ใบรับรองวิทยานิพนธ์

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์

วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

ปริญญา

วิทยาการคอมพิวเตอร์	วิทยาการคอมพิวเตอร์
สาขา	**ภาควิชา**

เรื่อง การหาตำแหน่งเด่นโดยใช้ความเปรียบต่างของสี

Saliency Localization Based on Color Contrast

นามผู้วิจัย นายสุชาติ วิจารณ์ปรีชา

ได้พิจารณาเห็นชอบโดย

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก .....

(อาจารย์ผกาเกษ วัตุยา, Dr.rer.nat)

หัวหน้าสายวิชา .....

(ผู้ช่วยศาสตราจารย์ศิริกร จันทร์นวล, M.S.)

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์รับรองแล้ว

.....

(รองศาสตราจารย์กัญจนา ธีระกุล, D.Agr.)

คณบดีบัณฑิตวิทยาลัย

วันที่ ..... **เดือน** ..... **พ.ศ.** .....

ลิขสิทธิ์ มหาวิทยาลัยเกษตรศาสตร์

วิทยานิพนธ์

เรื่อง

การหาตำแหน่งเด่นโดยใช้ความเปรียบต่างของสี

Saliency Localization Based on Color Contrast

โดย

นายสุชาติ วิจารณ์ปรีชา

เสนอ

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์

เพื่อความสมบูรณ์แห่งปริญญาวิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

พ.ศ. 2557

สุชาติ วิจารณ์ปรีชา 2557: การหาตำแหน่งเด่นโดยใช้ความเปรียบต่างของสี ปริญญาวิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์) สาขาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก: อาจารย์ผกาเกษ วัตุยา, Dr.rer.nat 70 หน้า

โดยทั่วไป วัตถุประสงค์ของโมเดลสำหรับการตัวจับสิ่งเด่น กับโมเดลสำหรับทำนายอาการที่ตาของมนุษย์มองเข้าไปตรงตำแหน่งของจุดภาพใด จุดภาพหนึ่งของภาพนำเข้า (fixation prediction) นั้นมีส่วนที่เกี่ยวเนื่องกัน โมเดลสำหรับการตรวจจับวัตถุเด่นนั้นมีจุดมุ่งหมายเพื่อที่จะสำรวจว่าที่จะเป็นค่าความถูกต้องเชิงผลบวกมากเท่าที่จะเป็นไปได้ ขณะที่โมเดลแบบ fixation prediction พยายามที่จะสร้างค่าความผิดพลาดเชิงบวกที่น้อย

ในงานนี้ ผู้วิจัยพยายามที่จะทำการรวมจุดแข็งของทั้งสองโมเดลเข้าด้วยกัน ผู้วิจัยได้สร้างสิ่งนี้โดย ขั้นแรก ทำการแทนที่คุณลักษณะการรับรู้ระดับสูงที่มีการใช้เป็นประจำในโมเดลแบบ fixation prediction ด้วยวิธีการใหม่ของผู้วิจัยในการระบุตำแหน่งของวัตถุเด่น เพื่อทำให้โมเดลมีความเหมาะสมสำหรับภาพของวัตถุประเภทต่างๆ ขั้นที่สอง ผู้วิจัยได้ทำการเรียนรู้โดยทำการฝึกโมเดลสำหรับการตรวจจับสิ่งเด่นด้วยข้อมูลการติดตามตามมนุษย์ที่มองเข้าไปตรงตำแหน่งของจุดภาพใด จุดภาพหนึ่งของภาพนำเข้า เพื่อสร้างโมเดลที่ตรงตามลักษณะของจากระบุตำแหน่งของตามนุษย์ที่ดี ที่ปราศจากความสนใจแบบบนลงล่าง (top-down attention)

ผู้วิจัยได้เสนอผลการวัดประสิทธิภาพของโมเดลใหม่ในการระบุตำแหน่งบนชุดข้อมูลการตรวจจับสิ่งเด่นและ ชุดข้อมูลการทำนายตำแหน่งของจุดภาพที่ตามนุษย์มองเข้าไปในภาพ อีกทั้งทำการเปรียบเทียบกับโมเดลการตรวจจับสิ่งเด่นในปัจจุบัน ผลการทดลองแสดงให้เห็นว่าประสิทธิภาพของโมเดลที่นำเสนอมีประสิทธิภาพที่ดีกว่าวิธีการอื่นในงานประยุกต์ของการตรวจจับวัตถุเด่น สำหรับในงานประยุกต์ของการทำนายตำแหน่งของจุดภาพที่ตามนุษย์มองเข้าไปในภาพ ผลลัพธ์แสดงให้เห็นว่าโมเดลที่นำเสนอนั้นมีความใกล้เคียงในด้านความถูกต้องกับคุณลักษณะระดับสูงในโมเดลต้นฉบับ แต่ใช้เวลาในทำการประมวลผลที่น้อยกว่า

_____ _____ ___ / ___ / ___

ลายมือชื่อนิสิต ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

Suchat Vijanprecha 2014: Saliency Localization Based on Color Contrast. Master of Science (Computer Science), Major Field: Computer Science, Department of Computer Science. Thesis Advisor: Miss Pakaket Wattuya, Dr.rer.nat. 70 pages.

Generally, the purpose of saliency detection models for saliency object detection and for fixation prediction is complementary. Saliency detection models for saliency object detection aim to discover as much as possible true positive, while saliency detection models for fixation prediction intend to generate few false positive.

In this work, we attempt to combine their strength together. We accomplish this by, firstly, replacing high-level features that frequently used in a fixation prediction model with our new saliency location map in order to make the model more general. Secondly, we train a saliency detection model with human eye tracking data in order to make the model correspond well to the human eye fixation (without the use of top-down attention).

We evaluate the performance of our new saliency location map on both saliency detection and fixation prediction datasets in comparison with six state-of-the-art saliency detection models. The experimental results show that the performance of our proposed method is superior to other methods in an application of saliency object detection. For fixation prediction application, the results show that our saliency location map performs comparable to the high-level features, but requires much less computation time.

_____ _____ ___ / ___ / ___

Student's signature Thesis Advisor's signature

กิตติกรรมประกาศ

ขอขอบคุณ ดร. ผกาเกษ วัตุยา ประธานกรรมการที่ปรึกษา ที่กรุณาเป็นที่ปรึกษาในงานวิจัย พร้อมทั้งคำแนะนำ ข้อเสนอแนะและให้กำลังใจในการทำงานวิจัยจนสำเร็จลุล่วงครั้งนี้และขอบพระคุณ อ.สุนทรี คุ้มไพโรจน์ กรรมการวิชาเอก และ ดร.อุษา สัมมาพันธ์ที่ได้ให้คำแนะนำและข้อเสนอแนะดีๆ มากมายในงานวิจัย

ขอขอบพระคุณบิดา มารดา ญาติพี่น้อง ที่ให้กำลังใจ ทุนทรัพย์ และอื่นๆ ขอบคุณเพื่อนๆ ชาววิทยาการคอมพิวเตอร์ทุกๆคน ที่ร่วมด้วยช่วยกันเรียน และข้อคิดเห็นดีๆ ในการทำงานเพื่อให้สำเร็จลุล่วงไปด้วยดี และขอขอบคุณ คุณรัชนุ บุษบาบาล และคุณอารียา จันทรขันตี ในการช่วยเหลือต่างๆ

กราบขอบพระคุณ สาขาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตบางเขน ที่ให้การสนับสนุนทุนวิจัยและทุนการศึกษา

และที่ขาดไปไม่ได้ ขอขอบพระคุณ อาจารย์ผู้ที่ได้ทำการสอนสั่ง ประสิทธิ์ประสาทความรู้ทุกอย่าง ขอขอบคุณคณะวิทยาศาสตร์ ภาควิชาวิทยาการคอมพิวเตอร์ ที่ให้ห้องเรียนให้ข้าพเจ้าได้ค้นคว้าหาความรู้ ขอขอบคุณมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตบางเขน ที่เป็นแหล่งให้ฝึกฝนหาความรู้เพิ่มเติม

สุชาติ วิจารณ์ปรีชา

มิถุนายน 2557

สารบัญ

หน้า

สารบัญตาราง

สารบัญภาพ

สารบัญภาพ (ต่อ)

สารบัญภาพ (ต่อ)

สารบัญภาพ (ต่อ)

คำอธิบายสัญลักษณ์และคำย่อ

Bottom-up	=	วิธีการแบบล่างขึ้นบน
Center-Bias	=	การให้ความสำคัญตรงกลางภาพ
False positive	=	ค่าความผิดพลาดเชิงบวก
False positive rate	=	อัตราค่าความผิดพลาดเชิงบวก
Fixation	=	อาการที่ตาของมนุษย์มองเข้าไปตรงตำแหน่งของจุดภาพใด จุดภาพหนึ่งของภาพนำเข้า
High-level feature	=	คุณลักษณะการรับรู้ระดับสูง
Hue	=	ค่าโทนสี
Intensity	=	ค่าความเข้มแสง
Low-level feature	=	คุณลักษณะการรับรู้ระดับล่าง
Meta-level	=	ระดับชั้นที่อยู่เบื้องหลัง
Pixel	=	จุดภาพ
Precision rate	=	อัตราค่าความถูกต้อง
Recall rate	=	อัตราค่าความความครบถ้วน
Saliency detection	=	การตรวจจับสิ่งเด่น
Saliency Location	=	ตำแหน่งของวัตถุเด่น
Saliency map	=	ภาพที่เก็บเฉพาะพื้นที่ของวัตถุเด่น หรือ แมพ
Saturation	=	ค่าความอิ่มตัวของสี
Smoothen image	=	การปรับภาพเพื่อให้ผลลัพธ์ที่ได้มีความเรียบ
True positive rate	=	อัตราค่าความถูกต้องเชิงบวก
Top-down	=	วิธีการแบบบนลงล่าง

การหาตำแหน่งเด่นโดยใช้ความเปรียบต่างของสี

Saliency Localization Based on Color Contrast

คำนำ

ปัญหาสำหรับการตรวจจับสิ่งเด่นในภาพ คือ ต้องการระบุว่าวัตถุหรือพื้นที่ ณ ตำแหน่งใดในภาพ เป็นส่วนที่เด่นที่สุด ซึ่งวัตถุหรือพื้นที่เด่นนั้นจะมีความเกี่ยวข้องกับระบบความสนใจในการมองเห็นของมนุษย์ เช่น มนุษย์สามารถบอกได้ว่าพื้นที่ไหนมีความน่าสนใจ เมื่อมองเข้าไปในภาพ นอกจากนี้งานทางด้านการตรวจจับวัตถุเด่นในภาพยังมีประโยชน์ในโปรแกรมประยุกต์ทางด้านคอมพิวเตอร์วิทัศน์ ตัวอย่างเช่น โปรแกรมทางด้านการบีบอัดภาพหรือวิดีโอ (Xiaokang *et al*., 2005) การรู้จำภาพ (Rutishauser *et al.*, 2004) และ การตัดแบ่งภาพ (Ko and Nam, 2006) การตรวจจับสิ่งเด่นสามารถให้ข้อมูลที่เป็นประโยชน์เกี่ยวกับตำแหน่งของสิ่งเด่นและน่าดึงดูดที่สุดในภาพ ซึ่งกระบวนการนี้ได้เข้าไปช่วยแก้ปัญหางานทางด้านคอมพิวเตอร์วิทัศน์ โดยช่วยให้ปัญหาเหล่านั้นมีการคำนวณที่รวดเร็วและมีความถูกต้องมากขึ้น ยิ่งกว่านั้นยังมีการนำเสนอโมเดลในงานทางด้านการตรวจจับวัตถุเด่นออกมาอย่างหลากหลายในปัจจุบัน

โดยทั่วไปโมเดลการตรวจจับวัตถุเด่นในภาพสามารถแบ่งออกได้เป็น 2 กลุ่มตามโมเดลของวิธีการคำนวณในด้านประสาทวิทยา (Neroscience) คือ โมเดลแบบล่างขึ้นบน (Bottom-up) และ โมเดลแบบบนลงล่าง (Top-down) ในส่วนของโมเดลแบบล่างขึ้นบน ภาพผลลัพธ์ที่แสดงเฉพาะส่วนของสิ่งเด่น (saliency map) จะอาศัยการคำนวณบนพื้นฐานของการรวมคุณลักษณะระดับล่าง (low-level feature) ตัวอย่างเช่น สี ลายพื้นผิว หรือ มุม ในภาพ เพื่อตรวจจับ และตัดแบ่งส่วนที่เป็นวัตถุหรือพื้นที่ที่มีความเด่นที่สุดในภาพ ซึ่งเป็นจุดมุ่งหมายหลักของโมเดลนี้ ประโยชน์หลักของโมเดลล่างขึ้นบน คือ ใช้เวลาในการคำนวณน้อย และเป็นเหตุผลที่ทำให้โมเดลแบบล่างขึ้นบนมีการนำไปประยุกต์ใช้ในโปรแกรมประยุกต์ทางด้านงานคอมพิวเตอร์วิทัศน์อย่างกว้างขวาง (Achanta and Susstrunk, 2009; Ko and Nam, 2006; Rutishauser *et al.*, 2004; Xiaokang *et al.*, 2005)

ในตัวโมเดลแบบบนลงล่าง จะอยู่บนพื้นฐานคุณลักษณะการรับรู้ระดับสูง (high-level feature) คือ คุณลักษณะจะขึ้นอยู่กับลักษณะของงานที่ต้องการนำไปประยุกต์ใช้ เช่น การตรวจจับใบหน้า รถยนต์ บุคคล หรือ ข้อความในภาพ เป็นต้น ในงานวิจัยของ Erdem, E. และ A. Erdem กับ Judd, T. และคณะ (Erdem and Erdem, 2013; Judd *et al*, 2009) เสนอว่าสิ่งเหล่านี้มักเป็นส่วนที่มนุษย์ให้ความสนใจสูงเมื่อเห็นภาพต่างๆ แต่ข้อเสียเปรียบหลักของโมเดลแบบบนลงล่าง คือ มีการคำนวณที่ซับซ้อนกว่า

สำหรับโมเดลที่มีคุณลักษณะในการทำนายตำแหน่งของความสนใจที่ตามนุษย์มองเข้าไปในภาพ สามารถเรียกอีกอย่างหนึ่งว่า โมเดลสำหรับการทำนายแบบ fixation จะประกอบไปด้วยคุณลักษณะการรับรู้ระดับสูง โดยแต่ละคุณลักษณะก็จะขึ้นอยู่กับงานแต่ละชนิด ไม่สามารถที่จะตรวจจับสิ่งอื่นนอกเหนือจากวัตถุประสงค์ได้ ดังนั้นผู้วิจัยจึงเสนอที่จะใช้คุณลักษณะที่ไม่ได้เจาะจงกับงานในการตรวจจับ เช่น คุณลักษณะการรับรู้ระดับล่างแทนคุณลักษณะการรับรู้ระดับสูงที่มีความเจาะจงต่อวัตถุ

ในงานวิจัยนี้ ผู้วิจัยได้เสนอวิธีใหม่สำหรับการประมาณตำแหน่งของวัตถุเด่นในภาพบนพื้นฐานความเปรียบต่างของสี เนื่องจากเป็นคุณลักษณะที่เด่นและง่ายที่จะดึงดูดความสนใจของมนุษย์ และมีผลกระทบที่สำคัญในการตรวจจับพื้นที่วัตถุเด่นในภาพ อีกทั้งมีการนำไปประยุกต์ใช้กับหลายงานวิจัยในอดีต (Erdem and Erdem, 2013; Hui *et al.*, 2012; Itti *et al.*, 1998; Judd *et al.*, 2009)

สิ่งที่เป็นความแตกต่างระหว่างสิ่งที่นำเสนอกับโมเดลในงานวิจัยที่มีในปัจจุบัน คือ งานวิจัยนี้สามารถสร้างคุณลักษณะที่เน้นเฉพาะพื้นที่ของวัตถุเด่นเท่านั้นโดยที่ปราศจากส่วนที่เป็นพื้นหลังหรือสิ่งที่ไม่สนใจ ยิ่งกว่านั้นยังสามารถนำไปประยุกต์กับโมเดลสำหรับการทำนายแบบ fixation ในงานวิจัยนี้ได้ใช้กระบวนการเรียนรู้โดยมีการแยกชุดข้อมูลฝึก จากชุดข้อมูลการติดตามตำแหน่งของตามนุษย์ในภาพ (eye tracking data) โดยตรง ซึ่งในขบวนการฝึก ผู้วิจัยได้นำวิธีการในงานวิจัยนี้รวมกับคุณลักษณะระดับล่าง และระดับกลาง ที่ได้เสนอไว้กับโมเดลสำหรับการทำนายแบบ fixationของ Judd, T. และคณะ (Judd *et al.*, 2009) โดยที่ปราศจากส่วนของคุณลักษณะระดับสูง ที่เป็นคุณลักษณะพื้นฐานในโมเดล

วัตถุประสงค์

เพื่อศึกษาและปรับปรุงความสามารถของโมเดลการตรวจจับสิ่งเด่น (saliency detection) และโมเดลสำหรับการทำนายแบบ fixation โดยใช้วิธีการรวมทั้งสองโมเดลเข้าด้วยกัน เพื่อสร้างวิธีการที่สามารถทำงานได้ดีกับชุดข้อมูลของทั้งสองโมเดล ในด้านประสิทธิภาพความถูกต้อง และเวลาในการประมวลผล

ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถนำไปประยุกต์ในงานวิจัยที่ต้องการแยกพื้นหน้า และพื้นหลัง ออกจากกันในส่วนเริ่มต้นของกระบวนการทำงาน
2. สามารถนำไปประยุกต์ใช้เพื่อเพิ่มประสิทธิภาพด้านความถูกต้อง และลดเวลาในการคำนวณในงานวิจัย ทางด้านคอมพิวเตอร์วิทัศน์ ที่เกี่ยวกับการลดขนาดภาพ การบีบอัด และ การตัดเฉพาะส่วนที่สนใจ
3. สามารถนำวิธีการระบุเฉพาะตำแหน่งของวัตถุเด่นในภาพ เป็นกระบวนการที่ช่วยกำจัดสิ่งที่ไม่สนใจออกจากผลลัพธ์ในโมเดลการตรวจจับวัตถุเด่น
4. ช่วยลดการประมวลผลในโมเดลการทำนายแบบ fixation

ขอบเขต

1. ข้อมูลสำหรับทดสอบประสิทธิภาพ แบ่งออกเป็นสามชุด (1) ภาพปกติที่มีวัตถุเด่นในภาพแค่หนึ่งวัตถุเท่านั้น จากงานวิจัยของ Tie, L. และคณะ (Tie *et al.*, 2011) โดยใช้ภาพผลเฉลยที่มีการตัดตามขอบของวัตถุเด่นในภาพด้วยมนุษย์ จากงานวิจัยของ Achanta, R. และคณะ (Achanta *et al.*, 2009a) (2) ภาพในชุดข้อมูล SED2 (Alpert *et al.*, 2007) โดยแต่ภาพในชุดข้อมูลจะมีวัตถุเด่นสองวัตถุประกอบอยู่ในภาพ (3) ภาพในชุดข้อมูล MIT (Judd *et al.*, 2009) สำหรับโมเดลการทำนายแบบ fixation
2. การทดสอบความถูกต้องของโมเดล จะใช้กราฟค่าความถูกต้องกับค่าความครบถ้วน ในการเปรียบเทียบกับโมเดลตรวจจับวัตถุเด่นในปัจจุบัน และกราฟความสัมพันธ์ระหว่างอัตราความถูกต้องเชิงบวก กับอัตราความผิดพลาดเชิงบวก หรือกราฟ Receiver Operator Characteristic (ROC) ที่เปรียบเทียบระหว่างโมเดลที่เสนอกับ โมเดลของ Judd (Judd *et al.*, 2009)

การตรวจเอกสาร

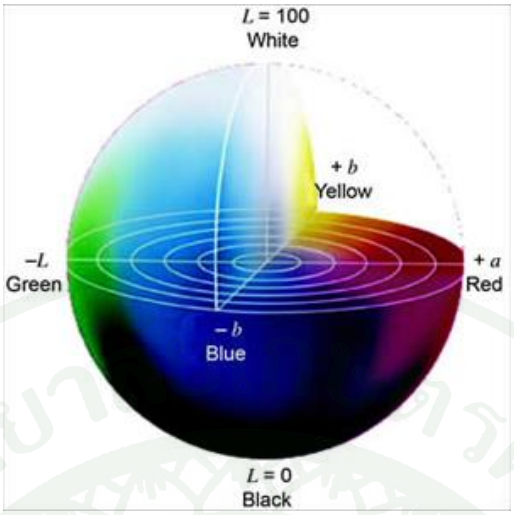
ทฤษฎีที่เกี่ยวข้องกับงานวิจัย

1. ระบบสี (Color System)

ในปี ค.ศ. 1666, Sir lsaac Newton พบว่าเมื่อลำแสงอาทิตย์ซึ่งเป็นแสงสีขาว (white light) ผ่านพรีซึม (prism) จะเป็น สเปคตรัม (spectrum) ที่ต่อเนื่องของสีตั้งแต่สีม่วงซึ่งอยู่ปลายด้านหนึ่ง ไปยังสีแดงที่อยู่ปลายอีกด้านหนึ่ง โดยสีที่อยู่ติดกัน จะมีลักษณะของสีที่ไม่ต่างกันมาก โดยแต่ละสีจะมีความถี่ของคลื่นแสงที่ต่างกัน ความยาวคลื่นของแสงที่ตามนุษย์มองเห็นได้จะอยู่ระหว่างคลื่นความถี่ 400 – 700 นาโนเมตร โดยมีเซลล์รูปกรวย (cones) เป็นตัวรับสีในตามนุษย์ จากคุณลักษณะในการดูดกลืนแสงของเซลล์รูปกรวยในตามนุษย์ ได้นำไปสู่การสร้างแบบจำลอง CIE (Commission Internationale d'Eclarage) Chromaticity ใน ค.ศ.1931 ซึ่งเป็นประโยชน์อย่างยิ่งต่อการพัฒนาวงการโทรทัศน์ แบบจำลองนี้แสดงให้เห็นถึงสีต่างๆ ที่มนุษย์สามารถมองเห็นได้

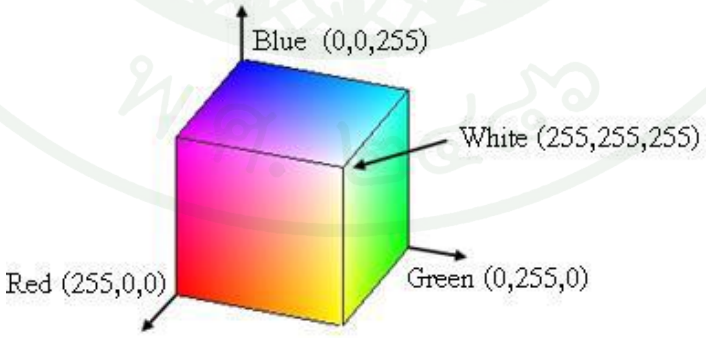
จากแบบจำลอง CIE ได้มีการกำหนด แม่สี (primary color) เพื่อให้เป็นมาตรฐานออกมา 3 สี คือ สีแดง ที่มีความยาวคลื่น 700 นาโนเมตร สีเขียว ที่มีความยาวคลื่น 546.1 นาโนเมตร และสีน้ำเงิน ที่มีความยาวคลื่น 435.8 นาโนเมตร จากการผสมของแม่สีทั้งสามยังสามารถทำให้เกิดสีผสม (secondary color) คือ สีแดงอมม่วง เกิดจาก การผสมกันระหว่าง สีแดงกับน้ำเงิน สีฟ้าอมเขียว เกิดจากการผสมกันระหว่าง สีเขียวกับน้ำเงิน และ สีเหลือง ที่เกิดจากการผสมกันระหว่างสีแดง กับสีเขียว ในระบบสีที่นิยมใช้ในการแสดงค่าสีเชิงตัวเลขที่สัมพันธ์กับการรับรู้สี คือ CIE LAB โดยค่า L* คือ ค่าความสว่าง ค่า a* แสดงความเป็นสีแดง – เขียว และค่า b* แสดงความเป็นสีเหลือง – น้ำเงิน ดังแสดงในภาพที่ 1 ซึ่งสีที่ต่างกันจะกระตุ้นให้เกิดการตอบสนองและจดจำสีได้ต่างกัน

การใช้สีในการประมวลผลภาพมีแรงจูงใจมาจากสองทฤษฎี คือ สีเป็นตัวบรรยายที่มีพลังที่ช่วยลดความยุ่งยากในการระบุวัตถุ และสามารถสกัดเหตุการณ์ที่เกิดในภาพ อย่างที่สอง มนุษย์สามารถมองเห็นสีได้หลายพันความเข้ม และเฉดสี เปรียบเทียบโดยประมาณ 24 บิตเฉดสีเทา และเป็นองค์ประกอบที่สำคัญที่จะใช้ในการวิเคราะห์ภาพ



ภาพที่ 1 โมเดลสี CIE LAB

การประมวลผลภาพสีสามารถแบ่งออกเป็นกลุ่มใหญ่ ได้เป็นสองกลุ่ม คือ การคำนวณสีจริงที่ได้มาจากภาพ RGB (แดง เขียว และ น้ำเงิน ตามลำดับ) หรือเรียกว่า full-color และกระบวนการกำหนดค่าสีให้กับค่าระดับสีโทนเทา โดยขึ้นอยู่กับเกณฑ์ที่กำหนด เรียกว่า pseudocolor ใน full-color อาจเรียกว่าเป็นระบบสี ซึ่งทำให้สะดวกในการระบุมาตรฐานของสีที่เป็นที่ยอมรับโดยทั่วไป เช่น จอภาพสี เครื่องสแกน และกล้องวิดีโอที่ใช้โมเดลสี RGB ส่วน pseudocolor นั้น ทำให้มนุษย์มีการมองเห็นภาพและตีความหมายของภาพที่ต้องการจะสื่อได้ดีขึ้น โดยตัวอย่างโคออร์ดิเนตคาร์ทีเซียนของสี่ RGB สามารถที่จะแสดงได้ดังภาพที่ 2

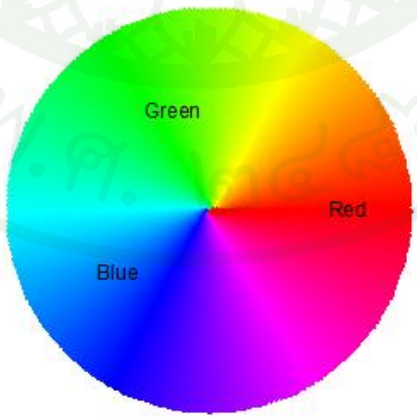


ภาพที่ 2 โคออร์ดิเนตคาร์ทีเซียนของสี RGB

ในทางปฏิบัติ จำนวนระดับความเข้มของสีในแต่ละแกนมีค่าจำกัด ขึ้นอยู่กับจำนวนบิตที่ใช้ เช่นหากใช้ 8 บิต (ให้ความเข้มสีที่แตกต่างกันได้ 256 ระดับ) เพื่อแทนองค์ประกอบของสีในแต่ละแกน การแสดงสีของจุดภาพหนึ่งจุด จะต้องใช้ทั้งหมด 24 บิต ซึ่งสามารถแสดงสีที่แตกต่างกันได้มากถึง 256×256×256 = 16,777,216 สี

แบบจำลองสี HSI (Hue-Saturation-Intensity) เป็นระบบพิกัดการแสดงสีอีกรูปแบบหนึ่งที่มีความสำคัญมากในทางปฏิบัติ มีหลักการและคุณลักษณะที่แตกต่างไปจากแบบจำลองสี RGB อย่างชัดเจน (วิทยากร และคณะ, 2555) โดยทฤษฎีแล้ว การพัฒนาแบบจำลองสี HSI มีวัตถุประสงค์หลักเพื่อให้ได้แบบจำลองสีที่เหมาะสมสอดคล้องกับธรรมชาติการมองเห็นของมนุษย์ ในขณะที่แบบจำลองสี RGB ได้รับการออกแบบเพื่อใช้ในการแสดงผลบนจอภาพคอมพิวเตอร์ โดยทั่วไปเราพบว่า ตามธรรมชาติมนุษย์สามารถมองเห็นรับรู้และเข้าใจสีได้ง่าย โดยพิจารณาจากค่าโทนสี (H) ค่าความอิ่มตัวของสี (S) และ ค่าความเข้มแสง (I) มากว่าการบรรยายสีในรูปของปริมาณสัดส่วนของสีแดง (R), เขียว (G), และน้ำเงิน (B)

ค่าโทนสี (hue) ในวงล้อสีจากภาพที่ 3 แต่ละโทนสามารถแสดงได้ในรูปของมุม 0 ถึง 360 องศา ยกตัวอย่างเช่น สีแดงกำหนดให้มีค่าเท่ากับ 0 องศา สีเขียวมีค่าเท่ากับ 120 องศา และสีน้ำเงินมีค่าเท่ากับ 240 องศา คำว่าโทนสี มีความหมายที่ต่างจากคำว่าสี เพราะว่าสีหมายรวมถึงทั้งค่าความเข้มแสง ค่าความอิ่มตัว และค่าโทนสีด้วย



ภาพที่ 3 ภาพวงล้อสี (color wheel)

ค่าความอิ่มตัวของสี (saturation) หมายถึงระดับความเข้มของค่าโทนสีที่มีอยู่ในสี โดยให้พิจารณาไล่ลำดับจากสีที่มีองค์ประกอบโทนสีบริสุทธิ์ทั้งหมด ไปยังสีที่มีองค์ประกอบของสีเทาเพิ่มมากขึ้นเรื่อยๆ จนในที่สุดกลายเป็นสีที่มีแต่องค์ประกอบของสีเทาทั้งหมด กำหนดให้ค่าความอิ่มตัวเป็น 0 ในกรณีที่เป็นสีเทาทั้งหมด คือ การไม่มีความอิ่มตัวของค่าสี และจะมีค่าสูงสุดเท่ากับ 1 เมื่อเป็นโทนสีบริสุทธิ์ ฉะนั้น ค่าความอิ่มตัวของสีจึงบ่งบอกถึงความบริสุทธิ์ของสี นั้นคือ ยิ่งมีค่าความอิ่มตัวมากก็หมายถึงมีความบริสุทธิ์มากหรือมีการปนของสีเทาในสัดส่วนที่น้อย แต่ถ้ามีการปนของสีเทาในสัดส่วนที่เพิ่มมากขึ้นก็จะมีค่าความอิ่มตัวน้อยลงตามลำดับ

ค่าความเข้มแสง (intensity) หมายถึง ปริมาณของความสว่างในสี นั่นคือ ถ้าสีที่พิจารณามีค่าความเข้มแสงมากจะมองดูสว่าง ในทางกลับกันสีที่มีค่าความเข้มแสงน้อยจะได้สีที่ดูมืด ค่าความเข้มแสงไม่เหมือนกับค่าความอิ่มตัวของสีเพราะปริมาณเนื้อโทนสีที่มีอยู่ในสีไม่ได้ลดลงหรือเพิ่มขึ้นแต่อย่างใดเพียงแต่ความสว่างของสีเพิ่มขึ้นเท่านั้น เราสามารถเปรียบเทียบค่าความเข้มแสงได้กับวงจร ปรับความสว่างของหลอดไฟในบ้าน กล่าวคือ แสงไฟมีความสว่างมากขึ้นโดยที่โทนสีของ หลอดไฟไม่ได้เปลี่ยนไปแต่อย่างใด หรือจะเทียบกับการปรับปุ่มความสว่างของจอมอนิเตอร์ก็ได้โดยจะเห็นว่าโทนสีของจอภาพไม่เปลี่ยนแปลงแต่มีความสว่างมากขึ้น

2. การตรวจจับวัตถุเด่นในภาพ (Saliency Detection)

การออกแบบตัวตรวจจับวัตถุเด่น หรือตัวตรวจจับจุดสนใจ ได้รับความสำคัญจากการศึกษาทางด้านคอมพิวเตอร์วิทัศน์ เป็นเวลาหลายทศวรรษ ตัวตรวจจับยังได้รับการปรับใช้อย่างกว้างขวางในโปรแกรมประยุกต์เช่น การติดตามวัตถุ (object tracking) การรู้จำ (recognition) และอีกมากมาย ในปัจจุบัน ตัวตรวจจับวัตถุเด่นมักจะอยู่ในขั้นตอนก่อนการประมวลผลของโปรแกรมประยุกต์เหล่านี้ ที่ช่วยประหยัดการคำนวณและเพิ่มความถูกต้อง อีกทั้งยังช่วยอำนวยความสะดวกในขั้นตอนการออกแบบที่ตามมา

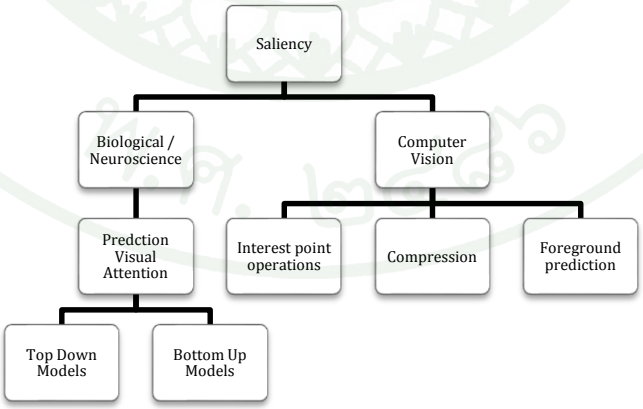
โมเดลการตรวจจับวัตถุเด่นในภาพสามารถแยกออกเป็นสามส่วนใหญ่ๆ ตามวิธีการคำนวณ คือ (1) วิธีการแบบล่างขึ้นบน ที่ใช้คุณสมบัติพื้นฐานของภาพ (low-level) เช่น ค่าความต่างสี ทิศทางและค่าความเข้มแสง เพื่อตรวจจับวัตถุเด่นในภาพ (2) วิธีการแบบบนลงล่าง ใช้คุณลักษณะระดับสูง เช่น ตัวตรวจจับใบหน้าหรือวัตถุที่เฉพาะเจาะจงกับโมเดลนั้นๆ (3) รวมวิธีการแบบล่างขึ้นบน กับ บนลงล่างเข้าด้วยกัน โดยสามารถสรุปเป็นความสัมพันธ์ของสิ่งเด่นบนภาพใน

งานทางด้านประสาทวิทยาศาสตร์กับคอมพิวเตอร์วิทัศน์ ดังภาพที่ 4 โดยส่วนของคอมพิวเตอร์วิทัศน์ ยังแสดงให้เห็นถึงภาพรวมว่าการตรวจจับวัตถุเด่นยังมีความเกี่ยวข้องในส่วนของตัวดำเนินการตรวจจับจุดที่สนใจ การทำนายส่วนของพื้นหน้าและการบีบอัดภาพ

ในการรวมคุณลักษณะต่างๆ เข้าด้วยกันสามารถแยกออกเป็นสองกลุ่มหลัก คือ

1. แบบเชิงเส้น ที่หาค่าน้ำหนักคุณลักษณะแต่ละอันแล้วนำมารวมกัน โดยผลรวมเชิงเส้นของคุณลักษณะ x และ y อาจเขียนเป็นนิพจน์ในรูป $ax + by$ โดย a และ b เป็นค่าน้ำหนักของคุณลักษณะแต่ละอันหลังจากการเรียนรู้ ยิ่งกว่านั้นวิธีการรวมคุณลักษณะแบบเชิงเส้นยังเป็นที่นิยมสำหรับโมเดลตรวจจับวัตถุเด่นในปัจจุบัน สำหรับตัวอย่างอัลกอริทึมการจำแนกประเภทแบบเชิงเส้นได้แก่ Perceptron, Least Squares Methods, Linear Regression และ Linear Support Vector Machines (SVMs)

2. แบบไม่เป็นเชิง (non-linear) ที่มีความเหมาะสมสำหรับการกระจายตัวของข้อมูล ที่อยู่ในรูปแบบเป็นกลุ่มและยากที่จะใช้การจำแนกประเภทแบบเชิงเส้นแบ่งข้อมูลออกจากกัน สำหรับวิธีรวมคุณลักษณะแบบไม่เป็นเชิงเส้น สามารถที่จะใช้อัลกอริทึม Nonlinear Regression, Adaboost และ Nonlinear Support Vector Machines เพื่อทำการคำนวณหาค่าน้ำหนักที่เหมาะสมในแต่ละคุณลักษณะหลังจากการเรียนรู้และทำการรวมเข้าด้วยกัน



ภาพที่ 4 ความสัมพันธ์ของการตรวจจับสิ่งเด่นบนภาพในงานทางด้านประสาทวิทยาศาสตร์ กับคอมพิวเตอร์วิทัศน์

สำหรับลักษณะของชุดข้อมูลและวิธีการวัดประสิทธิภาพของโมเดลการตรวจจับวัตถุเด่นในภาพที่ใช้ในปัจจุบัน สามารถแบ่งออกเป็นสองชุดหลักๆ คือ (1) ชุดข้อมูลของภาพที่เก็บเฉพาะพื้นที่ของวัตถุเด่น (saliency map) (2) ชุดข้อมูลที่ระบุตำแหน่งที่ตามนุษย์มองเข้าไปในภาพ (eye fixation)

วิธีการวัดประสิทธิภาพที่นิยมใช้ในงานวิจัยทั้งในอดีตและปัจจุบันจะอยู่บนพื้นฐานค่าความถูกต้อง (Precision) กับค่าความครบถ้วน (Recall) หรือจะเป็นกราฟความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) กับอัตราความผิดพลาดเชิงบวก (FPR) หรือเรียกอีกอย่างว่ากราฟ Receiver Operator Characteristic (ROC) ซึ่งสามารถแสดงสมการการคำนวณได้ดังนี้

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$TPR = \frac{\left(\frac{TP}{TP + FN}\right)}{N} \tag{3}$$

$$FPR = \frac{\left(\frac{FP}{FP + TN}\right)}{N} \tag{4}$$

เมื่อ *TP* คือ ค่าความถูกต้องเชิงบวก (true positive)
TN คือ ค่าความถูกต้องเชิงลบ (true negative)
FP คือ ค่าความผิดพลาดเชิงบวก (false positive)
FN คือ ค่าความผิดพลาดเชิงลบ (false negative)
N คือ จำนวนภาพทั้งหมดที่ใช้ในการทดสอบ

ความหมายสำหรับค่าความถูกต้องเชิงบวก คือ การที่ผลลัพธ์ที่ได้จากวิธีการทำนายว่าเป็นพื้นที่ของวัตถุเด่น แล้วเป็นพื้นที่ของวัตถุเด่นจริง ค่าความถูกต้องเชิงลบ หมายถึงการที่ผลลัพธ์ที่ได้จากวิธีการทำนายว่าเป็นพื้นที่ของวัตถุเด่น แล้วเป็นพื้นที่ของพื้นหลัง ค่าความผิดพลาดเชิงบวก หมายถึง การที่ผลลัพธ์ที่ได้จากวิธีการทำนายว่าเป็นพื้นที่ของพื้นหลัง แล้วเป็นพื้นที่ของพื้นหลังจริง และค่าความผิดพลาดเชิงลบ หมายถึง การที่ผลลัพธ์ที่ได้จากวิธีการทำนายว่าเป็นพื้นที่ของพื้นหลัง แล้วเป็นพื้นที่ของวัตถุเด่น

(a) (b) (c) (d)

ภาพที่ 5 ตัวอย่างภาพนำเข้าและผลเฉลยบนชุดข้อมูล (a) ภาพจากชุดข้อมูล MSRA (Tie *et al.*, 2011) (b) ผลเฉลยพื้นที่ของวัตถุเด่น (Achanta *et al.*, 2009a) (c-d) ภาพนำเข้าและผลเฉลยแบบ eye fixation (Judd *et al.*, 2009)

3. ตัวกรองภาพ (Image Filter)

การกรองข้อมูลภาพ คือ การนำภาพไปผ่านตัวกรองสัญญาณเพื่อให้ได้ภาพผลลัพธ์ออกมา โดยภาพผลลัพธ์ที่ได้จะมีคุณสมบัติแตกต่างจากภาพเริ่มต้น วัตถุประสงค์หลักของการกรองข้อมูลภาพ คือ การเน้น (Enhance) หรือลดทอน (Attenuate) คุณสมบัติบางประการของภาพเพื่อให้ได้ภาพที่มีคุณสมบัติตามต้องการและเหมาะกับการนำไปใช้งานต่อไป

การกรองข้อมูลภาพเป็นขั้นตอนที่ต่อจากขั้นตอนการได้มาของภาพผลลัพธ์การตรวจจับ ซึ่งเป็นขั้นตอนที่จำเป็นมาก เนื่องจากในการทำงานจริง ภาพที่ได้ส่วนใหญ่จะมีสัญญาณรบกวน (Noise) หรือสัญญาณไม่พึงประสงค์อื่นๆ รวมอยู่ในภาพด้วย การกรองข้อมูลภาพสามารถปรับปรุงให้ภาพมีคุณสมบัติที่ดีขึ้นเหมาะแก่การใช้งานหรือการนำไปประมวลผลในขั้นตอนต่อๆ ไป ซึ่งประเภทของตัวกรองที่นิยมใช้ มีอยู่ 2 ประเภท คือ ตัวกรองความถี่ต่ำ (Low Pass Filter) ซึ่งเป็นการกรองที่ใช้ในการปรับปรุงภาพเพื่อให้มีความเรียบ หรือเพื่อกำจัดสิ่งรบกวนออกจากภาพ และตัว

$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$
$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$
$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$
$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$
$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$	$\frac{1}{25}$

ภาพที่ 6 เคอร์เนลขนาด 3×3 และ 5×5 ที่ใช้กับตัวกรองค่าเฉลี่ย

กรองความถี่สูง (High Pass Filter) ซึ่งเป็นการกรองที่ใช้ในการปรับปรุงภาพเพื่อความคมของขอบวัตถุในภาพ ซึ่งทั้งสองวิธีขึ้นอยู่กับการกำหนด คอนโวลูชันเคอร์เนล (Convolution Kernel)

การกรองข้อมูลโดยการทำคอนโวลูชัน (Convolution) คือ การหาผลรวมค่าที่ได้จากการถ่วงน้ำหนักของค่าจุดภาพบริเวณรอบๆ ดังสมการที่ 5

$$h[x,y] = f[x,y] * g[x,y] \qquad (5)$$

เมื่อ $h[x,y]$ คือ ภาพผลลัพธ์หลังการทำคอนโวลูชัน

$f[x,y]$ คือ ภาพนำเข้า

$g[x,y]$ คือ เมตริกซ์ขนาด $n \times n$

โดยทำการคำนวณหาค่าเพื่อนำมาแทนค่าจุดภาพเดิม ซึ่งค่าที่ใช้ในการถ่วงน้ำหนักเป็นเมตริกซ์ขนาด $n \times n$ โดย n จะนิยมใช้เป็นเลขคี่ เช่น 3, 5, 7, 9 เป็นต้น เพราะว่าสามารถหาค่าจุดภาพตรงกลางได้ การกำหนดขนาดเคอร์เนล (Kernel) ถ้ามีขนาดใหญ่มาก จะส่งผลให้โปรแกรมทำงานช้าลง เนื่องจากมีการคำนวณที่มากและก็จะยิ่งมีผลทำให้ความเรียบและความพร่ามัวของภาพผลลัพธ์มีมากขึ้น

4. ค่าแปรปรวน (Variance)

ค่าแปรปรวน (มัลลิกา และคณะ, 2540) เป็นค่าที่ใช้วัดการกระจายที่นิยมใช้มากที่สุด โดยที่ค่าแปรปรวนพิจารณาจากผลรวมของค่าแตกต่างระหว่างค่าของข้อมูลแต่ละค่ากับค่าเฉลี่ยเลขคณิต ถ้าค่าแปรปรวนสูงมากแสดงว่าข้อมูลมีการกระจายมาก ค่าแตกต่างระหว่างค่าของข้อมูลกับค่าเฉลี่ยอาจจะเป็นค่าบวกหรือค่าลบขึ้นกับค่าเฉลี่ยจะมากกว่าหรือน้อยกว่าค่าของข้อมูล ผลรวมของค่าแตกต่างที่เป็นบวกและลบอาจจะกลายเป็นศูนย์ทำให้ผู้วิเคราะห์คาดว่าข้อมูลไม่มีการกระจาย จึงต้องกำหนดให้ค่าแปรปรวนเป็นผลรวมของค่าแตกต่างระหว่างข้อมูลแต่ละค่ากับค่าเฉลี่ยเลขคณิตยกกำลังสองแล้วหารด้วย N ซึ่งเป็นจำนวนข้อมูลทั้งหมดซึ่งมีสูตรดังต่อไปนี้

$$var = \frac{\sum (X_i - \mu)^2}{N} \qquad (6)$$

เมื่อ var คือ ค่าแปรปรวน

X_i คือ ค่าของข้อมูลแต่ละค่า

μ คือ ค่าเฉลี่ยเลขคณิต

N คือ จำนวนข้อมูลทั้งหมด

เนื่องจากค่าแปรปรวน คือ $\frac{\sum f_i (X_i - \mu)^2}{N}$ จึงทำให้หน่วยของค่าแปรปรวนยกกำลังสองไปด้วย เพราะทั้งค่าของข้อมูล (X) และค่าเฉลี่ยเลขคณิต ต่างก็มีหน่วยเดียวกัน เช่น ถ้าค่าของข้อมูลมีหน่วยเป็นบาท ค่าแปรปรวนจะมีหน่วยเป็น (บาท)^2 หรือถ้าค่าของข้อมูลมีหน่วยเป็นคน ค่าแปรปรวนจะมีหน่วยเป็น (คน)^2

งานวิจัยที่เกี่ยวข้อง

ในส่วนนี้จะนำเสนองานวิจัยที่เกี่ยวข้องในการทำวิจัย เพื่อแสดงให้เห็นถึงความสำคัญของคุณลักษณะต่างๆ ของภาพ ที่งานวิจัยทั้งในอดีตและปัจจุบันเลือกใช้สำหรับโมเดลที่ได้นำเสนอ หรือ วิธีการรวมของคุณลักษณะแต่ละชนิดเข้าด้วยกัน เพื่อให้ได้ตำแหน่งของสิ่งเด่น อีกทั้งยังได้อธิบายถึงชุดข้อมูลมาตรฐานต่างๆ ที่งานวิจัยในปัจจุบันใช้เพื่อประเมินประสิทธิภาพของโมเดล ซึ่งมีรายละเอียดดังนี้

งานวิจัยของ Itti *et al.,* (1998) เสนอโมเดลของระบบความสนใจในการมองเห็นของมนุษย์ โดยยึดหลักความแตกต่างระหว่างกรอบที่พิจารณากับพื้นที่รอบกรอบนั้นๆ ที่มีขนาดภาพแตกต่างกัน คุณลักษณะสำคัญได้มาจากในการพิจารณาที่ออกแบ่งออกเป็น 3 ชนิด คือ สี มุมและความเข้มแสง

งานวิจัยของ Ma and Zhang (2003) เสนอทางเลือกในการวิเคราะห์ทีละตำแหน่ง สำหรับสร้างผลลัพธ์ในการตรวจจับโดยใช้โมเดลการเติบโตแบบคลุมเครือ (fuzzy growth) สำหรับระบุพื้นที่ที่สนใจ (ROIs) Ma, Y.-F. และ H.-J. Zhang ได้ใช้วิธีการคำนวณที่ง่ายและชัดเจนในการสร้างภาพความละเอียดต่ำ ด้วยการพิจารณาคุณลักษณะสีแบบพื้นที่ โดยใช้ตัวกรองและมีการหาค่าเฉลี่ยเพื่อแทนที่เข้าไปทุกจุดในตัวกรอง

งานวิจัยของ Harel *et al*. (2006) ได้เสนอวิธีแสดงภาพเป็นกราฟที่เชื่อมต่อกันแบบเต็ม (fully-connected graph) น้ำหนักของแต่ละโหนดจะถูกกำหนดเป็นคุณลักษณะเชิงพื้นที่ระหว่างสองจุดภาพ ดังนั้น ถ้าจุดภาพสองจุดมีความแตกต่างกันมาก ก็จะทำให้มีค่าน้ำหนักมาก สำหรับกราฟนี้มีการกระจายสมดุลที่สะท้อนถึงเวลาที่สุ่มเดินที่ใช้ไปในการเน้นโหนดที่มีความแตกต่างจากโหนดอื่น ซึ่งการตรวจสอบนี้ได้สะท้อนวัตถุเด่นออกมาจากภาพ คุณลักษณะที่ใช้ประกอบไปด้วย (1) คุณสมบัติทางด้านมุม ที่ใช้ คือ ตัวกรองเกเบอร์ (Gabor filter) ใน 4 องศาที่ต่างกัน (2) ความเปรียบต่างของแสง (luminance contrast) ในพื้นที่ใกล้เคียงขนาด 80×80 พิกเซล ที่วัดสองขนาดของเชิงพื้นที่ ทั้งสิบสองภาพที่เก็บเฉพาะพื้นที่ของวัตถุเด่น ซึ่งเรียกว่า แมพ (map) นี้จะถูกลดขนาดไปเป็น 28×37 ในแต่ละแมพ โดย Harel, J. และคณะได้ทำการวัดประสิทธิภาพของโมเดลบนชุดข้อมูลของ Doves (Van Der Linde *et al.*, 2009) ที่เก็บภาพธรรมชาติในรูปแบบภาพโทนเทา

งานวิจัยของ Achanta, R. และคณะ (Achanta *et al.*, 2009a) ใช้ประโยชน์จากคุณสมบัติของสเปกตรัมโดเมนของภาพสำหรับการตรวจจับสิ่งเด่น ด้วยวิธีการปรับจูนความถี่ (frequency-tuned) แต่ในการดำเนินการแปลงแบบฟูรีเย (Fourier transform) Achanta, R. และคณะได้เสนอให้ใช้ตัวกรองแบบลำดับชั้น (stack of filters) เพื่อจับคุณสมบัติสเปกตรัมของภาพ ช่วงสีแต่ละช่วงจะผ่านตัวกรองของความแตกต่างของเกาส์เซียน (DoG) ที่เป็นตัวกรองประเภทแถบความถี่ (Bandpass filter) ที่คล้ายกับวิธีการของสเปกตรัมทั่วไป Achanta, R. และคณะตรวจจับวัตถุเด่นโดยลบข้อมูลเฉลี่ยบางส่วนจากความถี่ที่ได้จากภาพและหาค่าเฉลี่ยของจุดภาพเพื่อใช้เป็นค่าเฉลี่ยของข้อมูล แต่วิธีการของ Achanta, R. และคณะไม่เหมือนกับวิธีการของสเปกตรัมทั่วไปตรงที่ ได้รวมข้อมูลสีเข้าไปในวิธีการ สำหรับภาพนำเข้าจะถูกเปลี่ยนเป็นช่วงสีของ L*a*b* ซึ่งแมพสามารถหาได้จากแต่ละช่วงสี และรวมแมพทั้งสามด้วยวิธีการคำนวณระยะทางแบบยุคลิด สำหรับพารามิเตอร์ของตัวกรองจะถูกทำการเลือกโดยการปรับด้วยมือ เพื่อให้ได้แมพเฉพาะพื้นที่ของวัตถุเด่นที่ปราศจากพื้นหลัง Achanta, R. และคณะได้ใช้วิธีการย้ายตามค่าเฉลี่ยแบบแบ่งส่วน (mean-shift segmentation) ในช่วงสี L*a*b*

สำหรับการประเมินผล ผู้เขียนเห็นว่าผลเฉลยในงานของ Tie, L.และคณะ (Tie *et al.*, 2011) ที่เป็นกรอบสี่เหลี่ยมล้อมรอบวัตถุเด่นนั้นมีความห่างไกลความถูกต้องมาก จึงได้เสนอผลเฉลยแบบใหม่ เป็นภาพแบบทวิภาคที่ทำการตัดตามขอบของวัตถุเด่นโดยใช้มนุษย์

งานวิจัยของ Judd *et al.* (2009) ได้เสนอการตรวจจับวัตถุเด่นโดยการรวมชุดของคุณลักษณะในวงกว้าง Judd, T. และคณะได้เสนอการใช้คุณลักษณะที่แบ่งออกเป็น 3 ระดับ คือ ระดับล่าง (low-level) ระดับกลาง (mid-level) และระดับบน (high-level) ในคุณลักษณะระดับล่างเกิดขึ้นจากคุณลักษณะของ ค่าความเข้มแสง มุม และค่าเปรียบต่างของสี ที่ได้ระบุไว้ในงานของ Itti, L.และคณะ (Itti *et al.*, 1998) ยิ่งกว่านั้น Judd, T. และคณะได้รวมคุณลักษณะระยะทางไปยังศูนย์กลาง (distance to center) ตำแหน่งค่าพลังงานแบบพีรามิด (local energy pyramids) และความน่าจะเป็นของช่องสี (probability color channels) ที่คำนวณจากฮิสโทแกรมของสีแบบ 3 มิติกับตัวกรองมัธยฐาน (median filter) 6 ระดับ ในคุณลักษณะระดับกลางเกิดขึ้นจากการตรวจจับเส้นระดับสายตา (Horizon Line) ในภาพ สุดท้ายในคุณลักษณะระดับบน เกิดขึ้นจากการ การตรวจจับใบหน้าของ Viola-Jones (Viola and Jones, 2004) กับตัวตรวจจับรถยนต์ และบุคคลโดย Felzenszwalb, P. F.และ คณะ (Felzenszwalb *et al.*, 2010)

ในการรวมทุกคุณลักษณะเข้าด้วยกันมีการใช้ Support Vector Machine (SVMs) ที่เป็นรูปแบบเชิงเส้นในการจำแนกเพื่อกำหนดค่าน้ำหนักให้กับแต่ละคุณลักษณะที่ใช้อย่างเหมาะสม ก่อนการคำนวณของทุกคุณลักษณะ จะมีการปรับขนาดภาพให้เป็น 200×200 จุดภาพเสมอ

สำหรับการฝึกและประเมินผลของโมเดลดังกล่าว ผู้เขียนได้ทำการเก็บฐานข้อมูลจำนวน 1003 ภาพจากบริการแบ่งปันภาพถ่าย และเก็บฐานข้อมูลของการทำนายแบบ fixation ของมนุษย์ด้วยผู้ใช้จำนวน 15 คนสำหรับค่าเฉลยของข้อมูล สำหรับแต่ละภาพทดสอบ จะมีการจัดอันดับและเลือก 10 จุดภาพจากค่าบน 20% ของตำแหน่งวัตถุเด่นให้เป็นค่าบวก และ 10 จุดภาพจากค่าล่าง 70% ของพื้นที่ที่ไม่ใช้วัตถุเด่นเป็นค่าลบ ซึ่งจะไม่มีการเลือกกลุ่มตัวอย่างใดๆ ที่ห่างจากขอบของภาพ 10 จุดภาพ ทั้งนี้เพื่อความเสถียรของกลุ่มตัวอย่างทั้งสอง ข้อเสียเปรียบหลักสำหรับโมเดลนี้คือ มีการคำนวณที่ซับซ้อนและใช้เวลามากสำหรับตัวตรวจจับในส่วนของคุณลักษณะระดับสูง อีกทั้ง Judd, T. และคณะไม่ได้ ทำการทดสอบว่าโมเดลนี้เหมาะสมและให้ประสิทธิภาพกับงานทางด้านการตรวจจับวัตถุเด่นในภาพอย่างไร

งานวิจัยของ Tie *et al.* (2011) ผู้เขียนได้สร้างโมเดลที่อยู่บนพื้นฐานการประมวลผลโดยยึดหลักความแตกต่างระหว่างกรอบที่พิจารณากับพื้นที่รอบกรอบนั้น เหมือนกับในหลายๆ งานวิจัยที่เป็นพื้นฐานของวิธีการคำนวณหาวัตถุเด่น ผู้เขียนได้แบ่งการพิจารณาคุณลักษณะในการหาวัตถุเด่นออกเป็น 3 รูปแบบ คือ ทีละตำแหน่ง (Local) แบบพื้นที่ (Region) และแบบภาพรวม (Global) โดยยึดหลัก (1) ความเปรียบต่างของพื้นที่ในหลายขนาด (Multi–scale local contrast) (2) ค่าฮิสโทแกรมรอบจุดพิจารณา (Center – surround histogram) ที่สังเกตว่าฮิสโทแกรมของหน้าต่างที่วาดรอบวัตถุนั้นมีระดับที่สูงกว่าหน้าต่างที่วาดในพื้นที่เดียวกันที่รอบหน้าต่างของวัตถุอีกทีหรือไม่ (3) การกระจายตัวในตำแหน่งของสี (Color spatial distribution) คือ การนับจำนวนจุดภาพเพื่อพิจารณาบริบททั้งหมดของภาพ สำหรับวิธีการกระจายตัวในตำแหน่งของสีและค่าความแปรปรวนของแต่ละสีในโมเดลมีการใช้ Gaussian Mixture Model (GMM) สำหรับการสร้างผลลัพธ์สุดท้ายออกมามีการใช้อัลกอริทึมเรียนรู้ Condition Random Fields (CRFs) ที่มีความสามารถในการรวมคุณลักษณะทั้ง 3 เข้าด้วยกัน

สำหรับการฝึกและประเมินผล Tie, L และคณะได้เก็บรวบรวม 2 ชุดข้อมูล คือ (1) ภาพทั่วไปที่มีวัตถุเด่นอยู่ในภาพทั้งหมด 20000 ภาพ (2) ผลเฉลยของชุดข้อมูลที่เป็นกล่องสีเหลี่ยมรอบวัตถุเด่นในภาพ ภาพผลลัพธ์ที่ได้จาก CRFs จะถูกเปลี่ยนเป็นแมพ และทำการตัดแบ่งภาพโดยใช้ค่า

ขีดแบ่ง (threshold) และทำการตีกรอบรอบตำแหน่งของวัตถุเด่นที่ทำการตรวจจับได้ เพื่อที่จะสามารถนำไปเปรียบเทียบกับภาพผลเฉลย แต่ว่าวิธีนี้มีการใช้เวลาในการประมวลผลที่นานเนื่องมาจากอัลกอริทึมเรียนรู้ที่เลือกใช้ในงานวิจัย

ในงานวิจัยของ Borji *et al.* (2012) ได้มีการศึกษาวิธีการดำเนินการอย่างกว้างขวางเกี่ยวกับประสิทธิภาพของโมเดลการตรวจจับสิ่งเด่นในภาพ และโมเดลการทำนายแบบ fixation อีกทั้งยังทำสรุปการเปรียบเทียบประสิทธิภาพของโมเดลเหล่านั้นบนชุดข้อมูลสาธารณะต่างๆ ที่ทำการเผยแพร่โดยเจ้าของงานวิจัย ตามผลการทดลองได้สรุปว่า (1) โมเดลที่สร้างสำหรับ การทำนายแบบ fixation มักจะได้ประสิทธิภาพที่ต่ำกว่าโมเดลที่สร้างเพื่อตรวจจับและตัดแบ่งวัตถุที่เด่นที่สุดในภาพ (2) การรวมโมเดลที่ทำงานได้ดีในแต่ละชุดข้อมูลเข้าด้วยกันจะทำให้ความถูกต้องของโมเดลนั้นมีประสิทธิภาพในการการทำงานที่ดีขึ้น

งานวิจัยของ Hui *et al.* (2012) ได้เสนอวิธีที่ง่ายและมีประสิทธิภาพในการคำนวณจุดภาพของวัตถุเด่นโดยที่ความละเอียดภาพคงเดิม ในขั้นแรก Hui, Z. และคณะได้เสนอวิธีการสร้างภาพเพื่อใช้พิจารณาที่แบ่งออกเป็นสี่ช่วงสี คือ แดง เขียว น้ำเงิน และ เหลือง โดยมีการปรับแต่งวิธีการคำนวณช่วงสีในงานของ Itti, L. และคณะ (Itti *et al.*, 1998) แล้วทำการเลือกออกมาหนึ่งช่วงสีโดยพิจารณาช่วงสีที่มีข้อมูลมากที่สุดโดยใช้ค่าความแปรปรวนเป็นเกณฑ์การวัด จากนั้นจุดภาพของวัตถุเด่นจะถูกคำนวณผ่านการรวมกันระหว่างคุณลักษณะความเปรียบต่าง กับฟังก์ชั่นความสนใจเชิงพื้นที่ในแต่ละช่องสี (channel) เพื่อสร้างภาพที่เก็บเฉพาะพื้นที่ของวัตถุเด่นออกมา โดยผู้เขียนได้วัดประสิทธิภาพบนฐานข้อมูลของ Achanta, R. และคณะ (Achanta *et al.*, 2009a)

งานวิจัย Erdem and Erdem (2013) ได้เสนอวิธีการใหม่ในคำนวณแบบล่างขึ้นบน โดยใช้ค่าความแปรปรวนร่วมเชิงพื้นที่เป็นคุณลักษณะ ผู้เขียนใช้รูปแบบของเมทริกซ์ในการพิจารณาค่าความแปรรวนร่วมที่ได้จากการแตกคุณลักษณะของพื้นที่ในภาพ Erdem, E. และ A. Erdem ได้คำนวณตำแหน่งของวัตถุเด่นโดยตรง จากส่วนของภาพที่แบ่งภาพออกส่วนๆ เพื่อหาค่าความความแตกต่างระหว่างส่วนที่พิจารณากับส่วนที่อยู่บริเวณรอบๆ ซึ่งถ้าค่าความแตกต่างระหว่างส่วนที่พิจารณากับส่วนที่อยู่บริเวณรอบๆ มีค่าความแตกต่างกันมากกว่าค่าเฉลี่ยของทั้งภาพก็จะได้รับการทำนายเป็นพื้นที่ๆ มีวัตถุเด่นอยู่ ในขั้นสุดท้าย ผู้เขียนได้มีการเสนอให้รวมวิธีการที่ให้ความสำคัญตรงกลางภาพ (center bias) เพื่อเพิ่มประสิทธิภาพโมเดล

ผู้เขียนได้ทดสอบประสิทธิภาพบนสองชุด คือ (1) ชุดข้อมูลของ Bruce, N. and J. Tsotsos (Bruce and Tsotsos, 2005) ที่ประกอบไปด้วยภาพสีที่ถ่ายในเมืองในลักษณะกลางแจ้งและในร่ม จำนวน 120 ภาพ (2) ชุดข้อมูลของ Judd, T. และคณะ (Judd *et al.*, 2009)

ตารางที่ 1 สรุปงานวิจัยที่เกี่ยวข้องเรียงตามปีที่พิมพ์

ที่	ปี	ชื่องานวิจัย	สิ่งที่นำเสนอ	ผลการทดลอง
1	1998	A Model of Saliency-Based Visual Attention for Rapid Scene Analysis (Itti *et al.*, 1998)	โมเดลของระบบความสนใจในการมองเห็นของมนุษย์ โดยยึดหลักความแตกต่างระหว่างกรอบที่พิจารณากับพื้นที่รอบกรอบนั้นๆ	สามารถที่จะตรวจจับตำแหน่งของวัตถุเด่นอย่างคร่าวๆ ได้
2	2003	Contrast-based image attention analysis by using fuzzy growing (Ma and Zhang, 2003)	เสนอทางเลือกในการวิเคราะห์ที่ละตำแหน่ง สำหรับสร้างผลลัพธ์ในการตรวจจับโดยใช้โมเดลการเติบโตแบบคลุมเครือ (fuzzy growth) เพื่อระบุพื้นที่ที่สนใจ	สามารถที่จะตรวจจับตำแหน่งของวัตถุเด่นอย่างคร่าวๆ ได้
3	2007	Graph-based visual saliency (Harel *et al.*, 2006)	เสนอวิธีแสดงภาพเป็นกราฟที่เชื่อมต่อกันแบบเต็ม (fully-connected graph) น้ำหนักของแต่ละโหนดจะถูกกำหนดเป็นคุณลักษณะเชิงพื้นที่ระหว่างสองจุดภาพ ดังนั้น ถ้าจุดภาพสองจุดมีความแตกต่างกันมากก็จะทำให้ค่าน้ำหนักมีค่ามากเช่นกัน	มีการทำงานกับภาพบนชุดข้อมูลของ Doves (Van Der Linde, I. และคณะ, 2009) ได้ดี แต่ผลลัพธ์ยังคงมีส่วนที่ไม่สนใจที่มาก

ตารางที่ 1 (ต่อ)

ที่	ปี	ชื่องานวิจัย	สิ่งที่นำเสนอ	ผลการทดลอง
4	2009	Frequency-tuned salient region detection (Achanta *et al.*, 2009a)	ใช้ประโยชน์จากคุณสมบัติของสเปกตรัมโดเมนของภาพสำหรับการตรวจจับสิ่งเด่น ด้วยวิธีการปรับจูนความถี่	ให้ผลลัพธ์เป็นรูปร่างของวัตถุเด่น ที่ตัดแบ่งวัตถุออกจากพื้นหลังด้วยค่าขีดแบ่งที่ผู้ใช้กำหนด
5	2009	Learning to predict where humans look (Judd *et al.*, 2009)	เสนอการตรวจจับวัตถุเด่นโดยการรวมชุดของคุณลักษณะในวงกว้างที่แบ่งออกเป็น 3 ระดับ คือ ระดับล่าง ระดับกลาง และระดับบน อีกทั้งยังเสนอชุดข้อมูลสำหรับทดสอบการทำนายแบบ fixation	สามารถทำงานกับภาพในลักษณะที่เป็นการทำนายแบบ fixation ได้ดีแต่มีการประมวลผลที่ช้าเนื่องมาจากคุณลักษณะระดับสูง (ระดับบน)
6	2011	Learning to Detect a Salient Object (Tie *et al.*, 2011)	เสนอการพิจารณาคุณลักษณะในการหาวัตถุเด่นออกเป็น 3 รูปแบบ คือ ทีละตำแหน่ง (Local) แบบพื้นที่ (Region) และแบบภาพรวม (Global) โดยใช้ โมเดล CRFs ในการรวมคุณลักษณะทั้งสามเข้าด้วยกัน ทั้งยังเสนอชุดข้อมูล MSRA สำหรับการวัดประสิทธิภาพในโมเดลการตรวจจับพื้นที่เด่นในภาพ	สามารถตีกรอบสี่เหลี่ยมรอบวัตถุเด่นโดยวัดความถูกต้องจากการซ้อนทับกันของผลเฉลยที่ตีกรอบสี่เหลี่ยมรอบวัตถุเด่นโดยใช้คน 9 คนกับผลลัพธ์ที่ได้จากโมเดล
7	2012	Salient object detection: a benchmark(Borji *et al.*, 2012)	ศึกษา สรุป และเปรียบเทียบประสิทธิภาพของโมเดลการตรวจจับสิ่งเด่นในภาพ กับโมเดลการทำนายแบบ fixation ที่มีอยู่ในปัจจุบัน	การรวมโมเดลที่ทำงานได้ดีในแต่ละชุดข้อมูลเข้าด้วยกันจะทำให้ความถูกต้องของโมเดลนั้นมีประสิทธิภาพในการการทำงานที่ดีขึ้น

ตารางที่ 1 (ต่อ)

ที่	ปี	ชื่องานวิจัย	สิ่งที่นำเสนอ	ผลการทดลอง
8	2012	A simple and effective saliency detection approach (Hui *et al.*, 2012)	เสนอวิธีที่ง่ายและมีประสิทธิภาพในการคำนวณจุดภาพของวัตถุเด่นโดยที่ความละเอียดภาพคงเดิม	มีการทำงานที่รวดเร็วถึงแม้ว่าจะมีส่วนที่ไม่สนใจในภาพติดมากับผลลัพธ์
9	2013	Visual saliency estimation by nonlinearly integrating features using region covariances (Erdem and Erdem, 2013)	เสนอวิธีการใหม่ในการคำนวณแบบล่างขึ้นบน โดยใช้ค่าความแปรปรวนร่วมเชิงพื้นที่เป็นคุณลักษณะในการพิจารณา	สามารถที่จะทำงานกับภาพที่มีความซับซ้อนหรือมีวัตถุเด่นในภาพมากกว่าหนึ่งวัตถุได้

อุปกรณ์และวิธีการ

อุปกรณ์

1. ฮาร์ดแวร์

1.1. MacBook Pro CPU 2.2 GHz Core 2 Duo 4GB 1067 MHz DDR3 of RAM Hard disk: WD Black scorpion 7200 RPM 750 GB

1.2. Intel Core i7-3770 CPU 3.40 GHz 16 GB 1600 MHz DDR3 of RAM Mainboard: ASUSTek P8Z77-V LE PLUS Hard disk: Seagate Barracuda 7200 RPM 500 GB

2. ซอฟต์แวร์

2.1. Mac OS X version 10.8.5

2.2. Windows 7 Ultimate Service Pack 1 64-bit Operating System

2.3. MATLAB 2013b

วิธีการ

1. หลักการทำงาน

งานวิจัยนี้จะแบ่งการทำงานออกเป็น 2 ส่วน คือ ส่วนของวิธีการใหม่ในการคำนวณหาตำแหน่งของวัตถุเด่นในภาพและการสร้างโมเดลการเรียนรู้สำหรับตรวจจับวัตถุเด่น ซึ่งจะอธิบายในหัวข้อ 1.1 และ 1.2 ตามลำดับ

1.1 ส่วนของการคำนวณหาตำแหน่งของวัตถุเด่นในภาพ (Saliency Location Map)

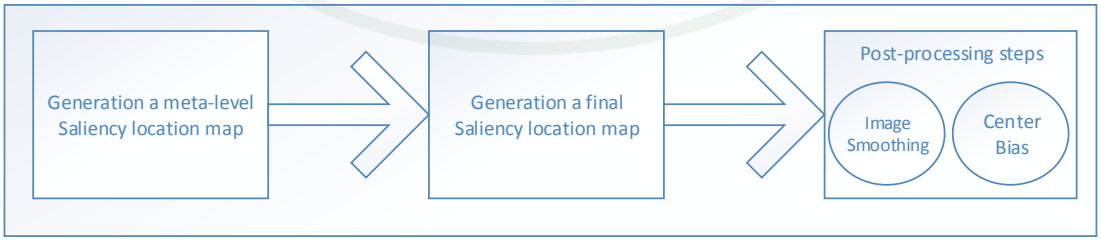
วิธีที่เสนอนี้มีการคำนวณตำแหน่งของวัตถุเด่นโดยใช้คุณลักษณะพื้นฐานของความเปรียบต่างของสี ในขั้นตอนแรกสำหรับการสร้างภาพที่เก็บเฉพาะตำแหน่งของวัตถุเด่นในระดับชั้นที่อยู่เบื้องหลัง ได้มีการปรับวิธีการตรวจจับวัตถุเด่นที่นำเสนอโดย Hui, Z. และคณะ (Hui *et al.*, 2012) ต่อจากนี้จะขอเรียกโมเดลที่นำเสนอโดย Hui, Z. และคณะ ว่าเป็นโมเดลของ Zhang ปกติแล้วผลลัพธ์จากขั้นตอนระดับชั้นที่อยู่เบื้องหลังจะประกอบไปด้วยส่วนที่ไม่สนใจที่ถูกทำนาย

เป็นจุดภาพของวัตถุเด่น ดังนั้นวัตถุประสงค์หลักอย่างหนึ่งของงานวิจัยนี้ คือ ต้องการที่จะลบส่วนที่ไม่สนใจออกจากระดับชั้นที่อยู่เบื้องหลังให้มากที่สุด เพื่อให้มีความเหมาะสมที่จะนำไปประยุกต์ในงานทางด้านการระบุตำแหน่งของวัตถุเด่น ซึ่งสามารถที่จะแสดงภาพโครงสร้างโมเดลของส่วนการคำนวณหาตำแหน่งของวัตถุเด่นในภาพ ดังในภาพที่ 7

ต่อจากนี้ขออ้างถึงวิธีการที่นำเสนอว่า MEV ที่มีความหมายมาจากค่าเฉลี่ยของความแปรปรวน (Mean of Variance) ซึ่งโมเดลที่นำเสนอนี้จะประกอบไปด้วย 3 ขั้นตอน ได้แก่ ขั้นตอนแรก คือ การสร้างในระดับชั้นที่อยู่เบื้องหลังสำหรับระบุตำแหน่งของวัตถุเด่น ขั้นตอนที่ 2 คือ การสร้างการระบุตำแหน่งของวัตถุเด่นขั้นสุดท้ายและขั้นตอนสุดท้าย คือ ขั้นตอนหลังการประมวลผล ดังมีรายละเอียดดังต่อไปนี้

1.1.1 การสร้างในระดับชั้นที่อยู่เบื้องหลังสำหรับระบุตำแหน่งของวัตถุเด่น (Generating a meta-level saliency location map)

ในบางงานวิจัย (Achanta *et al.*, 2009a; Erdem and Erdem, 2013; Harel *et al.*, 2006; Hui *et al.*, 2012; Itti *et al.*, 1998; Ma and Zhang, 2003; Tie *et al.*, 2011) ได้แสดงให้เห็นว่าคุณลักษณะของสีนั้นมีบทบาทสำคัญและยังเป็นองค์ประกอบหลักในทุกโมเดลการตรวจจับวัตถุเด่น บ่อยครั้งที่มีการใช้ปริภูมิสี Lab และ RGB ในปริภูมิสี Lab ได้ถูกออกแบบให้มีความใกล้เคียงกับการรับรู้ภาพของมนุษย์ โดยเฉพาะองค์ประกอบของ L ตรงกับความรับรู้ทางด้านความสว่างของมนุษย์ ส่วนโมเดลสี RGB เป็นพื้นฐานหลักในการออกแบบสำหรับการแสดงผลของภาพในระบบอิเล็กทรอนิกส์ แต่ RGB ยังเกี่ยวข้องอย่างใกล้ชิดกับการมองเห็นสีทั้งสามของมนุษย์ หรือความสามารถในการมองเห็นสีที่ต่างกันของมนุษย์



ภาพที่ 7 โครงสร้างโมเดลของส่วนการคำนวณหาตำแหน่งของวัตถุเด่นในภาพ

นอกเหนือจากองค์ประกอบช่วงสีแดง เขียว และน้ำเงิน Itti, L. และคณะ (Itti *et al.*, 1998) แนะนำว่าให้ใช้สีเหลือง เพิ่มเข้ามาเป็นองค์ประกอบสำคัญในการดึงดูดความสนใจของมนุษย์ ซึ่งสามารถเรียกปริภูมิสีแบบนี้ว่า RGBY ผู้วิจัยได้ทำการศึกษาและทดลองเพื่อวัดความสามารถของตัวตรวจจับวัตถุเด่นของทั้งสามปริภูมิสี คือ Lab, RGB และ RGBY ที่บรรยายโดย Hui, Z. และคณะ โดยใช้โมเดลการทำนายแบบ fixation ที่เสนอโดย Erdem, E. และ A. Erdem (Erdem and Erdem, 2013) ต่อจากนี้จะขอเรียกโมเดลที่นำเสนอโดย Erdem, E. และ A. Erdem ว่าเป็นโมเดลของ Erdem และโมเดลที่นำเสนอโดย Itti, L. และคณะว่าเป็นโมเดลของ Itti

จากตารางที่ 2 ผลการทดลองได้แสดงให้เห็นว่าระบบสี RGBY มีความสามารถในการตรวจจับวัตถุเด่นสูงที่สุด ทั้งนี้เนื่องมาจากมีค่าความถูกต้องสูงสุด ที่มากกว่าระบบสี LAB และ RGB ดังนั้นในงานวิจัยนี้ผู้วิจัยจึงตัดสินใจที่จะประยุกต์ใช้โมเดลสี RGBY ที่เหมือนกับการปรับช่องสี(four broadly-tuned) ที่บรรยายโดย Hui, Z. (Hui *et al.*, 2012)

$$R' = \left|R-(G+B)/2\right| \quad (1)$$

$$G' = \left|G-(R+B)/2\right| \quad (2)$$

$$B' = \left|B-(R+G)/2\right| \quad (3)$$

$$Y' = \left|(R+G)/2-\left|R-G\right|/2-B\right| \quad (4)$$

เมื่อ R คือ ค่าความเข้มแสงของจุดภาพในช่องสีแดง

G คือ ค่าความเข้มแสงของจุดภาพในช่องสีเขียว

B คือ ค่าความเข้มแสงของจุดภาพในช่องสีน้ำเงิน

R', G', B', Y' คือ สีที่ทำการปรับแต่งใหม่

ตารางที่ 2 แสดงค่าความถูกต้องสูงสุดในแต่ละปริภูมิสีหลังปรับแก้คุณลักษณะในโมเดลของ Erdem

วิธีการ	ค่าความถูกต้องสูงสุด
Erdem กับระบบสี RGBY	0.8183
Erdem กับระบบสี LAB	0.8010
Erdem กับระบบสี RGB	0.8041

สามารถสังเกตได้ว่าการปรับแต่งช่องสีที่เสนอในโมเดลของ Zhang กับโมเดลของ Ittiมีความแตกต่างระหว่างการคำนวณ คือโมเดลของ Zhang ได้ทำการพัฒนาวิธีการในโมเดลของ Ittiโดยใช้ค่าสมบูรณ์ในสมการปรับแต่งช่องสี วิธีการในโมเดลของ Itti แนะนำว่าค่าความเปลี่ยนแปลงของสีสันจะไม่สามารถรับรู้ได้ในค่าแสงที่น้อย ดังนั้น Itti, L. และคณะ (Itti *et al.*, 1998) จึงกำหนดทุกค่าที่น้อยหรือติดลบ ของ *R', G', B', Y'* เป็นศูนย์ ในทางกลับกัน วิธีการในโมเดลของ Zhang แนะนำว่าค่าที่เป็นลบที่มีค่ามากอาจทำให้เกิดผลกระทบที่สำคัญในการดึงดูดความสนใจของมนุษย์ ทำให้ค่าเหล่านี้ควรจะนำมาพิจารณา ในงานวิจัยนี้ผู้วิจัยได้ประยุกต์ใช้แนวคิดหลัง

จากภาพที่ 8 สามารถสังเกตได้ว่าในโมเดลของ Itti จะให้ข้อมูลสีของพื้นหลังมากกว่าตัววัตถุหลังจากการปรับของช่วงสีน้ำเงินดังแสดงในภาพที่ 8d ต่างกับโมเดลของ Zhang ที่ยังคงให้ข้อมูลสีตรงกับตำแหน่งของวัตถุ หลังจากการปรับช่วงสีน้ำเงินดังแสดงในภาพที่ 8h ยิ่งกว่านั้นจะเห็นได้ว่าผลลัพธ์หลังจากการปรับทุกช่วงสีในโมเดลของ Zhang ให้ผลลัพธ์ของค่าความเปรียบต่างของสีที่มีประสิทธิภาพมากกว่าวิธีการปรับช่วงสีที่เสนอในโมเดลของ Itti

เนื่องจากพื้นที่ของวัตถุเด่นมีแนวโน้มที่จะเกิดในตำแหน่งของความเปรียบต่างของสีที่มากในภาพ (high color contrast) ดังนั้น ช่วงสีที่มีความเปรียบต่างของสีที่มากจะถูกเลือกสำหรับการคำนวณวัตถุเด่น ในขั้นตอนต่อไป ช่องความเปรียบต่างของสีจะถูกระบุโดยความแปรปรวนของค่าความเข้มแสงของจุดภาพ (intensity) ในแต่ละช่องสี ช่องสีที่มีค่าความแปรปรวนสูงสุดจะถูกนำมาใช้สำหรับการคำนวณในระดับชั้นที่อยู่เบื้องหลังของการหาตำแหน่งของวัตถุเด่นในภาพ

$$var = \frac{1}{W \times H} \sum_i \sum_j (I(i,j) - \overline{I}_c)^2 \qquad (5)$$

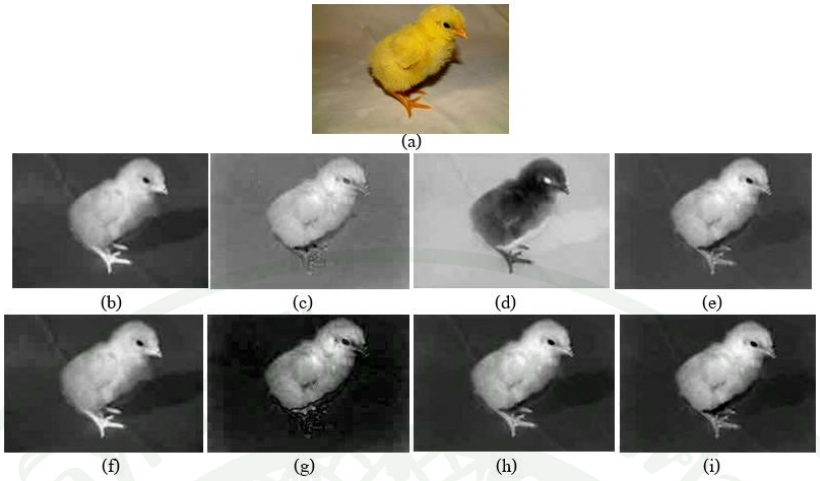
เมื่อ W คือ ขนาดความกว้างของภาพนำเข้า

H คือ ขนาดความสูงของภาพนำเข้า

$I(i,j)$ คือ ค่าความเข้มแสงของจุดภาพในตำแหน่งแถวที่ *i* หลักที่ *j* ของภาพนำเข้า

$\overline{I}_c$ คือ ค่าเฉลี่ยของความเข้มในภาพ

var คือ ค่าความแปรปรวน



ภาพที่ 8 แสดงผลลัพธ์ของการปรับช่วงสี (a) ภาพนำเข้า (b-e) ผลลัพธ์จากการปรับช่วงสีของ Itti, L. และคณะ (Itti *et al.*, 1998) (f-i) ผลลัพธ์จากการปรับช่วงสีของ Hui, Z. และคณะ (Hui *et al.*, 2012) โดยเรียง *R', G', B', Y'* ตามลำดับ

โดยธรรมชาติแล้ว สิ่งที่เด่นควรเกิดขึ้นในพื้นที่ที่มีค่าของความเปรียบต่างของสีสูง ระดับของความเปรียบต่างในแต่ละจุดภาพสามารถใช้ในการตรวจสอบความเป็นไปได้ของจุดภาพว่าเป็นจุดภาพของวัตถุเด่นหรือไม่ ตัวอย่างเช่นระดับที่สูงของความเปรียบต่างสามารถบ่งบอกถึงว่าความเป็นไปได้ที่จะเป็นสิ่งที่เด่นสูง ในงานวิจัยนี้ผู้วิจัยได้กำหนดระดับของความเปรียบต่างในแต่ละจุดภาพเป็นค่าความต่างในค่าสัมบูรณ์ ระหว่างค่าความเข้มแสงของจุดภาพในแต่ละตำแหน่ง กับค่าเฉลี่ยของความเข้มในภาพ ที่เสนอในโมเดลของ Zhang เป็น

$$I_{diff}(i,j) = \left| I(i,j) - \bar{I}_c \right| \tag{6}$$

เมื่อ $I_{diff}(i,j)$ คือ ค่าความต่างในค่าสัมบูรณ์

$I(i,j)$ คือ ค่าความเข้มแสงของจุดภาพในตำแหน่งแถวที่ i หลักที่ j ของภาพนำเข้า

$\bar{I}_c$ คือ ค่าเฉลี่ยของความเข้มในภาพ

เหตุผลที่อยู่เบื้องหลังการใช้สมการนี้ คือ ค่าจุดภาพที่ใกล้เคียงกับ $\bar{I}_c$ ควรจะมีความเป็นไปได้ต่ำที่จะเป็นจุดภาพของสิ่งเด่น (ความเปรียบต่างต่ำ) ในขณะที่จุดภาพที่มีความแตกต่างสูงกับ $\bar{I}_c$ ควรจะมีความเป็นไปได้สูงที่จะเป็นจุดภาพของสิ่งเด่น (ความเปรียบต่างสูง)

1.1.2 การสร้างการระบุตำแหน่งของวัตถุเด่นขั้นสุดท้าย (Generating a final saliency location map)

จากการระบุตำแหน่งของวัตถุเด่นในระดับชั้นที่อยู่เบื้องหลัง สามารถได้พื้นที่ที่ครอบคลุมสิ่งเด่นพอสมควร แต่น่าเสียดายที่ยังมีจุดภาพของสิ่งที่ไม่สนใจจำนวนมาก หรือเรียกได้ว่าเป็นค่าความผิดพลาดเชิงบวก (false positive) ตัวอย่างเช่น จุดภาพของพื้นหลังที่ถูกทำนายเป็นจุดภาพของวัตถุเด่น ดังนั้นผู้วิจัยจึงจำเป็นต้องลดจุดภาพที่เป็นความผิดพลาดเชิงบวกให้มากที่สุดที่จะเป็นไปได้เพื่อสร้างภาพของคุณลักษณะที่ดีในการประมาณตำแหน่งของวัตถุเด่น สำหรับวัตถุประสงค์นี้ ผู้วิจัยได้เสนอที่จะใช้ค่าความแปรปรวนของความต่างในค่าสัมบูรณ์ $I_{diff}(i,j)$ ที่ทำการสกัดส่วนพื้นที่ของผลลัพธ์ในระดับชั้นที่อยู่เบื้องหลัง ดังนั้นผู้วิจัยจึงเสนอการหาตำแหน่งของสิ่งเด่นโดยใช้ความเปรียบต่างของสี (Vijanprecha and Wattuya, P, 2014) เพื่อแก้ปัญหานี้

เพื่อที่จะทำการประมาณตำแหน่งของวัตถุเด่น ผู้วิจัยได้ปรับขนาดผลลัพธ์ของระดับชั้นที่อยู่เบื้องหลังเป็นขนาด 512×512 จุดภาพ และแบ่งภาพที่ระบุตำแหน่งของวัตถุเด่นในส่วนระดับชั้นที่อยู่เบื้องหลังออกเป็นส่วนๆ (patches) ขนาด 8×8 ที่ไม่ซ้อนกันเป็น p_i ซึ่งจากผลการทดลองได้แสดงให้เห็นว่าถ้าหน้าต่างมีขนาดที่ใหญ่กว่าที่กำหนดจะส่งผลให้การแยกแยะพื้นหลังมีความถูกต้องที่ลดลง ถึงแม้ว่าจะประหยัดเวลาในการทำงาน แต่ถ้าหน้าต่างมีขนาดเล็กกว่าที่กำหนดจะทำให้มีการประมวลผลที่นานและไม่เหมาะสมสำหรับการระบุตำแหน่งของวัตถุเด่น

สำหรับการพิจารณาหลังแบ่งภาพออกเป็นแต่ละส่วน ค่าความแปรปรวนในพื้นที่ $I_{diff}(i,j)$ ของทุกจุดภาพในแต่ละส่วนจะถูกคำนวณ ผู้วิจัยแสดงค่าความแปรปรวนในพื้นที่ $I_{diff}(i,j)$ แต่ละส่วน p_i เป็น var_{pi} ดังแสดงในสมการที่ 7 เนื่องจากความแปรปรวนในพื้นที่ var_{pi} สามารถจับการเปลี่ยนแปลงระดับความเข้มภายในแต่ละส่วนของภาพ ซึ่งส่วนที่มีการเปลี่ยนแปลงความเข้มสูงจะได้รับการพิจารณาว่าเป็นส่วนที่เด่น ในขณะที่ส่วนที่มีการเปลี่ยนแปลงค่าความเข้มแสงน้อย ก็จะได้รับการพิจารณาว่าเป็นส่วนพื้นหลัง

$$var_{pi} = \frac{1}{k \times k} \sum_{x}^{k} \sum_{y}^{k} (p_i(x,y) - S_{pi})^2 \qquad (7)$$

เมื่อ k คือ ขนาดความกว้างและความสูงของการแบ่งพื้นหลังออกเป็นส่วนๆ

$p_i(x,y)$ คือ ค่าความเข้มแสงของจุดภาพในตำแหน่งแถวที่ x หลักที่ y ของแต่ละส่วนหลังทำการแบ่งภาพ

S_{pi} คือ ค่าเฉลี่ยของความเข้มในแต่ละส่วนหลังทำการแบ่งภาพ

var_{pi} คือ ค่าความแปรปรวนในพื้นที่ $I_{diff}(i,j)$ ในแต่ละส่วน p_i

ในงานวิจัยนี้ผู้วิจัยตรวจสอบสิ่งเด่นของแต่ละส่วนหลังทำการแบ่งส่วนในระดับชั้นที่อยู่เบื้องหลัง โดยใช่ค่าเฉลี่ยของความแปรปรวนในพื้นที่ของทุกส่วน avg_var_{pi} ในแต่ละส่วนของสิ่งเด่นจะถูกคำนวณเพื่อหาค่าเฉลี่ยของ $I_{diff}(i,j)$ ทุกจุดภาพในส่วนที่ตัด s_{pi} ซึ่งค่า s_{pi} สามารถแสดงให้เห็นถึงระดับของส่วนที่เป็นสิ่งเด่น ถ้าค่า s_{pi} มีค่ามากแสดงว่ามีระดับเป็นส่วนที่เป็นพื้นที่สิ่งเด่นสูง

$$Sal(p_i) = \begin{cases} S_{pi} & if \ var_{pi} \geq avg\_var_{pi} \\ 0 & if \ var_{pi} < avg\_var_{pi} \end{cases} \quad (8)$$

เมื่อ $Sal(p_i)$ คือ ภาพแสดงตำแหน่งของวัตถุเด่นที่สร้างจากการรวมทุก p_i เข้าด้วยกัน

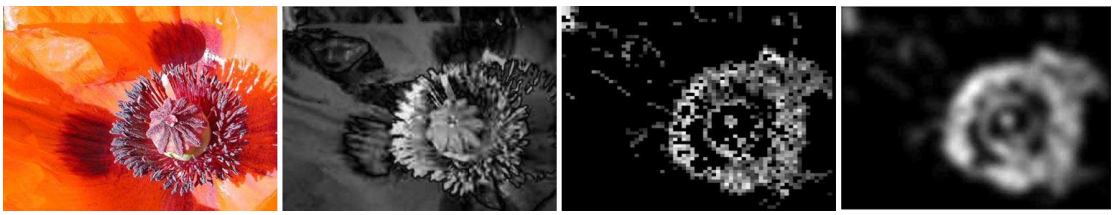
S_{pi} คือ ค่าเฉลี่ยของความเข้มในแต่ละส่วนหลังทำการแบ่งภาพ

var_{pi} คือ ค่าความแปรปรวนในพื้นที่ $I_{diff}(i,j)$ ในแต่ละส่วน p_i

avg_var_{pi} คือ ค่าเฉลี่ยของความแปรปรวนในพื้นที่ของทุกส่วน

ภาพแสดงตำแหน่งของวัตถุเด่นจะถูกสร้างโดยการกำหนด S_{pi} ไปในทุกจุดภาพของส่วนที่ตัดแบ่ง ถ้าค่า var_{pi} มากกว่า avg_var_{pi} ในส่วนของส่วนที่ตัดแบ่งที่มีค่าน้อยกว่า avg_var_{pi} จะถูกตั้งให้เป็นศูนย์ เพื่อแสดงเป็นพื้นที่ของพื้นหลัง

จากภาพที่ 9 แสดงตัวอย่างของผลลัพธ์การระบุตำแหน่งของวัตถุเด่นในภาพ ซึ่งสามารถสังเกตได้ว่าผลลัพธ์จากการระบุตำแหน่งขั้นสุดท้ายในภาพที่ 9c สามารถที่จะกำจัดส่วนพื้นที่ที่ไม่ใช้วัตถุเด่นในภาพออกจากผลลัพธ์ได้เป็นอย่างมาก หลังจากการสร้างใน ระดับชั้นที่อยู่เบื้องหลังดังแสดงในภาพที่ 9b



(a) (b) (c) (d)

ภาพที่ 9 ตัวอย่างของผลลัพธ์การระบุตำแหน่งของวัตถุเด่นในภาพ (a) ภาพนำเข้า (b) ภาพที่ได้หลังจากการสร้างในระดับชั้นที่อยู่เบื้องหลัง (c) ผลลัพธ์จากการระบุตำแหน่งขั้นสุดท้าย (d) ผลลัพธ์หลังการปรับภาพตำแหน่งของวัตถุเด่นให้เรียบและการให้ความสำคัญกับตำแหน่งกลางภาพ

1.1.3 ขั้นตอนหลังการประมวลผล (Post-processing steps)

สองขั้นตอนหลักสำหรับขั้นตอนหลังการประมวลผลที่นิยมใช้อย่างกว้างขวางในโมเดลการตรวจจับวัตถุเด่นที่เสนอในวรรณกรรม คือ การปรับภาพตำแหน่งของวัตถุเด่นให้เรียบและการให้ความสำคัญกับตำแหน่ง กลางภาพ

1.1.3.1 การปรับภาพตำแหน่งของวัตถุเด่นให้เรียบ

โดยในงานวิจัยนี้ ผลลัพธ์ของภาพระบุตำแหน่งวัตถุเด่นได้ทำการปรับเรียบด้วยวิธีคอนโวลูชั่นด้วยตัวกรองแบบเกาส์เซียน (สมการรูประฆังคว่ำ) ขนาดเล็ก เนื่องจากรูปร่างของตัวกรองง่ายต่อการกำหนด เมื่อกำหนดให้ $G(x,y)$ เป็นตัวกรองเกาส์เซียนในโดเมนสถานที่ (spatial domain) ฟังก์ชันตัวกรองเกาส์เซียนสามารถกำหนดได้โดย

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (9)$$

เมื่อ $G(x,y)$ คือ ตัวกรองเกาส์เซียนในโดเมนสถานที่

σ คือ ส่วนเบี่ยงเบนมาตรฐาน

x คือ ระยะทางจากจุดกำเนิดในแกนนอน

y คือ ระยะทางจากจุดกำเนิดในแกนตั้ง

จากสมการที่ 9 ผู้วิจัยได้กำหนดค่าส่วนเบี่ยงเบนมาตรฐานเป็นผลคูณระหว่างค่าคงที่ 0.02 และกำหนดขนาดกับความกว้างของภาพนำเข้า และกำหนดขนาดความกว้างของเมทริกซ์จัตุรัสเป็น 4 เท่าของส่วนเบี่ยงเบนมาตรฐาน ค่าดังกล่าวได้รับการศึกษาและถูกนำเสนอโดย Zhang (Hui *et al.*, 2012)

1.1.3.2 การให้ความสำคัญกับตำแหน่งกลางภาพ

เนื่องจากว่าการเพิ่มความสำคัญกับตำแหน่งกลางภาพให้กับภาพผลลัพธ์หลังการตรวจจับตำแหน่งของวัตถุเด่น ได้รับการพิสูจน์ว่าสามารถที่จะปรับปรุงความถูกต้องของผลลัพธ์ (Erdem and Erdem, 2013; Hui *et al.*, 2012; Judd *et al.*, 2009; Schauerte and Stiefelhagen, 2013) ที่เป็นอย่างนี้มากจากเหตุผลหลักที่ว่า บ่อยครั้งที่วัตถุเด่นมักจะตั้งอยู่ตรงกลางภาพ ตัวอย่างเช่น ลักษณะการเน้นความสำคัญและพฤติกรรมของนักถ่ายภาพ สำหรับวิธีในการให้ความสำคัญตรงกลางภาพ ในบางงานวิจัยได้ใช้เกาส์เซียน เพื่อให้ความสำคัญตรงกลางภาพ (Judd *et al.*, 2009) และวิธีการที่ซับซ้อนมากขึ้นเช่นการคำนวณการกระจายของสิ่งเด่นก่อนโดยใช้ชุดข้อมูลฝึกในการเรียนรู้ (Fu *et al.*, 2013) ในงานวิจัยนี้ผู้วิจัยใช้วิธีการเพิ่มความสำคัญกับตำแหน่งกลางภาพที่เสนอโดย Hui, Z. และคณะ (Hui *et al.*, 2012) ซึ่งการคำนวณขึ้นอยู่เพียงแค่ตำแหน่งระยะทางระหว่างแต่ละจุดภาพถึงตรงกลางภาพของภาพนำเข้า เมื่อกำหนดให้ $C(i,j)$ เป็นฟังก์ชันของตำแหน่งความสนใจที่สะท้อนถึงแนวโน้มของตำแหน่งที่ตามนุษย์ให้ความสนใจ ฟังก์ชันสำหรับการให้ความสำคัญกับตำแหน่งกลางภาพสามารถคำนวณได้โดย

$$C(i,j) = \frac{1}{1+d(i,j)} \Big/ \left(\frac{L}{2}\right) \tag{10}$$

เมื่อ $C(i,j)$ คือ ฟังก์ชันของตำแหน่งความสนใจที่สะท้อนถึงแนวโน้มของตำแหน่งที่ตามนุษย์ความสนใจ

$d(i,j)$ คือ ระยะทางระหว่างจุดภาพที่ตำแหน่ง *(i, j)* กับ ตำแหน่งตรงกลางภาพนำเข้า

L คือ ความยาวเส้นทแยงมุมของภาพนำเข้า

1.2 การประยุกต์ใช้โมเดล MEV ในโมเดลการเรียนรู้สำหรับตรวจจับวัตถุเด่น

เพื่อที่จะทำงานกับข้อมูลของการทำนายแบบ fixation ผู้วิจัยใช้วิธีการเรียนรู้ด้วย ข้อมูลฝึก ที่ทำการจำแนกจากการติดตามของตามนุษย์ที่มองเข้าไปในภาพโดยตรง (Judd *et al.*, 2009) ในขั้นตอนการฝึกสำหรับโมเดลการเรียนรู้ที่รวมกับโมเดลการตรวจจับวัตถุเด่น ผู้วิจัยได้รวมส่วนของการคำนวณหาตำแหน่งของวัตถุเด่นในภาพที่ทำการเสนอเข้ากับคุณลักษณะ ระดับล่าง (low-level) กับระดับกลาง (mid-level) ที่เสนอในงานวิจัยของ Judd (Judd *et al.*, 2009) และปราศจากคุณลักษณะระดับสูงในโมเดลของ Judd

อุปสรรคที่สำคัญของการใช้คุณลักษณะระดับสูง เช่น ตัวตรวจจับใบหน้า บุคคล หรือรถยนต์ คือ อัตราของค่าความผิดพลาดเชิงบวกตัวตรวจจับ ประสิทธิภาพในการตรวจจับสิ่งเด่นของงานวิจัยของ Judd (Judd *et al.*, 2009) ขึ้นอยู่กับอัตราของค่าความผิดพลาดเชิงบวกในตัวตรวจจับที่สูง ซึ่งเกิดจากเวลาที่ตัวตรวจจับวัตถุระบุ สิ่งที่ตรวจจับผิดพลาดถึงแม้ว่าจะไม่มีวัตถุที่ระบุอยู่ในภาพก็ตาม ซึ่งอัตราค่าความผิดพลาดเชิงบวกที่สูงอาจจะลดความถูกต้องของการตรวจจับวัตถุเด่นในภาพ นอกจากนี้ตัวตรวจจับวัตถุจะไม่มีประโยชน์มากเมื่อไม่มีวัตถุที่ระบุดังกล่าวในภาพ อีกทั้งยังใช้ค่าใช้จ่ายที่สูงสำหรับการคำนวณคุณลักษณะระดับสูงของตัวตรวจจับ

โดยผู้วิจัยมีแนวคิดที่จะนำโมเดล MEV เข้าไปแทนที่คุณลักษณะระดับสูง เพื่อลดเวลาในการคำนวณและมีความสามารถที่จะตรวจจับวัตถุเด่นในชุดข้อมูลภาพทั่วไป สำหรับวิธีการแทนที่คุณลักษณะระดับสูงในโมเดลการตรวจจับวัตถุเด่นด้วยโมเดล MEV ผู้วิจัยได้นำผลลัพธ์จากโมเดล MEV หรือแมพเข้าเป็นหนึ่งในคุณลักษณะของโมเดลการตรวจจับวัตถุเด่น ซึ่งมีผลทำให้ต้องมีกระบวนการเรียนรู้โมเดลการตรวจจับวัตถุเด่นใหม่

ข้อได้เปรียบหลักของการใช้วิธีการระบุตำแหน่งวัตถุเด่นในภาพที่นำเสนอ แทนที่คุณลักษณะระดับสูง คือ ต้องการเวลาในการคำนวณที่น้อยกว่าและยังทำให้การตรวจจับวัตถุเด่นทำงานได้ดีกับชุดข้อมูลภาพวัตถุเด่นทั่วไป

2. การวัดประสิทธิภาพของวิธีการทำงาน

สำหรับการวัดประสิทธิภาพของวิธีการทำงาน ผู้วิจัยได้ใช้วิธีการเปรียบเทียบที่เป็นที่ยอมรับและนิยมใช้ในงานวิจัยในปัจจุบัน คือ วิธีการวัดแบบอัตราของค่าความถูกต้องกับค่าความครบถ้วน กับความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) เพื่อทำการแสดงกราฟ ROC เนื่องจากช่วยในการเปรียบเทียบประสิทธิภาพของการโมเดลตรวจจับวัตถุเด่น จากเปรียบเทียบพื้นที่ใต้เส้นโค้งของการตรวจจับแต่ละชนิด ถ้าโมเดลมีพื้นที่ใต้กราฟมากนั้นแสดงว่ามีประสิทธิภาพที่สูง

2.1 วิธีการวัดแบบอัตราของค่าความถูกต้องกับค่าความครบถ้วน (Precision-Recall curve)

ผู้วิจัยได้เตรียมการเปรียบเทียบเพื่อวัดประสิทธิภาพของวิธีที่นำเสนอกับ 6 โมเดลที่เสนอในปัจจุบัน คือ Judd (Judd *et al.*, 2009), Erdem (Erdem and Erdem, 2013), IT (Itti *et al.*, 1998), Zhang (Hui *et al.*, 2012), GB (Harel *et al.*, 2006) และ AC (Achanta *et al.*, 2009a) บนฐานข้อมูลจำนวน 1000 ภาพ (Tie *et al.*, 2011) ที่มีผลเฉลยเป็นภาพแบบทวิภาค (Achanta *et al.*, 2009a) ยิ่งกว่านั้น ผู้วิจัยได้ทำการทดสอบบนชุดข้อมูล SED2 ที่มีวัตถุเด่นสองวัตถุประกอบอยู่ในภาพ โดยมีผลเฉลยเป็นภาพแบบทวิภาคกับโมเดลในปัจจุบัน คือ Zhang, Erdem และ Judd

ผู้วิจัยประเมินประสิทธิภาพของโมเดลที่นำเสนอโดยวิธีการวัดแบบอัตราของค่าความถูกต้องกับค่าความครบถ้วน ที่เสนอโนงานวิจัยของ Achanta, R. และคณะ (Achanta *et al.*, 2009a) ในการเปรียบเทียบของภาพที่แสดงเฉพาะส่วนเด่น ผู้วิจัยได้นำผลลัพธ์ของโมเดล IT, GB และ AC จำนวน 1000 ภาพ ที่ถูกเตรียมโดย Achata, R. และคณะ ในโมเดลของ Judd กับ Erdem ได้ผลของภาพที่แสดงเฉพาะส่วนเด่นจากการคำนวณด้วยซอร์สโค้ดที่ถูกนำเสนอโดยเจ้าของโมเดลบนอินเตอร์เน็ต ยกเว้นโมเดลของ Zhang ที่ถูกสร้างขึ้นมาใหม่ตามวิธีที่เสนอในเอกสารงานวิจัยโดยผู้วิจัย

พื้นฐานในการคำนวณอัตราของค่าความถูกต้องกับค่าความครบถ้วนได้รับจากอัตราของค่าความถูกต้อง กับอัตราของค่าความครบถ้วน ที่เปลี่ยนแปลงอย่างต่อเนื่องโดยค่าที่กำหนด (Threshold) ที่ละ 0.005 สำหรับแต่ละขั้นในช่วงกว้างของค่าเท่ากับ 0 ถึง 1 ค่าอัตราความถูกต้องแสดงถึงอัตราส่วนระหว่างพื้นที่ๆ ทำนายว่าเป็นพื้นที่เด่นแล้วถูกต้องกับพื้นที่ๆ ทำนายว่าเป็นพื้นที่

เด่นที่โมเดลทำได้ทั้งหมด ถ้าอัตราของความถูกต้อง เท่ากับ 1 แสดงว่า พื้นที่ๆ ทำนายมาได้ทั้งหมดเป็นคำตอบที่ถูกต้อง ค่าอัตราความครบถ้วน คือ อัตราส่วนระหว่างพื้นที่ๆ ทำนายว่าเป็นพื้นที่เด่นแล้วถูกต้อง กับพื้นที่ๆ ถูกต้องทั้งหมดของผลเฉลย ซึ่งเน้นให้คำตอบที่ถูกต้องออกมาให้ได้มากที่สุด

2.2 ความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) เพื่อทำการแสดงกราฟ ROC

ผู้วิจัยได้ทำการวัดประสิทธิภาพโมเดลที่นำเสนอกับโมเดลของ Judd (Judd *et al.*, 2009) บนชุดข้อมูล MIT (Judd *et al.*, 2009) ด้วยกราฟ ROC ทั้งนี้เพราะผู้วิจัยไม่สามารถที่จะเปรียบเทียบโมเดลที่นำเสนอในข้อมูลชุดแรกได้อย่างเดียว เพราะงานวิจัยที่นำมาเปรียบเทียบในข้อมูลชุดแรกมีวัตถุประสงค์ของงานที่ต่างจากโมเดลที่นำเสนอ ในชุดข้อมูล MSRA (Tie *et al.*, 2011) จะมีสิ่งเด่นอยู่ในภาพแค่สิ่งเดียว ทั้งนี้เพื่อป้องกันความกำกวมของผลลัพธ์ แต่ในชุดข้อมูล MIT จะประกอบไปด้วยตำแหน่งของวัตถุเด่นมากว่าหนึ่ง ขึ้นอยู่กับผลเฉลยของตำแหน่งที่ตามองลงไปในภาพ สำหรับความยุติธรรมในการเปรียบระหว่างโมเดลที่นำเสนอกับโมเดลของ Judd (Judd *et al.*, 2009) ผู้วิจัยได้สุ่มภาพจากชุดข้อมูล MIT จำนวน 100 ภาพ เพื่อใช้ในการฝึกสำหรับโมเดลที่นำเสนอกับโมเดลของ Judd (Judd *et al.*, 2009) ในส่วนข้อมูลทดสอบ ผู้วิจัยได้ใช้ทุกภาพที่อยู่ในชุดข้อมูล MIT สร้างกราฟ ROC ในการหาค่าน้ำหนักที่เหมาะสมในแต่ละคุณลักษณะ ผู้วิจัยได้ใช้ตัวจำแนก Support Vector Machines (SVMs) เหมือนในงานของ Judd (Judd *et al.*, 2009)

ผลและวิจารณ์

ผล

ผลลัพธ์ของขั้นตอนวิธีที่เสนอในงานวิจัยนี้ สามารถแบ่งออกเป็นสองส่วนหลักตามลักษณะของชุดข้อมูล (1) การทดลองที่ระบุพื้นที่วัตถุเด่นในภาพ (2) การทดลองที่เป็นการระบุตำแหน่งที่ตามนุษย์มองเข้าไปในภาพ สำหรับในส่วนรายงานการทดลอง ผู้วิจัยจะเรียกวิธีการที่นำเสนอว่า *JuddLM + MEV* ซึ่งมาจาก โมเดลของ Judd ที่ใช้คุณลักษณะระดับล่าง (Low-level) และระดับกลาง (Mid-level) กับวิธีการระบุตำแหน่งของวัตถุเด่นในภาพที่นำเสนอ ที่คำนวณบนพื้นฐานค่าเฉลี่ยของความแปรปรวน(Mean of Variance) โดยผลการทดลองสามารถแสดงรายละเอียดดังต่อไปนี้

1. การทดสอบประสิทธิภาพโมเดล MEV

1.1 การทดสอบประสิทธิภาพการระบุพื้นที่วัตถุเด่นในภาพ

สำหรับการทดสอบกับโมเดลการตรวจจับวัตถุเด่นนี้ ผู้วิจัยได้ใช้ฐานข้อมูล MSRA เนื่องจากว่าเป็นฐานข้อมูลมาตรฐานที่ใช้ในการวัดความถูกต้องของโมเดลการตรวจจับสิ่งเด่นในหลายงานวิจัยทั้งในอดีตและปัจจุบัน (Achanta *et al.*, 2009a; Hui *et al.*, 2012; Tie *et al.*, 2011) ยิ่งกว่านั้นยังเป็นฐานข้อมูลสาธารณะที่สามารถนำมาใช้สำหรับการเปรียบเทียบประสิทธิภาพ (Achanta *et al.*, 2009b) ลักษณะชุดข้อมูล MSRA (Tie *et al.*, 2011) จะมีตำแหน่งของวัตถุเด่นที่ชัดเจน กล่าวคือ ในภาพจะประกอบไปด้วยวัตถุเด่นเพียงสิ่งเดียว จึงให้ไม่เกิดความกำกวมในการจำแนกวัตถุเด่นออกจากพื้นหลัง อย่างที่กล่าวข้างต้น ผลเฉลยของ MSRA (Tie *et al.*, 2011) มีความห่างไกลความถูกต้องมาก ดังนั้นในงานวิจัยนี้จึงได้เลือกใช้ผลเฉลยที่เสนอโดย Achanta, R. และคณะ (Achanta *et al.*, 2009a) ยิ่งกว่านั้น ยังทำการทดสอบบนฐานข้อมูล SED2 (Alpert *et al.*, 2007) ที่ภาพประกอบไปด้วยวัตถุเด่นสองวัตถุในภาพ ทั้งนี้ผู้วิจัยต้องการแสดงให้เห็นถึงประสิทธิภาพของโมเดลที่นำเสนอนั้นสามารถที่จะทำงานกับภาพที่ประกอบไปด้วยวัตถุเด่นในภาพมากกว่าหนึ่งวัตถุ

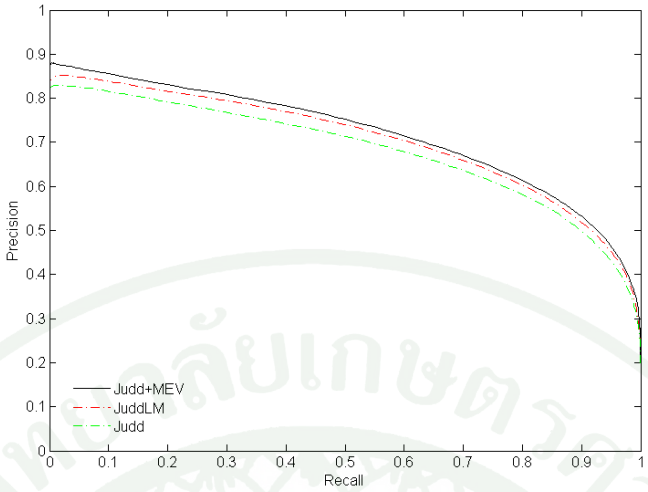
สำหรับวิธีที่ใช้ในเปรียบเทียบเพื่อวัดประสิทธิภาพของวิธีการที่นำเสนอ จะใช้พื้นฐานในการคำนวณอัตราของค่าความถูกต้องกับค่าความครบถ้วนที่กล่าวไว้ข้างต้น เปรียบเทียบกับวิธีในปัจจุบัน เพื่อแสดงให้เห็นว่าวิธีที่นำเสนอมีประสิทธิภาพที่ดีในด้านความถูกต้องบนชุดข้อมูล MSRA สำหรับการพิจารณากราฟผลลัพธ์ จะทำการเปรียบเทียบโมเดลในปัจจุบันกับโมเดลที่นำเสนอ แยกออกจากกัน ทั้งนี้เพื่อความง่ายในการพิจารณาในแต่ละกรณี

1.1.1 เปรียบเทียบประสิทธิภาพระหว่างโมเดลของ Judd กับโมเดลที่เสนอบนชุดข้อมูล MSRA

สำหรับวัตถุประสงค์ในการเปรียบเทียบประสิทธิภาพของกราฟผลลัพธ์ที่ได้จากโมเดลของ Judd กับโมเดลที่เสนอ ผู้วิจัยต้องการแสดงให้เห็นว่า หลังจากที่ได้ทำการแทนพื้นฐานคุณลักษณะระดับสูง ด้วยโมเดลที่นำเสนอลงในโมเดลของ Judd นั้นสามารถที่จะเพิ่มประสิทธิภาพในการตรวจจับพื้นที่ของวัตถุเด่นในชุดข้อมูล MSRA

จากภาพกราฟที่ 10 แสดงผลการทดสอบประสิทธิภาพในการตรวจจับพื้นที่เด่นของโมเดล กำหนดให้โมเดลของ Judd เป็นเส้นปะสีเขียว โมเดลของ Judd ที่ไม่มีคุณลักษณะระดับสูงเป็นเส้นปะสีแดง (JuddLM) และโมเดลที่นำเสนอเป็นเส้นทึบสีดำ (JuddLM+MEV) ซึ่งสามารถอธิบายได้ว่า ในช่วงที่ค่าความครบถ้วนมีค่าเป็นศูนย์ จะสังเกตได้ว่าโมเดลที่นำเสนอให้ค่าความถูกต้องที่สูงกว่าโมเดลของ Judd แสดงให้เห็นว่าโมเดลที่นำเสนอได้ข้อมูลที่เป็นพื้นที่เด่นแล้วมีความถูกต้องมากกว่าโมเดลของ Judd ยิ่งกว่านั้นพื้นที่ใต้กราฟของโมเดลที่นำเสนอยังมีค่ามากกว่าโมเดลของ Judd ซึ่งแสดงให้เห็นว่าโมเดลที่นำเสนอมีการทำนายผลลัพธ์ที่ถูกต้องมากกว่าในโมเดลของ Judd

เหตุผลหลักที่ทำให้โมเดลที่นำเสนอมีค่าความถูกต้องที่สูงกว่าโมเดลของ Judd คือ คุณลักษณะระดับสูงที่ใช้ในโมเดลของ Judd เช่น ตัวตรวจจับใบหน้า, บุคคล หรือ รถยนต์ นั้นมีการทำนายตำแหน่งของวัตถุที่ผิดพลาดบนชุดข้อมูล MSRA กล่าวคือ สามารถทำนายตำแหน่งของวัตถุเด่นออกจากตัวตรวจจับ ทั้งที่ไม่มีวัตถุที่ตรงกับวัตถุประสงค์ของตัวตรวจจับเหล่านั้น ซึ่งสามารถสังเกตได้จากผลกราฟของ JuddLM ในภาพที่ 10 โดยผู้วิจัยเรียกเหตุการณ์ในลักษณะนี้ว่าเป็นความผิดพลาดเชิงบวก

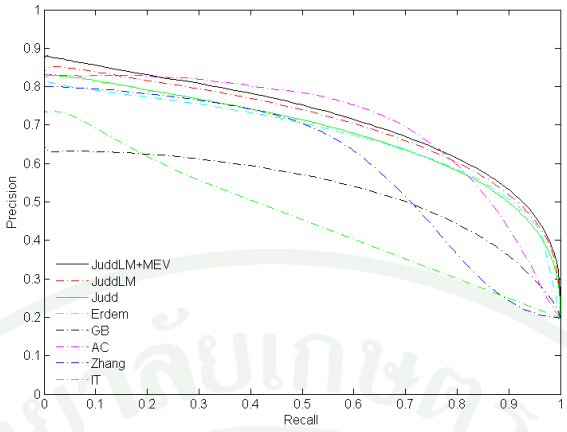


ภาพที่ 10 กราฟแสดงการเปรียบเทียบอัตราของค่าความถูกต้องกับค่าความครบถ้วนของโมเดลของ Judd โมเดลของ Judd ที่ไม่มีคุณลักษณะระดับสูง (JuddLM) กับโมเดลที่นำเสนอ (JuddLM+MEV) บนชุดข้อมูล MSRA

1.1.2 เปรียบเทียบประสิทธิภาพระหว่างโมเดลการตรวจจับวัตถุเด่นในปัจจุบัน กับโมเดลที่เสนอบนชุดข้อมูล MSRA

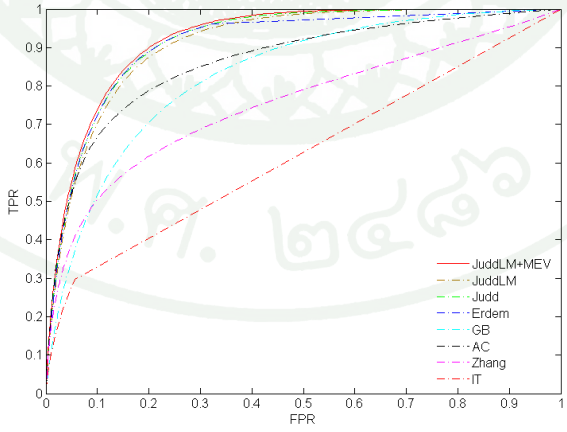
สำหรับการเปรียบเทียบระหว่างโมเดลการตรวจจับวัตถุเด่นในปัจจุบันกับโมเดลที่นำเสนอ มีวัตถุประสงค์เพื่อต้องการแสดงให้เห็นว่า โมเดลที่นำเสนอไม่เพียงแต่สามารถทำงานได้ดีในโมเดลสำหรับการทำนายแบบ fixation เท่านั้น แต่ยังสามารถที่จะนำไปประยุกต์ใช้ในการตรวจจับวัตถุเด่นได้อีกด้วย และประสิทธิภาพในการทำงานยังเป็นที่น่าพอใจเมื่อทำการเปรียบเทียบกับโมเดลการตรวจจับวัตถุเด่นในปัจจุบัน

จากภาพกราฟที่ 11 แสดงให้เห็นว่าโมเดลที่นำเสนอมีค่าความถูกต้องสูงสุดที่ ค่าความครบถ้วนเท่ากับศูนย์ ถึงแม้ว่าในช่วงที่ค่าความครบถ้วน มากกว่า 0.3 จะมีค่าความถูกต้อง ที่น้อยกว่าของกราฟโมเดล AC (Achanta *et al.*, 2009a) เพราะผลลัพธ์ในการตรวจจับวัตถุเด่นของโมเดลที่นำเสนอมีความกระจายตัวของพื้นที่ๆ เป็นวัตถุเด่นที่สูงกว่าของโมเดล AC (Achanta *et al.*, 2009a) แต่ทั้งนี้ไม่ได้ความว่าผลลัพธ์ที่ได้จากโมเดล AC (Achanta *et al.*, 2009a) จะมีความถูกต้องที่มากกว่าโมเดลที่นำเสนอ ซึ่งสามารถสังเกตได้จากช่วงกราฟที่มีค่ามากกว่า 0.8 เป็นต้นไปเพื่อที่บอกถึงประสิทธิภาพที่ชัดเจนในแต่ละโมเดลของการตรวจจับวัตถุเด่น ผู้วิจัยได้ใช้กราฟ ROC เป็นเครื่องมือใช้ในการบ่งบอกถึงพื้นที่ใต้โค้งของโมเดลตรวจจับแต่ละชนิด



ภาพที่ 11 กราฟแสดงการเปรียบเทียบอัตราของค่าความถูกต้องกับค่าความครบถ้วนของโมเดลการตรวจจับวัตถุเด่นในปัจจุบัน (Erdem, GB, AC, Zhang, IT) กับโมเดลที่นำเสนอ (JuddLM+MEV)

จากภาพกราฟที่ 12 แสดงความสัมพันธ์ระหว่างอัตราความถูกต้องเชิงบวก และอัตราความผิดพลาดเชิงบวก ของโมเดลการตรวจจับวัตถุเด่นในปัจจุบันกับโมเดลที่นำเสนอ จะสังเกตได้ว่า ในกราฟโมเดล AC มีพื้นที่ใต้กราฟ ROC ที่น้อยกว่าโมเดลที่นำเสนอ ซึ่งแสดงให้เห็นถึงโมเดลที่นำเสนอมีการทำนายตำแหน่งของวัตถุเด่นได้ถูกต้องมากกว่าและมีประสิทธิภาพที่ดีกว่าโมเดล AC (Achanta *et al.*, 2009a)



ภาพที่ 12 กราฟแสดงความสัมพันธ์ระหว่างอัตราความถูกต้องเชิงบวก (TPR) และอัตราความผิดพลาดเชิงบวก (FPR) เพื่อแสดงกราฟ ROC ของโมเดลการตรวจจับวัตถุเด่นในปัจจุบัน (Erdem, GB, AC, Zhang, IT) กับโมเดลที่นำเสนอ (JuddLM+MEV)

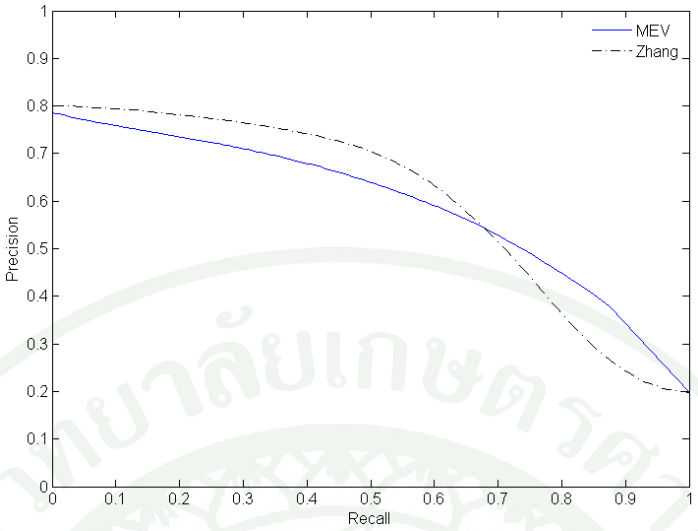
ข้อสังเกตจากเส้นกราฟในภาพที่ 12 สำหรับโมเดลของ Erdem กับโมเดลที่นำเสนอ ถึงแม้ว่าผลของพื้นที่ใต้กราฟ ROC จะมีค่าที่ใกล้เคียงกัน แต่ถ้าพิจารณาเส้นกราฟในภาพที่ 11 ที่แสดงกราฟอัตราของค่าความถูกต้องกับค่าความครบถ้วน แสดงให้เห็นว่าโมเดลที่นำเสนอได้ข้อมูลที่เป็นพื้นที่เด่นแล้วมีความถูกต้องมากกว่าโมเดลของ Erdem (Erdem and Erdem, 2013)

1.1.3 เปรียบเทียบประสิทธิระหว่างโมเดลการตรวจจับวัตถุเด่นของ Zhang กับวิธีการที่ใช้ระบุตำแหน่งของโมเดลที่นำเสนอ (MEV) บนชุดข้อมูล MSRA

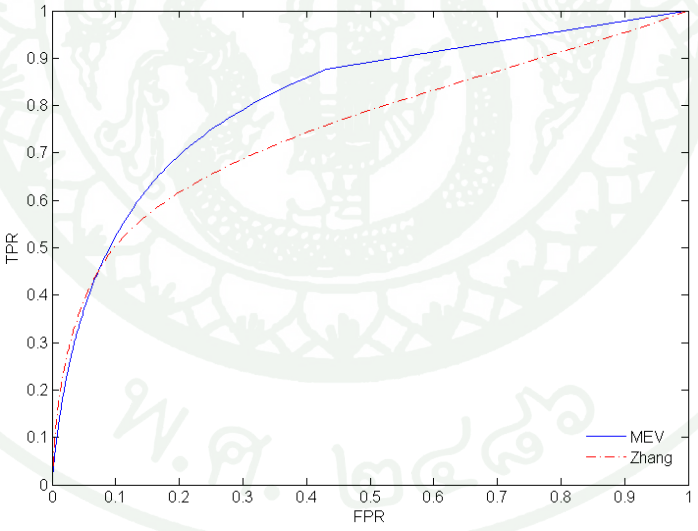
เนื่องจากโมเดลที่นำเสนอนี้มีพื้นฐานการระบุตำแหน่งของวัตถุเด่นมาจากโมเดลของ Zhang ดังนั้นจึงเป็นสิ่งสำคัญที่ต้องแสดงการเปรียบเทียบประสิทธิภาพระหว่างโมเดลของ Zhang กับวิธีการที่ใช้ระบุตำแหน่งของโมเดลที่นำเสนอ (MEV)

จากภาพกราฟที่ 13 แสดงให้เห็นว่าที่ค่าความครบถ้วนต่ำกว่า 0.65 เปอร์เซ็นต์ โมเดลที่นำเสนอจะมีค่าของความถูกต้องที่ต่ำกว่าโมเดลของ Zhang ถึงแม้ว่าจะมีค่าความครบถ้วนที่ต่ำกว่า แต่จุดมุ่งหมายหลักของการระบุตำแหน่งของวัตถุเด่นนั้นต่างจากโมเดลของ Zhang จุดมุ่งหมายหลักของโมเดลที่นำเสนอ คือ เพื่อต้องการระบุพื้นที่ที่ถูกต้องของวัตถุเด่นโดยมีข้อมูลพื้นหลังเข้ามาเกี่ยวข้องน้อยที่สุด จากการลดพื้นที่ที่ไม่เกี่ยวข้องหลังการตรวจจับในโมเดลของ Zhang ทำให้พื้นที่ของวัตถุเด่นลดลง ส่งผลให้กราฟของวิธีการที่ใช้ระบุตำแหน่งของโมเดลที่นำเสนอมีค่าน้อยกว่าโมเดลของ Zhang ในช่วงค่าความครบถ้วนที่ต่ำกว่า 0.65 เปอร์เซ็นต์ แต่ในช่วงที่ค่าความครบถ้วนมีค่ามากกว่า 0.65 เปอร์เซ็นต์ วิธีการที่ใช้ระบุตำแหน่งของโมเดลที่นำเสนอได้ให้ค่าที่สูงกว่า แสดงให้เห็นถึงผลลัพธ์ของตำแหน่งที่ถูกต้องหลังการระบุพื้นที่วัตถุเด่น

สำหรับการบรรยาย เพื่อแสดงให้เห็นถึงประสิทธิภาพของวิธีการที่ใช้ระบุตำแหน่งของโมเดลที่นำเสนอ กับโมเดลของ Zhang จะใช้กราฟแสดงความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) โดยมีผลลัพธ์ดังต่อไปนี้



ภาพที่ 13 กราฟแสดงการเปรียบเทียบอัตราของค่าความถูกต้องกับค่าความครบถ้วนของโมเดลการตรวจจับวัตถุเด่นของ Zhang กับวิธีการที่ใช้ระบุตำแหน่งของโมเดลที่นำเสนอ (MEV)



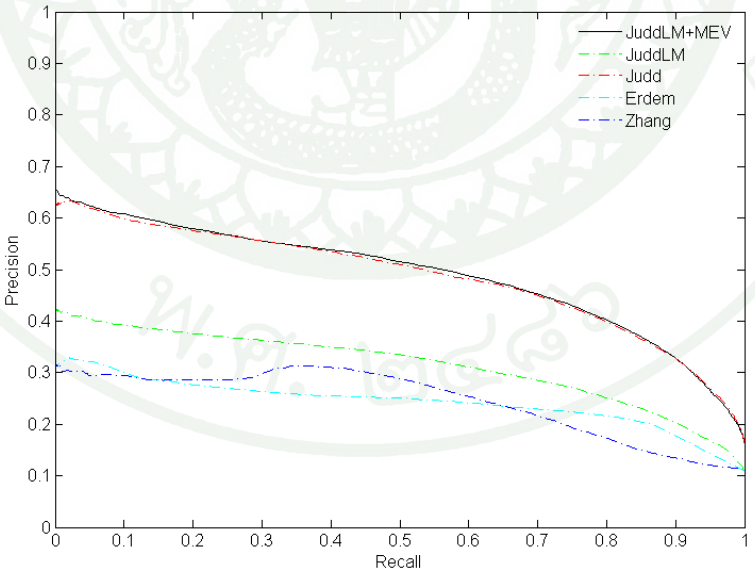
ภาพที่ 14 กราฟแสดงความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) เพื่อแสดงกราฟ ROC ระหว่างโมเดลของ Zhang กับวิธีการที่ใช้ระบุตำแหน่งของโมเดลที่นำเสนอ (MEV)

จากภาพกราฟที่ 14 แสดงความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) เพื่อแสดงกราฟ ROC ระหว่างโมเดลของ Zhang

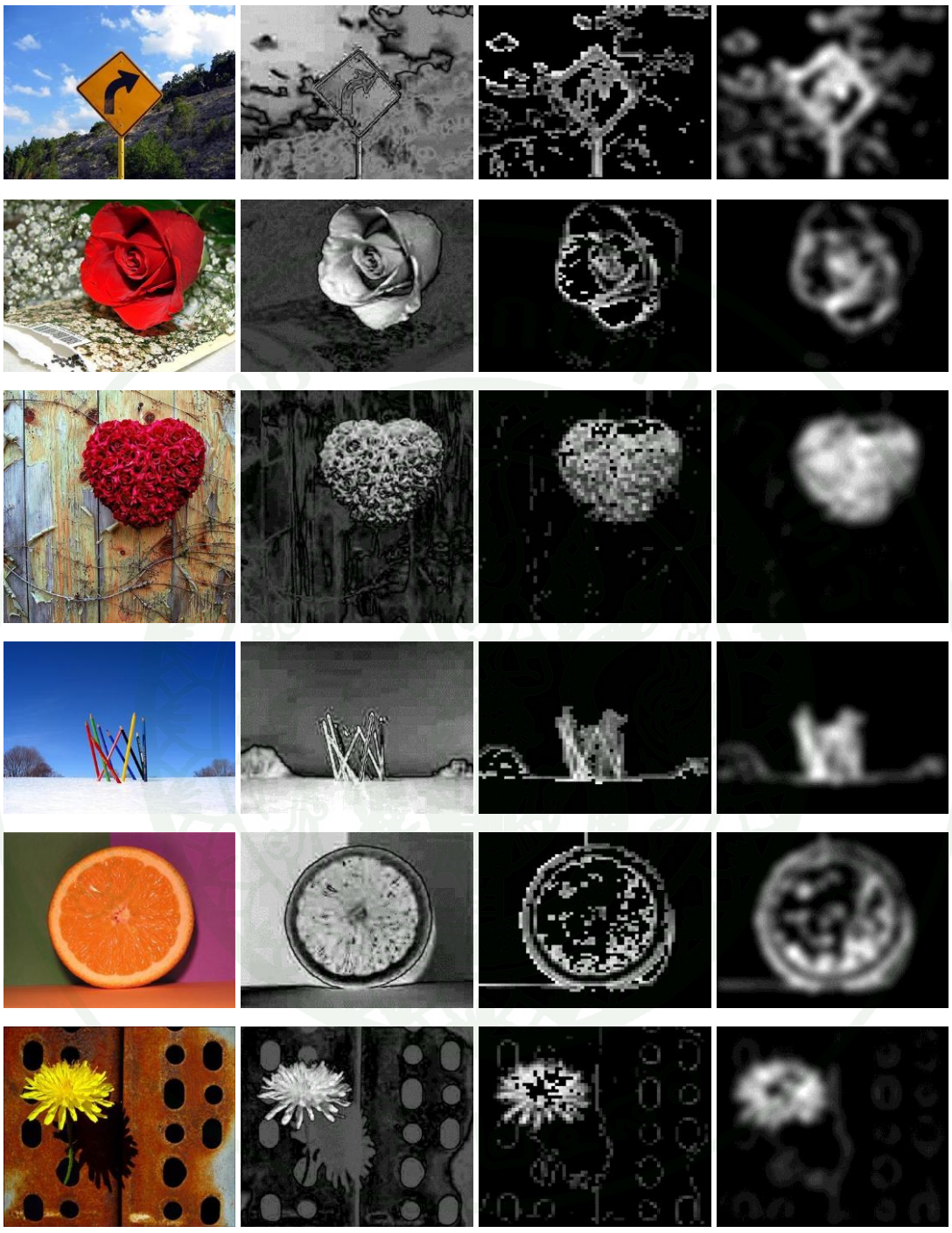
กับวิธีการที่ใช้ระบุตำแหน่งของโมเดลที่นำเสนอ (MEV) จะสังเกตได้ว่าในกราฟโมเดลของ Zhang มีพื้นที่ใต้กราฟ ROC ที่น้อยกว่าโมเดลที่นำเสนอ แสดงให้เห็นว่าวิธีการที่ใช้ระบุตำแหน่งของโมเดลที่นำเสนอ (MEV) มีการทำนายตำแหน่งของวัตถุเด่นได้ถูกต้อง และยังประกอบไปด้วยส่วนที่ไม่สนใจที่น้อยกว่าโมเดลของ Zhang จึงสรุปได้ว่าโมเดลที่นำเสนอให้ประสิทธิภาพที่ดีกว่าโมเดลของ Zhang (Hui *et al.*, 2012) จากการพิจารณากราฟ ROC

เพื่อแสดงให้เห็นว่าวิธีการที่นำเสนอ สามารถที่จะทำงานกับภาพที่มีวัตถุเด่นมากกว่าหนึ่งวัตถุ ผู้วิจัยจึงได้ทำการวัดประสิทธิภาพบนชุดข้อมูล SED2 ที่มีผลเฉลยเป็นภาพที่เก็บเฉพาะพื้นที่ของวัตถุเด่นที่สร้างจากมนุษย์ ดังแสดงในกราฟ

จากภาพกราฟที่ 15 แสดงให้เห็นว่า วิธีการที่นำเสนอนั้นมีความใกล้เคียงกับโมเดลของ Judd และดีกว่ากับโมเดลที่เหลือ ยิ่งกว่านั้นกราฟ JuddLM ได้แสดงให้เห็นว่าโมเดล MEV นั้นสามารถที่จะนำมาแทนที่คุณลักษณะระดับสูงในโมเดลของ Judd เพื่อที่จะช่วยให้การทำงานนั้นมีประสิทธิภาพคงเดิม สำหรับเหตุผลที่ผู้วิจัยไม่ได้ทำการเปรียบเทียบกับวิธีการของ IT GB และ AC เพราะไม่สามารถที่จะหาผลลัพธ์ของวิธีการเหล่านี้บนชุดข้อมูล SED2 ได้

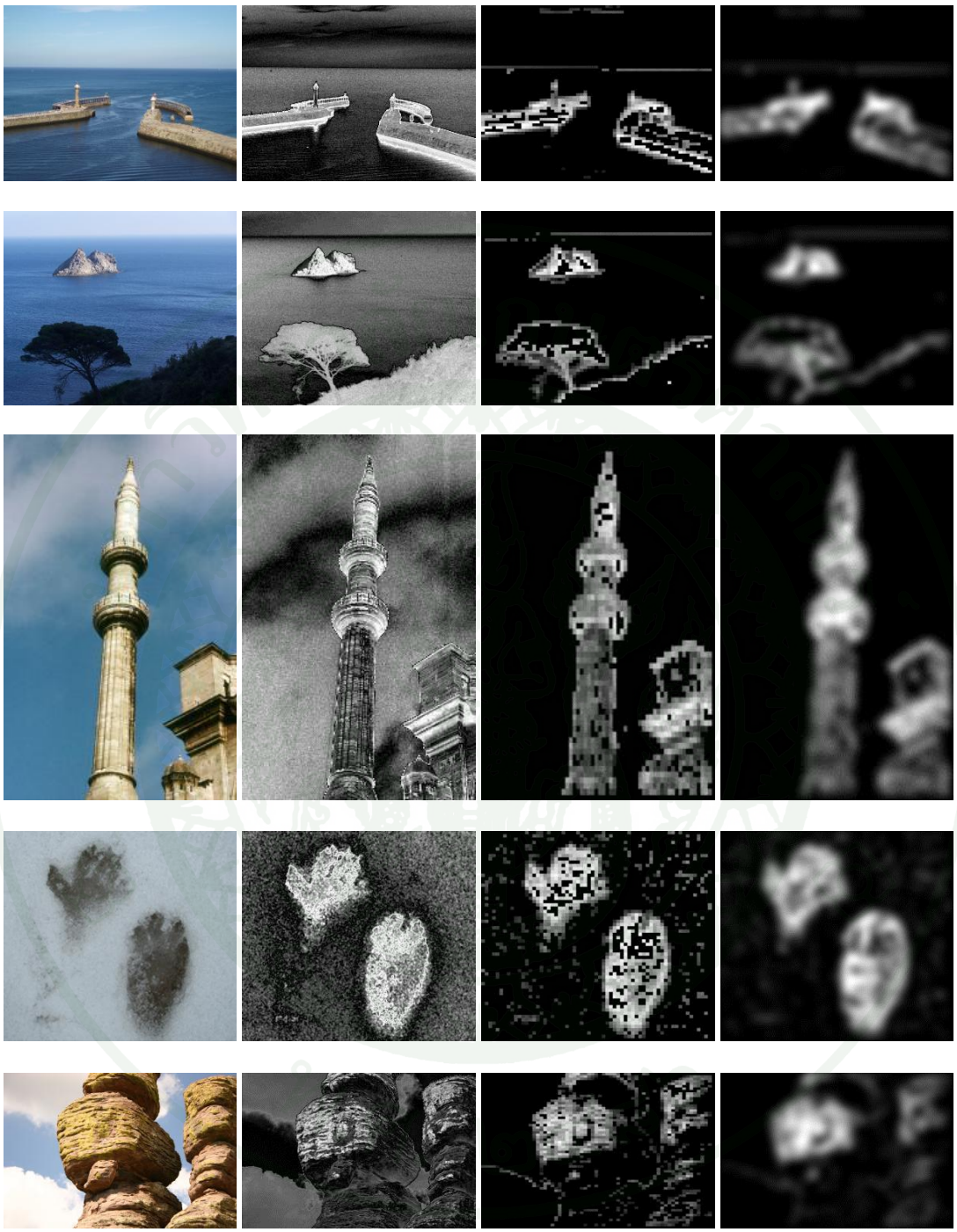


ภาพที่ 15 กราฟแสดงการเปรียบเทียบอัตราของค่าความถูกต้องกับค่าความครบถ้วนของวิธีการทั้งสามในปัจจุบัน กับวิธีการที่นำเสนอบนชุดข้อมูล SED2



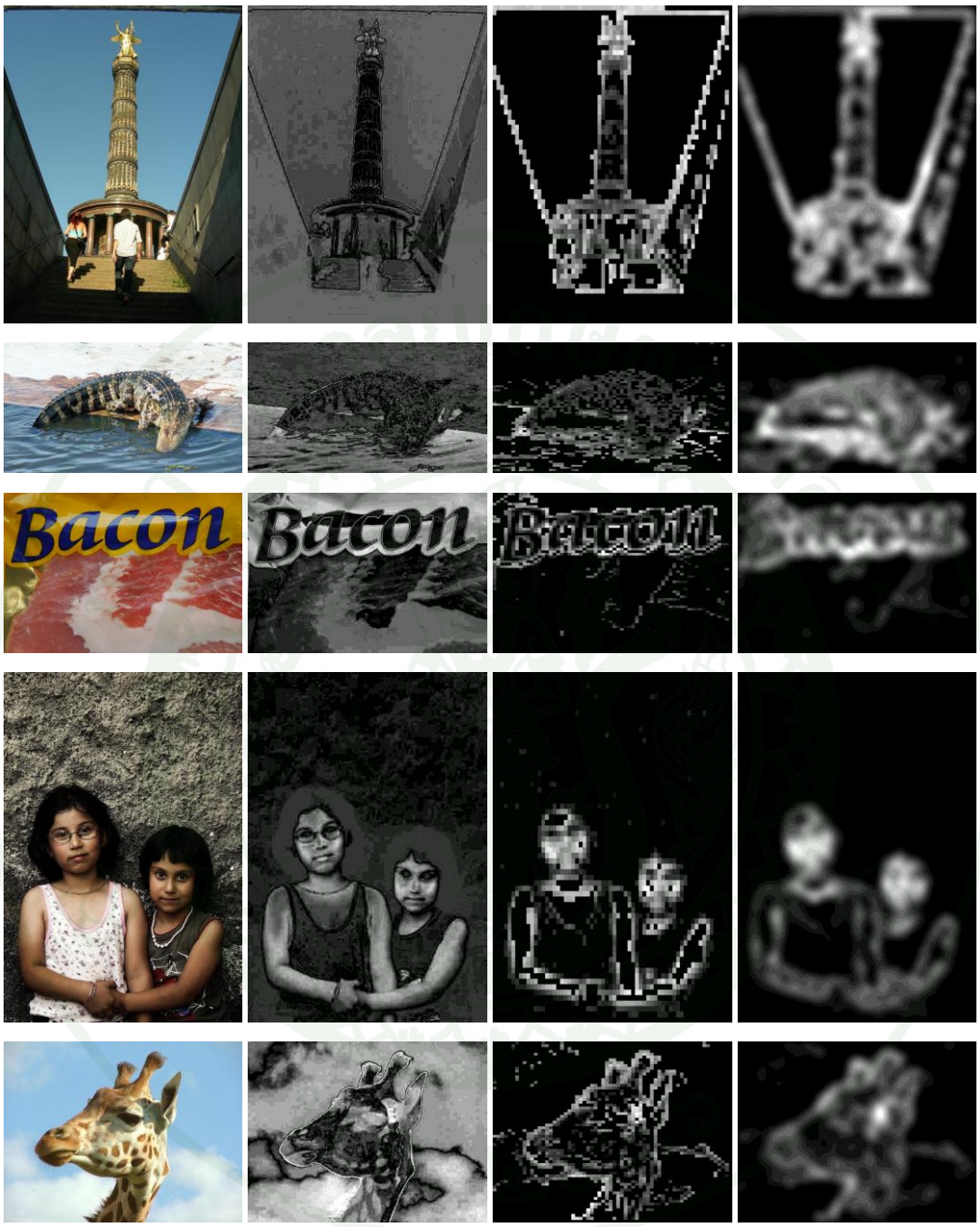
(a) (b) (c) (d)

ภาพที่ 16 ตัวอย่างของผลลัพธ์การระบุตำแหน่งของวัตถุเด่นในภาพบนชุดข้อมูล MSRA (a) ภาพนำเข้า (b) ภาพที่ได้หลังจากการสร้างในระดับชั้นที่อยู่เบื้องหลัง (c) ผลลัพธ์จากการระบุตำแหน่งขั้นสุดท้าย (d) ผลลัพธ์หลังการปรับภาพตำแหน่งของวัตถุเด่นให้เรียบและการให้ความสำคัญกับตำแหน่งกลางภาพ



(a) (b) (c) (d)

ภาพที่ 17 ตัวอย่างของผลลัพธ์การระบุตำแหน่งของวัตถุเด่นในภาพบนชุดข้อมูล SED2 (a) ภาพนำเข้า (b) ภาพที่ได้หลังจากการสร้างในระดับชั้นที่อยู่เบื้องหลัง (c) ผลลัพธ์จากการระบุตำแหน่งขั้นสุดท้าย (d) ผลลัพธ์หลังการปรับภาพตำแหน่งของวัตถุเด่นให้เรียบและการให้ความสำคัญกับตำแหน่งกลางภาพ



(a) (b) (c) (d)

ภาพที่ 18 ตัวอย่างของผลลัพธ์การระบุตำแหน่งของวัตถุเด่นในภาพบนชุดข้อมูล MIT (a) ภาพนำเข้า (b) ภาพที่ได้หลังจากการสร้างในระดับชั้นที่อยู่เบื้องหลัง (c) ผลลัพธ์จากการระบุตำแหน่งขั้นสุดท้าย (d) ผลลัพธ์หลังการปรับภาพตำแหน่งของวัตถุเด่นให้เรียบและการให้ความสำคัญกับตำแหน่งกลางภาพ

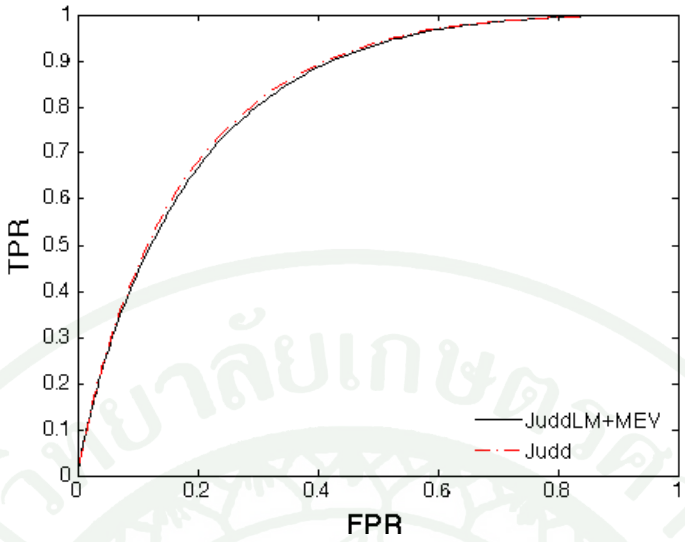
จากภาพที่ 16-18 ผู้วิจัยได้แสดงตัวอย่างของผลลัพธ์การระบุตำแหน่งของวัตถุเด่นในภาพบนชุดข้อมูล MSRA SED2 และ MIT ที่ได้ทำการแบ่งขั้นตอนตามโครงสร้างของโมเดลที่นำเสนอ เพื่อแสดงให้เห็นถึงลักษณะของภาพในชุดข้อมูลที่นำมาทดสอบ โดยสามารถสังเกตเห็นได้ว่าในบางภาพ โมเดลที่นำเสนอสามารถที่จะนำข้อมูลส่วนที่เป็นวัตถุเด่นกลับมาถึงแม้ว่าในระดับชั้นที่อยู่เบื้องหลังจะได้ส่วนที่เป็นพื้นหลัง

1.2. การทดสอบประสิทธิภาพการระบุตำแหน่งที่ตามนุษย์มองเข้าไปในภาพ

สำหรับการทดสอบกับโมเดลสำหรับการทำนายแบบ fixation นี้ ผู้วิจัยได้ใช้ฐานข้อมูล MIT ที่เสนอโดย Judd, T. และคณะ (Judd *et al.*, 2009) เนื่องจากว่าเป็นฐานข้อมูลมาตรฐานที่ใช้ในการวัดความถูกต้องของโมเดลสำหรับการทำนายแบบ fixation ในปัจจุบัน (Erdem and Erdem, 2013; Judd *et al.*, 2009) ยิ่งกว่านั้นชุดข้อมูล MIT ยังเป็นฐานข้อมูลสาธารณะที่สามารถนำมาใช้สำหรับการเปรียบเทียบประสิทธิภาพ (Zhu *et al.*, 2014) สำหรับลักษณะของชุดข้อมูล MIT จะประกอบไปด้วยภาพวิว 799 ภาพ และภาพบุคคล 288 ภาพ ที่ Judd, T. และคณะ (Judd *et al.*, 2009) ได้ทำการรวบจากฐานข้อมูลภาพสาธารณะ Flicker และ LabelMe รวมทั้งหมดจำนวน 1003 ภาพ

เพื่อที่จะทำงานบนโมเดลสำหรับการทำนายแบบ fixation ผู้วิจัยได้ทำการรวมวิธีการที่ใช้ระบุตำแหน่งของโมเดลที่นำเสนอ เข้ากับคุณลักษณะระดับล่าง และระดับกลาง ที่เสนอในโมเดลของ Judd (Judd *et al.*, 2009) โดยที่ปราศจากส่วนของคุณลักษณะระดับสูง ที่เป็นคุณลักษณะพื้นฐานในโมเดลของ Judd

สำหรับวิธีการเปรียบเทียบเพื่อวัดประสิทธิภาพของวิธีการที่นำเสนอ ผู้วิจัยได้ใช้การแสดงความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) เพื่อแสดงกราฟ ROC ที่กล่าวไว้ข้างต้น เปรียบเทียบกับโมเดลของ Judd (Judd *et al.*, 2009) เพื่อแสดงให้เห็นว่าวิธีที่นำเสนอสามารถนำไปแทนที่ในส่วนของคุณลักษณะระดับสูง ที่เป็นคุณลักษณะพื้นฐานในโมเดลของ Judd ซึ่งสามารถแสดงกราฟผลลัพธ์ดังต่อไปนี้



ภาพที่ 19 กราฟแสดงความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) เพื่อแสดงกราฟ ROC ระหว่างโมเดลของ Judd กับโมเดลที่นำเสนอ (JuddLM+MEV)

จากภาพกราฟที่ 19 แสดงความสัมพันธ์ระหว่างระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) เพื่อแสดงกราฟ ROC ระหว่างโมเดลของ Judd กับโมเดลที่นำเสนอ (JuddLM+MEV) จะสังเกตได้ว่า ประสิทธิภาพด้านความถูกต้องในการระบุตำแหน่งที่ตาคนมองเข้าไปในภาพของโมเดลที่นำเสนอนี้มีผลลัพธ์ที่ใกล้เคียงกับโมเดลของ Judd สำหรับประสิทธิภาพด้านเวลาที่ใช้ในการคำนวณ โมเดลที่นำเสนอนั้นใช้เวลาที่น้อยกว่าโมเดลของ Judd ถึง 20 เปอร์เซ็นต์ โดยทำการเทียบจากการคำนวณในชุดข้อมูล MIT จำนวน 1003 ภาพ ดังแสดงในตารางที่ 3 ทั้งนี้เนื่องมาจาก ในขั้นตอนของคุณลักษณะระดับสูง มีการใช้เวลาในการคำนวณที่มาก ไม่ว่าจะเป็นตัวตรวจจับบุคคล หรือ รถยนต์ ดังนั้นเมื่อทำการแทนที่ด้วยวิธีการที่ใช้ระบุตำแหน่งของโมเดลที่นำเสนอ (MEV) ที่มีการคำนวณที่ไม่ซับซ้อน จึงเป็นผลให้เวลาในการคำนวณลดลง

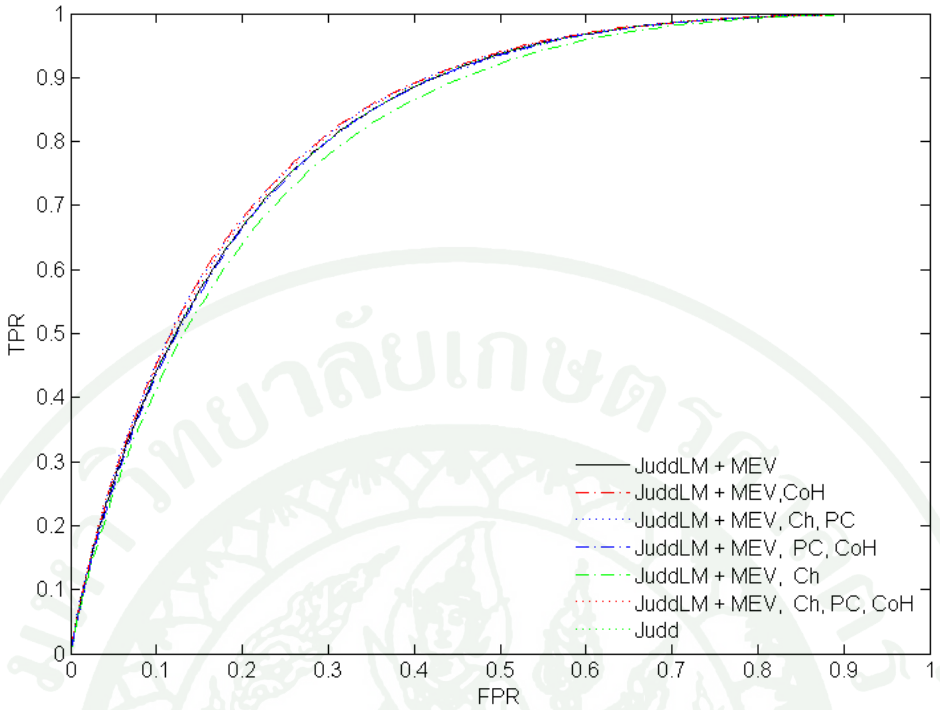
ตารางที่ 3 แสดงเวลาที่ใช้ในการคำนวณทั้งหมดบนชุดข้อมูล MIT จำนวน 1003 ภาพ

โมเดล	เวลาคำนวณ (วินาที)
Judd	21078.6
JuddLM+MEV	16817.7

เนื่องจากโมเดลของ Judd เป็นโมเดลที่ทำการรวมคุณลักษณะหลายชนิดเข้าด้วยกัน โดยคุณลักษณะที่สำคัญอย่างหนึ่งก็ คือ สี อย่างที่กล่าวในตอนต้น สามารถสังเกตได้จากค่าน้ำหนักในโมเดลของ Judd หลังการเรียนรู้ สำหรับคุณลักษณะสีที่โมเดลของ Judd ใช้ในการประมวลผลประกอบไปด้วย คุณลักษณะของสี RGB ที่แยกตามช่องสี (Channel – Ch) ความน่าจะเป็นของสี RGB ที่แยกตามช่องสี (Probability Channel - PC) และความน่าจะเป็นของสีที่ผ่านการเบลอที่ขนาดของเคอร์เนลแตกต่างกันโดยใช้วิธีการคำนวณฮิสโทแกรมแบบสามมิติ (Color Histogram - CoH)

เพื่อแสดงให้เห็นถึงประสิทธิภาพการทำงาน หลังจากที่ได้นำคุณลักษณะสีออกจากโมเดลที่นำเสนอ จะใช้การพิจารณาที่แยกออกเป็นแต่ละกรณีได้ดังนี้ โดยให้คำว่า “Ch” แทนคุณลักษณะของสี RGB ที่แยกตามช่องสี “PC” แทนคุณลักษณะความน่าจะเป็นของสี RGB ที่แยกตามช่องสี “CoH” แทนคุณลักษณะความน่าจะเป็นของสีที่ผ่านการเบลอที่ขนาดของเคอร์เนลแตกต่างกันโดยใช้วิธีการคำนวณฮิสโทแกรมแบบสามมิติ และ MEV แทนโมเดลที่ได้นำเสนอ

1. JuddLM + MEV คือ นำคุณลักษณะ Ch PC และ CoH ที่ใช้ในโมเดลของ Judd ออกจากโมเดลที่นำเสนอทั้งหมดโดยที่เหลือเฉพาะ MEV
2. JuddLM + MEV,CoH คือ นำคุณลักษณะ Ch และ PC ที่ใช้ในโมเดลของ Judd ออกจากโมเดลที่นำเสนอ โดยที่เหลือเฉพาะ MEV กับ CoH
3. JuddLM + MEV, Ch, PC คือ นำคุณลักษณะ CoH ที่ใช้ในโมเดลของ Judd ออกจากโมเดลที่นำเสนอ โดยที่เหลือเฉพาะ MEV Ch กับ PC
4. JuddLM + MEV, PC, CoH คือ นำคุณลักษณะ Ch ที่ใช้ในโมเดลของ Judd ออกจากโมเดลที่นำเสนอ โดยที่เหลือเฉพาะ MEV PC กับ CoH
5. JuddLM + MEV, Ch คือ นำคุณลักษณะ PC และ CoH ที่ใช้ในโมเดลของ Judd ออกจากโมเดลที่นำเสนอ โดยที่เหลือเฉพาะ MEV กับ Ch
6. JuddLM + MEV, Ch, PC, CoH คือ นำคุณลักษณะ Ch PC และ CoH ที่ใช้ในโมเดลของ Judd พิจารณาร่วมกับ MEV
7. โมเดลของ Judd
8. โมเดล MEV



ภาพที่ 20 กราฟแสดงความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) เพื่อแสดงกราฟ ROC หลังจากได้นำคุณลักษณะสีออกจากโมเดลที่นำเสนอที่ผู้วิจัยแบ่งออกเป็นแต่ละกรณี

จากภาพกราฟที่ 20 แสดงความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) เพื่อแสดงกราฟ ROC หลังจากได้นำคุณลักษณะสีออกจากโมเดลที่นำเสนอที่ผู้วิจัยแบ่งออกเป็นแต่ละกรณี สามารถสังเกตได้ว่าเส้นกราฟในกรณีที่ 3 นั้นมีพื้นที่ใต้กราฟที่น้อยที่สุดเมื่อเทียบกับกรณีอื่นๆ ทั้งนี้อาจเป็นเพราะ คุณลักษณะสีที่ใช้มีความผิดพลาดในการทำนายแบบ fixation แต่สิ่งที่ช่วยให้กราฟของโมเดลที่นำเสนอกลับขึ้นมาใกล้เคียงกับโมเดลของ Judd คือ คุณลักษณะ CoH ดังแสดงในตารางที่ 3 ที่ทำการสรุปการเปรียบเทียบพื้นที่ใต้กราฟ ROC สำหรับในแต่ละกรณี ดังแสดงในตารางที่ 4

ตารางที่ 4 สรุปการเปรียบเทียบพื้นที่ใต้กราฟ ROC สำหรับในแต่ละกรณี หลังนำคุณลักษณะสีเข้าร่วมพิจารณากับโมเดลที่นำเสนอ

กรณี	ชื่อ	คุณลักษณะสีที่นำเข้าพิจารณา			MEV	ค่าพื้นที่ใต้กราฟ (ROC)	เวลาเฉลี่ยในการคำนวณต่อภาพ (วินาที)
		Ch	PC	CoH			
1	JuddLM + MEV				✓	0.8221	**1.0935**
2	JuddLM + MEV,CoH			✓	✓	*0.8271*	10.4221
3	JuddLM + MEV, Ch, PC	✓	✓		✓	0.8107	1.2141
4	JuddLM + MEV, PC, CoH		✓		✓	0.8239	1.1044
5	JuddLM + MEV, Ch	✓			✓	0.8229	1.0996
6	JuddLM + MEV, Ch, PC, CoH	✓	✓	✓	✓	0.8235	10.4274
7	Judd	✓	✓	✓		**0.8285**	20.0786

จากตารางที่ 4 เครื่องหมาย ✓ ที่อยู่ในตาราง แสดงถึงการนำคุณลักษณะสีที่ใช้ในโมเดลของ Judd ทำการเข้าร่วมพิจารณากับโมเดลที่นำเสนอ ซึ่งสามารถสังเกตได้ว่า ค่าของพื้นที่ใต้กราฟ ROC หลังจากนำคุณลักษณะ Ch และ PC ออกจากโมเดลที่นำเสนอ ทำให้ค่าพื้นที่ใต้กราฟ ROC มีความใกล้เคียงกับโมเดลของ Judd ทั้งนี้อาจเนื่องมาจากคุณลักษณะสีทั้งสองให้ความผิดพลาดในการทำนายพื้นที่เด่นในภาพ เป็นผลให้พื้นที่ใต้กราฟ ROC โดยรวมมีค่าที่ลดลง สำหรับกรณีของโมเดลที่นำเสนอ มีค่าพื้นที่ใต้กราฟ ROC น้อยกว่ากรณีที่ 3 และ 4 เพราะมาจากสาเหตุในการรวมทั้ง 3 คุณลักษณะสีที่ใช้ในการประมวลผลเหมือนโมเดลของ Judd จึงสามารถสรุปได้ว่าควรที่จะนำคุณลักษณะทั้งสองออกจากโมเดล ทั้งนี้เพื่อเพิ่มประสิทธิภาพด้านความถูกต้องให้กับโมเดลที่นำเสนอ

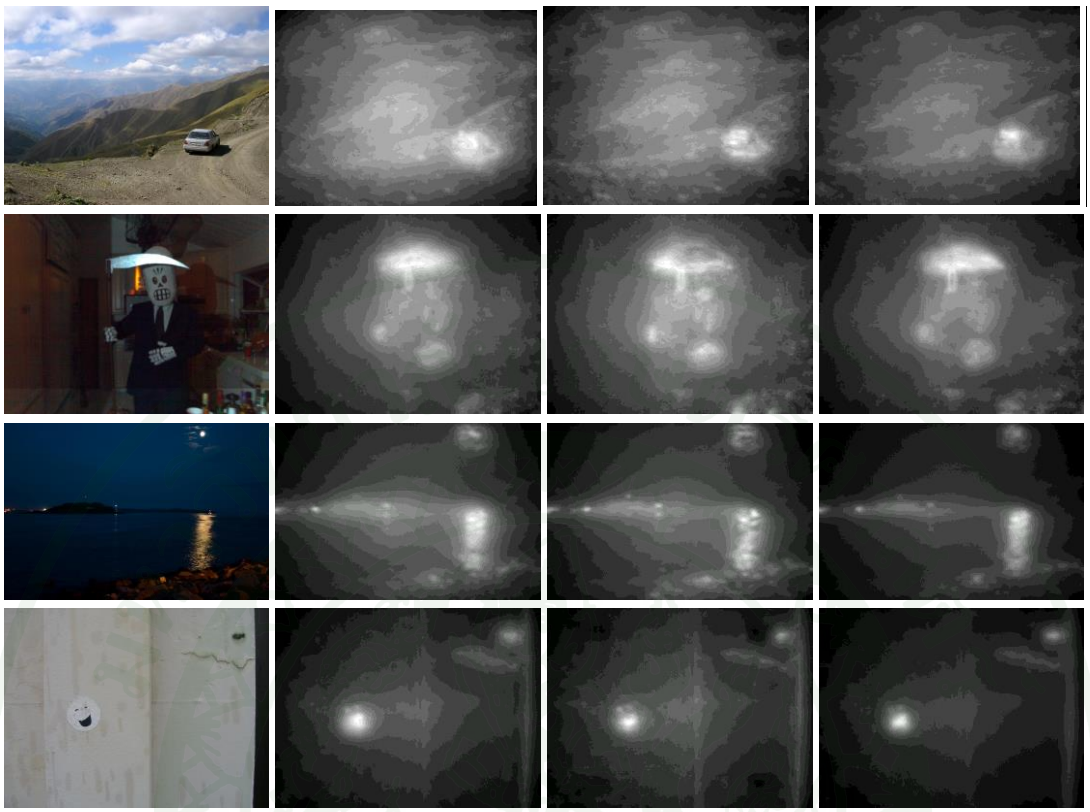
สำหรับเวลาที่ใช้ในการประมวลผล พบว่าในทุกกรณีที่ไม่มีการคำนวณ CoH ได้ใช้เวลาน้อยกว่าถึง 10 เท่า โดยเทียบจากเวลาในการคำนวณเฉลี่ยต่อภาพ เหตุผลเนื่องมาจากคุณลักษณะดังกล่าวต้องทำการเปรียบเทียบเคอร์เนลที่ขนาดแตกต่างกันถึง 4 ระดับ ได้แก่ 2×2, 4×4, 8×8, และ 16×16 จึงเป็นเหตุให้ใช้เวลาในการประมวลผลมาก

สำหรับการทดสอบเพื่อเปรียบเทียบประสิทธิภาพระหว่าง 3 โมเดล คือ (1) โมเดลของ Judd (2) โมเดลของ Judd ที่ปราศจากส่วนของคุณลักษณะระดับสูง (3) โมเดลที่นำเสนอ ผู้วิจัยได้ทำการเรียนรู้ด้วยโมเดลการเรียนรู้แบบ Support Vector Machines (SMVs) แบบเชิงเส้นเหมือนกับที่เสนอโดย Judd, T. และคณะ (Judd *et al.*, 2009) เพื่อหาค่าน้ำหนักที่เหมาะสมสำหรับแต่ละคุณลักษณะในทุกโมเดลใหม่ ด้วยชุดข้อมูลฝึกเดียวกัน ที่ได้จากการสุ่มจากชุดข้อมูล MIT จำนวน 903 ภาพ และทดสอบด้วยชุดข้อมูลทดสอบที่เหลือจากการสุ่มในชุดข้อมูลฝึกจำนวน 100 ภาพ ผู้วิจัยได้เก็บผลลัพธ์ที่ได้จากการทดสอบบนชุดข้อมูลทดสอบเป็นค่าของพื้นที่ใต้กราฟของ แต่ละภาพที่ใช้เป็นข้อมูลทดสอบ เพื่อความยุติธรรมในการเปรียบเทียบ ผู้วิจัยได้ใช้วิธีการคำนวณเพื่อหาพื้นใต้กราฟ ROC เหมือนกับวิธีที่ใช้ในงานวิจัยของ Judd (Judd *et al.*, 2009) ที่ได้กล่าวในส่วนของการตรวจเอกสาร ซึ่งผลที่ได้สามารถที่จะแสดงในตารางค่าเฉลี่ยของพื้นที่ใต้กราฟ ROC ที่ได้จากข้อมูลทดสอบได้ โดยโมเดลของ Judd ที่เอาคุณลักษณะระดับสูงออกไป เราจะให้ชื่อว่า JuddLM ดังตารางที่ 5

จากตารางที่ 5 สามารถอธิบายได้ว่า ประสิทธิภาพหลังการนำคุณลักษณะระดับสูงออกจากโมเดลของ Judd ทำให้ค่าเฉลี่ยของพื้นที่ใต้กราฟ ROC ลดลงเหมือนดังที่กล่าวไว้ในงานวิจัยของ Judd (Judd *et al.*, 2009) ที่ว่าคุณลักษณะระดับสูง มีความสำคัญที่จะช่วยระบุตำแหน่งที่ตามนุษย์มองเข้าไปในภาพ แต่ทั้งนี้เมื่อได้นำโมเดลที่นำเสนอเข้าแทนที่คุณลักษณะระดับสูง ผลปรากฏว่าค่าเฉลี่ยโดยรวมนั้นมีประสิทธิภาพที่ดีกว่าโมเดลของ Judd ที่ไม่มีคุณลักษณะระดับสูง (JuddLM) และใช้เวลาในการประมวลผลเพิ่มขึ้นเพียง 0.0763 วินาทีต่อภาพ ซึ่งแสดงให้เห็นว่าโมเดลที่นำเสนอได้มีส่วนที่ช่วยในการระบุตำแหน่งที่ตาคนมองเข้าไปในภาพ

ตารางที่ 5 แสดงค่าเฉลี่ยของค่าพื้นที่ใต้กราฟ ROC ที่ได้จากข้อมูลทดสอบบนชุดข้อมูล MIT

ลำดับ	โมเดล	เวลาเฉลี่ยในการคำนวณต่อภาพ (วินาที)	ค่าเฉลี่ย (ROC)
1	JuddLM+MEV	10.4274	0.8322
2	Judd (original)	20.0786	0.8358
3	JuddLM	10.3511	0.8315
4	JuddLM+Zhang	10.4077	0.8320



(a) (b) (c) (d)

ภาพที่ 21 แสดงผลลัพธ์การเปรียบเทียบระหว่าง Judd JuddLM และ JuddLM+MEV บนชุดข้อมูล MIT โดยที่ (a) เป็นภาพนำเข้า (b) ภาพผลลัพธ์จากโมเดล Judd (c) ภาพผลลัพธ์จากโมเดล JuddLM (d) ภาพผลลัพธ์จากโมเดลJuddLM+MEV

ถึงแม้ว่าตัวตรวจจับวัตถุหรือคุณลักษณะระดับสูง จะมีตัวตรวจจับหรือมีการเรียนรู้คุณลักษณะระดับสูง ที่ช่วยตรวจจับหา รถยนต์ บุคคลและใบหน้าบุคคล เพื่อช่วยในประสิทธิภาพด้านความถูกต้องสำหรับการทำนายแบบ fixation ในโมเดลของ Judd แต่อย่างที่กล่าวข้างต้นว่า ถ้าไม่มีวัตถุตรงตามการเรียนรู้ของตัวตรวจจับที่มีการสอนผ่านมา ก็อาจทำให้การทำนายผิดพลาดได้ ยกตัวอย่างเช่น ไม่มีวัตถุเหล่านี้ในภาพแต่มีการทำนายว่ามีเป็นต้น ดังนั้นในการศึกษาวิจัยนี้จึงทดสอบโดยการแทนที่ตัวตรวจจับวัตถุคุณลักษณะระดับสูงด้วยโมเดล MEV ที่นำเสนอ

จากภาพที่ 21 จะเห็นได้ว่าผลลัพธ์ที่ได้จากโมเดลของ Judd JuddLM และ JuddLM+MEV นั้นมีการกระจายตัวของส่วนที่ไม่สนใจที่มาก ทั้งนี้เกิดมาจากการรวมผลลัพธ์ที่ได้จากการคำนวณของหลายๆ คุณลักษณะที่โมเดลของ Judd ใช้พิจารณาเข้าด้วยกัน แต่จะสังเกตได้ว่าโมเดล

JuddLM+MEV นั้นจะมีการกระจายตัวของส่วนที่ไม่สนใจที่น้อยกว่าโมเดล JuddLM และยังมีการเน้นส่วนที่เป็นพื้นที่ของวัตถุเด่นอย่างเช่นเดียวกับโมเดลของ Judd ทั้งนี้เนื่องมาจากโมเดล MEV ที่นำไปประยุกต์ใช้นั้นมีการทำงานอย่างเช่นเดียวกับการทำงานในส่วนของคุณลักษณะระดับสูงในโมเดล Judd จึงทำให้ผลลัพธ์ที่ได้มีส่วนของสิ่งที่ไม่สนใจน้อยกว่าโมเดล JuddLM และยิ่งกว่านั้นยังได้เข้าไปช่วยในการเน้นส่วนที่เป็นพื้นที่ของวัตถุเด่นในภาพให้มีความถูกต้องมากกว่าโมเดล JuddLM อีกด้วย

2. การทดลองประยุกต์ใช้กับคุณลักษณะอื่นๆ รวมกับคุณลักษณะความเปรียบต่างของสี

ไม่เพียงแต่คุณลักษณะของสีเท่านั้น ที่งานวิจัยในปัจจุบัน ใช้ในการพิจารณาวัตถุเด่นในภาพ ยังมีคุณลักษณะอื่นๆ ไม่ว่าจะเป็นคุณลักษณะค่าความเข้มแสง ค่าโทนสี ค่าความอิ่มตัวของสี และ ค่าความสว่าง เป็นต้น โดยวิธีการคำนวณค่าความเข้มแสง สามารถหาได้จากค่าเฉลี่ยของค่าสี RGB ดังสมการที่ 11

$$I = \frac{R + G + B}{3} \tag{11}$$

สำหรับการคำนวณหาค่าโทนสี และค่าความอิ่มตัวของสี ของระบบสี HSV ผู้วิจัยได้ใช้ฟังก์ชันสูตรการแปลงระบบสี RBG เป็น HSV (rgb2hsv) ที่มีอยู่ในโปรแกรม MATLAB 2013b โดยมีขั้นตอนในฟังก์ชันสูตรการแปลงดังนี้

1. ทำการปรับช่วงกว้างของแต่ละช่องสีให้อยู่ในช่วง 0 ถึง 1 (R', G', B')
2. ทำการหาค่ามากสุด (C_{max}) น้อยสุด (C_{min}) และผลต่างระหว่างค่ามากสุดและน้อยสุด (Δ) ของช่องสีที่ได้ทำการปรับช่วงกว้างแล้ว
3. คำนวณหาค่าโทนสี (H) ดังสมการที่ 12

$$H = \begin{cases} 60° \times \left(\frac{G' - B'}{\Delta} mod 6\right) & if\ C_{max} = R' \\ 60° \times \left(\frac{B' - R'}{\Delta} + 2\right) & if\ C_{max} = G' \\ 60° \times \left(\frac{R' - G'}{\Delta} + 4\right) & if\ C_{max} = B' \end{cases} \tag{12}$$

4. คำนวณหาค่าความอิ่มตัวของสี (S) ดังสมการที่ 13

$$S = \begin{cases} 0 & if\ \Delta = 0 \\ \dfrac{\Delta}{C_{max}} & if\ \Delta \neq 0 \end{cases} \tag{13}$$

และสุดท้ายวิธีการหาค่าความสว่างของระบบสี CIE LAB ผู้วิจัยได้ใช้ฟังก์ชันสูตรการแปลงระบบสี RBG เป็น CIE LAB (makecform('srgb2lab')) ที่มีอยู่ในโปรแกรม MATLAB 2013b โดยมีขั้นตอนในฟังก์ชันสูตรการแปลงดังนี้

1. ทำการแปลงภาพจากระบบสี RGB เป็น XYZ จะมีเมตริกสำหรับการแปลงดังนี้ คือ

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.813 & 0.011 \\ 0.000 & 0.010 & 0.990 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{14}$$

2. จากนั้นทำการแปลงจากระบบสี XYZ เป็น CIE LAB ซึ่งต้องมีการอ้างจุดสีขาว (WhitePoint) บนแกน X Y Z (X_r, Y_r, Z_r) ในโปรแกรม MABLAB 2013b สำหรับสมการการแปลงภาพจากระบบสี XYZ เป็น CIE LAB สามารถแสดงได้ดังนี้

$$\begin{aligned} L &= 166 f_y - 16 \\ a &= 500\left(f_x - f_y\right) \\ b &= 200\left(f_y - f_z\right) \end{aligned} \tag{15}$$

โดยที่

$$\begin{aligned} f_x &= \begin{cases} \sqrt[3]{x_r} & if\ x_r > \varepsilon \\ \frac{kx_r + 16}{116} & if\ x_r \leq \varepsilon \end{cases} \\ f_x &= \begin{cases} \sqrt[3]{x_r} & if\ y_r > \varepsilon \\ \frac{ky_r + 16}{116} & if\ y_r \leq \varepsilon \end{cases} \\ f_x &= \begin{cases} \sqrt[3]{x_r} & if\ z_r > \varepsilon \\ \frac{kz_r + 16}{116} & if\ z_r \leq \varepsilon \end{cases} \end{aligned} \tag{16}$$

$$x_r = \frac{X}{X_r}$$
$$y_r = \frac{Y}{Y_r} \quad (17)$$
$$z_r = \frac{Z}{Z_r}$$

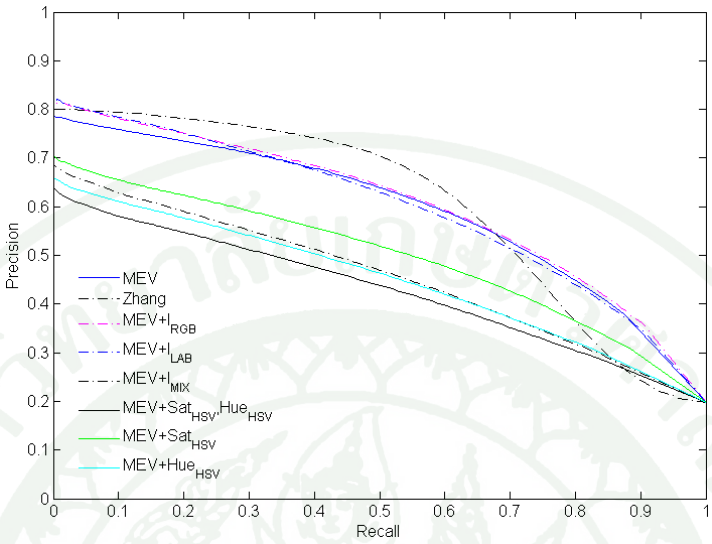
$$\varepsilon = \begin{cases} 0.008856 & \text{ค่าที่ใช้จริง} \\ \frac{216}{24389} & \text{ค่าตามนิยามของ } CIE \end{cases} \quad (18)$$
$$k = \begin{cases} 903.3 & \text{ค่าที่ใช้จริง} \\ \frac{24389}{27} & \text{ค่าตามนิยามของ } CIE \end{cases}$$

สำหรับการทดลองนี้ ผู้วิจัยต้องการแสดงให้เห็นว่า เมื่อทำการรวมคุณลักษณะดังกล่าว ที่ได้หลังจากการคำนวณในระบบสีต่างกัน จะมีผลอย่างไรต่อความถูกต้องของโมเดลที่นำเสนอ ชุดข้อมูลที่นำมาทดสอบมี 3 ชุด ได้แก่ MSRA MIT และ SED2 โดยมีการอ้างถึงชื่อของโมเดลที่ใช้ในการเปรียบเทียบดังตารางที่ 6

ตารางที่ 6 แสดงรายละเอียดของโมเดลสำหรับการทดลองประยุกต์ใช้กับคุณลักษณะอื่นๆ รวมกับคุณลักษณะความเปรียบต่างของสี

ลำดับ	ชื่อ	อธิบาย
1	MEV	วิธีการของคุณลักษณะที่นำเสนอ
2	Zhang	วิธีการของ Zhang (Hui *et al.*, 2012) ที่ได้นำเสนอ
3	MEV+I_{RGB}	การนำคุณลักษณะความเข้มของภาพนำเข้า ที่คำนวณจากระบบสี RGB ร่วมพิจารณากับ MEV
4	MEV+I_{LAB}	การนำคุณลักษณะในส่วนค่าความสว่าง ของระบบสี CIE LAB ร่วมพิจารณากับ MEV
5	MEV+I_{MIX}	การรวมคุณลักษณะทั้งหมดที่ใช้ไม่ว่าจะเป็น Hue กับ Saturation ของระบบสี HSV ค่าความสว่าง ของระบบสี CIE LAB และ ความเข้มของภาพนำเข้า ที่คำนวณจากระบบสี RGB ร่วมพิจารณากับ MEV
6	MEV+Sat_{HSV} ,Hue_{HSV}	การนำคุณลักษณะ โทนสี กับ ความอิ่มตัวของสี ของระบบสี HSV ร่วมพิจารณากับ MEV
7	MEV+Sat_{HSV}	การนำคุณลักษณะความอิ่มตัวของสี ของระบบสี HSV ร่วมพิจารณากับ MEV
8	MEV+Hue_{HSV}	การนำคุณลักษณะโทนสีของระบบสี HSV ร่วมพิจารณากับ MEV

2.1 การทดลองบนชุดข้อมูล MSRA

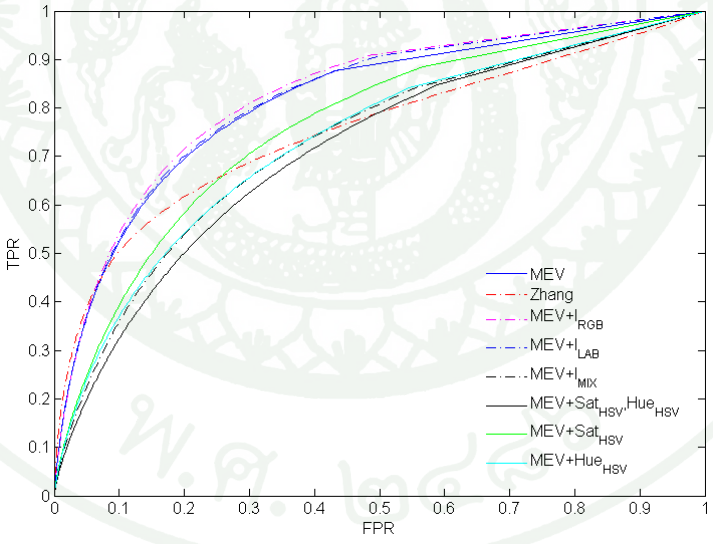


ภาพที่ 22 กราฟแสดงการเปรียบเทียบอัตราของค่าความถูกต้องกับค่าความครบถ้วนของโมเดลการตรวจจับวัตถุเด่นของ Zhang และวิธีการที่ใช้ระบุตำแหน่งของโมเดล MEV กับวิธีหลังจากทำการรวมคุณลักษณะเพิ่มเติมเข้ากับ MEV บนชุดข้อมูล MSRA

จากภาพกราฟที่ 22 แสดงให้เห็นว่าการนำคุณลักษณะความเข้มของภาพนำเข้า ที่คำนวณจากระบบสี RGB มาร่วมพิจารณากับ MEV กับการนำคุณลักษณะในส่วนค่าความสว่าง ที่คำนวณจากระบบสี CIE LAB ร่วมพิจารณากับ MEV ให้ค่าความถูกต้องที่สูงเมื่อเทียบกับโมเดล MEV ที่แสดงในภาพกราฟที่ 22 ซึ่งหมายความว่า ถ้ามีการรวมคุณลักษณะความเข้มของภาพนำเข้าที่คำนวณจากระบบสี RGB หรือ คุณลักษณะในส่วนค่าความสว่างของระบบสี CIE LAB เข้ากับคุณลักษณะของโมเดล MEV จะสามารถเพิ่มประสิทธิภาพความถูกต้องในการตรวจจับส่วนที่เป็นวัตถุเด่นในภาพ โดยมีเหตุผลเนื่องมาจาก ในบางครั้ง ตำแหน่งของวัตถุเด่น นั้นไม่ได้อิงกับช่วงสีใดช่วงสีหนึ่งใน R G B และ Y แต่เป็นค่าเฉลี่ยของช่วงสีทั้งสาม ซึ่งค่าดังกล่าวนี้สามารถหาได้จากค่าเฉลี่ยของช่วงสี RGB หรือเรียกอีกอย่างว่าค่าความเข้มแสงในระบบสี RGB ส่วนค่าความสว่างของระบบสี CIE LAB นั้นจะเกี่ยวกับความสว่างและความมืดของสี ที่มีลักษณะการคำนวณที่คล้ายกับค่าความเข้มแสงดังกล่าว ซึ่งถูกกำหนดค่าเป็นเปอร์เซ็นต์จาก 0% (สีดำ) ถึง 100% (สีขาว) ยิ่งมีเปอร์เซ็นต์มากจะทำให้สีนั้นสว่างมากขึ้น หรือกล่าวได้ว่าค่าเฉลี่ยของสีที่มากด้วย

ถือแม้ว่า ในปัจจุบันจะมีการใช้ค่าโทนสี กับค่าความอิ่มตัวของสี ในระบบสี HSV เป็นคุณลักษณะสำคัญในการพิจารณาวัตถุเด่นในภาพ แต่การนำมาประยุกต์ใช้สำหรับโมเดล MEV นั้นอาจไม่เหมาะสมสำหรับการนำมาร่วมพิจารณาเพื่อทำการตรวจจับวัตถุเด่น เหตุผลเนื่องมาจากค่าความแปรปรวนที่ได้จากคุณลักษณะของ HSV นั้นมีค่ามาก เมื่อทำการเทียบกับค่าความแปรปรวนของคุณลักษณะ R G B และ Y ที่ทำการพิจารณาร่วมกัน ดังนั้นจึงทำให้ผลลัพธ์ของการทำนาย เอนเอียงไปทางคุณลักษณะของ HSV เป็นส่วนใหญ่ จึงทำให้มีค่าความถูกต้องที่น้อยเมื่อทำการเทียบกับโมเดลที่ไม่ได้ทำการรวมคุณลักษณะของ HSV เข้ามาพิจารณา

เพื่อแสดงให้เห็นว่าคุณลักษณะความเข้มที่คำนวณจากระบบสี RGB และค่าความสว่างที่คำนวณจากระบบสี CIE LAB นั้นมีค่าความถูกต้องที่สูงกว่าโมเดล MEV ผู้วิจัยจึงทำการแสดงกราฟของพื้นที่ใต้โค้งดังภาพกราฟที่ 23 เพื่อแสดงให้เห็นถึงความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวกและอัตราค่าความผิดพลาดเชิงบวก ซึ่งจะเห็นได้ว่าพื้นที่ใต้กราฟของ MEV+I_{RGB} นั้นมีพื้นที่ใต้กราฟที่มากที่สุดดังแสดงในภาพกราฟที่ 23



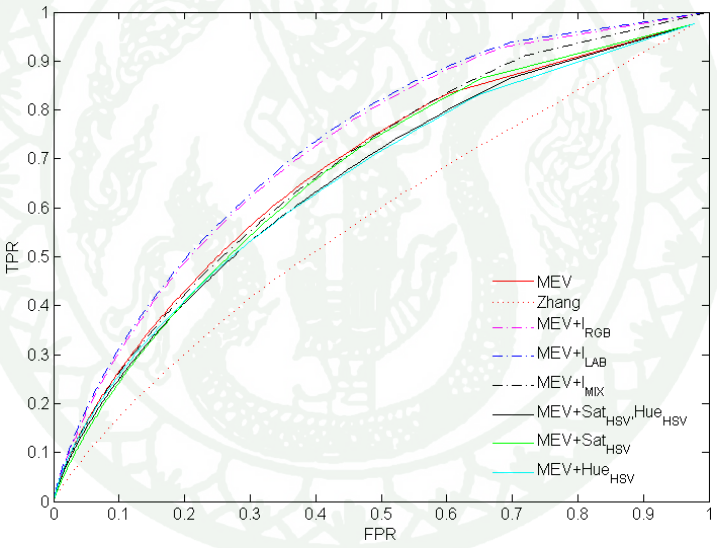
ภาพที่ 23 กราฟแสดงความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) เพื่อแสดงกราฟ ROC ของโมเดลการตรวจจับวัตถุเด่นของ Zhang และวิธีการที่ใช้ระบุตำแหน่งของโมเดล MEV กับวิธีหลังจากทำการรวมคุณลักษณะเพิ่มเติมเข้ากับ MEV บนชุดข้อมูล MSRA

(a) (b) (c) (d) (e) (f) (g) (h) (i)

ภาพที่ 24 แสดงผลลัพธ์หลังจากทำการรวมคุณลักษณะเข้ากับ MEV บนชุดข้อมูล MSRA ดังนี้ (a) ภาพนำเข้า (b) ผลลัพธ์ของ MEV (c) ผลลัพธ์ของ Zhang (d) ผลลัพธ์ของ MEV รวมกับค่า Hue และ Saturation ในระบบสี HSV (e) ผลลัพธ์ของ MEV รวมกับค่าความสว่าง ในระบบสี CIE LAB (f) ผลลัพธ์ของ MEV รวมกับค่าเฉลี่ยระหว่างสีแดง เขียว น้ำเงิน ในระบบสี RGB และ (g) ผลลัพธ์ของ MEV รวมกับค่า Hue และ Saturation ในระบบสี HIS กับ ค่าความสว่าง ในระบบสี CIE LAB และค่าเฉลี่ยระหว่างสีแดง เขียว น้ำเงิน ในระบบสี RGB เข้าด้วยกัน (h) ผลลัพธ์ของ MEV รวมกับค่า Hue ในระบบสี HSV (i) ผลลัพธ์ของ MEV รวมกับค่า Saturation ในระบบสี HSV

สำหรับตัวอย่างภาพผลลัพธ์ที่ได้จากการรวมคุณลักษณะเข้ากับ MEV บนชุดข้อมูล MSRA ในแต่ละวิธีการ ดังภาพที่ 24 จะสังเกตได้ว่า ผลลัพธ์ที่ได้จากการรวมกับค่าโทนสี และค่าความอิ่มตัวของสี นั้นจะมีรายละเอียดของภาพที่ไม่ได้เป็นสิ่งเด่นปรากฏอยู่ในผลลัพธ์การทำนายมากกว่า การรวมคุณลักษณะค่า ความเข้ม หรือค่าความสว่าง ของระบบสี RGB และ CIE LAB ตามลำดับ ซึ่งสิ่งนี้เป็นเหตุผลที่ว่าทำไมผลกราฟของ MEV+Sat_{HSV} ,Hue_{HSV} MEV+Sat_{HSV} MEV+Hue_{HSV} และ MEV+I_{MIX} ที่แสดงในภาพกราฟที่ 23 จึงมีค่าที่ต่ำกว่ากราฟ MEV+I_{RGB} กับ MEV+I_{LAB} เพื่อยืนยันว่าคุณสมบัติค่าความเข้มแสง กับค่าความสว่างนั้นช่วยให้ประสิทธิภาพของโมเดล MEV นั้นจริง ผู้วิจัยจึงได้ทำการทดลองบนชุดข้อมูล MIT

2.2 การทดลองบนชุดข้อมูล MIT



ภาพที่ 25 กราฟแสดงความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) เพื่อแสดงกราฟ ROC ของโมเดลการตรวจจับวัตถุเด่นของ Zhang และวิธีการที่ใช้ระบุตำแหน่งของโมเดล MEV กับวิธีหลังจากทำการรวมคุณลักษณะเพิ่มเติมเข้ากับ MEV บนชุดข้อมูล MIT

เนื่องจากผลเฉลยในชุดข้อมูล MIT เป็นแบบ fixation point จึงไม่สามารถที่จะทำการเปรียบเทียบอัตราค่าความถูกต้องกับค่าความครบถ้วนของโมเดลได้ ผู้วิจัยจึงได้ใช้กราฟแสดงความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก และอัตราค่าความผิดพลาดเชิงบวก ในการเปรียบเทียบผลลัพธ์ ซึ่งจากผลกราฟที่แสดงในภาพกราฟที่ 25 จะเห็นได้ว่าโมเดลที่มีการเพิ่ม

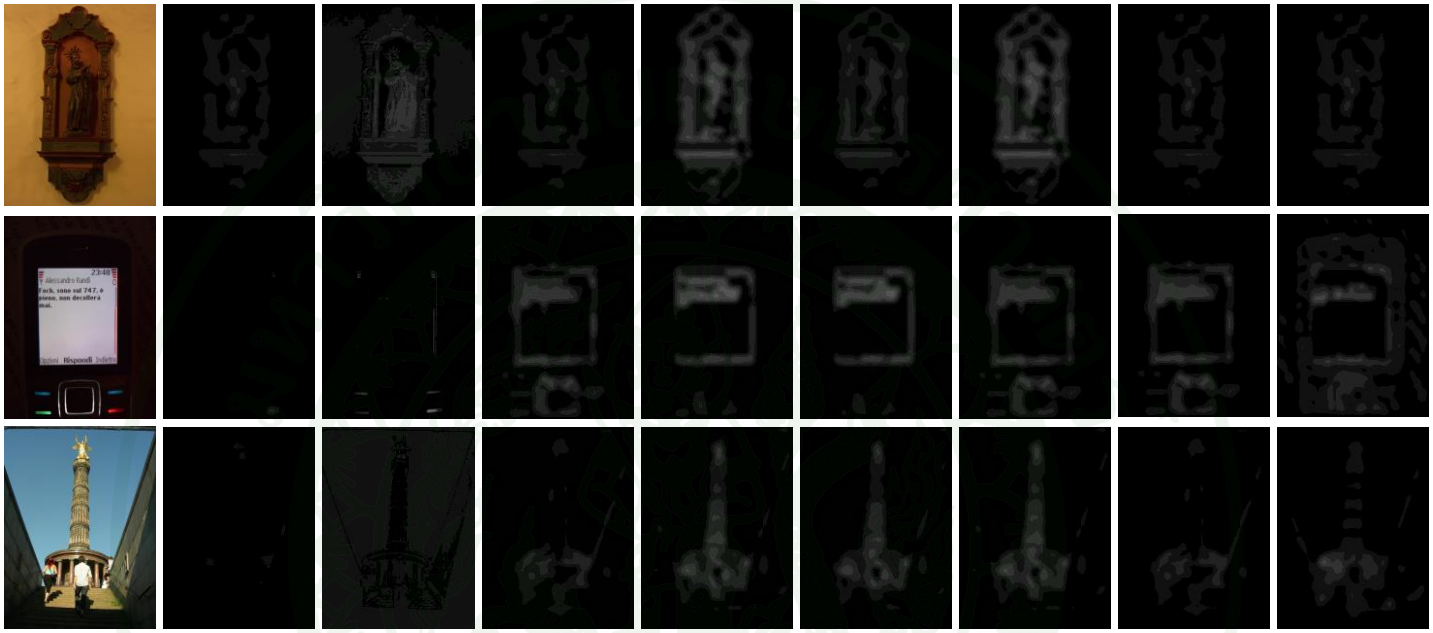
คุณลักษณะค่าความเข้มแสง ค่าความสว่างในโมเดล MEV มีพื้นที่ใต้กราฟที่สูง อย่างเช่นเดียวกับผลการเปรียบเทียบในชุดข้อมูล MSRA

โดยสามารถกล่าวได้ว่าโมเดลมีการเพิ่มคุณลักษณะค่าความเข้มแสง ค่าความสว่างในโมเดล MEV สามารถที่จะทำนายพื้นที่ที่เป็นวัตถุเด่นได้ถูกต้องและมีส่วนที่ไม่สนใจน้อยกว่าโมเดลที่ทำการเปรียบเทียบ และมีความเหมาะสมสำหรับการทำนายแบบ fixation point กว่าที่ไม่มีการเพิ่มคุณลักษณะค่าความเข้มแสง หรือค่าความสว่างเข้ามาพิจารณา ซึ่งภาพตัวอย่างของแต่ละวิธีการหลังเพิ่มคุณลักษณะสามารถแสดงได้ดังภาพที่ 26

ทั้งนี้ผู้วิจัยได้ทำการเปรียบเทียบระหว่างโมเดลที่นำเสนอในตอนต้นกับโมเดลที่ทำการรวมคุณลักษณะความเข้มที่คำนวณจากระบบสี RGB ที่รวมเข้ากับโมเดลสำหรับการทำนายแบบ fixation ดังแสดงในตารางที่ 7

ตารางที่ 7 แสดงการเปรียบเทียบโมเดลที่ทำการรวมคุณลักษณะความเข้มกับโมเดลที่นำเสนอด้วยค่าเฉลี่ยของค่าพื้นที่ใต้กราฟ ROC ที่ได้จากข้อมูลทดสอบ

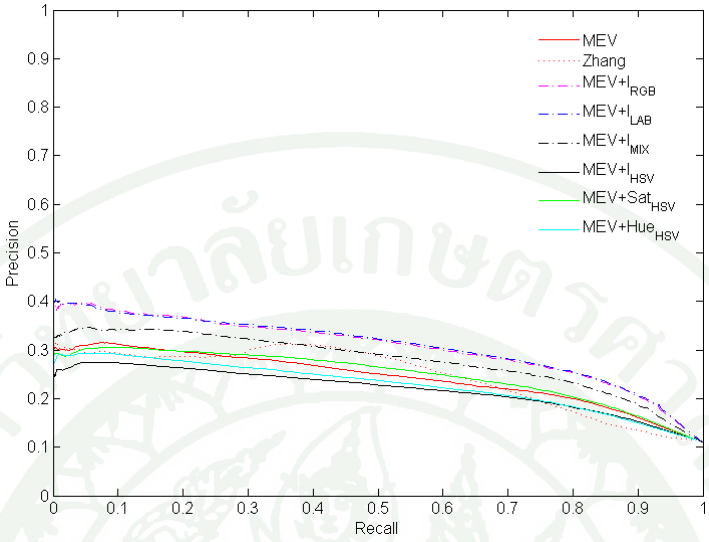
ลำดับ	โมเดล	เวลาเฉลี่ยในการคำนวณต่อภาพ (วินาที)	ค่าเฉลี่ย (ROC)
1	JuddLM+MEV	10.4274	0.8322
2	Judd (original)	20.0786	0.8358
3	JuddLM+MEV+I_{RGB}	10.4359	0.8324



(a) (b) (c) (d) (e) (f) (g) (h) (i)

ภาพที่ 26 แสดงผลลัทธ์หลังจากทำการรวมคุณลักษณะเพิ่มเติมเข้ากับ MEV บนชุดข้อมูล MIT ดังนี้ (a) ภาพนำเข้า (b) ผลลัพธ์ของ MEV (c) ผลลัพธ์ของ Zhang (d) ผลลัพธ์ของ MEV รวมกับค่า Hue และ Saturation ในระบบสี HSV (e) ผลลัพธ์ของ MEV รวมกับค่าความสว่าง ในระบบสี CIE LAB (f) ผลลัพธ์ของ MEV รวมกับค่าเฉลี่ยระหว่างสีแดง เขียว น้ำเงิน ในระบบสี RGB และ (g) ผลลัพธ์ของ MEV รวมกับค่า Hue และ Saturation ในระบบสี HIS กับ ค่าความสว่าง ในระบบสี CIE LAB และค่าเฉลี่ยระหว่างสีแดง เขียว น้ำเงิน ในระบบสี RGB เข้าด้วยกัน (h) ผลลัพธ์ของ MEV รวมกับค่า Hue ในระบบสี HSV (i) ผลลัพธ์ของ MEV รวมกับค่า Saturation ในระบบสี HSV

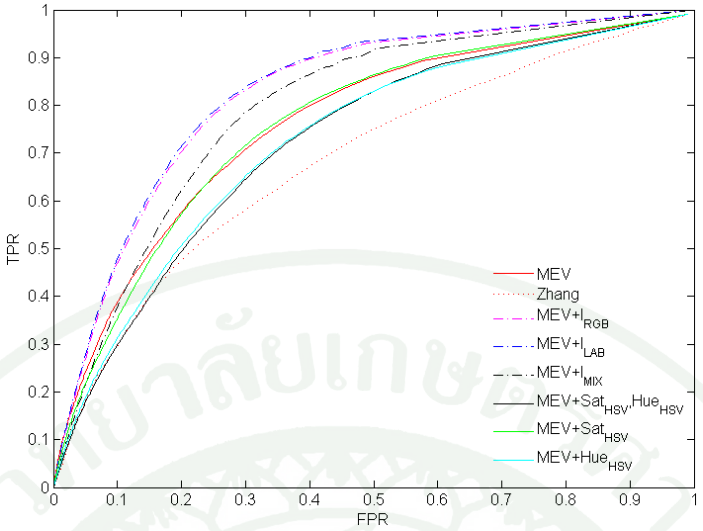
2.3 การทดลองบนชุดข้อมูล SED2



ภาพที่ 27 กราฟแสดงการเปรียบเทียบอัตราของค่าความถูกต้องกับค่าความครบถ้วนของโมเดลการตรวจจับวัตถุเด่นของ Zhang และวิธีการที่ใช้ระบุตำแหน่งของโมเดล MEV กับวิธีหลังจากทำการรวมคุณลักษณะเพิ่มเติมเข้ากับ MEV บนชุดข้อมูล SED2

สำหรับภาพกราฟที่ 27 ได้แสดงให้เห็นว่าการนำคุณลักษณะความเข้มของภาพนำเข้าที่คำนวณจากระบบสี RGB ร่วมพิจารณากับ MEV กับการนำคุณลักษณะในส่วนค่าความสว่าง ที่คำนวณจากระบบสี CIE LAB ร่วมพิจารณากับ MEV นั้นยังคงให้ค่าความถูกต้องกับค่าความครบถ้วนที่สูงอย่างเช่นเดียวกับในการทดลอง ที่ได้ทำการรวมคุณลักษณะความเข้มของภาพ กับค่าความสว่าง บนชุดข้อมูล MSRA และ MIT

สำหรับกราฟที่แสดงความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก และอัตราค่าความผิดพลาดเชิงบวก ของการทดลองบนชุดข้อมูล SED2 ยังเป็นสิ่งที่ตอกย้ำโมเดลที่ทำการเพิ่มคุณลักษณะดังกล่าวได้ให้ประสิทธิภาพที่ดีกว่า MEV ดังภาพกราฟที่ 28 โดยมีเหตุผลเช่นเดียวกับการทดลองการรวมคุณลักษณะความเข้มของภาพ กับค่าความสว่าง บนชุดข้อมูล MSRA และ MIT ที่ทำได้ผลการทดลองออกมาสูงกว่า MEV



ภาพที่ 28 กราฟแสดงความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) เพื่อแสดงกราฟ ROC ของโมเดลการตรวจจับวัตถุเด่นของ Zhang และวิธีการที่ใช้ระบุตำแหน่งของโมเดล MEV กับวิธีหลังจากทำการรวมคุณลักษณะเพิ่มเติมเข้ากับ MEV

จากการทดลองการเพิ่มคุณลักษณะ ค่าความเข้มแสง ค่าความสว่าง ค่าโทนสี ค่าความอิ่มตัวของสี หรือการเพิ่มคุณลักษณะทั้งมด เข้ากับโมเดล MEV และทำการทดสอบบนชุดข้อมูล MSRA MIT และ SED2 นั้นได้แสดงให้เห็นว่า การรวมคุณลักษณะความเข้มของภาพนำเข้า ที่คำนวณจากระบบสี RGB หรือ คุณลักษณะในส่วนค่าความสว่าง ของระบบสี CIE LAB เข้ากับคุณลักษณะของโมเดล MEV จะสามารถที่จะเพิ่มประสิทธิภาพความถูกต้องให้กับโมเดล MEV จากการเปรียบเทียบโดยใช้กราฟการเปรียบเทียบอัตราของค่าความถูกต้องกับค่าความครบถ้วน กับ กราฟความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR) และอัตราค่าความผิดพลาดเชิงบวก (FPR) ทั้งนี้ ผู้วิจัยสามารถสรุปได้ว่า ถึงแม้ว่าบางคุณลักษณะจะช่วยในการดึงข้อมูลที่หายไปกลับคืนมา แต่ก็ไม่ทุกครั้งที่ข้อมูลที่ทำการดึงกลับมานั้นจะมีประโยชน์และสามารถนำมาประยุกต์ใช้งานได้กับโมเดลที่นำเสนอได้อย่างมีประสิทธิภาพ อย่างเช่นเดียวกับคุณลักษณะของ HSV ที่ร่วมพิจารณา ซึ่งจากผลการทดลองที่แสดงในภาพกราฟที่ 25 ผู้วิจัยจึงเห็นควรที่จะทำการรวมคุณลักษณะความเข้มที่คำนวณจากระบบสี RGB เข้ากับคุณลักษณะที่นำเสนอ เพื่อเพิ่มประสิทธิภาพด้านความถูกต้องให้กับโมเดล MEV

เนื่องจากชุดข้อมูล MIT นั้นประกอบไปด้วยภาพที่มีวัตถุ ที่ตรงกับวัตถุประสงค์ของตัวตรวจจับในปริมาณมาก จึงทำให้คุณลักษณะระดับสูงมีผลต่อประสิทธิภาพความถูกต้องของโมเดล แต่ถ้าชุดข้อมูลนั้นมีภาพที่มีวัตถุ ที่ตรงกับวัตถุประสงค์ของตัวตรวจจับในปริมาณน้อย ตัวอย่างเช่น ชุดข้อมูล MSRA ก็จะทำให้ประสิทธิภาพด้านความถูกต้องของโมเดลลดน้อยลง

ข้อสังเกตอีกอย่างหนึ่งของตารางที่ 7 คือ กรณีที่มีการรวมค่าความเข้มแสงเข้าไปในโมเดลที่นำเสนอ ซึ่งผลที่ได้ออกมา แสดงให้เห็นว่าค่าเฉลี่ยพื้นที่ใต้กราฟของโมเดลที่ทำการรวมค่าความเข้มแสงเข้าไปนั้นมีค่าเฉลี่ยพื้นที่ใต้กราฟมากกว่า โมเดลที่นำเสนอ ทั้งนี้เป็นเพราะว่าคุณสมบัติความเข้มของสีนั้นได้มีส่วนที่จะช่วยในการระบุตำแหน่งของวัตถุเด่น

3. สรุปผลการประเมินประสิทธิภาพของโมเดล Judd และ JuddLM+MEV บนชุดข้อมูลทั้งสามชุด

เพื่อชี้ให้เห็นว่าโมเดล JuddLM+MEV นั้นให้ประสิทธิภาพที่ดีกว่าโมเดลของ Judd ในด้านประสิทธิภาพด้านความถูกต้องและทางด้านเวลาในการประมวลผลบนชุดข้อมูลทดสอบทั้งสองแบบ ไม่ว่าจะเป็น ชุดข้อมูลของภาพที่เก็บเฉพาะพื้นที่ของวัตถุเด่นได้แก่ MSRA, SED2 และ ชุดข้อมูลที่ระบุตำแหน่งที่ตามนุษย์มองเข้าไปในภาพ (MIT) ดังนั้นผู้วิจัยจึงได้เสนอตารางสรุปผลการประเมินดังกล่าว โดยใช้ค่า ROC กับ ค่า PR และเวลาในการประมวลผลที่สามารถหาได้จากเวลาเฉลี่ยที่ได้จากการคำนวณต่อภาพเป็นหน่วยวินาที โดยผลลัพธ์ที่แสดงในตารางที่ 8 ผู้วิจัยขอสรุปการเปรียบเทียบประสิทธิของ Judd กับ JuddLM+MEV โดยแบ่งออกเป็นแต่ละชุดข้อมูลดังนี้

ตารางที่ 8 สรุปการเปรียบเทียบประสิทธิภาพของ Judd กับ JuddLM+MEV บนชุดข้อมูล MSRA SED2 และ MIT

โมเดล / ชุดข้อมูล	เวลาเฉลี่ยในการคำนวณต่อภาพ (วินาที)		ค่า ROC		ค่า PR	
	Judd	JuddLM+MEV	Judd	JuddLM+MEV	Judd	JuddLM+MEV
MSRA	18.0908	**9.9435**	0.9177	**0.9229**	0.7054	**0.7189**
SED2	7.8256	**6.8282**	0.8580	**0.8626**	0.4857	**0.4893**
MIT	20.0786	**10.1274**	**0.8358**	0.8322	-	-

3.1 สรุปการเปรียบเทียบบนชุดข้อมูล MSRA

สำหรับการเปรียบเทียบบนชุดข้อมูล MSRA จะสังเกตได้ว่าโมเดล JuddLM+MEV นั้นให้ค่า ROC ที่สูงกว่าโมเดลของ Judd ถึง 0.0052 ดังแสดงในตารางที่ 8 ทั้งนี้จากผลในตารางที่ 8 แสดงให้เห็นว่าโมเดลที่นำเสนอสามารถที่จะทำนายพื้นที่ของวัตถุเด่นและมีความถูกต้องมากกว่าโมเดลของ Judd ซึ่งเหตุผลเนื่องมาจากภาพบนชุดข้อมูล MSRA นั้นมีความหลากหลายของวัตถุที่อยู่ในภาพ จึงทำให้คุณลักษณะระดับสูงที่ใช้ในโมเดลของ Judd นั้นมีการทำนายตำแหน่งของวัตถุที่ผิดพลาด หรือ ไม่สามารถที่จะทำนายตำแหน่งของวัตถุเด่นได้เลย เนื่องมาจากวัตถุดังกล่าวที่มีอยู่ในภาพนั้นไม่ตรงกับวัตถุประสงค์ของตัวตรวจจับที่มีอยู่ในโมเดลของ Judd ตัวอย่างเช่น ผู้วิจัยต้องการทำนายตำแหน่งรถยนต์ในภาพ โดยใช้คุณลักษณะระดับสูงที่อยู่ในโมเดลของ Judd เช่น โมเดลการตรวจจับใบหน้า ซึ่งผลลัพธ์ทางทฤษฎีแล้ว โมเดลดังกล่าวจะต้องทำนายว่าไม่มีใบหน้าอยู่ในภาพที่นำมาพิจารณา แต่ผลปรากฏว่า ในบางครั้งเกิดเหตุการณ์ที่โมเดลตรวจจับในหน้านั้นมีการทำนายไฟหน้ารถยนต์เป็นตำแหน่งของใบหน้า เป็นต้น สำหรับกรณีที่ทำนายออกมาแล้วไม่มีใบหน้าอยู่ในภาพที่พิจารณา ตัวโมเดลใบหน้าก็จะไม่มีประโยชน์ต่อโมเดลของ Judd อีกทั้งยังต้องเสียเวลาในการประมวลผลของคุณลักษณะระดับสูงดังกล่าวอีกด้วย

3.2 สรุปการเปรียบเทียบบนชุดข้อมูล SED2

สำหรับการเปรียบเทียบบนชุดข้อมูล SED2 จะเห็นได้ว่าโมเดล JuddLM+MEV นั้นมีค่า ROC ที่สูงกว่าโมเดลของ Judd ถึง 0.0046 ดังแสดงในตารางที่ 8 ยิ่งกว่านั้นยังมีประสิทธิภาพที่ดีกว่าโมเดลที่นำมาเปรียบเทียบ ดังแสดงในภาพกราฟที่ 15 ซึ่งเหตุผลหลักมาจากการทำนายวัตถุเด่นที่มีมากกว่าหนึ่งวัตถุในภาพนั้นมีลักษณะที่คล้ายกับการทำนายบนชุดข้อมูลการระบุตำแหน่งที่ตามนุษย์มองเข้าไปในภาพ โดยลักษณะของวัตถุในภาพบนชุดข้อมูลดังกล่าว จะประกอบไปด้วยวัตถุเด่นที่มากกว่าหนึ่งวัตถุเป็นส่วนใหญ่ อีกทั้งยังมีคุณลักษณะระดับสูงต่างๆ ช่วยทำนายตำแหน่งของวัตถุเด่นในภาพ จึงสามารถที่จะนำไปใช้กับชุดข้อมูลที่มีวัตถุเด่นที่มากกว่าหนึ่งวัตถุได้อย่างมีประสิทธิภาพมากกว่าโมเดลการทำนายตำแหน่งของพื้นที่วัตถุเด่นในภาพ

3.3 สรุปการเปรียบเทียบบนชุดข้อมูล MIT

เหตุผลหลักที่โมเดล JuddLM+MEV นั้นมีค่า ROC ที่น้อยกว่ากับโมเดลของ Judd ประมาณ 0.0036 ดังแสดงในตารางที่ 8 เนื่องมาจากข้อมูลบนชุดข้อมูล MIT นั้นประกอบไปด้วยใบหน้า รถยนต์ และบุคคล เป็นจำนวนมากบนชุดข้อมูล ดังนั้นจึงทำให้โมเดลในส่วนของคุณลักษณะระดับสูงมีส่วนช่วยในการทำนายพื้นที่ของวัตถุเด่นในภาพ แต่ถ้าสังเกตจากตารางที่ 3 จะเห็นได้ว่าโมเดล JuddLM+MEV ได้ใช้เวลาในการประมวลผลที่น้อยกว่าโมเดลของ Judd ถึง 20 เปอร์เซ็นต์

จากการทดลองเปรียบเทียบประสิทธิภาพโมเดลของ Judd และโมเดล JuddLM+MEV บนชุดข้อมูลทั้งสาม สามารถสรุปได้ว่า ผลลัพธ์ที่ได้จากคุณลักษณะระดับสูงจะมีลักษณะในการทำนายเพื่อที่จะระบุตำแหน่งของวัตถุเด่นในภาพอย่างคร่าวๆ ว่าอยู่ตรงส่วนไหนของภาพ โดยจะทำการตีกรอบสี่เหลี่ยมล้อมรอบวัตถุดังกล่าว ซึ่งผลลัพธ์ของการระบุตำแหน่งนี้ จะมีลักษณะที่คล้ายกับโมเดล MEV คือ ทำการระบุตำแหน่งของวัตถุเด่นในภาพอย่างคร่าวๆ เช่นกัน แต่ว่าโมเดล MEV นั้นจะทำการระบุตำแหน่งที่ดีกว่าคุณลักษณะระดับสูง ตรงที่สามารถทำนายตำแหน่งที่เป็นรูปร่างคร่าวๆ ของวัตถุ หรือ ตำแหน่งของวัตถุดังกล่าวไม่ใช่กรอบสี่เหลี่ยมเหมือนดังเช่นผลลัพธ์ที่ได้จากคุณลักษณะระดับสูง ซึ่งสิ่งนี้จะเข้าไปช่วยลดค่าความผิดพลาดในโมเดล JuddLM+MEV ยิ่งกว่านั้นโมเดล JuddLM+MEV ยังมีความเหมาะสมกับชุดข้อมูลภาพที่มีความหลากหลายของวัตถุในภาพกว่าโมเดลของ Judd ดังแสดงในภาพกราฟที่ 11 และ 15 เป็นต้น

วิจารณ์

ทั้งนี้งานวิจัย ได้ให้ความสำคัญในการกำจัดส่วนที่ไม่ได้เป็นสิ่งเด่นออกจากผลลัพธ์ของโมเดลการตรวจจับสิ่งเด่น เพื่อให้มีความเหมาะสมที่จะนำไปประยุกต์ใช้ในงานทางด้านการระบุตำแหน่งของวัตถุเด่น ซึ่งตัวชี้วัดที่งานวิจัยนี้ให้ความสำคัญ คือ ค่าอัตราความถูกต้องเชิงบวก ซึ่งหมายถึงข้อมูลที่เป็นพื้นที่ของวัตถุเด่นแล้วโมเดลได้ทำนายว่าเป็นพื้นที่ของวัตถุเด่น จากภาพกราฟที่ 12 พบว่าโมเดลที่นำเสนอจะให้ผลการทำนายที่มีความถูกต้องสูงกว่าโมเดลในปัจจุบัน บนชุดข้อมูล MSRA ที่มีผลเฉลยเป็นภาพแบบทวิภาคที่ทำการตัดตามขอบของวัตถุเด่นโดยใช้มนุษย์

สำหรับการเปรียบเทียบของอัตราของค่าความถูกต้องกับค่าความครบถ้วน จากภาพที่ 10 และภาพที่ 11 แสดงให้เห็นว่าโมเดลที่นำเสนอมีค่าความถูกต้อง ที่สูงที่สุดในช่วงที่ค่าความครบถ้วนไม่เกิน 0.25 จากค่าดังกล่าวสามารถที่จะบอกได้ว่า ถ้ามีการนำโมเดลไปประยุกต์ใช้ในงานวิจัยที่ต้องการแยกพื้นหน้าและพื้นหลังออกจากกัน หรือต้องการระบุตำแหน่งของวัตถุเด่นในภาพ จะต้องมีการกำหนดค่าขีดแบ่ง (Threshold) ประมาณ 25 เปอร์เซ็นต์ในช่วงบน เพื่อให้ได้ตำแหน่งของพื้นหน้าหรือวัตถุเด่นหลังการตรวจจับ ตัวอย่างสำหรับ 25 เปอร์เซ็นต์ในช่วงบน คือถ้าภาพมีพิสัยความเข้มของจุดภาพเท่ากับ 0 ถึง 255 ค่าขีดแบ่งสำหรับ 25 เปอร์เซ็นต์ในช่วงบนจะมีค่าเท่ากับ 63.75 จากการเทียบบัญญัติไตรยางศ์

จากค่าพื้นที่ใต้กราฟ ROC ที่เสนอในตารางที่ 4 พบว่าหลังจากนำคุณลักษณะของสี RGB ที่แยกตามช่องสีและความน่าจะเป็นของสี RGB ที่แยกตามช่องสี ออกจากโมเดลที่นำเสนอ ทำให้ค่าพื้นที่ใต้กราฟ ROC เพิ่มขึ้น ทั้งนี้อาจเนื่องมาจากคุณลักษณะสีทั้งสองให้ความผิดพลาดในการทำนายพื้นที่เด่นในภาพ เป็นผลให้พื้นที่ใต้กราฟ ROC โดยรวมมีค่าที่ลดลง และยังพบว่าถ้านำการคำนวณคุณลักษณะของความน่าจะเป็นของสีที่ผ่านการเบลอที่ขนาดของเคอร์เนลแตกต่างกันโดยใช้วิธีการคำนวณฮิสโทแกรมแบบสามมิติออก จะลดเวลาในการคำนวณลงถึง 10 เท่า โดยเทียบจากเวลาในการคำนวณต่อภาพ

ในการทำงานบนภาพที่มีความหลากหลายของวัตถุในภาพ โมเดลที่มีการใช้คุณลักษณะระดับสูง อาจไม่เหมาะสม ตัวอย่างเช่น ถ้าภาพบนชุดข้อมูลประกอบไปด้วยภาพสัตว์จำนวนมาก โมเดลของ Judd (Judd *et al.*, 2009) อาจจะไม่เหมาะสม เป็นต้น แต่สำหรับโมเดลที่เสนอสามารถตรวจจับวัตถุเด่นต่างๆ ในภาพได้โดยไม่จำกัดประเภทของวัตถุ ยิ่งกว่านั้นยังมีความซับซ้อนในการประมวลผลที่น้อยกว่าตัวตรวจจับวัตถุ จึงทำให้ประหยัดเวลาในการประมวลผลกว่าโมเดลของ Judd แต่ทว่า ค่าความเปรียบต่างของสีของวัตถุจะต้องมีค่าที่ชัดเจน

สรุปและข้อเสนอแนะ

สรุป

วิทยานิพนธ์ฉบับนี้ได้เสนอวิธีใหม่สำหรับการประมาณตำแหน่งของวัตถุเด่นในภาพบนพื้นฐานความเปรียบต่างของสี เนื่องจากเป็นคุณลักษณะที่เด่นและง่ายที่จะดึงดูดความสนใจของมนุษย์ และมีผลกระทบที่สำคัญในการตรวจจับพื้นที่วัตถุเด่นในภาพ โดยสิ่งที่เป็นความแตกต่างระหว่างสิ่งที่นำเสนอกับโมเดลในงานวิจัยที่มีในปัจจุบัน คือ งานวิจัยนี้สามารถสร้างคุณลักษณะที่เน้นเฉพาะพื้นที่ของวัตถุเด่นเท่านั้นโดยที่ปราศจากส่วนที่เป็นพื้นหลังหรือสิ่งที่ไม่สนใจ ยิ่งกว่านั้นยังสามารถนำไปประยุกต์กับโมเดลสำหรับการทำนายตำแหน่งที่ตาคนมองเข้าไปในภาพ

ประสิทธิภาพของการตรวจจับวัตถุเด่น ใช้กราฟชี้วัดมาตรฐานสองตัว คือ กราฟแสดงการเปรียบเทียบอัตราของค่าความถูกต้องกับค่าความครบถ้วนของโมเดล และกราฟ ROC ที่ได้จากความสัมพันธ์ระหว่างอัตราค่าความถูกต้องเชิงบวก (TPR)และอัตราค่าความผิดพลาดเชิงบวก (FPR)

ผลลัพธ์ของวิทยานิพนธ์นี้แสดงให้เห็นว่าโมเดลที่นำเสนอมีค่าความถูกต้องสูงสุดที่ ค่าความครบถ้วนเท่ากับศูนย์ ถึงแม้ว่าในช่วงที่ค่าความครบถ้วน มากกว่า 0.3 จะมีค่าความถูกต้อง ที่น้อยกว่าของกราฟโมเดล AC (Achanta *et al.*, 2009a) แต่ในกราฟโมเดล AC มีพื้นที่ใต้กราฟ ROC ที่น้อยกว่าโมเดลที่นำเสนอ ซึ่งแสดงให้เห็นถึงโมเดลที่นำเสนอมีการทำนายตำแหน่งของวัตถุเด่นได้ถูกต้องมากกว่าและมีประสิทธิภาพที่ดีกว่าโมเดล AC และเพื่อแสดงให้เห็นว่าวิธีการที่นำเสนอสามารถที่จะทำงานกับภาพที่มีวัตถุเด่นมากกว่าหนึ่งวัตถุ ผู้วิจัยจึงได้ทำการวัดประสิทธิภาพบนชุดข้อมูล SED2 จากผลการทดลองแสดงให้เห็นว่า วิธีการที่นำเสนอนั้นมีความใกล้เคียงกับโมเดลของ Judd และดีกว่ากับโมเดลในปัจจุบันที่นำมาทดสอบ สำหรับประสิทธิภาพด้านความถูกต้องในการระบุตำแหน่งที่ตาคนมองเข้าไปในภาพของโมเดลที่นำเสนอนี้มีผลลัพธ์ที่ใกล้เคียงกับโมเดลของ Judd สำหรับประสิทธิภาพด้านเวลาที่ใช้ในการคำนวณ โมเดลที่นำเสนอนั้นใช้เวลาที่น้อยกว่าโมเดลของ Judd ถึง 20 เปอร์เซ็นต์ โดยทำการเทียบจากการคำนวณในชุดข้อมูล MIT จำนวน 1003 ภาพ

สำหรับการนำค่าความเข้มแสง ที่ได้จากการคำนวณจากระบบสี RGB เข้ามาพิจารณาร่วมกับ MEV นั้นถือว่ามีความเหมาะสมเป็นอย่างยิ่ง โดยอิงจากผลการทดลองบน 3 ชุดข้อมูล ไม่ว่า

จะเป็น MSRA MIT และ SED2 ที่อยู่ในส่วนของการระบุพื้นที่วัตถุเด่นในภาพ นั้นได้ให้ค่าความถูกต้องที่มากกว่า MEV ยิ่งกว่านั้นยังทำให้ค่าพื้นที่ใต้กราฟของ Judd+MEV มีค่าที่มากขึ้นเช่นกันดังสังเกตได้จากตารางที่ 5

จากการดำเนินงานวิจัยที่ผ่านมา สามารถสรุปผลการวิจัยได้ว่า ระบบที่ออกแบบสามารถใช้งานได้ตามวัตถุประสงค์ที่ต้องการ และมีประสิทธิภาพที่เพียงพอที่จะนำไปประยุกต์ใช้งานให้เกิดประโยชน์ได้ รวมทั้งมีโอกาสในการพัฒนาให้มีประสิทธิภาพที่ดียิ่งขึ้นได้

ข้อเสนอแนะ

ข้อเสนอแนะสำหรับการนำไปใช้ประโยชน์ของงานวิจัย คือ สามารถที่จะนำไปใช้ในกระบวนการแยกพื้นหน้าและพื้นหลังออกจากกันในส่วนเริ่มต้นของกระบวนการทำงาน ทั้งนี้เพื่อเพิ่มประสิทธิภาพด้านความถูกต้องและช่วยลดเวลาในการคำนวณในงานวิจัยทางด้านคอมพิวเตอร์วิทัศน์ ที่เกี่ยวกับการลดขนาดภาพ การบีบอัด และการตัดเฉพาะส่วนที่สนใจออกจากภาพ

สำหรับการพัฒนาวิธีการตรวจจับวัตถุเด่นในภาพ โดยการเพิ่มคุณลักษณะต่างๆ เข้ามาช่วยในการพิจารณาวัตถุ อาจมีส่วนช่วยให้ประสิทธิภาพด้านความถูกต้องมีค่าที่มากยิ่งขึ้น แต่ก็ต้องแลกมากับเวลาที่ใช้ในการประมวลผล เช่นการเพิ่มค่าความเข้มแสงของภาพในโมเดล MEV ในวิธีการตรวจจับวัตถุเด่นในภาพ ดังแสดงผลลัพธ์ในตารางที่ 5 ดังนั้นสิ่งที่ควรคำนึงเป็นสำคัญ คือ ลักษณะของงานที่จะนำวิธีการตรวจจับวัตถุเด่นในภาพไปประยุกต์ใช้งาน

แนวทางสำหรับการวิจัย เพื่อพัฒนาวิธีการที่นำเสนอ อาจใช้วิธีการปรับเปลี่ยนคุณลักษณะหรือ วิธีการรวมวิธีการตรวจจับวัตถุเด่นหลายๆ วิธีการเข้าด้วยกัน เพื่อให้สามารถตรวจจับวัตถุเด่นได้แม่นยำขึ้นโดยใช้ทรัพยากรเหมาะสมที่สุด

เอกสารและสิ่งอ้างอิง

มัลลิกา บุนนาค, กัลยา ครองแก้ว, วัชราภรณ์ สุริยาภิวัฒน์ และน. รุ่งอุทัยศิริ. 2540. **สถิติ**. จุฬาลงกรณ์มหาวิทยาลัย,กรุงเทพฯ.

วิทยากร อัศดรวิเศษ, ไพศาล กิตติศุภกร, ภูมิพัฒ แสงอุดมเลิศ, ปาณิศา แก้วสวัสดิ์, วรากร ศรีเชวง ทรัพย์, นรรัตน์ วัฒนมงคล, ทับทิม อ่างแก้ว, สุเชษฐ์ ลิขิตเลอสรวง, จันทร์จิรา สินทนะโยธิน, วศินี หนุนภักดี, กำพล วรดิษฐ์, ยุรนันท์ ลิมปนันท์วดี, Annur Robithoh, สุวิทย์ นาคพีระยุทธ, ลัญฉกร วุฒิสิทธิกุลกิจ, อัจฉริยา สุริยะวงค์, พิสิฐ วนิชชานันท์, เอมอมร สุวิชากร และก. วรรณคง. 2555. **การประยุกต์ใช้ Matlab.** จุฬาลงกรณ์มหาวิทยาลัย,กรุงเทพฯ.

Achanta, R. and S. Susstrunk. 2009. Saliency Detection for Content-Aware Image Resizing, pp. 1005-1008. In **IEEE, International Conference on Image Processing.**

Achanta, R., S. Hemami, F. Estrada and S. Susstrunk. 2009a. Frequency-Tuned Salient Region Detection, pp. 1597-1604. In **IEEE, Computer Vision and Pattern Recognition.**

Achanta, R., S. Hemami, F. Estrada and S. Susstrunk. 2009b. **Frequency-Tuned Salient Region Detection**. Available Source: http://ivrgwww.epfl.ch/supplementary_material/RK_CVPR09/, 21 June 2012.

Alpert, S., M. Galun, R. Basri and A. Brandt. 2007. Image Segmentation by Probabilistic Bottom-up Aggregation and Cue Integration, pp. 1-8. In **Computer Vision and Pattern Recognition.**

Borji, A., D.N. Sihite and L. Itti. 2012. Salient Object Detection: A Benchmark, pp. 414-429. In **Proceedings of the 12th European conference on Computer Vision - Volume Part II.** Springer-Verlag, Florence, Italy.

Bruce, N. and J. Tsotsos. 2005. Saliency Based on Information Maximization, pp. 155-162. In **Advances in neural information processing systems**.

Erdem, E. and A. Erdem. 2013. Visual Saliency Estimation by Nonlinearly Integrating Features Using Region Covariances. **Journal of vision** 13 (4):1-20

Felzenszwalb, P.F., R.B. Girshick, D. McAllester and D. Ramanan. 2010. Object Detection with Discriminatively Trained Part-Based Models. **IEEE Transactions on Pattern Analysis and Machine Intelligence** 32 (9): 1627-1645.

Fu, K., C. Gong, J. Yang, Y. Zhou and I. Yu-Hua Gu. 2013. Superpixel Based Color Contrast and Color Distribution Driven Salient Object Detection. **Signal Processing: Image Communication** 28 (10): 1448-1463.

Harel, J., C. Koch and P. Perona. 2006. Graph-Based Visual Saliency, pp. 545-552. In **Advances in neural information processing systems.**

Hui, Z., W. Weiqiang, S. Guiping and D. Lijuan. 2012. A Simple and Effective Saliency Detection Approach, pp. 186-189. In **International Conference on Pattern Recognition (ICPR).**

Itti, L., C. Koch and E. Niebur. 1998. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. **IEEE Transactions on Pattern Analysis and Machine Intelligence** 20 (11): 1254-1259.

Judd, T., K. Ehinger, F. Durand and A. Torralba. 2009. Learning to Predict Where Humans Look, pp. 2106-2113. In **IEEE 12th international conference on Computer Vision**,

Ko, B.C. and J.-Y. Nam. 2006. Object-of-Interest Image Segmentation Based on Human Attention and Semantic Region Clustering. **JOSA A** 23 (10): 2462-2470.

Ma, Y.-F. and H.-J. Zhang. 2003. Contrast-Based Image Attention Analysis by Using Fuzzy Growing, pp. 374-381. In **Proceedings of the eleventh ACM international conference on Multimedia.**

Rutishauser, U., D. Walther, C. Koch and P. Perona. 2004. Is Bottom-up Attention Useful for Object Recognition?, pp. II-37-II-44 Vol.32. In **IEEE Computer Vision and Pattern Recognition.**

Schauerte, B. and R. Stiefelhagen. 2013. How the Distribution of Salient Objects in Images Influences Salient Object Detection, In **Proceedings of the 20th (ICIP) International Conference on Image Processing,**

Tie, L., Y. Zejian, S. Jian, W. Jingdong, Z. Nanning, T. Xiaoou and S. Heung-Yeung. 2011. Learning to Detect a Salient Object. **IEEE Transactions on Pattern Analysis and Machine Intelligence** 33 (2): 353-367.

Van Der Linde, I., U. Rajashekar, A. Bovik and L. Cormack. 2009. Doves: A Database of Visual Eye Movements. **Spaital vision** 22 (2): 161-177.

Vijanprecha, S. and Wattuya, P. 2014. Salient Location based on Color Contrast, doi:10.1117/12.2064429. **In Sixth International Conference on Digital Image Processing (ICDIP 2014)**.

Viola, P. and M.J. Jones. 2004. Robust Real-Time Face Detection. **International journal of computer vision** 57 (2): 137-154.

Xiaokang, Y., L. Weisi, L. Zhongkang, L. Xiao, S. Rahardja, O. EePing and Y. Susu. 2005. Rate Control for Videophone Using Local Perceptual Cues. **IEEE Transaction on Circuits and Systems for Video Technology** 15 (4): 496-507.

Zhu, G., Q. Wang and Y. Yuan. 2014. Tag-Saliency: Combining Bottom-up and Top-Down Information for Saliency Detection**. Computer Vision and Image Understanding** 118 (0): 40-49.

ประวัติการศึกษา และการทำงาน

ชื่อ –นามสกุล	สุชาติ วิจารณ์ปรีชา
วัน เดือน ปี ที่เกิด	วันที่ 12 พฤษภาคม 2532
สถานที่เกิด	นครสวรรค์
ประวัติการศึกษา	วท.บ. (วิทยาการคอมพิวเตอร์) มหาวิทยาลัยนเรศวร
ตำแหน่งหน้าที่การงานปัจจุบัน	-
สถานที่ทำงานปัจจุบัน	-
ผลงานดีเด่นและรางวัลทางวิชาการ	-
ทุนการศึกษาที่ได้รับ	ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขต บางเขน