ชื่อเรื่องวิทยานิพนธ์ ระบบแปลภาษามือภาษาไทยโดยใช้ยู-เซิร์ฟ

ผู้เขียน นายพลกฤต จันดา

ปริญญา วิศวกรรมศาสตรมหาบัณฑิต (วิศวกรรมคอมพิวเตอร์)

อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผศ.ดร.ศันสนีย์ เอื้อพันธ์วิริยะกุล

บทคัดย่อ

การสื่อสารในปัจจุบันได้แบ่งออกเป็น การฟัง การพูด การอ่าน การเขียน และรวมไปถึงการแสดงท่าทาง และผู้พิการทางหูและเป็นใบ้จะไม่สามารถสื่อความหมายโดยการพูดและการฟังได้ เหมือนคนปรกติทั่วไป หากบุคคลปกติต้องการที่จะสื่อสารกับผู้พิการทางหูและเป็นใบ้ นั้นบุคคลปกติจะต้องศึกษาเรียนรู้ลักษณะท่าทางของมือหรือภาษามือที่เป็นมาตรฐานภาษามือก่อนจึงสามารถที่จะสื่อสารได้ ซึ่งต้องใช้ระยะเวลาในการศึกษาและเกิดความยุ่งยากซับซ้อนในการเรียนรู้

ภาษามือนั้นสามารถแบ่งออกเป็น 2 ประเภทคือ การสะกดนิ้วมือ (Finger Spelling) จะใช้นิ้วมือสะกดเป็นตัวอักษรและภาษามือ (Sign Language) เป็นภาษาที่ใช้มือทำท่าทางต่างๆ และสีหน้าเพื่อบ่งบอกถึงคำแต่ละคำ ซึ่งสะดวก และรวดเร็ว มีความหมายใกล้เคียงกับธรรมชาติ โดยคนหูหนวกเป็นผู้สร้างและใช้ร่วมกันและบางครั้งกลุ่มคนที่จำเป็นต้องใช้ภาษามือเหล่านี้ ส่วนใหญ่จะสร้างภาษามือขึ้นมาใช้ในกลุ่มของตนเอง และเข้าใจกันเฉพาะบางกลุ่ม ดังนั้นจึงมีบางคำบางคำที่ไม่สามารถใช้สื่อความหมายให้เข้าใจตรงกันได้

วิทยานิพนธ์ฉบับนี้นำเสนอวิธีการแปลภาษามือภาษาไทย โดยใช้ภาษามือภาษาไทย 52 คำ และรู้จำท่าทางของมือโดยประยุกต์ใช้ ยู-เซิร์ฟ (Up-right Speed Robust Feature: U-SURF ร่วมกับวิธีการรู้จำต้นแบบโดยวิธีไดนามิกไทม์วอร์ปิง, คอนเทกซ์ดีเพนเด้นไดนามิกไทม์วอร์ปิงและ สตริงแกรมม่าฮาร์ด-ซีมีน เพื่อเปรียบเทียบการรู้จำท่าทางของมือและต้นแบบที่ให้ความถูกต้องดีที่สุด โดยใช้ผู้ทำการทดลองทั้งหมด 20 คนเก็บเป็นไฟล์วิดิทัศน์ แบ่งเป็น 4 กลุ่มคือ กลุ่มที่ 1 (จากผู้ทดลอง 5 คน) เพื่อใช้ทดสอบและเก็บเป็นฐานข้อมูลที่ใช้ในการเปรียบเทียบ (Signature Library)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

และต้นแบบของคำ ซึ่งขั้นตอนในการเก็บฐานข้อมูลที่ใช้ในการเปรียบเทียบจะทำโดยการบันทึกวีดิทัศน์ของผู้ทำการทดลองที่แสดงท่าทางของภาษามือแต่ละคำ แล้วทำการแยกเฟรมวีดิทัศน์ออกเป็นรูปภาพด้วยอัตรา 30 เฟรมต่อวินาทีและเลือกรูปโดยใช้เฟรมเว้นเฟรมของแต่ละคำของภาษามือแล้วทำการตัดเฉพาะบริเวณมือ (หากมืออยู่บริเวณหน้าจะทำการลบหน้าออกให้เหลือเฉพาะบริเวณมือ) สำหรับรูปที่ใช้ในการทดสอบเกิดจากการบันทึกวีดิทัศน์ของผู้ทำการทดสอบแต่ละคนแล้วทำการแยกเฟรมวีดิทัศน์ด้วยอัตรา 30 เฟรมต่อวินาทีและเลือกรูปโดยใช้เฟรมเว้นเฟรมของแต่ละคำของภาษามือเช่นกัน โดยผู้ทำการทดลองจะต้องสวมใส่เสื้อแขนยาวสีทึบ และมีฉากหลังเป็นสีทึบ กลุ่มที่ 2 (จากผู้ทดลอง 5 คน) ใช้สำหรับทดสอบและหาต้นแบบของคำ และกลุ่มที่ 3 (จากผู้ทดลอง 5)ใช้สำหรับการทดสอบเพื่อเปรียบเทียบผลความถูกต้องเท่านั้น กลุ่มที่ 4 (จากผู้ทดลอง 5 คน)ใช้สำหรับการทดลองแบบออนไลน์(การเพิ่มฐานข้อมูลที่ใช้ในการเปรียบเทียบ ต้นแบบของคำ และการทดสอบร่วมกับระบบเดิม) จำนวนของรูปที่ใช้เป็นฐานข้อมูลในการเปรียบเทียบมี 3,019 รูป และจำนวนรูปที่ใช้ในการทดลองไม่น้อยกว่า 680,100 รูป โดยแบ่งการทดลองออกเป็น 4 การทดลอง ได้แก่ 1.การทดลองการเปรียบเทียบระหว่างยู-เซิร์ฟและซิฟต์ 2.การทดลองภาษามือภาษาไทยทั้ง 52 คำ 3.การทดลองแบบออนไลน์ 4.การทดลองระบบโดยล่ามและผู้ที่มีความบกพร่องทางการได้ยิน

ในงานวิจัยนี้ได้ทำการเปรียบเทียบผลระหว่างยู-เซิร์ฟและซิฟต์กับวิธีการรู้จำต้นแบบโดยวิธีไดนามิกไทม์วอร์ปิง คอนเทกซ์ดีเพนเด้นไดนามิกไทม์วอร์ปิงและสตริงแกรมม่าฮาร์ด-ซีมีน จากคำภาษามือภาษาไทย 10 คำนั้นซิฟต์ให้ความถูกต้องเฉลี่ยร้อยละ 25.4%, 14.5% และ 41.0% จากทั้งสามวิธีตามลำดับแต่ผลลัพธ์ที่ได้จากยู-เซิร์ฟกับทั้ง 3 วิธีอยู่ที่ความถูกต้องเฉลี่ยร้อยละ 66.0%, 60.3% และ 86.66% ตามลำดับ สำหรับการทดลองของภาษามือทั้ง 52 คำโดยวิธีการของยู-เซิร์ฟกับวิธีการรู้จำต้นแบบโดยวิธีสตริงแกรมม่าฮาร์ด-ซีมีนให้ความถูกต้องเฉลี่ยที่ 81.58% และการทดลองแบบออนไลน์ให้ความถูกต้องเฉลี่ยที่ 83.33% ในขณะที่การทดลองโดยล่ามและผู้พิการทางหูความถูกต้องเฉลี่ย 51% ถึงแม้ว่ายู-เซิร์ฟนั้นจะให้ความถูกต้องสูงกว่าและเร็วกว่าซิฟต์ถึง 2-3 เท่าก็ตาม แต่ก็ยังมีความผิดพลาดของการจับคู่ระหว่างรูปทดสอบกับสัญลักษณ์อยู่ เนื่องจากรูปที่ใช้ทดสอบนั้นไม่มีกระบวนการประมวลผลรูปทดสอบก่อน และบางคำหรือบางสัญลักษณ์ในฐานข้อมูลที่ใช้ในการเปรียบเทียบค่อนข้างใกล้เคียงกันมาก

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

Thesis Title Thai Sign Language Translation System Using U-SURF

Author Mr.Phonkrit Chanda

Degree Master of Engineering (Computer Engineering)

Thesis Advisor Asst. Prof. Dr. Sansanee Auephanwiriyakul

ABSTRACT

Communication can be divided into listening, speaking, reading, writing and posture. The deaf cannot communicate by speaking and listening like normal people. When normal people want to communicate with the deaf, they need to learn the standard sign language. However, it is not an easy task and it takes a long time to learn.

There are 2 types of sign language, i.e., fingerspelling (a signer spells each character using his fingers) and sign language (a signer uses hand gesture and face expression to represent each word). It is significantly faster and more natural than the former. Sign language is sometimes specific into some private groups. Only people in the same group understand the true meaning of those signs.

In this research, we propose Thai sign language translation system. We implement the system for 52 signed words. The Up-right Speed-up Robust Feature (U-SURF) with dynamic time warping (DTW), context dependent dynamic time warping (CD-DTW) or string grammar hard C-means (SgHCM) is utilized to recognize the hand gestures. There are 20 subjects wearing a black shirt with long sleeves and standing in front of the dark background in the experiment. Five of them are the subjects used in the collection of signature library and the creation of prototypes. The video of each sign is recorded with 30 frames/second. Every other frame is extracted and called representative frame (Rframe). Only the hand part of Rframe is kept. Other

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

five subjects are used as a part of prototypes creation. The last ten subjects are used as a blind test data set with 5 for offline testing and the other 5 for online testing. There are 3,019 frames in the signature library. There are 680,100 frames in the experiment. We compare U-surf and SIFT using DTW, CD-DTW and SgHCM for 10 Thai sign words. The SIFT correct classification is around 25.4%, 14.5% and 41.0% on the average for each classifier. While the correct classfication from U-SURF is around 66.0%, 60.3%, and 86.66% on the average for each classifier. Then we implement U-SURF with SgHCM for 52 signed-words. The correct classification in this case is around 81.58% on the average. The correct classification for online-system is around 83.33%. We then implement this system with subjects who are real sign interpreters and real deaf people. The result is around 51% on the average.

Although U-SURF gives high accuracy and are approximately 2-3 times faster than SIFT. There are false matching between testing image and symbols in signature library because testing images are not preprocessed and some symbols in signature library are very similar.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved