

การประกอบรวมโครงร่างมนุษย์จากการตรวจจับการเคลื่อนไหว โดยใช้กล้อง Kinect หลายตัว

นราวุฒิ พัฒโนทัย, พรชัย มงคลนาม และ บัณฑิต วรรณภา

คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

126 ถนนประชาธิปไตย แขวงบางมด เขตทุ่งครุ กรุงเทพมหานคร 10140

Emails: 53441318@st.sit.kmutt.ac.th, {pornchai | bunthit}@sit.kmutt.ac.th

บทคัดย่อ

กล้อง Kinect สามารถตรวจจับโครงร่างของมนุษย์ได้อย่างถูกต้อง เฉพาะส่วนที่ไม่ถูกบดบัง ดังนั้นในทางปฏิบัติ การตรวจจับการเคลื่อนไหวของมนุษย์ จึงมีความจำเป็นต้องใช้กล้องสองตัวหรือมากกว่า เพื่อสร้างโครงร่างที่สมบูรณ์จากการตรวจจับ แต่การเพิ่มจำนวนกล้องทำให้ได้โครงร่างมากกว่าหนึ่งในการตรวจจับคนหนึ่งคน งานวิจัยนี้จึงนำเสนอวิธีการการประกอบรวมโครงร่างมนุษย์ที่ได้จากการตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect หลายตัว และนำเสนอระบบการตรวจจับการเคลื่อนไหวตามเวลาจริงที่รองรับการใช้กล้อง Kinect จำนวนมากได้ จากผลการทดสอบวิธีการที่นำเสนอปรากฏผลในระดับที่ยอมรับได้ และสามารถนำไปใช้ในงานอื่นๆ ได้ต่อไป

คำสำคัญ-- กล้อง Kinect; การตรวจจับการเคลื่อนไหว; การประกอบรวมโครงร่างมนุษย์; OpenNI

1. บทนำ

กล้อง Kinect เป็นอุปกรณ์เสริมการใช้งานเครื่องเล่น Xbox 360 [1] มีความสามารถในการตรวจจับร่างกายผู้เล่น และสามารถจับแฉกตำแหน่งของข้อต่อต่างๆ ของร่างกายผู้เล่นได้ โดยให้ข้อมูลตำแหน่งของข้อต่อในรูปแบบจุดพิกัดสามมิติ ดังนั้นจึงสามารถนำมาประยุกต์ใช้กับการตรวจจับการเคลื่อนไหวของร่างกายมนุษย์ แบบปราศจากการทำเครื่องหมายตามตำแหน่งต่างๆ บนร่างกายได้ (Markerless Motion Capture)

เมื่อใช้กล้อง Kinect ในการตรวจจับการเคลื่อนไหว ข้อมูลตำแหน่งของข้อต่อของร่างกายที่ได้จะมีความถูกต้องเฉพาะส่วนที่กล้องตรวจจับได้ หากบางส่วนของร่างกายถูกบดบัง ข้อมูลตำแหน่งของข้อต่อส่วนนั้นจะเป็นค่าประมาณ มีความถูกต้องน้อย การทำให้ทุกส่วนของร่างกายถูกตรวจจับได้ตลอดเวลา จึงต้องมีการเพิ่มจำนวนกล้องในการตรวจจับ และจัดวางกล้องในหลายๆ มุมมอง แต่การทำเช่นนี้จะทำให้ได้ข้อมูลตำแหน่งของข้อต่อมากกว่าหนึ่งชุดในการตรวจจับคนหนึ่งคน จึงจำเป็นต้องนำข้อมูลตำแหน่งที่ได้มาประมวลผลรวมกัน และทำการ

ทดแทนข้อมูลตำแหน่งที่ถูกบดบังในบางมุมมอง ด้วยข้อมูลตำแหน่งจากกล้องอื่นที่ไม่ถูกบดบัง เพื่อให้ได้ข้อมูลตำแหน่งของข้อต่อที่ประกอบกันเป็นโครงร่างมนุษย์ที่มีความถูกต้องสมบูรณ์ สามารถนำไปใช้ในการตรวจจับท่าทางและการเคลื่อนไหวได้ต่อไป

2. การตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect

การตรวจจับการเคลื่อนไหวของร่างกายมนุษย์โดยใช้กล้อง Kinect ในงานวิจัยนี้ ใช้ API ของ OpenNI [2] ในการให้ข้อมูลตำแหน่งของข้อต่อของผู้ที่ถูกตรวจจับได้ โดยตำแหน่งของจุดกำเนิด ($X=Y=Z = 0$) จะเป็นตำแหน่งของกล้องที่ใช้ในการตรวจจับ ตามที่แสดงในรูปที่ 1 ซึ่งแกน Z จะเป็นทิศทางที่กล้องตรวจจับ

ค่า X จะเป็นค่าแสดงระยะทางที่ห่างออกไปจากจุดกำเนิดตามแนวอนันต์สัมพัทธ์กับกล้องที่ตรวจจับ จะมีค่าเป็นบวกเมื่อตำแหน่งที่ถูกตรวจจับได้อยู่ทางด้านขวาของกล้อง ค่า Y จะเป็นค่าแสดงระยะทางที่ห่างออกไปจากจุดกำเนิดตามแนวตั้งที่กล้องตรวจจับ จะมีค่าเป็นบวกเมื่อตำแหน่งที่ถูกตรวจจับได้อยู่สูงกว่าตำแหน่งของกล้อง และค่า Z จะเป็นระยะทางที่ห่างออกไปจากกล้องโดยจะมีค่าเป็นบวกเสมอ

ตำแหน่งของข้อต่อที่ได้รับจาก API จะสัมพันธ์กับตำแหน่งของร่างกายมนุษย์ 15 ตำแหน่ง ตามที่แสดงในรูปที่ 2 เมื่อผู้ที่ถูกตรวจจับยื่นหน้าเข้าหากล้อง ตำแหน่งของข้อต่อข้างซ้ายและขวาจะสลับข้างกัน แตกต่างจากร่างกายจริง แต่ละตำแหน่งของข้อต่อจะมีค่า X, Y และ Z ในหน่วยมิลลิเมตร และมีฟิลด์ fConfidence แสดงสถานะการตรวจจับตำแหน่งข้อต่อนั้นๆ โดยจะมีค่าระหว่าง 0 – 1 ซึ่ง 1 หมายถึง สามารถตรวจจับได้ และหากน้อยกว่า 1 หมายถึง การตรวจจับเป็นการประมาณ

ในการตรวจจับการเคลื่อนไหว หากผู้ที่ถูกตรวจจับอยู่ภายในมุมกล้อง ห่างจากกล้องประมาณ 1.2 – 3.5 เมตร ค่าตำแหน่งของข้อต่อที่ถูกตรวจจับได้ จะมีค่าใกล้เคียงกับการวัดจริง และฟิลด์ fConfidence จะมีค่าเป็น 1 แต่หากมีบางข้อต่อที่กล้องไม่สามารถตรวจจับได้ หรือถูกบดบัง เช่น ยืนมือโพล่หลัง เป็นต้น ค่าตำแหน่งของข้อต่อนั้นจะเป็นค่าประมาณ และฟิลด์ fConfidence จะมีค่าน้อยกว่า 1



รูปที่ 1 แกน X, Y และ Z ของกล้อง Kinect [3]

3. งานวิจัย และทฤษฎีที่เกี่ยวข้อง

3.1. งานวิจัยที่เกี่ยวข้อง

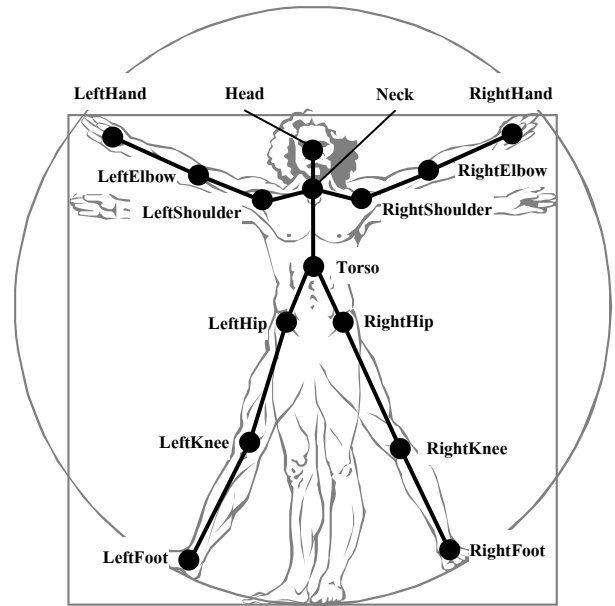
งานวิจัยของ D. Alexiadis, P. Kelly, P. Daras, N. E. O'Connor, T. Boubekeur, และ M. B. Moussa [4] ได้ทำการศึกษาเกี่ยวกับระบบการวัดประสิทธิภาพของนักเดิน โดยใช้กล้อง Kinect ในการตรวจจับการเคลื่อนไหว และใช้ OpenNI เป็น API ในการให้ข้อมูลการตรวจจับ งานวิจัยนี้ใช้การเปรียบเทียบตำแหน่งของข้อต่อของนักเดินสองคนที่ได้จากการตรวจจับตามเวลาจริงมาเป็นส่วนหนึ่งในการให้คะแนนการเดิน ผลการศึกษาของงานวิจัยนี้แสดงให้เห็นว่าสามารถนำข้อมูลตำแหน่งของข้อต่อไปใช้งานในด้านการตรวจจับการเคลื่อนไหวได้

งานวิจัยของ K. Berger, K. Ruhl, Y. Schroeder, C. Bruemmer, Al. Scholz, และ M. Magnor [5] ได้ทำการศึกษาเกี่ยวกับการตรวจจับการเคลื่อนไหวแบบปราศจากการทำเครื่องหมายตามตำแหน่งต่างๆ บนร่างกาย โดยใช้กล้อง Kinect หนึ่งถึงสี่ตัวในการตรวจจับ แล้วนำภาพสีที่แสดงความลึก (Depth Image) ที่ได้จากกล้องในหลายๆ มุมมอง นำไปคำนวณหาค่าตำแหน่งของข้อต่อของร่างกายผู้ถูกตรวจจับ จากนั้นนำตำแหน่งของข้อต่อไปสร้างเป็นโครงร่างมนุษย์ ซึ่งแตกต่างจากงานวิจัยนี้ที่นำตำแหน่งของข้อต่อที่ได้รับจาก API นำไปสร้างเป็นโครงร่างมนุษย์โดยตรง ผลการศึกษาของงานวิจัยนี้แสดงให้เห็นว่าการใช้กล้องหลายตัวเพื่อเพิ่มมุมมองในการตรวจจับ ทำให้สามารถติดตามตำแหน่งของข้อต่อได้อย่างสม่ำเสมอ และสามารถช่วยลดโอกาสที่ส่วนของร่างกายไม่ถูกตรวจจับได้อย่างมีนัยสำคัญ

3.2. การแปลงระบบพิกัด

เนื่องจากกล้อง Kinect แต่ละตัวที่ใช้ในการตรวจจับการเคลื่อนไหว จะใช้ตำแหน่งของกล้องเป็นตำแหน่งจุดกำเนิด ทำให้ตำแหน่งของข้อต่อตำแหน่งเดียวกันที่ถูกตรวจจับได้ มีค่าแตกต่างกัน เพราะมีตำแหน่งจุดกำเนิดที่เป็นตำแหน่งอ้างอิงต่างกัน ดังนั้น จึงต้องมีการแปลงระบบพิกัดของกล้องให้เป็นระบบพิกัดเดียวกัน เพื่อปรับตำแหน่งของข้อต่อให้มีตำแหน่งอ้างอิงเดียวกัน

การแปลงระบบพิกัด (Coordinate System Transformation) ในงานวิจัยนี้ จะใช้การหมุน (Rotation) และการย้าย (Translation) ในระบบพิกัดสามมิติ นำมาสร้างเมตริกซ์การแปลง (Transformation Matrix) [6] โดยนำเมตริกซ์การหมุนในแต่ละแกนมาคูณกัน ทำให้ได้เมตริกซ์การหมุนแบบ X-Y-Z Fixed Angles Rotation [7] ตามที่แสดง



รูปที่ 2 ตำแหน่งของข้อต่อที่สัมพันธ์กับตำแหน่งร่างกายมนุษย์ [3]

ในสมการที่ 2 จากนั้นนำเมตริกซ์การย้ายมาคูณกับเมตริกซ์การหมุน จะทำให้ได้เมตริกซ์การแปลงที่มีทั้งการย้ายและการหมุนพร้อมกัน ตามที่แสดงในสมการที่ 4

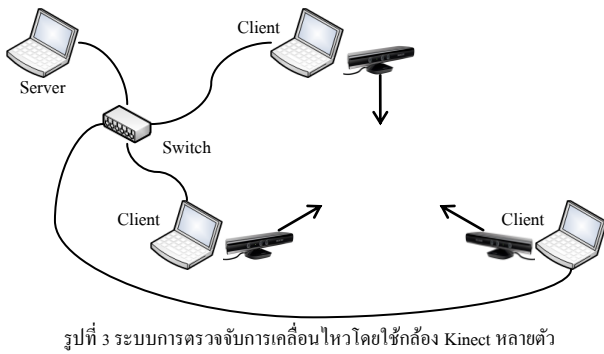
$$[transform] = [translate] \cdot ([rotate\ Z\text{-axis}] \cdot [rotate\ Y\text{-axis}] \cdot [rotate\ X\text{-axis}]) \quad (1)$$

$$= [translate] \cdot [X\text{-Y-Z Fixed Angles Rotation}] \quad (2)$$

$$= \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

$$= \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

เมื่อ r_{11} คือ $\cos(a_z) \cos(a_y)$
 r_{12} คือ $-\sin(a_z) \cos(a_x) + \cos(a_z) \sin(a_y) \sin(a_x)$
 r_{13} คือ $\sin(a_z) \sin(a_x) + \cos(a_z) \sin(a_y) \cos(a_x)$
 r_{21} คือ $\sin(a_z) \cos(a_y)$
 r_{22} คือ $\cos(a_z) \cos(a_x) + \sin(a_z) \sin(a_y) \sin(a_x)$
 r_{23} คือ $-\cos(a_z) \sin(a_x) + \sin(a_z) \sin(a_y) \cos(a_x)$
 r_{31} คือ $-\sin(a_y)$
 r_{32} คือ $\cos(a_y) \sin(a_x)$
 r_{33} คือ $\cos(a_y) \cos(a_x)$
 a_x, a_y และ a_z คือ ขนานมุมที่ใช้ในการหมุนในแต่ละแกน
 t_x, t_y และ t_z คือ ระยะการเลื่อนจากจุดกำเนิดในแต่ละแกน



4. วิธีการที่นำเสนอ

วิธีการประกอบรวมโครงร่างมนุษย์ที่ได้จากการตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect หลายตัว ประกอบด้วย 2 ขั้นตอน คือ การปรับข้อมูลตำแหน่งของข้อต่อ และการรวมข้อมูลตำแหน่งของข้อต่อ มีรายละเอียดดังนี้

4.1. การปรับข้อมูลตำแหน่งของข้อต่อ

การปรับข้อมูลตำแหน่งของข้อต่อให้มีตำแหน่งอ้างอิงเดียวกัน มีรายละเอียดดังนี้

- 1) กำหนดตำแหน่งใดๆ บนพื้นที่ที่ทำการตรวจจับ เป็นตำแหน่งสมมุติที่จะวางจุดกำเนิดของ Global Coordinate System (GCS) เพื่อให้กล้องทุกตัวใช้อ้างอิงตำแหน่งของกล้องจากจุดนี้ โดยทิศทางของแต่ละแกนของ GCS จะใช้ทิศทางเช่นเดียวกันกับแต่ละแกนของกล้อง Kinect ตามที่แสดงในรูปที่ 1
- 2) นำกล้องแต่ละตัวที่จะทำการตรวจจับมาวางที่จุดกำเนิด GCS โดยจัดวางให้จุดกำเนิดและทิศทางของแต่ละแกนของกล้องตรงกันกับจุดกำเนิดและทิศทางของแต่ละแกนของ GCS
- 3) ทำการหมุนกล้องตามเข็มนาฬิกาไปยังทิศทางที่ต้องการให้กล้องตรวจจับ และทำการวัดขนาดมุมที่ใช้ในการหมุนกล้องตามแกน X, Y และ Z ของจุดกำเนิด GCS ซึ่งจะได้อ่า a_x , a_y และ a_z โดยค่า a_x จะเป็นขนาดมุมที่เกิดจากการก้มหรือยกกล้อง ค่า a_y จะเป็นขนาดมุมที่เกิดจากการหันกล้อง และค่า a_z จะเป็นขนาดมุมที่เกิดจากการเอียงกล้อง
- 4) ทำการย้ายกล้องไปยังตำแหน่งที่ตั้งที่ต้องการให้กล้องตรวจจับ และทำการวัดระยะห่างตามแนวแกน X, Y และ Z ของจุดกำเนิด GCS ไปยังตำแหน่งจุดกำเนิดของกล้อง ณ ตำแหน่งที่ตั้ง ซึ่งจะได้ค่า t_x , t_y และ t_z หากการย้ายนั้นเป็นการย้ายที่ข้อนทิศทางค่าบวกของแกน X, Y และ Z ของจุดกำเนิด GCS ระยะที่วัดได้ให้มีค่าติดลบ
- 5) นำค่า a_x , a_y , a_z , t_x , t_y และ t_z ที่วัดได้มาแทนค่าในสมการที่ 4 เพื่อสร้างเมตริกซ์การแปลงของกล้องตัวนั้นๆ เมื่อนำค่าตำแหน่งของแต่ละข้อต่อที่กล้องตัวนั้นตรวจจับได้มาคูณกับเมตริกซ์การแปลงตามที่แสดงในสมการที่ 5 จะทำให้ได้ตำแหน่งของข้อต่อที่มีค่าอ้างอิงจากจุดกำเนิด GCS โดยมีค่า X, Y และ Z ตามที่แสดงในสมการที่ 7, 8 และ 9 ตามลำดับ

$$P' = [\text{transform}] \cdot P \tag{5}$$

$$\begin{bmatrix} P'_X \\ P'_Y \\ P'_Z \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_X \\ r_{21} & r_{22} & r_{23} & t_Y \\ r_{31} & r_{32} & r_{33} & t_Z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} P_X \\ P_Y \\ P_Z \\ 1 \end{bmatrix} \tag{6}$$

$$P'_X = r_{11} P_X + r_{12} P_Y + r_{13} P_Z + t_X \tag{7}$$

$$P'_Y = r_{21} P_X + r_{22} P_Y + r_{23} P_Z + t_Y \tag{8}$$

$$P'_Z = r_{31} P_X + r_{32} P_Y + r_{33} P_Z + t_Z \tag{9}$$

เมื่อ P คือ จุดพิกัดของข้อต่อใดๆ

P' คือ จุดพิกัดของข้อต่อนั้นหลังทำการปรับระบบพิกัด

4.2. การรวมข้อมูลตำแหน่งของข้อต่อ

หลังจากทำการปรับค่าตำแหน่งของทุก ข้อต่อที่ได้จากกล้องแต่ละตัวให้มีตำแหน่งอ้างอิงเดียวกัน คือตำแหน่งของจุดกำเนิด GCS แล้วค่าตำแหน่งของข้อต่อที่ได้จากกล้องแต่ละตัวจะมีค่าใกล้เคียงกัน จากนั้นจะนำค่าตำแหน่งของแต่ละข้อต่อจากกล้องทุก ตัว ที่ฟิลด์ iConfidence มีค่าเป็น 1 ในขณะนั้นๆ มาคำนวณหาค่าเฉลี่ย ตามที่แสดงในสมการที่ 10 ซึ่งจะทำได้ค่าตำแหน่งเฉลี่ยของข้อต่อนั้น นำมาใช้เป็นค่าตำแหน่งของข้อต่อที่ประกอบกันเป็นโครงร่างมนุษย์ชิ้นใหม่

$$Q_{(x,y,z)} = \left(\frac{1}{n} \sum_{i=1}^n P'_{X_i}, \frac{1}{n} \sum_{i=1}^n P'_{Y_i}, \frac{1}{n} \sum_{i=1}^n P'_{Z_i} \right) \tag{10}$$

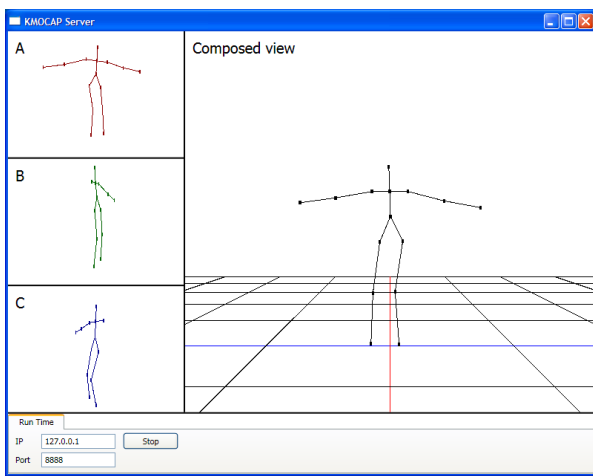
เมื่อ Q คือ จุดพิกัดเฉลี่ยของข้อต่อใดๆ ที่คำนวณจากจุดพิกัดของข้อต่อนั้นหลังจากทำการปรับระบบพิกัดตามกล้องแต่ละตัวแล้ว ที่ฟิลด์ iConfidence มีค่าเป็น 1

5. ระบบทางกายภาพและการเชื่อมต่อ

เนื่องจากข้อจำกัดในการใช้งานกล้อง Kinect ในขณะนี้ สามารถต่อและใช้งานได้เพียงหนึ่งกล้องต่อหนึ่ง USB Root Hub และคอมพิวเตอร์ส่วนใหญ่มีเพียงหนึ่ง USB Root Hub เช่นกัน จึงออกแบบระบบการตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect หลายตัว ให้เป็นแบบ Client – Server ตามที่แสดงในรูปที่ 3 โดย Client จะเป็นคอมพิวเตอร์ที่ต่อเข้ากับกล้อง Kinect และมีโปรแกรมที่ส่งข้อมูลการตรวจจับการเคลื่อนไหวที่ตรวจจับได้ไปยัง Server ที่เป็นคอมพิวเตอร์ที่มีโปรแกรมรับข้อมูลการตรวจจับจากกล้องแต่ละตัว และทำการประกอบรวมข้อมูลตำแหน่งของข้อต่อให้เป็นโครงร่างมนุษย์ตามเวลาจริง การออกแบบระบบให้เป็นแบบ Client – Server นี้ ทำให้สามารถเพิ่มจำนวน Client ซึ่งเป็นการเพิ่มจำนวนกล้องที่ใช้ในการตรวจจับได้ไม่จำกัด



รูปที่ 4 ตัวอย่างโปรแกรมส่วน Client



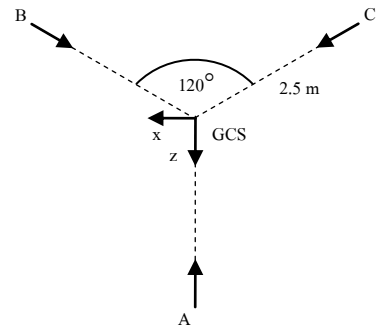
รูปที่ 5 ตัวอย่างโปรแกรมส่วน Server

ตารางที่ 1 ค่าของตัวแปรที่ใช้ในการสร้างเมตริกซ์การแปลง

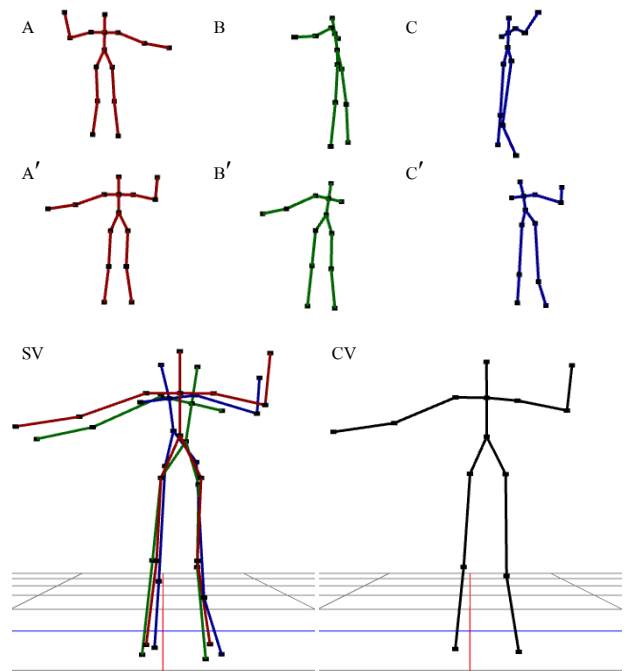
ตัวแปร	t_x	t_y	t_z	a_x	a_y	a_z
กล้อง A	0	0.6	2.50	0	180	0
กล้อง B	2.165	0.6	-1.25	0	300	0
กล้อง C	-2.165	0.6	-1.25	0	60	0

6. การทดสอบและผลการทดสอบ

งานวิจัยนี้ทำการทดสอบประสิทธิภาพของวิธีการและระบบที่นำเสนอ โดยการพัฒนาโปรแกรมทั้งฝั่ง Client และ Server ขึ้นมาด้วยภาษา C# ใช้ OpenNI เวอร์ชัน 1.5.2.7 ตามที่แสดงในรูปที่ 4 และ 5 และสร้างสถานการณ์การตรวจจับการเคลื่อนไหวขึ้นมา โดยใช้กล้อง Kinect สามตัวในการตรวจจับ ตั้งกล้องอยู่บนระนาบที่ขนานกับระนาบพื้นพื้นที่ทำการตรวจจับ สูงจากระนาบพื้น 0.6 เมตร ห่างออกไป 2.5 เมตร จากจุดกำเนิด GCS ซึ่งเป็นจุดกึ่งกลางพื้นที่ที่ทำการตรวจจับ และหันกล้องเข้าสู่ศูนย์กลาง โดยให้แกน Z ของแต่ละกล้องทำมุม 120 องศาต่อกัน ตามที่แสดงในรูปที่ 6 ค่าของตัวแปร t_x , t_y , t_z , a_x , a_y และ a_z ที่ใช้ในการสร้างเมตริกซ์การแปลงของกล้องแต่ละตัว จะมีค่าตามที่แสดงในตารางที่ 1



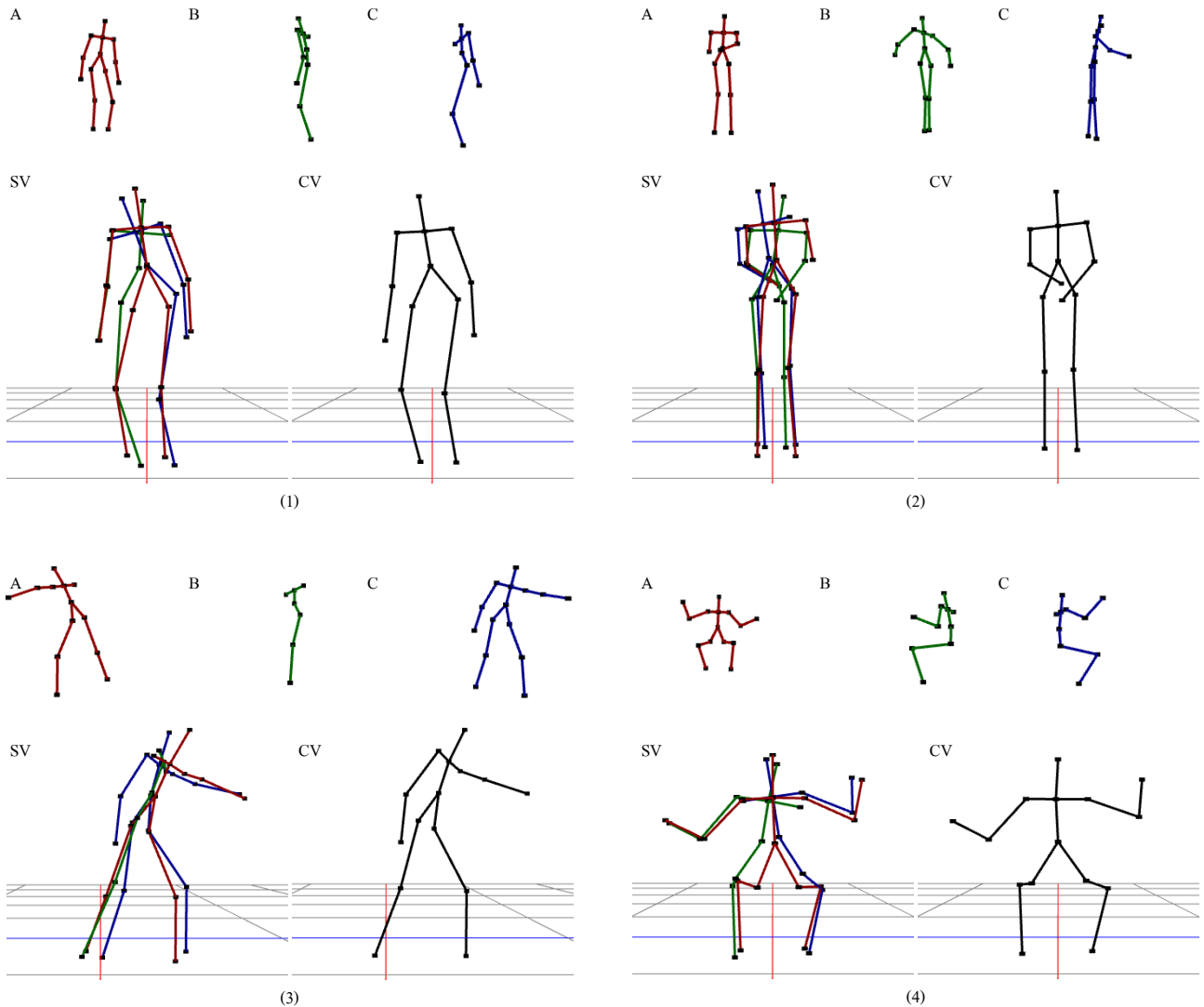
รูปที่ 6 แผนผังการจัดวางกล้องในการตรวจจับการเคลื่อนไหว



รูปที่ 7 โครงร่างมนุษย์ที่ได้จากการตรวจจับ แสดงตามแต่ละขั้นตอน

ตัวอย่างของโครงร่างมนุษย์ที่ได้จากสถานการณ์การตรวจจับการเคลื่อนไหวที่จัดทำขึ้นเมื่อผ่านขั้นตอนตามวิธีการที่นำเสนอ แสดงในรูปที่ 7 โดยแถวแรกจะเป็นภาพโครงร่างมนุษย์ที่กล้องแต่ละตัวตรวจจับได้จากตำแหน่งที่กล้องตั้ง โดยแสดงเฉพาะข้อต่อที่ฟิลด์ fConfidence มีค่าเท่ากับ 1 เท่านั้น แถวที่สองจะเป็นภาพโครงร่างมนุษย์ของแต่ละกล้องหลังจากที่ทำการปรับข้อมูลตำแหน่งของข้อต่อให้ใช้จุดกำเนิด GCS เป็นตำแหน่งอ้างอิงเดียวกันแล้ว (ภาพ A', B' และ C') และแถวสุดท้ายจะเป็นภาพของโครงร่างมนุษย์ที่ทำการปรับข้อมูลตำแหน่งของข้อต่อแล้วนำมาแสดงซ้อนทับกัน (Superimposed View : SV) และภาพโครงร่างมนุษย์ที่สร้างขึ้นจากตำแหน่งของข้อต่อเฉลี่ย (Composed View : CV)

ผลของวิธีการประกอบรวมโครงร่างมนุษย์ที่นำเสนอจากสถานการณ์การตรวจจับการเคลื่อนไหวที่จัดทำขึ้น แสดงในรูปที่ 8 โดยรูปที่ 8 (4) เป็นการตรวจจับเมื่อผู้ตรวจจับนั่งอยู่บนเก้าอี้



รูปที่ 8 โครงร่างมนุษย์ที่กล้องแต่ละตัวตรวจจับได้ และ โครงร่างมนุษย์ประกอบรวม

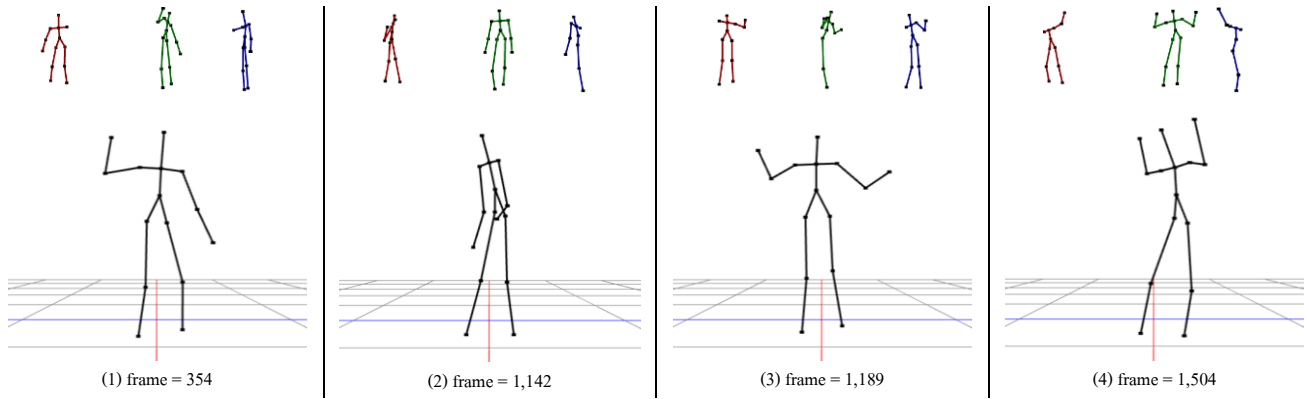
ตัวอย่างของการตรวจจับการเคลื่อนไหวแบบต่อเนื่องแสดงในรูปที่ 9 ซึ่งแถวแรกจะเป็นภาพโครงร่างมนุษย์ที่กล้องแต่ละตัวตรวจจับได้จากตำแหน่งที่กล้องตั้ง แสดงเรียงลำดับตามกล้อง A, B และ C แถวที่สองจะเป็นภาพโครงร่างมนุษย์ประกอบรวม การตรวจจับการเคลื่อนไหวนี้มีทั้งหมด 1,700 เฟรม เป็นเวลาประมาณ 1 นาที หากนำค่าตำแหน่งเฉลี่ยและค่าตำแหน่งที่กล้องแต่ละตัวตรวจจับได้หลังจากทำการปรับระบบพิกัดมาเปรียบเทียบกัน โดยใช้วิธีการคำนวณหาระยะห่างของค่าตำแหน่งระหว่างกัน ด้วยสมการที่ 11 ค่าเฉลี่ยระยะห่างของค่าตำแหน่งระหว่างกันของแต่ละข้อต่อจากการตรวจจับการเคลื่อนไหวนี้ แสดงในตารางที่ 2 ระยะห่างเฉลี่ยของค่าตำแหน่งระหว่างกันที่มากที่สุดในการตรวจจับนี้คือที่ตำแหน่ง RightShoulder ของกล้อง C มีระยะห่างเฉลี่ยประมาณ 11.0 เซนติเมตร และระยะห่างเฉลี่ยของค่าตำแหน่งระหว่างกันที่น้อยที่สุดในการตรวจจับนี้คือที่ตำแหน่ง LeftFoot ของกล้อง B มีระยะห่างเฉลี่ยประมาณ 4.2 เซนติเมตร

$$d = \sqrt{(Q_x - P_x)^2 + (Q_y - P_y)^2 + (Q_z - P_z)^2} \quad (11)$$

เมื่อ d คือ ระยะห่างระหว่างสองจุดใดๆ ในสามมิติ

7. สรุปและแนวทางพัฒนาต่อไป

งานวิจัยนี้นำเสนอวิธีการการประกอบรวม โครงร่างมนุษย์จากการตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect หลายตัว และนำเสนอระบบการตรวจจับการเคลื่อนไหวตามเวลาจริงที่รองรับกล้อง Kinect จำนวนมากได้จากผลของวิธีการที่นำเสนอแสดงให้เห็นว่า หลังจากทำการปรับระบบพิกัดของตำแหน่งของข้อต่อที่ได้จากการตรวจจับจากกล้องแต่ละตัวแล้วสามารถใช้ค่าเฉลี่ยของตำแหน่งของข้อต่อในการประกอบรวมโครงร่างมนุษย์ได้ และการใช้กล้องหลายตัวช่วยเพิ่มโอกาสให้ทุกข้อต่อถูกตรวจจับได้และมีค่าตำแหน่งของข้อต่ออยู่ตลอดเวลา



รูปที่ 9 โครงร่างมนุษย์ที่กล้องแต่ละตัวตรวจจับได้ และโครงร่างมนุษย์ประกอบรวม ในบางเฟรมจากการตรวจจับการเคลื่อนไหวต่อเนื่องเดียวกัน

ตารางที่ 2 ระยะเวลาเฉลี่ยระหว่างค่าตำแหน่งเฉลี่ยกับค่าตำแหน่งที่ได้จากกล้องแต่ละตัว หลังจากทำการปรับระบบพิกัดของแต่ละข้อต่อจากการตรวจจับการเคลื่อนไหวในรูปที่ 9

ข้อต่อ	ระยะเวลาเฉลี่ย (เซนติเมตร)		
	กล้อง A	กล้อง B	กล้อง C
Head	8.1	7.8	8.0
Neck	5.0	5.5	5.8
Torso	5.0	5.0	5.5
LeftHand	5.6	6.3	6.5
LeftElbow	4.5	6.3	6.0
LeftShoulder	6.3	5.5	6.1
RightHand	5.5	6.2	6.5
RightElbow	8.7	6.3	7.3
RightShoulder	10.8	9.9	11.0
LeftHip	8.5	6.7	7.7
LeftKnee	7.5	6.4	5.1
LeftFoot	5.0	4.2	4.3
RightHip	7.7	7.7	8.4
RightKnee	9.4	6.8	5.7
RightFoot	4.3	5.3	6.1

เนื่องจากความถูกต้องของตำแหน่งของข้อต่อจะขึ้นอยู่กับ การวัดค่าตำแหน่งและทิศทางที่ตรวจจับของกล้องแต่ละตัว และ API ที่ใช้ในการตรวจจับการเคลื่อนไหว แนวทางการพัฒนาต่อไปคือ การพัฒนาวิธีการหาตำแหน่งและทิศทางที่ตรวจจับของกล้องแต่ละตัวที่ใช้ในการตรวจจับการเคลื่อนไหวเดียวกันโดยอัตโนมัติ การพัฒนาส่วนเสริมของ API เพื่อคัดกรองและเพิ่มความถูกต้องของตำแหน่งของข้อต่อที่ได้จากการตรวจจับการเคลื่อนไหวก่อนนำไปทำการประกอบรวมโครงร่างมนุษย์ และทำการศึกษาการใช้กล้อง Kinect ในจำนวนที่เหมาะสมกับการตรวจจับการเคลื่อนไหวในด้านต่างๆ

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับการสนับสนุนจากโครงการส่งเสริมการวิจัยในอุดมศึกษา และการพัฒนามหาวิทยาลัยวิจัยแห่งชาติ (NRU) ของสำนักงานคณะกรรมการการอุดมศึกษา ขอขอบคุณมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรีที่สนับสนุนทุนเพชรพระจอมเกล้า “ทุนส่งเสริมด้านวิชาการดีเด่น” และขอขอบคุณคณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรีที่ให้การสนับสนุนงานวิจัยชิ้นนี้

เอกสารอ้างอิง

[1] Microsoft. “Kinect for Xbox 360 - Xbox.com”. www.xbox.com/en-GB/kinect, 2012.

[2] OpenNI. “Introducing OpenNI”. www.openni.org, 2012.

[3] OpenNI. “Downloads”. www.openni.org/Downloads.aspx, 2012.

[4] D. Alexiadis, P. Kelly, P. Daras, N. E. O'Connor, T. Boubekeur, and M. B. Moussa, “Evaluating a Dancer's Performance using Kinect-based Skeleton Tracking”. *Proceedings of the 19th ACM international conference on Multimedia*, pp. 659-662, Nov. 2011.

[5] K. Berger, K. Ruhl, Y. Schroeder, C. Brummer, Al. Scholz, and M. Magnor, “Markerless Motion Capture using multiple Color-Depth Sensors”. *Proc. Vision, Modeling and Visualization (VMV) 2011*, pp. 317-324, Oct. 2011.

[6] J. Vince, *Mathematics for Computer Graphics*, 3rd ed. London: Springer, 2010.

[7] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*, Springer, 2008.