

การพัฒนาาระบบฝึกภาษามือไทยด้วยอุปกรณ์เซนเซอร์เพื่อตรวจสอบการเคลื่อนไหวไคนค

The Development of Thai Sign Language Training System using Kinect

วนารัตน์ จุฬพันธ์ทอง สุรเดช จิตประไพกุลศาล และไพศาล มุณีสว่าง*

ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

มหาวิทยาลัยนเรศวร พิษณุโลก 65000

Email: wanaratj15@gmail.com, suradetj@gmail.com, *paisarnmu@nu.ac.th

บทคัดย่อ: ระบบฝึกภาษามือไทยด้วยอุปกรณ์ไคนคพัฒนาขึ้นสำหรับผู้ที่ต้องการฝึกการใช้ท่าทางในภาษามือไทย เพื่อช่วยให้ผู้ฝึกมีการเคลื่อนไหวของมือและแขน ได้ถูกต้องตามท่าทางของผู้เชี่ยวชาญมากยิ่งขึ้น การทำงานของระบบประกอบด้วยการวิเคราะห์การเคลื่อนไหวจากเซนเซอร์ของไคนค การรู้จำท่าทางภาษามือต่อเนื่องโดยประยุกต์ใช้จากหลักการ SOFM Tf-Idf และ Bayesian และแสดงผลเป็นแอนิเมชันสามมิติเปรียบเทียบระหว่างผู้ฝึกและผู้เชี่ยวชาญพร้อมคะแนนเพื่อประเมินความถูกต้องในการแสดงท่าทาง การทดลองขั้นต้นพบว่ามีความถูกต้องในการรู้จำท่าทาง 64.81% สามารถรู้จำท่าทางเมื่อแสดงท่าทางภาษามือไทยแบบต่อเนื่อง ให้คะแนนความถูกต้องในการแสดงท่าทางภาษามือไทยของผู้ฝึก และแสดงผลเป็นสามมิติได้

คำสำคัญ: ภาษามือไทย, ไคนค

Abstract: Thai sign language (TSL) training system with Kinect is developed for the beginner of Thai sign language in training in order to improve trainee movement more correctly. The system framework includes gesture detection from Kinect motion sensor, TSL gesture recognition using SOFM, Tf-Idf and Bayesian, and TSL gesture assessment. The primary experimental result shows the average accuracy of recognition is 64.81%. The system is able to recognize the continuing gesture and give the assessment score of the performance. It also provide visualize 3D animation of user language action.

Keywords: Thai Sign Language, Kinect

1. บทนำ

ในการใช้ภาษามือเพื่อการสื่อสารสำหรับผู้ที่มีปัญหาทางการได้ยินในประเทศไทย พบว่าแต่ละท้องถิ่นมีการใช้งานเฉพาะกลุ่ม การใช้ท่าทางในการสื่อความหมายที่ต่างกันในนี้อาจทำให้เกิดความไม่เข้าใจกันระหว่างการสื่อสารได้ การกำหนดภาษามือไทยจึงเป็นสิ่งจำเป็น โดย กสทช. (2555) ได้พัฒนาพจนานุกรมอิเล็กทรอนิกส์ภาษามือไทยตามพื้นฐานธรรมชาติภาษามือไทย โดยรวบรวมคำที่ใช้สื่อสารในชีวิตประจำวัน เพื่อช่วยให้มีการใช้ท่าทางการสื่อสารภาษามือไทยที่ตรงกัน

เพื่อที่จะสามารถสื่อสารภาษามือไทยได้อย่างมีประสิทธิภาพ ทำให้ต้องมีการเรียนรู้ท่าทางภาษามือไทยที่ตรงกัน สำหรับผู้ที่ต้องการเรียนรู้การใช้ท่าทางภาษามือไทย ไม่ว่าจะเป็นผู้ที่เคยใช้ภาษามือมาก่อนหรือผู้ที่เริ่มต้นฝึกก็ตาม พบว่ามีปัญหาสำคัญที่อาจเกิดขึ้น คือความกังวลเกี่ยวกับการเคลื่อนไหวของร่างกายเพื่อแสดงท่าทางที่ถูกต้อง ซึ่งเป็นอีกความท้าทายหนึ่งที่ผู้วิจัยให้ความสนใจ โดยผู้วิจัยมีแนวคิดในการพัฒนาระบบฝึกภาษามือไทย เพื่อช่วยประเมินความถูกต้องของผู้ฝึกในการใช้มือและแขนเคลื่อนไหวท่าทางในการสร้างคำเพื่อสื่อความหมาย ผู้วิจัยได้นำเทคโนโลยีของอุปกรณ์ไคเนมาช่วยตรวจจับความเคลื่อนไหว โดยระบบต้นแบบที่สร้างขึ้นจะทำการวิเคราะห์การเคลื่อนไหวของผู้ฝึกโดยอัตโนมัติ รู้จำท่าทางของภาษามือไทย ให้คะแนนความถูกต้องในท่าทางของผู้ฝึกเปรียบเทียบกับฐานข้อมูลของผู้เชี่ยวชาญ และแสดงผลออกมาเป็นแอนิเมชันสามมิติ

2. ทบทวนวรรณกรรม

ในการศึกษาระบบที่เกี่ยวข้องกับภาษามือ พบว่ามีการใช้อุปกรณ์เพื่อรับข้อมูลหลายประเภท เช่น การแปลภาษามือโดยใช้ถุงมืออิเล็กทรอนิกส์ (Kau et al., 2008) ข้อดีของการใช้อุปกรณ์นำเข้านี้คือสามารถตรวจจับความเคลื่อนไหวของนิ้วได้ ทำให้การจำแนกท่าทางที่แตกต่างกันเพียงการแสดงนิ้วมือ สามารถทำได้ดี อย่างไรก็ตามถุงมือนั้นมักเป็นอุปกรณ์ที่พัฒนาเฉพาะขึ้นมา ทำให้การนำอุปกรณ์นี้ไปใช้จริงได้ยากกว่าอุปกรณ์นำเข้าอื่น ดังนั้นจึงมีงานวิจัยที่นำเสนอการรู้จำภาษามือโดยใช้กล้องถ่ายภาพ (Kumarage et al., 2011) ซึ่งเป็นอุปกรณ์ที่มีอยู่ทั่วไป โดยข้อมูลที่นำเข้านั้นมีทั้งเป็นภาพนิ่งและภาพต่อเนื่อง ข้อเสียของอุปกรณ์ประเภทนี้คือการนำข้อมูลไปประมวลผลเพื่อหาการเคลื่อนไหวในเชิงลึกทำได้ยาก การรู้จำภาษามือด้วยไคเนค (Tiwari et al., 2015) ซึ่งเป็นกล้องที่สามารถตรวจจับการเคลื่อนไหวในเชิงลึกได้ จึงเป็นอุปกรณ์หนึ่งในปัจจุบันที่นิยมนำมาใช้ในการรู้จำท่าทาง ไม่เฉพาะเพียงภาษามือเท่านั้น เช่น การรู้จำท่าทางเพื่อใช้ในระบบ

การฝึกเดินบิลเล่ (Kyan et al, 2015) ระบบช่วยฟื้นฟูการเคลื่อนไหวของผู้ป่วย (Shapi et al., 2015) เป็นต้น

3. การดำเนินการวิจัย

3.1. อุปกรณ์ไคเนค

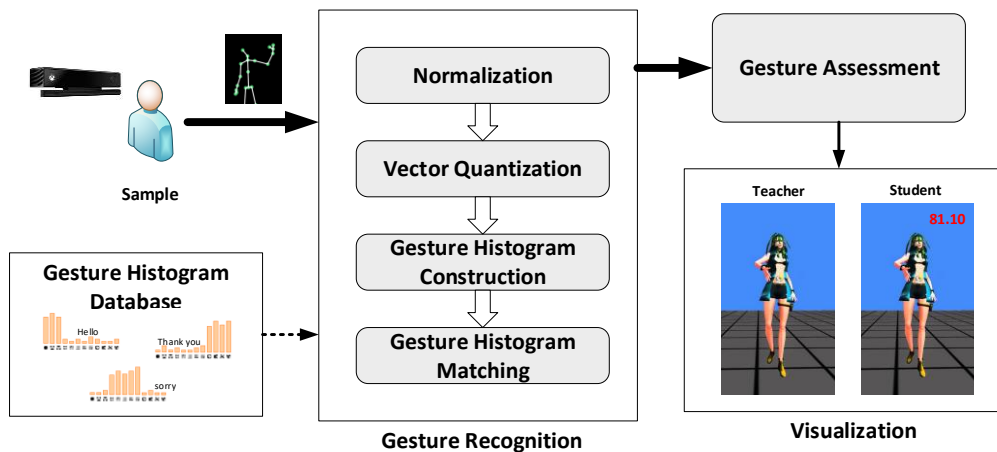
ไคเนค (Microsoft, 2016) เป็นอุปกรณ์ประเภทกล้องที่พัฒนาโดยบริษัทไมโครซอฟต์ ภายในไคเนคประกอบด้วยอุปกรณ์ฉายแสงอินฟราเรด กล้องวัดความลึกของภาพ กล้องวิดีโอ ไมโครโฟน และเซ็นเซอร์ประมวลผล เซ็นเซอร์ของไคเนคสามารถตรวจจับความเคลื่อนไหวของผู้ใช้และแสดงผลเป็นการเคลื่อนไหวของข้อต่อของมนุษย์ในลักษณะของโครงกระดูก (Skeleton) ซึ่งปัจจุบันไคเนคเวอร์ชัน 2 สามารถตรวจจับความเคลื่อนไหวและแสดงผลผ่านข้อต่อละเอียดถึง 25 ข้อต่อ และรองรับการเชื่อมต่อกับระบบปฏิบัติการวินโดวส์ 8 ขึ้นไปผ่าน USB 3.0 รูปที่ 1 แสดงอุปกรณ์ไคเนค



รูปที่ 1 อุปกรณ์ไคเนคเวอร์ชัน 2 (Microsoft, 2016)

3.2. กรอบแนวคิดของระบบ

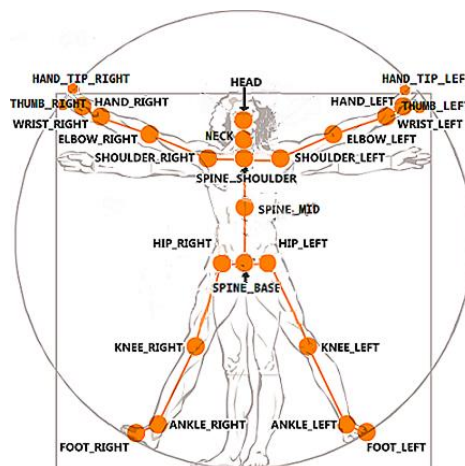
ระบบฝึกภาษาไทยด้วยอุปกรณ์ไคเนคได้รับข้อมูลนำเข้าเป็นลักษณะข้อมูลแบบเวกเตอร์ซึ่งระบุตำแหน่งของข้อต่อ 25 ส่วน จากนั้นระบบจะใช้ข้อมูลเพื่อคาดการณ์ท่าทางภาษามือของผู้ใช้ผ่านกระบวนการรู้จำท่าทาง โดยใช้ข้อมูลท่าทางภาษามือของผู้เชี่ยวชาญที่จัดเก็บในฐานข้อมูล สุดท้ายระบบประเมินความถูกต้องของการทำท่าทาง โดยเปรียบเทียบท่าทางภาษามือระหว่างผู้ใช้และผู้เชี่ยวชาญ สรุปผลเป็นคะแนนและแสดงผลผ่านแอนิเมชันสามมิติ กรอบแนวคิดของระบบแสดงดังรูปที่ 2



รูปที่ 2 กรอบแนวคิดของระบบ

3.3. การตรวจจับการเคลื่อนไหวของผู้ใช้

การตรวจจับการเคลื่อนไหวของผู้ใช้ ใช้ข้อมูลตำแหน่งของข้อต่อทั้ง 25 ส่วน ในแกน x y และ z ที่ประมวลผลมาจากเซนเซอร์ของไคเนค โดยระบบจะเก็บข้อมูลตำแหน่งที่เปลี่ยนแปลงไปของแต่ละข้อต่อตามการเคลื่อนไหวของผู้ใช้ เพื่อนำไปวิเคราะห์ในกระบวนการต่อไป ข้อต่อทั้ง 25 ส่วนของไคเนคเวอร์ชัน 2 แสดงในรูปที่ 3



รูปที่ 3 ข้อต่อ 25 ส่วนในไคเนคเวอร์ชัน 2 (Microsoft, 2016)

3.4. การรู้จำท่าทางภาษามือ

3.4.1. Normalization

ขณะที่ผู้ใช้แสดงท่าทางของภาษามือ ระบบจะเก็บข้อมูลตำแหน่งของข้อต่อทั้ง 25 ส่วนต่อเฟรม จากเริ่มต้นจนเสร็จสิ้นการทำท่าทาง จากนั้นนำข้อมูลผ่านกระบวนการ Normalization

โดยให้ข้อต่อส่วนฐานกระดูกสันหลัง (Spine Base) เป็นตำแหน่งอ้างอิง เพื่อนำไปใช้ในการเปรียบเทียบประสิทธิภาพการรู้จำ ข้อมูลตำแหน่งต่อเฟรม จัดเก็บในรูปแบบเวกเตอร์ในสมการที่ 1

$$\bar{x} = \{x_i, i = 1, \dots, 75\} \quad (1)$$

โดย x_i คือตำแหน่งของข้อต่อ โดยจัดเก็บทั้ง 25 ข้อต่อในแกน x y และ z

3.4.2. Vector Quantization

ในกระบวนการ Vector Quantization ผู้วิจัยสร้างคลาสของข้อมูลโดยใช้หลักการ SOFM (Self-Organizing Feature Maps) จากอัลกอริทึมของ Bashir (2006) โดย SOFM จะเรียนรู้ผ่านชุดข้อมูลการแสดงท่าทางภาษามือของผู้เชี่ยวชาญ และกำหนดคลาส (นิวรอน) ของค่าที่สำคัญของข้อมูลที่เกิดขึ้นในท่าทางภาษามือ จากนั้นทำการ Vector Quantization โดยเลือกคลาสจาก SOFM ที่เหมาะสมกับข้อมูล และสร้างชุดลำดับคลาสสำหรับแต่ละท่าทางของภาษามือ โดย S_c คือชุดลำดับคลาสสำหรับท่าทางภาษามือที่ c ดังสมการที่ 2

$$S_c = \{n_i, i = 1, \dots, f_c\} \quad (2)$$

เมื่อ n_i คือ คลาสที่เฟรม i โดยแต่ละท่าทางภาษามือมีจำนวน f_c เฟรม

3.4.3. Gesture Histogram Construction

ชุดลำดับคลาสที่ได้จากกระบวนการก่อนหน้า จะนำมาสร้างเป็นฮิสโตแกรม โดยผู้วิจัยได้แบ่งวิธีการสร้างออกเป็น 4 แบบ ดังนี้

1. BoW : ใช้หลักการของ bag-of-words โดยนับความถี่ของแต่ละคลาสที่เกิดขึ้น ฮิสโตแกรมของ BoW แสดงดังสมการที่ 3 โดย $H_{c,n}$ คือ จำนวนความถี่ของคลาสที่ n ในท่าทางภาษามือที่ c

$$BoW_c = H_{c,n}, 1 \leq n \leq N \quad (3)$$

2. Sparse Code (SC) : คล้ายวิธีการแบบ BoW แต่จะพิจารณาการเกิดขึ้นของคลาส BoW โดยฮิสโตแกรมของ SC แสดงดังสมการที่ 5 ซึ่งใช้ค่า $H_{c,n}$ เช่นเดียวกับสมการที่ 3

$$C_{c,n} = \begin{cases} 1, & \text{if } H_{c,n} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

$$SC_c = C_{c,n}, 1 \leq n \leq N \quad (5)$$

3. Transition (T) : ใช้หลักการ bag-of-segments โดยพิจารณาความถี่ของการส่งผ่านของคลาส จากชุดลำดับคลาสในเฟรมที่ i ไป j ฮิสโตแกรมของ T แสดงดังสมการที่ 6 โดย $H_{c,n(i \rightarrow j)}$ คือ จำนวนความถี่ของคลาสที่ n ในท่าทางภาษามือที่ c จากชุดลำดับคลาสในเฟรมที่ i ไป j

$$T_c = H_{c,n(i \rightarrow j)}, 1 \leq n \leq N, 1 \leq i < j \leq f_c \quad (6)$$

4. Transition Sparse Code (TSC) : คล้ายวิธีการแบบ T แต่จะการใช้การพิจารณาการเกิดขึ้นของการส่งผ่านของคลาส ฮิสโตแกรมของ TSC แสดงในสมการที่ 8 ซึ่งใช้ค่า $H_{c,n(i \rightarrow j)}$ เช่นเดียวกับสมการที่ 6

$$TS_{c,n(i \rightarrow j)} = \begin{cases} 1, & \text{if } H_{c,n(i \rightarrow j)} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$TSC_c = TS_{c,n(i \rightarrow j)}, 1 \leq n \leq N, 1 \leq i < j \leq f_c \quad (8)$$

อย่างไรก็ตามในการทดสอบที่ 4.3 พบว่าท่าทางภาษามือที่ค่อนข้างใกล้เคียงกันจะมีค่าคลาสในฮิสโตแกรมที่คล้ายกัน ผู้วิจัยจึงประยุกต์ใช้หลักการ Tf-Idf (term frequency-inverse document frequency) เพื่อให้น้ำหนักค่าของคลาสที่สำคัญที่ปรากฏในแต่ละฮิสโตแกรม ซึ่งสามารถช่วยเพิ่มประสิทธิภาพการรู้จำท่าทางโดยรวมได้ดีขึ้น

3.4.4. Gesture Histogram Matching

กระบวนการจับคู่ฮิสโตแกรมเป็นขั้นตอนสุดท้ายของการรู้จำท่าทางภาษามือ เมื่อได้ฮิสโตแกรมของท่าทางภาษามือที่ผู้ใช้ให้ข้อมูลการเคลื่อนไหวผ่านไคเนค จากกระบวนการก่อนหน้าแล้ว ระบบจะใช้วิธีการ Histogram Intersection ดังสมการที่ 9 และคำนวณค่าความน่าจะเป็น

จากหลักการ Bayesian ดังสมการที่ 10 เพื่อคาดการณ์ท่าทางภาษามือของผู้ใช้ เทียบกับฐานข้อมูลฮิสโตแกรมที่สร้างจากท่าทางภาษามือของผู้เชี่ยวชาญ

$$HI(h_s, h_c) = \sum_n \min [h_{s,n}, h_{c,n}] \quad (9)$$

เมื่อ h_s คือ ฮิสโตแกรมของผู้ใช้ h_c คือฮิสโตแกรมของผู้เชี่ยวชาญในท่าทางภาษามือที่ c และ n คือ ค่าของคลาสที่ n ตามวิธีการสร้างฮิสโตแกรมในหัวข้อ 3.4.3

$$P_t(c | h_s) = \frac{P_t(h_s | c)P_t(c)}{\sum P_t(h_s | c)P_t(c)} \quad (10)$$

เมื่อ $P_t(c | h_s)$ คือ ความน่าจะเป็นที่ฮิสโตแกรมของผู้ใช้ h_s จะเป็นท่าทางภาษามือที่ c โดย $P_t(h_s | c)$ และ $P_t(c)$ คำนวณจากสมการที่ 11 และ 12

$$P_t(h_s | c) = HI(h_s, h_c) \quad (11)$$

$$P_t(c) = \begin{cases} \frac{1}{K} & \text{if } t = t_0 \\ \frac{P_{t-1}(c | h_s) \cdot HI(h_s, h_c)}{\sum_K P_{t-1}(c | h_s) \cdot HI(h_s, h_c)} & \text{otherwise} \end{cases} \quad (12)$$

ในขั้นต้นระบบจะเลือกท่าทางภาษามือที่มีค่าความน่าจะเป็นมากที่สุด

3.5. การแสดงผลและประเมินผลท่าทางภาษามือ

ท่าทางภาษามือจะแสดงผลเป็นแอนิเมชันสามมิติ โดยแสดงท่าทางภาษามือของผู้เชี่ยวชาญที่ระบบได้คาดการณ์ในกระบวนการรู้จำ และท่าทางภาษามือของผู้ฝึกที่ได้ให้ข้อมูลผ่านไคเนค โดยจะแสดงผลเปรียบเทียบกันพร้อมค่าคะแนนความถูกต้องในการแสดงท่าทางของผู้ฝึก

การประเมินความถูกต้องในการแสดงท่าทางของผู้ฝึกใช้การความแตกต่างตำแหน่งของข้อต่อทั้ง 25 ส่วนต่อเวลา ระหว่างผู้เชี่ยวชาญและผู้ฝึก โดยในขั้นต้นให้เวลาในการแสดงท่าทางของผู้เชี่ยวชาญและผู้ฝึกมีค่าเท่ากัน

4. ผลการศึกษาและอภิปรายผล

4.1. ทำทางภาษามือไทย

ทำทางภาษามือที่นำมาทดสอบ ผู้วิจัยเลือกภาษามือที่แสดงความหมายของคำหรือวลีจำนวน 9 ทำทาง ที่มีการเคลื่อนไหวไม่ซับซ้อนมากนัก ได้แก่ 1.สวัสดี 2.ขอบคุณ 3.ขอโทษ 4.บอกลา 5.โชคดี 6.ไม่เป็นไร 7.คิด 8.เข้าใจ และ 9.ฉลาด

4.2. ข้อมูลของกลุ่มตัวอย่าง

ในการสร้าง SOFM และฐานข้อมูล ผู้วิจัยเก็บข้อมูลทำทางภาษามือจากผู้เชี่ยวชาญหนึ่งท่าน จำนวนสิบชุดข้อมูลในแต่ละทำทางภาษามือ โดยในข้อมูลหนึ่งชุด จะให้ผู้เชี่ยวชาญยืนอยู่ด้านหน้าอุปกรณ์ไคเนคระยะ 1.5 และ 2 เมตร โดยยืนในมุม 45 90 และ 135 องศากับอุปกรณ์

ส่วนข้อมูลที่นำมาทดสอบ มาจากการรับข้อมูลผ่านไคเนคของสองผู้ใช้ ผู้ใช้ละหนึ่งชุดข้อมูลในแต่ละทำทางในการทดสอบ 4.3 และหนึ่งผู้ใช้ จำนวนหนึ่งชุดข้อมูลแสดงทำทางต่อเนื่องจำนวนเก้าทำทาง ซึ่งทดสอบภายในห้องปฏิบัติการที่มีแสงเพียงพอ

4.3. การทดสอบประสิทธิภาพในการรู้จำทำทาง

เพื่อหาประสิทธิภาพในการรู้จำทำทางภาษามือ ได้ทำการทดลองโดยให้กลุ่มตัวอย่างแสดงทำทางภาษามือตามที่กำหนดไว้ 9 ทำทางและเปรียบเทียบผลกับการสร้างฮิสโตแกรมทั้ง 4 แบบคือ Bag-of-Words Sparse Code Transition และ Transition Sparse Code ระบบจะคาดการณ์ทำทางภาษามือ โดยเลือกจากทำทางที่มีค่า HI ที่คำนวณในสมการที่ 9 มากที่สุด ผลลัพธ์ของความถูกต้องในการรู้จำทำทางแสดงดังตารางที่ 1

ค่าเฉลี่ยความถูกต้องในการรู้จำประสิทธิภาพทั้งหมด 9 ทำทาง พบว่าวิธีการแบบ BoW และ TSC มีประสิทธิภาพสูงสุดเท่ากับ 68.52% ส่วนวิธีการแบบ T และ SC มีค่าเฉลี่ยความถูกต้องเท่ากับ 62.96% และ 53.70% ตามลำดับ เช่นเดียวกันกับเมื่อใช้หลักการ Tf-Idf ในตารางที่ 2 พบว่าวิธีการแบบ BoW และ TSC ยังคงมีประสิทธิภาพสูงสุด

เมื่อพิจารณาค่าเฉลี่ยความถูกต้องแต่ละทำทางภาษามือ พบว่าทำทาง “โชคดี” มีค่าความถูกต้องที่ต่ำกว่าค่าเฉลี่ย เนื่องจากมีตำแหน่งในการเคลื่อนไหวของมือใกล้เคียงกับคำว่า “บอกลา” แสดงดังรูปที่ 4 นอกจากนี้ทำทาง “คิด” “เข้าใจ” และ “ฉลาด” มีค่าต่ำกว่าค่าเฉลี่ยเล็กน้อย เนื่องจากทั้ง 3 ทำทาง มีการเคลื่อนไหวของมือไม่มาก และลักษณะการเคลื่อนไหวค่อนข้างใกล้เคียงกัน อย่างไรก็ตามเมื่อได้เพิ่มหลักการ Tf-Idf เพื่อให้ค่าน้ำหนักของคลาสในฮิสโตแกรม พบว่าสามารถช่วยเพิ่ม

ประสิทธิภาพการรู้จำในท่าทางที่ใกล้เคียงกันมากขึ้น โดยสามารถรู้จำท่าทางของ “คิด” และ “ฉลาด” ได้เพิ่มขึ้น

ตารางที่ 1 ค่าเฉลี่ยความถูกต้องในการรู้จำท่าทางภาษามือ

ท่าทางภาษามือ	ค่าเฉลี่ยความถูกต้องในการรู้จำท่าทางภาษามือ (%)				
	BoW	SC	T	TSC	รวมทั้งหมด
1.สวัสดี	66.67	66.67	83.33	83.33	75
2.ขอบคุณ	100	50	83.33	83.33	79.17
3.ขอโทษ	83.33	83.33	83.33	83.33	83.33
4.บอกลา	83.33	83.33	83.33	83.33	83.33
5.โชคดี	16.67	33.33	16.67	16.67	20.84
6.ไม่เป็นไร	83.33	50	83.33	83.33	75
7.คิด	100	66.67	16.67	16.67	50
8.เข้าใจ	16.67	33.33	66.67	83.33	50
9.ฉลาด	16.67	33.33	50	83.33	45.83
รวมทั้งหมด	68.52	53.70	62.96	68.52	63.43

ตารางที่ 2 ค่าเฉลี่ยความถูกต้องในการรู้จำท่าทางภาษามือเมื่อใช้ Tf-Idf

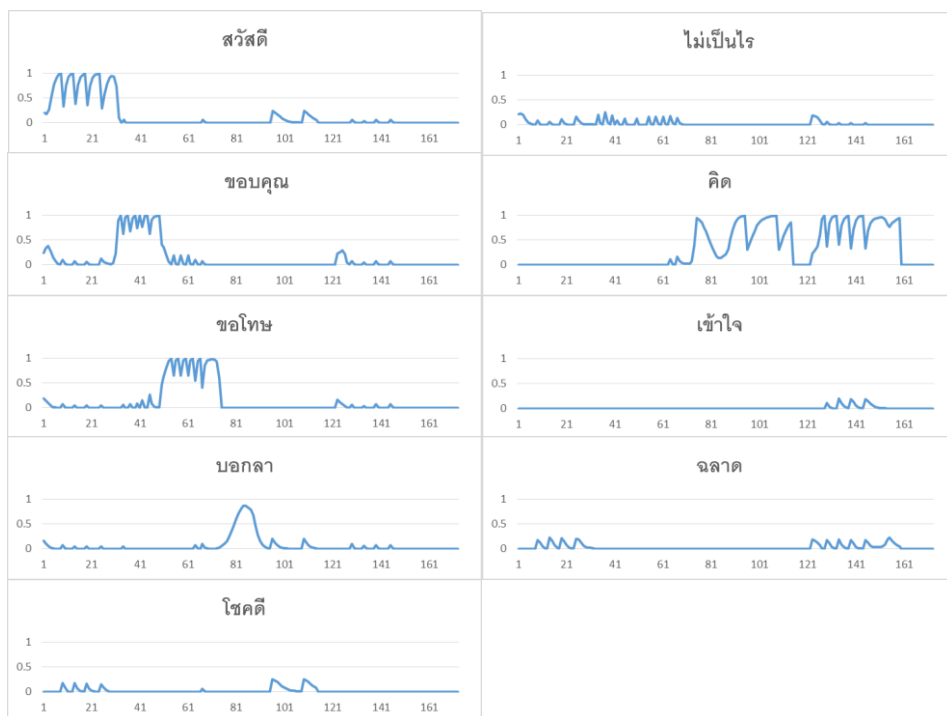
ท่าทางภาษามือ	ค่าเฉลี่ยความถูกต้องในการรู้จำท่าทางภาษามือ (%)				
	BoW	SC	T	TSC	รวมทั้งหมด
1.สวัสดี	83.33	83.33	66.67	83.33	79.17
2.ขอบคุณ	100	83.33	100	83.33	91.67
3.ขอโทษ	83.33	33.33	83.33	66.67	66.67
4.บอกลา	83.33	50	83.33	83.33	75
5.โชคดี	33.33	0	0	33.33	16.67
6.ไม่เป็นไร	83.33	83.33	83.33	100	87.5
7.คิด	66.67	66.67	66.67	66.67	66.67
8.เข้าใจ	16.67	16.67	33.33	50	29.17
9.ฉลาด	83.33	66.67	66.67	66.67	70.84
รวมทั้งหมด	70.37	53.70	64.81	70.37	64.81



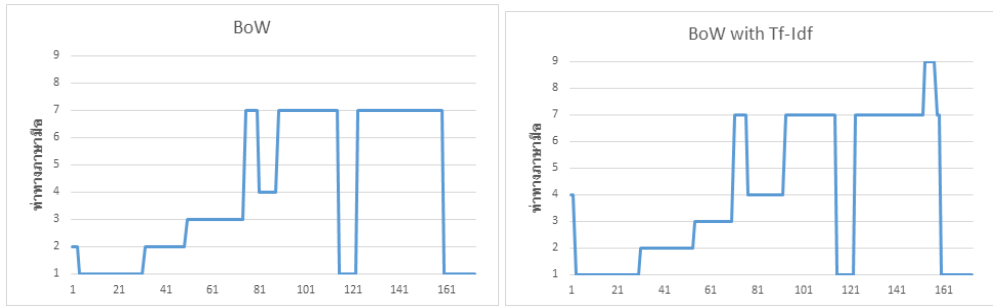
รูปที่ 4 ท่าทาง “ซอคติ” (ด้านซ้าย) และ “บอกลา” (ด้านขวา)

4.4. การทดสอบประสิทธิภาพในการรู้จำท่าทางต่อเนื่อง

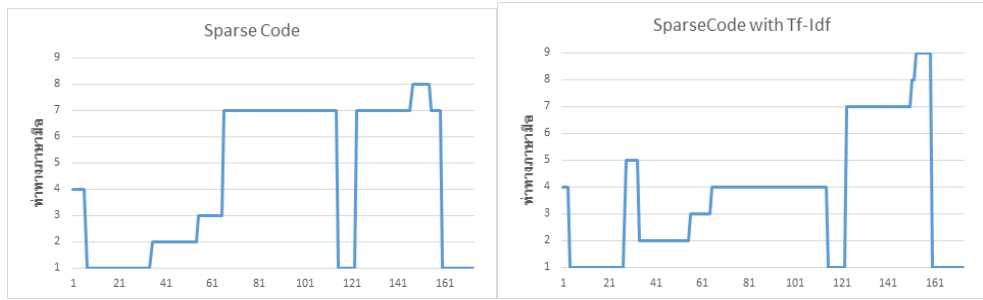
การทำท่าทางต่อเนื่อง ระบบจะคาดการณ์ท่าทางภาษามือทันทีเมื่อผู้ใช้เริ่มต้นให้ข้อมูลการเคลื่อนไหวผ่านทางไคเนค ทำการทดลองโดยให้กลุ่มตัวอย่างแสดงท่าทางภาษามือตามที่กำหนดไว้ 9 ท่าทางต่อเนื่องกัน และเปรียบเทียบผลกับการสร้างฮิสโตแกรมทั้ง 4 แบบคือ Bag-of-Words Sparse Code Transition และ Transition Sparse Code เช่นเดียวกัน ระบบจะคาดการณ์ท่าทางภาษามือ โดยเลือกจากท่าทางที่มีค่าความน่าจะเป็นที่คำนวณในสมการที่ 10 มากที่สุด ค่าความน่าจะเป็นของท่าทางภาษามือทั้ง 9 ท่าทางของวิธีการแบบ BoW แสดงดังรูปที่ 5 และผลลัพธ์ของการรู้จำท่าทางแสดงดังรูปที่ 6



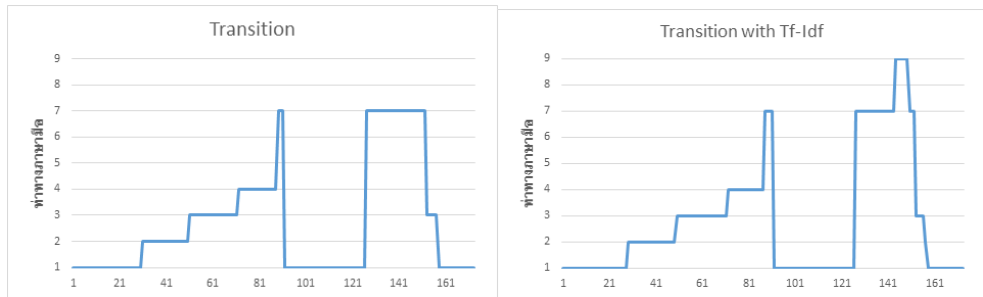
รูปที่ 5 ค่าความน่าจะเป็นของทั้ง 9 ท่าทาง แบบ BoW



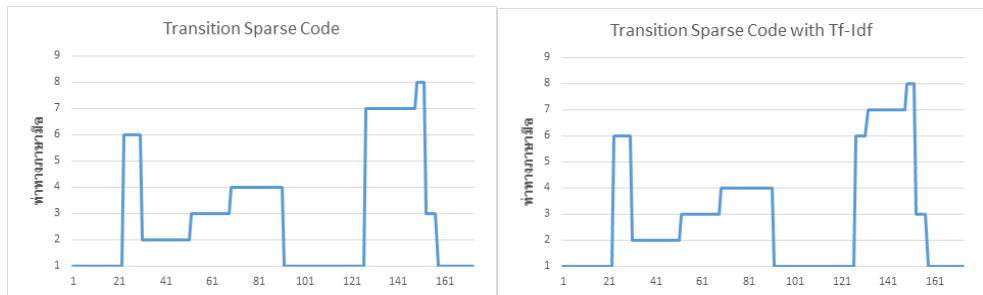
(ก)



(ข)



(ค)



(ง)

รูปที่ 6 ผลลัพธ์ของการรู้จำท่าทางของวิธีการ 4 แบบ

ผลลัพธ์ของการรู้จำท่าทางภาษามือทั้ง 4 แบบ พบว่าระบบมีแนวโน้มว่าจะคาดการณ์ท่าทาง “โชคดี” “คิด” “เข้าใจ” และ “ฉลาด” ยังไม่ถูกต้อง เนื่องจากระบบยังไม่สามารถแยกท่าทาง

ที่มีการเคลื่อนไหวที่ใกล้เคียงกัน อย่างไรก็ตามเมื่อเพิ่มหลักการ Tf-Idf พบว่าระบบสามารถคาดการณ์ได้มีประสิทธิภาพมากขึ้น โดยสามารถคาดการณ์ท่าทาง “ฉลาด” ได้ดีขึ้นในการสร้างรหัสโปรแกรมแบบ Bag-of-Words Sparse Code และ Transition

5. สรุปผลการวิจัย

ระบบฝึกภาษามือไทยด้วยอุปกรณ์โคเนคที่ได้พัฒนาขึ้น สามารถรู้จำท่าทางภาษามือที่แตกต่างกันได้ อย่างไรก็ตามในท่าทางที่มีการเคลื่อนไหวน้อยและลักษณะการเคลื่อนไหวที่ใกล้เคียงกัน ยังมีประสิทธิภาพในการรู้จำต่ำ ในงานวิจัยขั้นถัดไปจะดำเนินการปรับปรุงประสิทธิภาพการรู้จำ รวมถึงการแสดงผลผ่านแอนิเมชันสามมิติ และการประเมินความถูกต้องของผู้ใช้ในการแสดงท่าทางภาษามือ

เอกสารอ้างอิง

- กสทช. (2555). พจนานุกรมภาษามือไทยอิเล็กทรอนิกส์. Retrieved from <http://e-tsl.com/>
- Bashir Magomedov. (2006). Self-Organizing Feature Maps - Kohonen maps. Retrieved September 10, 2015, from <http://www.codeproject.com/Articles/16273/Self-Organizing-Feature-Maps-Kohonen-maps>
- Kau, L.-J., Su, W.-L., Yu, P.-J., & Wei, S.-J. (2015). A real-time portable sign language translation system. In 2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS) (pp. 1–4).
- Kumarage, D., Fernando, S., Fernando, P., Madushanka, D., & Samarasinghe, R. (2011). Real-time sign language gesture recognition using still-image comparison motion recognition. In 2011 6th IEEE International Conference on Industrial and Information Systems (ICIIS) (pp. 169–174).
- Kyan, M., Sun, G., Li, H., Zhong, L., Muneesawang, P., Dong, N., Guan, L. (2015). An Approach to Ballet Dance Training Through MS Kinect and Visualization in a CAVE Virtual Reality Environment. ACM Trans. Intell. Syst. Technol., 6(2), 23:1–23:37.
- Microsoft. (2016). Meet Kinect for Windows. Retrieved from <https://dev.windows.com/en-us/kinect>

- Shapi, A., Bahari, N. N., Arshad, H., Zin, N. A. M., & Mahayuddin, Z. R. (2015). Rehabilitation exercise game model for post-stroke using Microsoft Kinect camera. In 2015 2nd International Conference on Biomedical Engineering (ICoBE) (pp. 1–6).
- Tiwari, V., Anand, V., Keskar, A. G., & Satpute, V. R. (2015). Sign language recognition through kinect based depth images and neural network. In 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 194–198).