



รายงานวิจัยฉบับสมบูรณ์

โครงการ การพิสูจน์บุคคลโดยวิธีการวิเคราะห์ข้อมูล 3 มิติ แบบทันทีทันใด  
ผ่านระบบเน็ตเวิร์คเส้นใยแก้ว

รองศาสตราจารย์ ดร.ไพศาล มณีสว่าง


ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนครสวรรค์

กันยายน 2560

มหาวิทยาลัยนครสวรรค์  
เลขที่ 05 ๒๙ 2560  
103/1655  
จ ๑ ๑A  
623  
๗๑๑๖5  
2560

# รายงานวิจัยฉบับสมบูรณ์

โครงการ การพิสูจน์บุคคลโดยวิธีการวิเคราะห์ข้อมูล 3 มิติ แบบทันทีทันใด  
ผ่านระบบเน็ตเวิร์คเส้นใยแก้ว

- 
1. รศ.ดร.ไพศาล มณีสว่าง      สังกัด ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์  
คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร

สนับสนุนโดยงบประมาณรายได้ กองทุนวิจัย มหาวิทยาลัยนเรศวร  
ปีงบประมาณ 2559

# สารบัญ

บทคัดย่อ.....	3
Abstract.....	4
Executive Summary.....	5
บทที่ 1 .....	6
บทนำ.....	6
บทที่ 2 .....	8
กรอบแนวคิดทฤษฎี และงานวิจัยที่เกี่ยวข้อง .....	8
การทบทวนวรรณกรรม/สารสนเทศ (Information) ที่เกี่ยวข้อง .....	10
วัตถุประสงค์ของการวิจัย .....	11
ประโยชน์ที่คาดว่าจะได้รับ .....	11
ขอบเขตการวิจัย .....	11
วิธีดำเนินการวิจัย.....	12
บทที่ 3 .....	13
วิธีดำเนินการวิจัย .....	13
บทที่ 4 .....	17
ผลการวิจัย.....	17
บทที่ 5 .....	21
สรุปผลการวิจัย.....	21
บรรณานุกรม .....	22
ภาคผนวก .....	23

## บทคัดย่อ

งานวิจัยนี้นำเสนอเทคนิคการพิสูจน์บุคคลด้วยวิธีการอินครีเมนต์ไดนามิกโทมวอร์ปิง (IDTW) โดยวิเคราะห์ข้อมูล 3 มิติที่บันทึกได้จากการตรวจจับการเคลื่อนไหวของร่างกาย Microsoft's Kinect® ข้อมูลที่นำมาวิเคราะห์จำเป็นต้องมีคุณลักษณะเด่นทางชีววิทยาที่ชัดเจน และสามารถสกัดหรือดึงค่าของโครงร่างหรือโครงกระดูก (skeleton) ของร่างกาย ซึ่งเป็นชุดข้อมูลที่แสดงจุดสำคัญที่เป็นประโยชน์ต่อการพิสูจน์บุคคล เพื่อใช้ข้อมูลเหล่านั้นเปรียบเทียบความคล้ายกันกับลำดับข้อมูลการเคลื่อนไหวที่ถูกบันทึกไว้ในฐานข้อมูลก่อนหน้า จากผลการทดลองพบว่าวิธีการ IDTW สามารถตรวจสอบการรู้จำของบุคคลจากท่าทางการเดินได้อย่างมีประสิทธิภาพ และใช้ระยะเวลาในการประมวลผลลดลงเมื่อเปรียบเทียบกับวิธีการดั้งเดิม ซึ่งสามารถใช้เทคนิคที่นำเสนอไปประยุกต์ใช้ในการพิสูจน์บุคคลที่มีการประมวลผลแบบทันทีทันใดหรือแบบออนไลน์ได้



## Abstract

In this research we propose a method of early recognition of a subject person based on Incremental Dynamic Time Warping (IDTW) by identification of personal gait patterns recorded in a video recorded via a 3D depth sensing camera; Microsoft's Kinect<sup>®</sup>. A video camera recording of a person is connected by Fiber to the home (FTTH) technology to a cloud data center. The video is analyzed in real time and the online automated personal identification system provides immediate identification of the subject person based on comparison with stored data. The experimental results show that the IDTW algorithm increases the efficiency of recognizing the subject, even with incomplete data patterns.



## Executive Summary

ปัจจุบันอุตสาหกรรมการรักษาความปลอดภัยพัฒนาไปอย่างรวดเร็วและมีอัตราเติบโตอย่างต่อเนื่อง เนื่องจากได้รับแรงกระตุ้นจากวิกฤตภัยคุกคามจากการก่อการร้ายจากทั่วโลกและความก้าวหน้าทางเทคโนโลยี นอกจากนี้ความกังวลถึงความปลอดภัยในทรัพย์สินเนื่องจากรายงานมิจฉาชีพเพิ่มมากขึ้น ทำให้มีการติดตั้งอุปกรณ์รักษาความปลอดภัยเพิ่มขึ้นนั่นเอง เป็นสาเหตุให้การพัฒนาระบบตรวจสอบหรือระบบรักษาความปลอดภัยมีความต้องการมากขึ้นเช่นกัน โดยเทคนิคของการค้นหาบุคคลและระบุตัวตนนั้นจำเป็นต้องใช้ข้อมูลคุณลักษณะเด่นทางชีววิทยา (Biologic features) ที่คัดกรองจากสื่อดิจิทัลต่าง ๆ

ข้อมูลคุณลักษณะเด่นทางชีววิทยาจำเป็นต้องมีการบันทึกข้อมูลจากอุปกรณ์ตรวจจับที่สามารถสกัดหรือดึงค่าของโครงร่างหรือโครงกระดูก (skeleton) ของร่างกาย ซึ่งเป็นชุดข้อมูลที่แสดงจุดสำคัญที่เป็นประโยชน์ต่อการพิสูจน์บุคคลและจำเป็นต้องมีการบันทึกหรือตรวจจับความเคลื่อนไหวของร่างกาย (Motion tracking) ที่แม่นยำและสามารถแยกส่วนประกอบของร่างกายเป็นส่วน ๆ ได้ ทั้งนี้เพื่อนำข้อมูลดังกล่าวมาวิเคราะห์ท่าทางการเดินของแต่ละบุคคลได้ ดังนั้นกระบวนการแยกส่วนเฉพาะของท่าทางการเดินจากภาพวิดีโอจะต้องอาศัยการคำนวณที่ซับซ้อนและมักจะทำให้เกิดความผิดพลาดได้ง่าย นอกจากนี้เทคนิคที่ใช้ในการตรวจสอบและเปรียบเทียบข้อมูลคุณลักษณะเด่นทางชีววิทยาของแต่ละบุคคล จะต้องให้ผลลัพธ์ที่มีความแม่นยำ มีประสิทธิภาพตอบสนองการทำงานได้อย่างรวดเร็ว

โครงการวิจัยนี้ผู้วิจัยทำการประยุกต์ใช้วิธีการอินครีเมนทัลไดนามิกไทม์วอร์ปปีง (IDTW) เป็นขั้นตอนวิธีสำหรับการเปรียบเทียบความคล้าย (Similarity Measure) ของลำดับที่มีความแตกต่างกันในด้านเวลาหรือความเร็ว ในการวิเคราะห์ข้อมูลที่บันทึกข้อมูลจากอุปกรณ์ตรวจจับ Kinect sensor จากนั้นจะทำการศึกษาเพื่อหาข้อสรุปเกี่ยวกับปัจจัยที่เกี่ยวข้องต่าง ๆ รวมไปถึงการทดสอบประสิทธิภาพการทำงานของวิธีการรู้จำบุคคลจากท่าทางการเดิน รวมไปถึงทดสอบความแม่นยำ เพื่อนำไปพัฒนาระบบพิสูจน์บุคคลอัตโนมัติจากท่าทางการเดินที่มีความถูกต้องแม่นยำ มีประสิทธิภาพต่อไป

รายงานนี้เป็นส่วนหนึ่งของโครงการวิจัยภายใต้การสนับสนุนโดยงบประมาณรายได้ กองทุนวิจัย มหาวิทยาลัยนเรศวร ประจำปีงบประมาณ 2559 โดยคณะผู้วิจัยได้รับการอนุมัติให้ทำงานวิจัยนี้ โดยมีชื่อโครงการ และรายละเอียดเกี่ยวกับโครงการวิจัยดังต่อไปนี้

ชื่อโครงการ (ภาษาไทย) การพิสูจน์บุคคลโดยวิธีการวิเคราะห์ข้อมูล 3 มิติ แบบทันทีทันใดผ่านระบบเน็ตเวิร์ค  
เส้นใยแก้ว

(ภาษาอังกฤษ) Real-time Person Identification via 3-dimensional data analysis method on the optical fiber network

คณะผู้วิจัย(ระบุสังกัดภาควิชา) และสัดส่วนที่ทำงานวิจัย (%)

หัวหน้าโครงการวิจัย

รองศาสตราจารย์ ดร.ไพศาล มณีสว่าง (100%)

อาจารย์สังกัดภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร

สถานที่จัดทำโครงการวิจัย

ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร

ความสำคัญและที่มาของปัญหาที่ทำการวิจัย

ปัจจุบันอุตสาหกรรมการรักษาความปลอดภัยพัฒนาไปอย่างรวดเร็วและมีอัตราเติบโตอย่างต่อเนื่อง เนื่องจากได้รับแรงกระตุ้นจากวิกฤตภัยคุกคามจากการก่อการร้ายจากทั่วโลกและความก้าวหน้าทางเทคโนโลยี โดยปัจจัยหนุนสำคัญคือการขยายตัวของความต้องการทั้งจากลูกค้าภาครัฐและภาคเอกชน โดยเฉพาะโครงการที่อยู่อาศัย ทั้งบ้านจัดสรร อาคารชุด อาคารสำนักงาน โรงงานอุตสาหกรรมในนิคมอุตสาหกรรม รวมไปถึงพื้นที่สาธารณะสำคัญต่าง ๆ เช่น สนามบิน สถานีขนส่งมวลชน ห้างสรรพสินค้า และโรงพยาบาล เป็นต้น นอกจากนี้ ความกังวลถึงความปลอดภัยในทรัพย์สินเนื่องจากรายงานมีอาชญากรรมเพิ่มขึ้น ทำให้มีการติดตั้งอุปกรณ์รักษาความปลอดภัยเพิ่มขึ้นนั่นเอง เป็นสาเหตุให้การพัฒนาระบบตรวจสอบหรือระบบรักษาความปลอดภัยมีความต้องการมากขึ้นเช่นกัน ซึ่งเทคโนโลยีการวิเคราะห์เนื้อหาของสื่อมัลติมีเดียเพื่อสร้างดรชนี่ เป็นอีกเทคโนโลยีหนึ่งที่สามารถนำมาใช้ในการค้นหาบุคคลและระบุตัวตนได้

ในเทคนิคของการค้นหาบุคคลและระบุตัวตนนั้นจำเป็นต้องใช้ข้อมูลคุณลักษณะเด่นทางชีววิทยา (Biologic features) ที่คัดกรองจากสื่อดิจิทัลต่าง ๆ เช่น ภาพถ่ายหรือวิดีโอของบุคคล ซึ่งในขั้นตอนการค้นหา บุคคลระบบคอมพิวเตอร์จะทำการวัดและวิเคราะห์ข้อมูลจากคุณลักษณะเด่นทางชีววิทยาของแต่ละบุคคลเพื่อนำไปสู่การรู้จักบุคคลแบบอัตโนมัติ ระบบการพิสูจน์บุคคลนั้นสามารถแบ่งได้เป็น 2 ประเภทคือ Active system และ Passive system สำหรับระบบการพิสูจน์บุคคลแบบแอ็กทีฟนั้น บุคคลจำเป็นต้องสัมผัสหรือติดต่อกับเครื่องรับรู้ (Sensor) ตัวอย่างของระบบนี้ได้แก่ ระบบการพิสูจน์บุคคลจากลายนิ้วมือ (Fingerprint system) และระบบการพิสูจน์บุคคลจากม่านตา เป็นต้น ส่วนระบบการพิสูจน์บุคคลแบบพาสซีฟไม่จำเป็นต้องให้บุคคลสัมผัสกับอุปกรณ์รับรู้ เพื่อการเก็บข้อมูล ในระบบนี้บุคคลเป้าหมายอาจจะไม่ได้ตระหนักว่าตนเองกำลังถูกระบบการพิสูจน์ทางชีววิทยาตรวจสอบอยู่ ตัวอย่างของระบบนี้ได้แก่ การรู้จำเสียงและการรู้จำใบหน้า จะเห็นได้ว่าระบบการพิสูจน์บุคคลทั้งสองประเภทจำเป็นต้องใช้เครื่องรับรู้ในบันทึกข้อมูลคุณลักษณะเด่นทางชีววิทยา แต่อย่างไรก็ตาม

ระบบการพิสูจน์บุคคลทั้งสองประเภทที่กล่าวมาก็ยังมีข้อจำกัดและปัญหา กล่าวคือ ในระบบการพิสูจน์บุคคลแบบแอ็กทีฟ บุคคลที่จะรับการตรวจสอบจำเป็นต้องสัมผัสกับเครื่องรับรู้เพื่อให้ระบบสามารถบันทึกข้อมูลได้ ในขณะที่ระบบการพิสูจน์บุคคลแบบพาสซีฟ การวิเคราะห์หรือค้นหาตำแหน่งของคุณลักษณะเด่นทางชีววิทยาที่ละเอียดและถูกต้อง จำเป็นต้องมีความไวต่อสภาพแวดล้อมของการบันทึกข้อมูล ได้แก่ ค่าความสว่างหรือเข้มของแสง และสิ่งกีดขวางต่อการบันทึกใบหน้าบุคคล เช่น แว่น ผม หรือหมวก เป็นต้น ซึ่งอาจส่งผลให้ประสิทธิภาพการรู้จำของระบบลดลง จากข้อจำกัดดังกล่าวส่งผลให้งานวิจัยในปัจจุบันมีความสนใจในระบบการพิสูจน์บุคคลแบบพาสซีฟที่มีการวัดค่าคุณลักษณะเด่นทางชีวภาพที่คัดกรองจากรูปร่างของบุคคลและท่าทางการเดิน (Gait gesture) สำหรับวิธีการประมวลผลสัญญาณภาพและการวิเคราะห์ภาพที่ซับซ้อน จะต้องประกอบไปด้วยการสกัดหรือดึงค่าของโครงร่างหรือโครงกระดูก (skeleton) ของร่างกาย ซึ่งเป็นชุดข้อมูลที่แสดงจุดสำคัญที่เป็นประโยชน์ต่อการพิสูจน์บุคคลและจำเป็นต้องมีการบันทึกหรือตรวจจับความเคลื่อนไหวของร่างกาย (Motion tracking) ที่แม่นยำและสามารถแยกส่วนประกอบของร่างกายเป็นส่วน ๆ ได้ ทั้งนี้เพื่อนำข้อมูลดังกล่าวมาวิเคราะห์ท่าทางการเดินของแต่ละบุคคลได้ ดังนั้นกระบวนการแยกส่วนเฉพาะของท่าทางการเดินจากภาพวิดีโอจะต้องอาศัยการคำนวณที่ซับซ้อนและมักจะทำให้เกิดความผิดพลาดได้ง่าย

จากปัญหาที่เกิดขึ้นจากการวิเคราะห์ข้อมูลภาพและวิดีโอซึ่งต้องการการคำนวณที่ซับซ้อนนั้น นักวิจัยในปัจจุบันนิยมใช้เครื่องรับรู้แบบ 3 มิติ ในการบันทึกท่าทางของบุคคล โดยบริษัทไมโครซอฟท์ (Microsoft Corporation) ได้เสนอเครื่องรับรู้แบบ 3 มิติ แบบใหม่ที่มีชื่อว่า Kinect sensor ซึ่งถูกนำมาใช้ในกล่องวิดีโอเกมแบบ X-Box 360 ซึ่งเป็นเครื่องที่ทำให้ผู้ใช้สามารถควบคุมจอร์เจส ไกลแบบอัตโนมัติโดยใช้ท่าทาง เครื่องดังกล่าวมีอุปกรณ์ตรวจจับและส่วนต่อประสานโปรแกรมประยุกต์ (Application Programming Interface หรือ API) ซึ่งจะทำการค้นหาและระบุตำแหน่งต่างๆ โครงกระดูกของผู้ใช้หรือกลุ่มของผู้ที่ปรากฏอยู่ตรงตำแหน่งด้านหน้ากล้องของอุปกรณ์รับรู้ งานวิจัยนี้จึงนำเสนอวิธีการพิสูจน์บุคคลจากการเคลื่อนไหว โดยอาศัยข้อมูลทุกส่วนของร่างกายที่ได้จาก Kinect sensor ซึ่งเป็นคุณลักษณะเด่นทางชีววิทยานอกเหนือจากชนิดอื่น ๆ ที่ได้กล่าวไว้ข้างต้น โดยงานวิจัยนี้จะประยุกต์ใช้วิธีปัญญาเชิงคำนวณ (Computational intelligence method) เพื่อตรวจสอบและเปรียบเทียบข้อมูลคุณลักษณะเด่นทางชีววิทยาของแต่ละบุคคล เพื่อจัดทำระบบพิสูจน์บุคคลอัตโนมัติจากท่าทางการเดินที่มีความถูกต้องแม่นยำ มีประสิทธิภาพ และประมวลผลได้อย่างรวดเร็ว



กรอบแนวคิดทฤษฎี และงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงรายละเอียดเกี่ยวกับความสำคัญและที่มาของปัญหาวิจัย รวมไปถึงการทบทวนวรรณกรรมของการศึกษาที่เกี่ยวข้อง และรายละเอียดโดยรวมของวิธีการพิสูจน์บุคคลและเทคนิคการรู้จำท่าทางเพื่อนำไปใช้ในการจัดท่าระบบพิสูจน์บุคคลอัตโนมัติจากท่าทางการเดิน ซึ่งรายละเอียดเป็นดังนี้

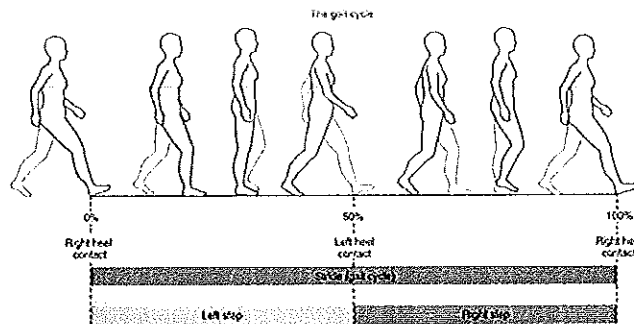
กรอบแนวคิดในการทำวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อนำเสนอวิธีการพิสูจน์บุคคลจากการเคลื่อนไหว และจัดท่าระบบพิสูจน์บุคคลอัตโนมัติจากท่าทางการเดิน โดยประยุกต์ใช้วิธีปัญญาเชิงคำนวณ (Computational intelligence method) เพื่อตรวจสอบและเปรียบเทียบข้อมูลคุณลักษณะเด่นทางชีววิทยาของแต่ละบุคคล ที่บันทึกได้จากอุปกรณ์ตรวจจับความเคลื่อนไหว 3 มิติ (Kinect sensor) ผู้วิจัยจึงได้รวบรวมเอกสารและงานวิจัยที่เกี่ยวข้องในประเด็นสำคัญ ดังนี้

2.1 ท่าทางการเดินของมนุษย์ (Human gait gesture)

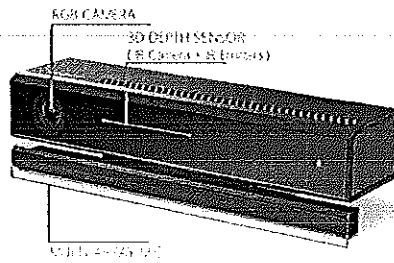
บุคลิกภาพของแต่ละคนจะเป็นสิ่งประจำตัวของคน ๆ นั้นที่ทำให้แตกต่างจากคนอื่น และมีหลายสิ่งหลายอย่างที่จะประกอบกันทำให้คนแต่ละคนมีบุคลิกภาพเป็นของตัวเอง [1] ซึ่งเป็นผลมาจากการทำงานประสานกันของสมองที่ขึ้นอยู่กับพันธุกรรม และประสบการณ์ที่ได้รับจากสิ่งแวดล้อม ซึ่งท่าทางการเคลื่อนไหวหรือท่าทางการเดินจัดได้ว่าเป็นบุคลิกภาพทางด้านกายภาพรูปแบบหนึ่ง

การเดินคือการเคลื่อนไหวของขา 2 ข้างสลับกัน (Alternate) อย่างเป็นจังหวะ (Rhythmic) และมีการเคลื่อนไหวของลำตัวเพื่อให้จุดศูนย์ถ่วง (Center of gravity) ของร่างกายเคลื่อนที่ไปด้านหน้า การเดินนั้นจะต้องเป็นลักษณะวงจร (Cycle) หรือเรียกว่า วงจรของการเดิน (Gait cycle) ซึ่งจะครอบคลุมระยะ Stance phase คือระยะที่ขาข้างนั้นวางอยู่บนพื้น และระยะ Swing phase คือระยะที่ขาข้างนั้นยกลอยขึ้นจากพื้น ดังรูปที่ 1 ซึ่งวงจรของการเดินจะมีการเคลื่อนไหวของจุดข้อต่อต่าง ๆ ในร่างกายและจะมีรูปแบบการเคลื่อนที่ (Movement Pattern) ที่เหมือนกันในทุกรอบของวงจรของการเดิน ทำให้สามารถใช้รูปแบบท่าทางการเดินบ่งบอกถึงบุคลิกภาพเฉพาะของบุคคลจากท่าทางการเดินได้



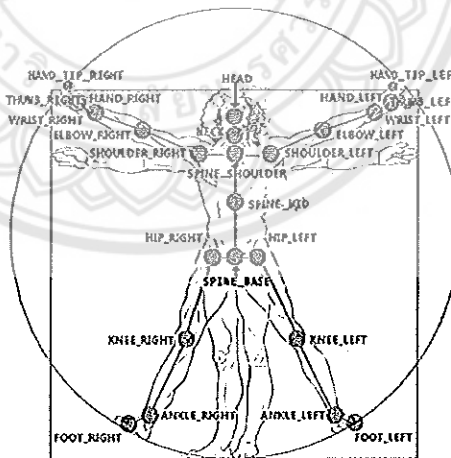
รูปที่ 1 วงจรของการเดิน (Gait cycle) [1]

## 2.2 อุปกรณ์ตรวจจับการเคลื่อนไหว (Microsoft Kinect Sensor)



รูปที่ 2 Microsoft Kinect sensor v2

กล้อง Kinect เป็นอุปกรณ์ในการบันทึกภาพเชิงลึก (Depth camera) ถูกพัฒนาโดยบริษัทไมโครซอฟต์ และมี Library functions ที่ช่วยในการแยกแยะโครงสร้างร่างกายของมนุษย์และสามารถค้นหาจุดต่าง ๆ บนร่างกาย (Joint) ได้มากถึง 25 จุด (ดังแสดงในรูปที่ 3) แล้วสามารถนำไปสร้างเป็นโครงร่างจำลอง (Skeleton) ในการตรวจจับการเคลื่อนไหวของร่างกาย (Motion tracking) ซึ่งสามารถช่วยลดเวลาในการหาวิธีการตรวจจับโครงสร้างของร่างกายและการตรวจจับการเคลื่อนไหว [2, 3] ส่วนประกอบที่สำคัญของกล้อง Kinect คือ 1) 3D depth sensor ซึ่งประกอบไปด้วย Infrared projector ทำหน้าที่ฉายแสงอินฟราเรดซึ่งมองด้วยตาเปล่าไม่เห็นออกมาเป็นแพทเทิร์นจุด และมี Infrared camera ที่ทำหน้าที่รับแสงอินฟราเรดที่ถูกฉายออกไปเพื่อใช้ในการสร้างภาพที่บอกความลึกต้นของวัตถุ (Depth map) 2) RGB camera เป็นกล้องถ่ายภาพเคลื่อนไหวใช้รับข้อมูลสี และช่วยในระบบจดจำใบหน้า (Face recognition) และ 3) Multi-array microphone มีหน้าที่ใช้รับเสียงและมีส่วนช่วยในการระบุตำแหน่งของผู้ใช้ว่าอยู่ที่ไหน มีไมโครโฟนทั้งหมดจำนวน 4 ตัว สามารถใช้แยกแยะเสียงของผู้เล่นและเสียงรบกวนภายนอกออกจากกันได้

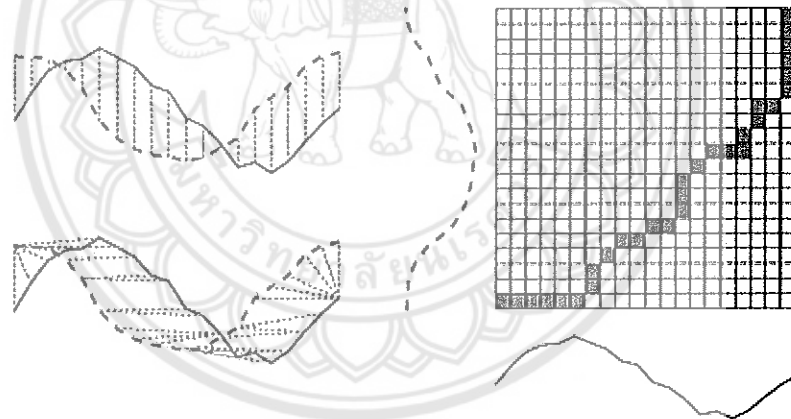


รูปที่ 3 โครงร่างจำลองที่ได้จาก Kinect sensor จำนวน 25 จุด (Microsoft SDK Skeleton) [4]

ข้อดีของ Kinect คือเป็นอุปกรณ์ที่มีราคาไม่แพงเมื่อเปรียบเทียบกับอุปกรณ์ตรวจจับค่าความเคลื่อนไหวแบบอื่น พร้อมทั้งสามารถนำไปใช้งานได้อย่างสะดวกและให้ความแม่นยำในการตรวจจับค่าความเคลื่อนไหวของร่างกายในระนาบ 3 มิติ [5] งานวิจัยนี้จึงได้นำกล้อง Kinect มาบันทึกข้อมูลท่าทางการเคลื่อนไหวของท่าทางการเคลื่อนไหวของแต่ละบุคคล เพื่อนำข้อมูลไปวิเคราะห์ในขั้นตอนต่อไป

### 2.3 Incremental Dynamic Time Warping (IDTW) algorithm

อินครีเมนทัลไดนามิกไทม์วอร์ปิง (IDTW) เป็นขั้นตอนวิธีสำหรับการเปรียบเทียบความคล้าย (Similarity Measure) ของลำดับที่มีความแตกต่างกันในด้านเวลาหรือความเร็ว เช่น รูปแบบการเดินของคน ๆ หนึ่งจะถูกนับว่ามีความคล้ายไม่ว่าคน ๆ นั้นจะเดินอย่างรวดเร็ว เดินอย่างเชื่องช้า หรือแม้แต่เดินด้วยความเร่ง เมื่อพิจารณาจากผู้สังเกตเดียวกัน ซึ่งไดนามิกไทม์วอร์ปิงสามารถนำไปประยุกต์ได้กับวิดีโอ เสียง และภาพ รวมไปถึงข้อมูลต่างๆที่สามารถแปลงให้อยู่ในรูปของข้อมูลเชิงเส้นได้ [6, 10] โดย IDTW ได้ถูกพัฒนาต่อยอดมาจากวิธีการไดนามิกไทม์วอร์ปิง (DTW) [9] ผลลัพธ์จากการคำนวณด้วยไดนามิกไทม์วอร์ปิงประกอบด้วย 2 ส่วน คือ ระยะทาง (Distance) ระหว่างลำดับเวลาที่นำมาวัด และการปรับแนวระหว่างลำดับทั้งสอง ซึ่งระยะทางหมายถึงความเหมือนกันระหว่างลำดับข้อมูลทั้งสอง หากระยะทางมีค่ามากแสดงว่าลำดับข้อมูลทั้งสองมีความแตกต่างกันมาก ลำดับข้อมูลที่มีความเหมือนกันมากจะมีค่าระยะทางใกล้ค่าศูนย์ อีกส่วนคือการปรับแนว (Alignment) ซึ่งจะแสดงถึงการจับคู่กันของแต่ละจุดข้อมูลที่มีตำแหน่งสอดคล้องกันซึ่งเป็นกุญแจสำคัญในการประมวลผล ดังแสดงในรูปที่ 4



รูปที่ 4 วิธีการเปรียบเทียบลำดับข้อมูลและการปรับแนวของไดนามิกไทม์วอร์ปิง

### การทบทวนวรรณกรรม/สารสนเทศ (Information) ที่เกี่ยวข้อง

เทคโนโลยีการวิเคราะห์คอนเทนต์ของสื่อมัลติมีเดียเพื่อสร้างครรชนหรือคำอธิบายไฟล์มัลติมีเดีย มีประโยชน์แก่การสืบค้นข้อมูลแบบอัตโนมัติจากฐานข้อมูลขนาดใหญ่ ซึ่งเทคโนโลยีนี้สามารถนำมาใช้ในการค้นหาบุคคล และการพิสูจน์บุคคลจากฐานข้อมูลที่ถูกเก็บไว้ในลักษณะต่างๆ เช่น Cloud datacenter เป็นต้น ซึ่งมีความวิจัยในปัจจุบันสนใจระบบการพิสูจน์บุคคลที่มีการวัดค่าคุณลักษณะเด่นทางชีววิทยาที่คัดกรองจากรูปร่างของบุคคลและท่าทางการเดิน โดยงานวิจัยของ Murray คณะ [7] ได้ใช้เทคนิคการพิสูจน์บุคคลจากท่าทางการเดิน

เนื่องจากมีงานวิจัยที่แสดงว่าบุคคลมีรอบของการเดินที่เป็นเฉพาะบุคคล ทำทางการเดินของบุคคลมีองค์ประกอบมากกว่า 20 องค์ประกอบที่สามารถรวมกันเป็นต้นฉบับเพื่อใช้ในการแยกแยะบุคคลได้ วิธีการนี้ต้องการส่วนประมวลผลวิดีโอเพื่อตรวจจับการเคลื่อนไหวของร่างกายที่แม่นยำ มีตัวอย่างงานวิจัย [8] นำภาพวิดีโอที่บันทึกจาก Depth images camera มาประมวลผลแยกสกัดส่วนต่าง ๆ ของร่างกาย เพื่อนำข้อมูลที่แสดงส่วนของร่างกายมาวิเคราะห์รูปแบบของท่าทางการเดิน ดังนั้นข้อมูลวิดีโอที่ประกอบไปด้วยการเคลื่อนไหวที่หลากหลายและมีความยากในระดับที่แตกต่างกัน กระบวนการแยกส่วนเฉพาะของท่าทางการเดินจากวิดีโอจะต้องอาศัยการคำนวณที่ซับซ้อนและสามารถเกิดความผิดพลาดได้ง่าย

จากปัญหาข้างต้นได้มีงานวิจัยของ Khan และคณะ [9] ได้นำเสนอเทคนิคการตรวจสอบท่าทางเวชศาสตร์ฟื้นฟูสำหรับผู้ป่วยที่ต้องการฟื้นฟูสมรรถภาพของร่างกาย โดยใช้อุปกรณ์ตรวจจับการเคลื่อนไหว Microsoft Kinect ในการบันทึกและแบ่งแยกสกัดส่วนของร่างกายของผู้ป่วยที่ทำการฝึกท่าต่าง ๆ ให้ถูกต้องตามหลัก และประยุกต์ใช้เทคนิคการเปรียบเทียบอินครีเมนทัลไดนามิกใหม่เวอร์ปิงในการวัดค่าความเหมือนระหว่างผู้ป่วยกับท่าทางต้นฉบับ ทำให้ผู้ส่วนสามารถทำท่าทางกายภาพบำบัดได้ด้วยตนเองและถูกต้อง

#### วัตถุประสงค์ของการวิจัย

1. เพื่อสร้างแอปพลิเคชันใหม่ให้แก่กลุ่มบริษัทผู้ให้บริการอินเทอร์เน็ตความเร็วสูง เพื่อเป็นบริการเสริมของการใช้บริการอินเทอร์เน็ตความเร็วสูงบน FTTH
2. เพื่อพัฒนาระบบใหม่สำหรับการวิเคราะห์คอนเทนต์ ทำดรรชนี และสืบค้นข้อมูลแบบทันทีทันใดจากฐานข้อมูลขนาดใหญ่บนคลาวด์ ซึ่งระบบนี้จะสามารถใช้ในการค้นหาและพิสูจน์บุคคลแบบอัตโนมัติ อย่างถูกต้องและมีประสิทธิภาพ

#### ประโยชน์ที่คาดว่าจะได้รับ

ได้ระบบใหม่สำหรับการวิเคราะห์คอนเทนต์ทำดรรชนีและสืบค้นข้อมูลแบบทันทีทันใดจากฐานข้อมูลขนาดใหญ่บนคลาวด์ ซึ่งระบบนี้จะสามารถใช้ในการค้นหาและพิสูจน์บุคคลแบบอัตโนมัติได้อย่างถูกต้องและมีประสิทธิภาพ ระบบดังกล่าวจะนำไปสร้างเป็นบริการเสริมให้แก่ผู้ให้บริการอินเทอร์เน็ตความเร็วสูง โดยจะสามารถนำระบบไปใช้กับบริษัทและผู้ประกอบการการให้บริการอินเทอร์เน็ต (ISP: Internet Service Provider)

#### ขอบเขตการวิจัย

โครงการวิจัยนี้เน้นการศึกษาลักษณะการทำงานของอัลกอริทึมการการรู้จำท่าทางเคลื่อนไหวของบุคคลกลุ่มตัวอย่าง จำนวน 37 คน และจัดทำระบบพิสูจน์บุคคลอัตโนมัติจากท่าทางการเดินที่มีความถูกต้องแม่นยำ มีประสิทธิภาพและมีประมวณผลได้อย่างรวดเร็ว เมื่อเปรียบเทียบกับวิธีแบบดั้งเดิม โดยข้อมูลที่บันทึกจากจากอุปกรณ์ตรวจจับการเคลื่อนไหว (Kinect sensor) จะถูกวิเคราะห์และเปรียบเทียบข้อมูลใช้วิธีอินครีเมนทัลไดนามิกใหม่เวอร์ปิง (IDTW) เพื่อตรวจสอบและเปรียบเทียบลำดับข้อมูลท่าทางการเดินของแต่ละบุคคล

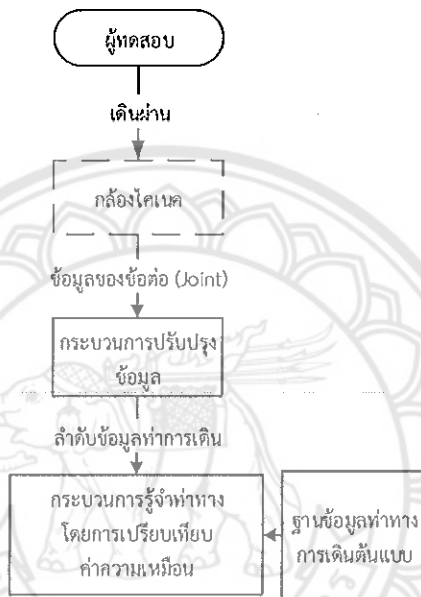
## วิธีดำเนินการวิจัย

1. ศึกษาและค้นคว้าทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง
2. ศึกษาวิธีการกำหนดแผนการทดลอง
3. เก็บข้อมูลตัวอย่าง
4. พัฒนาระบบตรวจสอบและเปรียบเทียบลำดับข้อมูลจากท่าทางการเดิน
5. ศึกษาเปรียบเทียบประสิทธิภาพของอัลกอริทึมระหว่าง DTW และ IDTW โดยพิจารณาประสิทธิภาพของผลลัพธ์และระยะเวลาที่ใช้ในการประมวลผล
6. พิจารณาความแม่นยำของระบบ
7. ประมวลผลจากโปรแกรมเพื่อเก็บรวบรวมข้อมูล และทำการวิเคราะห์ข้อมูลเพื่อหาข้อสรุป
8. สรุปผลและจัดทำบทความเพื่อตีพิมพ์เผยแพร่



วิธีดำเนินการวิจัย

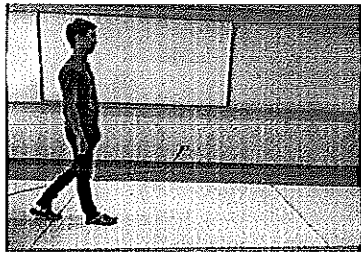
ในบทนี้ผู้วิจัยจะกล่าวถึงกระบวนการทำงานของระบบหุ่นยนต์อัตโนมัติจากท่าทางการเดิน จะมีการบันทึกข้อมูลท่าทางการเดินของผู้ทดสอบไว้ก่อนล่วงหน้าเพื่อใช้สำหรับเป็นฐานข้อมูลในการเปรียบเทียบท่าทางการเดิน เมื่อมีบุคคลเดินผ่านอุปกรณ์ตรวจจับความเคลื่อนไหว 3 มิติ (Kinect sensor) ระบบจะนำข้อมูลที่บันทึกได้เข้าสู่กระบวนการปรับปรุงข้อมูลเพื่อให้ได้ข้อมูลที่เป็นมาตรฐานเดียวกัน จากนั้นจะทำการวิเคราะห์และประมวลผลข้อมูลที่รับเข้ามาใหม่กับข้อมูลที่ถูกบันทึกไว้ก่อนหน้าในฐานข้อมูล เพื่อตรวจสอบและเปรียบเทียบข้อมูลคุณลักษณะเด่นทางชีววิทยาของแต่ละบุคคล ดังแสดงการทำงานในรูปที่ 5



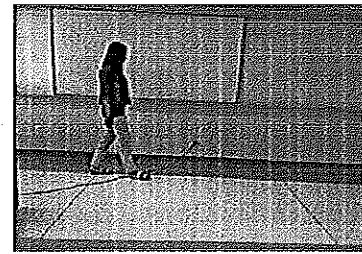
รูปที่ 5 ขั้นตอนการทำงาน

3.1 การบันทึกข้อมูลท่าทางการเดิน (Movement recording)

งานวิจัยนี้ได้ทำการทดลองเก็บข้อมูลท่าทางการเดินของบุคคล จำนวน 37 คน โดยแต่ละคนจะถูกบันทึกท่าทางการเดินจากระยะที่ห่างจากกล้อง Kinect ใน 2 ระดับคือ ระยะใกล้ 2 เมตร (Near record) ดังแสดงในภาพที่ 6(a) และระยะไกล 3 เมตร (Far record) ดังแสดงในภาพที่ 6(b) เพื่อใช้สำหรับเป็นฐานข้อมูลในการเปรียบเทียบ

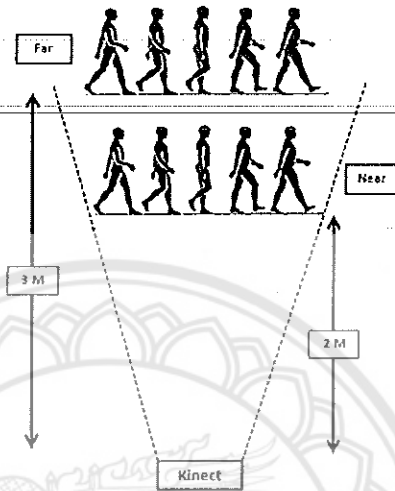


(a) ตัวอย่างเฟรมการบันทึกที่ระยะใกล้ (Near record)



(b) ตัวอย่างเฟรมการบันทึกที่ระยะไกล (Far record)

รูปที่ 6 ตัวอย่างภาพการบันทึกข้อมูลท่าทางการเดินของบุคคล



รูปที่ 7 The experimental setting

ลำดับข้อมูลที่ถูกบันทึกได้จาก Kinect จะมีลักษณะรูปแบบข้อมูลเป็น 3 มิติ กล่าวคือ ณ ตำแหน่งใด ๆ ของลำดับข้อมูลจะประกอบไปด้วยตำแหน่งของค่า  $(x, y, z)$  ของแต่ละจุดข้อต่อ (Joint) รวมกัน 26 จุด ดังในภาพที่ 3 ดังนั้นจะได้ลำดับของข้อมูลดังต่อไปนี้

$$\begin{aligned}
 \text{Sequence}(t_0) &= \begin{bmatrix} x_0 & x_1 & \dots & x_{25} \\ y_0 & y_1 & \dots & y_{25} \\ z_0 & z_1 & \dots & z_{25} \end{bmatrix} \\
 \text{Sequence}(t_1) &= \begin{bmatrix} x_0 & x_1 & \dots & x_{25} \\ y_0 & y_1 & \dots & y_{25} \\ z_0 & z_1 & \dots & z_{25} \end{bmatrix} \\
 &\vdots \\
 \text{Sequence}(t_N) &= \begin{bmatrix} x_0 & x_1 & \dots & x_{25} \\ y_0 & y_1 & \dots & y_{25} \\ z_0 & z_1 & \dots & z_{25} \end{bmatrix}
 \end{aligned}$$

ดังนั้นจะได้ชุดข้อมูล (Dataset) ของลำดับท่าทางการเดินของบุคคลเป็น

$$S = \{X_0, X_1, \dots, X_{25}\}$$

เมื่อ  $S$  คือลำดับของข้อมูลท่าทางการเดิน  
 $t$  คือเวลาใด ๆ  
 $X$  คือตำแหน่งของจุดข้อต่อ ( $X_i = [x_i, y_i, z_i]$ )

### 3.2 กระบวนการปรับปรุงข้อมูล (Data Normalization)

ข้อมูลที่ได้มาจาก Kinect นั้นไม่สามารถนำมาใช้เปรียบเทียบได้โดยตรง เนื่องจากระยะห่างของจุดตามร่างกายแต่ละคนที่บันทึกมานั้นมีอัตราส่วนที่ไม่เท่ากัน ทำให้เวลานำข้อมูลไปเปรียบเทียบกันอาจทำให้เกิดความผิดพลาดได้ ดังนั้นจึงจำเป็นต้องมีการปรับข้อมูลให้อยู่ในอัตราส่วนเดียวกันก่อน โดยกำหนดให้ตำแหน่งที่ Spine base ในรูปที่ 3 เป็นจุดเริ่มต้น จากนั้นนำข้อต่อจากทุกจุดมาหาผลต่างด้วยค่าของตำแหน่งของ Spine base [9] ดังสมการที่ 1

$$S'_i = \frac{S_i - S_0}{\|S_i - S_0\|^2}, i = 1, \dots, 25 \quad (1)$$

เมื่อ  $S'_i$  คือตำแหน่งของจุดข้อต่อตามร่างกายที่ผ่านการกระบวนการปรับปรุงข้อมูลแล้ว  
 $S_0$  คือค่าของตำแหน่งข้อต่อที่จุด Spine base  
 $S_i$  คือค่าของตำแหน่งข้อต่อที่จุดอื่น ๆ  
 $i$  คือหมายเลขของจุดข้อต่อทั้งหมด

### 3.3 กระบวนการรู้จำ (Recognition process)

งานวิจัยนี้ได้ใช้เทคนิคอินครีเมนทัลไดนามิกไทม์วอร์ปิง (IDTW) ในการเปรียบเทียบข้อมูลลำดับของท่าทางการเดิน ซึ่งหลักการทำงานในขั้นแรกทำการหาจุดเริ่มต้นและจุดสิ้นสุดของลำดับของท่าทางการเดินคือ การหาเฟรมที่เป็นเฟรมเริ่มต้นของท่าทางและเฟรมที่สิ้นสุดของท่าทาง จากนั้นนำลำดับมาเปรียบเทียบกับลำดับของท่าทางการเดินต้นแบบที่ได้บันทึกไว้ก่อนหน้านี้เพื่อหาค่าความแตกต่าง โดยการเปรียบเทียบลำดับของข้อมูลที่มีค่าคะแนนความแตกต่างน้อยที่สุดก็แสดงว่าลำดับข้อมูลนั้นมีความคล้ายกัน ซึ่งมีขั้นตอนการทำงานดังต่อไปนี้

ขั้นตอนที่ 1 กำหนดค่าเริ่มต้นดังต่อไปนี้

$U$  คือ ลำดับข้อมูลท่าทางการเดิน ณ เวลาปัจจุบัน มีขนาดเท่ากับ  $N$

$E$  คือ ลำดับข้อมูลท่าทางการเดินที่ถูกบันทึกเป็นต้นแบบ มีขนาดเท่ากับ  $M$

$G$  คือ ตารางเมทริกซ์ขนาด  $M \times N$

$V$  คือ เฟรมถัดไปของลำดับข้อมูลท่าทางการเดิน  $U$

ขั้นตอนที่ 2 เพิ่มเฟรม :  $Q = N + 1$  และ  $U_Q = V$  เมื่อ  $U_Q$  คือเฟรมที่  $Q$  ของลำดับข้อมูล  $U$

ขั้นตอนที่ 3 คำนวณระยะห่างระหว่าง  $U$  กับ  $E^j, j = 1, \dots, M$  โดยที่ทุก  $E^j$  คือลำดับข้อมูล  $E$  ณ เฟรมที่  $j$

ขั้นตอนที่ 4 ค่า Minimum คำนวณจากสมการ (2)

$$D_{IDTW}(U, E) = \min_{j=1, \dots, M} D(U, E^j) \quad (2)$$

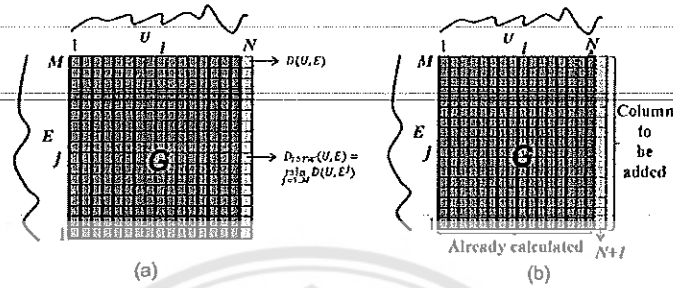


ขั้นตอนที่ 5 ท้าวานซ้ำในขั้นตอนที่ 2 ถึง 4 สำหรับทุกลำดับข้อมูลใหม่ และปรับค่าระยะห่าง IDTW ที่คำนวณได้แล้วเก็บ

ค่าไว้ในตารางเมทริกซ์  $G$  จำนวนจากสมการ (2)

$$G = \min(G(1 \dots M, Q)) / Q \quad (3)$$

เมื่อพิจารณาจากขั้นตอนการทำงานจะเห็นได้ว่าเฟรมถัดไปของลำดับข้อมูลจะถูกคำนวณ เมื่อเพิ่มคอลัมน์ในตาราง เมทริกซ์  $G$  สำหรับเก็บค่าระยะห่างที่ปรับปรุงใหม่ทำให้สามารถคำนวณหาค่า Minimum ได้อย่างรวดเร็ว [9] ดังแสดงตัวอย่าง ตารางเมทริกซ์  $G$  ในรูปที่ 8

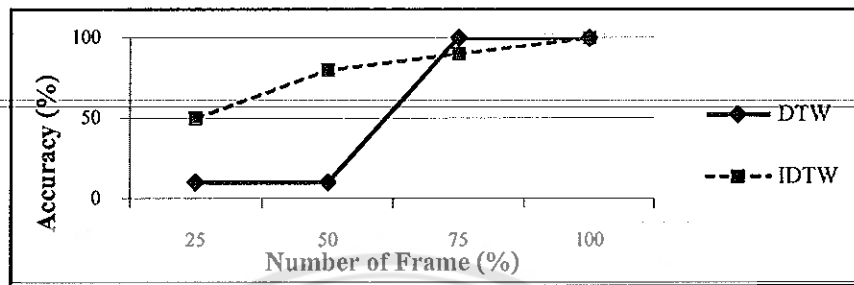


รูปที่ 8 (a) ตารางเมทริกซ์  $G$  เปรียบเทียบระหว่างการคำนวณค่า DTW และ IDTW

(b) การเพิ่มคอลัมน์ในตาราง เมทริกซ์  $G$  ค่าระยะห่างของ IDTW

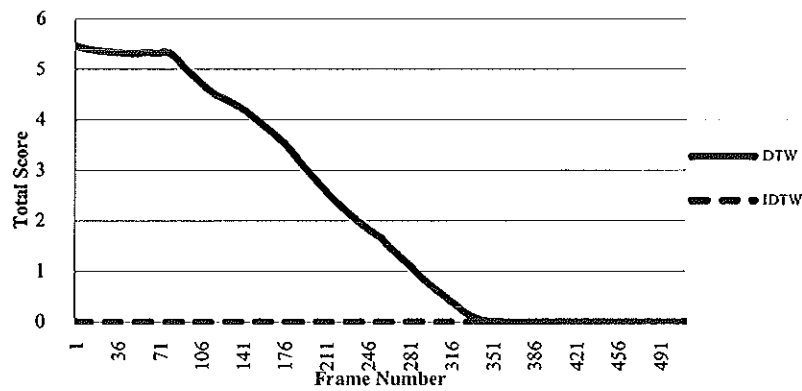


เนื้อหาในบทนี้ จะกล่าวถึงผลการทดสอบประสิทธิภาพในการทำงานของการรู้จำท่าทางการเดินของบุคคล โดยเปรียบเทียบระหว่างวิธีการ ไดนามิกไทม์วอร์ปปีง (DTW) และวิธีการอินครีเมนต์ไดนามิกไทม์วอร์ปปีง (IDTW) และทดสอบความถูกต้องของผลลัพธ์ที่ได้จากการรู้จำท่าทางการเดินของบุคคลด้วยวิธีการอินครีเมนต์ไดนามิกไทม์วอร์ปปีง (IDTW) ซึ่งได้ผลการทดลองดังนี้



รูปที่ 9 เปรียบเทียบประสิทธิภาพในการทำงานระหว่าง DTW และ IDTW [11]

จากรูปที่ 9 แสดงการเปรียบเทียบประสิทธิภาพในการทำงานระหว่าง DTW และ IDTW โดยประเมินจากค่าความถูกต้องของการรู้จำท่าทางการเดินกับจำนวนเฟรมที่ใช้ในการประมวลผล จากการทดลองพบว่าทั้ง 2 วิธีการสามารถให้ผลลัพธ์การรู้จำท่าทางการเดินได้อย่างถูกต้องร้อยละ 100 เมื่อประมวลผลจำนวนเฟรมของลำดับข้อมูลครบทั้งหมด แต่วิธีการ IDTW สามารถให้จำนวนผลลัพธ์การรู้จำท่าทางการเดินได้รวดเร็วกว่าวิธีการ DTW จากรูปที่ 9 จะเห็นได้ว่าเมื่อมีการประมวลผลจำนวนเฟรมเพียงร้อยละ 25 ของจำนวนเฟรมทั้งหมด วิธีการ IDTW สามารถให้ผลลัพธ์การรู้จำท่าทางการเดินได้มากกว่าร้อยละ 50 และเพิ่มสูงขึ้นเรื่อย ๆ แต่วิธีการ DTW จำเป็นมีการประมวลผลจำนวนเฟรมถึงมากกว่าร้อยละ 50 ของจำนวนเฟรมทั้งหมด จึงจะสามารถให้ผลลัพธ์การรู้จำท่าทางการเดินได้ดี และเมื่อทำการเปรียบเทียบผลค่าระยะห่างของลำดับข้อมูลระหว่าง DTW และ IDTW นั้นพบว่าวิธีการ IDTW สามารถคำนวณหาค่าระยะห่างของลำดับข้อมูลที่เหมือนกันได้ตั้งแต่จำนวนเฟรมแรก ๆ ของลำดับข้อมูล ซึ่งวิธีการ DTW นั้นสามารถคำนวณหาเหมือนกันของลำดับข้อมูลได้ในช่วงท้ายเฟรมของลำดับข้อมูล นั้นหมายความว่า วิธีการ IDTW สามารถหาคำตอบในการรู้จำท่าทางการเดินได้รวดเร็วกว่าวิธีการ DTW นั้นเอง ดังแสดงในรูปที่ 10



รูปที่ 10 เปรียบเทียบการคำนวณหาค่าระยะห่างของลำดับข้อมูลระหว่าง DTW และ IDTW [11]

สำหรับทดสอบความถูกต้องของผลลัพธ์ที่ได้จากการรู้จำท่าทางการเดินของบุคคลด้วยวิธีการอินคริเมนต์

ไดนามิกใหม่วอร์ปิง (IDTW) จากผู้ทดสอบจำนวน 37 คน จากผลในตารางที่ 1 พบว่าระบบสามารถการรู้จำท่าทางการเดินของบุคคลที่บันทึกในระยะไกล (Far record) ได้ถูกต้องถึง 24 คน จากผู้สอบทั้งหมด 37 คน หรือคิดเป็นร้อยละ 65 แต่ถ้าหากพิจารณาผลของคำตอบ 5 คำตอบแรกที่มีค่าระยะห่างน้อยที่สุด พบว่าระบบสามารถการรู้จำท่าทางการเดินของบุคคลที่บันทึกในระยะไกลได้ถูกต้องมากขึ้นเป็น 35 คน จากผู้สอบทั้งหมด 37 คน หรือคิดเป็นร้อยละ 95 ส่วนผลการรู้จำท่าทางการเดินของบุคคลที่บันทึกในระยะใกล้ (Near record) มีความถูกต้องถึง 27 คน จากผู้สอบทั้งหมด 37 คน หรือคิดเป็นร้อยละ 73 แต่ถ้าหากพิจารณาผลของคำตอบ 5 คำตอบแรกที่มีค่าระยะห่างน้อยที่สุด พบว่าระบบสามารถการรู้จำท่าทางการเดินของบุคคลที่บันทึกในระยะใกล้ได้ถูกต้องมากขึ้นเป็น 34 คน จากผู้สอบทั้งหมด 37 คน หรือคิดเป็นร้อยละ 92

จะเห็นได้ว่าเมื่อมีการพิจารณาผลของคำตอบ 5 คำตอบแรก (Top 5) ที่มีค่าระยะห่างน้อยที่สุด ทั้งการบันทึกในระยะใกล้และบันทึกในระยะไกล ส่งผลให้ความถูกต้องของผลลัพธ์ที่ได้จากการรู้จำเพิ่มขึ้น ผู้วิจัยจึงได้นำผลลัพธ์จากทั้งการบันทึกในระยะใกล้และบันทึกในระยะไกลมาหาค่าเฉลี่ย [12, 13] โดยคำนวณจากสมการ (4) ส่งผลให้ความถูกต้องของผลลัพธ์ที่ได้จากการรู้จำท่าทางการเดินเพิ่มขึ้นเป็น 29 จากผู้สอบทั้งหมด 37 คน หรือคิดเป็นร้อยละ 78 และพิจารณาผลของคำตอบ 5 คำตอบแรกที่มีค่าระยะห่างน้อยที่สุดเป็น 36 คน จากผู้สอบทั้งหมด 37 คน หรือคิดเป็นร้อยละ 97 ดังแสดงในตารางที่ 1

$$IDTW_{Best\_Match} = \operatorname{argmin}_{k=1, \dots, K} \left( \frac{IDTW_{near_k} + IDTW_{far_k}}{2} \right) \quad (4)$$

- เมื่อ  $IDTW_{near}$  คือค่า IDTW สำหรับการบันทึกในระยะใกล้  
 $IDTW_{far}$  คือค่า IDTW สำหรับการบันทึกในระยะไกล  
 $K$  คือจำนวนไฟล์ทั้งหมด

ตารางที่ 1 ผลการรู้จำท่าทางการเดินของแต่ละบุคคลด้วยวิธีการอินครีเมนที่ไดนามิกใหม่เวอร์บปีง (IDTW)

ลำดับ	Referen ce	Result (Far)		Result (Near)		Result (Average)	
		Top 1	Top 5	Top 1	Top 5	Top 1	Top 5
1	ก	✓	✓	✓	✓	✓	✓
2	ข	✓	✓	✓	✓	✓	✓
3	ค	✓	✓	✓	✓	✓	✓
4	ง	✗	✓	✗	✓	✗	✓
5	จ	✓	✓	✓	✓	✓	✓
6	ฉ	✗	✓	✗	✓	✗	✓
7	ช	✗	✓	✓	✓	✓	✓
8	ซ	✓	✓	✗	✓	✓	✓
9	ฌ	✓	✓	✓	✓	✓	✓
10	ญ	✓	✓	✓	✓	✓	✓
11	ฎ	✓	✓	✓	✓	✓	✓
12	ฏ	✗	✗	✗	✗	✗	✗
13	ฐ	✗	✓	✓	✓	✓	✓
14	ฑ	✓	✓	✓	✓	✓	✓
15	ฒ	✗	✓	✓	✓	✓	✓
16	ณ	✓	✓	✓	✓	✓	✓
17	ด	✓	✓	✗	✓	✓	✓
18	ต	✗	✗	✗	✓	✗	✓
19	ถ	✗	✗	✓	✓	✓	✓
20	ท	✓	✓	✓	✓	✓	✓
21	ธ	✓	✓	✗	✓	✓	✓
22	น	✓	✓	✓	✓	✓	✓
23	บ	✓	✓	✓	✓	✓	✓
24	ป	✗	✗	✓	✓	✗	✓
25	ผ	✓	✓	✓	✓	✓	✓
26	ฝ	✗	✗	✓	✓	✗	✓
27	พ	✓	✓	✓	✓	✓	✓
28	ฟ	✓	✓	✓	✓	✓	✓
29	ภ	✓	✓	✓	✓	✓	✓
30	ม	✗	✓	✓	✓	✓	✓
31	ย	✓	✓	✗	✓	✓	✓
32	ร	✓	✓	✓	✓	✓	✓
33	ล	✗	✓	✗	✗	✗	✓
34	ว	✓	✓	✓	✓	✓	✓

ลำดับ	Reference	Result (Far)		Result (Near)		Result (Average)	
		Top 1	Top 5	Top 1	Top 5	Top 1	Top 5
35	ศ	x	✓	x	x	x	✓
36	ช	✓	✓	✓	✓	✓	✓
37	ส	✓	✓	✓	✓	✓	✓
Accuracy		24/37	35/37	27/37	34/37	29/37	36/37
		65%	95%	73%	92%	78%	97%



## บทที่ 5

### สรุปผลการวิจัย

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อนำเสนอวิธีการพิสูจน์บุคคลอัตโนมัติจากท่าทางการเดิน จากนั้นวัดประสิทธิภาพของวิธีการที่นำเสนอโดยพิจารณาประสิทธิภาพในการพยากรณ์รู้จำบุคคลและระยะเวลาที่ใช้ในการประมวลผล เพื่อหาข้อสรุปและข้อเสนอแนะในการเพิ่มประสิทธิภาพของอัลกอริทึม IDTW ในการสร้างระบบพิสูจน์บุคคลอัตโนมัติจากท่าทางการเดินที่เหมาะสมให้มีประสิทธิภาพมากยิ่งขึ้น

#### สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้ได้นำเสนอวิธีการพิสูจน์บุคคลอัตโนมัติจากท่าทางการเดิน โดยประยุกต์ใช้วิธีการเปรียบเทียบความคล้ายของลำดับที่มีความแตกต่างกันในด้านเวลาหรือความเร็วที่เรียกว่า Incremental Dynamic Time Warping (IDTW) เพื่อตรวจสอบและเปรียบเทียบข้อมูลคุณลักษณะเด่นทางชีววิทยาของแต่ละบุคคล ที่บันทึกได้จากอุปกรณ์ตรวจจับความเคลื่อนไหว 3 มิติ (Microsoft's Kinect<sup>®</sup>) จากผลการทดลองพบว่าวิธีการ IDTW สามารถหาคำตอบในการรู้จำท่าทางการเดินได้รวดเร็วกว่าวิธีการ DTW เนื่องจากผลของคำตอบที่ได้มีการลู่เข้าหาผลลัพธ์ได้ตั้งแต่จำนวนเฟรมช่วงแรก ๆ ของลำดับข้อมูล โดยไม่ต้องประมวลผลครบจำนวนเฟรมของลำดับข้อมูล และเมื่อพิจารณาความถูกต้องของผลลัพธ์ที่ได้จากการรู้จำท่าทางการเดินของบุคคลด้วยวิธีการ IDTW ที่นำเสนอใหม่มีค่าสูงถึงร้อยละ 78 และหากพิจารณาผลของคำตอบ 5 คำตอบแรกที่มีค่าระยะห่างน้อยสุด พบว่าค่าความถูกต้องของผลลัพธ์สูงถึงร้อยละ 97 ซึ่งสามารถนำเทคนิควิธีนี้ไปประยุกต์ใช้กับระบบพิสูจน์บุคคลอัตโนมัติได้อย่างมีประสิทธิภาพ และสามารถนำไปประยุกต์ใช้สร้างเป็นบริการเสริมให้แก่ผู้ใช้บริการอินเทอร์เน็ตความเร็วสูง งานวิจัยสามารถเพิ่มตัวอย่างที่ใช้ในการทดสอบในท่าทางการเดินให้มีความหลากหลายมากยิ่งขึ้นเพื่อความเชื่อมั่นและความแม่นยำในการรู้จำท่าทางการเดินต่อไป

## บรรณานุกรม

1. Neumann, Donald A. *Kinesiology of the Musculoskeletal System-E-Book: Foundations for Rehabilitation*. Elsevier Health Sciences, 2013.
2. Preis, Johannes, et al. "Gait recognition with kinect." 1st international workshop on kinect in pervasive computing. New Castle, UK, 2012.
3. Patsadu, Orasa, Chakarida Nukoolkit, and Bunthit Watanapa. "Human gesture recognition using Kinect camera." In *Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference on*, pp. 28-32. IEEE, 2012.
4. Microsoft Developer Network. JointType Enumeration [Online]. Available at <https://msdn.microsoft.com>, Accessed Sep-18, 2017.
5. Munsell, Brent C., Andrew Temlyakov, Chengzheng Qu, and Song Wang. "Person identification using full-body motion and anthropometric biometrics from kinect videos." In *Computer Vision-ECCV 2012. Workshops and Demonstrations*, pp. 91-100. Springer Berlin Heidelberg, 2012.
6. Górecki, Tomasz, and Maciej Łuczak. "Multivariate time series classification with parametric derivative dynamic time warping." *Expert Systems with Applications* 42.5 (2015): 2305-2312.
7. Murray, M. Pat, A. Bernard Drought, and Ross C. Kory. "Walking patterns of normal men." *JBJS* 46.2 (1964): 335-360.
8. Sivapalan, Sabesan, et al. "Gait energy volumes and frontal gait recognition using depth images." *Biometrics (UCB), 2011 International Joint Conference on*. IEEE, 2011.
9. Khan, Naimul Mefraz, et al. "A visual evaluation framework for in-home physical rehabilitation." *Multimedia (ISM), 2014 IEEE International Symposium on*. IEEE, 2014.
10. Reyes, Miguel, Gabriel Domínguez, and Sergio Escalera. "Feature weighting in dynamic time warping for gesture recognition in depth data." *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011.
11. Wisrut Kwankhoom and Paisarn Muneesawang. "Early Recognition of Personal Identification Based on IDTW via the FTTH network". *National Conference on Knowledge and Digital Society*, 2015, p.334-345.
12. Kwankhoom, Wisrut, and Paisarn Muneesawang. "Person Re-identification Using 3D Data Analysis Method and Kinect Sensor." *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* 9.2-5 (2017): 151-154.
13. Kwankhoom, Wisrut, and Paisarn Muneesawang. "An Incremental Dynamic Time Warping for person re-identification." *Computer Science and Software Engineering (JCSSE), 2017 14th International Joint Conference on*. IEEE, 2017.

## ภาคผนวก

---

1. บทความวิจัยที่ได้รับการตอบรับตีพิมพ์ในวารสารระดับนานาชาติที่มีค่า Impact Factor
2. บทความวิจัยที่นำเสนอในการประชุมวิชาการในระดับนานาชาติ ที่มีการตีพิมพ์บทความบน Proceedings





# Person Re-identification Using 3D Data Analysis Method and Kinect Sensor

Wisrut Kwankhoom, Paisarn Muneesawang

Department of Electrical and Computer Engineering, Faculty of Engineering,  
Naresuan University, Phitsanulok 65000, Thailand.

paisarnmu@nu.ac.th

**Abstract**—Automated personal identification systems, such as personal facial recognition systems and automated motor vehicle registration number checking, are examples of public security protection systems. The area of personal identification for security purposes is of growing interest for security assessment in public places, and airports, as examples, now becoming an imperative matter for research in the Internet netscape. We propose a method of immediate recognition of a subject person, based on Incremental Dynamic Time Warping (IDTW) which identifies personal gait patterns recorded via a 3D depth sensing camera such as in Microsoft's Kinect® version 2, by analyzing a dataset of gait gestures derived from a sample of 16 people. The experimental results show that the IDTW algorithm increases the efficiency of recognizing at 81%.

**Index Terms**—Personal Identification; Gesture Recognition; IDTW; Kinect Camera; Computer Visions.

## I. INTRODUCTION

Multimedia analysis technology for indexing or explaining multimedia files provides significant advantages for information retrieval from huge databases. This technology can be applied for personal identification based on a video database. The science of personal physical behavior pattern searching and indexing is based on the biological features of a subject person that are able to be filtered from multimedia files (image or video) recording imagery of that person. Specific physical aspects of the person are analyzed and are measured and analyzed by a computer based on standard, generic data of biological features for recognition and identification of the person.

Gesture recognition is an important part of communication between humans and computers. Interactive human body movement and gesture tracking is the basis of many applications including gaming, human-computer interaction, telepresence, health-care and security [1]. Generally, the recognition result is acquired after the movement or gesture has been detected and completed.

In this paper, we propose an analysis method for recognition of gait gestures and its application to person identification.

In the person identification process, the skeletal joints of the person can be analyzed to solve the limited problem of face recognition by using the biological features of the gait gestures. There has also been research into the analysis of the human personal cycle of gait gestures [2,3]. Gait gesture analysis requires more than twenty components to identify a person, and uses the kinematics of the person; movement and rhythm of walking, for example. Compared with analyzing the human face, using gait gesture recognition is the more

effective method for person identification. Person re-identification has to analyze joint of skeleton

However, the gait gesture recognition methods need video images to track the motion correctly and a segmentation algorithm to separate the human body image area from the overall media image. The human body data area is the only part of the image used to analyze gait gestures [4] and therefore must be isolated from the overall image. This is difficult due to the movement of the image in

the video. Segmenting a moving gesture from the video requires a high degree of computation and can easily be mistaken.

Given the problems of analyzing images and videos, researchers have found that using a 3-dimensional data sensor to record human gait gesture is the preferred option. In 2010, Microsoft Company produced a 3D depth sensing camera called the 'Kinect sensor' which uses X-Box 360 videogame. This player can act as a remote control for gaming by using the gesture of the players. The camera has a sensor and an application programming interface (API) for searching and representing the body as a skeletal image. It identifies the body joints of a player who is standing in front of the sensor. The Kinect camera has the great advantage of being inexpensive but easy to use and gives high accuracy of detection of the motion, in 3D [5].

This paper presents a method for person identification using gait gestures. The computational method uses the anthropometric biological features in the 3D data, captured by the Kinect 3D camera, and analyzes that data.

The method of gait pattern recognition that we propose is based on the Incremental Dynamic Time Warping (IDTW) algorithm. This an extension of the classic Dynamic Time Warping (DTW) algorithm by providing an accurate comparison between the incoming, incomplete data and the complete sample data already available in the image database.

The preliminary results of our work indicate that the proposed system is capable of successful person identification when incomplete data is available.

## II. RESEARCH METHODOLOGY

This section describes the process of recording the human gait patterns to enable the creation of the skeletal models of the sample people, to compute the feature vectors, and to measure similarity, as shown in Figure 1.

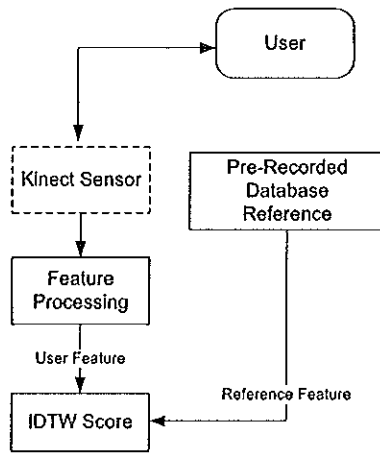
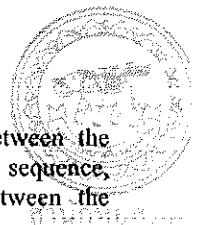


Figure 1: Block diagram of the proposed method for person re-identification

A. Kinect® Camera version 2

The Kinect camera is a human motion tracking peripheral for the Xbox 360 console from Microsoft. It was the first applied in gaming systems but it has been found to have many other applications and uses, such as human motion and features recognition, 3D model reconstruction, robot navigation, medical applications and dance training [6-8].

Kinect version 2 can extract many aspects of human body poses from the 3D depth images, representing the body as a skeletal image with identifiable body joints. The positions being tracked and recorded are more anatomically correct and stable the longer the activity proceeds, given that more video data is being recorded. Overall, twenty-five body joints can be represented, as shown in Figure 2.

From the Skeleton representation shown in Figure 2, the Spine base joint is used as the origin of the skeleton and the relative positions of the other joints are calculated from this joint.

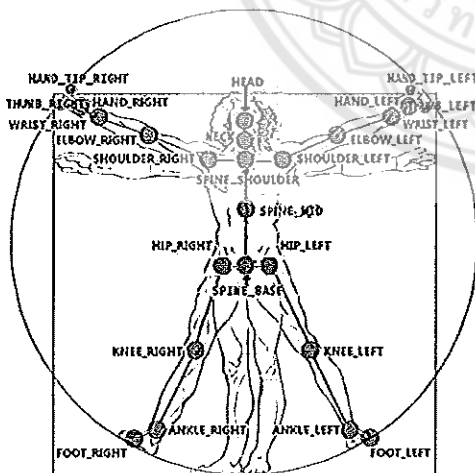


Figure 2: Microsoft SDK Skeleton representation [9]

B. Incremental Dynamic Time Warping (IDTW)

IDTW is an algorithm for comparing video image sequences taken at varying speeds and over different time periods by compressing or expanding them in the time domain. This is an extension of the DTW as described in [10]. We used the stored complete data of skeletal features as the reference sequence, as the benchmark, and did a real-time comparison between the new sequence with the reference sequence data. The algorithm calculates a distance score

between sequences, providing a comparison between the incomplete sequence and the reference complete sequence, thereby indicating the degree of sameness between the images.

IDTW algorithm works by comparing each sequence with the best starting segment of reference. It terminates with the result after applying the full (reference) sequence with all possible frames, computing the DTW distances for every possible comparison and reporting the minimum distance as follows [7]:

- I. Initialize parameters:  $U$  is the user sequence up to current time (length  $N$ ),  $E$  is the reference full sequence (length  $M$ ),  $G$  is the matrix ( $M \times N$ ) cumulative cost matrix up to current time, and  $V$  is next frame in user sequence.
- II. Inserting a new frame:  $Q = N + 1$  and  $U_Q = V$ , where  $U_Q$  is the  $Q$ -th frame of  $U$ .
- III. The DTW distance is calculated for all  $E^j, j = 1, \dots, M$  where each  $E^j$  is the reference sequence  $E$  truncated at  $j^{\text{th}}$  frame, i.e.,  $D(U, E^j), j = 1, \dots, M$  is the DTW distance obtained by the optimal warping path.
- IV. The minimum is found using the following formula:

$$D_{IDTW}(U, E) = \min_{j=1, \dots, M} D(U, E^j) \quad (1)$$

- V. Step II and IV are iterated for each the new sequence and updated IDTW distance (the cost matrix  $G$ ) by using the following formula:

$$G = \min(G(1 \dots M, Q)) / Q \quad (2)$$

Note that the components of DTW from previous states are reused as the new sequence movement advances, to optimize computation time.

Considering the steps, we notice that the next frame of the user sequence will be input to the process. Then just one column is added to the cumulative cost matrix  $G$ . This column is filled in line with a chosen updated rule in a similar way to classic DTW. Finally, the normalized minimum value from the newly added column of the cost matrix  $G$  comes out as the current IDTW distance.

III. EXPERIMENTAL RESULTS

We obtained a dataset of gait gestures from 16 people by each person walking at a different distance from the Kinect sensor, at the Near Recording (R1) distance of two meters and the Far Recording (R2) distance of three meters, and another record distance between Near Recording and Far Recording for testing data (R3) as shown in Figure 3.

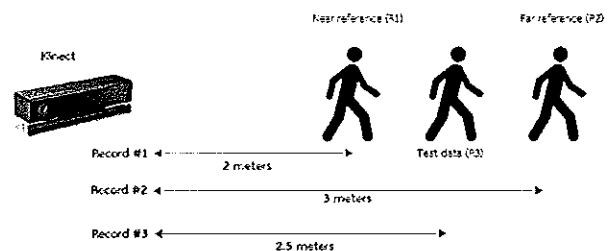


Figure 3: The experimental setting

In this experiment, we selected the twenty joints shown in Figure 2, as a way to evaluate the accuracy of the program. They are SpineMid, Neck, Head, LeftShoulder, LeftElbow, LeftWrist, LeftHand, RightShoulder, RightElbow, RightWrist, RightH, HipLeft, KneeLeft, AnkleLeft, FootLeft, HipRight, KneeRight, AnkleRight, FootRight, and SpineShoulder. Since the movement of each gesture is strongly related to these joints, the application of this image data can classify person gait gestures effectively.

The experimental results as shown in Table 1 which shows the results of the IDTW algorithm in recognizing the between Near record and Far record. For each query instance, the IDTW scores were calculated and nearest neighbor search was applied to select the best match.

Table 1  
Recognition results of Near record and Far record

No.	Ref.	Result (Far)	Result (Near)
1	a	a2 ✓	a1 ✓
2	ad	ad2 ✓	ad1 ✓
3	ant	ant2 ✓	ant1 ✓
4	ap	pae2 x	sol x
5	b3	b2 ✓	b1 ✓
6	bo	va2 x	kanil x
7	bu	ta2 x	bul ✓
8	ch	ch2 ✓	sakul x
9	cu	cu2 ✓	cul ✓
10	ja	ja2 ✓	jal ✓
11	ka	ka2 ✓	kal ✓
12	kani	si2 x	tal x
13	ko	va2 x	kol ✓
14	mo	mo2 ✓	mol ✓
15	mum	tum2 x	mum1 ✓
16	n	n12 ✓	n11 ✓
Accuracy		10/16 63%	12/16 75%

From the results in Table 1, it can be seen that the accuracy of recognition can be improved by using the average IDTW scores from the Near record and the Far record, as shown in Table 2. These were calculated by the following formula:

$$Best\_Match = \operatorname{argmin}_{k=1..K} \left( \frac{IDTW_{near_k} + IDTW_{far_k}}{2} \right) \quad (3)$$

where *IDTW* near is the IDTW score of Near record, *IDTW*far is the IDTW score of Far record, and *K* is the total number of files.

Usually, when the skeleton files from the Far record are compared against those of the Near record, each containing the same reference gestures of the same person, the IDTW score should be close to zero. However, as demonstrated in our experiments, it is very difficult to achieve an IDTW score of zero, or close to zero. Thus, it is important to obtain the best possible recordings and images for inclusion in the reference gesture database.

Some query instances (as shown in Table 1) provided positive and negative results; for example, persons numbered 4, 6 and 12. The application of equation 2 reduced the confusion when making decision with nearest neighbor search on Far and Near records. The new results are represented in Table 2 and the average result was 81%.

Figure 4 shows a graphical comparison of distance scores of both the classic DTW and IDTW algorithms for each frame during the comparison tests between two data sequences. It shows that the distance scores approach zero at the last data frame for both algorithms, which means that both algorithms can ultimately correctly

identify the subject with enough data. However, the IDTW provides the answer faster than the DTW.

In practice, recorded data sequences may be incomplete, or the query sequence is incomplete and may not be effective for comparison with the completed reference data. In such cases, the IDTW algorithm was superior in recognition with incomplete recorded sequences. Moreover, the IDTW algorithm can support the real-time classification because it can issue the comparing result within a second.

Table 2  
Recognition results when calculated average IDTW scores between the Near record and the Far record

No.	Ref.	Result (Average)
1	a	a ✓
2	ad	ad ✓
3	ant	ant ✓
4	ap	pae x
5	b3	b ✓
6	bo	va x
7	bu	bu ✓
8	ch	ch ✓
9	cu	cu ✓
10	ja	Ja ✓
11	ka	ka ✓
12	kani	si x
13	ko	ko ✓
14	mo	mo ✓
15	mum	mum ✓
16	n	n ✓
Accuracy		13/16 81%

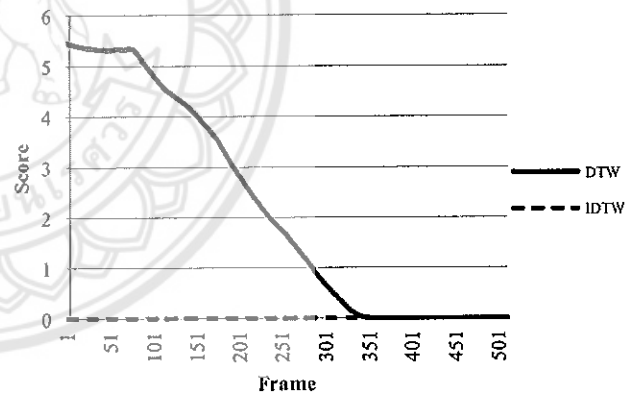


Figure 4: Comparison of distance scores between DTW and IDTW

#### IV. CONCLUSIONS

We have proposed and tested a method for early recognition of human gait gestures, for person identification, using the Kinect camera. The Incremental Dynamic Time Warping (IDTW) algorithm was applied to calculate the similarity between human gait gestures. We showed that the IDTW algorithm can calculate the distance score between two sequences faster and with greater accuracy, than the classic DTW. The IDTW algorithm accurately identified the test subject 81% of the time. A larger volume of data will enhance this outcome, and we will develop the image database to achieve this end. We will also move the database to be remotely stored in a 'cloud' system.

REFERENCES

- [1] Fothergill, S., Mentis, H., Kohli, P. and Nowozin, S. Instructing people for training gestural interactive systems. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, (2012).
- [2] Maquet, Paul GJ. Biomechanics of the knee: with application to the pathogenesis and the surgical treatment of osteoarthritis. Springer Science & Business Media, (2012).
- [3] Dikovski, Bojan, Gjorgji Madjarov, and Dejan Gjorgjevikj. Evaluation of different feature sets for gait recognition using skeletal data from Kinect. Information and Communication Technology, Electronics and Microelectronics (MIPRO), 37th International Convention on. IEEE, (2014).
- [4] Iosifidis, Alexandros, Anastasios Tefas, Nikolaos Nikolaidis, and Ioannis Pitas. Multi-view human movement recognition based on fuzzy distances and linear discriminant analysis. Computer Vision and Image Understanding, 116.3 (2012): 347-360.
- [5] Munsell, B. C., Temlyakov, A., Qu, C., and Wang, S. Person identification using full-body motion and anthropometric biometrics from kinect videos. In Computer Vision—ECCV-2012. Workshops and Demonstrations. Springer Berlin Heidelberg, (2012) 91-100.
- [6] Naimul Mefraz Khan, Stephen Lin, Ling Guan, and Baining Guo. A Visual Evaluation Framework for In-Home Physical Rehabilitation. IEEE International Symposium on Multimedia, (2014).
- [7] M. Kyan, G. Sun, H. Li, L. Zhong, P. Muneesawang, N. Dong, B. Elder, and L. Guan, An Approach to Ballet Dance Training through MS Kinect and Visualization in a CAVE Virtual Reality Environment. Special Issue on Visual Understanding with RGB-D Sensors, ACM Transactions on Intelligent Systems and Technology (TIST), Vol. 6(2), Article ID 23 (2015).
- [8] W. Kwankhoom and P. Muneesawang. Recognition of Standard Thai Traditional Dance Through 3D Data Analysis. Naresuan University Engineering Journal, Vol.11, No.2 (2016) 75-84.
- [9] Microsoft Developer Network. JointType Enumeration. Available at <https://msdn.microsoft.com/en-us/library/microsoft.kinect.jointtype.aspx>. Accessed Nov 11, (2016).
- [10] Miguel Reyes, Gabriel Dominguez, and Sergio Escalera. Feature Weighting in Dynamic Time Warping for Gesture Recognition in Depth Data. IEEE International Conference on Computer Vision, (2011).



# An Incremental Dynamic Time Warping for Person Re-identification

Wisrut Kwankhoom and Paisarn Muneesawang

Department of Electrical and Computer Engineering, Faculty of Engineering,  
Naresuan University, Phitsanulok 65000, Thailand

**Abstract**— This paper presents principles and techniques of a human gesture recognition algorithm for person identification which identifies personal gait patterns recorded with a 3D depth sensing camera, in this case the Microsoft Kinect<sup>®</sup> version 2. The recorded images are analyzed against a dataset of gait gestures derived from a sample of 37 people. We compared two algorithms for analyzing movement trajectories; Sparse code and Incremental Dynamic Time Warping (IDTW). Experimental results show that the methods have an encouraging performance. When comparing the accuracy of algorithms, IDTW gave better recognition results than the Sparse code method.

**Keywords**- *Gesture recognition; Personal identification; recognition techniques; Sparse code; IDTW*

## I. INTRODUCTION

With an increasing emphasis on security, automated personal identification based on biometrics has been receiving extensive attention over the past decade. Human gesture recognition is an emerging biometric recognition approach that has been an active topic in both research and practical applications. Personal identification systems can be classified as either being Active systems or Passive systems. Personal identification is an active system whereby the personal physical and biological data from persons are analyzed via biological means such as fingerprint recognition or retinal scanning. Passive systems analyze personal data from the person by facial recognition or by voice print recognition. The analysis results from such passive systems depend on the analysis of clear data without noise interference or by other visual or aural obstacles that interfere with and reduce the efficacy of detection.

Gesture recognition has become an important part of communication between humans and computers. Interactive human body movement and gesture tracking is the basis of many applications including gaming, human-computer interaction, telepresence, health-care and security. Generally, the recognition result is acquired after the movement or gesture has been detected and completed. Therefore, if these are to be available for analysis, they need to be captured, recorded and stored. When long gesture or gait patterns are being analyzed, lengthy analysis is required.

In this research we created and tested a system for gait gesture analysis. Two recognition algorithms, Sparse code and IDTW were tested for our system as the gait pattern recognition process that we propose, in which the subject in motion is recorded using the Kinect<sup>®</sup> 3D depth sensing camera. From the captured video image of the person passing in front of a sensor,

twenty five points at particular joints on the upper and lower skeleton of the body are identified and tracked and compared against sample data already available in a database.

## II. LITERATURE REVIEW

Gesture recognition is an ideal example of multidisciplinary research. There are different tools for gesture recognition, based on approaches including statistical modeling, computer vision and pattern recognition, image processing, connectionist systems, etc. Statistical modeling has been the prominent method for gesture recognition, but there has been interest in other methods of gesture recognition, such as:

W. Kwankhoom and P. Muneesawang [1] proposed the recognition of Standard Thai Traditional Dance gestures through 3D Data obtained with a Kinect<sup>®</sup> sensor by using Self-Organizing Mapping (SOM) to create gesture templates and applying the Posture Transition Sparse Code (PTSC) for calculating gesture similarity. The method performs well in recognizing standard Thai traditional dance gestures in real-time environments and provides high accuracy in classification.

Shimada et al [2], proposed gesture recognition by using an artificial neural network (Self-Organizing Map: SOM) and the Sparse code index for the gesture recognition, and was able to achieve better than 98% accuracy. This approach could be developed as part of an interactive recognition system.

Munsell et al [3] proposed personal identification from person-specific walking and running motions captured by the Kinect<sup>®</sup> sensor. In this study, the researchers found that body segments could be correctly isolated and identified, and body motions and movements classified from images from the Kinect<sup>®</sup> sensor. The approach described in this paper was stated to be more successful and accurate than previous methods.

Khan et al [4] developed a method for in-home physical rehabilitation, where a user can visually evaluate his/her own performance compared to that of an expert by using the Kinect<sup>®</sup> sensor. The application applies the Incremental Dynamic Time Warping (IDTW) algorithm for comparing sequences. This paper also proposed the method of mapping the IDTW measurements to a color-coded skeleton frame for evaluation.

In the current research the concept of using the Kinect<sup>®</sup> sensor for data recording, together with the Sparse code and IDTW methods for analysis of the data for gait gesture recognition was developed. This is the unique contribution of the research, the combination of these three elements in a single system. The system is an extension of the research of Khan et al (ibid) with the objective of implemented a new approach to gait gesture recognition in a system with enhanced performance.

### III. PROPOSED METHODOLOGY

This section examines relevant theory relating to the recognition of gait gesture and the process of recording the human gait patterns to enable the creation of skeletal models of the sample people. This data is normalized to compute the feature vectors, and to measure similarity. The concept of data normalization is described at length below.

#### A. Human gait gesture

Gait is characterized as a facilitated cyclic mix of activities that effect the motion of the individual's body [5]. The movement of the human body presents a specific pattern that can identify an individual by their walking behavior. The walking 'activities' are shown in Figure 1. There is a sequence of alternative footsteps identified in every movement of human walking. The motion of the two feet produces a cyclic action which is unique to an individual [6], whether walking, jogging, sprinting, ascending stairs, stretching, relaxing or bending down to pick up an object etc. Jumping from the current position to another position also falls under human locomotion. Human locomotion actions may be cyclic or non-cyclic.

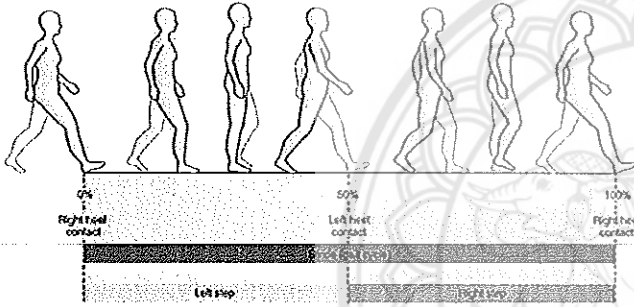


Figure 1. The gait cycle [7]

#### B. Kinect® camera version 2

The Kinect® camera is a human motion tracking peripheral for the Xbox 360 console from Microsoft. It was first applied in gaming systems but was found to have many other applications and uses, such as human motion and features recognition, 3D model reconstruction, robot navigation, medical applications and dance training.

Kinect® version 2 can extract many aspects of human body poses from the 3D depth images, representing the body as a skeletal image with identifiable body joints. The positions being tracked and recorded are more anatomically correct and stable the longer the activity proceeds, given that more video data is being recorded. Overall, twenty-five body joints can be represented, as shown in Figure 2.

From the Skeleton representation, the Spine base joint is used as the origin of the skeleton and the relative positions of the other joints are calculated from this joint and we use the values of each body joint to analyze the movement gesture.

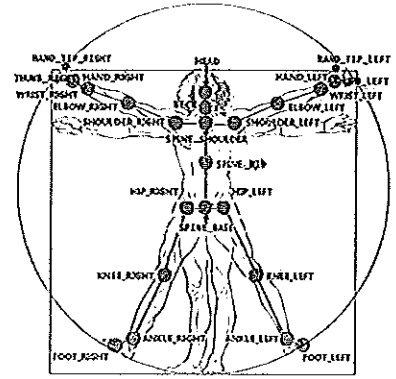


Figure 2. Microsoft SDK Skeleton representation [8]

The data sequence from Kinect® will include 3D position  $(x, y, z)$  of 26 joints at any time.

$$Sequence(t_0) = \begin{bmatrix} x_0 & x_1 & \dots & x_{25} \\ y_0 & y_1 & \dots & y_{25} \\ z_0 & z_1 & \dots & z_{25} \end{bmatrix}$$

$$Sequence(t_1) = \begin{bmatrix} x_0 & x_1 & \dots & x_{25} \\ y_0 & y_1 & \dots & y_{25} \\ z_0 & z_1 & \dots & z_{25} \end{bmatrix}$$

⋮

$$Sequence(t_N) = \begin{bmatrix} x_0 & x_1 & \dots & x_{25} \\ y_0 & y_1 & \dots & y_{25} \\ z_0 & z_1 & \dots & z_{25} \end{bmatrix}$$

This research obtained a dataset of gait gestures from sample people so we will have the sequences:

$$S = \{X_0, X_1, \dots, X_{25}\}$$

where  $S$  is a data sequence and  $t$  is the time and  $X$  is joint positions ( $X_i = [x_i, y_i, z_i]$ )

#### C. Data Normalization

The data from the Kinect® sensor cannot be used directly because the body ratio varies depending on distance from the camera and thus size of image. The data therefore must be calculated to conform to a common standard [4], which we term Data Normalization. In this data normalization process we use the Spine base (shown in Figure 2) as the point of origin and calculate distances between the joints and from the Spine base according to the following:

$$S'_i = \frac{S_i - S_0}{\|S_i - S_0\|^2}, i = 1, \dots, 25 \quad (1)$$

where  $S'_i$  is a normalize body joint,  $S_0$  is the Spine base,  $S_i$  is another joint, and  $i$  is the number of joints.

#### D. Self-Organizing Map

The approach used in performing posture recognition in this application is the self-organizing map (SOM). SOM is a type of artificial neural network that is trained using unsupervised

learning vector quantization to create a map which is used for finding the result of an input space [9]. Figure 3 shows an example of SOM where the map is adjusted by a competitive learning rule. The winning node and its neighbors are moved according to the rule. The approach is described by the following algorithm:

**Algorithm 1: Self-organizing map**

**input:** training set  $S' = \{X_0, X_2, \dots, X_n, \dots, X_N\}$

**output:** weight vector  $W = \{W_0, W_2, \dots, W_m, \dots, W_M\}$

**initialize** the weight vectors by selecting vectors from  $X$

**repeat**

    Random select  $X_n$  from set  $S'$

    Find all distance  $d$  from  $X_n$  to all  $W_m$

    Find winning node using  $\min_{m=1, \dots, M} d_m$

    Update all nodes and update parameters

**until** nodes cannot be moved

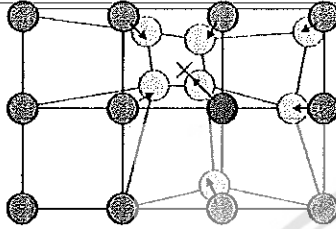


Figure 3. Neuron updating in SOM

The application defines 64 output nodes (also known as neurons) for learning in SOM which are applied to obtain the result sequence of each class. By doubling the number of output nodes for SOM, the recognition precision will decrease dramatically. After all learned neurons are assigned, Sparse code is used for indexing trainer's postures. The Boolean matrix is created for each posture and mapped to the SOM result, as shown in Figure 4. Any tester's posture will be first mapped to the SOM result, then compared to all of the trainer's Sparse code.

Figure 4 shows the steps in the process of gesture sequences mapping to SOM to create a class pattern sequence of  $I(1), I(2), I(3), I(4)$ , wherein every frame of a gesture is mapped to SOM node. This is also called the Coding Process, which produces the class pattern of sequences which can be used to create a template histogram for the gesture class.

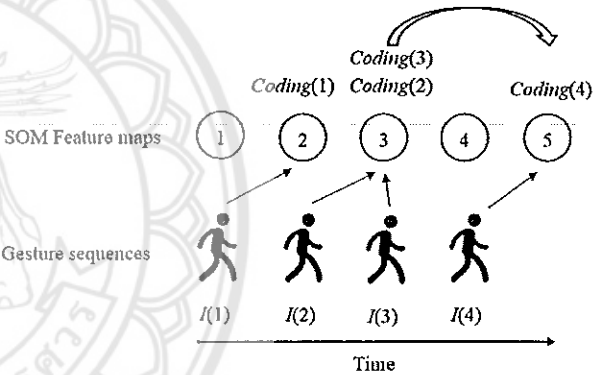
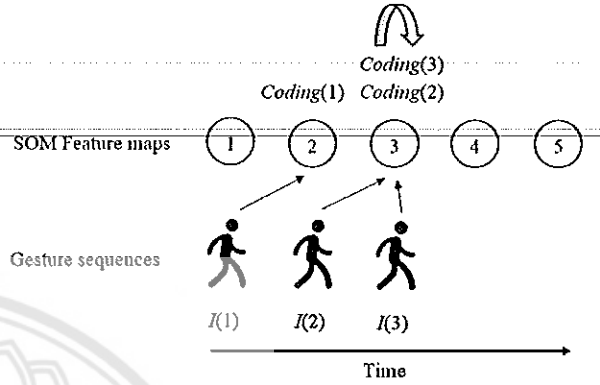
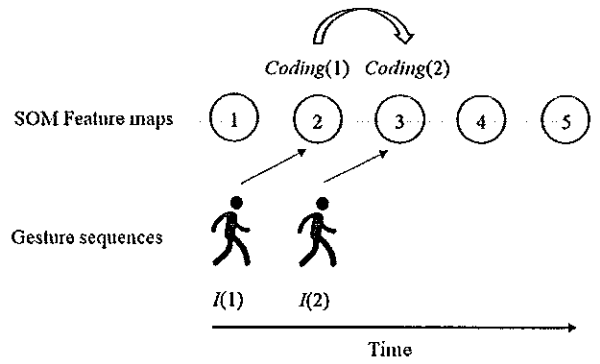
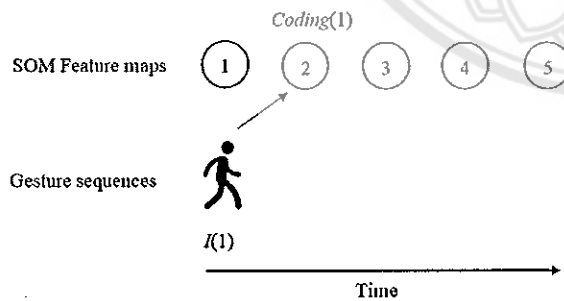


Figure 4. Sample of data encryption for 4 frames class pattern sequence of  $I(1), I(2), I(3), I(4)$

### E. Gesture Histogram

After a class pattern sequence has been calculated from vector quantization by using SOM, we can create a histogram by using Sparse code (SC) which is the representation of items by the strong activation of a relatively small set of neurons. For each stimulus, this is a different subset of all available neurons. In this first descriptor, a histogram is formed for each instance  $n$  in the original's gesture sequence  $H_{c,n} = hist(S_{c,n})$ . A template histogram for the gesture class may also be formed by summing over the set of  $H_{c,n}$ , where  $c = 1, 2, \dots, K$  represents the set of gesture classes. The histogram can be calculated as [10]:

$$C_{c,n} = \begin{cases} 1, & \text{if } H_{c,n} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$$SC_c = C_{c,n}, 1 \leq n \leq N \quad (3)$$

Figure 4 the data encryption process which creates a template histogram for the gesture class, as shown in Figure 5. The histograms for each gesture will be used for comparing with other gesture sequences, and to identify matched or unmatched sequence.

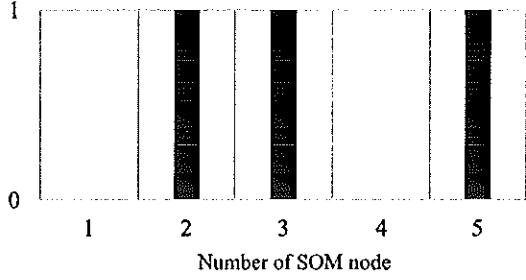


Figure 5. Sample of Sparse code histogram

#### F. Incremental Dynamic Time Warping (IDTW)

IDTW is an algorithm for comparing video image sequences taken at varying speeds and over different time periods by compressing or expanding them in the time domain. This is an extension of the DTW as described in [4, 11]. We used the stored complete data of skeletal features as the reference sequence and benchmark, and did a real-time comparison between the new sequences with the reference sequence data. The algorithm calculates a distance score between sequences, providing a comparison between the incomplete sequence and the reference complete sequence, thereby indicating the degree of sameness between the images.

IDTW works by comparing each sequence with the best starting segment of reference. It terminates with the result after applying the full (reference) sequence with all possible frames, computing the DTW distances for every possible comparison and reporting the minimum distance as follows:

- I. Initialize parameters:  $U$  is the user sequence up to current time (length  $N$ ),  $E$  is the reference full sequence (length  $M$ ),  $G$  is the matrix ( $M \times N$ ) cumulative cost matrix up to current time, and  $V$  is next frame in user sequence.
- II. Inserting a new frame:  $Q = N + 1$  and  $U_Q = V$ , where  $U_Q$  is the  $Q$ -th frame of  $U$ .
- III. The DTW distance is calculated for all  $E^j, j = 1, \dots, M$  where each  $E^j$  is the reference sequence  $E$  truncated at  $j^{\text{th}}$  frame, i.e.,  $D(U, E^j), j = 1, \dots, M$  is the DTW distance obtained by the optimal warping path.
- IV. The minimum is found using the following formula:

$$D_{IDTW}(U, E) = \min_{j=1, \dots, M} D(U, E^j) \quad (4)$$

- V. Step II and IV are iterated for each the new sequence and updated IDTW distance (the cost matrix  $G$ ) by using the following formula:

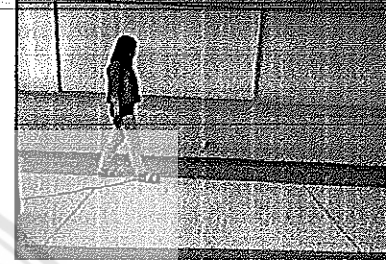
$$G = \min(G(1 \dots M, Q)) / Q \quad (5)$$

Note that the components of DTW from previous states are reused as the new sequence movement advances, to optimize computation time.

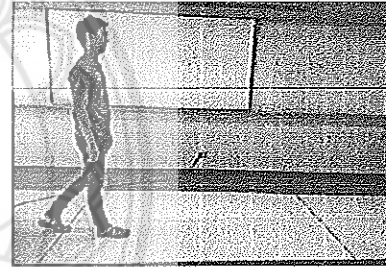
Considering the steps, we notice that the next frame of the user sequence will be input to the process. Then just one column is added to the cumulative cost matrix  $G$ . This column is filled in line with a chosen updated rule in a similar way to classic DTW. Finally, the normalized minimum value from the newly added column of the cost matrix  $G$  comes out as the current IDTW distance.

#### IV. EXPERIMENT RESULTS

We obtained a dataset of gait gestures from 37 people by each person walking at a different distance from the Kinect<sup>®</sup> sensor (Figure 6), at the Near Recording distance of two meters and the Far Record distance of three meters, as shown in Figure 7.



(a) Far record



(b) Near record

Figure 6. Example of side-view gait data distance of recording

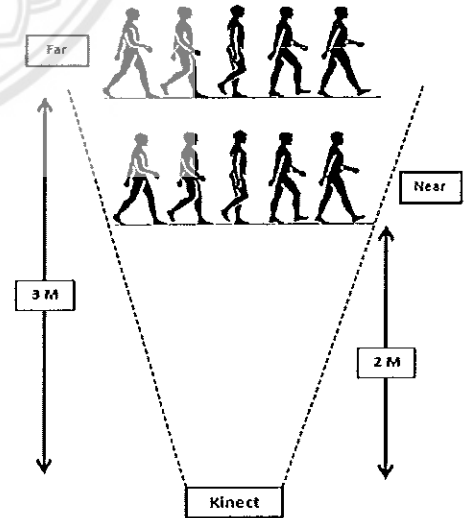


Figure 7. The experimental setting



The accuracy of the IDTW algorithm was also compared with Sparse code algorithm. The results are shown in Table 1. Top1 means the best result, and Top5 means the best five results.

TABLE I. RESULT OF REGCONITION

Recognition Accuracy	Sparse code		IDTW	
	Top1	Top5	Top1	Top5
Far record	8 of 37	19 of 37	24 of 37	35 of 37
	22%	51%	65%	95%
Near record	10 of 37	20 of 37	27 of 37	34 of 37
	27%	54%	73%	92%

Table 1 shows the recognition accuracy of Top1 and Top5 solutions of each algorithm. In the Far recording dataset, the Sparse code algorithm achieved 22% accuracy in Top1 and 51% accuracy for the Top5 results. For the Near record dataset, 27% accuracy in Top1 and 54% accuracy in the Top5 results were achieved.

The results when applying IDTW were, for the Far record dataset, 65% accuracy in Top1 and 95% in the Top5 results, and For the Near record dataset 73% accuracy in Top1 and 92% in the Top5 results were achieved.

The results of IDTW are shown in Table 1. Table 2 illustrates that the accuracy of recognition can be improved by using the average results from the Near record and the Far record datasets. These were calculated by the following formula:

$$IDTW_{Best\_Match} = \operatorname{argmin}_{k=1,\dots,K} \left( \frac{IDTW_{near_k} + IDTW_{far_k}}{2} \right) \quad (6)$$

where  $IDTW_{near}$  is the IDTW score of Near record,  $IDTW_{far}$  is the IDTW score of Far record, and K is the total number of files.

The results from the application of our new IDTW algorithm (Formula 6) are represented in Table 2, and show an improved recognition accuracy of up to 78% for the Far record and 97% for the Near record.

TABLE II. AVERAGE RESULT OF REGCONITION FOR IDTW ALGORITHM

Recognition Accuracy	IDTW algorithm	
	Top1	Top5
Average	29 of 37	36 of 37
	78%	97%

## V. CONCLUSION

We proposed the Incremental Dynamic Time Warping (IDTW) algorithm for human gait recognition through 3D data analysis from Microsoft's Kinect® sensor. The IDTW algorithm was able to recognize human gait gestures with a higher level of accuracy than the Sparse code algorithm alone because of the ability of the IDTW algorithm to compare each sequence with the best starting segment of the reference image and the result is the distance and the best alignment between the data for these two sequences, taking into account the variation in speed and time of the two separate sequence data sets, which SOM cannot achieve. It is suggested that future development should include a larger volume data.

## ACKNOWLEDGEMENT

Many thanks to Mr. Roy Morien of the Naresuan University Language Centre for his editing assistance and advice on English expression in this document.

## REFERENCES

- [1] W. Kwankhoom and P. Muneesawang. "Recognition of Standard Thai Traditional Dance Through 3D Data Analysis", Naresuan University Engineering Journal, Vol.11, No.2, 2016.
- [2] Shimada, Atsushi, and Rin-ichiro Taniguchi. "Gesture recognition using sparse code of hierarchical SOM." Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE, 2008.
- [3] B.C. Munsell, Andrew Temlyakov, Chengzheng Qu, and Song Wang. "Person identification using full-body motion and anthropometric biometrics from Kinect® videos." In Computer Vision-ECCV 2012. Workshops and Demonstrations, pp. 91-100. Springer Berlin Heidelberg, 2012.
- [4] Khan, Naimul Mefraz, Stephen Lin, Ling Guan, and Baining Guo. "A Visual Evaluation Framework for In-Home Physical Rehabilitation." In Multimedia (ISM), 2014 IEEE International Symposium on, pp. 237-240. IEEE, 2014.
- [5] Neumann, Donald A. "Kinesiology of the musculoskeletal system: foundations for rehabilitation", Elsevier Health Sciences, 2013.
- [6] M. Jhapate, L. Choudhary and R. Singh Pippal, "A Review on Triangle based Techniques in Biometric Gait Recognition", International Journal of Computer Applications, vol. 121, no. 7, pp. 6-9, 2015.
- [7] P. Gangotia, G. Shrivastava and J. Pal Singh, "An Approach of Human Identification based on Gait", International Journal of Computer Applications, vol. 93, no. 1, pp. 33-36, 2014.
- [8] Microsoft Developer Network. JointType Enumeration. Available at <https://msdn.microsoft.com>, Accessed Feb 14, 2017.
- [9] Kawashima M, Shimada A, Taniguchi R, "Early recognition of gesture patterns using sparse code of self-organizing map", 7th International Workshop on Self-Organizing Maps, 2009.
- [10] M. Kyan, G. Sun, H. Li, L. Zhong, P. Muneesawang, N. Dong, B. Elder, and L. Guan, "An Approach to Ballet Dance Training through MS Kinect® and Visualization in a CAVE Virtual Reality Environment", in Special Issue on Visual Understanding with RGB-D Sensors, ACM Transactions on Intelligent Systems and Technology (TIST), Vol. 6(2), Article ID 23, 2015.
- [11] Reyes, Miguel, Gabriel Domínguez, and Sergio Escalera. "Featureweighting in dynamic timewarping for gesture recognition in depth data." In Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, pp. 1182-1188. IEEE, 2011.