื่อ และ นามสกุล ของผู้เข้าร่วมงานวิจัยทุกคนจะไม่ถูกนำเสนอสู่สาธารณะเพื่อเป็นการคงไว้ซึ่งสิทธิเสรีภาพส่วนบุคคล

ข้าพเจ้าซึ่ง ได้ลงนามที่ด้านล่างของหนังสือฉบับนี้ ได้รับคำอธิบายอย่างชัดเจนจนเป็นที่พอใจจากผู้วิจัยถึงวัตถุประสงค์และขั้นตอนการวิจัย และประโยชน์ซึ่งจะเกิดขึ้นจากการวิจัยเรื่องนี้แล้ว

ข้าพเจ้าเข้าร่วมการวิจัยครั้งนี้ด้วยความสมัครใจ และข้าพเจ้ามีสิทธิ จะถอนตัวออกจากการวิจัยเมื่อไรก็ได้ตามความประสงค์ ข้าพเจ้ายินดีเข้าร่วมการวิจัยครั้งนี้ ภายใต้งบเงื่อนไขที่ระบุไว้ในเอกสารข้อมูลสำหรับกลุ่มประชากรหรือผู้มีส่วนร่วมในการวิจัย

ลงนามผู้มีส่วนร่วมในการวิจัย

ตัวบรรจง

วันที่

การรับรองจากผู้วิจัย

ข้าพเจ้ารับรองว่า ข้อมูลที่ได้รับจะถูกใช้เพื่อวัตถุประสงค์ในการวิจัยตามที่ได้ระบุไว้ในที่นั้นเว้นเสียจากจะได้รับ ความยินยอมให้ใช้ได้ในวัตถุประสงค์อื่น และข้าพเจ้าจะเก็บรักษาข้อมูลนี้เป็นความลับและไม่เปิดเผยข้อมูลเฉพาะตัวของผู้มีส่วนร่วมในการวิจัยเว้นเสียจากจะได้รับ ความยินยอมให้ดำเนินการได้

กรุณาลงนามหน้าถัดไป

ลงนามผู้วิจัย

ตัวบรรจง

วันที่

ภาคผนวก ข
ผลงานที่ได้รับการตีพิมพ์

Postural Classification using Kinect

Thammarsat Visutarrom, Pornchai Mongkolnam, and Jonathan Hoyin Chan

The 2014 International Computer Science and Engineering Conference (ICSEC2014)

July 30 - August 1, 2014

Postural Classification using Kinect

Thammarsat Visutarrom, Pornchai Mongkolnam, and Jonathan Hoyin Chan

School of Information Technology
King Mongkut's University of Technology Thonburi
Bangkok 10140, Thailand
{pornchai | jonathan}@sit.kmutt.ac.th

Abstract—This research focuses on the comparison of posture recognition, using a data mining classification approach on the skeleton data stream obtained from Kinect camera. We classified four standard postures including *Stand*, *Sit*, *Sit on floor* and *Lie Down*. We compared six classifiers, namely, decision tree, neural network, naïve Bayes, support vector machine, logistic regression and random forest in order to find a suitable classifier. Our best results can correctly classify the postures with 97.88% accuracy, 97.40% sensitivity, and 0.991 ROC area under curve using Max-Min normalization with a decision tree classifier on four transformed attributes. Our future work will use the knowledge obtained to classify a wider range of postures of the elderly while watching television, to be a part of a bigger effort to monitor and study elderly behavior at home.

Keywords— data mining; elderly; Kinect camera; postural classification

I. INTRODUCTION

There are two definitions of ageing society as defined by the United Nations: when people in the society aged 65 years old or over comprise more than 7%, or 60 years old or over comprise more than 10% of a population. From the second definition, the world became an ageing society since 2000 when the number of people aged 60 years old or over made up about 10% of the world's population [1-2] this number is expected to rise to 21.1% by year 2050, as can be seen in Table I. Thailand, in particular, became an ageing society since 2007 when it had about 11.0% ageing population [3]. Although the ageing population increases globally, elderly people are being taken care less and less. Our main reason is that more and more family members work outside of the home. In addition, physical decline and less financial support are the other factors that largely keep the elderly people at home. They tend to be alone in daytime at best or alone the whole day at worst. One technology that could be used to help to improve this quality of life and to help family members to understand them better is a software system that could track daily activities engage them in various social activities. Specifically, tracking postures of the elderly people while watching television is of our interest and is the focus of our work. We would like to learn what most elderly people spend a substantial time at home with doing while watching television.

Like many researchers who pay attention to behavioral tracking of elderly people. In this paper, we compare the efficacy of classification accuracy in postural classification, using Kinect camera.

TABLE I. PERCENTAGE OF POPULATION AGED 60 YEARS OR OVER [2].

Region/Country	Year 1950	Year 2000	Year 2050
World	8.2%	10.0%	21.1%
Asia	6.8%	8.8%	22.6%
Europe	12.1%	20.3%	36.6%
China	7.5%	10.1%	29.9%
India	5.6%	7.6%	20.6%
Japan	7.7%	23.2%	42.3%
Thailand	5.0%	8.1%	27.1%
USA	12.5%	16.1%	26.9%

The knowledge and methodology obtained from this experiment is used to classify the basic postures of elderly while watching television, which is a part of tracking elderly behavior.

II. LITERATURE REVIEW

A. Relationship between Elderly People and Televisions

Health and financial problems of elderly people tend to keep them staying more at home and meeting fewer people. When they meet fewer people and spend less time socially interacting, they start to lose social contacts, become more isolated and unaware of current affairs. In addition, they may spend more time alone because the household family members are too busy working outside. As a result, watching television has become a preferred activity for elderly people because it can help them see the world from their homes. Roles of televisions were studied by Reid [4] in "Lifeline or Leisure?: TV's Role in the Lives of the Elderly". He compared television to a window that made them see many things happening around the world, and preserved the sense of participation in society and resisted the feelings of isolation and loneliness [5]. Several researchers have found elderly people love to spend many hours per day in front of televisions [6-8]. In addition, many researches focus on developing a television system appropriate for elderly people. For instance, in 2007 Nakajima et al. [9] developed a simple and inexpensive system or telemonitoring system of television's operating state for remotely located families to improve smooth communications between elderly people, who live alone, and their families. In 2012, Choomkasean et al. [10] proposed a conceptual model of a multimedia delivery to televisions of the elderly people. The system could receive multimedia to the elderly people via televisions, which is used

to remind them of upcoming events or to allow them to conveniently communicate with their family members and friends. Consequently, it helped sustain relationships between the elderly people and their family members and friends, which could promote mental health in the long run and improve the quality of the elderly.

B. Gestural and Postural Detection

Two dimension image processing is the most popular technique used to detect human gestures or postures using a simple Web camera or a video camera. However, the technique still has the problem of separating a desired object from a complex background. On the other hand, 3D imaging device can provide a better way to extract a human body out of a complex background. However, it has been in limited use for many years mainly due to its high cost and low accuracy in detecting and extracting desired objects.

In 2010, Kinect cameras and Kinect library functions were released by Microsoft [11-13]. It is a game controller that is controlled by gestures and spoken commands of the user. Kinect library functions can separate a human body from a complex background in any kind of lighting conditions and analyze each presumed point of the human body joints in order to build a visual body structure (skeleton). Each joint consists of a 3D coordinate of X, Y, and Z position. Kinect camera has two competing library functions: OpenNI and Microsoft SDK, which is chosen for this research due to its effectiveness. Microsoft SDK library has two detection modes. One is a default Stand tracking mode; this mode can detect all body joints (20 skeletal joints). Another is a Seated tracking mode; this mode can detect the upper shoulder joints (10 skeletal joints) as shown in Fig. 1. The Kinect system includes an RGB camera, a depth sensor and a multi-array microphone as shown in Fig. 2. The RGB camera is placed between the depth sensors. The camera is used to record 2D image. The depth sensor is used to record depth image for analysis of the skeleton. In addition, the depth sensor system in Kinect camera is of infrared type. Therefore the Kinect camera still works with 3D mode in dark areas. The multi-array microphone is placed horizontally on the Kinect camera body which is used to record sound commands from users.

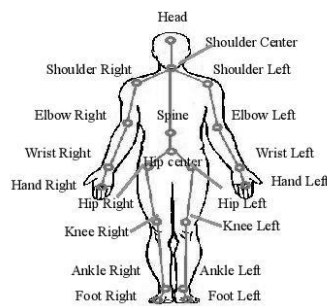


Fig. 1. Twenty Skeletal Joints.

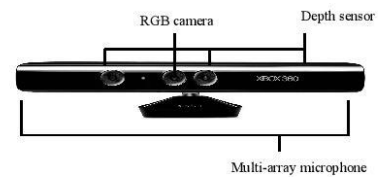


Fig. 2. Kinect components [13].

The ability of Kinect in extracting a 3D human skeleton and its competitive price when compared with other 3D imaging devices are the main reasons why many researches use it to capture and analyze human gestures and postures.

In 2012 Patsadu et al. [14] presented human gesture recognition with Kinect system using a data mining classification technique. Their data set was extracted from input of twenty body-joint positions. The average accuracy of all classification methods was 93.72%, and that can confirm capability of the Kinect camera for use for human gesture recognition. Also in 2012, Liu et al. [15] introduced a view independent posture recognition framework by presenting a body orientation classification system. The posture was used in classification of 5 standing postures. Their results showed that the framework was robust to variant viewpoint with high accuracy.

In 2013, Kaenchan et al. [16] proposed automatic data gathering with multiple Kinect cameras from several viewpoints to create complete skeleton to solve the problem that some parts of the human body may be obscured by objects.

In 2014, Paliyawan et al. [17] presented a system for monitoring office workers in order to prevent Office Workers Syndrome using Kinect camera. The system can alert a user when it is time to relax and provide a daily summary report used to track working behavior of the user.

C. Behavioral Tracking

Behavioral tracking of elderly people can help family members, physicians, and other health practitioners understand the elderly people better. Knowing more about the behaviors and activities of the elderly has been one of the most sought-after studies in imminent ageing societies. Technology could be used to monitor and keep records of the daily lives of the elderly, especially while being at home, to help us better understand and innovate technologies to serve them.

In 2010, Cardile et al. [18] introduced a computer vision-based architecture for remote elderly monitoring, which was built upon a network of wireless camera sensors for tracking activities of the elderly people. The work was motivated by the necessity of unobtrusive remote caring systems for elderly people.

In 2012, Fahim et al. [19] developed an application to track daily life activities tracked by smart phones, which could send messages to remind elderly people to complete some remaining work, take drugs, schedule some activities, etc. In addition, the system can track activities of the elderly people

in each day to provide useful data to family members and care givers. Also in 2012, Correa et al. [20] proposed a navigation system for elderly care applications based on wireless sensor networks, which monitored emergency situations of the elderly people. The system added some functionalities in order to obtain a full health care system for increasing standard of lives.

In 2013, Kim et al. [21] proposed a real-time emergency alarm system that could monitor motions of elderly people. The system could send out alert messages and images quickly to the designated people when an emergency situation occurred to the elderly people who lived alone.

III. PROCESS OVERVIEW

Our experiment was separated into two parts. Part one is data preprocessing that is done prior to postural classification. Part two is classification that would compare results from each candidate classifier models. An outline of the experimental procedure is shown in Fig. 3. We used Kinect XBOX 360 and Microsoft SDK version 1.8 working in default standing mode that can give raw data of 20 joint positions. In the experiment, we used only 10 joints including head, shoulder left, shoulder right, hip left, hip right, hip center, knee left, knee right, foot left and foot right with data transformations instead of using all available 20 joints.

The data used in our training set are collected from ten people (5 males and 5 females) with heights between 150-180 centimeters and with various body sizes. Each subject was asked to perform different postures (*Stand*, *Sit*, *Sit on Floor*, and *Lie Down*) for 4-6 minutes per round. The subjects were asked to be natural and try not to be stationary only for the duration to emulate the real television watching process. Five rounds were performed for each subject in front of the Kinect cameras, which were about 1.8-3.0 meters away. The Kinect cameras captured about 25-31 frames per second. Fig. 4 shows four different postures of a sample subject. The training data set consists of 30,088 instances (each instance comprises the four attributes mentioned in Section IV), with 7,697 instances for *Stand*, 8,848 instances for *Sit*, 5,819 instances for *Sit on Floor*, and 7,724 instances for *Lie Down*.

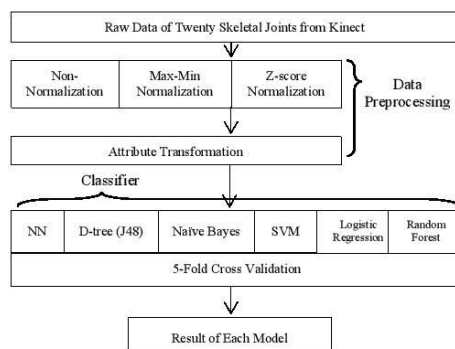


Fig. 3. Process Overview.

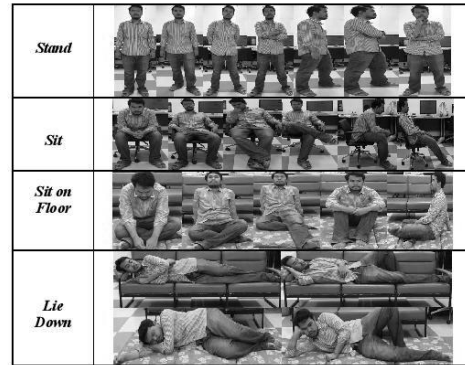


Fig. 4. Different Posture Examples.

IV. DATA PREPROCESSING

In the step of data preprocessing, we compared two techniques of normalization to one without normalization of the raw data. The two normalization techniques used were Max-Min and Z-score. The corresponding normalization equations are shown as (1) and (2), respectively.

$$J' = \frac{J - \text{Min}}{\text{Max} - \text{Min}}, \quad (1)$$

$$J' = \frac{J - \bar{J}}{\sigma}, \quad (2)$$

where J' is the normalized data; Min and Max of (1) are the minimum and maximum values in each skeleton joint position training set; $\bar{J}$ and σ in (2) are the mean and the standard deviation of each skeleton joint position, respectively.

The data used in classification have four transformed attributes, including angle of left knee, angle of right knee, distance from hip to a room floor, and the aspect ratio. These attributes are provided as follows.

A. Aspect ratio of height and width of the depth image

Each posture gives a different aspect ratio of the width and the height as shown in Fig. 5. Therefore, this ratio was selected to be one of the attributes for classification.

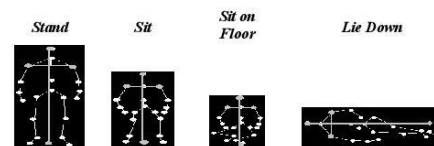


Fig. 5. Aspect Ratio of Each Posture.

The aspect ratio values as in (3)-(5) are derived from the ratio of distance between maximum and minimum values of x and y positions of only 5 joints including head, shoulder left, shoulder right, foot left and foot right.

$$\text{Width} = \text{Max } x - \text{Min } x, \quad (3)$$

$$\text{Height} = \text{Max } y - \text{Min } y, \quad (4)$$

$$\text{Aspect Ratio} = \frac{\text{Width}}{\text{Height}}. \quad (5)$$

However, using the aspect ratio only cannot give a high accuracy. In this research, we added three attributes including angle of left knee, angle of right knee, and distance from hip to a room floor. Angles of left knee and right knee can help to differentiate between *sit* and *stand*. When people stand, the angles of their knees are larger. On the other hand, when they sit, the angles of their knees are smaller as shown in Fig. 6. In addition, we used distance from hip to room floor for helping to classify *sit* and *sit on floor*. When a subject changes posture to *sit on floor*, the distance of hip to room floor will decrease as shown in Fig. 7.

B. Angles of left knee and right knee

Angles, θ , of the left knee and the right knee are used to analyze *Sit* or *Stand* postures. Both angles of the left knee and right knee are computed using (6)-(8), where X, Y, and Z are the coordinates of the knee joint and hip joint positions.

$$A = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2}, \quad (6)$$

$$B = |Y_1 - Y_2|, \quad (7)$$

$$\theta = \sin^{-1} \left(\frac{A}{B} \right). \quad (8)$$

C. Distance from hip to room floor

Distance from hip to a room floor is calculated using (9), where A, B, C, and D are coefficients of the floor plane, and X, Y, and Z are the coordinates of the hip center.

$$\text{Distance} = \frac{AX + BY + CZ + D}{\sqrt{A^2 + B^2 + C^2}}. \quad (9)$$



Fig. 6. Angles of left knee and right knee.



Fig. 7. Distance from hip to a room floor.

V. CLASSIFICATION DATA

Postural analysis used attribute data from the data recorder. We selected 6 models including neural network (NN), decision tree (D-tree), naïve Bayes, support vector machine (SVM), random forest, and logistic regression to compare the quality of data analysis using 5-fold cross-validation. The Weka data mining tool was used to implement these six models.

A. Classification Models

1) Neural Network

In this research we used simple multilayer perceptron (MLP) which uses back-propagation algorithm in learning [22-23] for predicting class member of postures (*Stand*, *Sit*, *Sit on Floor*, and *Lie Down*). Structural of neural network in our experiment have three layers (input layer, hidden layer, and output layer) with 4, 10, and 4 nodes, respectively.

2) Decision Tree

D-tree is used to classify data from class label, which yields output as a flowchart-like tree structure [14, 17, 22-23]. The J48 is used in our work to classify the data as a set of decision nodes and leaf nodes. Each leaf node shows a class outcome label of postures.

3) Naïve Bayes

Naïve Bayes is a statistical classifier based on the Bayes's theorem, which predicts class membership based on conditional probability [22-23]. The nodes in a Bayesian model are created from the given training data. Building model times of naïve Bayes are extremely fast when compared to other methods.

4) Random Forest

Random forest is a classifier consisting of a collection of tree structured classifiers, which uses the tree bagging algorithm in learning model for predicting class member [24]. In our experiment, we used the default number of trees provided in Weka.

5) Logistic Regression

Logistic regression is the well-known statistical model that uses logistic function for predicting class member by comparing the probability values between the categorical dependent variable and independent variable of each instance [24]. In our experiment, we used the multinomial logistic regression algorithm in learning for predicting class member.

6) Support Vector Machine

SVM can classify both linear and nonlinear data [22-23]. The SVM learner also supports multiple-class problems by

computing the hyper-plane between each class and the rest. We used SVM polynomial kernel to classify class member of postures. In our experiment, we used SVM with polynomial kernel to classify data set. The degree of polynomial kernel was 3.

VI. RESULTS

This section compares the results of the six classification methods used in classifying different postures. In the experiment we used 5-fold cross-validation to determine performance of each method with the results as given in Table II. Differences of the building model time of each classification method are not significant in this experiment. The wrong predictions mainly occurred with *Lie Down* because Kinect library function sometimes gave error joint detections on the body while people lay down.

From Table II, the neural network, D-tree, random forest and support vector machine methods almost have same results and give better results than the naïve Bayes and logistic regression methods; and Max-Min normalization is the best one of all results for the D-tree. Therefore, we select the D-tree with Max-Min normalization in our postural classifications.

As a comparison, we also used data of twenty skeletal joints to build classification models without transformation. The results are shown in Table III. It is clear that using only transformed attributes based on domain knowledge provided better performance.

TABLE II. RESULT OF EACH MODEL WITH DATA TRANSFORMATION.

Non-Normalization			
Model	ROC Area	Sensitivity	Accuracy
NN	0.981	97.20 %	97.68 %
D-tree (J48)	0.971	97.30 %	97.32 %
Naïve Bayes	0.869	89.90 %	89.90 %
SVM	0.967	97.10 %	97.47 %
Logistic Regression	0.961	96.00 %	95.95 %
Random Forest	0.975	97.80 %	97.75 %
Max-Min Normalization			
Model	ROC Area	Sensitivity	Accuracy
NN	0.983	97.60 %	97.87 %
D-tree (J48)	0.991	97.40 %	97.88 %
Naïve Bayes	0.994	95.70 %	85.69 %
SVM	0.971	97.00 %	97.52 %
Logistic Regression	0.970	96.80 %	96.78 %
Random Forest	0.975	97.90 %	97.88 %
Z-score Normalization			
Model	ROC Area	Sensitivity	Accuracy
NN	0.981	97.00 %	97.68 %
D-tree (J48)	0.969	97.30 %	97.32 %
Naïve Bayes	0.899	89.00 %	89.90 %
SVM	0.969	97.00 %	97.50 %
Logistic Regression	0.961	96.40 %	95.95 %
Random Forest	0.975	97.80 %	97.80 %

TABLE III. RESULTS WITH TWENTY SKELETAL JOINTS WITHOUT DATA TRANSFORMATION.

Non-Normalization			
Model	ROC Area	Sensitivity	Accuracy
NN	0.943	94.25 %	94.20 %
D-tree (J48)	0.981	96.20 %	96.16 %
Naïve Bayes	0.917	77.50 %	77.50 %
SVM	0.945	95.42 %	95.30 %
Logistic Regression	0.957	84.10 %	84.13 %
Random Forest	0.997	94.50 %	90.01 %
Max-Min Normalization			
Model	ROC Area	Sensitivity	Accuracy
NN	0.944	94.35 %	94.35 %
D-tree (J48)	0.981	96.20 %	96.19 %
Naïve Bayes	0.917	77.50 %	77.50 %
SVM	0.952	95.70 %	95.47 %
Logistic Regression	0.841	95.70 %	84.13 %
Random Forest	0.937	92.40 %	91.43 %
Z-score Normalization			
Model	ROC Area	Sensitivity	Accuracy
NN	0.945	94.25 %	94.25 %
D-tree (J48)	0.985	97.00 %	97.03 %
Naïve Bayes	0.918	77.70 %	77.74 %
SVM	0.940	95.60 %	95.20 %
Logistic Regression	0.841	92.70 %	84.13 %
Random Forest	0.965	92.80 %	91.46 %

VII. CONCLUSION AND FUTURE WORK

In this paper, we presented a system for postural detection and classification of elderly people while watching televisions. We used Kinect camera to record postures of elderly people. The input data have four attributes (angles of the left knee and right knee, distance from hip to a room floor, and aspect ratio). The appropriate distance of the camera from the person is about 1.8 to 3.0 meters. In the experiment, we selected six models to compare the results of postural classification and selected optimal suitable classifier by using 5-fold cross-validation. The six classification methods were NN, SVM, D-tree, logistic regression, random forest and naïve Bayes. The most suitable model was found to be D-tree with Max-Min normalization of the transformed.

In the near future, we plan to include a wider range of postures such as half-lying down, slouching, and so on, into our postural classification system so that we could learn more about the behaviors of the elderly people while they watch televisions at home. These obtained data would be very useful to their family members, physicians, and other health practitioners. Moreover, we could include a facial recognition system to make us understand the elderly people's expressions and emotions even better, say, while they watch comedy shows or suspense movies. Above all, more and more technologies could be used to monitor and keep records of the daily lives of the elderly people, especially while being at home, to help us better understand and innovate technologies to serve them. At the end of the day we may ask ourselves, "Are we prepared for the imminent ageing society yet?"

Acknowledgment

This work is supported by the Higher Education Research Promotion and National Research University Project, Thailand's Office of the Higher Education Commission, and School of Information Technology at KMUTT. We thank all volunteers and the SIT's D-Lab students and staff for their kind help.

References

- [1] United Nations, "World Population Ageing: 1950-2050," [Online] www.un.org/esa/population/publications/worldageing19502050, [May 1, 2014].
- [2] United Nations, "World Population Ageing 2007," [Online] www.un.org/esa/population/publications/WPA2007/wpp2007.htm, [May 1, 2014].
- [3] Thailand's National Statistical Office, "Information Services," [Online] service.nso.go.th/nso/nso_center/project/table/files/S-elderly/2550/000/00_S-elderly_2550_000_000000_00200.xls, [May 16, 2014].
- [4] K. Reid, "Lifeline or Leisure?: TV's Role in the Lives of the Elderly," [Online] <http://www.medialit.org/reading-room/lifeline-or-leisure-tvs-role-lives-elderly>, [May 1, 2014].
- [5] M. Graney and E. Graney, "Communications Activity Substitutions in Aging," *Journal of Communication*, Vol. 24, Issue 4, 1974, pp. 88-96.
- [6] S. Kensaku, "Older People and Television Viewing in Japan," *Journal of NHK Broadcasting Studies*, No. 8, 2010, pp. 63-94.
- [7] B. Oestlund, B. Jönsson and P. Waller, "Watching Television in Later Life: A deeper understanding of the meaning of TV viewing for design in geriatric contexts," *Scandinavian Journal of Caring Sciences*, Vol. 24, Issue 2, 2010, pp. 233-243.
- [8] N. Sachiko, K. Toshiyuki and M. Emi, "Television as the Most 'Relaxing' Medium: From the Time Use Survey on Television and Moods," *Journal of NHK Broadcasting Studies*, No. 8, 2010, pp. 95-122.
- [9] K. Nakajima, H. Matsui and K. Sasaki, "Telenmonitoring system of television's operating state for remotely located families," *Information Technology Applications in Biomedicine (ITAB2007)*, the 6th International Special Topic Conference on, Tokyo, Japan, Nov. 8-11, 2007, pp. 186-188.
- [10] J. Choomkasean, P. Mongkolnam and J. H. Chan, "Multimedia Delivery for Elderly People: A Conceptual Model," *The 7th International Conference on Advances in Information Technology (IAIT2012)*, Bangkok, Thailand, Dec. 6-7, 2012, pp. 58-69.
- [11] Microsoft, "MSDN – the Microsoft Developer Network," [Online] <http://msdn.microsoft.com>, [May 1, 2014].
- [12] J. Webb and J. Ashley, "Beginning Kinect programming with the Microsoft Kinect SDK," [E-book], Apress©, pp. 1-321.
- [13] Microsoft Developer Network, "Kinect for Windows Architecture" [Online] <http://msdn.microsoft.com/en-us/library/jj131023.aspx>, [16 June, 2014].
- [14] O. Patsadu, C. Nukoolkit and B. Watanapa, "Human Gesture Recognition Using Kinect Camera," *Computer Science and Software Engineering (JCSSE)*, 2012 International Joint Conference on, Bangkok, Thailand, May 30, 2012-Jun. 1, 2012, pp. 28-32.
- [15] Y. Liu, Z. Zhang, A. Li and M. Wang, "View Independent Human Posture Identification using Kinect," *Biomedical Engineering and Informatics (BMEI)*, 2012 the 5th International Conference on, Chongqing, China, Oct. 16-18, 2012, pp. 1590-1593.
- [16] S. Kaenchan, P. Mongkolnam, B. Watanapa, and S. Sathienpong, "Automatic Multiple Kinect Cameras Setting for Simple Walking Posture Analysis", *The 17th International Computer Science and Engineering Conference*, Bangkok, Thailand, Sep. 4-6, 2013, pp. 250-254.
- [17] P. Paliyawan, C. Nukoolkit and P. Mongkolnam, "Prolonged Sitting Detection for Office Workers Syndrome Prevention Using Kinect," 2014 *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2014 the 11th International Conference on, Nakhon Ratchasima, Thailand, May 14-17, 2014.
- [18] F. Cardile, G. Iannizzotto and F. La Rosa, "A Vision-Based System for Elderly Patients Monitoring," *Human System Interactions (HSI)*, 2010 3rd Conference on, Rzeszów, Poland, May 13-15, 2010, pp. 195-202.
- [19] M. Fahim, I. Fatima, S. Lee and Y. K. Lee, "Daily Life Activity Tracking Application for Smart Homes using Android Smartphone," *Advanced Communication Technology (ICACT)*, 2012 the 14th International Conference on, PyeongChang, South Korea, Feb. 19-22, 2012, pp. 241-245.
- [20] A. Correa, A. Morell, M. Barcelo and J. L. Vicario, "Navigation system for elderly care applications based on wireless sensor networks," *Signal Processing Conference (EUSIPCO)*, 2012 Proceedings of the 20th European, Bucharest, Romania, Aug. 27-31, 2012, pp. 210-214.
- [21] J. T. Kim, J. Y. Soh, S. H. Kim and K. Y. Chung, "Emergency Situation Alarm System Motion using Tracking of People Like Elderly Live Alone," *Information Science and Applications (ICISA)*, 2013 International Conference on, Suwon, South Korea, Jun. 24-26, 2013, pp. 1-4.
- [22] J. Han and M. Kamber, "Data mining concepts and techniques," Morgan Kaufmann Publishers, 2nd Edition, 2006, pp. 285-378.
- [23] The University of Waikato, "Weka 3-Data Mining with Open Source Machine Learning Software in Java," [Online] <http://www.cs.waikato.ac.nz/ml/weka/>, [May 1, 2014].
- [24] T. Hastie, R. Tibshirani and J. Friedman, "The Elements of Statistical Learning Data Mining, Inference, and Prediction", 2nd Edition, 2008, pp. 119-603.

ภาคผนวก ค
ผลงานที่ได้รับการตีพิมพ์

**Multiple-stage Classification of Human Poses while Watching
Television**

Thammarsat Visutarrom, Pornchai Mongkolnam, and Jonathan Hoyin Chan

2014 2nd International Symposium on Computational and Business Intelligence (ISCBI-II 2014)

December 7-8, 2014

Multiple-stage Classification of Human Poses while Watching Television

Thammarsat Visutarrom, Pornchai Mongkolnam, and Jonathan H. Chan

Data and Knowledge Laboratory
School of Information Technology
King Mongkut's University of Technology Thonburi
Bangkok 10140, Thailand
E-mail: jonathan@sit.kmutt.ac.th

Abstract— We compared the accuracy measure between a single-stage classifier model and a multiple-stage classifier model in postural classifications using Kinect. Postural training sets were collected from Kinect's skeletal data streams, based on some of the common human postures during television watching. Three types of training sets were used, including Kinect's raw skeletal training set, skeletons with attribute selection training set, and skeletal position transformation training set. We selected four learning models, namely, neural network, naïve Bayes, logistic regression, and decision tree, for learning our data sets and classifying a testing set to find the appropriate learning model. The best accuracy value of our experiment was 87.68 % by using skeletal position transformation training set with neural network. In the future, we will apply our technique and methodology to track elderly behaviors while they are watching television.

Keywords— *postural classification; Kinect; data transformation; multiple-stage classifier; television watching*

I. INTRODUCTION

Human posture classification is a very challenging research which has been an ongoing study for a long time by researchers in many different fields such as medical science, sport science, and social science. Current technology allows for better human posture classification with a high accuracy. In 2010, the Kinect system was released by Microsoft. Soon afterwards, it had become popular among many researchers for human body detection and human postural classification, largely due to its capability in separating the human body from a complex background. In addition, Kinect has a built-in software library that helps to separate each skeletal joint and enables the analysis of a whole skeleton. It can detect up to twenty joints, as shown in Fig. 1, each with three positional coordinates X, Y and Z. Thus there are a total of 60 raw attributes for each image or frame.

Unlike most other researchers, we use Kinect to classify human postures while watching television. In this paper, we investigate two classification architectures: a single-stage classifier model and a multiple-stage classifier model. Both models are given three types of training sets, including raw skeletal training set, skeleton with attributes selection training set, and skeletal position transformation training set. The data mining techniques used in our experiment include four learning models of back propagation neural network

(NN), naïve Bayes (NB), logistic regression (LR), and decision tree (J48).

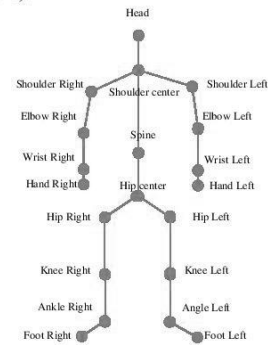


Figure 1. Twenty joints obtained from Kinect.

II. LITERATURE REVIEW

Human posture detection can be separated into two techniques: 2D image and 3D image processing [1-3]. The 2D image processing was popular in the past because it used pixel color information to analyze the data. Consequently, the performance of its postural classification heavily depended on environmental factors such as light conditions and background complexity.

Using 3D image processing can decrease the background complexity problem that occurs with 2D images. The reason being 3D image processing provides depth data that can help to better separate each object. In addition, 3D image processing with infrared sensors could be used in dark rooms, making it suitable for even complex human postural analyses.

In 2003, Cohen and Li [4] presented the human postural classification with SVM modeling using the 3D visual-hull constructed from a set of silhouette input data. The system returned the classified human body postures in the form of thumbnail images.

In 2007, Wu and Aghajan [5] proposed the method of human posture estimation in a multiple camera network by using the concept of an opportunistic fusion framework that

comprised three dimensions of space, time, and feature levels to obtain a 3D human skeleton.

In 2010, Kinect [6-9] was launched by Microsoft. It was a 3D camera game controller which was controlled by the human gestures of a player. Currently there are two software libraries for analyzing skeletal joints, i.e., Microsoft SDK and OpenNI. The OpenNI library can analyze 15 main joints of a human body. The Microsoft SDK library has two detection modes. The first mode is the default stand tracking that can analyze 20 joints of the body, and the second mode is the seated tracking mode that can analyze 10 upper shoulder joints of the human body. Kinect has three input components, including depth image sensors, RGB camera, and a multi-array microphone as shown in Fig. 2. With the ability to separate the human body from a complex background and in analyzing the human skeletal joints, Kinect has been applied by many researchers in their studies since its inception.

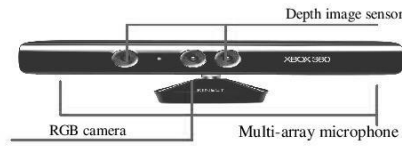


Figure 2. Kinect components [10].

In 2011, Hüke and Khalifa [11] presented a real-time gesture classification system to classify the dancing gestures from moving skeletal joint data obtained from Kinect. The accuracy of their system was 96.9% using the approximate 4-second record of the skeletal motion.

In 2014, Visutaron et al. [12] presented a system for simple postural detection and classification of elderly people while watching television. Their experiments have four standard postures including stand, sit, sit on floor and lie down. They selected six models to compare the results of postural classification and selected the most suitable classifier by using 5-fold cross-validation. The six classification methods used were neural network, support vector machine, decision tree, logistic regression, random forest and naïve Bayes. The best accuracy was 97.88%, obtained by decision tree with Max-Min normalization technique.

In 2014, Dai et al. [13] proposed a machine learning and vision-based method for elderly fall detection using statistical human posture sequence modeling. A series of laboratory simulated falls and activities of daily livings (ADLs) were performed and recorded by Kinect. Hidden Markov Models were used for modeling the fall posture sequences and distinguishing different fall activities and ADLs. The average fall recognition rate was above 80%.

III. PROCESS OVERVIEW

Our experiment can be separated into two parts: data preparation and data classification. The workflow is shown

in Fig. 3. Data preparation is used to prepare the training sets before the learning of each model. Data classification is to compare the postural classification result of each training-set against a testing dataset. We use Kinect XBOX 360 with the Microsoft SDK library version 1.8 with the default stand tracking mode to detect a human body having 20 skeletal joints. The distance for test subjects is between 1.8-3.0 meters away from Kinect during the data collection process.

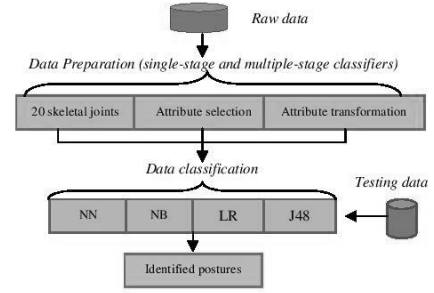


Figure 3. Experimental workflow.

Training sets were recorded from seven subjects, including four males and three females. Their ages are between 26-65 years old, with height between 150-180 centimeters and various body sizes. Our training sets have a total of 110,751 instances. The testing data has 20,868 instances obtained from another three subjects. Their ages are between 30-45 years old, with height between 155-180 centimeters. The details of number of instances in each posture for both training set and testing set are shown in Table I. There are a total of 18 postural class memberships as summarized in Table I. These postures are also illustrated in Fig. 4 as four main class memberships of stand, sit, sit on floor and lie down.

TABLE I. NUMBER OF INSTANCES IN EACH POSTURE.

Posture	Number of instances	
	Training set	Testing set
Normal stand	6584	985
Stand both hands up	6325	1445
Stand left hand up	6575	856
Stand right hand up	5892	1345
Stand lean forward	6320	964
Sit straight back	5981	869
Sit lean forward	5888	1250
Sit lean backward	5123	958
Sit both hands up	5320	1345
Sit left hand up	6120	945
Sit right hand up	5369	1450
Sit on floor straight back	5842	1065
Sit on floor lean forward	6923	1425
Sit on floor lean backward	6320	954
Sit on floor both hands up	6790	1356
Sit on floor left hand up	6952	1475
Sit on floor right hand up	5602	1185
Lie down in various manners	6825	996

Class membership						
Stand						
	Normal stand	Stand both hands up	Stand left hand up	Stand right hand up	Stand lean forward	
Sit						
	Sit straight back	Sit lean forward	Sit lean backward	Sit both hands up	Sit left hand up	Sit right hand up
Sit on floor						
	Sit on floor straight back	Sit on floor lean forward	Sit on floor lean backward	Sit on floor both hands up	Sit on floor left hand up	Sit on floor right hand up
Lie down						
	Lie down in various manners					

Figure 4. Class membership.

A. Data preparation

This step is to prepare the training sets before being used in learning of each model. In our experiment, we considered two types of classifier architectures. First is a single-stage classifier and second is a multiple-stage classifier.

A single-stage classifier model uses all training data packaged in one set. Therefore, the single-stage classifier has only one model in the prediction data as shown in Fig. 5 (a). Consequently, if the training data are large, it is very time consuming in creating the model.

On the other hand, a multiple-stage classifier would partition the data into multiple training sets in the first stage, and that would help to reduce the time to create the model, depending on the number of partitioned training sets. Therefore, a multiple-stage classifier requires many models working together for prediction as shown in Fig. 5 (b).

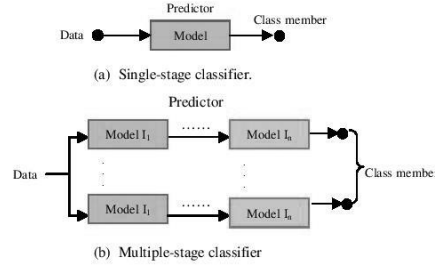


Figure 5. Single-stage classifier and multiple-stage classifier.

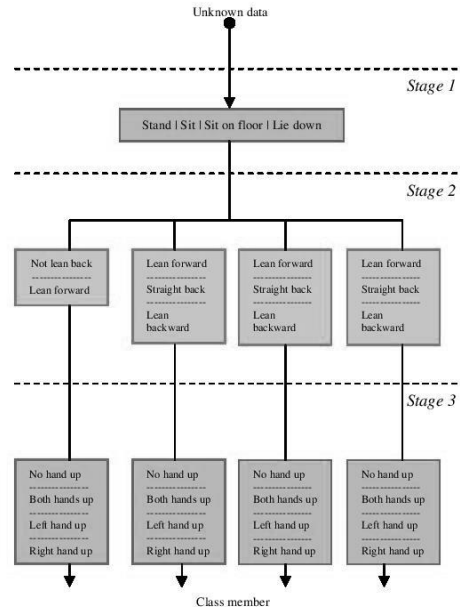


Figure 6. Multiple-stage classifier model.

In this work, we separate the classification task according to target classes of training. The classifier is divided into three stages as shown in Fig. 6. The first stage is used to classify four main postures, including stand, sit, sit on floor, and lie down. The second and third stages are used to classify sub-postures. In particular, the second stage is used to classify the *back* posture. The *back* posture is used to classify lean forward, straight back, or lean backward. The third stage is used to classify *arm* posture. The *arm* posture is used to classify hand-up, including both hands up, left hand up, and right hand up.

Both single-stage classifier and multiple-stage classifier have three types of training data set as follows.

1) Raw skeletal training set is the training data which use all 20 skeletal joint positions. Each joint has three coordinates X, Y, and Z. Each coordinate is considered one attribute; therefore, this training set has 60 total attributes.

2) Skeleton with attribute selection training set is the reduced number of attributes and reduced time for finding relationships between attributes. One advantage of the attribute reduction (feature selection) is that less time is needed in creating the model. We compare two-attribute selection techniques in order to reduce the dimensionality from the 60 attributes. They are the gain ratio attribute evaluator and the information gain attribute evaluator. The ranker method of feature selection is used in our training set to obtain the top 10, 20, 30, 40, 50 attributes as features. In this work, both the single-stage classifier and the multiple-stage classifier use the same ranker attributes.

3) Skeletal position transformation training set is the training set with transformed data from the skeletal joints. Using raw data may make training set have complexity that affect in learning of classifier and result in error accumulation due to redundant attributes. Attributes transformation is a technique that help to reduce the complexity of attributes in training set. Therefore, we select this technique to compare accuracy with other training sets. Equations used in transformed attributes are the equations that are used to observe the changes of each part of the human body. Transformation data can help to reduce the dimensionality of the problem. In addition, transformation data can help increase the relationships in training and increase the accuracy of prediction of the learning model. We select 12 skeletal joints to create nine attributes in the training set, including shoulder center, shoulder left, shoulder right, elbow left, elbow right, wrist left, wrist right, hip center, hip left, hip right, knee left, and knee right. The nine attributes include angle knee left, angle knee right, aspect ratio of height and width, distance from hip to room floor, back status, hand_SWL, hand_SWR, hand_SEL, and hand_SER.

For stand, sit, sit on floor and lie down postures, they have different aspect ratio of the width and the height. Therefore, the aspect ratio of width and height was selected

to be one of the attributes for classification. We found the aspect ratio can help to classify lie down posture from other postures as well. However, using the aspect ratio only cannot give a high accuracy with all postures. We added three transformed attributes including angle of left knee, angle of right knee, and distance from hip to the room floor for improving the prediction of stand, sit and sit on floor. The angles of left knee and right knee can help to identify stand and sit. When people stand, the angles at their knee joints are larger. On the other hand, when they sit, the angles of their knees are smaller. Distance from hip to room floor can help to identify sit and sit on floor. When people sit on floor, distance between hip to room floor will be less than sit on sofa and chair. In addition, we classify hand up and back characteristics of each posture for postural classification to provide even more details. Back status is used to classify the backward characteristics, including lean back forward, straight back and lean back backward. hand_SWL, hand_SWR, hand_SEL, and hand_SER are used to classify hand up of each posture that has 3 characteristics including hand up right, hand up left and both hands up. The equations used to find the values of those attributes are given below in Eqns. (1) – (12).

Aspect ratio of width and height

$$Width = Max_x - Min_x, \quad (1)$$

$$Height = Max_y - Min_y, \quad (2)$$

$$Aspect_Ratio = \frac{Width}{Height}, \quad (3)$$

Max and Min is the maximum and minimum values of X and Y coordinates of only five skeleton joints from head, shoulder left, shoulder right, foot left, and foot right.

Angles of left knee and right knee

$$A = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2}, \quad (4)$$

$$B = |Y_1 - Y_2|, \quad (5)$$

$$\theta = \sin^{-1}\left(\frac{A}{B}\right). \quad (6)$$

X_1 , Y_1 and Z_1 are coordinates of hip left and right. X_2 , Y_2 and Z_2 are values of knee left and right. Angles of left knee and right knee can help to identify stand and sit.

Distance from hip to room floor

$$Distance = \frac{AX_3 + BY_3 + CZ_3 + D}{\sqrt{A^2 + B^2 + C^2}}, \quad (8)$$

X_3 , Y_3 and Z_3 are the coordinates of the hip center. A, B, C and D are coefficients of the floor plane. Distance from hip to room floor can help to identify sit and sit on floor.

Back

$$\text{Back status} = Z_4 - Z_3, \quad (7)$$

Z_4 is Z coordinate of shoulder center and Z_3 is Z coordinate of hip center. Back is used to classify the backward characteristics that include lean back forward, straight back and lean back backward.

Hand up

$$\text{Hand_SWL} = Y_4 - Y_6, \quad (9)$$

$$\text{Hand_SWR} = Y_5 - Y_7, \quad (10)$$

$$\text{Hand_SEL} = Y_4 - Y_8, \quad (11)$$

$$\text{Hand_SER} = Y_5 - Y_9, \quad (12)$$

Y_4 and Y_5 are the Y coordinates of shoulder left and shoulder right. Y_6 and Y_7 are the Y coordinates of wrist left and wrist right. Y_8 and Y_9 are Y coordinates of elbow left and elbow right. All attributes in (9)–(12) are used to classify hand up right, hand up left and both hands up. Hand_SWL and Hand_SWR are differences between the Y coordinates of shoulder and wrist of the left hand side and the right hand side, respectively. Hand_SEL and Hand_SER are differences between the Y coordinates of shoulder and elbow of the left hand side and the right hand side, respectively.

B. Data classification

In this work, we select four learning models, including neural network, naïve Bayes, logistic regression, and decision tree; all of which are available in the Weka data mining tool [14].

Neural network [15–16] used in the experiment is the simple multilayer perceptron (MLP) which uses the back-propagation algorithm in learning. Structure of the neural network in the experiment comprises three layers, including input layer, hidden layer, and output layer. The input and output layer values depend on the amount of attributes and the amount of class memberships of each training set. Hidden layer value used is the default value provided by the Weka data mining tool.

Decision tree [15–16] is used to classify data from class label, yielding the output as a flowchart-like tree structure. In the experiment, we used J48 algorithm to classify class membership as a set of decision nodes and leaf nodes. Each leaf node represents a class outcome label.

Naïve Bayes [15–16] is a simple probability classifier based on the Bayes' theorem with the assumption of independence between every pair of features. The nodes in the Bayesian model are created from given training data. Each node counts the number of rows per attribute value per class for nominal attributes and calculates the Gaussian distribution for numerical attributes.

Finally, we also used the multinomial logistic [15–16] regression algorithm for predicting class member and use the default parameter provided in Weka data mining tool.

IV. RESULTS

Table II provides the average accuracy result of each training model, against the testing data. We found that the skeletal training set of both the single-stage classifier model and multiple-stage classifier model produced poor results with the prediction testing set falling between 19% and 26%. Attributes selection of both the single-stage classifier model and multiple-stage classifier model increased the accuracy of the prediction testing set slightly. However, the results were still poor in the range of 15% to 34%, with better performance falling in the top 30 ranked attributes with the values between 22% and 34%. Interestingly, we found Y coordinate was ranked in the top group that was important in our training set and X coordinate was ranked in the unimportant group with the training set. On the other hand, the skeletal position transformation training set produced better values than the skeletal training set and the skeleton with attributes selection training set. The skeletal position transformation training set of the single-stage classifier gave the accuracy results between 40% and 60%, and skeletal position transformation training set of the multiple-stage classifier gave the best results in the range of 79% to 87%. The learning model that gave the best results was neural network with skeletal position transformation training set using a multiple-stage classifier, with the average accuracy result of 87.68%.

The accuracy of each individual postural prediction using the skeletal position transformation training set, against the testing set, is shown in Table III. The results show that the performance of the individual posture models is quite similar except for the noticeably poorer accuracy for the lie down posture. The best classifier was neural network in all cases. This is consistent with previous works in medical data classification using classic machine learning techniques that show neural networks performs at least as well as other techniques and quite often outperforms the others [17]. One possible reason in our case is that there are more adjustable parameters that can be tuned in Weka.

In summary, attributes transformation is the technique that helped reduce the complexity of attributes in training set and increase the accuracy of the data prediction of each model. In addition, the reduced number of attributes helped decrease the time needed to create the classifier model. Most error prediction occurred with the lie down posture. In addition, we found the error of predicting postures occurred while people changed postures, or during the transition period. As a result, postural transitions should also be considered in order to detect and predict accurate postures while watching television.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we are interested in studying the differences of accuracy values of the single-stage classifier model and multiple-stage classifier model of postural classification. Postural data were recorded from Kinect. From the experiment results, the data transformation training set of multiple-stage classifier with the neural network learning model gave the best results.

In the near future, we want to apply our knowledge and methodology found herein for developing the postural classification of elderly people while watching television, which is the main pastime/activity of the elderly. More sub-postures such as stretching hands to the sides may be added. Being able to detect postures and subsequent behaviors can help their family members, physicians, and other health practitioners understand the elderly even more. As a result, better services and technologies could be created and offered to improve their quality of lives.

ACKNOWLEDGMENT

This work has been supported by the Higher Education Research Promotion and National Research University Project, Thailand's Office of the Higher Education Commission, and School of Information Technology at KMUTT. We thank all volunteers and D-Lab students and staff for their kind help and useful discussions.

REFERENCES

- [1] Y. Sun, M.H. Li, J.S. Hu and E.L. Wang, "2D RECOVERY OF HUMAN POSTURE," Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on, Beijing, China, Nov. 4-5, 2002, pp. 1638 - 1640.
- [2] K. Takahashi and S. Sugakawa, "Remarks on human posture classification using self-organizing map," Systems, Man and Cybernetics, 2004 IEEE International Conference on, Netherlands, Oct. 10-13, 2004, pp. 2623 - 2628.
- [3] C.C. Li and Y.Y. Chen, "Human Posture Recognition by Simple Rules," Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on, Southampton, United Kingdom, Apr. 10-12, 2006, pp. 351 - 356.
- [4] I. Cohen and H. Li, "Inference of Human Postures by Classification of 3D Human Body Shape", Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on, Nice, France, Oct. 17, 2003, pp. 74 - 81.
- [5] C. Wu and H. Aghajan, "Model-based Human Posture Estimation for Gesture Analysis in an Opportunistic Fusion Smart Camera Network", Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on, London, England, Sept. 5-7, 2007, pp. 453 - 458.
- [6] Microsoft, "Kinect for Windows Architecture," [Online] <http://msdn.microsoft.com/en-us/library/jj131023.aspx>, [Sep. 18, 2014].
- [7] M. Ruta, F. Scioscia, M. Di Summa, S. Ieva, E. Di Sciascio and M. Sacco, "Semantic matchmaking for Kinect-based posture and gesture recognition," Semantic Computing (ICSC), 2014 IEEE International Conference on, Hue Royal City, Vietnam, Aug. 1-3, 2012, pp. 309 - 312.
- [8] Y. Liu, Z. Zhang, A. Li and M. Wang, "View Independent Human Posture Identification using Kinect," Biomedical Engineering and Informatics (BMEI), 2012 the 5th International Conference on, Chongqing, China, Oct. 16-18, 2012, pp. 1590-1593.
- [9] H.P.H. Shum, E.S.L. Ho, Y. Jiang and S. Takagi, "Real-Time Posture Reconstruction for Microsoft Kinect," Cybernetics, IEEE Transactions on, vol. 43, Oct. 2013, pp. 1357 - 1369.
- [10] Xbox, "Kinect for Xbox 360," [Online] <http://www.xbox.com/en-US/kinect>, [Sep. 18 2014].
- [11] K. K. Hike and O.O. Khalifa, "Comparison of supervised and unsupervised learning classifiers for human posture recognition", Computer and Communication Engineering (ICCCE), 2010 International Conference on, Kuala Lumpur, Malaysia, May 1-12, 2010, pp. 1-6.
- [12] T. Visutarnom, P. Mongkolnam, and J. H. Chan, "Postural Classification using Kinect," 2014 International Computer Science and Engineering Conference, Khon Kaen, Thailand, Jul. 30 - Aug. 1, 2014, pp. 346 - 351.
- [13] X. Dai, M. Wu, B. Davidson, M. Mahoor and Z. Jun, "Image-Based Fall Detection with Human Posture Sequence Modeling," Healthcare Informatics (ICHI), 2013 IEEE International Conference on, Philadelphia, Pennsylvania, USA, Sept. 9-11, 2013, pp. 376 - 381.
- [14] The University of Waikato, "Weka 3-Data Mining with Open Source Machine Learning Software in Java," [Online] <http://www.cs.waikato.ac.nz/ml/weka/>, [May 1, 2014].
- [15] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning - Data Mining, Inference, and Prediction, 2nd ed. Springer, 2008, pp. 119-603.
- [16] J. Han and M. Kamber, Data mining concepts and techniques, 2nd ed. Morgan Kaufmann Publishers, 2006, pp. 285-378.
- [17] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," J. Biomed. Inform. vol. 35, pp. 352-359, October 2002.

TABLE II. AVERAGE ACCURACY OF 18 POSTURES USING DIFFERENT FEATURE SELECTION METHODS

Single-Stage classifier					Multiple-stage classifier				
	NN	NB	LGT	J48		NN	NB	LGT	J48
All 20 joints	25.78	19.79	22.80	20.89	All 20 joints	26.89	18.57	20.04	18.65
Gain 10	16.89	15.89	21.90	20.70	Gain 10	17.89	11.87	18.76	19.00
Gain 20	32.78	23.45	25.55	25.89	Gain 20	24.56	17.02	19.01	20.01
Gain 30	33.89	26.50	27.45	28.61	Gain 30	34.05	22.34	28.23	22.45
Gain 40	28.45	22.00	25.12	24.87	Gain 40	27.46	21.03	26.45	20.34
Gain 50	24.30	21.00	20.56	20.86	Gain 50	26.04	20.67	24.08	19.04
Info-Gain 10	16.70	15.81	21.90	20.76	Info-Gain 10	22.49	17.57	21.00	17.48
Info-Gain 20	32.06	22.80	25.45	26.89	Info-Gain 20	23.93	20.89	22.48	19.78
Info-Gain 30	32.45	24.32	25.45	24.31	Info-Gain 30	29.59	21.89	23.90	21.90
Info-Gain 40	28.76	22.04	25.50	25.12	Info-Gain 40	25.67	20.56	22.56	19.87
Info-Gain 50	23.50	21.45	19.55	21.56	Info-Gain 50	23.90	19.34	20.45	19.87
Attribute transformation	62.78	50.34	60.47	58.46	Attribute transformation	87.68	79.89	85.67	84.69

TABLE III. ACCURACY OF EACH POSTURE USING TRANSFORMED ATTRIBUTES

Single-Stage classifier					Multiple-stage classifier				
Posture	NN	NB	LGT	J48	Posture	NN	NB	LGT	J48
Normal stand	66.23	52.14	62.04	60.12	Normal stand	89.36	83.26	88.65	86.56
Stand both hands up	66.23	52.02	61.59	60.37	Stand both hands up	89.56	83.25	88.58	86.45
Stand left hand up	66.78	52.01	61.89	59.34	Stand left hand up	89.53	83.14	88.56	85.14
Stand right hand up	67.32	51.48	61.59	59.32	Stand right hand up	89.66	82.36	87.48	85.36
Stand lean forward	61.65	51.21	61.32	59.38	Stand lean forward	89.79	82.34	86.32	85.47
Sit straight back	60.25	51.42	61.56	59.12	Sit straight back	88.69	80.24	86.64	85.36
Sit lean forward	61.08	50.12	61.25	58.36	Sit lean forward	88.89	80.53	86.32	85.69
Sit lean backward	65.36	51.28	61.01	58.01	Sit lean backward	87.96	81.56	86.95	85.36
Sit both hands up	61.45	51.24	61.04	58.47	Sit both hands up	87.21	79.45	85.36	85.32
Sit left hand up	62.54	50.18	60.48	58.21	Sit left hand up	87.45	79.36	85.99	85.15
Sit right hand up	63.51	50.13	60.48	58.12	Sit right hand up	86.35	79.85	85.86	84.26
Sit on floor straight back	61.79	50.48	60.78	58.02	Sit on floor straight back	87.69	79.45	85.36	84.89
Sit on floor lean forward	60.35	50.23	59.62	58.15	Sit on floor lean forward	87.65	79.25	84.26	84.56
Sit on floor lean backward	62.35	50.02	59.89	58.36	Sit on floor lean backward	87.36	78.54	84.38	84.36
Sit on floor both hands up	61.23	50.45	59.42	58.69	Sit on floor both hands up	87.45	78.56	84.35	84.57
Sit on floor left hand up	65.25	50.36	59.24	58.25	Sit on floor left hand up	87.35	78.98	84.15	83.53
Sit on floor right hand up	61.48	50.23	59.67	58.36	Sit on floor right hand up	87.39	78.36	84.28	83.63
Lie down in various manners	55.21	41.2	55.63	53.68	Lie down in various manners	78.98	69.58	78.64	78.79

ประวัติผู้วิจัย

ชื่อ — สกุล

นาย ธรรมศาสตร์วิสุทธิธารมณ

วัน เดือน ปีเกิด

1 กรกฎาคม 2532

ประวัติการศึกษา

ระดับมัธยมศึกษา

โรงเรียนภูเก็ตวิทยาลัย พ.ศ. 2550

ระดับปริญญาตรี

วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์
มหาวิทยาลัยสงขลานครินทร์ พ.ศ. 2554

ระดับปริญญาโท

วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมซอฟต์แวร์
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี พ.ศ. 2557

ทุนการศึกษา หรือทุนวิจัย

ทุนอุดหนุนสนับสนุนการวิจัยระดับบัณฑิตศึกษา
ทบวงมหาวิทยาลัย ปีงบประมาณ 2556

ผลงานที่ได้รับการตีพิมพ์

1. ธรรมศาสตร์ วิสุทธิธารมณ, พรชัย มงคลนาม และโจนาธาน โอ
ยีน ชาน, 2557, "การจำแนกท่าทางขณะรับชมโทรทัศน์โดยใช้
กล้องคิเนค," **The 10th National Conference on Computing
and Information Technology**, ครั้งที่ 10, 8-9 พฤษภาคม 2557,
โรงแรมอัสสนาลากูนา, อ. ภูเก็ต, จ. ภูเก็ต, หน้า 583-588 .
2. Visutarrom, T., Mongkolnam, P. and Chan, J. H., 2014,
"Postural Classification using Kinect," **The 2014
International Computer Science and Engineering
Conference (ICSEC2014)**, July 30 - August 1, Pullman Khon
Kaen Raja Orchid, Khon Kaen, Thailand, pp. 346-351.

ผลงานที่ได้รับการตีพิมพ์

3. Visutarrom, T., Mongkolnam, P. and Chan, J. H., 2014, "Multiple-stage Classification of Human Poses while Watching Television," **2014 2nd International Symposium on Computational and Business Intelligence (ISCBI-II 2014)**, December 7-8, The Ashtan Sarovar Portico, New Delhi, India, (Accepted).