การรู้จำใบหน้าและเสียงพูดแบบประสานกันเพื่อระบุตัวบุคคล
ด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชัน

โดย นายศักรินทร์ ศรีคุณ
นางสาวธัญญาเรศ คำสุดที

บทคัดย่อ

โครงงานนี้นำเสนอการระบุตัวบุคคลด้วยการเรียนรู้จำภาพใบหน้าและเสียงพูดแบบประสานกัน ซึ่งเป็นตัวเลือกในการรักษาความปลอดภัยและเก็บรวมรวมข้อมูลหรือนำไปประยุกต์ใช้งานในด้านอื่นๆได้ โดยระบบในการรู้จำใช้ทฤษฎีของโครงข่ายประสาทเทียมแบบคอนโวลูชั่น (Convolutional Neural Networks: CNNs) ในการเรียนรู้จากข้อมูลรูปภาพใบหน้าและเสียงพูด ขั้นตอนแรกจะเก็บรวมรวมข้อมูลภาพใบหน้าและเสียงพูดของกลุ่มตัวอย่างนักศึกษาวิศวกรรมไฟฟ้า ชั้นปีที่ 4 จำนวน 50 คนแบ่งเป็นชาย 28 คน หญิง 22 คน โดยการเก็บภาพใบหน้าและเสียงพูด 3 ประโยคได้แก่ สวัสดีครับ/ค่ะ ขอบคุณครับ/ค่ะ ลาก่อนครับ/ค่ะ ข้อมูลภาพใบหน้าเป็นข้อมูลภาพโดยตรงแต่ข้อมูลเสียงพูดต้องนำไฟล์เสียงไปแปลงเป็นข้อมูลเชิงรูปภาพก่อนที่จะนำไปใช้ในกระบวนการรู้จำโดยใช้ทฤษฎีของ (Mel-Frequency Cepstral Coefficients: MFCCs) หลังจากที่ข้อมูลเป็นเชิงรูปภาพทั้งหมดแล้ว นำข้อมูลที่ได้ไปเทรนโมเดลในการรู้จำใบหน้าและเสียงพูดให้มีค่า Accuracy สูงที่สุดเพื่อความแม่นยำถูกต้อง การทดสอบผลการทำนายโมเดลแบ่ง 3 ส่วน ประกอบด้วย 1.การทดสอบแบบการบันทึกภาพถ่ายและเสียงพูด 2.การทดสอบแบบเรียลไทม์ผ่านกล้องเว็บแคม (webcam) และไมโครโฟน 3.การทดสอบการทำงานด้วยบอร์ด Respberry Pi ผลการทดสอบโดยรวมพบว่าระบบสามารถระบุตัวบุคคลได้ถูกต้องเกิน 80% ซึ่งมีแนวโน้มที่สามารถนำไปประยุกต์ใช้งานในชีวิตจริงได้

Coordinated Face and Speech Recognition for Person Identification by Convolutional Neural Networks

By Mr. Sakkarin Srikhun
Miss. Thanyares Khamsudtee

ABSTRACT

This project presents the coordinated face and speech recognition for person Identification by convolutional neural networks. It can be an option for security and data collection or other applications. The recognition system uses the theory of convolutional neural networks (CNNs) for learning from the image data of face and speech. In this project, the first step is to collect the facial and speech image data of a sample of 50 fourth-year electrical engineering students which consist of 28 males and 22 females. The collect information of face image and three speech sentences which are "สวัสดีครับ/ค่ะ" (Hello), "ขอบคุณครับ/ค่ะ"(Thank you) and "ลาก่อนครับ/ค่ะ"(Goodbye). The facial image data can be directly image data, but the speech data must be converted from the audio waveform to image data based on Mel-frequency cepstral coefficients (MFCCs). Next, the all-data files are taken to train the recognition model in order to get the highest accuracy. To validate the proposed method, the test is divided into three parts that consist of pre-taken photo and pre-recorded speech, real-time test by web camera and microphone, and real-time test by Raspberry Pi. The results show that the identification of proposed method achieve accuracy more than 80%, which has a potential to be used in real applications.