บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในการสร้างระบบการสืบค้นประวัติทางการแพทย์โดยใช้เทคนิครู้จำสำหรับหุ่นยนต์ดูแลผู้สูงอายุ ในบทความต่อไปนี้จะกล่าวถึงความรู้พื้นฐานของเทคโนโลยีและเครื่องมือที่ใช้ในการพัฒนา ดังต่อไปนี้ การตรวจจับใบหน้า (Face Detection) วิธีการในการตรวจจับใบหน้า, การรู้จำใบหน้า (Face Recognition) การวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis) Microsoft Cognitive Services

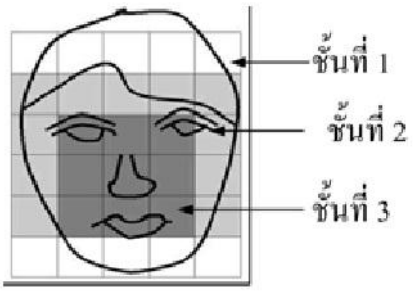
การตรวจจับใบหน้า (Face Detection)

การตรวจจับใบหน้า (Face Detection) คือ กระบวนการค้นหาใบหน้าของบุคคลจากภาพหรือวีดีโอหลังจากนั้นก็ทำการประมวลผลภาพใบหน้าที่ได้ สำหรับขั้นตอนถัดไปเพื่อให้ภาพใบหน้าที่ตรวจจับได้ง่ายต่อการจำแนก และอัลกอริทึมที่ใช้ในการตรวจจับใบหน้าในปัจจุบันก็มีอยู่ด้วยกันหลายวิธีซึ่งอัลริทึมในการตรวจจับใบหน้าที่ดีนั้น มีส่วนช่วยในการจำแนกใบหน้าได้แม่นยำและรวดเร็วขึ้นเป็นอย่างมาก

วิธีการพื้นฐานในการตรวจจับใบหน้า

วิธีการพื้นฐานในการตรวจจับใบหน้าในปัจจุบันสามารถแบ่งออกได้เป็น 4 กลุ่ม จำแนกได้ตามวิธีดังนี้

1. วิธีเชิงความรู้ (Knowledge-based methods) เป็นวิธีการพิจารณาความสัมพันธ์ระหว่างลักษณะเด่นเฉพาะต่างๆบนใบหน้า ที่มีตำแหน่งและองค์ประกอบหลักพื้นฐานตายตัวบนใบหน้า เช่นตัวอย่างงานวิจัยของ Yang and Huang (1994) [1] ศึกษาวิธีเชิงความรู้แบบลำดับชั้น (hierarchical) ซึ่งพิจารณาแบ่งเป็น3 ลำดับชั้น ดังแสดงในภาพที่ 2.1



ภาพที่ 2.1 ขอบเขตการพิจารณาแต่ละลำดับชั้น

ที่มา : (http://kokzard.blogspot.com/2012/04/face-recognition-algorithm.html)

ชั้นแรก พิจารณาหาตำแหน่งความน่าจะเป็นของใบหน้าโดยรวม ชั้นที่ 2 พิจารณาโดยค่าฮิสโตแกรม (histogram)ร่วมกับการหาเส้นขอบ ชั้นที่ 3 พิจารณาลักษณะเด่นภายในของใบหน้าเช่น ความสมมาตรกันของตาทั้ง 2 ข้างตำแหน่งของจมูกและปากในแนวตั้งและแนวนอน

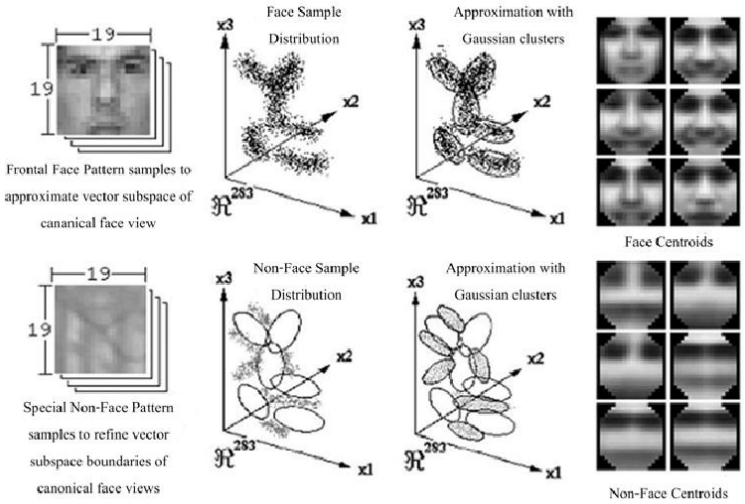
2. วิธีเชิงลักษณะ (Feature-based methods) เป็นการใช้อัลกอริทึมพิจารณาลักษณะเด่นและโครง สร้าง ของใบหน้า รวมทั้งความเปลี่ยนแปลงขององค์ประกอบต่างๆของภาพเช่น แสงเงา เช่น ตัวอย่างงานวิจัยของ Chetverikov and Lerch (1993) [2] ใช้ความแตกต่างของแสงเงาและเส้นแนวเป็นแบบจำลองในการค้นหาใบหน้า โดยแบบจำลองนั้นประกอบด้วยจุดความสว่างน้อย (จุดมืด) 2 จุดเพื่อแสดงดวงตาและจุดความสว่างมา (จุดอ่อน) 3 จุด เพื่อแสดงโหนกแก้มและจมูกแล้วหาความสัมพันธ์ของระยะห่างและตำแหน่งของจุดต่างเพื่อคัดเลือกแบบหน้าที่เหมาะสม ระบบนี้มีข้อจำกัดอยู่ที่เมื่อแสงเงาของสภาวะแวดล้อมเปลี่ยนแปลงจะทำให้ประสิทธิภาพในการค้นหาเปลี่ยนไป

3. วิธีเทียบเคียงแผ่นแบบ (Template matching methods) เป็นการเปรียบเทียบภาพที่ต้องการค้นหากับโครงสร้างแบบจำลองของใบหน้ามาตรฐาน โดยเก็บข้อมูลความสัมพันธ์อย่างอิสระของส่วนต่างๆ บนใบหน้าได้แก่โครงสร้างใบหน้า ตา จมูก และปาก ในท่าหน้าตรง ตัวอย่างเช่น Sakai, Nagao and Fujibayashi, (1969) [3] เสนอโครงสร้างแบบจำลองแผ่นแบบย่อย (Subtemplates model) โดยใช้ตัวกรองโซเบล (Sobel filter) หาเส้นขอบ เพื่อหาตำแหน่งความน่าจะเป็นของส่วนย่อยต่างๆบนใบหน้าที่สามารถเข้ากันได้ดีที่สุดกับแบบจำลองแผ่นแบบย่อยให้เป็นตำแหน่งของใบหน้าที่ต้องการ

4. วิธีเชิงลักษณะปรากฏ (Appearance-based methods) เป็นการเปรียบเทียบภาพที่ต้อง การค้นหากับโครงสร้างแบบจำลองของใบหน้าที่ทำการเรียนรู้และฝึกสอนให้ระบบจดจำและนำความรู้

ในฐานข้อมูลมาใช้ในการพิจารณาโดยสามารถแบ่งย่อยได้หลายวิธีด้วยกันในที่นี้จะทำการเสนอเพียงวิธีการที่ได้รับความนิยมใช้งานกันอย่างกว้างขวาง 2 วิธีดังนี้

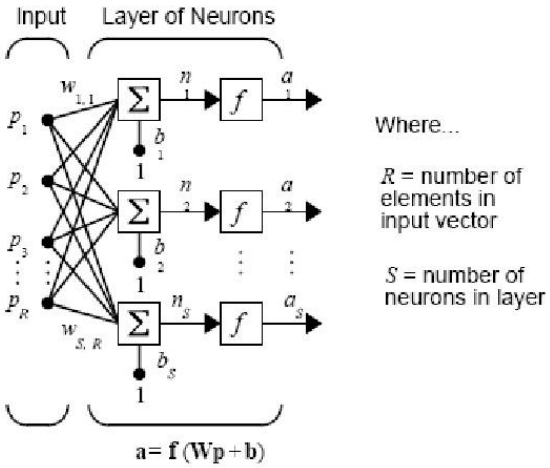
4.1 วิธีหน้าลักษณะเฉพาะ (Eigenface Methods) เป็นการวิเคราะห์องค์ประกอบพื้นฐานของใบหน้า ด้วยกระบวนการทางสถิติของใบหน้าที่หลากหลาย โดยหน้าลักษณะเฉพาะคือเซตของเวกเตอร์ลักษณะเฉพาะ (Eigen vector) ที่สามารถหาได้จากเมทริกซ์ความแปรปรวน ร่วมเกี่ยว (Covariance matrix) สร้างเป็นแบบจำลองของใบหน้าที่รวมเอาลักษณะเด่นต่างๆของภาพใบหน้าตัวอย่างมาร่วมกันเพื่อหาค่าเฉพาะเจาะจงขององค์ประกอบบนใบหน้า ซึ่งในแต่ละบุคคลจะเป็นค่าเฉพาะของบุคคลนั้น ดังแสดงภาพที่ 2.2



ภาพที่ 2.2 การกระจายตัวของข้อมูลตัวอย่าง

ที่มา : (http://kokzard.blogspot.com/2012/04/face-recognition-algorithm.html)

4.2 โครงข่ายประสาทเทียม (Artificial Neural Networks) เป็นการฝึกสอนโครงข่ายด้วยโครงสร้างใบหน้าที่มีความซับซ้อน โดยทำการปรับค่าน้ำหนักประสาทเพื่อให้ค่าความผิดพลาดกำลังสองเฉลี่ยมีค่าลดลง ความถูกต้องของโครงข่ายประสาทเทียมขึ้นกับตัวอย่างที่ทำการฝึกสอน จำนวนชั้นนิวรอนและจำนวนนิวรอนที่เหมาะสม ดังแสดงในภาพที่ 2.3



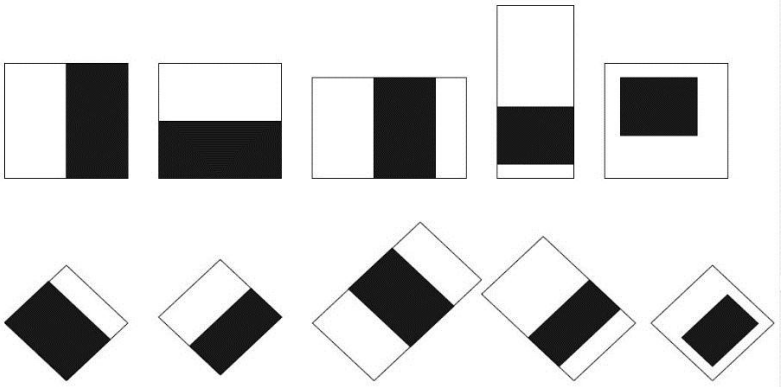
ภาพที่ 2.3 ตัวอย่างโครงข่ายประสาทเทียม

ที่มา : (http://kokzard.blogspot.com/2012/04/face-recognition-algorithm.html)

วิธีการตรวจจับใบหน้าของ Viola-Jones

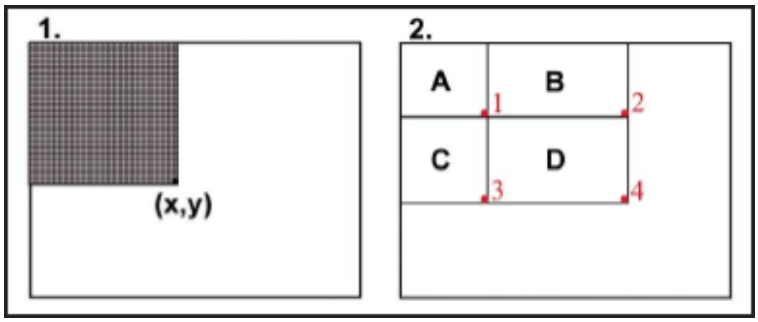
Paul viola และ Michael J. Jones [4]ได้คิดค้นวิธีการหนึ่งที่ใช้ในการตรวจจับใบหน้าที่มีความสามารถในการประมวลผลได้รวดเร็วและมีอัตราความถูกต้องในการตรวจหาสูง คือ Viola-Jones method โดยเทคนิคการตรวจจับใบหน้า Viola-Jones

หลักการพื้นฐานของอัลกอรอทึมของ Viola-Jones คือ การนำภาพที่ต้องการตรวจหาใบหน้ามาแบ่งเป็นภาพย่อย (Sub-window) จากนั้นภาพย่อยดังกล่าวจะถูกพิจารณาเป็นภาพอินพุตของกระบวนการตรวจหาใบหน้า เทคนิคทั่วไปในการตรวจหาใบหน้าจะทำการปรับขนาดของภาพอินพุตแตกต่างกันหลาย ๆ ขนาด และใช้ตัวตรวจจับ (Detector) ที่มีขนาดคงที่ค้นหาวัตถุ ข้อเสียของวิธีนี้ คือ ใช้เวลาในการคำนวณไม่คงที่ แล้ว Viola-Jones จึงได้เสนอวิธีใหม่โดยใช้การปรับขนาดตัวตรวจจับแทนการปรับขนาดรูปภาพอินพุต และใช้ตัวตรวจจับทำการตรวจจับใบหน้าหลายๆรอบ แต่ละรอบใช้ขนาดตัวตรวจจับที่แตกต่างกัน ซึ่งเมื่อทำการเปรียบเทียบกับวิธีเดิมพบว่าใช้เวลาคำนวณไม่แตกต่างกันมากหนัก ดังนั้น Viola-Jonesได้คิดค้นตัวตรวจหาที่ใช้จำนวนครั้งในการคำนวณคงที่แม้จะมีขนาดของภาพแตกต่างกัน โดยตัวตรวจหาดังกล่าวนี้สร้างขึ้นโดยใช้ features ของ Haar wavelets และ Integral Image



ภาพที่ 2.4 Examples of the Haar features

ที่มา : (https://software.intel.com/en-us/node/504530)



ภาพที่ 2.5 The Integral Image

ที่มา : (https://software.intel.com/en-us/node/504530)

การรู้จำใบหน้า (Face Recognition)

การรู้จำใบหน้า (Face Recognition) คือ กระบวนการที่ได้นำภาพใบหน้าที่ตรวจจับได้และประมวลผลแล้วจากขั้นตอนการตรวจจับใบหน้ามาเปรียบเทียบกับฐานข้อมูลของใบหน้าเพื่อระบุว่าใบหน้าที่ตรวจจับได้ตรงกับบุคคลใด

การวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis)

การวิเคราะห์องค์ประกอบหลักหรือ PCA [5] เป็นวิธีทางสถิติที่ใช้สร้างเมตริกซ์ของความแปรปรวนร่วม (Covariance Matrix) จากข้อมูลภาพ โดยจะถูกนำไปใช้ในการสร้างภาพใบหน้าไอเกน (Eigen faces) แสดงดังภาพที่ 2.6

$$A = A_{m\times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & \cdots & a_{mn} \end{bmatrix}$$

ภาพที่ 2.6 การแปลงภาพในรูปภาพในรูปแบบเมตริกซ์ไปเป็นภาพในรูปแบบเวกเตอร์ภาพในรูปแบบเมตริกขนาด N^2xM

ที่มา : (https://wikimedia.org/api/rest_v1/media/math/render/svg/)

โดยข้อมูลภาพใบหน้าขนาด NxN พิกเซล จะถูกแปลงให้อยู่ในรูปแบบของเวกเตอร์ 1 มิติ ขนาด N^2x1 และสร้างเป็นเมตริกซ์จนถึงเวกเตอร์ของชุดข้อมูลที่ m จะเป็นคอลัมน์ที่ m ของเมตริกซ์ ดังนั้นจะได้เมตริกซ์ A ที่มีมิติ n x m โดยที่ m หมายถึง ข้อมูลภาพที่ m ละ n หมายถึงมิติของชุดข้อมูลนั้นๆ และสร้างเมตริกซ์ของความแปรปรวนร่วม (Covariance Matrix) จากข้อมูลภาพ คำนวณหาค่าไอเกนเวคเตอร์ (v) และค่าไอเกน (μ) ของเมตริกซ์ความแปรปรวนร่วมสร้างเป็นภาพใบหน้าไอเกนขึ้นมา

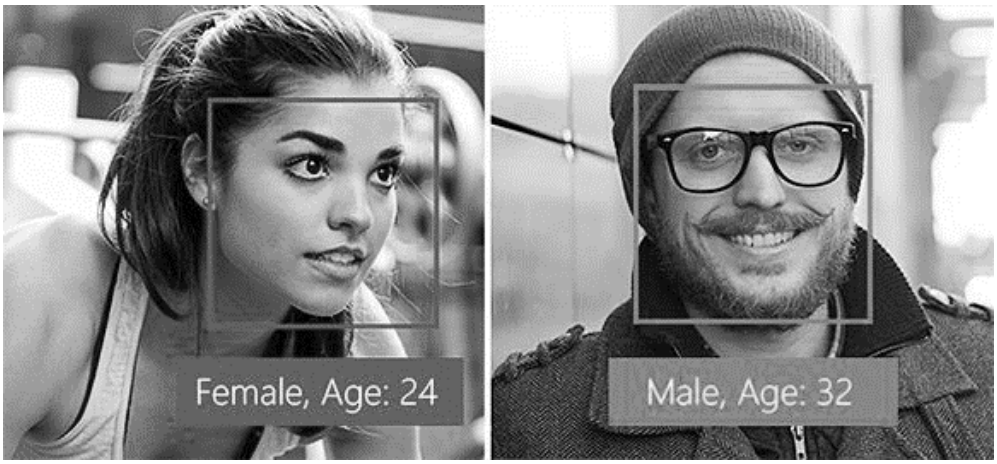
การเรียนรู้ของเครื่อง (Machine Learning)

เป็นสาขาหนึ่งของปัญญาประดิษฐ์ที่พัฒนามาจากการศึกษาการรู้จำแบบ เกี่ยวข้องกับการศึกษาและการสร้างอัลกอริทึมที่สามารถเรียนรู้ข้อมูลและทำนายข้อมูลได้ อัลกอริทึมนั้นจะทำงานโดยอาศัยโมเดลที่สร้างมาจากชุดข้อมูลตัวอย่างขาเข้าเพื่อการทำนายหรือตัดสินใจในภายหลัง แทนที่จะทำงานตามลำดับของคำสั่งโปรแกรมคอมพิวเตอร์

วิธีการเรียนรู้ที่ใช้เป็นการเรียนรู้ด้วยความคล้าย (similarity and metric learning) ในการเรียนรู้ด้วยความคล้ายนั้น เครื่องจะมีตัวอย่างของคู่ที่ถูกมองว่าคล้ายมากและคู่ที่ถูกมองว่าคล้ายน้อย เครื่องจะต้องหาฟังก์ชันความคล้ายออกมาที่สามารถทำนายได้ว่าวัตถุใหม่นั้นมีความคล้ายมากน้อยเพียงใด มักใช้ในระบบแนะนำ (Recommendation System)

ไมโครซอฟต์คอนนิทิฟเซอร์วิส (Microsoft Cognitive Services)

เฟซทเอพีไอ (Face API) สามารถตรวจจับใบหน้าได้ถึง 64 หน้าคนที่มีความเที่ยงตรงสูงที่หันหน้าเข้าหาตำแหน่งในภาพ และภาพสามารถระบุได้โดยไฟล์รูปภาพหรือ ยูอาร์แอล (URL) การแสดงผลรูปกรอบสี่เหลี่ยมบนใบหน้า เพื่อแจ้งให้ทราบว่ามีการหันหน้าเข้าหาตำแหน่งในภาพจะถูกส่งกลับมาพร้อมกับแต่ละใบหน้าที่ตรวจพบ อีกทางเลือกหนึ่งคือให้การตรวจจับใบหน้าจะแยกเป็นชุดข้อมูลที่เกี่ยวข้องกับใบหน้าเช่น ท่าทาง , เพศ , อายุ , ดวงตาและแว่นตา



ภาพที่ 2.7 การตรวจจับใบหน้าของ Microsoft Cognitive Services

ที่มา : (https://docs.microsoft.com/en-us/azure/cognitive-services/face/images/face.detection.jpg)

การใช้จดจำใบหน้า มีการใช้กันอย่างแพร่หลายในหลายสถานการณ์ รวมถึงการรักษาความปลอดภัยที่มีอยู่ตามธรรมชาติ , อินเตอร์เฟซผู้ใช้ที่ใช้การวิเคราะห์เนื้อหาภาพและการจัดการแอปพลิเคชันบนมือถือ และการออกแบบหุ่นยนต์ มีฟังก์ชั่นในการทำงานอยู่ 4 แบบ ดังนี้ การใช้การจดจำใบหน้าจะมีฟังก์ชันการตรวจสอบใบหน้า : ทำให้การค้นหาใบหน้าที่คล้ายกันและเห็นหน้าการจัดกลุ่มและการระบุตัวตนของบุคคลนั้นได้

1. การยืนยันใบหน้า (Face Verification)

การตรวจสอบความถูกต้องของ Face API จะนำรูปใบหน้าที่จัดเก็บอยู่ในฐานข้อมูลกับใบหน้าที่ตรวจพบ มาทำการตรวจสอบความถูกต้องหรือการตรวจสอบความถูกต้องจากที่หนึ่งไปยังใบหน้าที่ตรวจพบวัตถุหนึ่งคน

2. ค้นหาใบหน้าที่คล้ายกัน (Finding Similar Face)

เป้าหมายที่ระบุใบหน้าที่ตรวจพบและการตั้งค่าของผู้สมัครงานที่หันหน้าเข้าหาเพื่อค้นหาด้วยการให้บริการของเราพบว่ามีการตั้งค่าขนาดเล็กที่มีใบหน้าที่มีลักษณะคล้ายกันมากที่สุดไปที่ใบหน้าของชิ้นงาน โหมดการทำงานทั้งสองให้ แม็ชเฟซ (matchFace) , แม็ชเฟซท (matchPerson) และได้รับการสนับสนุนโหมดที่คล้ายกันใบหน้า แม็ชเพอร์ชัน (matchPerson) หลังจากการนำตัวบุคคลเดียวกันกับเกณฑ์ขั้นต่ำที่ได้มาจากใบหน้า - ตรวจสอบโหมด แม็ชเฟซ (matchFace) จะละเว้นค่าเกณฑ์ของบุคคลเดียวกันที่ได้รับและส่งกลับด้านบนใบหน้าของผู้สมัครงานที่คล้ายกัน นำตัวอย่างต่อไปนี้ให้ใบหน้าของกลุ่มบุคคลจะแสดงอยู่ในรายการ ดังแสดงภาพที่ 2.8

ภาพที่ 2.8 ภาพกลุ่มตัวบุคคล

ที่มา : (https://docs.microsoft.com/en-us/azure/cognitive-services/face/images/facefindsimilar.candidates.jpg)

ภาพที่ 2.9 ใบหน้าที่ใช้เปรียบเทียบ

ที่มา : (https://docs.microsoft.com/en-us/azure/cognitive-services/face/images/facefindsimilar.queryface.jpg)

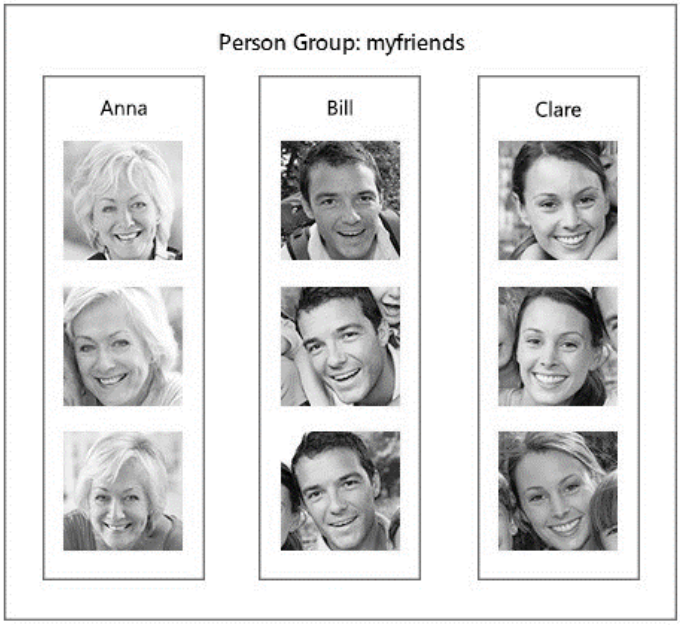
จากภาพที่ 2.9 ในการค้นหาใบหน้าที่คล้ายกันทั้งหมด 4 โหมด แล้วแม็ชเพอร์ชัน (matchPerson) ส่ง (a) และ (b) ซึ่งเป็นของบุคคลเดียวกันกับใบหน้าที่สืบค้นโหมด แม็ชเฟซ (matchFace) ส่งกลับค่า (a) (b) (c) และ (d) ให้ตรง 4 กลุ่มบุคคลแม้ว่าความคล้ายกันต่ำ

3. การจัดกลุ่มใบหน้า (Face Grouping)

ได้รับการตั้งค่าอย่างใดอย่างหนึ่งของใบหน้าที่ไม่รู้จักใบหน้าที่จัดกลุ่ม เฟซทเอพีไอ (Face API) โดยอัตโนมัติซึ่งจะแบ่งเป็นหลายกลุ่มตามความคล้ายกัน แต่ละกลุ่มเป็นกลุ่มย่อยของที่เหมาะสมที่แยกออกจากกันที่ไม่รู้จักใบหน้าที่ตั้งค่าเดิมและมีใบหน้าที่คล้ายกัน และใบหน้าทั้งหมดที่อยู่ในกลุ่มเดียวกันสามารถได้รับการพิจารณาว่าเป็นของบุคคลคนเดียวกับออบเจกต์

4. การระบุใบหน้า (Face Identification)

เฟซทเอพีไอ (Face API) ที่ใบหน้าที่สามารถใช้ในการระบุตัวบุคคลตามใบหน้าที่ตรวจพบและคนที่ฐานข้อมูล บุคคลที่ได้รับการกำหนดให้เป็นกลุ่ม ซึ่งจะต้องได้รับการสร้างขึ้นในการล่วงหน้าและสามารถแก้ไขได้เมื่อเวลาผ่านไป



ภาพที่ 2.10 ภาพกลุ่มบุคคลชื่อว่า “เพื่อนของฉัน”

ที่มา : (https://docs.microsoft.com/en-us/azure/cognitive-services/face/images/person.group.clare.jpg)

จากภาพที่ 2.10 เป็นตัวอย่างของกลุ่มบุคคลที่มีชื่อว่า " เพื่อนของฉัน " แต่ละกลุ่มอาจมีได้ถึง 1,000 คนที่ออบเจกต์ ในขณะเดียวกันให้แต่ละบุคคลที่ออบเจกต์สามารถมีหนึ่งหรือหลายใบหน้าที่ลงทะเบียนไว้ หลังจากที่มีการสร้างกลุ่มบุคคลที่ผ่านการฝึกอบรมและได้รับการระบุตัวตนสามารถทำได้กับกลุ่มและใบหน้าที่ตรวจพบใหม่ หากใบหน้าที่ถูกระบุว่าเป็นบุคคลที่ออบเจกต์ในกลุ่มบุคคลที่ออบเจกต์จะถูกส่งคืน

การสังเคราะห์เสียง (Text to Speech)

ซอฟต์แวร์สังเคราะห์เสียง เป็นเทคโนโลยีที่สามารถสร้างเสียงคำพูด ได้ตามความต้องการ ซึ่งในการใช้งานส่วนใหญ่จะต้องใช้งานร่วมกับเทคโนโลยีด้านการประมวลผลภาษา (Language Processing Technology) ทำให้ได้เทคโนโลยีสังเคราะห์เสียงจากข้อความ (Text-to-Speech Synthesis: TTS) ที่สามารถนำไปประยุกต์ใช้กับข้อความภาษาไทย เพื่อหาวิธีอ่านข้อความแล้วแปลงข้อความจากตัวหนังสือภาษาไทยให้เป็นเสียงพูดภาษาไทย ซอฟต์แวร์สังเคราะห์เสียงพูดภาษาไทยคุณภาพสูงสามารถสังเคราะห์เสียงพูดภาษาไทยได้ทุกคำ เนื่องจากมีส่วนวิเคราะห์คำอ่านที่สามารถวิเคราะห์ได้แม้แต่คำที่ไม่เคยปรากฏในพจนานุกรม นอกจากนี้ผู้ใช้สามารถเพิ่มคำเฉพาะเช่นชื่อบุคคล พร้อมทั้งกำหนดคำอ่านได้อย่างอิสระ เพื่อให้ซอฟต์แวร์สามารถแปลงข้อความมาเป็นเสียงพูดได้ตรงกับความต้องการของผู้ใช้

งานวิจัยที่เกี่ยวข้อง

รุสลี่ สุทธวีร์กูล และ วิไลพร แซ่ลี้ (2554) กล่าวว่า การตรวจจับใบหน้าด้วยวิธีการพื้นฐานของการจำลองรูปแบบ Haar-like ตั้งแต่อดีตจนถึงปัจจุบัน วิธีการนี้ได้ถูกนำเสนอเป็นครั้งแรกโดย Viola-Jones ในปี 2001 วิธีการตรวจจับใบหน้าของ Viola-Jones ประกอบด้วย 3 ขั้นตอน คือ การคำนวณการจำลองรูปแบบ Haar-like ด้วย Integral Image การค้นหาการจำลองรูปแบบ Haar-like ด้วย Adaboost และการรวมตัวจำแนกกลุ่มแบบต่อเรียง (Cascade Classifier) ซึ่งในการใช้การจำลองรูปแบบ Haar-like นั้นถือว่ามีความสำคัญต่อความแม่นยำที่สุดเพราะเป็นเครื่องมือที่ใช้ในการดึงลักษณะเด่นจากใบหน้า

นาย วิรัตน์ พรหมเมศ และ นายนพพณ เลิศชู วงศา (2558) กล่าวว่า ด้วยเทคนิค Eigen face และการปรับเปลี่ยนสภาวะแสง เป็นการประมวลผลภาพใบหน้าบุคคล เพื่อใช้กับระบบเปิดปิดประตูอัตโนมัติ ระบบประกอบด้วย computer และระบบฝังตัว raspberry pi raspberry pi ทำหน้าที่เป็น client ที่ติดต่อกับกล้องและส่งสัญญาณภาพจากกล้องไปยังคอมพิวเตอร์รวมถึงส่งสัญญาณควบคุมอุปกรณ์ต่างๆ computer ทำหน้าที่เป็น server ประมวลผลข้อมูลภาพที่มาจาก raspberry pi