

## การตรวจจับใบหน้าด้วยวิธีการพื้นฐานของการจำลองรูปแบบ Haar-like

### Face Detection based-on Haar-like Features

รุสลี สุทธีวีร์กุล วิลไลพร แซ่ลี

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ บางซื่อ กรุงเทพฯ ประเทศไทย

E-mail: ruslees@kmutnb.ac.th, wilaiporn.lee@gmail.com

#### บทคัดย่อ

บทความวิชาการฉบับนี้เป็นบททบทวนเอกสาร และงานวิจัยเกี่ยวกับการตรวจจับใบหน้าด้วยวิธีการพื้นฐานของการจำลองรูปแบบ 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 นั้นถือว่ามีผลต่อความแม่นยำที่สุด เพราะเป็นเครื่องมือที่ใช้ในการดึงลักษณะเด่นจากใบหน้า จึงมีงานวิจัยเป็นจำนวนมากที่มุ่งเน้นในการพัฒนาการจำลองรูปแบบ Haar-like ดังนั้น บทความฉบับนี้จึงทำการรวบรวมและสรุปงานวิจัยที่มุ่งเน้นในการเพิ่มหรือปรับปรุงรูปร่างของการจำลองรูปแบบ Haar-like ซึ่งเป็นผลให้เพิ่มความเร็วและความแม่นยำในการตรวจจับใบหน้า หรือเพิ่มความสามารถในการตรวจจับใบหน้าในมุมอื่น ๆ นอกเหนือจากใบหน้าตรง

**คำสำคัญ:** การตรวจจับใบหน้า เทคนิค Viola-Jones การจำลองรูปแบบ Haar-like

#### ABSTRACT

This article presents a review of papers and researches on face detection based-on Haar-like features. The Haar-like features for face detection is proposed by Viola and Jones since 2001. This technique consists of 3 steps: Integral Image, Adaboost and Cascade Classifier. Since the Haar-like features is the most effective application used to extract features from faces, there are many researches aiming to apply the Haar-like features to achieve higher performance. The article also focuses on research papers which extend or adapt formations of the Haar-like features to get higher speed and accuracy or to detect faces in poses other than frontal faces.

**Keyword:** Face detection, Viola-Jones algorithm, Haar-like features

#### 1. บทนำ

ด้วยเทคโนโลยีการสื่อสาร และระบบสารสนเทศที่ได้รับการพัฒนาขึ้นในปัจจุบัน ส่งผลให้วิธีการต่อประสานระหว่างมนุษย์และคอมพิวเตอร์ (Human

Machine Interface) ได้รับการพัฒนา โดยที่ไม่ได้ถูกจำกัดด้วยอุปกรณ์เดิม ๆ เช่น คีย์บอร์ด เมาส์ หรือจอภาพ นอกจากนั้นปัจจัยทางด้านราคาของคอมพิวเตอร์ส่วนบุคคล และระบบสมองกลฝังตัว

(Embedded Systems) ที่ลดลงอย่างต่อเนื่อง ก็ยังส่งผลให้ระบบการมองของคอมพิวเตอร์ (Computer Vision) สามารถพัฒนาใช้ในระดับผู้ใช้งานทั่วไปมากยิ่งขึ้น ดังนั้นการขยายตัวของงานวิจัยที่เกี่ยวข้องกับใบหน้ามนุษย์จึงมีจำนวนมากขึ้น โดยเฉพาะในด้านการระบุตัวตนด้วยใบหน้า (Face Recognition) สถานะของหน้า (Face Status) หรือการแสดงออกทางใบหน้า (Face Expression)

สำหรับขั้นตอนการตรวจจับใบหน้านั้นถือเป็นขั้นตอนแรกของการวิเคราะห์ใบหน้า ซึ่งเป็นขั้นตอนที่มีความสำคัญ และมีปัจจัยที่มีผลกระทบมากมาย เช่น สเกล ตำแหน่ง ทิศทาง ลักษณะสีหน้า การถูกบดบัง และสภาพแสง ซึ่งทั้งหมดนี้จะมีผลต่อลักษณะการปรากฏของใบหน้าในรูป อีกทั้งในการนำไปใช้งานจริงยังต้องการความเร็ว และความแม่นยำที่สูงอีกด้วย จึงเป็นสิ่งท้าทายและมีงานวิจัยจำนวนมากที่นำเสนอวิธีการแก้ปัญหา และเพิ่มประสิทธิภาพการตรวจจับใบหน้า

ในบทความวิชาการฉบับนี้กล่าวถึงวิธีการพื้นฐานในการตรวจจับใบหน้าในหัวข้อที่ 2 ซึ่งแบ่งเป็น 4 กลุ่ม จำแนกตามวิธีการ คือ วิธีการ Knowledge-based วิธีการ Feature invariant วิธีการ Template Matching และวิธีการ Appearance - based โดยในงานวิจัยที่เกี่ยวข้องในรอบทศวรรษที่ผ่านมาวิธี Appearance-based เป็นวิธีที่มีการพัฒนามากที่สุด และเป็นที่ยอมรับสามารถนำมาใช้งานได้จริงในหลายๆ ด้าน ในหัวข้อที่ 3 กล่าวถึงวิธีการตรวจจับใบหน้าของ Viola-Jones ที่มีประสิทธิภาพสูงโดยนำเสนอเทคนิค 3 ขั้นตอน คือ การคำนวณรูปแบบการจำลองด้วย Integral Image การค้นหารูปแบบการจำลองด้วย Adaboost และการรวมตัวจำแนกกลุ่มแบบต่อเรียง ในส่วนงานวิจัยที่เกี่ยวข้องกับการศึกษาและพัฒนาลักษณะการจำลองรูปแบบ Haar-like ได้กล่าวในหัวข้อที่ 4 และในหัวข้อสุดท้ายเป็นบทสรุปของบทความ

## 2. วิธีการพื้นฐานในการตรวจจับใบหน้า

วิธีการพื้นฐานในการตรวจจับใบหน้าในปัจจุบันสามารถแบ่งได้เป็น 4 กลุ่ม [1] จำแนกตามวิธีการได้ดังนี้

### 2.1 Knowledge-based methods

วิธีการนี้ใช้กฎเกณฑ์พื้นฐานความรู้ของมนุษย์ของส่วนประกอบสำคัญบนใบหน้า เช่น ในงานวิจัยของ Yang ในปี 1994 [2] และ งานวิจัยของ Kotropoulos และ Pitas ในปี 1997 [3] ได้ใช้กฎเกณฑ์ประกอบในการตัดสินใจ โดยที่สังเกตจากโพรไฟล์ความเข้มของพิกเซลภาพทั้งในแนวตั้ง และแนวนอน ซึ่งในรูปที่มีใบหน้าจะต้องประกอบด้วยลักษณะของดวงตา 2 ดวง ซึ่งอยู่ในตำแหน่งที่สมมาตรกันในแนวระนาบทั้งสองด้าน จะต้องมียอดหนึ่งจมูก และหนึ่งปาก ซึ่งจุดสำคัญดังกล่าวจะมีความสัมพันธ์กันโดยใช้ระยะทางและตำแหน่งตามกฎเกณฑ์ที่ได้ตั้งไว้

อุปสรรคที่สำคัญของวิธีการนี้ คือ การหากฎเกณฑ์ที่จำเพาะเจาะจงซึ่งใช้ในการจำแนกหน้าของมนุษย์นั้นทำได้ยาก เพราะถ้าหากกฎเกณฑ์นั้นระบุรายละเอียดมากเกินไปจะมีผลทำให้การตรวจจับใบหน้าทำได้ยาก (เนื่องจากไม่มีใบหน้าใดผ่านเกณฑ์เลย) หรือถ้าเกณฑ์ของใบหน้านั้นรายละเอียดน้อยเกินไปก็จะทำให้การตรวจจับใบหน้าเกิดข้อผิดพลาดขึ้นได้ง่าย (จะตรวจจับสิ่งที่ไม่ใช่ใบหน้าด้วย) และนอกเหนือจากนั้นในการหากฎเกณฑ์สำหรับการตรวจจับใบหน้าที่มีตำแหน่งทิศทางนอกเหนือไปจากใบหน้าตรงนั้นก็ยิ่งทำได้ยากและซับซ้อนมาก

### 2.2 Feature invariant approaches

วิธีการนี้ใช้สำหรับการค้นหาลักษณะเด่น (Feature) ที่แตกต่างกันของแต่ละใบหน้าในการตรวจจับสมมุติฐานของวิธีดังกล่าวตั้งอยู่บนพื้นฐานของการสังเกตของมนุษย์ซึ่งสามารถที่จะจดจำวัตถุใด ๆ ได้ไม่ว่าจะมีการเปลี่ยนแปลงตำแหน่งการจัดวาง หรือสภาพแสงที่แตกต่างกัน เป็นต้น ซึ่งจะต้องมีคุณลักษณะบางอย่างที่ไม่ขึ้นอยู่กับผลของการเปลี่ยนแปลงดังกล่าว ในการตรวจหาลักษณะเด่นบนใบหน้า มีวิธีการอยู่หลายวิธีที่มีการนำเสนอ เช่น ในงานวิจัยของ Yow และ Cipolla ในปี 1998 [4] ลักษณะเด่นของขนคิ้ว ดวงตา จมูก ปาก และเส้นผม จะถูกดึง (Extract) ออกมา โดยการตรวจจับขอบ (Edge Detection) จากนั้นจึงใช้โมเดลทางสถิติในการอธิบายถึงความสัมพันธ์ และยืนยันการตรวจพบใบหน้า

ปัญหาหนึ่งของวิธีการนี้คือ คุณลักษณะบนใบหน้าสามารถถูกทำลายได้โดยง่ายด้วยสภาพแสง สัญญาณรบกวน และการถูกบดบัง อีกทั้งขอบของจุดเด่นบนใบหน้าอาจจะตรวจพบได้ยากในบางกลุ่มใบหน้า รวมถึงสภาพแสงเงาก็ยังมีผลกระทบโดยตรงกับการตรวจจับขอบของวัตถุ

### 2.3 Template matching methods

วิธีการนี้ใช้แม่แบบมาตรฐาน (Standard Template) ของใบหน้า (โดยปกติเป็นหน้าตรง) ที่ถูกกำหนดเองด้วยมือ (Manual) หรือ กำหนดโดยฟังก์ชัน เช่น ในงานวิจัยของ Scassellati ในปี 1998 [5] ได้กำหนดแม่แบบของใบหน้าซึ่งประกอบด้วย 16 พื้นที่ และ 23 ความสัมพันธ์ ใบหน้าที่รับเข้ามาจะนำมาหาค่าสหสัมพันธ์ (Correlation Value) กับใบหน้าที่เป็นรูปแบบมาตรฐาน (Standard Pattern) ในส่วนของโครงหน้า ดวงตา จมูก และปาก อย่างอิสระจากกัน การปรากฏอยู่ของใบหน้าในรูปภาพนั้นหาได้จากค่าสหสัมพันธ์เป็นหลัก วิธีการนี้มีข้อดีคือค่อนข้างทำได้ง่าย แต่ประสิทธิภาพการตรวจจับใบหน้าที่ยังไม่ดีเท่าที่ควรซึ่งเกิดจากผลของขนาด ตำแหน่งการวาง และรูปทรงของใบหน้าที่แตกต่างจากแม่แบบมาตรฐาน

### 2.4 Appearance-based methods

วิธีการนี้จะใช้โมเดลที่ได้มาจากการเรียนรู้ของกลุ่มของรูปภาพตัวอย่าง ทั้งที่เป็นใบหน้า และไม่ใช่อใบหน้า โดยทั่วไปแล้ววิธีการนี้จะอยู่บนพื้นฐานของเทคนิคการวิเคราะห์ทางสถิติ และการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งใช้ในการหาคุณลักษณะความสัมพันธ์ของรูปใบหน้า และรูปที่ไม่ใช่ใบหน้า คุณลักษณะที่ได้รับการเรียนรู้จะถูกจัดในรูปแบบของแบบจำลองการกระจาย (Distribution Models) หรือ ฟังก์ชันการจำแนก (Discriminant Functions) ในขณะที่เดียวกันการลดมิติ (Dimension Reduction) ก็ถูกนำมาใช้เพื่อเพิ่มประสิทธิภาพในการคำนวณ และการตรวจจับใบหน้าที่อีกด้วย เช่น ในงานวิจัยของ Sung และ Poggio ในปี 1998 [6] ได้ประมาณฟังก์ชันความหนาแน่นของรูปแบบหน้า และไม่ใช่อหน้าด้วยการใช้กลุ่มของเกาส์เซียน (Gaussian)

จากการศึกษาในเรื่องการตรวจจับใบหน้าในรูปทศวรรษที่ผ่านมาพบว่าวิธีการ Appearance-based แสดงให้เห็นถึงประสิทธิภาพที่ดีกว่าวิธีการอื่น ๆ โดยเฉพาะงานวิจัยที่นำเสนอโดย Viola และ Jones ในปี 2001 [7-8] มีความสามารถที่จะตรวจจับใบหน้าที่อย่างรวดเร็ว มีประสิทธิภาพ และสามารถนำไปใช้ได้งานได้จริง เช่น ใช้ในกล้องถ่ายภาพดิจิทัล และใช้ในซอฟต์แวร์จัดการรูปภาพ เป็นต้น

## 3. วิธีการตรวจจับใบหน้าของ Viola-Jones

Paul viola และ Michael J. Jones ได้นำเสนอเทคนิคการตรวจจับใบหน้าที่มีความเร็วและมีความถูกต้องในการตรวจจับสูงในปี 2001 [7-8] เทคนิคการตรวจจับใบหน้าของ Viola-Jones เป็นเทคนิคที่ได้รับการยอมรับและรู้จักในงานวิจัยเรื่องการตรวจจับใบหน้าเป็นอย่างมาก [9-17] โดยเทคนิคการตรวจจับใบหน้า Viola-Jones นี้สามารถแบ่งออกเป็น 3 ขั้นตอนการคำนวณรูปแบบการจำลองด้วย Integral Image การค้นหาแบบจำลองด้วย Adaboost และการรวมตัวจำแนกกลุ่มแบบต่อเรียง (Cascaded Classifier) โดยแต่ละขั้นตอนมีรายละเอียดดังต่อไปนี้

หลักการพื้นฐานของเทคนิคการตรวจจับใบหน้าของ Viola-Jones คือ การนำภาพที่ต้องการตรวจหาใบหน้าที่มาแบ่งเป็นภาพย่อย (Sub-window) จากนั้นภาพย่อยดังกล่าวจะถูกพิจารณาเป็นภาพอินพุตของกระบวนการตรวจหาใบหน้า เทคนิคทั่วไปในการตรวจหาใบหน้าที่จะทำการปรับขนาดของภาพอินพุตแตกต่างกันหลาย ๆ ขนาด และใช้ตัวตรวจจับ (Detector) ที่มีขนาดคงที่ค้นหาวัตถุ ข้อเสียของการตรวจจับใบหน้าที่แบบนี้คือ ระยะเวลาในการคำนวณไม่คงที่ ดังนั้น Viola-Jones จึงเสนอเทคนิคการตรวจจับใบหน้าที่ใหม่ โดยการใช้การจำลองรูปแบบ Haar-like เป็นตัวตรวจจับ ทำการปรับขนาดของตัวตรวจจับแทนการปรับขนาดภาพอินพุต และใช้ตัวตรวจจับทำการตรวจจับใบหน้าหลาย ๆ รอบ โดยแต่ละรอบใช้ขนาดของตัวตรวจจับแตกต่างกัน ซึ่งเมื่อทำการเปรียบเทียบกับวิธีเดิมพบว่า เวลาที่ใช้ในการคำนวณไม่ได้แตกต่าง

กันมาก แต่ใช้เวลาในการคำนวณการตรวจจับภาพ  
ใบหน้าแต่ละรอบมีค่าคงที่ แม้ขนาดของตัวตรวจจับจะ  
แตกต่างกันก็ตาม เทคนิคที่ Viola-Jones นำเสนอนี้จะ  
ทำการคำนวณการจำลองรูปแบบ Haar-like ด้วย  
Integral Image

### 3.1 ขั้นตอน Integral image

ตัวอย่างของการจำลองรูปแบบ Haar-like ที่ใช้ใน  
เทคนิคการตรวจจับใบหน้าของ Viola-Jones แสดงใน  
รูปที่ 1 โดยแต่ละลักษณะเด่นจะประกอบด้วยพื้นที่  
สี่เหลี่ยม 2 ประเภทคือ ส่วนที่แรงเงา และส่วนที่ไม่ได้  
แรงเงา การหาค่าการจำลองรูปแบบ Haar-like คือการ  
หาผลต่างระหว่างความเข้มในส่วนที่แรงเงากับส่วนที่  
ไม่ได้แรงเงา จากนั้นนำผลลัพธ์ที่ได้ไปเปรียบเทียบกับ  
ค่าขีดแบ่ง (Threshold) กับขั้ว (Polarity) ที่ใช้ในการ  
ตัดสินใจว่าภาพที่รับเข้ามาควรถูกจัดให้เป็นบวก  
(ภาพใบหน้า) หรือเป็นลบ (ไม่ใช่ภาพใบหน้า) ผลลัพธ์  
ที่ได้จะถูกนำไปพิจารณาในขั้นตอนต่อไป

ตัวอย่างการจำลองรูปแบบ Harr-like ที่แสดงใน  
รูปที่ 1 มี 3 แบบคือ การจำลองรูปแบบ Haar-like ที่  
ประกอบด้วยพื้นที่สี่เหลี่ยมสองภาพหรือ Two-rectangle  
feature (รูปที่ 1ก และ 1ข) การจำลองรูปแบบ Haar-  
like ที่ประกอบด้วยพื้นที่สี่เหลี่ยมสามภาพ หรือ  
Three-rectangle feature (รูปที่ 1ค) และการจำลอง  
รูปแบบ Haar-like ที่ประกอบด้วยพื้นที่สี่เหลี่ยมสี่ภาพ  
หรือ Four-rectangle feature (รูปที่ 1ง)

ในรูปที่ 2 เป็นวิธีการทั่วไปในการหาค่าผลรวม  
ความเข้มของพื้นที่สี่เหลี่ยมใด ๆ ในรูปภาพ สามารถ  
เขียนแทนได้ด้วยสมการต่อไปนี้

$$i_s(x_1, x_2, y_1, y_2) = \sum_{x=x_1}^{x_2} \sum_{y=y_1}^{y_2} i(x, y) \quad (1)$$

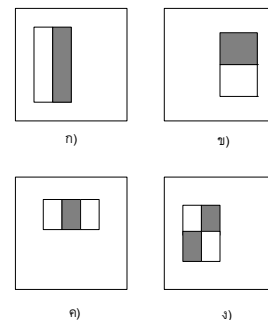
โดยที่  $i(x, y)$  แทนค่าความเข้มที่จุด  $x$  และ  $y$   
ในรูปภาพ  $i_s(x_1, x_2, y_1, y_2)$  แทนผลรวมความเข้ม  
ภายในพื้นที่สี่เหลี่ยม โดยที่  $x_1$   $x_2$   $y_1$  และ  $y_2$  เป็น  
พิกัดมุมทั้งสี่ของรูปสี่เหลี่ยม จากสมการที่ 1 จะเห็น  
ได้ว่าเวลาที่ใช้ในการคำนวณนั้นจะขึ้นอยู่กับขนาดของ  
รูปสี่เหลี่ยม

ในการเพิ่มประสิทธิภาพ Viola และ Jones ได้  
นำเสนอให้ใช้เทคนิคที่เรียกว่า Integral Image ในการ  
คำนวณการจำลองรูปแบบ Haar-like เทคนิค Integra  
image คือ การรวมความเข้ม (Intensity) ของแต่ละ  
พิกเซลเข้าด้วยกัน รูปที่ 3ก) แสดง Integral image  
ของรูป 3ข) การคำนวณหา Integral image ที่จุด  
 $(x, y)$  สามารถเขียนแทนได้ด้วยสมการต่อไปนี้

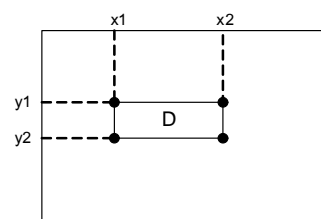
$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (2)$$

เมื่อ  $ii(x, y)$  คือ ค่าของ Integral image ที่ตำแหน่ง  
จุดที่ตำแหน่งจุด  $(x, y)$  และ  $i(x', y')$  คือ ค่าความเข้ม  
ในแต่ละพิกเซลของภาพต้นฉบับ

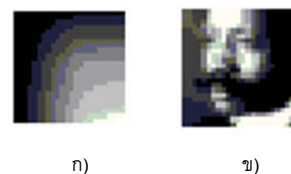
ตัวอย่างการคำนวณหาค่าผลรวมความเข้มของ  
พื้นที่สี่เหลี่ยม D โดยใช้ค่า Integral image จำนวน 4  
จุด ผลลัพธ์ที่ได้คือ  $4+1-(2+3)$  ดังแสดงในรูปที่ 4 ซึ่ง  
จะเห็นได้ว่าไม่ว่าสี่เหลี่ยมจะมีขนาดเท่าใดก็จะใช้เวลา  
ในการคำนวณค่าผลรวมความเข้มคงที่



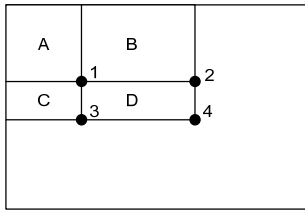
รูปที่ 1 ตัวอย่างการจำลองรูปแบบ Haar-like [7-8]



รูปที่ 2 ตัวอย่างการคำนวณความเข้มในพื้นที่สี่เหลี่ยม  
D โดยวิธีการทั่วไป



รูปที่ 3 ตัวอย่างภาพ Integral image



รูปที่ 4 ตัวอย่างการคำนวณผลรวมความเข้มในพื้นที่สี่เหลี่ยม D โดยวิธี Integral image

กำหนดให้การจำลองรูปแบบ Haar-like ที่ได้หลังจากทำการเปรียบเทียบกับค่าขีดแบ่งเรียกว่า ตัวจำแนกแบบอ่อนแอ (Weak Classifier) เนื่องจากการจำลองรูปแบบ Haar-like ที่ได้รับจะให้คำตอบที่ถูกต้องมากกว่าการเดาสุ่มเพียงเล็กน้อย ซึ่งการหาตัวจำแนกแบบอ่อนแอสามารถเขียนแทนด้วยสมการต่อไปนี้

$$h(x_i, f, p, \theta) = \begin{cases} 1 & \text{if } pf(x_i) < p\theta \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

เมื่อ  $x$  คือ ภาพย่อย โดยในงานวิจัยนี้ [7-8] ได้กำหนดให้มีขนาดเท่ากับ  $24 \times 24$  พิกเซล ส่วน  $h(x, f, p, \theta)$  คือ ตัวจำแนกแบบอ่อนแอที่พิจารณาจากการจำลองรูปแบบ Haar-like ( $f$ ) เทียบกับค่าขีดแบ่ง ( $\theta$ ) และ ( $p$ ) คือ ขั้วซึ่งเป็นตัวกำหนดทิศของสมการ

### 3.2 ขั้นตอนการเรียนรู้ด้วยวิธี Adaboost

เทคนิคการตรวจจับใบหน้าของ Viola-Jones ในขั้นตอนนี้จะนำการจำลองรูปแบบ Haar-like ที่ได้จากขั้นตอนแรกมาเข้ากระบวนการเรียนรู้ของเครื่อง (Machine Learning) ที่เรียกว่า “Adaptive Boost” หรือ “Adaboost” ซึ่งกระบวนการนี้ใช้การเร่งหาตัวจำแนกแบบอ่อนแอ และกำหนดค่าถ่วงน้ำหนักที่ทำให้ค่าผิดพลาดน้อยที่สุดในแต่ละรอบของกระบวนการ เพื่อสร้างตัวจำแนกแบบแข็งแกร่ง (Strong Classifier)

ขั้นตอนการเรียนรู้ด้วย Adaboost มีดังต่อไปนี้

1. พิจารณาเลือกกลุ่มภาพตัวอย่างจำนวน  $N$  ภาพ เพื่อใช้ในเรียนรู้ กลุ่มตัวอย่างภาพสามารถเขียนได้ดังสมการข้างล่าง

$$S = \{(x_i, z_i), i = 1, 2, \dots, N\} \quad (4)$$

เมื่อ  $x_i$  คือ ภาพตัวอย่างที่  $i$  และ  $z_i$  คือ ค่าที่ระบุว่าภาพตัวอย่างเป็นภาพที่เป็นใบหน้าคน ( $z_i = 1$ ) หรือไม่ใช่ใบหน้าคน ( $z_i = 0$ )

2. กำหนดจำนวนตัวจำแนกแบบอ่อนแอมีค่าเท่ากับ  $t = 1, 2, \dots, T$  รอบ
3. กำหนดค่าถ่วงน้ำหนักเริ่มต้น ( $w_{t,i}$ ) ตัวจำแนกแบบอ่อนแอตัวแรก (เมื่อ  $t = 1$ ) มีค่าเท่ากับ  $w_{1,i} = \frac{1}{2m}$ ;  $\frac{1}{2l}$  สำหรับ  $z_i = 0, 1$  เมื่อ  $m$  และ  $l$  คือ จำนวนภาพที่เป็นใบหน้าและไม่ใช่ใบหน้าตามลำดับ
4. ขั้นตอนการพิจารณาตัวจำแนกแบบอ่อนแอทีละค่ามีดังต่อไปนี้

a. ทำการนอร์มัลไลซ์ (Normalize) ค่าถ่วงน้ำหนัก  $w_{t,i} \approx \frac{w_{t,i}}{\sum_{j=1}^n w_{t,i}}$

b. ค้นหาตัวจำแนกแบบอ่อนแอที่ทำให้ค่าความผิดพลาด  $\epsilon_t$  ต่ำที่สุดตามสมการต่อไปนี้

$$\epsilon_t = \min_{f,p,\theta} \sum_i w_i |h(x_i, f, p, \theta) - z_i| \quad (5)$$

c. กำหนดให้  $h_t(x) = h(x, f_t, p_t, \theta_t)$  เมื่อ  $f_t, p_t$  และ  $\theta_t$  คือค่าที่ทำให้ได้ค่าต่ำที่สุดของ  $\epsilon_t$

d. จากนั้นทำการปรับปรุงค่าถ่วงน้ำหนักของแต่ละตัวอย่างใหม่มีค่าเท่ากับ

$$w_{t+1,i} = w_{t,i} \beta_t^{1-\epsilon_t}$$

เมื่อ  $\epsilon_t = 0$  ถ้าตัวอย่าง  $x_i$  ถูกคัดเลือกอย่างถูกต้อง และ  $\epsilon_t = 1$  ในกรณีอื่นๆ และค่า  $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$

5. ผลลัพธ์สุดท้ายที่ได้คือ ตัวจำแนกแบบแข็งแกร่งมีค่าเท่ากับ

$$C(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

เมื่อ  $\alpha_t = \log \frac{1}{\beta_t}$

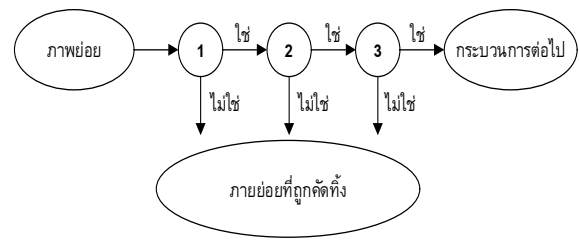
รูปที่ 5 แสดงตัวอย่างของการจำลองรูปแบบ Haar-like สองลำดับแรกที่ได้จากกระบวนการเรียนรู้ด้วยวิธี Adaboost จากรูปที่ 5 สังเกตเห็นว่าการจำลองรูปแบบ Haar-like แรกเป็นการจำลองรูปแบบ Haar-like ที่ได้จากความแตกต่างบริเวณดวงตาซึ่งจะมีค่าบริเวณส่วนบนของโหนกแก้ม สำหรับการจำลองรูปแบบ Haar-like ตัวที่สองได้จากความแตกต่างบริเวณดวงตาสองข้างซึ่งจะมีค่าบริเวณสันจมูก



รูปที่ 5 แสดงตัวอย่างของตัวจำแนกแบบแข็งแกร่งที่ได้จากกระบวนการเรียนรู้ด้วย Adaboost

### 3.3 ขั้นตอนการรวมตัวจำแนกกลุ่มแบบต่อเรียง

ขั้นตอนที่ 3 ที่ Viola-Jones นำเสนอเป็นเทคนิคที่ใช้ในการเพิ่มประสิทธิภาพในการตรวจจับให้มีความถูกต้อง และยังใช้เวลาในการคำนวณลดลงด้วย โดยแนวความคิดของขั้นตอนนี้ คือ การสร้างการรวมตัวจำแนกกลุ่มแบบต่อเรียง (Cascaded Classifier) โดยเทคนิคนี้จะนำตัวจำแนก (Classifier) หลาย ๆ ตัวต่อกันเป็นลำดับ ดังแสดงในรูปที่ 6 โดยที่ตัวจำแนกในลำดับต้น ๆ มักจะมีความซับซ้อนที่น้อยกว่า อาจสร้างมาจากตัวจำแนกแบบอ่อนแอกว่าไม่ก็ตัวเพื่อลดปริมาณการคำนวณและลดระดับของค่าแบ่งขีด (Threshold) ลงเพื่อเพิ่มลดการความผิดพลาดในการตรวจจับไม่พบใบหน้า ซึ่งภาพย่อยที่ถูกจัดประเภทเป็นไม่ใช่ใบหน้า (Non-face) จะถูกทิ้งไปเป็นจำนวนมากในลำดับต้น ๆ ในทางตรงข้าม ถ้าภาพย่อยนั้นถูกจำแนกว่ามีโอกาสเป็นภาพใบหน้า จะถูกส่งต่อไปยังตัวจำแนกตัวถัดไปซึ่งมีความซับซ้อนสูงขึ้นตามลำดับ และกล่าวได้ว่ายังมีจำนวนชั้นของตัวจำแนกมากเท่าใด โอกาสที่ภาพย่อยจะเป็นใบหน้าจะยิ่งมีมากขึ้น

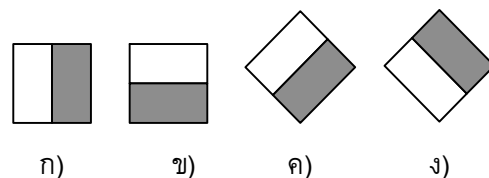


รูปที่ 6 แสดงขั้นตอนการรวมตัวจำแนกกลุ่มแบบต่อเรียง

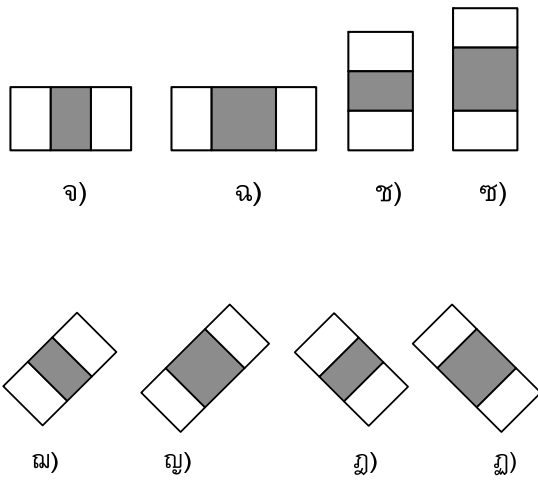
## 4. งานวิจัยที่ทำการศึกษและพัฒนาหลักเกณฑ์การจำลองรูปแบบ Haar-like

เทคนิคการตรวจจับใบหน้าของ Viola-Jones ที่ได้กล่าวไปนั้น สิ่งที่ส่งผลถึงประสิทธิภาพความถูกต้องในการตรวจจับใบหน้ามากที่สุด คือ การใช้การจำลองรูปแบบ Haar-like ดังนั้น ในบทความวิจัยนี้จึงมุ่งเน้นที่จะทำการศึกษางานวิจัยที่ทำการศึกษาต่อในเฉพาะส่วนของการจำลองรูปแบบ Haar-like

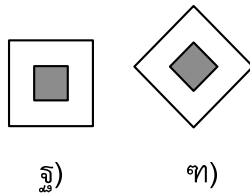
ในปี 2002 Lienhart และ Maydt [10] เสนอเทคนิคการตรวจจับใบหน้าต่อจากเทคนิคการตรวจจับใบหน้าของ Viola-Jones ที่เสนอในปี 2001 [7-8] โดย Lienhart และ Maydt นำเสนอการจำลองรูปแบบ Haar-like ใหม่ โดยนำการจำลองรูปแบบ Haar-like ที่ Viola-Jones นำเสนอในปี 2001 มาทำการหมุน 45 องศา และมีการนำเสนอการจำลองรูปแบบ Haar-like เพิ่มเติมดังแสดงในรูปที่ 7 รูปที่ 8 และ รูปที่ 9 จากผลการทดลองพบว่า การใช้การจำลองรูปแบบ Haar-like ที่นำเสนอสามารถเพิ่มประสิทธิภาพความถูกต้องในการตรวจจับภาพได้เพิ่มขึ้นเมื่อเทียบกับเทคนิคเดิม 10%



รูปที่ 7 การจำลองรูปแบบ Haar-like สำหรับตรวจจับขอบ (Edge features) ที่ถูกพิจารณาโดย Lienhart และ Maydt [10]

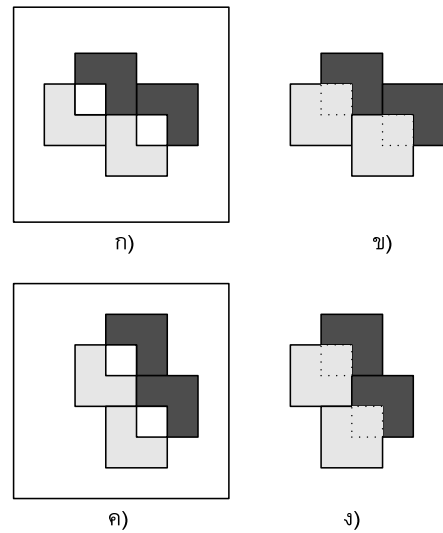


รูปที่ 8 การจำลองรูปแบบ Haar-like สำหรับตรวจจับเส้น (Line features) ที่ถูกพิจารณาโดย Lienhart และ Maydt [10]



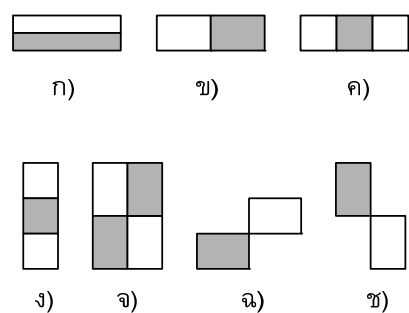
รูปที่ 9 การจำลองรูปแบบ Haar-like สำหรับตรวจจับจุดกึ่งกลาง (Center-surround features) ที่นำเสนอโดย Lienhart และ Maydt [10]

ในปี 2003 Viola และ Jones [11] ได้เสนอการจำลองรูปแบบ Haar-like แบบใหม่เพื่อแก้ไขข้อจำกัดของวิธีเดิมที่ได้นำเสนอในปี 2001 คือ ภาพใบหน้าที่จะนำมาตรวจจับจะต้องเป็นภาพที่ถ่ายใบหน้าด้านหน้า (ไม่ใช่ใบหน้าหันด้านข้าง) และมีลักษณะตั้งตรง (ไม่ใช่ใบหน้าเอียงซ้ายหรือขวา) โดยการจำลองรูปแบบ Haar-like ที่นำเสนอจะนำการจำลองรูปแบบ Haar-like เดิมมาต่อกันดังแสดงในรูปที่ 10 ซึ่งผลการทดลองแสดงให้เห็นว่าสามารถตรวจจับใบหน้าได้ทั้งใบหน้าด้านหน้าและด้านข้าง รวมทั้งยังสามารถตรวจจับใบหน้าที่ทำมุมเอียง (ไม่ตั้งตรง) ได้อีกด้วย

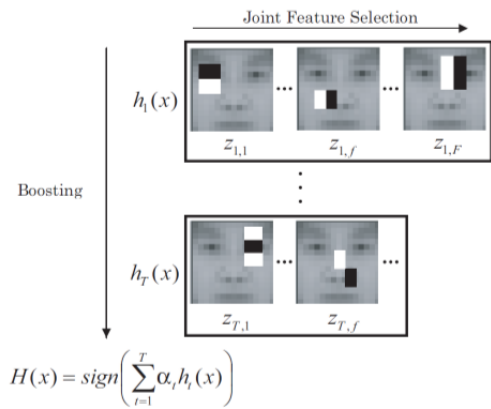


รูปที่ 10 ตัวอย่างการจำลองรูปแบบ Haar-like ที่นำเสนอโดย Viola และ Jones [11] เพื่อให้สามารถตรวจจับใบหน้าได้หลากหลายมากขึ้น

ในปี 2005 Mita และ Kaneko [12] ได้นำเสนอการจำลองรูปแบบ Haar-like ชนิดใหม่เรียกว่า การจำลองรูปแบบ Haar-like แบบร่วม (Joint Haar-like Feature) เพื่อปรับปรุงข้อจำกัดอีกประการของการจำลองรูปแบบ Haar-like วิธีเดิมๆ คือเรื่องความไม่ทนทานต่อสัญญาณรบกวนและการอ่อนไหวต่อการเปลี่ยนแปลงของแสง โดยการจำลองรูปแบบ Haar-like แบบร่วมที่นำเสนอขึ้นมาใหม่นี้สามารถแก้ปัญหาที่กล่าวมาได้ รูปที่ 11 และรูปที่ 12 แสดงตัวอย่างการจำลองรูปแบบ Haar-like ที่พิจารณาและแสดงภาพรวมของเทคนิคใหม่ที่นำเสนอตามลำดับ



รูปที่ 11 ตัวอย่างของการจำลองรูปแบบ Haar-like แบบร่วมที่นำเสนอโดย Mita และ Kaneko [12]



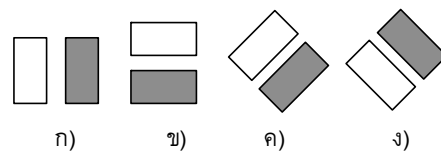
รูปที่ 12 ภาพรวมของเทคนิคการจำลองรูปแบบ Haar-like แบบร่วมที่นำเสนอโดย Mita และ Kaneko [12]

ในปี 2006 Wilson และ Fernandez [13] ทำการประยุกต์ใช้เทคนิค Haar classifier ในการตรวจจับรายละเอียดบนใบหน้าเช่น ตา คิ้ว จมูก ปาก เป็นต้น โดยการแบ่งแยกการตรวจจับใบหน้าแบบนี้ทำให้ความผิดพลาดในการตรวจจับลดลงและความเร็วที่ใช้ในการตรวจจับลดลงด้วยเนื่องจากพื้นที่ที่ต้องทำการตรวจสอบมีขนาดเล็กลง โดยในงานวิจัยนี้ใช้การจำลองของรูปแบบ Haar-like เหมือนกับที่นำเสนอโดย Lienhart และ Maydt [10] ดังแสดงในรูปที่ 7 ถึงรูปที่ 9 ต่อมาในปี 2007 Chen และ Liu [14] นำเสนอการปรับปรุงการตรวจจับใบหน้าด้วยการจำลองรูปแบบ Haar-like กับภาพสี ซึ่งแตกต่างจากเดิมที่จะทำการตรวจจับเฉพาะภาพที่เป็น Gray scaled โดยการจำลองรูปแบบ Haar-like ที่ใช้จะเป็นการจำลองรูปแบบ Haar-like ที่นำเสนอโดย Viola และ Jones ในปี 2001 [7-8]

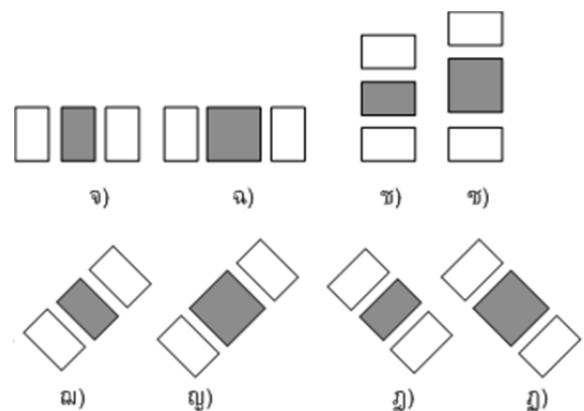
ในปี 2009 Khac และคณะ [15] ทำการนำเสนอการจำลองรูปแบบ Haar-like แบบใหม่เรียกว่า Variance based Haar-like ที่พัฒนาต่อจาก Viola-Jones ในปี 2001 โดยการจำลองรูปแบบ Haar-like แบบใหม่ที่น่าเสนอนี้จะทำการคำนวณหาค่าความแปรปรวนและค่าเฉลี่ยของพื้นที่สี่เหลี่ยมที่ได้จากวิธี Integral image และในงานวิจัยนี้ได้นำเสนอเทคนิค support vector machine (SVM) ใช้แทนวิธีการเรียนรู้แบบ Adaboost เพื่อให้เหมาะสมกับการจำลองรูปแบบ Haar-like แบบ

ใหม่ที่น่าเสนอ จากผลการทดลองที่ได้พบว่า การตรวจจับใบหน้าใช้เวลาในการตรวจจับน้อยลงและมีค่าความถูกต้องในการตรวจจับใบหน้าสูงกว่าเทคนิคเดิม 4.21%

ปี 2010 Lu และคณะ [16] นำเสนอการจำลองรูปแบบ Haar-like แบบใหม่ ที่เรียกว่า Separate Haar Feature แนวความคิดของการจำลองรูปแบบ Haar-like แบบใหม่คือ ไม่ต้องสนใจพื้นที่ระหว่างพื้นที่สี่เหลี่ยมของการจำลองรูปแบบ Haar-like รูปที่ 13 รูปที่ 14 และรูปที่ 15 แสดงการจำลองรูปแบบ Haar-like ใหม่ จากรูปจะพบว่าการจำลองรูปแบบ Haar-like ใหม่จะมีช่องว่างระหว่างพื้นที่ที่แรงเงากับพื้นที่ที่ไม่ได้แรงเงา ซึ่งผลการทดลองแสดงให้เห็นว่า การจำลองรูปแบบ Haar-like ที่นำเสนอใช้เวลาในการตรวจจับใบหน้าน้อยกว่าวิธีเดิม อีกทั้งยังมีความถูกต้องในการตรวจสอบใบหน้าด้วย

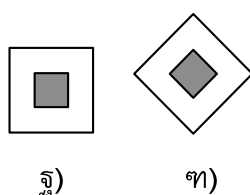


รูปที่ 13 การจำลองรูปแบบ Haar-like สำหรับตรวจจับขอบ (Edge features) ที่ถูกพิจารณาโดย Ning และคณะ [16]



รูปที่ 14 การจำลองรูปแบบ Haar-like สำหรับตรวจจับเส้น (Line features) ที่ถูกพิจารณา Ning และคณะ [16]





รูปที่ 15 การจำลองรูปแบบ Haar-like สำหรับตรวจจับจุดกึ่งกลาง (Center-surround features) ที่นำเสนอโดย โดย Ning และคณะ [16]

ในปี 2011 Pavani และคณะ [17] นำเสนอการจำลองรูปแบบ Haar-like ที่ถูกถ่วงน้ำหนักอย่างเหมาะสม โดยค่าถ่วงน้ำหนักที่ใส่เข้าในการจำลองรูปแบบ Haar-like ทุกแบบจะต้องมีค่าผลรวมเท่ากับศูนย์เสมอ ซึ่งจากผลการทดสอบพบว่า การจำลองรูปแบบ Haar-like ที่นำเสนอใช้เวลาในการตรวจจับใบหน้าน้อยกว่าการใช้การจำลองรูปแบบ Haar-like แบบเดิม [7-8] นอกจากนี้ ความถูกต้องในการตรวจสอบใบหน้าที่ยังมีมากกว่าอีกด้วย

ถึงแม้ว่าในหลายงานวิจัยได้อ้างว่าสามารถเพิ่มประสิทธิภาพ และความแม่นยำในการตรวจจับใบหน้าได้ แต่ในการเปรียบเทียบเชิงตัวเลขของแต่ละงานวิจัยนั้นไม่สามารถทำได้ เนื่องจากในหลายงานวิจัยใช้ฐานข้อมูลใบหน้าที่ต่างกันสรุปได้ดังตารางที่ 1

ตารางที่ 1 แสดงรายชื่อฐานข้อมูลใบหน้า และการอ้างอิง

ฐานข้อมูล	เอกสารอ้างอิง
MIT	[7, 8, 14]
CMU	[7, 8, 11, 14]
Yale	[12, 15]
XM2VTS	[12, 17]
FERET	[13]
UMIST	[16]
AR	[17]
Equinox	[17]
Bioid	[17]

## 5. สรุป

บทความวิชาการฉบับนี้กล่าวถึงการตรวจจับใบหน้าด้วยวิธีการพื้นฐานของการจำลองรูปแบบ Haar-like ที่ถูกนำเสนอโดย Viola-Jones ในปี 2001 การจำลองรูปแบบ Haar-like นั้นมีความสำคัญต่อความถูกต้องในการตรวจจับใบหน้ามากที่สุด เพราะเป็นเครื่องมือที่ใช้ในการดึงลักษณะเด่นจากใบหน้า

งานวิจัยจำนวนมากที่มุ่งเน้นในการปรับปรุงและพัฒนาการจำลองรูปแบบ Haar-like โดยงานวิจัยส่วนใหญ่จะนำเสนอการจำลองรูปแบบ Haar-like แบบใหม่โดยมุ่งเน้นที่จะลดเวลาในการคำนวณในขั้นตอนการหาตัวจำแนกแบบอ่อนแอซึ่งเป็นขั้นตอนสำคัญขั้นตอนหนึ่งในการตรวจจับใบหน้า นอกจากการปรับปรุงเรื่องเวลาแล้ว การจำลองรูปแบบ Haar-like ที่นำเสนอขึ้นมาใหม่ยังแสดงให้เห็นอีกว่า ความถูกต้องในการตรวจจับใบหน้าที่มีมากขึ้นเมื่อเปรียบเทียบกับกรณีที่ใช้การจำลองรูปแบบ Haar-like แบบเดิม

## 6. เอกสารอ้างอิง

- [1] M. H. Yang, D. J. Kriegman and N. Ahuja, "Detecting Faces in Images: A survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, Jan 2002.
- [2] G. Yang and T. S. Huang, "Human Face Detection in A Complex Back Ground," Pattern Recognition, vol. 27, no. 1, pp. 53-63, 1994.
- [3] C. Kotropoulos and I. Pitas, "Rule-Based Face Detection in Frontal Views," Proc. Int'l Conf. Acoustics, Speech and Signal Processing, vol. 4, pp. 2537-2540, 1997.
- [4] K. C. Yow and R. Cipolla, "A Probabilistic Framework for Perceptual Grouping of Features for Human Face Detection," Proc. Second Int'l Conf. Automatic Face and Gesture Recognition, pp. 16-21, 1996.

- [5] B. Scassellati, "Eye Finding via Face Detection for a Foveated, Active Vision System," Proc. 15<sup>th</sup> Nat'l Conf. Artificial Intelligence, 1998.
- [6] K. K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 39-51, Jan 1998.
- [7] P. Viola and M. Jones, "Rapid object detection using a Boosted cascade of simple features," Proc. Int'l Conf. Computer Vision and Pattern Recognition, pp. 1-9, 2001.
- [8] P. Viola and M. Jones, "Robust real-time face detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, 2004.
- [9] C. Zhang and Z. Zhang, "A Survey of Recent Advances in Face Detection," Technical Report TR-2010-66, MERL, June 2010.
- [10] R. Lienhart and J. Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection," Proc. Int'l Conf. Image Processing, pp. 900-903, 2002.
- [11] P. Viola and M. Jones, "Fast Multi-view Face Detection," Technical Report TR-2003-96, MERL, June 2003.
- [12] T. Mita, T. Kaneko and O. Hori, "Joint Haar-like Features for Face Detection," Proc. Int'l Conf. Computer Vision, pp. 1619-1626, 2005.
- [13] P. I. Wilson and J. Fernandez, "Facial Feature Detection using Haar Classifiers," Journal of Computing Sciences in Colleges, vol. 21, no. 4, pp. 128-133, April 2006.
- [14] D. S. Chen and Z. K. Liu, "Generalized Haar-like Features for Fast Face Detection," Proc. Int'l Conf. Machine Learning Cybernetics, pp. 2131-2135, 2007.
- [15] C. N. Khac, J. H. Park and H. Y. Jung, "Face Detection using Variance based Haar-like feature and SVM," World Academy of Science, Engineering and Technology 60, 2009.
- [16] N. J. Y. Lu, S. Tang and S. Goto, "Rapid Face Detection using a Multi-mode Cascade and Separate Haar Feature," Proc. Int'l Symp. Intelligent Signal Processing and Communication Systems, Dec. 2010.
- [17] S. K. Pavani, D. Delgado and A. F. Fragni, "Haar-like features with optimally weighted rectangles for rapid object detection," Pattern Recognition, vol. 43, pp. 160-172, 2010.