บทความวิจัย

การรู้จำท่าทางมือสำหรับตรวจจับความมีชีวิตของผู้ใช้แบบทันทีทันใดในแอปพลิเคชันโทรศัพท์เคลื่อนที่โดยใช้การเรียนรู้เชิงลึก

วรเมธ เลิศศิวนนท์ และ ฐิติรัตน์ ศิริบวรรัตนกุล*
คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 0 2727 3067 อีเมล: thitirat@as.nida.ac.th DOI: 10.14416/j.kmutnb.2021.06.003
รับเมื่อ 23 กันยายน 2563 แก้ไขเมื่อ 10 พฤศจิกายน 2563 ตอบรับเมื่อ 25 พฤศจิกายน 2563 เผยแพร่ออนไลน์ 10 มิถุนายน 2564

บทคัดย่อ

ระบบระบุตัวตนด้วยใบหน้าซึ่งเป็นที่นิยมใช้กันในปัจจุบัน โดยเฉพาะในแอปพลิเคชันบนโทรศัพท์เคลื่อนที่สมาร์ตโฟนนั้น มีจุดอ่อนที่สามารถถูกโจมตีได้ด้วยวิธีการต่างๆ เช่น การใช้รูปภาพใบหน้าสองมิติ หรือใช้แบบจำลองใบหน้าที่พิมพ์ด้วยเครื่องพิมพ์สามมิติมาแสดงที่หน้ากล้องเพื่อหลอกระบบว่าเป็นใบหน้าของบุคคลนั้นๆ เพื่อป้องกันการถูกโจมตีในลักษณะดังกล่าว แอปพลิเคชันส่วนใหญ่จึงมักมีการตรวจสอบด้วยวิธีการอื่นเพิ่มเติม เพื่อให้แน่ใจว่าใบหน้าที่เห็นปรากฏอยู่หน้ากล้องนั้นเป็นใบหน้าของบุคคลจริงที่มีชีวิต หรือเป็นใบหน้าที่ไม่มีชีวิตของภาพถ่ายหรือรูปปั้น ซึ่งกระบวนการนี้เรียกว่า "การตรวจสอบความมีชีวิตของผู้ใช้งาน" งานวิจัยชิ้นนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาความเป็นไปได้ของการนำท่าทางสัญลักษณ์มือมาใช้ในการตรวจสอบความมีชีวิตของผู้ใช้งาน 2) เพื่อเสริมความปลอดภัยให้กับระบบระบุตัวตนด้วยใบหน้าของแอปพลิเคชันในโทรศัพท์เคลื่อนที่สมาร์ตโฟน โดยนอกจากการทดลอง และพัฒนาแบบจำลองการเรียนรู้เชิงลึกที่สามารถแยกแยะท่าทางสัญลักษณ์มือหกท่าได้แล้ว ผู้วิจัยยังได้ทำการพัฒนาระบบต้นแบบบนโทรศัพท์เคลื่อนที่สมาร์ตโฟนที่รวมเอาระบบตรวจจับใบหน้าเข้ากับการใช้ท่าทางของมือเพื่อตรวจสอบความมีชีวิตของผู้ใช้ ทั้งนี้ระบบต้นแบบดังกล่าวถูกนำไปทดลองใช้เพื่อศึกษาประสบการณ์การใช้งานจากผู้ใช้จำนวน 40 ราย ในช่วงอายุต่างๆ กัน

คำสำคัญ: การเรียนรู้เชิงลึก การวิเคราะห์รูปภาพ การรู้จำสัญลักษณ์มือ การระบุตัวตนด้วยใบหน้า การตรวจสอบความมีชีวิต แอปพลิเคชัน โทรศัพท์เคลื่อนที่

การอ้างอิงบทความ: วรเมธ เลิศศิวนนท์ และ ฐิติรัตน์ ศิริบวรรัตนกุล, "การรู้จำท่าทางมือสำหรับตรวจจับความมีชีวิตของผู้ใช้แบบทันทีทันใดในแอปพลิเคชันโทรศัพท์เคลื่อนที่โดยใช้การเรียนรู้เชิงลึก," *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 32, ฉบับที่ 1, หน้า 153–163, ม.ค.–มี.ค. 2565.

Research Article

Hand Gesture Recognition for Real-time Liveness Detection in Mobile Phone Applications Using Deep Learning

Woramet Lertsiwanont and Thitirat Siriborvornratanakul*
Graduate School of Applied Statistics, National Institute of Development Administration, Bangkok, Thailand

* Corresponding Author, Tel. 0 2727 3067, E–mail: thitirat@as.nida.ac.th
DOI: 10.14416/j.kmutnb.2021.06.003
Received 23 September 2020; Revised 10 November 2020; Accepted 25 November 2020; Published online: 10 June 2021

Abstract

Recently vision-based face identification systems have become popular in mobile phone applications as they introduce easy and non-intrusive ways of human identification. Despite of their popularity, these systems can be easily tricked by 2D printed images or 3D printed models of faces. To prevent such attacks, most vision-based face identification systems enhance their security by involving additional liveness detection methods; this is for the purpose of checking whether the face as seen by camera is a face of living people or a face of non-living 2D images, 3D printed models or other statues. In this research, we conduct a feasibility study regarding usages of real-time hand gestures for liveness detection in smartphone-based face identification systems. Our work includes not only developing a robust in-depth learning model for real-time hand gesture recognition, but also creating a smartphone-based prototype application. This prototype application has been brought to test with 40 different smartphone users from various ranges of ages, allowing us to evaluate on-production technical efficiency, user satisfaction, and use acceptance.

Keywords: Deep Learning, Image Analytics, Hand Gesture Recognition, Face Identification, Liveness Detection, Mobile Phone Application

Please cite this article as: W. Lertsiwanont and T. Siriborvornratanakul, "Hand gesture recognition for real-time liveness detection in mobile phone applications using deep learning," *The Journal of KMUTNB*, vol. 32, no. 1, pp. 153–163, Jan.–Mar. 2022 (in Thai).

1. บทนำ

ด้วยการพัฒนาแบบก้าวกระโดดของเทคโนโลยีการวิเคราะห์รูปภาพ (Image Analytics) โดยเฉพาะในช่วงหนึ่งทศวรรษที่ผ่านมา ทำให้ความแม่นยำในการวิเคราะห์รูปภาพด้วยคอมพิวเตอร์เพิ่มสูงขึ้นจนถึงจุดที่สามารถถูกนำมาใช้งานได้จริงในเชิงพาณิชย์ และเชิงอุตสาหกรรม โดยในบรรดานั้นระบบวิเคราะห์ที่ใช้กล้องเป็นเครื่องมือเก็บข้อมูลเพื่อระบุตัวตนของบุคคลหนึ่งๆ ผ่านภาพถ่ายใบหน้า (Vision-based Face Identification) ถือเป็นระบบที่ได้รับความนิยมสูง มีการใช้งานอย่างแพร่หลายในต่างประเทศรวมถึงในประเทศไทย ซึ่งโดยมากจะเป็นลักษณะของการใช้ระบบเพื่อยืนยันตัวตนของบุคคลในการเข้าใช้งานอุปกรณ์ หรือแอปพลิเคชันที่ต้องการความปลอดภัย บ้างก็ใช้เพื่อระบุตัวตนของบุคคลที่เข้าออกสถานที่หนึ่งๆ โดยอัตโนมัติทดแทนการใช้ระบบตอกบัตรหรือระบบสแกนลายนิ้วมือ

อย่างไรก็ตาม ความแม่นยำของระบบระบุตัวตนบุคคลจากภาพใบหน้านั้น นอกจากจะขึ้นกับความฉลาดของซอฟต์แวร์ซึ่งเป็นปัญญาประดิษฐ์ (Artificial Intelligence; AI) แบบหนึ่งแล้ว ประสิทธิภาพของตัวฮาร์ดแวร์กล้อง และฮาร์ดแวร์ส่วนการประมวลผลของตัวอุปกรณ์เองก็มีผลด้วยเช่นกัน ซึ่งความท้าทายสำหรับแอปพลิเคชันสมาร์ตโฟนในปัจจุบันเกิดจากความหลากหลายของการประกอบ และรุ่นของฮาร์ดแวร์ของสมาร์ตโฟนในท้องตลาด ที่ทำให้ผู้พัฒนาระบบไม่สามารถกำหนดมาตรฐาน หรือรูปแบบประมวลผลหรือเงื่อนไขการตรวจสอบเพื่อความปลอดภัยเพียงแบบเดียวแต่สามารถใช้ได้กับสมาร์ตโฟนทุกรุ่นได้

ตัวอย่างความแตกต่างของฮาร์ดแวร์ในสมาร์ตโฟนที่ส่งผลกระทบต่อระดับความปลอดภัยในการระบุตัวตนจากใบหน้านั้น เช่น ไอโฟนรุ่นก่อนหน้ารุ่น X ที่มีฮาร์ดแวร์ส่วนกล้องที่ประกอบด้วยกล้องหน้าและกล้องหลัง ซึ่งทั้งคู่ทำงานภายใต้ช่วงแสงที่คนทั่วไปสามารถมองเห็นได้ตามปกติ (Visible Light Spectrum) ภาพที่ถ่ายได้ก็เป็นภาพสีสองมิติธรรมดา ด้วยฮาร์ดแวร์กล้องที่มีความสามารถจำกัดนี้ทำให้ผู้ไม่ประสงค์ดีสามารถใช้เครื่องพิมพ์สีตามท้องตลาดทั่วไป พิมพ์ภาพใบหน้าบุคคลที่ต้องการลงบนแผ่นกระดาษ แล้วแสดงแผ่นกระดาษดังกล่าวต่อหน้ากล้องเพื่อหลอกระบบระบุตัวตนด้วยใบหน้าบนไอโฟนรุ่นเก่าเหล่านี้ได้ แต่ทั้งนี้ไอโฟนตั้งแต่รุ่น X เป็นต้นไปมีการเปลี่ยนแปลงส่วนของฮาร์ดแวร์กล้องขนานใหญ่ โดยเพิ่มกล้องถ่ายภาพในช่วงแสงอินฟราเรด และเครื่องฉายแสงอินฟราเรดแบบจุดชนิดมุมกว้างแบบพิเศษเข้าไป ทำให้การระบุตัวตนบุคคลจากภาพใบหน้าเปลี่ยนจากการวิเคราะห์ภาพสีสองมิติเป็นการถอดรหัสจุดบนภาพอินฟราเรดที่ถ่ายได้ ซึ่งผลลัพธ์นั้น ทำให้สามารถวิเคราะห์รูปร่างใบหน้าคนได้ละเอียดในระดับสามมิติ เพิ่มความแม่นยำในการระบุตัวตนจากใบหน้าให้สูงขึ้น และไม่สามารถใช้การพิมพ์ภาพใบหน้าลงบนกระดาษสองมิติแบนๆ มาหลอกไอโฟนที่ใช้ระบบลักษณะนี้ได้

ความแม่นยำที่หลากหลายซึ่งเป็นผลมาจากความแตกต่างของฮาร์ดแวร์สมาร์ตโฟนแต่ละรุ่นแต่ละยี่ห้อนี้ สะท้อนให้การทดลองของสำนักข่าว Forbes [1] ที่ทดสอบสมาร์ตโฟน 5 รุ่น ในท้องตลาดแล้วพบว่า แบบจำลองสามมิติของศีรษะมนุษย์ ซึ่งพิมพ์ออกมาด้วยเครื่องพิมพ์สามมิตินั้นสามารถหลอกระบบระบุตัวตนจากใบหน้า เพื่อปลดล็อกสมาร์ตโฟนระบบปฏิบัติการแอนดรอยด์ได้ 4 รุ่น ได้แก่ รุ่น LG G7 ThinQ รุ่น Samsung Galaxy S9 รุ่น Samsung Galaxy Note 8 และรุ่น One Plus 6 โดยในการทดลองนี้มีเพียงไอโฟนรุ่น X เท่านั้นที่ไม่ถูกหลอก อย่างไรก็ตาม อ้างอิงจากสำนักข่าว Reuters [2] มีนักวิจัยชาวเวียตนามสามารถหลอกระบบระบุตัวตนด้วยใบหน้าของไอโฟนรุ่น X ได้สำเร็จ โดยการใช้หน้ากากที่พิมพ์จากเครื่องพิมพ์สามมิติร่วมกับการใช้ซิลิโคน และเทปกาวแปะประกอบกันบนใบหน้า

จากตัวอย่างที่ยกไปจะเห็นว่าการพึ่งพาความสามารถของฮาร์ดแวร์เพียงอย่างเดียวนั้น สุ่มเสี่ยงต่อการเปิดช่องโหว่ให้กับผู้ไม่ประสงค์ดี ที่อาจอาศัยจุดอ่อน หรือข้อจำกัดของฮาร์ดแวร์ในการเข้าโจมตีระบบ ดังนั้นจึงเป็นเรื่องจำเป็นที่ผู้พัฒนาจะต้องคำนึงถึงการใช้ซอฟต์แวร์ร่วมตรวจสอบ เพื่อเสริมความปลอดภัยเข้าไปอีกชั้นหนึ่ง โดยหนึ่งในวิธีซึ่งเป็นที่นิยมใช้ในแอปพลิเคชันต่างๆ คือ การใช้ซอฟต์แวร์ตรวจจับความมีชีวิตของผู้ใช้ (Liveness Detection) เพื่อแยกแยะว่าใบหน้าที่กล้องมองเห็นอยู่นี้ เป็นใบหน้าของบุคคลจริงที่มีชีวิต

หรือเป็นใบหน้าที่ไม่มีชีวิตของภาพถ่ายหรือรูปปั้นกันแน่

1.1 เทคนิคการตรวจจับความมีชีวิตของผู้ใช้

เทคนิคการตรวจจับความมีชีวิตของผู้ใช้นั้น สามารถแบ่งออกได้เป็น 3 ประเภทใหญ่คือ 1) แบบที่ให้ผู้ใช้ทำอะไรบางอย่างตามคำสั่งในระบบ (Active Liveness Detection) เช่น ให้กระพริบตา ยิ้ม หรือเอียงศีรษะ ข้อเสียของวิธีนี้คือ ผู้ใช้จะรู้ตัวว่ากำลังถูกระบบตรวจสอบอยู่ และผู้ไม่ประสงค์ดีก็สามารถเรียนรู้เงื่อนไขในการตรวจสอบของระบบได้โดยง่าย 2) แบบที่ใช้การวิเคราะห์อยู่เบื้องหลัง (Passive Liveness Detection) ซึ่งจะเป็นการตรวจสอบที่ไม่ทำให้ผู้ใช้รู้ตัวว่ากำลังถูกระบบตรวจสอบอยู่ อาทิ การตรวจสอบการเคลื่อนไหวของดวงตา การตรวจสอบสภาพผิวบนใบหน้า 3) แบบผสมผสาน (Hybrid) ที่ใช้เทคนิคการตรวจจับหลายวิธีร่วมกัน เช่น ตรวจจับการเคลื่อนไหวของดวงตาพร้อมกับบอกให้ผู้ใช้ยิ้ม หรือตรวจสอบการเคลื่อนไหวของดวงตาพร้อมกับตรวจสอบสภาพผิวหน้า

สำหรับเทคนิคการตรวจจับความมีชีวิตของผู้ใช้ที่พบได้บ่อยในแอปพลิเคชันสมาร์ตโฟน คือ การใช้การเคลื่อนไหวบนใบหน้าหรือการขยับศีรษะ อาทิ ให้ผู้ใช้กะพริบตา อ้าปาก เอียงศีรษะซ้ายขวา หันศีรษะซ้ายขวา ให้อ่านข้อความตามที่ระบบกำหนด หรือการใช้แสงที่ส่องจากหน้าจอสมาร์ตโฟนส่องกระทบใบหน้าผู้ใช้ เพื่อตรวจสอบลักษณะของผิวจากแสงสะท้อนใบหน้า ตัวอย่างของแอปพลิเคชันในประเทศไทยที่ใช้เทคนิคลักษณะนี้ได้แก่ แอปพลิเคชัน SCB Easy ที่ภายหลังการถ่ายรูปใบหน้าของผู้ที่ต้องการเปิดบัญชีแล้วยังมีการใช้การเคลื่อนไหวบนใบหน้าเพื่อยืนยันตัวตนอีกครั้ง โดยแอปพลิเคชันจะขอให้ผู้ใช้งานเคลื่อนไหวใบหน้าตามคำสั่งที่กำหนด เช่น หลับตาข้างใดข้างหนึ่ง เอียงศีรษะไปด้านใดด้านหนึ่ง (ข้อมูลจาก YouTube: SCB Thailand วิดีโอคลิปชื่อ SCB EASY EKYC https://www.youtube.com/watch?v=kCLk7CCOZp0) อีกตัวอย่างหนึ่งคือแอปพลิเคชันของรัฐบาลไทยอย่างเป๋าตัง ณ ตอนที่ถูกใช้ในการลงทะเบียนโครงการชิมช้อปใช้ ซึ่งในขั้นตอนการยืนยันตัวตนด้วยใบหน้าของแอปเป๋าตัง จะมีการเปลี่ยนหน้าจอเป็นสีต่างๆ อาทิ สีน้ำเงิน สีเขียว แล้ววิเคราะห์ภาพใบหน้าที่สะท้อนกับแสงสีนั้นๆ เพื่อให้แน่ใจว่าใบหน้าที่เห็นนั้นเป็นใบหน้าคนจริงๆ

จากการทบทวนวรรณกรรมในอดีต ผู้วิจัยแทบไม่พบงานวิจัยหรือแอปพลิเคชันสมาร์ตโฟนใดที่นำสัญลักษณ์หรือท่าทางมือ (Hand Gestures) มาใช้เพื่อตรวจสอบความมีชีวิตของผู้ใช้ในระบบระบุตัวตนจากใบหน้า ในนัยยะหนึ่งนั้น เป็นที่เข้าใจได้ว่าการใช้การเคลื่อนไหวใบหน้า และศีรษะมีความสะดวกมากกว่า เนื่องจากสามารถใช้ระบุตัวตนของผู้ใช้และตรวจสอบความมีชีวิตไปได้พร้อมๆ กัน แต่ผู้วิจัยก็เล็งเห็นถึงประโยชน์ของสัญลักษณ์ท่าทางมือที่มีความหลากหลายของท่าทาง และรูปแบบการเคลื่อนไหวมากกว่าใบหน้า และศีรษะ ว่าหากสามารถเสริมสัญลักษณ์มือเข้าไปในระบบได้น่าจะช่วยให้ระบบมีวิธีการตรวจจับความมีชีวิตของผู้ใช้ได้หลากหลายขึ้น ทำให้เป็นการยากยิ่งขึ้นที่ผู้ไม่ประสงค์ดีจะสามารถเรียนรู้วิธีตรวจสอบเพื่อโจมตีระบบได้ จึงเป็นที่มาของงานวิจัยชิ้นนี้ที่ต้องการศึกษาความเป็นไปได้และประสบการณ์ของผู้ใช้สมาร์ตโฟน (User Experience) ในการใช้สัญลักษณ์มือเพื่อตรวจจับความมีชีวิตร่วมไปกับการใช้ระบบรู้จำใบหน้า

1.2 เทคนิคการรู้จำท่าทางและสัญลักษณ์มือจากภาพ

เทคนิคการรู้จำสัญลักษณ์มือ (Hand Gesture Recognition) โดยใช้การวิเคราะห์จากรูปภาพที่ถ่ายในช่วงแสงปกติที่ตามนุษย์มองเห็นนั้น มีประวัติการค้นคว้าวิจัยที่ยาวนานและมีเทคนิคที่หลากหลาย อาทิ งานของ [3], [9] ที่ใช้เทคนิคการวิเคราะห์สีผิวของคนร่วมกับ Bayesian Classifier ในการแยกแยะมือออกจากพื้นหลัง จากนั้นจึงใช้เทคนิค Curve Fitting และ Clustering และ Particle Filter ร่วมกัน เพื่อให้สามารถค้นหาตำแหน่งของปลายนิ้วได้โดยมีความทนทานต่อสภาพแสงที่แตกต่างและการที่ปลายนิ้วอาจถูกวัตถุอื่นในภาพบดบังไปบ้างในบางช่วงเวลา หรือในงานของ [4] ที่ศึกษาเทคนิคการวิเคราะห์ภาษามือในภาษาอารบิกด้วยการใช้ DSIFT (Dense Scale Invariant Feature Transform) ร่วมกับ Bag of Visual Words (BoVM) และการเรียนรู้ของเครื่องจักรแบบ Support Vector Machine (SVM) ซึ่งผลลัพธ์ที่ได้สามารถวิเคราะห์สัญลักษณ์มือได้แม้จะอยู่ใน

รูปที่ 1 ตัวอย่างสัญลักษณ์มือการนับเลขศูนย์ถึงห้าในงานวิจัยนี้ โดยจะเป็นมือซ้ายหรือขวาหรือพื้นหลังแบบใดก็ได้

สิ่งแวดล้อมที่ซับซ้อน แสงที่หลากหลาย ระยะห่าง และองศาการวางตัวของมือในภาพที่แตกต่าง อีกทั้งยังสามารถทำงานได้รวดเร็วแม้จะเป็นบนระบบแบบฝังตัว (Embedded System)

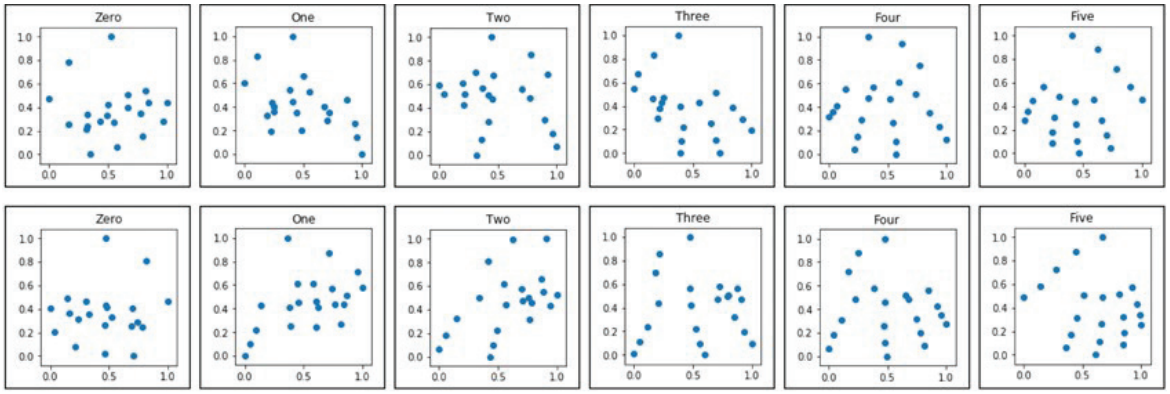
งานวิจัยที่กล่าวถึงไปในย่อหน้าก่อนเป็นลักษณะที่เรียกว่า Handcrafted Features หรือการที่มนุษย์ต้องเป็นผู้ออกแบบเวกเตอร์ลักษณะเด่นเชิงภาพ (Visual-based Feature Vector) ด้วยตนเอง แต่ด้วยความก้าวหน้าของเทคนิคการเรียนรู้เชิงลึก (Deep Learning) ในปัจจุบัน โดยเฉพาะโครงข่ายประสาทเทียมแบบคอนโวลูชัน (ConvNet; Convolutional Neural Network) ซึ่งมีความสามารถในการออกแบบสร้างเวกเตอร์ลักษณะเด่นเชิงภาพได้อัตโนมัติผ่านการเรียนรู้จากชุดข้อมูลรูปภาพจำนวนมาก ทำให้ในระยะหลังมีแนวโน้มของการนำการเรียนรู้เชิงลึก และ ConvNet มาเป็นส่วนหนึ่งเพื่อสร้างระบบรู้จำสัญลักษณ์มือเพิ่มมากขึ้น ตัวอย่างเช่น ในงานวิจัย [5]–[7] ที่ใช้การเรียนรู้เชิงลึกซึ่งรวม ConvNet มาเป็นส่วนหนึ่งในการพัฒนาระบบรู้จำท่าทางสัญลักษณ์มือ ทั้งนี้ในบรรดางานวิจัยที่ใช้แบบจำลองการเรียนรู้เชิงลึกสำหรับวิเคราะห์สัญลักษณ์มือ เทคนิคที่นำเสนอในเฟรมเวิร์ก MediaPipe โดยทีมวิจัยจากกูเกิล [8] ถือเป็นเทคนิคที่โดดเด่นมาก เนื่องจากสามารถวิเคราะห์หาตำแหน่งของข้อนิ้วมือทุกข้อ ทั้งมือซ้ายและมือขวาของทุกสภาพสีผิวได้อย่างแม่นยำในพิกัดสามมิติ อีกทั้งยังทำงานได้ในทุกพื้นหลัง ทุกสภาพแสง ทุกระยะห่างจากกล้อง ทุกองศาการวางตัวของมือ รองรับการทำงานกับมือหลายมือในภาพ (Multiple Hands) ทนต่อการเคลื่อนไหวแบบต่อเนื่องของมือและนิ้ว และได้ความเร็วการทำงานในระดับเรียลไทม์บนสมาร์ตโฟน

2. วัสดุ อุปกรณ์และวิธีการวิจัย

2.1 การพัฒนาแบบจำลองสำหรับตรวจจับท่าทางของมือ

ในส่วนของการพัฒนาระบบรู้จำสัญลักษณ์มือจากภาพที่มองเห็นโดยกล้องหน้าของสมาร์ตโฟนซึ่งเป็นเป้าหมายของงานวิจัยนี้ ผู้วิจัยเลือกที่จะเริ่มต้นจากสัญลักษณ์มือพื้นฐานที่ทุกคนรู้จักดีอย่างการนับเลขศูนย์ถึงห้าดังรูปที่ 1 อีกทั้งผู้วิจัยยังตัดสินใจที่จะนำเอาแบบจำลองที่ถูกสร้างไว้อย่างดีแล้วของทีมวิจัยจากกูเกิลใน MediaPipe [8] มาใช้ต่อยอดเพื่อให้สามารถสร้างระบบต้นแบบที่นำไปสู่การทดสอบความมีชีวิตของผู้ใช้ในขั้นต่อไปได้อย่างรวดเร็ว โดย ณ ขณะที่ผู้วิจัยทำการพัฒนาระบบนี้ MediaPipe ให้ผลลัพธ์ของมือข้างหนึ่งๆ ออกมาเป็นพิกัดสามมิติของจุดจำนวน 21 จุด ดังรูปที่ 2 โดยแต่ละจุดคือ ตัวแทนของจุดสังเกตของมือและนิ้ว (Hand Landmarks) ซึ่งออกแบบมาโดยทีมวิจัยจากกูเกิล

อย่างไรก็ตาม พิกัด 21 จุด ของมือดังกล่าวจำเป็นจะต้องผ่านการวิเคราะห์อีกชั้นเพื่อให้สามารถสรุปได้ว่าเป็นการนับเลขอะไรระหว่างศูนย์ถึงห้า โดยในขั้นตอนนี้ผู้วิจัยเลือกที่จะสร้าง และสอนแบบจำลองโครงข่ายประสาทเทียมแบบลึก (Deep Multi-Layer Perceptron; MLP) ให้รับอินพุตเป็นพิกัดทั้ง 21 จุด จาก MediaPipe และให้เอาต์พุตเป็นผลลัพธ์ที่บอกว่าสัญลักษณ์มือนี้คือการนับเลขอะไรในศูนย์ถึงห้า เพื่อการนี้ผู้วิจัยเริ่มต้นจากการเตรียมข้อมูลเพื่อใช้สอน (Train) และทดสอบ (Validate) แบบจำลอง โดย



รูปที่ 2 ตัวอย่างของพิกัดจุดสามมิติจำนวน 21 จุด ในงานวิจัยนี้ซึ่งเป็นผลลัพธ์จากการใช้เฟรมเวิร์ก MediaPipe เพื่อวิเคราะห์มือหนึ่งข้างซึ่งกำลังนับเลขศูนย์ถึงห้า (จากคอลัมน์ซ้ายไปขวาตามลำดับ) ทั้งนี้เนื่องจากการตั้งค่าแกน Y ของ MediaPipe กับโปรแกรมของผู้วิจัยแตกต่างกัน ผลลัพธ์ในภาพจึงเป็นภาพที่กลับหัวบนล่างดังที่ปรากฏ

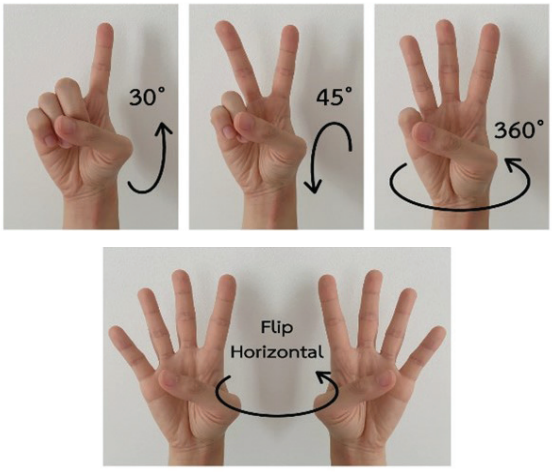
ทำการเก็บรูปภาพข้อมูลดิบจากผู้ใช้ 30 คน คนละ 12 รูป (ตัวอย่างในรูปที่ 1) แล้วนำรูปภาพทั้งหมดมาผ่าน MediaPipe แปลงให้เป็นพิกัดจุดสามมิติ

จากนั้นผู้วิจัยใช้เทคนิคเพิ่มจำนวนข้อมูล (Data Augmentation) โดยจำลองพฤติกรรมการหมุนของมือตามแกน X และแกน Y ในงานวิจัยนี้ผู้วิจัยทำการหมุนจุดในแนวแกน X ทีละ 5° ในช่วงการหมุน –30° ถึง 45° และหมุนจุดในแนวแกน Y ช่วง 0° ถึง 360° ด้วยสมการการหมุนตามแกน X และแกน Y ดังแสดงในสมการที่ (1)

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\theta & -sin\theta \\ 0 & sin\theta & cos\theta \end{bmatrix}$$

$$R_y(\theta) = \begin{bmatrix} cos\theta & 0 & sin\theta \\ 0 & 1 & 0 \\ -sin\theta & 0 & cos\theta \end{bmatrix} \quad (1)$$

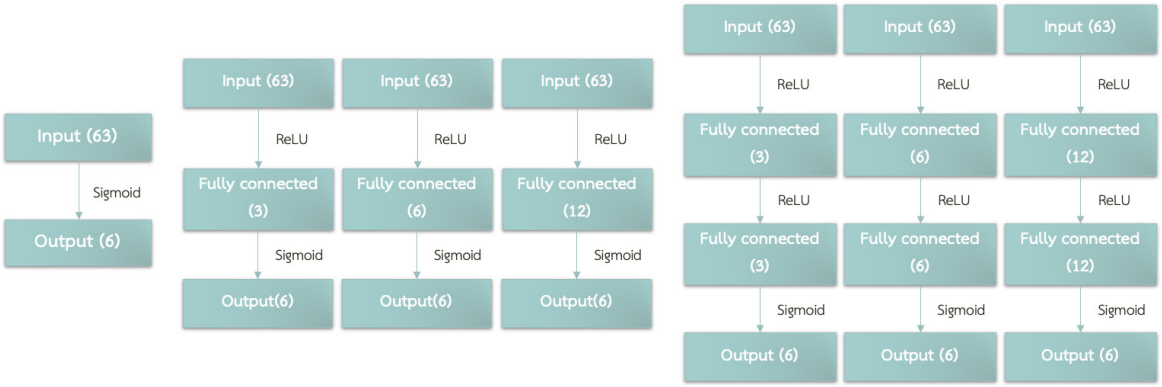
นอกจากนี้ผู้วิจัยยังเพิ่มการกลับด้านจุดตามแนวนอน (Flip Horizontal) หรือการเปลี่ยนค่า X ของจุดให้ติดลบ เพื่อจำลองการกลับข้างของมือซ้ายและมือขวาด้วย ทั้งนี้เทคนิคการเพิ่มข้อมูลทั้งหมดสรุปอยู่ในรูปที่ 3 ด้วยเทคนิคนี้ทำให้ข้อมูลดิบจำนวน 30 คน × 12 รูป = 360 ข้อมูล ที่ผู้วิจัยเก็บได้ เพิ่มจำนวนขึ้นเป็น 884,736 ข้อมูล ซึ่งถูกแบ่งเป็น



รูปที่ 3 เทคนิคการเพิ่มจำนวนข้อมูลด้วยการหมุนจุดตามแนวแกน X และ Y (ภาพแถวบน) และการกลับจุดตามแนวนอน (ภาพล่าง)

ข้อมูลสำหรับใช้สอน (Training Set) จำนวน 663,552 ข้อมูล (ข้อมูลของคน 25 คน) และข้อมูลสำหรับใช้ทดสอบ (Validation Set) จำนวน 221,184 ข้อมูล (ข้อมูลของคนอีก 5 คนที่เหลือ)

ก่อนจะนำข้อมูลทั้งหมดส่งให้แบบจำลอง MLP ผู้วิจัยทำ Data Normalization ด้วยสมการที่ (2) เพื่อแปลงพิกัดจุดทั้งหมดทุกค่าให้มีค่าอยู่ระหว่าง 0.0 ถึง 1.0



รูปที่ 4 สถาปัตยกรรมของแบบจำลอง MLP ทั้ง 7 แบบ ที่ถูกทดลองในงานวิจัยนี้

$$X_{norm} = (X - X_{min}) / (X_{max} - X_{min})$$
$$Y_{norm} = (Y - Y_{min}) / (Y_{max} - Y_{min})$$
$$Z_{norm} = (Z - Z_{min}) / (Z_{max} - Z_{min}) \quad (2)$$

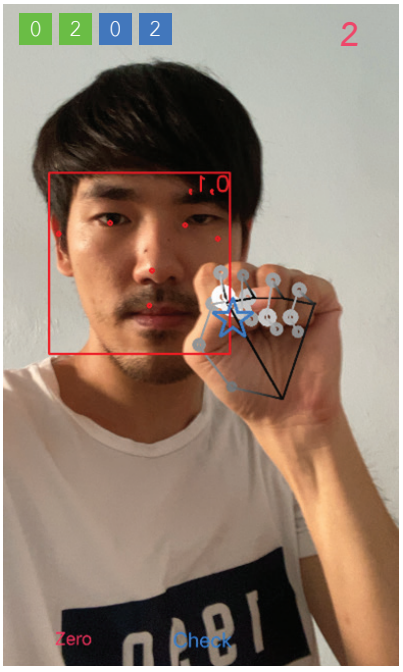
จากนั้นผู้วิจัยทำการทดลองเพื่อหาสถาปัตยกรรม MLP ที่เหมาะสมที่สุดสำหรับงานนี้ โดยได้ทำการทดลองทั้งหมด 7 รูปแบบ ดังสรุปในรูปที่ 4 แต่ละรูปแบบทำการทดลองด้วย Optimizer สองตัว ได้แก่ Stochastic Gradient Descent (SGD: learning_rate=0.01, batch_size=32) และ Adaptive Moment Estimation (Adam: learning_rate=0.001, beta_1=0.9, beta_2=0.999, batch_size=32) รวมเป็นทั้งหมด 14 การทดลอง

ในแต่ละการทดลองผู้วิจัยใช้ Mean Squared Error (MSE) Loss และใช้การสอนเพียง 5 Epochs เนื่องจากเป็นจำนวนที่ทำให้กราฟ Loss เริ่มแบนราบแล้ว หากทำการสอนด้วยจำนวน Epoch มากกว่านี้อาจทำให้แบบจำลอง Overfit ได้ สำหรับทรัพยากรหน่วยประมวลผลที่ใช้ในการสอนแบบจำลองแต่ละตัว ผู้วิจัยเลือกใช้ Tensor Processing Unit (TPU) ฟรีของ Google Colaboratory (Google Colab) โดยเป็นการรันใช้งานในช่วงเดือนมกราคมถึงมีนาคม ค.ศ. 2020

2.2 การพัฒนาแอปพลิเคชันต้นแบบบนสมาร์ตโฟน

เมื่อได้แบบจำลอง MLP สำหรับแยกแยะสัญลักษณ์มือของการนับศูนย์ถึงห้าที่ดีที่สุดจากการทดลองในหัวข้อที่ 2.1 แล้ว ขั้นตอนถัดมาคือการสร้างแอปพลิเคชันต้นแบบบนสมาร์ตโฟน เพื่อการนี้ผู้วิจัยใช้ภาษา C++ ในการเขียนโปรแกรมเพื่อเชื่อมต่อ MediaPipe เข้ากับแอปพลิเคชันที่สร้างเองบนระบบปฏิบัติการ iOS โดยในระบบต้นแบบนี้ผู้วิจัยเลือกที่จะใช้ระบบตรวจจับใบหน้า (Face Detector) ของกูเกิลชื่อ BlazeFace [10] ทดแทนระบบยืนยันตัวตนด้วยใบหน้าของจริงบนสมาร์ตโฟนไปก่อน เพื่อให้สร้างระบบได้ง่ายและผู้ใช้ยังคงได้สัมผัสประสบการณ์จริงของสมาร์ตโฟนที่ผสานการวิเคราะห์ใบหน้า และการวิเคราะห์สัญลักษณ์มือแบบเรียลไทม์ไว้ด้วยกัน ทั้งนี้จุดเด่นของ BlazeFace คือ เป็นแบบจำลองการเรียนรู้เชิงลึกที่ถูกออกแบบมาให้ค้นหาใบหน้าจากรูปภาพได้รวดเร็วบนหน่วยประมวลผลของสมาร์ตโฟน

ตัวของแอปพลิเคชันต้นแบบบนโทรศัพท์เคลื่อนที่นั้นถูกพัฒนาขึ้นมาด้วยภาษา Objective-C โดยมีการทำงานทั้งหมด 3 ส่วน ได้แก่ 1) ส่วนของการทำ Active Liveness Detection ที่ผู้วิจัยใช้การสุ่มตัวเลขขึ้นมา 4 ตัว (ค่าระหว่าง 0 ถึง 5) สำหรับให้ผู้ใช้ทำสัญลักษณ์มือตามตัวเลขแต่ละตัวไปตามลำดับ 2) ส่วนของการสุ่มจุดบนหน้าจอของแอปพลิเคชันเพื่อให้ผู้ใช้ขยับมือที่ทำสัญลักษณ์ไปอยู่ในจุดดังกล่าว และ 3) ส่วนของตัวกำหนดเวลาว่าผู้ใช้ต้องทำตามเงื่อนไขที่กำหนดบนหน้าจอให้แล้วเสร็จภายในระยะเวลาเท่าใด รูปที่ 5 แสดงตัวอย่างการทำงานของแอปพลิเคชันต้นแบบโดยกรอบสี่เหลี่ยมสีแดง คือ ผลลัพธ์ของใบหน้าที่ค้นพบจาก BlazeFace



รูปที่ 5 หน้าจอสมาร์ตโฟนระหว่างการใช้งานระบบต้นแบบ

สำหรับการนำแอปพลิเคชันต้นแบบนี้ไปทดลอง เพื่อศึกษาว่าการทำ Active Liveness Detection ด้วยสัญลักษณ์มือร่วมกับการใช้ระบบตรวจจับใบหน้าบนแอปพลิเคชันสมาร์ตโฟนนั้น จะให้ประสบการณ์ของผู้ใช้เป็นอย่างไร ผู้วิจัยได้ทำการทดสอบกับบุคคลทั่วไปจำนวน 40 คน รายละเอียดดังตารางที่ 1 และตารางที่ 2 โดยการทดลองนี้เป็นการเข้าร่วมแบบสมัครใจ ที่มีเงื่อนไขการเข้าร่วมเพียงว่าผู้เข้าร่วม คือ ผู้ที่สามารถใช้งานแอปพลิเคชันบนสมาร์ตโฟนได้ และไม่มีความผิดปกติหรือความพิการที่เกี่ยวข้องกับมือทั้งสองข้าง

ตารางที่ 1 สรุปเพศและช่วงอายุของผู้เข้าร่วมทดสอบ

ช่วงอายุ	เพศชาย	เพศหญิง	รวม
18–30 ปี	14	8	22
31–40 ปี	7	3	10
41–50 ปี	3	2	5
50 ปีขึ้นไป	2	1	3
รวม	26 (65%)	14 (35%)	40 (100%)

ตารางที่ 2 ประสบการณ์การใช้แอปพลิเคชันสมาร์ตโฟนในการยืนยันตัวตนด้วยใบหน้า

ช่วงอายุ	เคย	ไม่เคย	รวม
18–30 ปี	12	10	22
31–40 ปี	4	6	10
41–50 ปี	3	2	5
50 ปีขึ้นไป	0	3	3
รวม	19 (47.5%)	21 (52.5%)	40 (100%)

3. ผลการทดลอง

3.1 แบบจำลองสำหรับตรวจจับท่าทางของมือ

ตารางที่ 3 สรุปเวลาที่ใช้ในการสอนแบบจำลองทั้งหมด 14 รูปแบบ เป็นจำนวน 5 Epochs ส่วนตารางที่ 4 และ 5 คือ ผลลัพธ์ของ Training Loss และ Validation Loss ใน Epoch ที่ 5 ซึ่งเป็น Epoch สุดท้าย จากผลการทดลองทั้งหมดจะเห็นว่าแบบจำลองที่ดีที่สุด (Validation Loss ต่ำที่สุดที่ 0.0286) คือ แบบจำลองที่มี Hidden Layer 2 ชั้น แต่ละชั้นมี 12 โหนด และใช้ Adam Optimizer

อย่างไรก็ตาม เมื่อผู้วิจัยนำแบบจำลองที่ดีที่สุดนี้ไปทดสอบกับผู้ใช้จริงจำนวน 40 ราย (รายละเอียดการทดลองในหัวข้อที่ 2.2) พบว่า ในสัญลักษณ์มือที่ไม่ซับซ้อน และไม่มีความกำกวมคล้ายคลึงกับสัญลักษณ์อื่น (เช่น การนับศูนย์ ซึ่งเป็นการกำมือเพียงสัญลักษณ์เดียว) จะมีความแม่นยำมากกว่าสัญลักษณ์มืออื่น ส่วนสัญลักษณ์มือที่แบบจำลองที่ผู้วิจัยสร้างขึ้นยังทายผิดอยู่บ่อยๆ คือ สัญลักษณ์มือของการนับเลขสี่ ที่มักจะถูกทำนายผิดไปเป็นการนับเลขสามหรือเลขห้า โดยผู้วิจัยตั้งสมมติฐานว่าการทำนายผิดนี้น่าจะเกิดจากข้อมูลที่ใช้ในการสอนตัวแบบ MLP ยังมีความหลากหลายไม่พอ เช่น กรณีการนับเลขห้าที่นิ้วโป้งของผู้ใช้บางคนอาจแนบติดกับนิ้วชี้จนทำให้ดูคล้ายคลึงกับการนับเลขสี่ หรือกรณีที่ผู้ใช้ทำสัญลักษณ์เลขสาม ซึ่งนิ้วก้อยของผู้ใช้บางคนแม้จะพับลงแล้วแต่ส่วนข้อนิ้วก็ยังโผล่สูงขึ้นมาจนทำให้ระบบมองผิดเป็นการนับเลขสี่ได้ ข้อเสนอแนะของผู้วิจัยสำหรับแก้ไขปัญหาการพยากรณ์ผิดนี้ คือ การเก็บข้อมูลดิบสำหรับสร้างตัวแบบให้หลากหลายมากขึ้น โดยเฉพาะสัญลักษณ์มือ

ที่มีความกำกวมคล้ายคลึงกันซึ่งต้องเน้นเป็นพิเศษ ต้องเก็บข้อมูลในมุมมองที่หลากหลายให้มากขึ้นอีก

ตารางที่ 3 สรุปเวลาที่ใช้ในการสอน (Train) แบบจำลอง MLP ทั้ง 14 รูปแบบ

แบบจำลอง MLP	เวลาที่ใช้ (วินาที)	
	SGD	ADAM
ไม่มี Hidden Layer	75.126	103.598
Hidden Layer 1 ชั้น ชั้นละ 3 โหนด	83.434	105.170
Hidden Layer 1 ชั้น ชั้นละ 6 โหนด	74.478	108.607
Hidden Layer 1 ชั้น ชั้นละ 12 โหนด	80.452	107.818
Hidden Layer 2 ชั้น ชั้นละ 3 โหนด	93.976	119.257
Hidden Layer 2 ชั้น ชั้นละ 6 โหนด	91.387	129.253
Hidden Layer 2 ชั้น ชั้นละ 12 โหนด	98.378	135.346

ตารางที่ 4 สรุป Training Loss ที่ Epoch สุดท้ายของ แบบจำลอง MLP ทั้ง 14 รูปแบบ

แบบจำลอง MLP	Training Loss	
	SGD	ADAM
ไม่มี Hidden Layer	0.0768	0.0226
Hidden Layer 1 ชั้น ชั้นละ 3 โหนด	0.1040	0.0200
Hidden Layer 1 ชั้น ชั้นละ 6 โหนด	0.0502	0.0062
Hidden Layer 1 ชั้น ชั้นละ 12 โหนด	0.0515	0.0002
Hidden Layer 2 ชั้น ชั้นละ 3 โหนด	0.0603	0.0190
Hidden Layer 2 ชั้น ชั้นละ 6 โหนด	0.0409	0.0016
Hidden Layer 2 ชั้น ชั้นละ 12 โหนด	0.0431	0.0001

ตารางที่ 5 สรุป Validation Loss ที่ Epoch สุดท้ายของ แบบจำลอง MLP ทั้ง 14 รูปแบบ

แบบจำลอง MLP	Validation Loss	
	SGD	ADAM
ไม่มี Hidden Layer	0.0840	0.0412
Hidden Layer 1 ชั้น ชั้นละ 3 โหนด	0.1060	0.0433
Hidden Layer 1 ชั้น ชั้นละ 6 โหนด	0.0631	0.0335
Hidden Layer 1 ชั้น ชั้นละ 12 โหนด	0.0610	0.0297
Hidden Layer 2 ชั้น ชั้นละ 3 โหนด	0.0725	0.0404
Hidden Layer 2 ชั้น ชั้นละ 6 โหนด	0.0600	0.0363
Hidden Layer 2 ชั้น ชั้นละ 12 โหนด	0.0611	0.0286

รูปที่ 6 การทดลองแอปพลิเคชันต้นแบบกับบุคคลทั่วไป

3.2 แอปพลิเคชันต้นแบบบนสมาร์ตโฟน

รูปที่ 6 แสดงการทดลองแอปพลิเคชันต้นแบบกับบุคคลทั่วไป ในแง่ความเร็วของระบบนั้น จากการทดสอบบนไอโฟน 11 Pro พบว่า MediaPipe และ BlazeFace สามารถทำงานร่วมกันได้เป็นอย่างดี โดยได้ความเร็วเฉลี่ยสูงถึง 60 เฟรมต่อวินาที (Frames Per Second; fps) จึงสรุปได้ว่าระบบที่นำเสนอนี้สามารถทำงานได้จริงในแบบเรียลไทม์บนหน่วยประมวลผลของสมาร์ตโฟน

ตารางที่ 6 สรุปผลจากแบบสอบถามหลังการใช้ระบบของผู้ใช้ 40 คน สำหรับคำถามที่ว่า “ระบบมีระดับความยากง่ายในการใช้งานเป็นอย่างไร” โดยสามารถสรุปสาเหตุที่ผู้ใช้ตอบว่า “ระบบใช้ยาก” ได้สองสาเหตุใหญ่ คือ ในกลุ่มของผู้มีอายุ 41 ปีขึ้นไป ให้เหตุผลไปในทางเดียวกันว่าระบบมีขั้นตอนที่ยุ่งยาก และซับซ้อนมากขึ้นกว่าปกติ และในกลุ่มคนที่อายุน้อยกว่า 41 ปี ให้เหตุผลว่าระบบนี้จะถือว่าใช้งานยากถ้าต้องใช้ในที่สาธารณะหรือต้องใช้งานบ่อยๆ

ตารางที่ 6 สรุปความคิดเห็นของผู้เข้าร่วมทดสอบในประเด็นเรื่องความยากง่ายของระบบต้นแบบ

ช่วงอายุ	ใช้ง่าย	ปานกลาง	ใช้ยาก	รวม
18–30 ปี	11	9	2	22
31–40 ปี	5	3	2	10
41–50 ปี	1	1	3	5
50 ปีขึ้นไป	0	0	3	3
รวม	17 (42.5%)	13 (32.5%)	10 (25%)	40 (100%)

ในส่วนของตารางที่ 7 ซึ่งเป็นการสรุปผลสำหรับคำถามที่ว่า “ผู้ใช้คิดว่าการมีสัญลักษณ์มือเพิ่มเข้ามาสามารถช่วยให้การยืนยันตัวตนด้วยใบหน้ามีความปลอดภัยมากขึ้นหรือไม่” โดยสำหรับผู้ใช้ที่คิดว่าสัญลักษณ์มือช่วยให้ระบบปลอดภัยมากขึ้น ให้ความเห็นว่าเนื่องจากขั้นตอนการตรวจสอบมีมากขึ้นจึงน่าจะปลอดภัยมากขึ้น สำหรับผู้ใช้ที่คิดว่าน่าจะปลอดภัยเท่าๆ เดิม ให้ความเห็นว่าหากระบบระบุตัวตนด้วยใบหน้าถูกโจมตีได้ ก็คงไม่ใช่เรื่องยากที่ระบบที่ใช้สัญลักษณ์มือจะถูกโจมตีได้เหมือนกันและอาจจะโจมตีได้ง่ายกว่าใบหน้าเสียอีก ในส่วนของผู้ใช้ที่คิดว่าการมีสัญลักษณ์มือเพิ่มเข้ามาจะทำให้ระบบยืนยันตัวตนปลอดภัยน้อยลงนั้น ให้เหตุผลว่าสัญลักษณ์มือสามารถถูกโจมตีหรือหลอกได้ง่ายกว่าใบหน้า หากเอามาใช้ร่วมกันก็จะพลอยทำให้ประสิทธิภาพโดยรวมในการยืนยันตัวตนลดลง

ตารางที่ 7 สรุปความคิดเห็นของผู้เข้าร่วมทดสอบในประเด็นเรื่องความปลอดภัยในการยืนยันตัวตนด้วยระบบต้นแบบ

ช่วงอายุ	ปลอดภัยน้อยลง	ปลอดภัยเท่าเดิม	ปลอดภัยมากขึ้น	รวม
18–30 ปี	3	5	14	22
31–40 ปี	1	3	6	10
41–50 ปี	0	1	4	5
50 ปีขึ้นไป	0	0	3	3
รวม	4 (10%)	9 (22.5%)	27 (67.5%)	40 (100%)

4. สรุป

งานวิจัยนี้มีจุดประสงค์เพื่อทดสอบความเป็นไปได้ และประสบการณ์ของผู้ใช้ในสถานการณ์ที่สัญลักษณ์มือถูกนำมาใช้เพื่อเพิ่มความปลอดภัยให้กับระบบระบุตัวตนด้วยใบหน้าบนสมาร์ตโฟน ในงานนี้ผู้วิจัยได้พัฒนาแบบจำลองการเรียนรู้เชิงลึกที่ผสานการใช้งาน MediaPipe ของกูเกิลร่วมกับแบบจำลองโครงข่ายประสาทเทียมแบบลึกที่ผู้วิจัยพัฒนาเอง ทำให้สามารถแยกแยะสัญลักษณ์มือของการนับเลขศูนย์ถึงห้าได้ ผลลัพธ์ที่ได้ผู้วิจัยนำมาใช้ร่วมกับระบบตรวจจับใบหน้าชื่อ BlazeFace สร้างเป็นแอปพลิเคชันต้นแบบบนไอโฟน 11 Pro และนำไปทดสอบกับผู้ใช้จำนวน 40 คน ผลการทดสอบกับผู้ใช้จริงพบว่า ในแง่ของประสิทธิภาพความเร็วนั้นไม่มีปัญหาใดๆ แต่แบบจำลองยังทายสัญลักษณ์มือที่กำกวมบางอันผิด และในส่วนประสบการณ์ของผู้ใช้นั้นก็มีประเด็นเรื่องขั้นตอนของระบบที่มากขึ้นทำให้การใช้งานยากขึ้น และความไม่แน่ใจของผู้ใช้บางส่วนถึงระดับความปลอดภัยเชื่อถือได้ของระบบรู้จำสัญลักษณ์มือ

ในลำดับถัดไปของการวิจัย ผู้วิจัยมีความเห็นว่านอกจากการพัฒนาส่วนของแบบจำลองการเรียนรู้เชิงลึกให้รองรับท่าทางมือที่หลากหลายขึ้นแล้ว ยังควรให้ความสำคัญกับการวัดระดับความปลอดภัยของระบบระบุตัวตนด้วยใบหน้าในสถานการณ์การใช้งานต่างๆ รวมถึงการวัดระดับความใช้งานง่ายสำหรับผู้ใช้ไปพร้อมๆ กัน เพื่อให้สามารถหาจุดสมดุลที่ลงตัวที่สุดทั้งในแง่ความปลอดภัยและความใช้งานง่ายเหมาะสำหรับผู้ใช้งานโทรศัพท์เคลื่อนที่สมาร์ตโฟนที่เป็นบุคคลทั่วไปในหลากหลายช่วงอายุ

เอกสารอ้างอิง

[1] T. Brewster. (2020, September). *We broke into a bunch of Android phones with a 3D-printed head*. [Online]. Available: https://www.forbes.com/sites/thomasbrewster/2018/12/13/we-broke-into-a-bunch-of-android-phones-with-a-3d-printed-head/#36a4a5521330

[2] M. Nguyen. (2017, September). *Vietnamese researcher shows iPhone X face ID ‘hack’*, Reuters. [Online]. Available: https://www.reuters.com/article/us-apple-vietnam-hack-idUSKBN1DE1TH

[3] C. Kerdvibulvech, “Human hand motion recognition using an extended particle filter,” in *Proceedings AMDO 2014: Articulated Motion and Deformable Objects*, 2014, pp. 77–81.

[4] W. Abadi, M. Fezari, and R. Hamdi, “Bag of

Visualwords and Chi-squared kernel support vector machine: A way to improve hand gesture recognition," in *Proceedings Proceedings of the International Conference on Intelligent Information Processing, Security and Advanced Communication*, 2015, pp. 1–5.

[5] W. Wu, M. Shi, T. Wu, D. Zhao, S. Zhang, and J. Li, "Real-time hand gesture recognition based on deep learning in complex environments," presented at the Chinese Control And Decision Conference (CCDC), Nanchang, China, 2019.

[6] P. S. Neethu, R. Suguna, and D. Sathish. "An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks," *Soft Computing*, vol. 24, pp. 15239–15248, 2020.

[7] M. Rungruanganukul and T. Siriborvornratanakul, "Deep learning based gesture classification for hand physical therapy interactive program," in *Proceedings HCII 2020: Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Posture, Motion and Health*, 2020, pp. 349–358.

[8] V. Bazarevsky and F. Zhang. (2020, September). *On-Device, Real-Time Hand Tracking with MediaPipe*. [Online]. Available: https://ai.googleblog.com/2019/08/on-device-real-time-hand-tracking-with.html

[9] C. Kerdvibulvech, "A methodology for hand and finger motion analysis using adaptive probabilistic models," *Eurasip Journal on Embedded Systems*, vol. 18, 2014.

[10] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann, "BlazeFace: Sub-millisecond neural face detection on mobile GPUs," presented at CVPR Workshop on Computer Vision for Augmented and Virtual Reality, Long Beach, CA, USA, 2019.