การทำนายอารมณ์ของมนุษย์ที่มีต่อรูปภาพนามธรรมโดยใช้คุณลักษณะของภาพและเครื่องมือติดตามการมองเห็น

Prediction of Human Emotions toward Abstract Images by Image Features and Eye Tracking Device

กิติ์สุชาต พสุภา[1], ภาณวี ฉัตรค้ำจุนเจริญ[2], โชติรส วุฒิเลิศเดชา[3]

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

1 ซอยฉลองกรุง 1 เขตลาดกระบัง แขวงลาดกระบัง กรุงเทพมหานคร 10520

E-mail: [1]kitsuchart@it.kmit.ac.th, [2]panawee.c@gmail.com, [3]chotiroswuttilertdeshar@gmail.com

ABSTRACT – Currently, emotion semantic search technology can support users to access data in the database. This can cover user's desirable which focuses on emotion concept. Given an image to different users, users' emotion stimulated by the image might be different due to different areas of interest. This paper presents a novel approach to increase the accuracy of emotion based image classification by combining eye movement data with basic image feature. The results show that combining eye movement data together with color feature can yield better classification performance than using color feature alone.

KEYWORDS -- Eye Movements; Implicit Feedback; Emotion; Image Retrieval

บทคัดย่อ -- ปัจจุบันการสืบค้นเชิงอารมณ์เป็นเทคโนโลยีที่จะสนับสนุนให้ผู้ใช้สามารถเข้าถึงข้อมูลในคลังข้อมูลโดยครอบคลุมความต้องการของผู้ใช้ได้มากขึ้น โดยเน้นทางด้านอารมณ์ ถึงแม้ว่ารูปภาพรูปเดียวกันก็ตาม ผู้ใช้ต่างกัน อารมณ์ของผู้ใช้ที่ถูกกระตุ้นโดยรูปภาพนั้นอาจจะแตกต่างกัน ขึ้นอยู่กับส่วนของภาพที่ผู้ใช้มองและสนใจ บทความนี้นำเสนอการเพิ่มประสิทธิภาพของการจำแนกประเภทรูปภาพเชิงอารมณ์ของผู้ใช้ โดยการใช้ประโยชน์จากข้อมูลการเคลื่อนไหวของตาที่เก็บจากผู้ใช้ขณะมองรูปภาพ กับคุณลักษณะพื้นฐานของรูปภาพ จากการทดลองพบว่าข้อมูลการเคลื่อนไหวของตาที่มีต่อรูปภาพนั้นสามารถเพิ่มประสิทธิภาพในการจำแนกประเภทอารมณ์ได้ดีกว่าการใช้เพียงคุณลักษณะสีพื้นฐาน

คำสำคัญ – การเคลื่อนไหวของตา; การป้อนกลับโดยปริยาย; อารมณ์; การค้นคืนรูป

1. บทนำ

การค้นคืนภาพ (Image Retrieval) คือ การค้นหารูปภาพที่ผู้ใช้ต้องการจากรูปภาพที่มีในฐานข้อมูล ในอดีตระบบการค้นคืนภาพจะเป็นการค้นคืนโดยใช้คำสำคัญ (Keyword) ที่ถูกแท็ก (Tag) กับรูปภาพแต่ละรูปในฐานข้อมูล ในการแท็กรูปภาพทั้งหมดที่มีอยู่ในฐานข้อมูล ต้องใช้แรงงานมาก จึงส่งผลให้มีการพัฒนาระบบเพื่อการแท็กรูปภาพอัตโนมัติ [1] หรือใช้การค้นคืนภาพแบบอิงเนื้อหา (Content Based Image Retrieval; CBIR) โดยการเปรียบเทียบภาพคำถาม (Query Image) กับรูปภาพทั้งหมดที่อยู่ในฐานข้อมูล ตัวอย่างของระบบ CBIR คือ PicSOM [2], Google Images, Yahoo! Images [3] เป็นต้น จากนั้นในปี 2006 ได้มีการนำเสนอระบบการค้นคืนภาพจากอารมณ์ของรูปภาพ (Emotion Sematic Image Retrieval; ESIR) [4] และเริ่มมีงานวิจัยที่ศึกษาเกี่ยวกับ ESIR มากขึ้น เช่น การศึกษาคุณลักษณะ (Feature) ในการจำแนกอารมณ์รูปภาพโดยใช้ทฤษฎีด้านจิตวิทยาและศิลปะในการหาอารมณ์รูปภาพ [5] การศึกษาอารมณ์ของมนุษย์ที่มีต่อรูปภาพจากคุณลักษณะพื้นฐานของรูปภาพนามธรรม (Abstract Art) [6] และการศึกษาคุณลักษณะของรูปภาพที่มีผลต่อการจำแนกอารมณ์โดยใช้

Multiple Kernel Learning (MKL) โดยพื้นฐานของคุณลักษณะที่ใช้คือ สี, รูปร่างและพื้นผิว [7]

ในระบบ CBIR จะต้องมีการป้อนกลับความเกี่ยวข้อง (Relevance Feedback) เพื่อประเมินตัวอย่างรูปภาพที่ค้นคืนโดยระบบโดยผู้ใช้ นั่นคือการคลิกเมาส์ (Mouse Click) ซึ่งการคลิกเมาส์นั้นจัดเป็นการป้อนกลับโดยชัดเจน (Explicit Feedback) ซึ่งเกิดจากความตั้งใจของผู้ใช้ ซึ่งต้องใช้แรงงานในการป้อนกลับ ในปี 2009 ได้มีการเริ่มพัฒนาระบบค้นคืนโดยเริ่มจากการเรียงลำดับของรูปภาพโดยใช้การเคลื่อนไหวของตา [8,9] งานวิจัยนี้ใช้ประโยชน์ของการป้อนกลับโดยปริยาย (Implicit Feedback) ซึ่งเกิดจากความไม่ตั้งใจในขณะที่กำลังเรียงรูปภาพ ในการเพิ่มประสิทธิภาพของระบบ จากนั้นจึงมีการพัฒนาระบบค้นคืนภาพที่ใช้ทั้งการป้อนกลับโดยชัดเจนและการป้อนกลับโดยปริยายที่เรียกว่าระบบ PinView [10]

จากสำนวน “Beauty is in the eye of the beholder” หรือ “ความสวยขึ้นอยู่กับคนมอง” จึงทำให้เกิดคำถามวิจัยที่ว่ารูปภาพหนึ่งรูป และมีคนมองหลายคน อารมณ์ที่ถูกกระตุ้นโดยรูปภาพของแต่ละผู้มองนั้นอาจไม่เหมือนกัน ขึ้นอยู่กับพื้นที่ที่ผู้ใช้กำลังมองรูปภาพนั้นอยู่ งานวิจัยนี้จึงได้นำเสนอการจำแนกอารมณ์ของรูปภาพโดยใช้คุณลักษณะของภาพและการเคลื่อนไหวของตา โดยมีสมมติฐานที่ว่า การใช้คุณลักษณะของภาพร่วมกับข้อมูลการเคลื่อนไหวของตานั้นสามารถเพิ่มประสิทธิภาพในการจำแนกอารมณ์ของผู้ใช้ที่มีต่อรูปภาพจากการใช้คุณลักษณะของภาพอย่างเดียวได้

2. ทฤษฎีที่เกี่ยวข้อง

2.1 อารมณ์ (Emotions)

อารมณ์คือความรู้สึกที่ได้รับการกระทบจากสิ่งเร้า ซึ่งการแบ่งอารมณ์พื้นฐานของมนุษย์สามารถอธิบายได้ 2 แนวทาง คือ

1) Discrete Approach – อารมณ์แต่ละอารมณ์นั้นแตกต่างกันอย่างสิ้นเชิง โดยแต่ละอารมณ์จะมีระดับความเข้มของตัวเอง ทฤษฎีต่าง ๆ ที่อธิบายอารมณ์โดยใช้แนวทางนี้ เช่น [11,12] ในงานวิจัยนี้จะแบ่งอารมณ์โดยใช้แนวทางของ Ekman [13] ซึ่งแบ่งอารมณ์ของมนุษย์ออกเป็น 6 อารมณ์ นั่นคือ โกรธ (Anger), ขยะแขยง (Disgust), กลัว (Fear), ดีใจ (Happiness), เสียใจ (Sadness) และประหลาดใจ (Surprise) อารมณ์ดังกล่าวได้ถูกจัดกลุ่มใหม่เป็น 4 อารมณ์ คือ ดีใจ, เสียใจ, โกรธและกลัว โดยแบ่งตามกล้ามเนื้อการแสดงสีหน้าของแต่ละอารมณ์ โดยพบว่ากล้ามเนื้อของการแสดงสีหน้าของมนุษย์จากอารมณ์ประหลาดใจมีลักษณะคล้ายกับกลัว เช่นเดียวกับอารมณ์ขยะแขยงและโกรธ [14]

2) Dimensional Approach – อารมณ์ต่าง ๆ เกิดขึ้นได้จากการผสมกันระหว่างอารมณ์พื้นฐาน ซึ่งอารมณ์พื้นฐานแต่ละอารมณ์จะมีระดับความเข้มของอารมณ์ด้วย ตัวอย่างเช่น Plutchik [15] ได้แบ่งอารมณ์พื้นฐานออกเป็น 8 อารมณ์ คือ สุข (Joy), เสียใจ (Sadness), โกรธ (Anger), กลัว (Fear), ไว้วางใจ (Trust), รังเกียจ (Disgust), ประหลาดใจ (Surprise), และคาดหวัง (Anticipation) ซึ่งอารมณ์พื้นฐานเหล่านี้สามารถผสมเป็นอารมณ์ใหม่ได้ เช่น สุข-ไว้วางใจ จะได้ อารมณ์รัก (Love)

2.2. รูปภาพนามธรรม (Abstract Art)

รูปภาพนามธรรม คือ รูปภาพที่แสดงถึงอารมณ์หรือความรู้สึกของผู้วาด วาซีลี คันดินสกี (Wassily Kandinsky) ผู้ให้กำเนิดศิลปะแนวนามธรรมกล่าวว่า สิ่งที่สำคัญสำหรับศิลปะนามธรรม คือ สีและรูปทรงที่ใช้ในการถ่ายทอดโดยคำนึงถึงความรู้สึกภายนอกและภายใน กล่าวคือ การถ่ายทอดรูปทรงต่าง ๆ ให้กลมกลืนด้วยสี ลักษณะผิว หรือ ความเด่นชัดของรูปภาพ เป็นต้น เพื่อให้เกิดอารมณ์และความรู้สึก [16] ลักษณะของรูปนามธรรมจะไม่สื่อความหมายโดยตรงเป็นรูป ผู้ดูต้องใช้ความรู้สึกและจินตนาการ

อารมณ์ที่เกิดจากการกระตุ้นจากรูปนามธรรมนั้นเกิดได้จากการตีความหมายของแต่ละผู้มอง ซึ่งขึ้นอยู่กับพื้นที่ของรูปภาพที่ผู้มองต้องการจะมองด้วย ดังนั้นรูปภาพรูปเดียวกัน ผู้มองคนละคน อาจมีอารมณ์ที่ต่างกัน ดังนั้นอุปกรณ์การติดตามการมองจึงถูกนำมาใช้เพื่อเก็บข้อมูลพฤติกรรมการมองของผู้ร่วมเก็บข้อมูล และนำมาวิเคราะห์

2.3. การป้อนกลับความเกี่ยวข้อง (Relevance Feedback)

เป็นการให้ข้อมูลป้อนกลับสู่ระบบค้นคืน เพื่อประเมินว่าข้อมูลที่ถูกค้นคืนจากระบบนั้นมีความเกี่ยวข้องกับคำสอบถาม (Query) หรือไม่ เพื่อนำไปปรับเปลี่ยนคำสอบใหม่ เพื่อเพิ่มประสิทธิภาพของการค้นคืนข้อมูลที่ต้องการในครั้งต่อไป

ยกตัวอย่างในกรณีของระบบการค้นคืนภาพ หลังจากระบบได้ค้นคืนและแสดงรูปภาพให้กับผู้ใช้นั้น ผู้ใช้สามารถให้ข้อมูลป้อนกลับกับระบบ โดยการคลิ๊กที่รูปภาพที่ผู้ใช้ต้องการหรือ

ใกล้เคียงความต้องการ เพื่อให้ระบบได้นำข้อมูลนี้ไปใช้ในการปรับปรุงการค้นคืนรูปภาพให้กับผู้ใช้ในรอบถัดไป การป้อนกลับสามารถแบ่งออกเป็น 2 ประเภท คือ

1) การป้อนกลับโดยชัดเจน เป็นการป้อนกลับของผู้ใช้ระบบที่เกิดขึ้นโดยความตั้งใจ ซึ่งจำเป็นต้องใช้เวลาและแรงงาน เช่น การคลิกเมาส์ การเคลื่อนไหวของตาโดยความตั้งใจ เสียง และ ท่าทาง

2) การป้อนกลับโดยปริยาย เป็นการป้อนกลับของผู้ใช้ระบบที่เกิดจากความไม่ตั้งใจ ซึ่งเกิดโดยอัตโนมัติขณะที่มนุษย์กำลังทำงานใด ๆ อยู่ การป้อนกลับชนิดนี้จำเป็นต้องถูกประมวลผลก่อนนำไปใช้ เช่น อัตราการเต้นของหัวใจ ความดันโลหิต อุณหภูมิของร่างกาย และ การเคลื่อนไหวของตาโดยความไม่ตั้งใจ (ขณะที่กำลังอ่านหนังสือ หรือ หาข้อมูลบนเว็บเบราเซอร์)

2.4. ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine; SVM)

เป็นอัลกอริทึมการเรียนรู้ของเครื่องชนิดหนึ่งที่ได้รับความนิยมอย่างแพร่หลายในการจำแนกประเภทข้อมูลและถูกนำมาใช้ในการศึกษาอารมณ์ของรูปภาพ [5, 6, 7, 17] SVM จะแบ่งแยกข้อมูลโดยการใช้ระนาบหรือไฮเปอร์เพลน (Hyperplane) แบ่งข้อมูลแบบเชิงเส้น

กำหนดให้ x_i คือเวกเตอร์ตัวอย่างที่ i ที่คู่กับค่าเป้าหมาย y_i ซึ่ง $y_i \in \{-1,1\}$, w คือเวกเตอร์ค่าน้ำหนัก, b คือค่าความโน้มเอียง (Bias), ξ_i คือค่าความผิดพลาดของการจำแนกประเภท, และ C คือตัวแปรเรกิวเลอไรเซชัน (Reguralisation) โดยระนาบนั้นจะถูกสร้างขึ้นโดยการจะพยายามทำให้ระยะห่างระหว่างสองกลุ่มนั้นมากที่สุด และ ความผิดพลาดในการแบ่งกลุ่มน้อยที่สุดดังสมการที่ (1)

$$\arg\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_i \xi_i \qquad (1)$$

ซึ่งอยู่ภายใต้ $y_i(w \cdot x_i + b) \geq 1 - \xi_i$ และ $\xi_i \geq 0$ โดยที่มีฟังก์ชันในการตัดสินใจ คือ

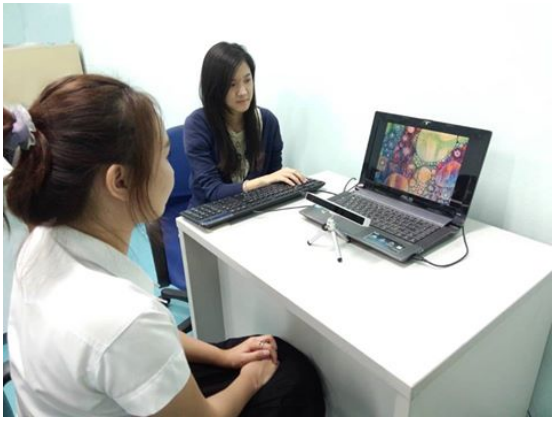
$$f(x) = \text{sign}(w^T x + b) \qquad (2)$$

ในโลกแห่งความเป็นจริงข้อมูลไม่ได้เป็นเชิงเส้นเสมอไป จึงมีการใช้เทคนิคที่เรียกว่า “เคอร์เนล” (Kernel) โดยใช้ฟังก์ชันเพื่อแปลงจากปริภูมิคุณลักษณะ (Feature Space) ปัจจุบันไปยังปริภูมิคุณลักษณะใหม่ที่สามารถแบ่งข้อมูลด้วยระนาบได้ เช่น Polynomial, Radial Basis Function เป็นต้น

3. การเก็บข้อมูล

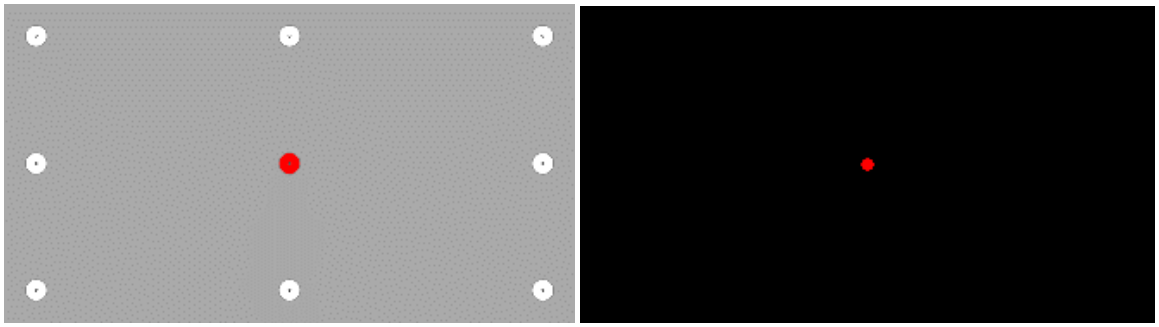
ในการทดลอง ได้มีการเก็บข้อมูลอารมณ์ของมนุษย์ที่มีต่อรูปภาพ และการเคลื่อนไหวของตา เพื่อใช้ในการวิเคราะห์ จากนักศึกษาระดับปริญญาตรี ซึ่งมีอายุ 18-22 ปี จำนวนทั้งหมด 10 คน แบ่งเป็น ชาย 5 คน หญิง 5 คน คนละ 100 รูป ผู้ร่วมการทดลองมีสายตาที่ปกติ และ ไม่ได้ใส่แว่นตา

ในการเก็บข้อมูล จะมีโปรแกรมที่พัฒนาขึ้นเพื่อแสดงรูปภาพจากฐานข้อมูล โดยทำงานร่วมกับอุปกรณ์ติดตามการมองเห็น โดยจะให้ผู้ร่วมเก็บข้อมูลนั่งประจำเครื่องคอมพิวเตอร์ขนาดพกพา 13 นิ้ว โดยการแสดงผลมีความละเอียดเท่ากับ 1366×768 พิกเซล และผู้ควบคุมการเก็บข้อมูลจะอธิบายขั้นตอนและวิธีการเก็บข้อมูลให้กับผู้เข้าร่วม โดยผู้เข้าร่วมเก็บข้อมูลจะบอกอารมณ์ที่เกิดขึ้นจากการมองแต่ละรูปภาพที่แสดงบนหน้าจอ และผู้ควบคุมการทดลองจะบันทึกข้อมูลที่ได้โดยใช้คีย์บอร์ด ดังรูปที่ 1.



รูปที่ 1. การเก็บการเคลื่อนไหวของตาโดยเครื่องติดตามการมอง และ อารมณ์ของผู้ร่วมการทดลองที่มีต่อรูปภาพ

ก่อนเริ่มการเก็บข้อมูล มีการแจ้งให้ผู้ร่วมการทดลองพยายามให้ศีรษะอยู่นิ่ง ๆ และจะต้องผ่านการปรับเทียบ (Calibration) ของอุปกรณ์การติดตามการมอง โดยใช้ 9 จุด (กลางหน้าจอ 1 จุด และ 8 จุดที่ขอบหรือมุมหน้าจอ) ดังรูปที่ 2 (ซ้าย) และก่อนที่จะให้ผู้ร่วมการทดลองดูรูปภาพใหม่ จะมีการแสดงรูปที่มีจุดอ้างอิง (Reference) ตรงกลางหน้าจอเป็นเวลา 3 วินาทีทุกครั้ง ดังรูปที่ 2 (ขวา) เพื่อให้การเคลื่อนเคลื่อนไหวของตาเริ่มต้นที่จุด ๆ เดียวกันทุกครั้ง ในการทดลองไม่มี



รูปที่ 2. ตัวอย่างหน้าจอ (ซ้าย) 9-point Calibration และ (ขวา) จุดอ้างอิง เพื่อให้ผู้ทดสอบเริ่มมองที่จุดเดียวกัน

รูปที่ 3. ตัวอย่างรูปนามธรรมที่ถูกเลือกจากงานวิจัย

กำหนดเวลาในการมองรูปภาพ รูปภาพจะเปลี่ยนก็ต่อเมื่อมีการบันทึกข้อมูลอารมณ์

อุปกรณ์การติดตามการมองที่ใช้คือ Eye Tribe Tracker สามารถเชื่อมต่อกับคอมพิวเตอร์หรือแท็บเล็ตด้วย USB3.0 โดยมีอัตราของการเก็บข้อมูลเท่ากับ 30 Hz และมีความแม่นยำที่ 0.5°-1° [18]

งานวิจัยชิ้นนี้ใช้รูปนามธรรมจาก [5] ซึ่งประกอบไปด้วยรูป 228 รูป แบ่งออกเป็น 8 อารมณ์ คือความสนุก (Amusement), ตื่นเต้น (Excitement), พอใจ(Contentment), เสียใจ (Sad), โกรธ (Anger), น่าเกรงขาม (Awe), กลัว (Fear), และขยะแขยง (Disgust) เราจึงจัดกลุ่มอารมณ์รูปภาพที่มีอารมณ์ใกล้เคียงกันให้เหลือเพียง 4 อารมณ์ เพื่อลดความซับซ้อนให้กับผู้ร่วมเก็บข้อมูลในการตัดสินใจ โดยอ้างอิงจาก [19] ซึ่งแบ่งอารมณ์อย่างมีโครงสร้างเป็น 3 ลำดับ คือ Primary Emotions, Secondary Emotions และ Tertiary Emotions โดยรวมรูปภาพที่มีอารมณ์คล้ายคลึงกัน ที่อยู่ในชั้น Primary Emotions เดียวกัน ให้เป็นอารมณ์พื้นฐาน 4 อารมณ์ เช่นเดียวกับ [14] คือ (1) ดีใจ – {สนุก, ตื่นเต้น, พอใจ}, (2) เสียใจ (3) โกรธ (4) กลัว – {กลัว, ขยะแขยง} เนื่องจากรูปที่มีอารมณ์น่าเกรงขามไม่ได้ถูกกล่าวไว้ใน [19] เราจึงไม่ได้สนใจรูปในกลุ่มนี้

จากนั้นจึงเลือกรูปภาพจำนวน 100 รูป โดยเลือกรูปที่มีอารมณ์ทั้ง 4 อารมณ์ อารมณ์ละ 25 รูป วิธีการเลือกรูปภาพนั้นเลือกจากคะแนนสูงสุดของแต่ละอารมณ์จากการโหวตของผู้ใช้ในงานวิจัยชิ้นนั้น ตัวอย่างของรูปภาพที่ถูกเลือกมาบางส่วนแสดงดังรูปที่ 3.

จากการเก็บข้อมูลนั้นพบว่ารูปเดียวกัน ดังรูปที่ 4(ก) คนมองต่างกัน อารมณ์ที่ได้ต่างกันเช่นกัน จากการมองรูปภาพของผู้เข้าร่วม-1 ให้ผลอารมณ์รูปภาพ คือ กลัว ดังรูปที่ 4(ข) ผู้เข้าร่วม-2 ให้ผลอารมณ์รูปภาพ คือ โกรธ ดังรูปที่ 4(ค) และผู้เข้าร่วม-3 ให้ผลอารมณ์รูปภาพ คือ เสียใจ ดังรูปที่ 4(ง)

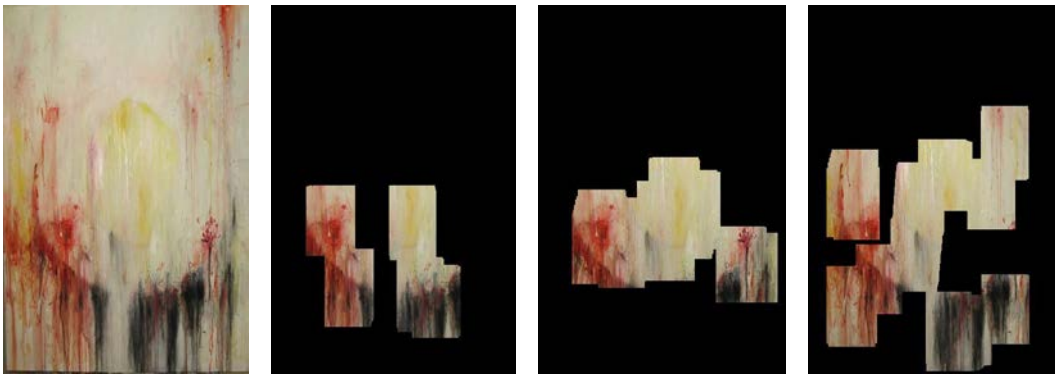
4. การสกัดคุณลักษณะ

ในงานวิจัยนี้เราสกัดคุณลักษณะสีของรูปภาพเป็นหลัก โดยใช้การแจกแจงความถี่ของสี (Colour Histogram) โดยการเปรียบเทียบคุณลักษณะ 3 ชนิด คือ การแจกแจงความถี่ของสีจาก (ก) รูปภาพต้นฉบับ (Original Image; F1) (ข) รูปภาพที่มีข้อมูลการมอง (Original Image+Eye Movements; F2) และ (ค) รูปภาพที่มีการมองโดยการเบลอด้วยฟังก์ชันเกาส์เซียน (Original Image+Eye Movements+Gaussian Blur; F3) ดังรูปที่ 5.

4.1 รูปต้นฉบับ

คุณลักษณะของแต่ละรูปถูกสกัดโดยใช้การแจกแจงความถี่ของสี RGB ที่มีอยู่ในรูปภาพ โดยทดลองใช้ขนาดของช่วงต่างกัน นั่นคือ {8, 16, 32, 64}-bin ในการทดลอง โดยที่ขนาดของคุณลักษณะที่ได้จะเท่ากับ

$$n_{feature} = n_{bin}{}^{3} \quad (3)$$



(ก) รูปต้นฉบับ (ข) กลัว (ค) โกรธ (ง) เสียใจ

รูปที่ 4. ตัวอย่างรูปภาพ และ ส่วนที่ถูกมองของผู้เข้าร่วม ที่มีอารมณ์แตกต่างกัน

(ก) F1 (ข) F2 (ค) F3

รูปที่ 5. ตัวอย่างรูปภาพ (ก) Original Image (F1) (ข) Original Image + Eye Movements (F2) และ (ค) Original Image + Eye Movements + Gaussian Blur (F3)

4.2 รูปภาพที่มีข้อมูลการมอง

ภาพในส่วนนี้จะขึ้นอยู่กับตำแหน่งการมองของแต่ละผู้ร่วมเก็บข้อมูล โดยใช้ฟังก์ชันความน่าจะเป็นการกระจายของตัวแปรสุ่มเกาส์เซียนที่มีขนาดต่าง ๆ คือ 100×100, 125×125 และ 150×150 กำหนดให้ $I(x,y)$ คือ รูปภาพ โดยที่มี (x,y) เป็นพิกัดของรูปภาพ และ $E(x,y)$ คือ ข้อมูลการมองของรูปภาพ จากนั้นจึงใช้ฟังก์ชันเกาส์เซียนกับข้อมูลการมองของรูปภาพ จาก

$$g(x_s) = e^{-\frac{x_s^2}{2\sigma^2}} \quad (4)$$

โดยที่ x_s คือขนาดของเกาส์เซียน ในที่นี้คือ 100, 125 และ 150 σ คือ ส่วนเบี่ยงเบนมาตรฐานของการกระจายเกาส์เซียน ซึ่งมีค่าเท่ากับ $\frac{x_s}{6}$ โดยนำเกาส์เซียนที่สร้างไปวางไว้ที่พิกัดตามข้อมูลการมองทุก ๆ จุดของ $E(x,y)$ จะได้ข้อมูลการมองของรูปภาพที่มีการกระจายตัวแบบเกาส์เซียน $G(x,y)$ ซึ่งสามารถนำมาใช้แสดงถึงการมองของมนุษย์ได้ [20] จากนั้นจึงนำไปแทนในฟังก์ชัน (5) เพื่อสร้างตัวกรองสำหรับตำแหน่งรูปภาพที่ถูกผู้ใช้มอง

$$H(x,y) = \begin{cases} 1 & if\ G(x,y) > 0 \\ 0 & if\ G(x,y) = 0 \end{cases} \quad (5)$$

ดังนั้นรูปที่ผ่านตัวกรองคือ

$$I_{F2}(x,y) = H(x,y) \times I(x,y) \quad (6)$$

โดยที่ $\times$ คือการคูณในระดับสมาชิกของเมตริกซ์

4.3 รูปภาพที่มีการมองโดยการเบลอด้วยฟังก์ชันเกาส์เซียน

ในส่วนนี้จะมีวิธีการสร้างเช่นเดียวกับรูปภาพในหัวข้อ 4.2 โดยใช้ $G(x,y)$ เป็นตัวกรองในการสร้างรูปใหม่

$$I_{F3}(x,y) = G(x,y) \times I(x,y) \quad (7)$$

$G(x,y)$ นั้นจะมีค่าตั้งแต่ 0 ถึง 1 หากค่าของตัวกรองตำแหน่งนั้นเข้าใกล้ศูนย์ ตำแหน่งภาพนั้นจะมืด แต่ถ้าค่าของตัวกรองตำแหน่งนั้นเข้าใกล้ 1 ตำแหน่งรูปภาพนั้นจะสว่าง

5. การทดลอง

จากนั้นเราได้เปรียบเทียบคุณลักษณะที่ได้ทั้ง 3 ชนิด โดยการทดลองเป็นโมเดลของแต่ละผู้ใช้ จำนวน10 คน โดยทดลองเปลี่ยนค่าจำนวนช่วงของการแจกแจงความถี่ คือ 8, 16, 32 และ 64 bin และ ขนาดของเกาส์เซียน คือ 100×100, 125×125 และ 150×150 จากนั้นข้อมูลได้ถูกเรียนรู้และทดสอบโดยใช้ SVM พร้อมกับ Linear Kernel โดยมีพารามิเตอร์ที่ต้องการปรับค่าเพียงตัวเดียว นั่นคือ C ของ SVM ตั้งแต่ 10^{-6}–10^{4} เพื่อให้ได้ผลการทำนายที่ดีที่สุด โดยใช้เทคนิค Leave-one-out Cross-validation (LOO-CV) เนื่องจากข้อมูลที่ใช้ทดสอบโมเดลผู้ใช้มีเพียงแค่ 100 รูปภาพ ซึ่งมีจำนวนน้อย

เนื่องจากข้อมูลประกอบไปด้วย 4 ประเภท คือ ดีใจ, เสียใจ, โกรธ และกลัว เราจึงเลือกการแบ่งกลุ่มข้อมูลโดยใช้ Multiclass Classification ชนิด 1-vs-all ข้อมูลที่ได้ถูกนำมานอร์มัลไลเซชัน (Normalization) ให้มีค่าเฉลี่ยเท่ากับศูนย์และค่าเบี่ยงเบนมาตรฐานเท่ากับหนึ่ง

6. ผลการทดลอง

ในงานวิจัยนี้ ค่าร้อยละความถูกต้องเฉลี่ยของผู้ใช้ทั้ง 10 คน ได้ถูกใช้ในการวัดและเปรียบเทียบประสิทธิภาพของการจำแนกอารมณ์ โดยพารามิเตอร์ C และ σ ได้ถูกเลือกโดยพิจารณาจากร้อยละความถูกต้องเฉลี่ย ซึ่งได้ถูกรายงานในตารางที่ 1 และทดสอบสมมติฐานการวิจัยโดยใช้ Paired t-test ดังตารางที่ 2

เนื่องจากอารมณ์ทั้ง 4 อารมณ์ของผู้ร่วมการทดลองนั้นไม่สมดุลกัน (Unbalanced Data) คุณลักษณะทั้ง 3 ประเภทที่นำเสนอในหัวข้อที่ 4 ได้ถูกเปรียบเทียบกับ Baseline (BL) นั่นคือ ในกรณีที่ SVM ทำการทำนายได้เพียงอารมณ์ส่วนมากเพียงอารมณ์เดียว

จากตารางที่ 1 พบว่า F1, F2 และ F3 นั้นให้ประสิทธิภาพที่ดีกว่า BL ในทุกกรณี (p<0.05 ยกเว้น F1 ในกรณี 64-bin ที่ p=0.110 ดังตารางที่ 2) ยิ่งไปกว่านั้น F2 และ F3 ยังดีกว่า F1 ในทุกกรณีเช่นเดียวกัน แต่มีเพียงกรณี 32-bin และ 64-bin ที่มีนัยสำคัญทางสถิติที่ระดับ p<0.05 และ F3 สามารถให้ประสิทธิภาพที่ดีกว่า F2 ในกรณีของ 64-bin ด้วย (p=0.129)

จากนั้นได้ทดลองเลือกโมเดลโดยปรับพารามิเตอร์ C, σ และ n_{bin} ของแต่ละผู้ใช้ โดยแสดงผลการทดลองเป็นรายบุคคลดังตารางที่ 3 และการทดสอบสมมติฐานการวิจัยโดยใช้ Paired t-test ซึ่งจะเห็นได้ว่า F1, F2, และ F3 นั้นให้ผลการทดลองที่ดีกว่า BL เช่นเดิม (p<0.01) และการใช้ประโยชน์ของการเคลื่อนไหวของตา (F2, F3) นั้นสามารถเพิ่มประสิทธิภาพจาก F1 ได้ (p<0.05) ยิ่งไปกว่านั้น F3 นั้นมีประสิทธิภาพที่ดีกว่า F2 อย่างมีนัยสำคัญทางสถิติที่ p=0.12 อย่างไรก็ตามใน

ตารางที่ 1. แสดงค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของผลการทำนายอารมณ์แต่ละ bin

n_{bin}	ความถูกต้องเฉลี่ย (%)			
	BL	F1	F2	F3
8	35.0	42.6±6.6	43.4±7.9	**45.1±5.3**
16		39.6±8.1	**42.3±7.4**	41.4±6.6
32		38.4±7.7	41.8±6.4	**43.1±6.4**
64		37.7±8.1	41.7±5.6	**43.9±8.0**

ตารางที่ 2. ค่า p-value จากการทำ Pair t-test ของประสิทธิภาพในการทำนายอารมณ์

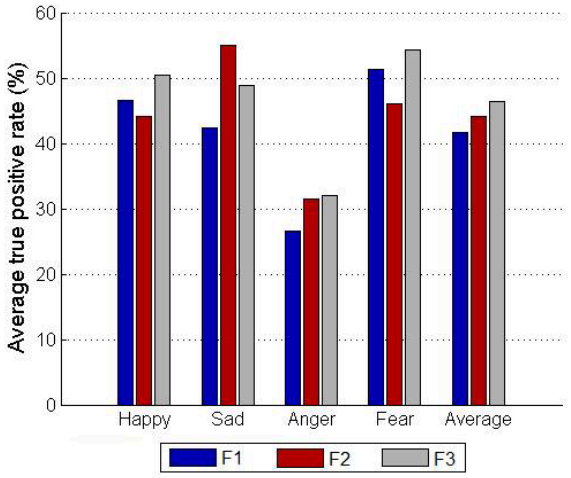
n_{bin}	p-value					
	F1-vs-BL	F2-vs-BL	F3-vs-BL	F1-vs-F2	F1-vs-F3	F2-vs-F3
8	0.004	0.001	0.001	0.663	0.203	0.497
16	0.050	0.007	0.001	0.060	0.293	0.573
32	0.048	0.011	0.001	0.033	0.006	0.297
64	0.110	0.001	0.002	0.012	0.011	0.129

ตารางที่ 3. ค่าความถูกต้องการทำนายอารมณ์ของแต่ละผู้ใช้

User	ความถูกต้องเฉลี่ย (%)			
	BL	F1	F2	F3
1	46.0	54.0	**58.0**	55.0
2	36.0	39.0	39.0	**47.0**
3	31.0	35.0	39.0	**47.0**
4	34.0	38.0	**44.0**	40.0
5	28.0	45.0	46.0	**47.0**
6	35.0	43.0	48.0	**49.0**
7	31.0	**48.0**	44.0	46.0
8	49.0	50.0	52.0	**54.0**
9	33.0	37.0	46.0	**52.0**
10	27.0	40.0	41.0	**42.0**
Avg.	35.0	42.9	45.7	**47.9**

ผู้ใช้ที่ 7 การเคลื่อนไหวของตานั้นอาจจะไม่มีประโยชน์เนื่องมาจากประสิทธิภาพของ F2 และ F3 นั้นน้อยกว่า F1

รูปที่ 6 แสดงค่าเฉลี่ยของ True Positive ของแต่ละอารมณ์และทุกอารมณ์ของทั้ง F1, F2 และ F3 จะเห็นได้ว่า F3 จะมีผลดีกว่า F1 ในทุกอารมณ์ ขณะที่ F2 นั้นดีกว่า F1 เพียงแค่อารมณ์เสียใจและโกรธเท่านั้น อย่างไรก็ตามในภาพรวมนั้น F3 นั้นให้ผลดีที่สุด ตามมาด้วย F2 และ F1 ตามลำดับ



รูปที่ 6. ค่าเฉลี่ย True Positive ของการทำนายอารมณ์ในแต่ละอารมณ์

7. สรุปผลการทดลอง

บทความนี้นำเสนอการจำแนกประเภทของรูปภาพเชิงอารมณ์ คือ ดีใจ, เสียใจ, โกรธ และกลัว โดยใช้ข้อมูล การเคลื่อนไหวของตาที่มีต่อรูปภาพ ร่วมกับคุณลักษณะสีพื้นฐาน จากการทดลองสรุปได้ว่า การใช้ข้อมูลการเคลื่อนไหวของตาที่มีต่อรูปภาพนั้น สามารถเพิ่มประสิทธิภาพของการจำแนกประเภทของภาพเชิงอารมณ์สำหรับโมเดลของแต่ละผู้ใช้ได้อย่างมีนัยสำคัญทางสถิติ อย่างไรก็ตาม งานวิจัยนี้ พิจารณาเพียงแค่ข้อมูลการเคลื่อนไหวของตาในการทดลอง ตัวแปรอื่น ๆ เช่น อายุ เพศ พื้นฐานการศึกษา อาชีพ ซึ่งไม่ได้พิจารณาก็อาจจะมีผลด้วยเช่นกัน

เนื่องจากในการทดลองได้ทดสอบเพียงแค่ Linear Kernel หากใช้ Non-linear Kernel อื่น ๆ ก็อาจจะช่วยเพิ่มประสิทธิภาพของการจำแนกประเภทก็เป็นไปได้

นอกจากการใช้คุณลักษณะสีพื้นฐานร่วมกับการเคลื่อนไหวของตาในการจำแนกอารมณ์ คุณลักษณะอื่น ๆ อาจจะมีผลต่อประสิทธิ์ภาพด้วย เช่น คุณลักษณะทางพื้นผิว (Texture) คุณลักษณะทางรูปร่าง (Shape) ในงานวิจัยนี้ รูปภาพที่ใช้ในการทดลองมีเพียงแค่รูปภาพนามธรรม หากใช้รูปถ่ายที่มีเนื้อหาเรื่องราวในการทดลองคุณลักษณะสีพื้นฐานอาจจะไม่มีผลต่อการจำแนกอารมณ์ก็เป็นไปได้

เอกสารอ้างอิง

[1] L. Ye, P. Ogunbona and J. Wang “Image Content Annotation Based on Visual Features”, In: Proceeding of International Symposium on Multimedia (ISM’2006), 11-13 Dec 2006, San Diego, USA, pp. 62–69, 2006.

[2] J. Laaksonen, M. Koskela, and E. Oja "PicSOM Self-organizing Image Retrieval with MPEG-7 Content Descriptors", IEEE Transactions on Neural Networks, 13(4), pp. 841–853, 2002.

[3] R. Datta, J. Li, and J. Z. Wang, “Content-based Image Retrieval – Approaches and Trends of the New Age”, In: Proceedings of ACM International Workshop on Multimedia Information Retrieval (MIR’2015), 10-11 Nov 2005, Singapore, pp. 253–262, 2005.

[4] W. Weining, Y. Yinlin, and J. Shengming, “Image Retrieval by Emotional Semantics: A Study of Emotional Space and Feature Extraction”, In: Proceeding of IEEE International Conference on Systems, Man and Cybernetics (SMC’2006), 8-11 Oct 2006, Taipei, Taiwan, pp. 3534–3539, 2006.

[5] J. Machajdik, and A. Hanbury, “Affective Image Classitication Using Features Inspired by Psychology and Art Theory,” In: Proceeding of ACM International Conference on Multimedia (MM’2010), 25-29 Oct 2010, Firenze, Italy, pp. 83–92, 2010.

[6] H. Zhang, E. Augilius, T. Honkela, J. Laaksonen, H. Gamper, and H. Alene, “Analyzing Emotional Semantics of Abstract Art Using Low-level Image Features”, In: Proceeding of International Symposium on Intelligent Data Analysis (IDA’2011), 29-31 Oct 2011, Porto, Portugal, pp. 413–423, 2011.

[7] H. Zhang, M. Gönen, Z. Yang and E. Oja, “Predicting Emotional States of Images Using Bayesian Multiple Kernel Learning”, In: Proceeding of International Conference on Neural Information Processing (ICONIP’2013), 3-7 Nov 2013, Daegu, Korea, pp. 274–-282, 2013.

[8] K. Pasupa, C. J. Saunders, S. Szedmak, A. Klami, S. Kaski, and S. R. Gunn, “Learning to Rank Images from Eye movements”, In: Proceeding of 2009 IEEE 12th International Conference on Computer Vision (ICCV'2009) Workshops on Human-Computer Interaction (HCI'2009), 27 Sep-4 Oct 2009, Kyoto, Japan, pp. 2009–2016, 2009.

[9] D. R. Hardoon and K. Pasupa “Image Ranking with Implicit Feedback from Eye Movements,” In: Proceedings of the 6th Biennial Symposium on Eye Tracking Research & Applications (ETRA'2010), 22-24 Mar 2010, Austin, USA, pp. 291–-298, 2010.

[10] P. Auer, Z. Hussain, S. Kaski, A. Klami, J. Kujala, J. Laaksonen, A. P. Leung, K. Pasupa, and J. Shawe-Taylor “Pinview: Implicit Feedback in Content-Based Image Retrieval, In: Proceeding of Workshop on Applications of Pattern Analysis (WAPA'2010), 1-2 Sep 2010, Cumberland Lodge, UK, pp 51–57, 2010.

[11] C. E. Izard “Basic Emotions, Relations Among Emotions, and Emotion-Cognition Relations”, Psychological Review, 99(3), pp. 561–565, 1992.

[12] K. Vytal and S. Hamann “Neuroimaging Support for Discrete Neural Correlates of Basic Emotions: A Voxel-based Meta-analysis,” Cognitive Neuroscience, 22(12), pp. 2864–2885, 2010.

[13] P. Ekman “Universals and Cultural Differences in Facial Expressions of Emotion,” Nebraska Symposium on Motivation, 19, pp. 207-282, 1972.

[14] R. E. Jack, O. G.B. Garrod, and P. G. Schyns “Dynamic Facial Expressions of Emotion Transmit an Evolving Hierarchy of Signals Over Time”, Current Biology, 24(2), pp. 187–192, 2014.

[15] R. Plutchik and H. Kellerman, “Emotion: Theory, Research and Experience,” Psychological Medicine, 11(1), pp. 207, 1980.

[16] D. W. Galenson “Two Paths to Abstract Art Kandinsky and Malevich”, Technical Report, National Bureau of Economic Research, No. 12403, 2006.

[17] Y. Wu, C. Bauckhage and C. Thurau “The Good, the Bad, and the Ugly: Predicting Aesthetic Image Labels”, In: Proceedings of 20th International Conference on Pattern Recognition (ICPR’2010). 23-26 Aug 2010, Istanbul, Turkey. pp. 1586–1589, 2010.

[18] The Eye Tribe Aps, The Eye Tribe, Available at https://theeyetribe.com.

[19] P. Shaver, J., Schwartz, D., Kirson, C., O'Connor “Emotional Knowledge: Further Exploration of a Prototype Approach,” In: Emotions in Social Psychology: Essential Readings, pp. 26-56, 2001.

[20] E. C. Chang, S., Mallat, C., Yap “Wavelet Foveation,” Appl. Comput. Harmon. Anal., 9, pp. 312-335, 2000