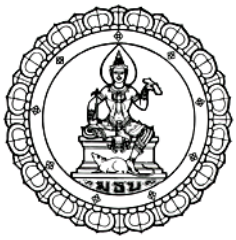


รายงานการวิจัย

เรื่อง

การจัดกลุ่มความหมายภาพส่วนบุคคลด้วยข้อความเหตุการณ์กิจกรรม

Semantic Clustering for Personal Image with Context of Activity Events

โดย

นัศพ์ชาณัณ ชินปัญช์ธนะ

มหาวิทยาลัยธุรกิจบัณฑิตย์

รายงานผลการวิจัยนี้ได้รับทุนอุดหนุนจากมหาวิทยาลัยธุรกิจบัณฑิตย์

พ.ศ. 2557

ชื่อเรื่อง : การจัดกลุ่มความหมายภาพส่วนบุคคลด้วยข้อความเหตุการณ์กิจกรรม
ผู้วิจัย : นัศพ์ชาณัณ ชินปัญช์ธนะ **สถาบัน** : มหาวิทยาลัยธุรกิจบัณฑิตย์
ปีที่พิมพ์ : พุทธศักราช 2558 **สถานที่พิมพ์** : มหาวิทยาลัยธุรกิจบัณฑิตย์
แหล่งที่เก็บรายงานการวิจัยฉบับสมบูรณ์ : ศูนย์วิจัยมหาวิทยาลัยธุรกิจบัณฑิตย์
จำนวนหน้าวิจัย : 67 หน้า **ลิขสิทธิ์** : มหาวิทยาลัยธุรกิจบัณฑิตย์
คำสำคัญ : การประมวลผลภาพ, การจัดกลุ่ม, ภาพเหตุการณ์กิจกรรม

บทคัดย่อ

การจำแนกความหมายภาพเป็นหัวข้องานวิจัยที่ยังคงท้าทายอย่างมากในสาขาการประมวลผลภาพ มีนักวิจัยหลายกลุ่มพยายามปรบปรุงวิธีการเพื่อแก้ไขปัญหาของการแทนความหมายภาพด้วยการประมวลผลภาพระดับต่ำที่มีการใช้สถิติเข้ามาช่วยเพื่ออธิบายภาพ แต่อย่างไรวิธีการเหล่านั้นไม่สามารถที่จะนำมาใช้แทนความหมายของภาพได้อย่างแท้จริง ดังนั้นในงานวิจัยนี้จึงได้พยายามใช้เทคนิคของการรับรู้ของมนุษย์ และการสกัดภาพขนาดของวัตถุด้วยโครงสร้างสเกลตรีตรอน

งานวิจัยนี้ได้นำเสนอวิธีการใหม่สำหรับการจัดกลุ่มภาพส่วนบุคลด้วยข้อความเหตุการณ์กิจกรรมวิธีการนี้มีทั้งหมด 4 ส่วนคือ (1) การจัดเก็บข้อมูล (2) การแทนค่าวัตถุของภาพ (3) การจัดกลุ่มเหตุการณ์กิจกรร (4) แสดงผลที่ได้จากการจัดกลุ่มภาพ วิธีที่นำเสนอใหม่นี้สามารถจำแนกความหมายของภาพได้ดีกว่าวิธีอื่นๆ และได้ค่าความถูกต้องสูงถึง 72.11%

Title : Semantic Clustering for Personal Image with Context of Activity Events

Researcher : Nutchanun Chinpanthana **Institution** : Dhurakijpundit University.

Year of Publication : 2015 **Publisher** : Dhurakijpundit University.

Sources : Dhurakijpundit University Research Center.

Number of Pages : 67 Pages **Copyright** : Dhurakijpundit University.

Keyword : Image processing, Clustering, Activity Event Images

Abstract

Semantic clustering is an active problem in multimedia personal photo collections. Many researchers have attempted to improve semantic models by using high-level concept based on keyword annotation. The model is rather rudimentary and it does not specific enough for representing the meaning of images. Therefore, we are applying the concept of human perception for extracting the objects. The structural skeleton is used to extract the object components and image meaning.

In this paper, we present a technique of the semantic clustering for personal images with context activity events. The approach is composed of four stages: (1) data collection, (2) object representation, (3) activity events clustering and (4) shows a semantic search result. The experimental results indicate that our proposed approach offers significant performance improvements in the cluster of activity events, with the maximum of 72.11%.

กิตติกรรมประกาศ

งานวิจัยชิ้นนี้สำเร็จลุล่วงได้ทั้งนี้เพราะได้รับความอนุเคราะห์ การสนับสนุน และแรงผลักดัน และอีกหลายฝ่ายที่ได้ให้ความช่วยเหลือเอื้อเฟื้อข้อมูลที่เป็นประโยชน์ในการทดลองโปรแกรม บันทึกผลการทดลอง และงานวิจัยนี้จะไม่สมบูรณ์ได้ หากไม่ได้ข้อเสนอแนะที่เป็นประโยชน์จากท่านอาจารย์ผู้ทรงคุณวุฒิ ผู้ซึ่งให้คำแนะนำที่ดีและมุมมองที่ผู้วิจัยได้นำมาปรับปรุงให้สมบูรณ์ยิ่งขึ้น

นอกจากนี้ใคร่ขอขอบคุณมหาวิทยาลัยธุรกิจบัณฑิตย์ ผู้ให้ทุนสนับสนุนงานวิจัยชิ้นนี้ หากงานวิจัยเล่มนี้มีข้อผิดพลาด ประการใดขออภัยไว้ ณ ที่นี้ด้วย

นัศพ์ชาณัณ ชินปัญช์ธนะ

กันยายน 2558

สารบัญ

สารบัญตาราง

ตารางที่หน้า

สารบัญรูปภาพ

รูปภาพที่ **หน้า**

บทที่ 1
บทนำ

1.1 ความเป็นมาของปัญหา

การประมวลผลภาพ จะอยู่ในรูปของการประมวลผลสัญญาณแอนะล็อก (analog) โดยใช้อุปกรณ์จับภาพแต่งภาพที่ถ่ายออกมายังคงเป็นภาพขาวดำและไม่ชัดเจนเท่าที่ควร ทำให้ส่วนใหญ่ภาพถ่ายมักจะถูกถ่ายเฉพาะภาพเหตุการณ์ที่สำคัญหรือบุคคลสำคัญเท่านั้น เนื่องจาก อุปกรณ์ที่นำมาใช้ในการถ่ายภาพ หรือบันทึกภาพนั้นซับซ้อนและมีความยุ่งยากในการใช้งาน รวมทั้งราคาสูง แต่อย่างไรก็ตามได้มีการพัฒนาปรับปรุงเทคโนโลยีเพื่อตอบรับกับความต้องการที่เพิ่มขึ้น ทำให้อุปกรณ์ถ่ายภาพในปัจจุบันมีราคาถูกลง และสะดวกต่อการใช้งานมาก และพัฒนาต่อเนื่องไปเป็นอุปกรณ์ถ่ายภาพแบบดิจิทัล (digital) ทำให้ภาพถ่ายดิจิทัล มีจำนวนเพิ่มมากขึ้นอย่างต่อเนื่อง ด้วยเหตุนี้ทำให้เกิดปัญหาในการจัดเก็บข้อมูลภาพรวมถึงการค้นคืนข้อมูลภาพ (image retrieval) ที่เพิ่มมากขึ้นอย่างต่อเนื่อง จนทำให้เกิดคำถามเดิม ๆว่าควรจะใช้วิธีการใดจึงจะสามารถจัดเก็บอย่างมีระบบและสามารถที่จะค้นหาภาพที่ต้องการได้อย่างถูกต้อง

การประมวลผลภาพดิจิทัล (digital image processing) เป็นสาขาที่ได้รับความนิยมในการวิจัยค่อนข้างมาก ทั้งทางด้านภาพ (still image) และ วิดีโอ (video) เพราะสามารถที่จะนำไปประยุกต์ได้หลายด้าน เช่น ทางด้านการแพทย์ ทางด้านอุตสาหกรรม หรือนำมาประยุกต์ใช้ในส่วนของทางด้านการจราจร ในการตรวจจับทะเบียนรถยนต์ เป็นต้น หัวข้อหลักในงานวิจัยแขนงนี้จะเกี่ยวเนื่องกันทั้งทางด้านการสร้างขั้นตอน (algorithm) เพื่อนำมาใช้ในการจำแนกหรือสืบค้นข้อมูลภาพ, การแทนความหมายของข้อมูลภาพ (image representation) และวิธีการใช้เครื่องมือในการจำแนกภาพ (classification method) [Pedro,2007] ซึ่งปัญหาทั้งหมดที่เกิดขึ้นเพียงเพื่อนำมาใช้ในการจำแนกหรือสืบค้นข้อมูลภาพ เพื่อให้ได้ภาพคำตอบที่ตรงตามความต้องการมากที่สุด สิ่งที่กล่าวมาข้างต้นเป็นเพียงส่วนหนึ่งที่นำการประมวลผลภาพดิจิทัลเข้ามาปรับปรุงและไปประยุกต์ใช้งาน

ปัจจุบันมีงานวิจัยหลายกลุ่ม [Qian Huang,1995];[Vailaya A.,2001];[W. Ma,1997];[C. Carson,1999] พยายามนำเทคนิคต่าง ๆเข้ามาใช้ในการค้นหาภาพเพื่อให้ผลลัพธ์ตรงกับความต้องการของผู้ใช้ เช่น งานวิจัยที่ใช้ข้อมูลจากการเขียนอัลกอริทึมเพื่อสกัดข้อมูลออกจากภาพแล้วนำมาใช้เป็นข้อมูลในการค้นคืนภาพ กลุ่มแรกนี้จะถูกเรียกว่า กระบวนการค้นคืนภาพตามเนื้อหา

สาระ (Content Based Image Retrieval: CBIR) จะเป็นการค้นคืนด้วยด้วยการรวมกันของลักษณะเฉพาะ (feature) จากการสกัดลักษณะเฉพาะจากภาพ (feature extraction) ออกมาเป็นส่วนๆ รูปแบบของการใช้ข้อมูลเหล่านี้มักถูกเรียกว่า ลักษณะเฉพาะระดับต่ำ (low-level features) เช่น สี (colour) รูปทรง (shape) และ พื้นผิว (texture) เป็นต้น ความพยายามในการค้นหาข้อมูลภาพเริ่มต้น ล้วนใช้ข้อมูลคุณลักษณะเฉพาะ ความพยายามในการค้นหาข้อมูลภาพเริ่มมาจากการสกัดคุณลักษณะข้อมูลภายในภาพ (feature extraction) ออกมาเป็นข้อมูลที่สามารถคำนวณได้ ข้อมูลส่วนนี้มักจะถูกเรียกกันว่าข้อมูลภาพระดับต่ำหรือ low-level features โดยข้อมูลนี้จะถูกนำมาใช้สำหรับการสืบค้นข้อมูลภาพแต่อย่างไรก็ตามการวิจัยในส่วนแรกนี้ยังคงมีการปรับปรุงพัฒนาอย่างต่อเนื่อง

การแทนคำอธิบายประกอบภาพ (Annotation) เป็นงานวิจัยอีกกลุ่มที่พยายามจะใช้เทคนิคของการเข้าใจความหมายของภาพแทน การสืบค้นแบบข้างต้น เป็นการประมวลผลภาพระดับสูง (high level image processing) งานวิจัยในกลุ่มนี้พยายามที่จะมองข้อมูลบนภาพเป็นวัตถุ (object) ที่มีความหมายและแทนวัตถุนั้นๆ ด้วยคำหลัก (keyword) บนภาพ เรียกว่า การแท็ก (tagging) หรือการให้ความหมายของวัตถุบนภาพเป็น ชื่อวัตถุ หรือคำศัพท์ ที่สอดคล้องกันเช่น “grass” , “plant” , “boat”, “sky” เป็นต้น [Galleguillos C., 2010] และใช้ความหมายหรือคำศัพท์นั้นเพื่อทำการสืบค้นข้อมูลแทน จะได้ผลที่ค่อนข้างดีกว่า แต่ขึ้นอยู่กับว่าขั้นตอนวิธีที่ถูกนำมาใช้นั้นจะเป็นลักษณะใด เช่น การใช้คำสำคัญของการแท็กหลายแท็ก (multiple label) [K.Narnard 2003][M. Johnson 2005] หรือการสร้างความสัมพันธ์ของวัตถุบนภาพด้วยโมเดลเหตุการณ์ (event model) [Joo-Hwee Lim, 2003] เช่นการเชื่อมวัตถุด้วยคำต่างๆ เช่น “touch”, “ontop” เป็นต้น บางงานวิจัยได้พยายามใส่ข้อความที่บรรยายความหมายของภาพ (context) [Mathias Lux, 2003];[Mathias Lux, 2009] ในหัวข้อที่สอดคล้องกับภาพ เช่น “birthday party of uncle Adam” หรือ “a picture showing a barking dog” รูปแบบของการใช้ข้อมูลเป็นคำหลักเหล่านี้มักถูกเรียกว่า ลักษณะเฉพาะระดับสูง (high-level features) เมื่อเปรียบเทียบผลลัพธ์ของการค้นคืนที่ค่อนข้างแม่นยำมากกว่าการใช้ฟีเจอร์ระดับต่ำ แต่อย่างไรก็ตามยังขึ้นอยู่กับว่าอัลกอริทึมที่ถูกนำมาใช้ร่วมด้วยนั้นจะเป็นลักษณะใด เช่น การใช้เทคนิคหาความสัมพันธ์ของแท็กชื่อวัตถุบนภาพ [Benitez A.B, 2001];[R. Zhao, 2002];[Philippe Mulhem,2002] ซึ่งเป็นการสอดคล้องกันด้วยความหมายตามพจนานุกรม การใช้ความสัมพันธ์ของความหมายที่เหมือนกันของคำ (synonym) การค้นหาภาพด้วยเทคนิคนี้จะได้ผลลัพธ์ที่ค่อนข้างขึ้นกับคำหลักที่ถูกให้ความหมายไว้ในภาพ หรือวัตถุบนภาพส่วนใหญ่

สำหรับการค้นคืนแต่ก็ยังถูกจำกัดด้วยการแท็กข้อมูลในภาพส่วนบุคคล เพื่อทำให้การค้นหาได้ความหมายภาพที่เพิ่มขึ้น M. Johnson and R. Cipolla [M. Johnson 2005] ได้พยายามจัดกลุ่มภาพที่มีวัตถุเหมือนกันอยู่ในกลุ่มเดียวกันแทนที่จะเป็นการแยกทีละภาพ I. Simon และคณะ [I. Simon 2007] ได้พยายามจัดกลุ่มภาพในลักษณะเดียวกันโดยที่จะใช้ข้อมูลสถานที่ (location) ของภาพช่วยในการจัดกลุ่ม นอกจากนี้การจัดกลุ่มตามชื่อเรื่อง (topic) หรือภาพที่เกิดขึ้นเป็นภาพธรรมชาติ อาจจะมีการจัดกลุ่มตาม เหตุการณ์ของธรรมชาติ (natural event) แต่บางงานวิจัยจะมีการใช้วิธีการเรียนรู้แม่แบบ ความคิด (learning concept templates) [Y. Wu 2007] เพื่อจัดกลุ่มภาพและเพิ่มส่วนของการย้อนกลับความสอดคล้องความสัมพันธ์ (relevance feedback) [Y. Wu 2006] เพื่อช่วยการการแก้ไขกลุ่มของภาพในการค้นคืนให้ได้ตรงตามกลุ่มเป้าหมายที่ดีขึ้น การค้นหาภาพด้วยเทคนิคที่กล่าวมานี้จะได้ผลลัพธ์ที่ขึ้นกับคำศัพท์ที่ถูกแท็กไว้บนภาพโดยยิ่งมีการแท็กข้อมูลบนภาพมากยิ่งสามารถหาความเหมือนกันบนภาพมากขึ้นเท่านั้น แต่ในความเป็นจริงแล้วภาพในปัจจุบันส่วนใหญ่ที่มีการจัดเก็บกันมากมายนั้นเป็นภาพส่วนบุคคล (Personal Photos) [Xin Xu 2013][M.S. Ryoo 2009] ดังนั้นในงานวิจัยนี้จึงนำเสนอการจัดแบ่งกลุ่มภาพส่วนบุคคลโดยใช้ข้อความเหตุการณ์กิจกรรม (activity events) ของมนุษย์เพื่อช่วยให้การจัดแบ่งกลุ่มได้ตรงตามความหมายของภาพมากขึ้น และทำเปรียบเทียบความเหมือนกันของความหมายภาพด้วยการขั้นตอนวิธีการจัดกลุ่มกิจกรรม (clustering algorithm) ตัวอย่างเหตุการณ์กิจกรรม เช่น ขับรถ (driving), studying (เรียน), eating (กินข้าว), ประชุม (meeting) เป็นต้น

1.2 วัตถุประสงค์

1. เพื่อพัฒนาเทคนิคใหม่ที่สามารถนำมาใช้ในการจัดกลุ่มความหมายภาพส่วนบุคคลด้วยข้อความเหตุการณ์กิจกรรม
2. สามารถนำไปประยุกต์กับโปรแกรมประยุกต์การค้นคืนภาพได้

1.3 สมมติฐาน

งานวิจัยนี้นำเสนอการแบ่งกลุ่มความหมายของภาพตามกิจกรรมเหตุการณ์ โดยใช้ความสัมพันธ์ (Relationship) ของการแท็กข้อมูลภาพจากเหตุการณ์กิจกรรม (activity event) ของมนุษย์ และวัตถุที่ปรากฏในภาพ เพื่อใช้เป็นลักษณะเฉพาะสำหรับข้อมูลการทดลอง และทำการจัดกลุ่มตามกิจกรรมของมนุษย์เพื่อแสดงถึงความหมายของภาพ

1.4 ขอบเขตของการวิจัย

1. ข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลภาพดิจิทัลที่เห็นกิจกรรมของมนุษย์ชัดเจน
2. ข้อมูลคำหลักบนภาพจะถูกจัดเก็บเป็นข้อมูลเข้าก่อนเริ่มการประมวลผล
3. ข้อมูลหรือวัตถุบนภาพที่เป็นอินพุตบนภาพจะถูกแท็ก เป็นคำศัพท์เข้ามาก่อน
4. ในการสร้างแบบจำลองภาพ (image representation) ข้อมูลจะถูกจัดเก็บมาในรูปแบบของฟีเจอร์ข้อมูลไว้แล้ว
5. ภาพจากคลังภาพจะถูกตั้งค่าเริ่มต้นโดยการใส่ข้อมูลวัตถุบนภาพพร้อมความสัมพันธ์ภาพไว้ก่อนแล้ว
6. ภาพที่ถูกนำมาใช้เป็นข้อมูลทดลองจากคลังภาพ
7. การเก็บผลข้อมูลเบื้องต้นของความหมายของภาพ ใช้กลุ่มนักศึกษา และบุคคลทั่วไปเป็นกลุ่มบุคคลที่ตัดสิน ความหมายของภาพสำหรับการทดลองในคลังภาพที่กล่าวมา

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ผลต่อสังคม

1. เพื่อนำเอากระบวนการนี้มาประยุกต์ใช้กับการค้นหาภาพบนห้องสมุดดิจิทัลได้
2. เพื่อได้แนวทางการแปลและการตีความหมายภาพ
3. ช่วยให้การใช้คำศัพท์ในการค้นหาข้อมูลภาพได้ผลตามความหมายของภาพมากขึ้น

1.5.2 ผลต่อมหาวิทยาลัย

1. สร้างกลุ่มนักวิจัยที่เป็นลักษณะของสาขาทางด้าน Images Processing ในสาขาย่อย Semantic Image Processing
2. เพื่อสนับสนุนและเสริมสร้างความรู้ความสามารถของการพัฒนาตนเอง และ กลุ่มงานในสาขาเทคโนโลยีสารสนเทศ
3. เพื่อเพิ่มตัวชี้วัดให้กับองค์กร สถาบันในส่วนของงานวิจัยและพัฒนา
4. สร้างชื่อเสียงให้กับมหาวิทยาลัย เมื่อมีบทความวิจัยลงในวารสารต่างประเทศ บทความวิจัยที่นำเสนอในการประชุมระดับชาติ และนานาชาติ

1.5.3 ผลต่อกลุ่มผู้วิจัย

1. พบแนวทางในการคิดค้นสิ่งใหม่ ๆในการจัดกลุ่มและกระบวนการแปลความหมายของภาพ

2. สามารถนำผลวิจัยมาเขียนบทความลงวารสารนานาชาติ และร่วมประชุมวิชาการระดับชาติและนานาชาติ
3. พัฒนาความรู้ใหม่ๆ หลักการแนวคิดใหม่ ให้เกิดขึ้นในกลุ่มของนักวิจัย

1.6 นิยามคำศัพท์

1. การประมวลผลภาพ (image processing) หมายถึง การนำภาพมาผ่านกระบวนการเพื่อประมวลผลสัญญาณบนสัญญาณ 2 มิติ เช่น ภาพนิ่ง หรือภาพวีดิทัศน์และนำมาใช้งาน
2. การค้นคืนภาพ (image retrieval) หมายถึง การสืบค้นภาพจากฐานข้อมูลภาพดิจิทัลเพื่อให้ได้ภาพที่ต้องการ
3. การจัดกลุ่มภาพ (image clustering) หมายถึง การแบ่งกลุ่มเป็นกลุ่มตามคุณลักษณะของฟีเจอร์ที่คล้ายคลึงกัน
4. ลักษณะเฉพาะ (feature) หมายถึงเป็นฟีเจอร์ หรือตัวแปร ที่ถูกสกัดออกมาจากภาพ เช่น สี (color) ลวดลาย (texture) หรือ รูปทรง (shape) รวมทั้ง วัตถุ ที่ปรากฏบนภาพเพื่อนำมาใช้ในการสืบค้นข้อมูลต่อไป
5. วัตถุ (object) หมายถึง ส่วนของวัตถุบนภาพ เช่น คน ต้นไม้ รถยนต์ เป็นต้น
6. คำศัพท์ (keyword) หมายถึง คำที่มีความหมายได้ใจความสามารถใช้แทนวัตถุบนภาพ
7. การแท็ก (tag) คือการกำหนดคำ หรือ คำศัพท์บนภาพ หรือเรียกว่าการกำหนดฉลาก (Labeled) ภาพด้วยคำศัพท์
8. เหตุการณ์กิจกรรม (activity event) คือเหตุการณ์ที่มนุษย์กระทำและเกิดขึ้นบ่อยๆ เป็นประจำ

บทที่ 2
ทฤษฎีและงานวิจัย

2.1 ทฤษฎีเบื้องต้น

วิวัฒนาการของเครื่องมือและอุปกรณ์ถ่ายภาพดิจิทัลได้พัฒนาอย่างรวดเร็วจนทำให้ภาพถ่ายภาพดิจิทัล มีจำนวนเพิ่มมากขึ้น ปัญหาที่ตามมาก็คือการจัดเก็บข้อมูลภาพที่เพิ่มมากขึ้นอย่างไร้ขีดจำกัดนี้จะทำอย่างไรจึงจะสามารถจัดเก็บอย่างมีระบบและสามารถสืบค้นข้อมูลภาพ และจำแนกข้อมูลภาพให้ตรงตามความหมายของภาพที่ต้องการของผู้ใช้มากที่สุด ทำให้งานวิจัยในปัจจุบันที่เกี่ยวข้องกับการค้นคืนรวมทั้งการจัดกลุ่มภาพให้ตรงกับความต้องการเพิ่มมากขึ้น รวมไปถึงระยะเวลาในการสืบค้นที่น้อยลงกับปริมาณของภาพที่เพิ่มทวีคูณ ดังนั้นปัญหาดังกล่าวมาข้างต้นนั้นจึงได้รับความสนใจจากนักวิจัยหลายกลุ่ม ซึ่งเป็นงานด้านการประมวลผลภาพ (image processing) ด้านการค้นคืนสารสนเทศ (image retrieval) เป็นอีกหนทางหนึ่งในการแก้ปัญหาดังกล่าว เนื่องจากสามารถช่วยให้การค้นหาข้อมูลกระทำได้โดยสะดวกยิ่งขึ้นรวมถึง การค้นหาภาพและการจำแนกประเภทข้อมูลภาพ (image classification) เพื่อคัดเลือกภาพ เพื่อให้ตรงตามความต้องการของผู้ใช้ยิ่งมีความจำเป็นอย่างยิ่ง

สำหรับงานวิจัยทางการประมวลผลภาพในการค้นคืนสารสนเทศกลุ่มแรก ๆจะมีการค้นคืนตามคุณลักษณะพื้นฐานของภาพที่ถูกสกัดคุณลักษณะด้วยอัลกอริทึมต่าง ๆ ยกตัวอย่างเช่น สี (color) พื้นผิว (texture) รูปทรง (shape) เป็นต้น กระบวนการนี้ถูกเรียกว่า การประมวลผลภาพระดับต่ำ (low-level image processing) [Jain A.K, 1996];[Smeulders, A. W. M, 2000] กระบวนการนี้สามารถค้นหาภาพได้ตามคุณลักษณะพื้นฐานที่นำไปสืบค้น โดยภาพผลลัพธ์ส่วนใหญ่มักจะเป็นภาพที่มีคุณลักษณะไม่ซับซ้อนมากนัก เช่น โทนสี หรือ รูปทรงที่แตกต่างกันอย่างเด่นชัด ดังภาพที่ 2.1 แสดงการค้นคืน ของ SIMPLIcity [Jia Li,2003];[James Z. Wang,2001] ด้วยคุณลักษณะฟีเจอร์ระดับต่ำด้วยสี ลวดลาย และตำแหน่งของพื้นที่ของภาพ จากผลลัพธ์จะสังเกตว่าผลลัพธ์ของภาพเป็นภาพที่มีโทนสีคล้ายกันเป็นหลัก แต่มีลักษณะวัตถุที่แตกต่างกันอย่างสิ้นเชิงที่อยู่ในหมวดหมู่เดียวกัน คือ ภาพชายหาด หรือ ชายทะเล แต่มีคุณลักษณะของโทนสี และลวดลาย ที่แตกต่างกันอย่างเห็นได้ชัดเจน เมื่อนำมาจำแนกด้วยการประมวลผลภาพระดับต่ำ (low-level feature) แล้วนั้นค่อนข้างยากที่จะจัดให้หมวดหมู่เดียวกัน

ภาพต้นฉบับ ภาพต้นฉบับที่ถูกแปลงเป็นฟีเจอร์

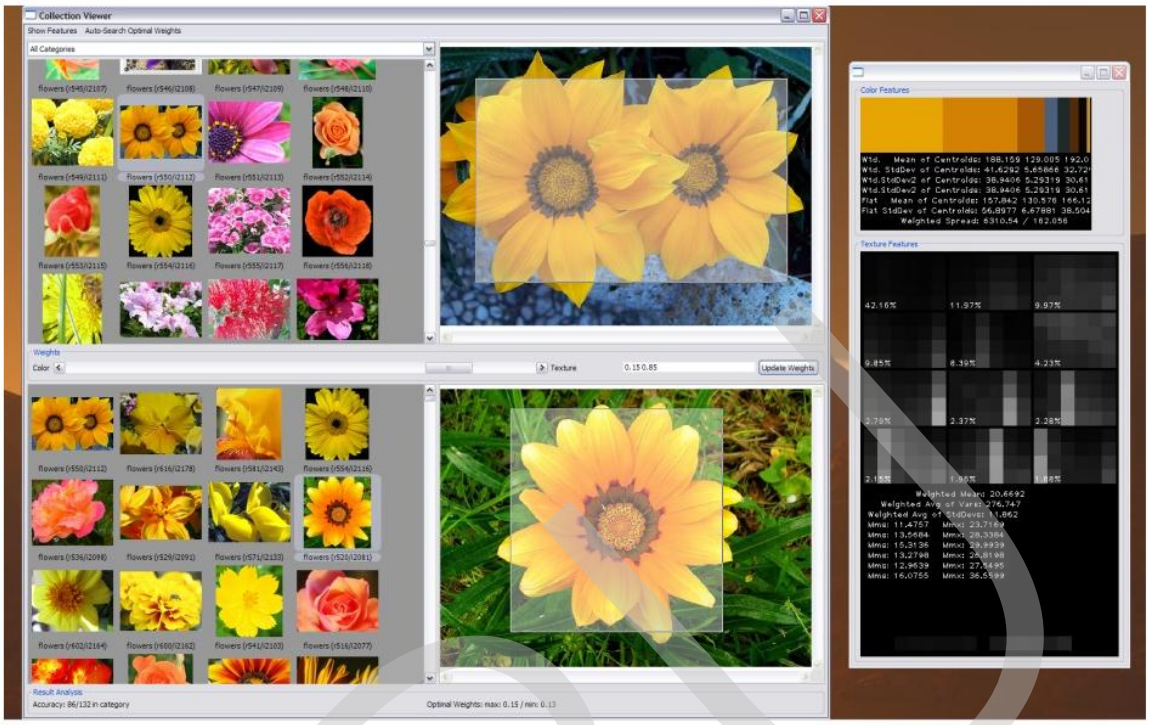
ก. ภาพต้นฉบับสำหรับการค้นคืนที่ถูกแปลงเป็นฟีเจอร์ระดับต่ำ

ข. ผลลัพธ์ ของการค้นคืนด้วยภาพต้นฉบับ ก.

ค. ผลลัพธ์ ของการค้นคืนด้วย

ภาพที่ 2.1 การประมวลผลแบบค้นคืนด้วยคุณลักษณะฟีเจอร์ระดับต่ำ[1]

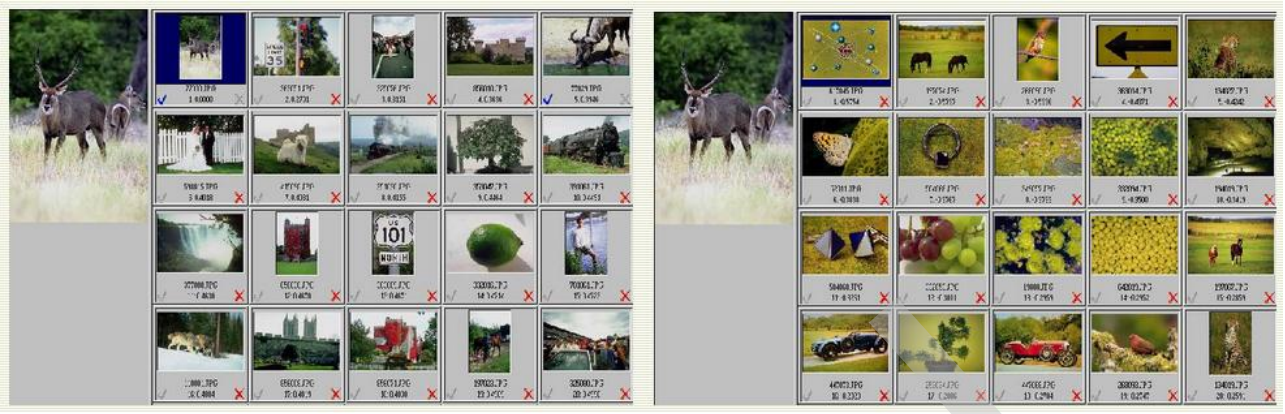
1 อ้างอิง http://alipr.com/cgi-bin/zwang/regionsearch_show.cgi และ http://wang.ist.psu.edu/IMAGE/ ค้นคืนเมื่อวันที่ 19 ส.ค. 2558



ภาพที่ 2.2 ตัวอย่างการประมวลผลแบบค้นคืนด้วยสีและลวดลาย [David Liu 2007] [2]

แต่อย่างไรก็ตามได้มีกลุ่มนักวิจัยที่พยายามปรับปรุงแปลงอัลกอริทึมด้วยการประมวลผลภาพระดับต่ำ เพื่อทำการค้นคืนภาพที่มีลักษณะฟีเจอร์ที่ใกล้เคียงกับภาพที่ต้องการมากที่สุด [M. Flickner, 1995]; [W. Ma, 1997][David Liu 2007] การปรับปรุงเทคนิควิธีการเพื่อให้กระบวนการค้นคืนภาพได้ผลลัพธ์ที่ถูกต้องเพิ่มมากขึ้น ด้วยการนำวิธีการมาผสมผสานกันระหว่างคุณลักษณะเพื่อให้สามารถวิเคราะห์ในรูปแบบที่ซับซ้อนได้มากขึ้น เช่น การรวมเทคนิคด้วยคุณลักษณะสีและรูปทรงของภาพเพื่อทำการค้นคืนภาพ [P.S. Hiremath, 2007] การใช้สีและลวดลายของวัตถุ [David Liu 2007] ดังแสดงผลลัพธ์ในภาพที่ 2.2 หรือมีการใช้อัลกอริทึมเพื่อทำการสกัดข้อมูลภาพเป็นฟีเจอร์ และนำฟีเจอร์นำมาใช้ในการค้นคืนภาพในรูปแบบที่แตกต่างกันออกได้

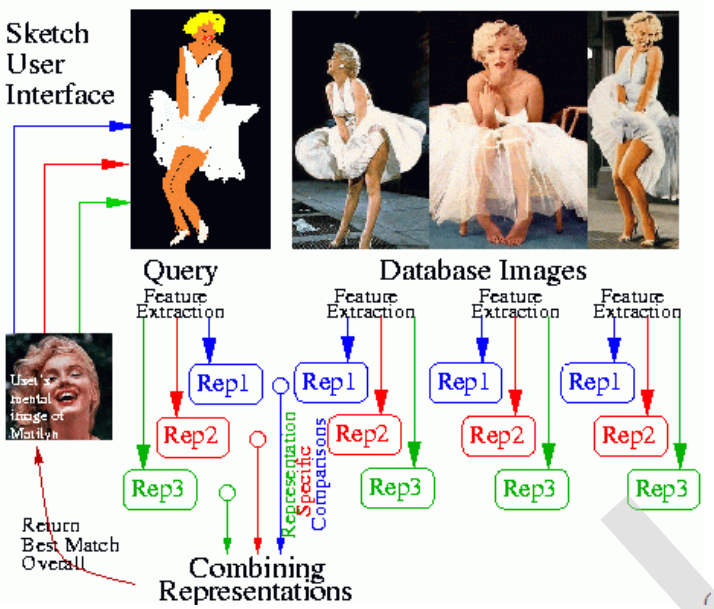
2 อ้างอิง http://iceboundflame.com/projects/content-based-image-retrieval ค้นคืนเมื่อวันที่ 19 ส.ค. 2558



ภาพที่ 2.3 ตัวอย่างการค้นคืนแบบย้อนกลับ[3]

ในบางกลุ่มงานวิจัย [J. Han 2004] ได้มีการใช้การคืนกลับของภาพเข้ามาช่วยเผื่อให้ผลของภาพได้ดีขึ้นและตรงกับจุดประสงค์ของผู้ใช้งานมากขึ้นแต่อย่างไรก็ตามวิธีการของการค้นคืนยังใช้คุณสมบัติของข้อมูลระดับเช่นเดิม ดังแสดงในภาพที่ 2.3 แสดงผลลัพธ์ของการค้นคืน และได้มีการพัฒนาวิธีการป้อนข้อมูลเพื่อใช้ในการค้นคืนเรื่อย ๆ เช่น วาดโครงร่าง (Sketch) [Huda Abdulaali ,2014] [Manolis Kamvysselis,1999] เพื่อใช้เป็นข้อมูลในการค้นคืนภาพ ดังแสดงในภาพที่ 2.4 จะเห็นว่าเมื่อมีการวาดโครงร่าง จะทำการสืบค้นด้วยการแบ่งข้อมูลภาพออกเป็นส่วน ๆ (Segment) ซึ่งในงานวิจัยนี้ได้พยายามแบ่งภาพด้วยสีและใช้การรวมคุณลักษณะต่าง ๆของภาพเข้าด้วยกันเพื่อให้ผลลัพธ์ทที่ได้สมบูรณ์ แต่ในความเป็นจริงแล้วนั้นลักษณะการมองภาพของคนโดยทั่วไปเป็นการมองจากความหมายของภาพ หรือมองจากชนิดของวัตถุของภาพ ตัวอย่างเช่น ภายในภาพประกอบด้วยวัตถุหลายชนิดได้แก่ ทะเล และ ท้องฟ้า เพียงอย่างเดียว แต่บางภาพอาจจะประกอบด้วยวัตถุอื่น ๆ มนุษย์ สุนัข หรืออาคาร เป็นภาพประเภทเดียวกันหรือภาพที่สื่อความหมายอย่างเดียวกัน โดยที่ไม่จำเป็นต้องมีคุณลักษณะสีหรือรูปทรงแบบเดียวกันก็สามารถเป็นภาพชนิดเดียวกันได้ ดังนั้นในการค้นคืนที่ใช้คุณลักษณะฟีเจอร์ระดับต่ำเพียงอย่างเดียว ทำให้ได้ผลลัพธ์ส่วนใหญ่ตรงกับคุณลักษณะของฟีเจอร์ที่สกัดมาแต่ไม่ได้ตรงกับความหมายที่เกิดภายในภาพที่ต้องการอย่างแท้จริง

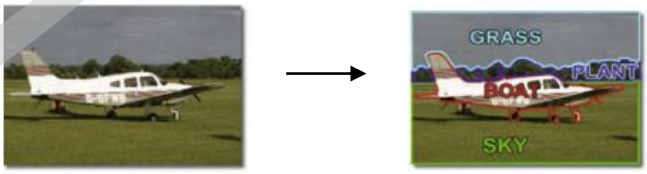
[3] อ้างอิง http://ivp.ee.cuhk.edu.hk/projects/demo/cbir/demo2.html ค้นคืนเมื่อวันที่ 9 ส.ค. 2558



ก. การวาดโครงร่างต้นแบบเพื่อใช้ค้นคืนภาพ ข. การแบ่งภาพ

ภาพที่ 2.4 ตัวอย่างขององค์ประกอบของภาพ[4]

การวิจัยในยุคแรกเป็นความพยายามในการค้นหาข้อมูลภาพเริ่มมาจากการสกัดข้อมูลภายในภาพ (feature extraction) ออกมาเป็นส่วน ๆแล้วใช้ข้อมูลระดับต่ำ เช่น สี (colour) รูปทรง (shape) และ พื้นผิว (texture) เป็นต้น เพื่อทำการสืบค้นข้อมูลภาพ ผลลัพธ์ที่ได้เป็นชุดของภาพที่เกิดจากการเปรียบเทียบจากข้อมูลจากฟีเจอร์ต่าง ๆตามคุณลักษณะของฟีเจอร์ที่สกัดได้เท่านั้น ทำให้ต้องมีการปรับปรุงพัฒนาอย่างต่อเนื่อง ด้วยเหตุนี้เองทำให้มีกลุ่มนักวิจัยที่สนใจในเรื่องของ การจัดกลุ่มภาพตามความหมายเรียกว่า semantic image เป็นการจัดกลุ่มภาพตามความหมายของภาพโดยแปลความหมายของภาพจากองค์ประกอบ หรือวัตถุ (object) ที่ปรากฏในภาพนั้น ซึ่งมีวิธีการค้นคืนที่แตกต่างกันเพื่อให้ได้ผลลัพธ์ที่ตรงตามความต้องการมากที่สุด



ภาพที่ 2.5 ตัวอย่างการแท็กองค์ประกอบของภาพ[5]

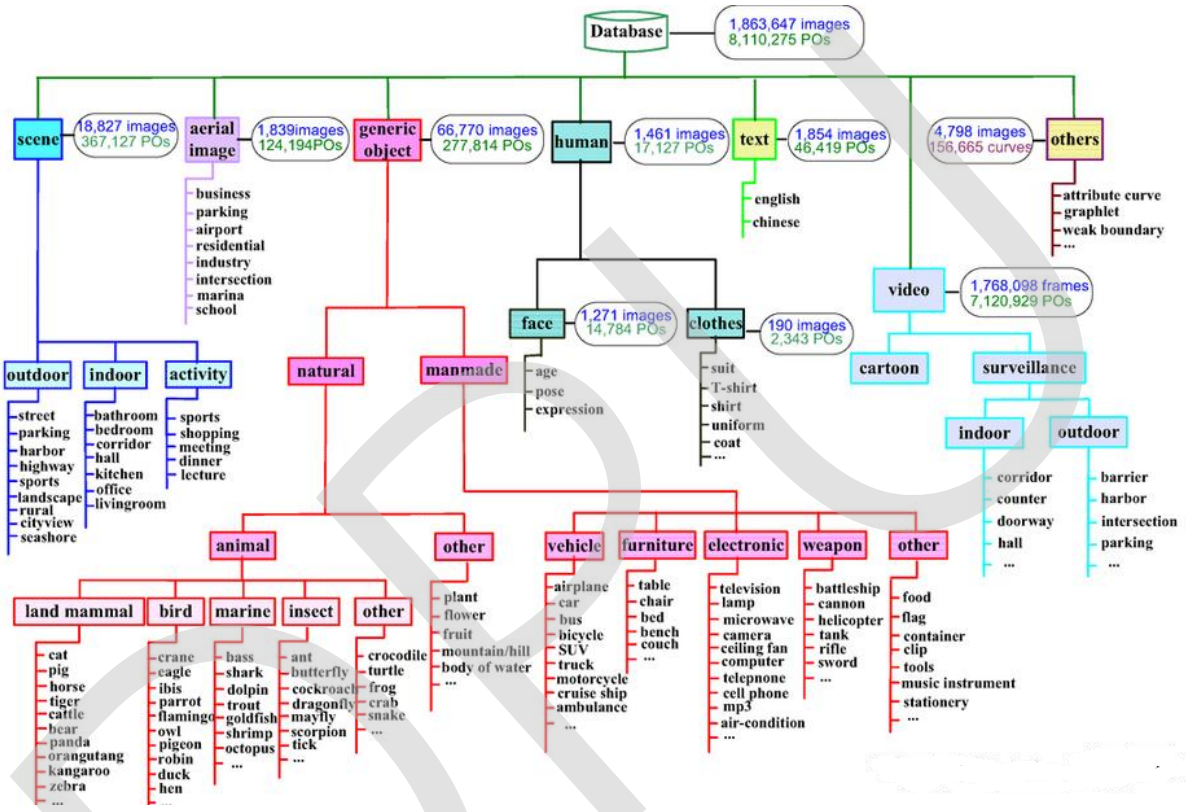
[4] อ้างอิง http://web.mit.edu/manoli/www/imagina/imagina.html ค้นคืนเมื่อวันที่ 9 ส.ค. 2558

[5] อ้างอิง http://vision.ucsd.edu/project/context-based-object-categorization ค้นคืนเมื่อวันที่ 9 ส.ค. 2558

แต่อย่างไรก็ตามได้มีงานวิจัยอีกกลุ่ม ที่พยายามจะใช้เทคนิคของการเข้าใจความหมายของภาพแทน การสืบค้นแบบข้างต้น เรียกว่า การประมวลผลภาพระดับสูง (high level image processing) งานวิจัยในกลุ่มนี้พยายามที่จะมองข้อมูลบนภาพเป็นวัตถุ (object) ที่มีความหมาย [Benitez A.B, 2002];[Galleguillos C., 2010] และแทนวัตถุนั้น ๆ ด้วยคำหลัก (keyword) บนภาพ เรียกว่า การแท็ก (tag) หรือการให้ความหมายของวัตถุบนภาพเป็น ชื่อวัตถุ หรือคำศัพท์ ที่สอดคล้องกันเช่น “grass” , “plant” , “boat”, “sky” เป็นต้น ดังแสดงในภาพที่ 2.5 [Galleguillos C., 2010] และใช้ความหมายหรือคำศัพท์นั้นเพื่อทำการสืบค้นข้อมูลแทน ซึ่งเป็นการใช้ความหมายของคำศัพท์ที่มีความสอดคล้องกันด้วยความหมายตามพจนานุกรม หรือในลักษณะใช้ความสัมพันธ์ของความหมายที่เหมือนกันของคำหลัก (synonym) [Benitez, A.B., 2002];[Kobus B., 2001];[Philippe M., 2002] เข้ามาใช้ในการค้นคืนข้อมูลภาพเพื่อให้ผลลัพธ์ตรงกับความต้องการของผู้ใช้ ยกตัวอย่างเช่น “stone” มีความหมายสอดคล้องกันกับ “rock” เป็นต้น Benjamin Yao [Benjamin Yao 2009] ได้สร้างความสัมพันธ์ของกลุ่มคำศัพท์โดยจัดหมวดหมู่ของคำศัพท์ที่มีความเกี่ยวข้องเข้าด้วยกันดังแสดงในภาพที่ 2.6 และจะได้ผลที่ค่อนข้างดีกว่า แต่ขึ้นอยู่กับว่าอัลกอริทึมที่ถูกนำมาใช้นั้นจะเป็นลักษณะใด

นักวิจัยบางกลุ่มใช้เทคนิคเพื่อสร้างความสัมพันธ์ (relationship) ของแท็กชื่อวัตถุบนภาพ [Benitez A.B, 2001]; Philippe M., 2002] ด้วยโมเดลเหตุการณ์ (event model) [Joo-Hwee L., 2003] เช่น การเชื่อมวัตถุด้วยคำต่าง ๆ เช่น “touch” “on” “top” เป็นต้น บางงานวิจัยได้พยายามใส่ข้อความ (context) เพื่อบรรยายความหมายของภาพ [Mathias Lux, 2003];[Mathias Lux, 2009] ในหัวข้อที่สอดคล้องกับภาพนั้น ๆ เช่น “birthday party of uncle Adam” หรือ “a picture showing a barking dog” แต่ภาพที่นำมาใช้จัดเก็บนั้นมักจะเป็นภาพส่วนตัว (personal images) ทำให้การค้นคืนจำกัดเพราะคำบรรยายส่วนใหญ่จะเป็นคำเฉพาะเจาะจงจึงไม่นิยมเท่าที่ควร บางงานวิจัยพยายามที่จะใส่ข้อมูลเหตุการณ์ต่าง ๆจนครบถ้วนในรูปแบบของ ใคร ทำอะไร ที่ไหน เมื่อไหร่ (who, what, when, where) ดังนั้นในการใส่ข้อมูลบางครั้ง เป็นข้อมูลที่นอกเหนือ หรือเกินความจำเป็นโดยใช่เหตุ ข้อมูลเหล่านี้อาจจะไม่มีความจำเป็นเลยสำหรับการสืบค้นข้อมูล อาจจะเป็นข้อมูลที่มีความเป็นส่วนตัวจนเกินไป และใช้คำศัพท์ซ้ำซ้อน ฟุ่มเฟือย ทำให้การค้นคืนยิ่งเกิดความสับสนปนเป กันของภาพผลลัพธ์ที่ได้มา จากภาพที่ 2.7 มีการนำคำศัพท์หลักที่ถูกแท็กลงในภาพเป็นข้อมูลที่ใช้ในการค้นหาซึ่งผลลัพธ์ของข้อมูลที่ได้มาจะมีความสอดคล้องกันของคำศัพท์ที่ถูกแท็กลงในแต่ละภาพ การค้นหาภาพด้วยเทคนิคนี้จะได้ผลลัพธ์ที่ขึ้นกับคำศัพท์ที่ถูกแท็กไว้บนภาพยิ่งมีการแท็กข้อมูล

บนภาพมากยิ่งสามารถหาความเหมือนกันบนภาพมากขึ้นเท่านั้น แต่ในความเป็นจริงแล้ว การแท็กข้อมูลบนภาพในปัจจุบันนั้นเป็นเพียงการหาคำศัพท์ที่ต้องการบนภาพ แต่ไม่ได้ให้ความหมายภาพโดยรวม ความหมายของภาพคือการนำวัตถุที่ปรากฏบนภาพมารวมกันเพื่อวิเคราะห์จากความคิดของมนุษย์เพื่อให้ได้ผลลัพธ์คือคำศัพท์ใหม่ที่แทนความหมายของภาพทั้งภาพ

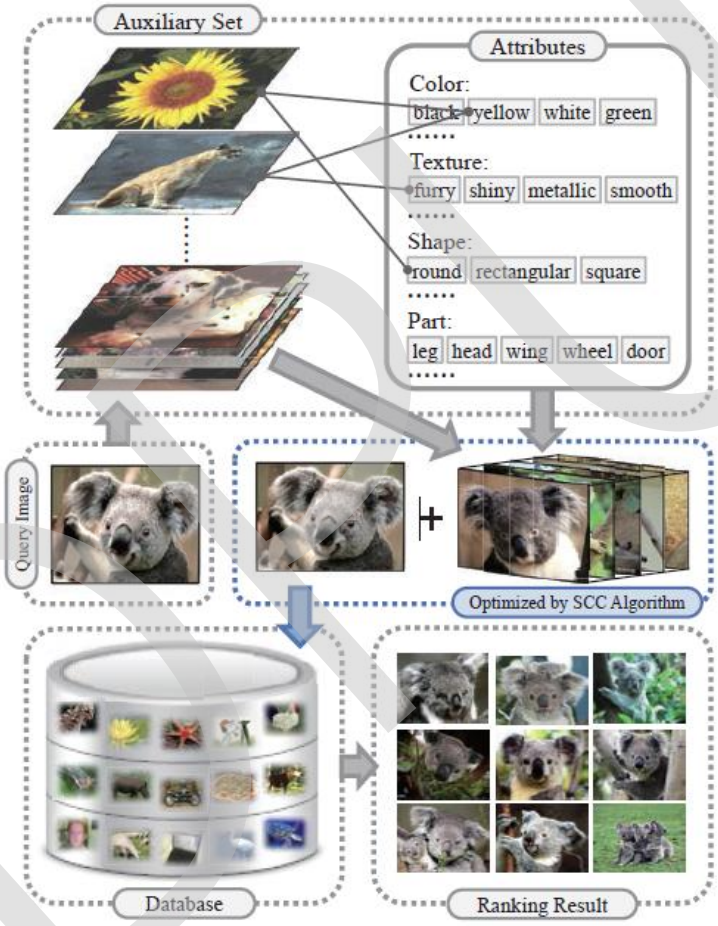


ภาพที่ 2.6 ตัวอย่างโครงสร้างแท็กคำหลักของภาพ[6]

เนื่องจากกลุ่มภาพถ่ายส่วนบุคคลมีจำนวนเพิ่มขึ้นอย่างรวดเร็วดังนั้นข้อมูลภาพถ่ายจึงมีจำนวนเพิ่มขึ้นเช่นกัน ทำให้มีกลุ่มงานวิจัยบางกลุ่มให้ความสนใจในการค้นคืนภาพส่วนบุคคลด้วยวิธีการต่าง ๆ เช่น Wei Yang [Wei Yang,2014] ได้ใช้วิธีการสกัดข้อมูลด้วยการแบ่งส่วนของภาพด้วยสีและพื้นที่ติดกันจากชุดเสื้อผ้าของบุคคลในภาพและทำการรู้จำเสื้อผ้าว่าเป็นประเภทใด เช่น เข็มขัด กางเกง เสื้อชุด และสัดส่วนของร่างกาย เช่น ศีรษะ ผม ผิว เป็นต้น ซึ่งผลลัพธ์และวิธีการแสดงในภาพที่ 2.8 [Nutchanun C.,2013] ใช้ท่าทางของมนุษย์สกัดข้อมูลด้วยค่าของพลังงาน และนำไปผสมรวมกันกับคำศัพท์ที่แท็กมาจากเพื่อจำแนกความหมายของภาพดังแสดงในภาพที่ 2.9

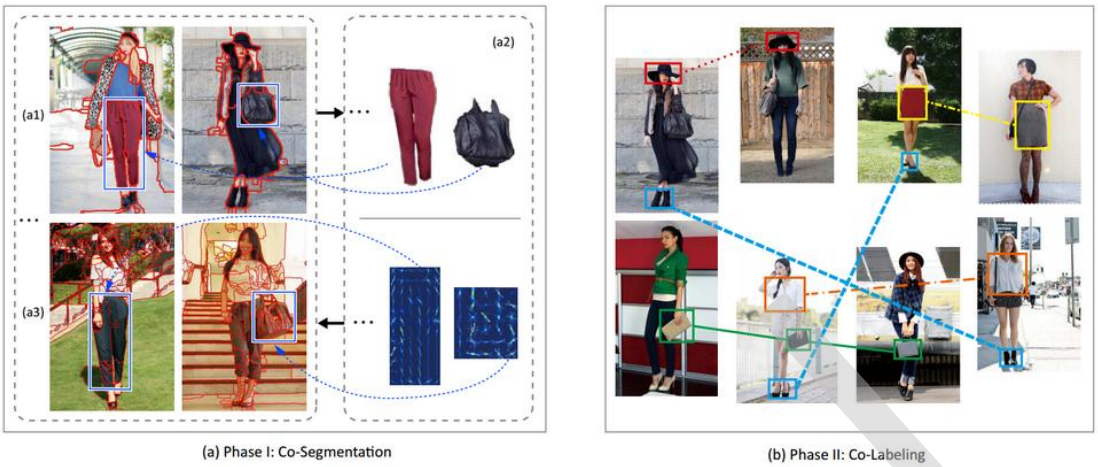
[6] อ้างอิง http://www.stat.ucla.edu/~zyyao/projects/I2T.htm ค้นคืนเมื่อวันที่ 12 ส.ค. 2558

[Chenliang Xu, 2015] ได้กำหนดกลุ่มของคำศัพท์ที่แท็กจากภาพเพื่อจัดเป็นคู่ของคำศัพท์ที่มีความเกี่ยวข้องกันโดยที่มีการเรียนรู้กิจกรรมในกลุ่มของ adult, cat, dog, baby, ball และ car ที่เป็นภาพถูกต้องและภาพที่ไม่เข้ากลุ่มและสามารถจำแนกเป็นกลุ่มของกิจกรรมได้ตามลักษณะของกลุ่ม actor ซึ่งถ้าเป็กลุ่ม adult สามารถจำแนกกิจกรรมได้ดังนี้ climb, crawl, at, jump, roll, run, walk และอื่น ๆ จากผลการทำลองนั้นสามารถจำแนกได้ตามกิจกรรมตาม actor ที่กำหนดได้ แต่อย่างไรก็ตามกิจกรรมที่จำแนกออกมานั้นเป็นเพียงลักษณะของการแสดงออกทางท่าทางเท่านั้น

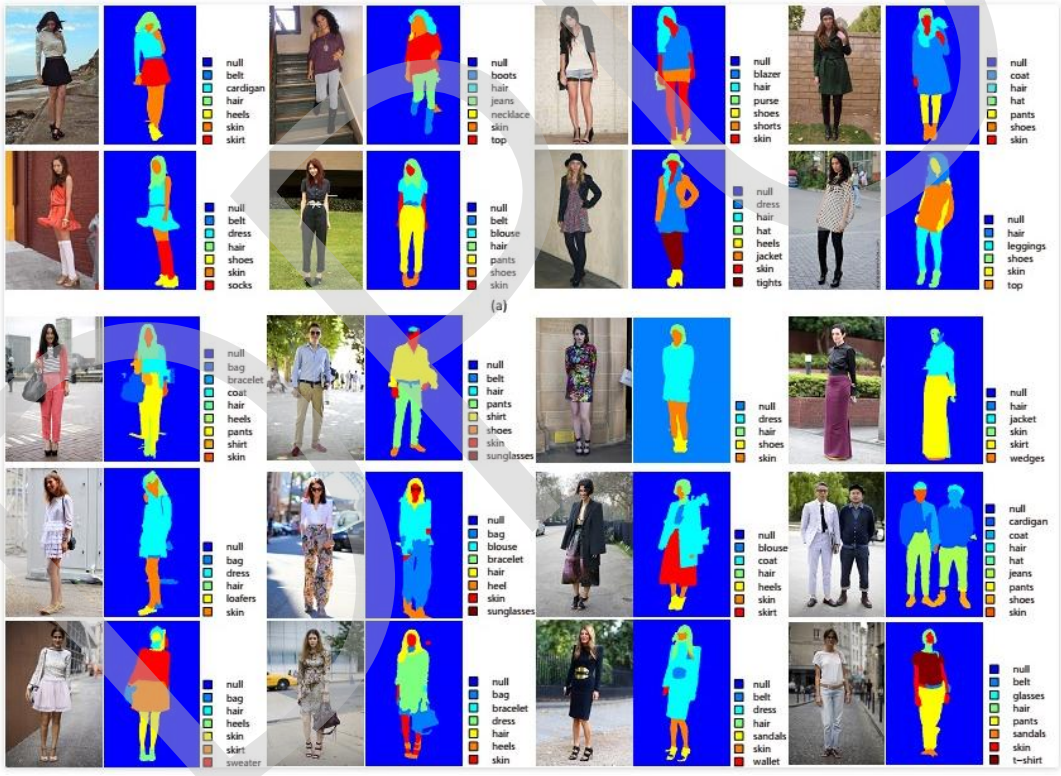


ภาพที่ 2.7 ตัวอย่างงานวิจัยที่ใช้คำศัพท์แท็กลงในภาพ [7]

7 http://lxiongh.com/2014/06/02/ImageRetrieval/ ค้นคืนเมื่อวันที่ 12 ส.ค. 2558



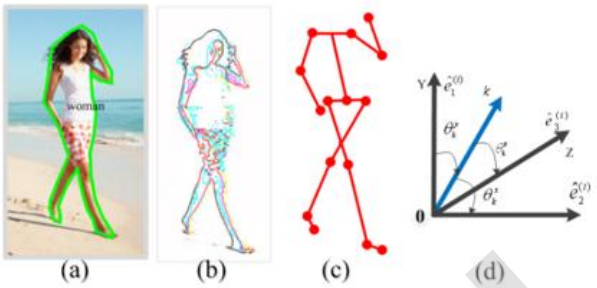
ก. การแบ่งส่วนของภาพด้วยพื้นผิวและสี



ข.ผลลัพธ์ของการรู้จำวัตถุในภาพ

ภาพที่ 2.8 ตัวอย่างการค้นคืนภาพส่วนบุคคล[8]

[8] http://vision.sysu.edu.cn/projects/clothing-co-parsing/ ค้นคืนเมื่อวันที่ 18 ส.ค. 2558



ก.การสกัดข้อมูลพลังงานจากโครงร่างมนุษย์

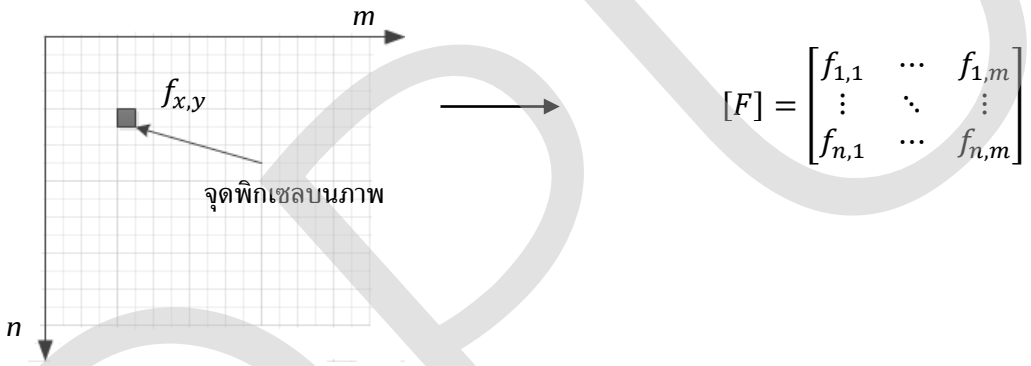
ข.ผลลัพธ์ของการจำแนกท่าทางมนุษย์ด้วยพลังงาน

ภาพที่ 2.9 ตัวอย่างของการจำแนกท่าทางมนุษย์ของภาพส่วนบุคคล

สำหรับการวิเคราะห์ความหมายของภาพจะได้จากการรับรู้ของมนุษย์ที่มองภาพนั้น ภาพยังมีวิธีการที่ผสมผสานระหว่างการใช้แท็กคำหลักบนภาพกับการสร้างความสัมพันธ์ระหว่างวัตถุด้วยมัลติเปิลเคอร์แนล (multiple kernel) ของพื้นที่วัตถุบนภาพ [Galleguillos C., 2011];[Galleguillos C., 2010] ผลลัพธ์ที่ได้จะเป็นกลุ่มของคำหลักที่เกิดขึ้นบนภาพเสียส่วนใหญ่ และมีนักวิจัยบางกลุ่มพยายามที่จะแก้ไขถึงความยุ่งยากลำบากในการใส่ข้อมูลบนภาพ (image annotation) และจำกัดขอบเขตคำศัพท์ของภาพให้รัดกุมยิ่งขึ้น ปัจจุบันการพัฒนาซอฟต์แวร์เพื่อให้เกิดความสะดวกสบายในการแท็กข้อมูล และข้อมูลที่ถูกนำมาเป็นคำศัพท์นั้นแบบแผนและโครงสร้างที่แน่นอน ยกตัวอย่าง เช่น Caliph & Emir [Mathias Lux, 2009], Annosearch [Xin-Jing, 2008], CAMEL

[Apostol Paul N., 2001] เป็นต้น ส่วนใหญ่จะมีความสนใจในการแทนค่าความหมาย (Semantic representation) โดยขึ้นกับภาษาและโครงสร้างการแทนค่าเพื่อทำให้มนุษย์ และเครื่องคอมพิวเตอร์เกิดความเข้าใจในการประมวลผล มีนักวิจัยหลายกลุ่มพยายามที่จะแก้ไข และใช้หลากหลายวิธีเพื่อที่จะเชื่อมโยงให้ได้ความหมายของภาพส่วนบุคคลที่แสดงถึงความหมายของภาพที่สอดคล้องกับกิจกรรมอย่างแท้จริง

ดังนั้นในงานวิจัยนี้จึงนำเสนอการจัดแบ่งกลุ่มภาพส่วนบุคคลจากวัตถุภายในภาพที่แท็กเป็นคำศัพท์ที่ถูกกำหนดไว้ลงในแต่ละกลุ่มตามเหตุการณ์กิจกรรม (activity events) ของมนุษย์ เพื่อช่วยให้การจัดแบ่งกลุ่มได้ตรงตามความหมายของภาพมากขึ้น และทำเปรียบเทียบความเหมือนกันของความหมายภาพด้วยการขั้นตอนวิธีการจัดกลุ่มกิจกรรม (clustering algorithm)



ภาพที่ 2.10 การแสดงตำแหน่งของพิกเซลบนภาพดิจิทัลด้วยเมทริกซ์

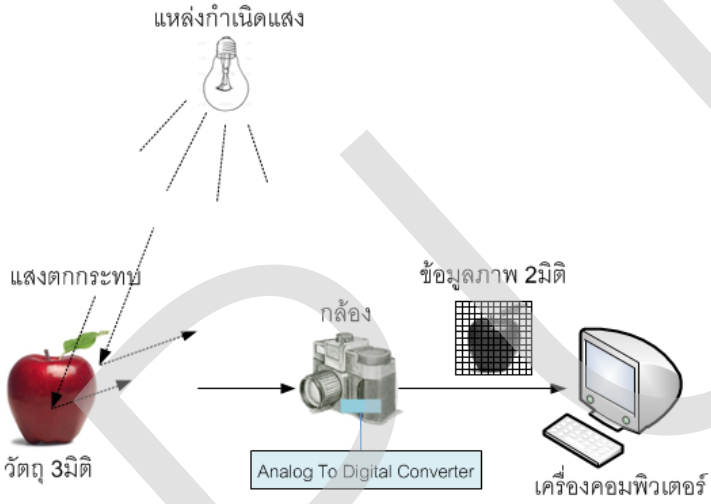
2.2 การประมวลผลภาพดิจิทัล

ภาพดิจิทัล (Digital Image) [R. C. Gonzalez, 2014] คือการเก็บบันทึกหรือแสดงผลในสิ่งที่เห็นหรือรับรู้ จะแสดงการบันทึกออกมาในลักษณะสองมิติ (2 Dimension) ในหน่วยที่เรียกว่าพิกเซล (pixel) ดังนั้นสามารถนิยามภาพดิจิทัลในรูปแบบของฟังก์ชัน สองมิติได้เป็น $f(x, y)$ โดยที่ x และ yเป็นตำแหน่งพิกัดของภาพ ดังแสดงในภาพที่ 2.10

2.2.1 การได้มาของภาพดิจิทัล

ภาพดิจิทัลจะได้จากการถ่ายภาพวัตถุ ซึ่งวัตถุที่อยู่ในโลกของความเป็นจริงสามารถจับต้องได้คือวัตถุในรูปแบบสามมิติ ดังนั้นภาพผ่านเลนส์เข้าสู่อุปกรณ์อิเลคทรอนิคส์ที่มีการรับรู้ทางด้านแสงโดยอุปกรณ์นี้จะเป็นชิ้นส่วนของกล้องถ่ายรูปที่จับแสงตกกระทบวัตถุ และสะท้อนกลับออกมาผ่านตัว

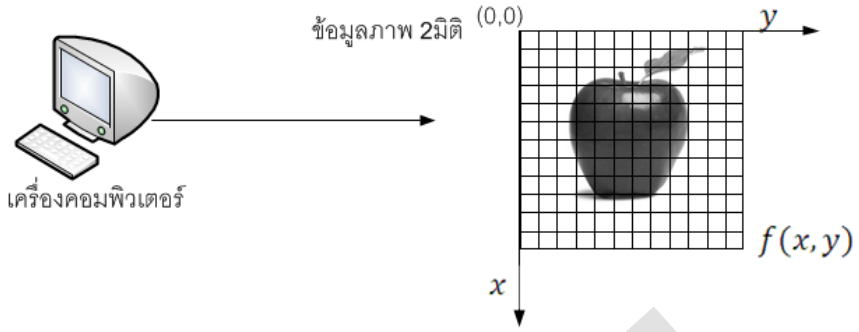
แปลงสัญญาณแอนะล็อกเป็นตัวเลข (Analog To Digital Converter) เพื่อทำการบันทึกเป็นข้อมูลภาพที่อยู่ในรูปแบบสองมิติ (ด้านเดียวของวัตถุ) ค่าที่ถูกบันทึกเป็นค่าความเข้มของแสง (Intensity) โดยปกติแล้วจะมีการแปลงเพื่อให้สอดคล้องกับการทำงานของระบบฮาร์ดแวร์ของเครื่องคอมพิวเตอร์ได้หลายขนาด เช่น 8บิต 16บิต 24บิต หรือ 36บิต ขึ้นกับการแสดงโทนสีของภาพและทำการจัดเก็บข้อมูลภาพลงในหน่วยความจำของคอมพิวเตอร์ ดังแสดงการจัดเก็บภาพดิจิทัลในรูปที่ 2.11



รูปที่ 2.11 การจัดเก็บข้อมูลภาพดิจิทัล

2.2.2 แบบจำลองภาพดิจิทัล

เมื่อภาพดิจิทัลถูกจัดเก็บลงในเครื่องคอมพิวเตอร์จะมีการจองหน่วยความจำในรูปของตัวแปร อะเรย์ (Array) ที่มีการกำหนดให้ตำแหน่ง (0,0) เป็นจุดกำเนิด (Coordinate Origin) จะอยู่ทางด้านซ้ายมือสุดของสเกล หรือด้านบนของ การจัดลำดับตำแหน่งของจุดภาพจะเรียงจากซ้ายไปขวาและจากบนลงล่างแต่ละเส้นจุด เรียกว่าพิกเซล (Pixel) ภายในพิกเซลจะถูกเก็บค่าความเข้มแสง (ค่าของสี) และสามารถเขียนให้อยู่ในรูปของฟังชันก์ของภาพสามารถแสดงในรูปสมการได้ $i = f(x, y)$ โดย (x, y) แทนพิกัดในแบบจำลองของภาพ (Image model) และ i แทนค่าความสว่างหรือความเข้มของแสง ดังแสดงในรูปที่ 2.12



รูปที่ 2.12 การจัดเก็บข้อมูลภาพดิจิทัล

โดยส่วนใหญ่แล้วการจัดเก็บข้อมูลภาพลงบนเครื่องคอมพิวเตอร์ที่มีลักษณะเป็นอะเรย์จะถูกแทนค่าเพื่อทำการประมวลผลบนเครื่องคอมพิวเตอร์ในรูปแบบของเมตริกซ์ (Matrix)กำหนดให้ภาพในรูปฟังก์ชัน $f(x,y)$ ที่มีขนาด $m \times n$ โดยที่ m แทน แถว และ n แทน คอลัมน์ สามารถเขียนให้อยู่ในรูปของเมตริกซ์ได้ดังนี้

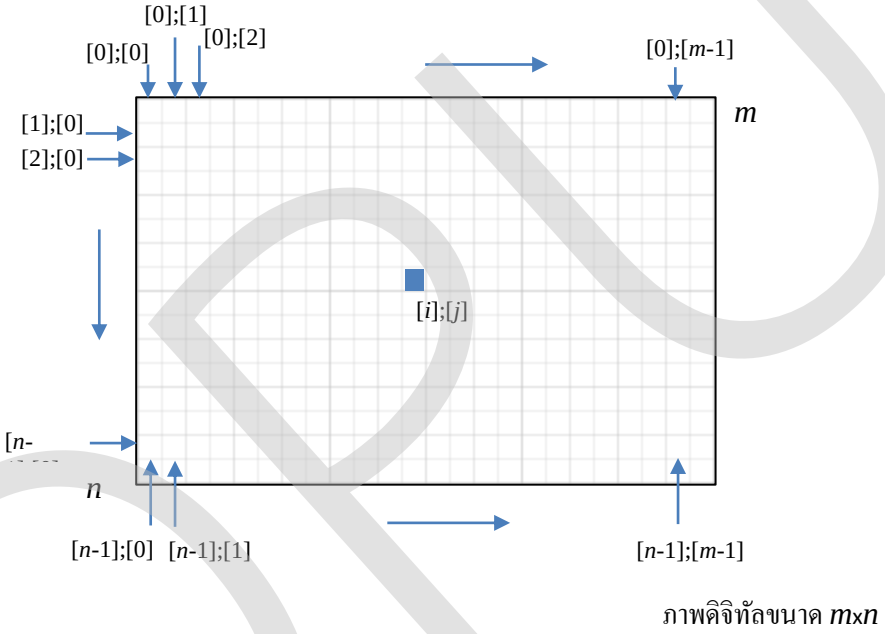
$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,n-1) \\ f(1,0) & f(1,1) & & f(1,n-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(m-1,0) & f(m-1,1) & \cdots & f(m-1,n-1) \end{bmatrix}$$

เมื่อ x, y เป็นพิกัดใด ๆภายในภาพและแอมพลิจูดของ f คือค่าความเข้มแสงของภาพที่พิกัดใด ๆเป็นค่าจำกัด (Finite value) จึงเรียกรูปภาพนี้ว่า ภาพดิจิทัล (Digital Image)

ในหน่วยความจำ จะทำการจัดเนื้อที่ในการเก็บภาพ สามารถคำนวณได้จาก $m \times n \times b$ เมื่อ b เป็นจำนวนเต็มที่แทนจำนวนบิตของข้อมูลในแต่ละจุดภาพ ตัวอย่างถ้า b มีค่าเท่ากับ 8 บิต จะสามารถเก็บความแตกต่างของระดับสีที่เป็นไปสูงสุด 256 ระดับ ค่า m และ n จะเป็นตัวบอกถึงความละเอียดของภาพ สำหรับคอมพิวเตอร์ทั่วไปในระบบ VGA (Video Graphic Array) จะมีขนาด 640×480, 800×600 และ 1024×768 จุด เป็นต้น การกำหนดความละเอียดจะขึ้นอยู่กับงานที่จะใช้ ในงานบางอย่างใช้ความละเอียดเพียง 30×50 จุด ก็พอซึ่งความละเอียดนั้นจะขึ้นกับงานที่จะใช้ ในบางงานจะใช้ความละเอียดถึง 1000×1000 จุด ก็ยังไม่พอ จากภาพที่ 2.13 สมมติให้ภาพแทนเป็นตัวแปรชื่อ x เป็นตัวแปรแบบอะเรย์ขนาด $m \times n$ (m แทน แถว และ n แทน คอลัมน์) ที่ใช้เก็บภาพขนาด $m \times n$ จุด และ ค่าของความเข้มของแสง (ค่าความสว่าง) ของจุดภาพในแถวที่ 5 คอลัมน์ที่ 4 จะตรงกับค่าของข้อมูล x เป็นรูปของ (5,4) จะเห็นว่าใช้ตำแหน่งของจุดภาพทั้งสองแกน

เป็นตัวชี้ค่าข้อมูลในอะเรย์ ดังแสดงการเรียงตัวของข้อมูลบนอะเรย์ในภาพด้านบน และการกำหนดความละเอียดของภาพ (image resolution) จากการกำหนดขนาดของพิกเซลตัวอย่างเช่น 1 ไมครอนต่อพิกเซล (µm/pix) 1 มิลลิเมตรต่อพิกเซล (mm/pix) เป็นต้น ในงานที่ต้องการทราบตำแหน่งหรือขนาดของวัตถุที่วัดเป็นค่าจริง เราสามารถที่จะคำนวณได้จาก

$$Resolution = \frac{\text{Field of vision in Y direction (mm)}}{\text{Number of pixels in Y direction}}$$



ภาพที่ 2.13 การจัดเก็บและเรียกใช้ข้อมูลภาพดิจิทัลลงในเมทริกซ์

โดยปกติแล้วในการเก็บข้อมูลภาพโดยเครื่องมือต่าง ๆ จะเก็บตามมาตรฐานของโทรทัศน์ซึ่งมีอัตราส่วน x ต่อ y เท่ากับ 4:3 สำหรับเครื่องมือเก็บข้อมูลภาพที่ไม่เป็นไปตามอัตราส่วน 4:3 เมื่อนำภาพนี้ไปแสดงในจอภาพมาตรฐาน จะทำให้ภาพที่แสดงนั้นมีขนาดของจุดภาพไม่เป็นสี่เหลี่ยมจัตุรัส เช่น ในบางระบบอาจจะใช้ความละเอียดในการแสดง เท่ากับ 640×580 ซึ่งจะทำให้ขนาดของจุดภาพที่ได้มีขนาดของด้านกว้างมีความยาวมากกว่าด้านสูง เมื่อมีการกำหนดให้ขนาดของบิตต่อจุดมากขึ้น จะทำให้จำนวนของสีมากขึ้นด้วย ตัวอย่างเช่น 1 บิต = 21 จะได้ 4 สี 2 บิต = 22 จะได้ 4 สี 4 บิต = 24 จะได้ 16 สี 8 บิต = 28 จะได้ 256 สี 16 บิต = 216 จะได้ 65536 สี เป็นต้น

จากคุณลักษณะของภาพดิจิทัลที่กล่าวมาข้างต้นนั้น เป็นการเก็บข้อมูลภาพเป็นแบบ เมตริกซ์ที่มีถึง 3 ระนาบ (dimension) การประมวลผลภาพดิจิทัล (digital image processing) เป็นการเรียกใช้ขั้นตอนหรือกระบวนการที่มากระทำบนภาพ โดยมีวัตถุประสงค์เพื่อปรับปรุงคุณภาพของภาพ ให้ได้ภาพใหม่ที่มีคุณสมบัติตามวัตถุประสงค์ที่นำไปใช้งาน เช่น การปรับให้ภาพมีความคมชัดมากขึ้น (enhancement) หรือการบีบอัดข้อมูลภาพ (compression) เพื่อประหยัดพื้นที่ในการจัดเก็บข้อมูล หรือการฝังลายน้ำ (watermark) เพื่อป้องกันการลักลอบการใช้ภาพที่ไม่ได้รับอนุญาตเป็นต้น สำหรับการประมวลผลภาพระดับสูงด้วยคอมพิวเตอร์ สามารถทำได้โดย นำภาพที่ได้มาจากกล้องหรือ image source ต่างๆ ซึ่งเป็นสัญญาณอนาล็อก แล้วนำมาแปลงเป็นสัญญาณดิจิตอลที่มีลักษณะเป็นรหัสเชิงตัวเลขฐานสอง (binary) ประกอบด้วยตัวเลข 0 และ 1 ที่สามารถใช้รูปแบบทางคณิตศาสตร์เข้ามาช่วยในการคำนวณและการประมวลผลข้อมูลภาพ การดำเนินงานวิจัยสำหรับการจัดกลุ่มความหมายของภาพได้มีผู้วิจัยจำนวนมากที่ศึกษาและทำการทดลองโดยทั่วไป แบ่งเป็น 2 ระดับ [A. Gupta, 1997] ดังนี้

2.2.3 การประมวลภาพเบื้องต้น

สำหรับขั้นตอนการประมวลภาพเบื้องต้น (image preprocessing) ในรูปแบบของการหาบริเวณที่ต้องการในรูปแบบอัตโนมัติ ยังคงเป็นงานวิจัยที่ยังหาข้อยุติไม่ได้ โดยเฉพาะที่เป็นลักษณะของ ระบบเรียลไทม์ (real time) ด้วยแล้วนั้นจะต้องคำนึงถึงเวลาในการคำนวณของอัลกอริทึม (computational cost of algorithm) ที่เป็นสิ่งที่จำเป็นค่อนข้างมาก สำหรับระบบเรียลไทม์ ที่มีการแปรเปลี่ยนรูปแบบของพื้นหลัง (modeling of background) จะทำให้เกิดกระบวนการในการสกัดวัตถุที่ต้องการขึ้นมา ซึ่งเทคนิคที่เรียกว่า Gaussians Mixture Model [Chris Stauffer, 2000] เป็นเทคนิคที่ค่อนข้างประสบความสำเร็จมาก แต่อย่างไรก็ตามยังคงต้องมีส่วนภาพพื้นหลังที่ไม่มีการเปลี่ยนแปลงมากนัก เพราะฉะนั้นปัญหาที่เกิดขึ้นยังคงเกิดขึ้นเมื่อพื้นหลังมีการแปรเปลี่ยนและยังคงเป็นปัญหาที่ยังคงต้องมีการแก้ไข ฉะนั้นในงานวิจัยนี้จึงข้ามในส่วนการวิเคราะห์การแปรเปลี่ยนของพื้นหลังที่ไม่คงที่ และได้มีการนำเทคนิคของฮีสโตแกรม (Histogram matching) [F.Porikli, 2005]; [D.Comaniciu, 2003];[D. Comaniciu, 2000] ที่ไม่ได้มีการนำส่วนของพื้นหลังเข้ามาเกี่ยวข้อง เพื่อทำการเปรียบเทียบแต่อย่างไรก็ตาม ด้วยคุณลักษณะของเทคนิคฮีสโตแกรมไม่สามารถที่จะรองรับตำแหน่งของพิกเซล (pixel location) ทำให้ยังคงมีปัญหาเมื่อมีวัตถุที่ทับซ้อนกัน จนไม่สามารถที่จะประมวลผลได้อย่างถูกต้อง เทคนิคสหสัมพันธ์ (Correlation) [A.J. Lipton, 199จะมี

การทำงานที่ไม่สูญเสียในข้อมูลส่วนของสเปเชียลทำให้สามารถทดแทนวิธีการของฮีสโตแกรมได้ และได้มีการใช้เทคนิคสหสัมพันธ์เพื่อทำการหาขอบของวัตถุซึ่งวิธีการนี้ถูกเรียกว่า Edge-Enhanced Normalized Correlation (EENC) [Javed Ahmed, 2008] เพื่อทำการแก้ไขปัญหาต่าง ๆ ที่เกิดจากการหาวัตถุที่เกิดจากการทับซ้อนกัน หรือมีปัญหาจากสัญญาณรบกวน หรือวัตถุที่มีการเปลี่ยนแปลงทิศทางจากการหมุน เป็นต้น และการใช้เทคนิค EENC สามารถทำงานได้ดีสำหรับการแมชชิ่งส่วนของพื้นที่ (matching region) พร้อมทั้งยังเป็นแทมเพลต (template) ที่ใช้ทำงานได้อย่างรวดเร็ว

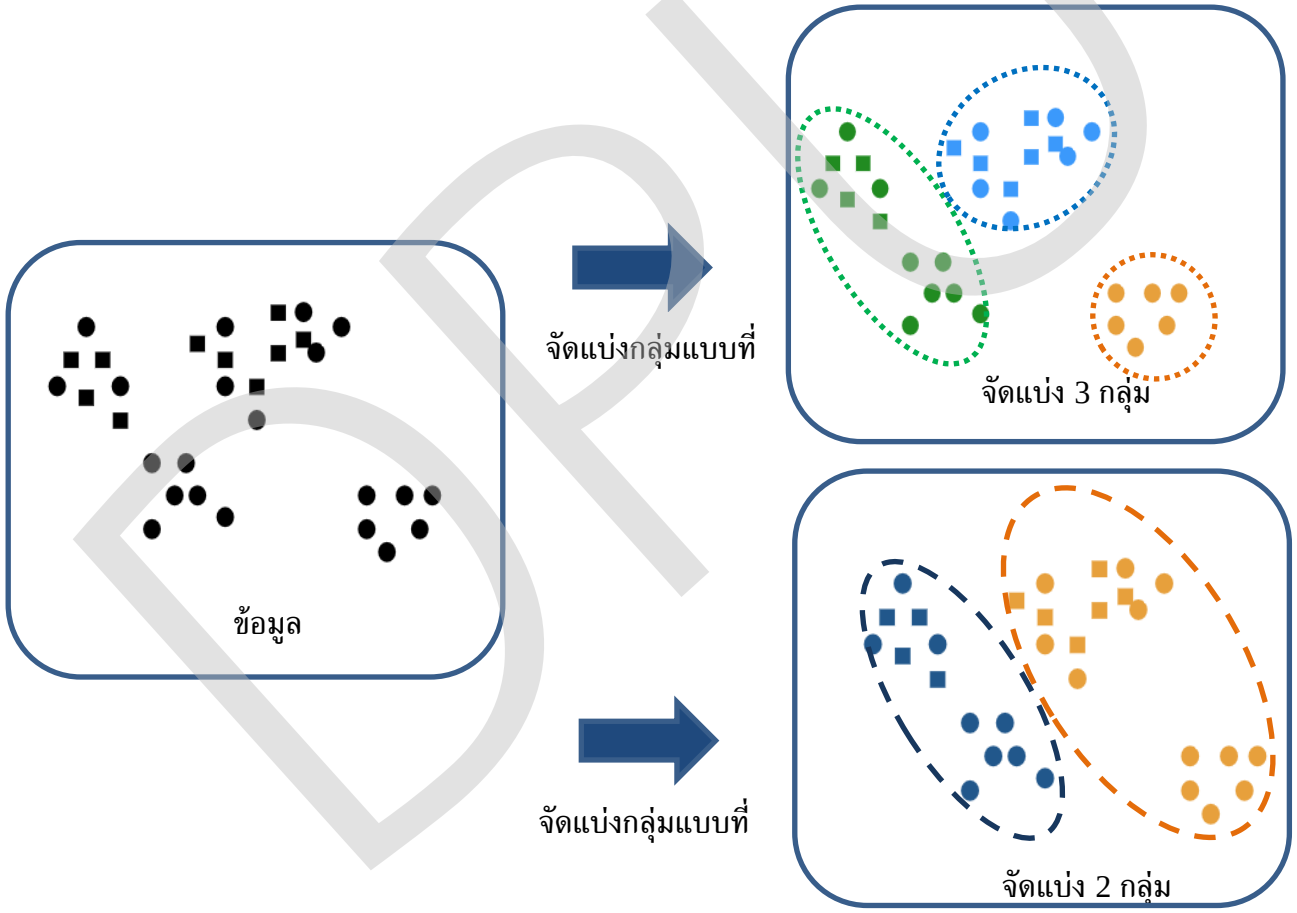
2.2.4 การประมวลผลภาพระดับสูง

การประมวลผลภาพระดับสูง (high-level image processing) เป็นการใช้ข้อมูลฟีเจอร์หรือผลลัพธ์จากการประมวลผลข้างต้นเพื่อผ่านกระบวนการ หรืออัลกอริทึมทำให้คอมพิวเตอร์รู้จักและเข้าใจภาพ (image understanding) รวมทั้งการหาความหมายของภาพ (semantic image) ได้ เพราะฉะนั้นการประมวลผลภาพระดับสูงจำเป็นต้องใช้ข้อมูลที่ได้มาจากการประมวลผลภาพระดับต่ำ ดังนั้นจะเห็นได้ว่าการประมวลผลภาพระดับต่ำมีความสำคัญมากสำหรับการทำให้คอมพิวเตอร์รู้จักและเข้าใจภาพได้ โดยส่วนใหญ่กลุ่มนักวิจัยพยายามที่จะข้ามขั้นตอนการประมวลผลภาพระดับต่ำไป เนื่องจากยังไม่มีความสมบูรณ์เท่าที่ควรและพยายามที่เข้าใช้การประมวลผลภาพระดับสูงอย่างเดียว แต่อย่างไรก็ตามการวิจัยทั้งสองกลุ่มนี้ยังคงมีการพัฒนาการวิจัยอย่างต่อเนื่องเพื่อนำผลลัพธ์ หรือฟีเจอร์ต่าง ๆ เข้ามาทำการค้นคืนภาพ (image retrieval) หรือ การจำแนกข้อมูลภาพ (image classification) รวมทั้งการแยกแยะความหมายของภาพ (semantic image classification) ซึ่งเป็นหัวข้อหลักในการทำวิจัยครั้งนี้

2.3 การจัดกลุ่มข้อมูลภาพ

การจัดกลุ่ม (Clustering) เป็นวิธีการที่พิจารณาข้อมูลแต่ละแถวเสมือนเป็นวัตถุ (object) ซึ่งจะมีหลักการเหมือนกับการจำแนกประเภทข้อมูล คือจะทำการแบ่งข้อมูลออกเป็นกลุ่ม (คลัสเตอร์) โดยจะจัดให้ข้อมูลที่มีความคล้ายคลึงกันอยู่ในคลัสเตอร์เดียวกัน และข้อมูลที่อยู่ต่างคลัสเตอร์กันจะมีความคล้ายคลึงกันน้อยที่สุด ซึ่งความเหมือนหรือต่างกันสามารถเปรียบเทียบได้กับความ

ใกล้ชิดกันของวัตถุใด ๆ โดยใช้ระยะทางเป็นตัวชี้วัด คุณภาพของแต่ละคลัสเตอร์สามารถอธิบายได้จากเส้นผ่านศูนย์กลางของคลัสเตอร์ (diameter) ซึ่งแสดงระยะห่างมากสุดของวัตถุสองชิ้นที่อยู่ในคลัสเตอร์เดียวกันแต่ละคลัสเตอร์จะมีตัวแทนที่สามารถแทนวัตถุทุกชิ้นของคลัสเตอร์นั้นได้ เช่น การใช้จุดศูนย์กลางคลัสเตอร์(centroid) แทนคลัสเตอร์นั้นสำหรับบางเทคนิคตัวแทนของคลัสเตอร์อาจมีได้หลายตัวแทน ทั้งนี้ขึ้นอยู่กับความเหมาะสมของแต่ละเทคนิคที่เลือกใช้ เทคนิคการจัดกลุ่มหรือการจำแนกเพื่อแบ่งกลุ่มตั้งแต่ 2 กลุ่มขึ้นไปเพื่อให้แต่ละกลุ่มมีคุณลักษณะที่เหมือนหรือคล้ายกันมากที่สุดนั้น จะต้องมีการพิจารณาเลือกลักษณะหรือตัวแปรที่จะนำมาใช้ในการแบ่งกลุ่มเพื่อให้ได้ผลลัพธ์หรือกลุ่มที่ต้องการ ทฤษฎีพื้นฐานของการจัดแบ่งกลุ่มมีดังนี้



ภาพที่ 2.14 แบบจำลองการจัดแบ่งกลุ่ม

2.3.1 การจัดกลุ่มแบบ K-means

การจัดกลุ่มแบบ K-means Clustering (K-means Clustering) [Zhang, Hsu and Dayal, 1999] เป็นการจัดกลุ่มด้วย อัลกอริทึม การเรียนรู้โดยไม่มีผู้สอนที่ง่ายที่สุด โดยอัลกอริทึม $K-means$ จะตัดแบ่ง (Partition) วัตถุออกเป็น Kกลุ่ม โดยแทนแต่ละกลุ่มด้วยค่าเฉลี่ยของกลุ่ม ซึ่งใช้เป็นจุดศูนย์กลาง (centroid) ของกลุ่มในการวัดระยะห่างของข้อมูลในกลุ่มเดียวกัน ในขั้นแรกของการจัดกลุ่มโดยการหาค่าเฉลี่ยแบบเคย์ต้องกำหนดจำนวนกลุ่มที่ต้องการ และกำหนดจุดศูนย์กลางเริ่มต้นจำนวน Kด สิ่งสำคัญในการกำหนดจุดศูนย์กลางเริ่มต้นของแต่ละกลุ่มนี้ ควรจะถูกกำหนดด้วยวิธีที่เหมาะสม เพราะตำแหน่งจุดศูนย์กลางเริ่มต้นที่แตกต่างกันทำให้ได้ผลลัพธ์สุดท้ายแตก ต่างกัน ดังนั้นในทางที่ดีควรจะกำหนดจุดศูนย์กลางนี้ให้หางจากจุดศูนย์กลางอื่น ๆ ขั้นตอนต่อไปคือสร้างกลุ่มข้อมูลและความสัมพันธ์กับจุดศูนย์กลางที่ใกล้มากที่สุด โดยแต่ละจุดจะถูกกำหนดไปยังจุดศูนย์กลางที่ใกล้เคียงที่สุดจนครบหมดทุกจุด และคำนวณจุดศูนย์กลางใหม่ โดยการหาค่าเฉลี่ยทุกวัตถุที่อยู่ในกลุ่ม หากจุดศูนย์กลางในแต่ละกลุ่มถูกเปลี่ยนตำแหน่ง จะได้จุดมีความสัมพันธ์กับกลุ่มใหม่และใกล้กับจุดศูนย์กลางใหม่ ทำซ้ำแบบนี้ไปเรื่อย ๆ ผลลัพธ์จากการทำซ้ำแบบนี้ทำให้จุดศูนย์กลางเปลี่ยนตำแหน่งทุกรอบ จนกระทั่งจุดศูนย์กลางจำนวน Kจุด ไม่มีการเปลี่ยนแปลงจึงจะสิ้นสุดกระบวนการ ทำให้ข้อมูลถูกจัดกลุ่มอย่างสมบูรณ์และข้อมูลไม่มีการเปลี่ยนกลุ่มอีกต่อไป ดังนั้นค่าเฉลี่ยของข้อมูลที่ถูกจัดให้อยู่ในคลัสเตอร์เดียวกันเป็นตัวแทนของทุกข้อมูลในคลัสเตอร์นั้น

อัลกอริทึมการจัดกลุ่มแบบ $K-means$

ขั้นตอนที่ 1 กำหนดจำนวนกลุ่ม โดยการสุ่มค่ากลางมาจำนวน j ชุด แทนด้วย C_j เมื่อ j แทนจำนวนกลุ่ม

ขั้นตอนที่ 2 นำวัตถุทั้งหมดจัดเข้ากลุ่มที่มีจุดศูนย์กลางที่อยู่ใกล้วัตถุนั้นมากที่สุด โดยคำนวณจากการวัดระยะห่างระหว่างจุดที่น้อยที่สุด โดยทำการคำนวณหาค่าระยะทางแบบยูคลิเดียน (Euclidean Distance)ระหว่างชุดข้อมูลเข้า (X_j)กับค่ากลาง (C_k) แล้วนำข้อมูล X_jนั้นไปใส่ไว้เป็นสมาชิกในกลุ่ม j ที่มีระยะทางที่น้อยที่สุดจะเป็นกลุ่มที่ชนะ

$$j = argmin\|x_j - c_j\|$$

ขั้นตอนที่ 3 คำนวณจุดศูนย์กลาง Kจุดใหม่ โดยหาจากค่าเฉลี่ยทุกวัตถุที่อยู่ในกลุ่ม โดยทำการคำนวณค่ากลางสำหรับแต่ละกลุ่มใหม่ โดยเฉลี่ยจากข้อมูลที่เป็นสมาชิกของกลุ่มนั้นทุก c_j

$$c_j = \frac{1}{q_j}\sum_{x_j \in \theta_j} x_i$$

ขั้นตอนที่ 4 คำนวณหาค่าระยะทางรวมทั้งหมดแทนด้วย D โดยเป็นค่าระยะทางระหว่างค่ากลางของกลุ่มกับสมาชิกของกลุ่มนั้น ๆ เมื่อให้ I_{ij} คือ ค่าความเป็นสมาชิกของข้อมูล ต่อกลุ่มที่ j ดังนี้

$$D = \sum_{i=1}^{N} \frac{k}{\sum_{j=1}^{k} 1/d_{ij}}$$

เมื่อกำหนดให้ $I_{ij} = 1$, x_iเป็นสมาชิกในθ_j

ขั้นตอนที่ 5 ทำซ้ำ (2-4) จนกระทั่งเงื่อนไข ระยะทางรวมไม่มีการเปลี่ยนแปลงแล้วหรือเปลี่ยนแปลงน้อยมาก

$$1 - \frac{D^{old}}{D^{new}} \leq \propto \text{ เมื่อ } \propto \text{ มีค่าน้อยมาก}$$

- ข้อเสียของทฤษฎี $K - means$

ปัญหาที่สำคัญ คือ การระบุจำนวนกลุ่มของ $K - means$ ว่าต้องการให้แบ่งกี่กลุ่ม จึงจะได้คำตอบที่เหมาะสมที่สุด เป็นเรื่องทำได้ยากเมื่อผู้ใช้ไม่มีความรู้พื้นฐานเกี่ยวกับข้อมูลที่ใช้ การระบุค่า Kที่ไม่เหมาะสมอาจทำให้กระบวนการจัดกลุ่มข้อมูลใช้เวลานานขึ้น แต่อาจจะได้คลัสเตอร์ที่ไม่มีคุณภาพได้ ในงานวิจัยของผมนั้นเลือกวิธีแก้ไขปัญหานี้โดยการทดลองให้ค่า K เปลี่ยนไปเรื่อย ๆ กับข้อมูลเรียนรู้ (Training dataset) และเมื่อได้จุดที่ให้ค่าความถูกต้องสูงที่สุดจึงนำค่านั้นมาใช้กับข้อมูลทดสอบ (Test dataset) ซึ่งวิธีนี้ก็ยังมีจุดอ่อน คือ ถ้าเป็นกรณีที่ข้อมูลของเราไม่มีข้อมูลเรียนรู้เลย ทำให้เราไม่สามารถค้นหาค่า K ได้

การทำงานของ $K - means$ จะมีประสิทธิภาพสูงก็ต่อเมื่อข้อมูลเกาะกลุ่มกันหนาแน่น แต่ละกลุ่มแยกจากกันอย่างชัดเจน และความหนาแน่นของข้อมูลในแต่ละกลุ่มใกล้เคียงกัน จุดเด่นของ $K - means$ คือง่ายและสามารถใช้ได้กับข้อมูลหลายประเภท และยังมีประสิทธิภาพในด้านความเร็ว แต่จุดด้อยของ $K - means$ ก็พบว่ายังไม่เหมาะสมกับข้อมูลทุกประเภท และไม่สามารถจัดการกลุ่มที่มีรูปร่างไม่เป็นรูปทรงกลมหรือกลุ่มที่มีขนาดหรือความหนาแน่นแตกต่างกันได้ นอกจากนี้ $K - means$ ยังถูกจำกัดสำหรับข้อมูลที่มีตัวแทนข้อมูลที่คลุมเครือหรือไม่ชัดเจน

การใช้งาน $K - means$ นั้นสามารถใช้งานได้ดีตราบที่สามารถหาค่าเฉลี่ยของคลัสเตอร์ได้ ทำให้เราทราบว่าข้อมูลที่เหมาะสมนั้นต้องเป็นข้อมูลตัวเลขเท่านั้น แต่ความเป็นจริงแล้วข้อมูลที่อยู่ในรูปแบบตัวอักษรก็เป็นข้อมูลที่สำคัญและจำเป็นที่ต้องใช้การวิเคราะห์ด้วย ซึ่งถ้าใช้วิธีการแก้ไขโดย

แปลงรูปแบบตัวอักษรให้อยู่ในรูปแบบตัวเลขก็มีโอกาสสูงที่ทำให้เกิดความผิดพลาดในเชิงความหมายได้

2.3.2 การจัดกลุ่มข้อมูลแบบเคฮาร์โมนิกมีน

การจัดกลุ่มข้อมูลแบบเคฮาร์โมนิกมีน (K-Harmonic Means: KHM) [Zhang et al, 2000] เป็นวิธีที่ปรับปรุงพัฒนามาจากวิธี $K-means$ โดยใช้ฟังก์ชันคิดค่าเฉลี่ยแบบค่าเฉลี่ยฮาร์โมนิกแทนการใช้ฟังก์ชัน Min วิธีการจัดกลุ่มแบบ KHM มีขั้นตอนวิธีดังต่อไปนี้

อัลกอริทึมการจัดกลุ่มแบบ KHM

ขั้นตอนที่ 1 สุ่มค่ากลางจำนวน j กลุ่ม แทนด้วย C_j

ขั้นตอนที่ 2 คำนวณหาค่าฟังก์ชั่นเป้าหมาย (Objective function value according) โดยแทนระยะห่างระหว่างชุดข้อมูลเข้าและค่ากลางด้วยสมการดังนี้

$$d_{ij} = \left\| X_i - C_j^p \right\|$$

$$KHM(X,C) = \sum_{i=1}^{N} \frac{k}{\sum_{j=1}^{k} 1/d_{ij}}$$

โดยที่กำหนดให้ j แทนจำนวนกลุ่ม C_j แทนค่ากลางของกลุ่ม X_i แทนด้วยข้อมูลและ N

แทนด้วย จำนวนข้อมูล

ขั้นตอนที่ 3 คำนวณหาค่าความเป็นสมาชิก ของข้อมูลทุกตัว Xi กับค่ากลางแต่ละกลุ่ม C_j แทนด้วย $m(\, C_j |\, X_j)$ คำนวณได้จากสมการดังต่อไปนี้

$$m(\, C_j |\, X_j) = \frac{1/d_{ij}^2}{\sum_{j=1}^{k} 1/d_{ij}^2}$$

ขั้นตอนที่ 4 คำนวณหาน้ำหนักของข้อมูล

$$m(\, X_j) = \frac{\sum_{j=1}^{k} 1/d_{ij}^2}{\left(\sum_{j=1}^{k} (1/d_{ij}^2)\right)^2}$$

ขั้นตอนที่ 5 คำนวณหาค่ากลางใหม่จากสมการดังต่อไปนี้

$$C_j = \frac{\sum_{i=1}^{N} m(C_j|X_i) W_i X_i}{\sum_{i=1}^{N} m(C_j|X_i) W_i}$$

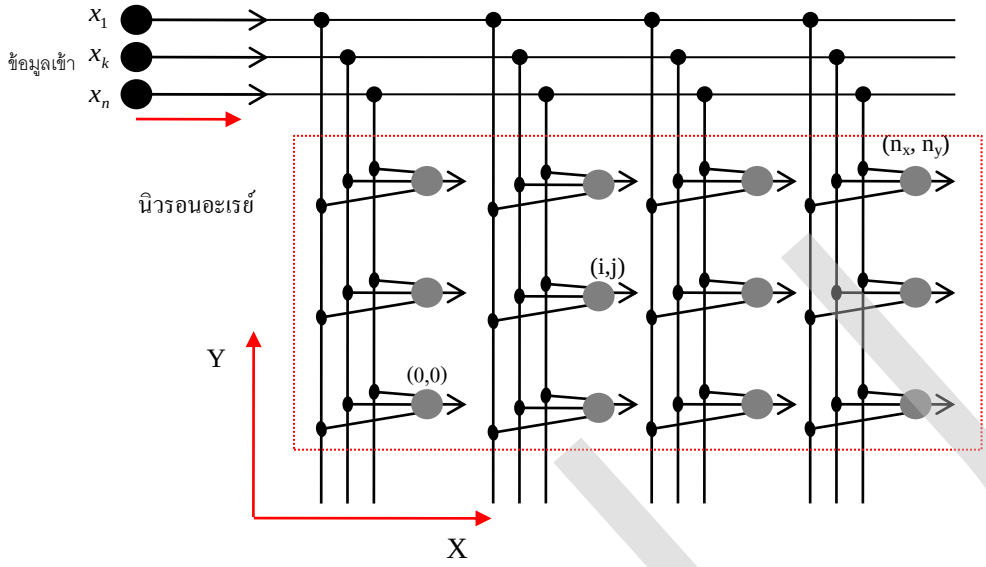
ขั้นตอนที่ 6 วนรอบทำซ้ำจนกระทั่งค่า $KHM(X,C)$ ไม่เปลี่ยนแปลง ให้ข้อมูล X_i อยู่กลุ่ม j เมื่อค่าข้อมูลมีค่าความเป็นสมาชิกในกลุ่มนั้นมากที่สุด

- ข้อเสียของทฤษฎี KHM

ปัญหาที่สำคัญ คือ KHM เป็นอัลกอริทึมที่มักจะเกิดปัญหาในการค้นหาตำแหน่งที่ดีที่สุดที่เหมาะสมที่สุด ซึ่งจะขึ้นกับการคำนวณที่มีความสัมพันธ์กับข้อมูลในกลุ่มที่เกิดขึ้น แต่อย่างไรก็ตาม การคำนวณของ KHM จะมีความเร็วกว่าแต่ผลเสียของ KHM จะเกิดติดขัดกับกลุ่มข้อมูลภายในมากกว่า ซึ่ง KHM มักจะถูกใช้คำนวณร่วมกับ Particle Swarm Optimization จะทำให้คุณภาพของการแบ่งกลุ่มดีขึ้น

2.3.3 ทฤษฎีแผนผังการจัดระเบียบตัวเอง

ทฤษฎีแผนผังการจัดระเบียบตัวเอง (Self-Organizing Maps : SOM) [T. Kohonen, 1995] [T. Mitchell,1997] เป็นอัลกอริทึมนิวรอนเน็ตเวิร์คที่นิยมใช้มากที่สุด โดยแนวคิดของ SOM คือ การทำซ้ำข้อมูลเพื่อหาค่าของน้ำหนักของข้อมูลที่มีอยู่ทั้งหมดตามจำนวนกลุ่มที่ต้องการ เป็นวิธีเกี่ยวกับการจัดกลุ่มด้วยตัวเองโดยใช้โครงสร้างตาข่ายระบบประสาท เป็นการเรียนรู้แบบไม่มีผู้สอน มีลักษณะการเรียนรู้ แบบเครือข่าย 2 ชั้น เป็นการจัดข้อมูลนำเข้าหลายมิติ ให้อยู่ในรูปแบบของข้อมูลสองมิติคือการจัดกลุ่มของข้อมูลที่มีลักษณะคล้ายกันจะอยู่ในโหนดใกล้ที่เคียงกัน โดยโครงสร้างการทำงานของ SOM ชั้นแรกทำหน้าที่นำเข้าข้อมูลและจัดส่งข้อมูลให้แก่นิวรอนชั้นที่สองทุกโหนดในตาราง หรือ อะเรย์นิวรอน (array of neurons) ระหว่างชั้นจะเชื่อมต่อกันด้วยค่าน้ำหนัก (weight) จากนั้นข้อมูลจะถูกส่งไปยังนิวรอนในชั้นที่สอง เพื่อทำการเปรียบเทียบว่าใกล้เคียงกับค่ากลางกลุ่มใดมากที่สุด แต่ละโหนดในชั้นนี้จะมีความสัมพันธ์กันแบบ เพื่อนบ้าน (neighborhood relation) ทำให้เกิดเป็นรูปแบบ 2 มิติ จากนั้นจะทำการปรับค่า น้ำหนักของตัวที่เป็นผู้ชนะ (winner) ทำให้ข้อมูลที่ข้อมูลเข้า เข้ามาเกิดการปรับเปลี่ยน และเมื่อผ่านการเรียนรู้ไปหลายๆรอบจะทำให้ ได้กลุ่มข้อมูลออกมาเป็นผลลัพธ์ จากนั้นจึงนำผลลัพธ์ที่ได้นั้นมาผ่านกระบวนการ Visualization เพื่อแสดงผลลัพธ์ที่ได้นั้นออกมาเป็นกราฟชนิดต่างๆต่อไป



ภาพที่ 2.15 โหนดการเรียนรู้ของแผนผังการจัดระเบียบตัวเอง

ขั้นตอนการทำงานของทฤษฎีแผนผังการจัดระเบียบตัวเอง SOM

1. ทำการจัดกลุ่มของข้อมูลนำเข้าให้อยู่ในกลุ่มข้อมูลที่ใกล้เคียง ทำการ Normalize ค่าข้อมูลแต่ละตัว เพื่อให้ได้ค่ากลางเป็น $x_i = \frac{x_i - \min}{\max - \min}$, กำหนดให้ $x_1, x_2, ..., x_n$ เป็นข้อมูลที่ถูกนำเข้า n คือจำนวนตัวอย่างข้อมูลเข้า min คือ ค่าที่น้อยที่สุดของชุดข้อมูล max คือ ค่าที่มากที่สุดของชุดข้อมูล
2. กำหนดโครงสร้างโครงข่ายประสาทเทียมในลักษณะสองมิติ กำหนด แนวแกน x และแกน y กำหนดชุดของข้อมูล สุ่มค่าน้ำหนักเริ่มต้น ในโครงข่ายดังแสดงในภาพที่ 2.15
3. สุ่มค่าเริ่มต้นให้กับค่ากลางของกลุ่ม (cluster center) $W = \{w_1, w_2, ..., w_p\}$, โดยที่ p คือจำนวนกลุ่ม เวกเตอร์ค่ากลางคือ $W_j = \{w_{j1}, w_{j2}, ..., w_{jm}\}$, เมื่อ $1 \le j \le p$
4. คำนวณหาระยะทางระหว่างโครงข่ายโดย ยูคลิเดียน (Euclidean distance) เพื่อหาผู้ชนะ (winner) ซึ่งจะหาได้จากโครงข่ายที่ใช้ระยะทางที่ได้จากการคำนวณ เพื่อหาค่าน้อยที่สุด

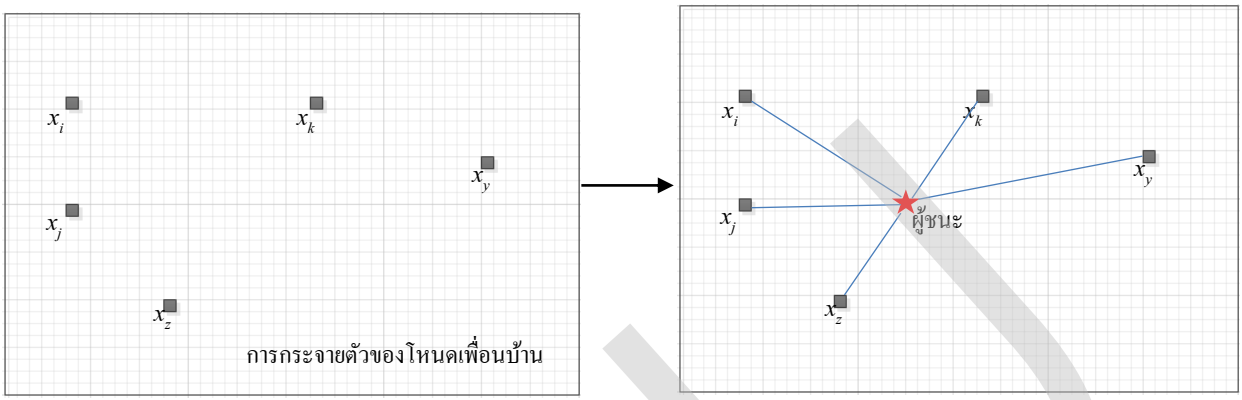
$$C(k_1, k_2) = \min_{i,j} C_{i,j}$$

เมื่อ k_1, k_2 คือดัชนีของโหนดผู้ชนะ

$$C_{i,j}^2 = \left\| x - w_j \right\|^2$$

$$C_{i,j}^2 = \sum_{i=1}^{w} (x_i - w_{ij})^2$$

เมื่อ $C_{i,j}$ ค่าความต่างระหว่างข้อมูลนำเข้า x_i กับเวกเตอร์น้ำหนัก $w_{i,j}$



ภาพที่ 2.16 การหาผู้ชนะสำหรับการกระจายตัวของข้อมูล x_i หรือโหนดเพื่อนบ้าน

5. คำนวณค่ากลางสำหรับกลุ่มที่เป็น ค่าผู้ชนะใหม่

$$w_j(t+1) = w(t) + \eta(x_i - w_j(t))$$

6. ทำซ้ำขั้นตอนที่ 4-5 จนครบชุดข้อมูลนำเข้า หรือจนกว่าค่าน้ำหนักเริ่มจะคงที่
7. โหนดที่เป็นผู้ชนะจะสอดคล้องตามสูตร $C(k_1,k_2) = \min_{i,j} C_{i,j}$ เมื่อ k_1,k_2 ดัชนีของโหนดที่เป็นผู้ชนะ
8. โหนดที่เป็นโหนดเพื่อนบ้านจะถูกกำหนดโดย $h(\rho,t) = \exp(\frac{\rho^2}{2\sigma^2(t)})$
9. หาระยะห่างระหว่างโหนดนั้น ๆ กับ โหนดที่เป็นผู้ชนะ $\rho = \sqrt{(k_1 - i)^2 + (k_2 - j)^2}$
10. หาโหนดเพื่อนบ้านที่ใกล้เคียงที่สุด โดยขอบเขตจะลดลงตามเวลาตามรูปแบบดังนี้

$$h(\rho,t) = \exp\left(-\frac{\rho^2}{\sigma^2(t)}\right)\left(1 - \frac{2}{\sigma^2(t)}\rho^2\right)$$

$$h(\rho) = \begin{cases} 1, & |\rho| \le a, \\ -\frac{1}{3}, & a < |\rho| \le 3a, \\ 0, & |\rho| < 3a, \end{cases}$$

11. ปรับค่าน้ำหนัก ของแต่ละโหนดด้วย

$$W_{i,j}(t+1) = W_{i,j}(t) + a(t)h(\rho,t)(X^l(t) - W(t))$$

- ข้อเสียของทฤษฎีแผนผังการจัดระเบียบตัวเอง SOM

การเรียนรู้นี้ไม่ทราบค่าที่ต้องการจึงไม่สามารถจะหยุดการเรียนรู้เมื่อไร ต้องมีการปรับค่าน้ำหนักไปเรื่อย ๆ ปัญหานี้สามารถแก้ไขได้ โดยอาศัยเงื่อนไขหรือหลักเกณฑ์บางอย่างในการใช้เป็นเครื่องมือในการหยุดการเรียนรู้

ค่ากลางบางตัวซึ่งเกิดจากการสุ่ม อาจจะตกในบริเวณที่ห่างจากกลุ่มค่ากลางตัวอื่นมาก ซึ่งไม่สามารถเป็น ตัวชนะได้เลย จึงต้องเพิ่มกระบวนการในการแก้ปัญหานี้

ไม่สามารถ กำหนดจำนวนชุดข้อมูลที่นำมาจัดกลุ่ม ควรมีกี่กลุ่มถึงจะเหมาะสม จึงควรมีการศึกษาวิจัยต่อไป

2.4 การวัดประสิทธิภาพ

การวัดประสิทธิภาพ (evaluation) เป็นขั้นตอนสุดท้าย เพื่อทำการตรวจสอบวิธีการที่ทำการทดลองมาข้างต้นว่ามีประสิทธิภาพมากหรือน้อยเพียงใดเมื่อนำมาใช้งานจริง จะเป็นขั้นตอนที่สำคัญ เพราะการนำวิธีการที่นำเสนอไปข้างต้นมาใช้งานได้นั้นจะต้องสอดคล้องกับความต้องการ จึงต้องมีการทดสอบศักยภาพการนำไปใช้ สถาปัตยกรรมที่ใช้ตัววัดความสำเร็จหลังการนำไปใช้หากนำไปใช้แล้วไม่ประสบผลสำเร็จต้องย้อนกลับไปเริ่มกระบวนการแรกใหม่ จึงต้องมีการประเมินผลก่อนการใช้งาน ในการประเมินนั้นกระทำได้โดยการวัดประสิทธิภาพของการจัดกลุ่มภาพมักจะถูกพิจารณาเป็นค่าของความถูกต้องของแต่ละกลุ่มข้อมูลซึ่งจะประกอบด้วย การวัดค่าความแม่นยำ ค่าความระลึก ค่าความถูกต้อง และ F-measure [R. O. Duda,1973] โดยยกตัวอย่างของค่าที่เกิดขึ้นจากตารางที่ 2.1

		ค่าทำนาย (predicted)	
		ปฏิเสธ (false/negative)	ยอมรับ (true/positive)
ค่าความจริง (actual)	ปฏิเสธ (negative)	a	b
	ยอมรับ (positive)	c	d
ค่าความถูกต้อง (accuracy)		Acc	

ตารางที่ 2.1 การวัดประสิทธิภาพ

- ค่าความแม่นยำ (false positive rate / Precision: Prec.) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องจากจำนวนภาพทั้งหมดที่ทำการค้นหามาได้

$$\Pr = \frac{a}{(a+b)}, a+b>0$$

- ค่าความระลึก (true positive rate / Recall: Re) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องจากจำนวนภาพที่ถูกต้องทั้งหมด

$$\mathrm{Re} = \frac{a}{(a+c)}, a+c>0$$

- ค่าความถูกต้อง (accuracy: Acc) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องทั้งหมดจากจำนวนภาพที่มีอยู่

$$Acc = \frac{(a+d)}{(a+b+c+d)}$$

- ค่า F-measure เป็นการวัดค่าความสัมพันธ์ระหว่างค่าความระลึกและค่าความแม่นยำในเชิงฮาร์โมนิค (harmonic) เหมาะสำหรับฐานข้อมูลสารสนเทศที่มีขนาดใหญ่มาก และมักจะไม่ทราบว่าข้อมูลภาพที่ถูกต้องทั้งหมดมีอยู่เท่าใด ทำให้ต้องทำการประมาณโดยใช้การสุ่มตัวอย่าง (sampling) ตามหลักทางสถิติหรือด้วยวิธีอื่นด้วย โดยทั่วไปจะเป็นการหาค่า F-measure ซึ่งแสดงสูตรได้ดังนี้

$$F = \frac{2(\Pr \cdot \mathrm{Re})}{(\Pr + \mathrm{Re})}$$

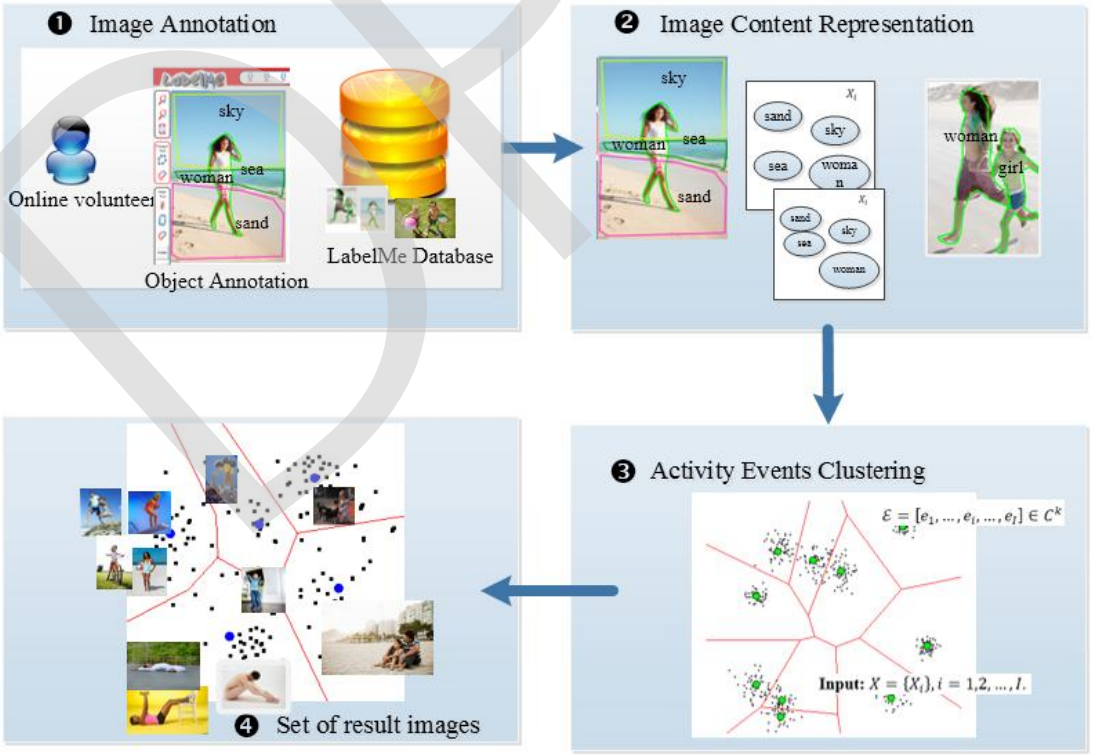
บทที่ 3
ขั้นตอนวิธีการดำเนินงานวิจัย

ในการศึกษางานวิจัยในครั้งนี้ได้มีการเก็บรวบรวมข้อมูลภาพและคัดเลือกภาพที่เหมาะสมเพื่อเตรียมเป็นข้อมูลภาพเบื้องต้น ดังนั้นข้อมูลภาพที่เตรียมพร้อมจะสามารถเข้าสู่กระบวนการประมวลผลภาพ โดยในงานวิจัยจะมีการแบ่งขั้นตอนวิธีการดำเนินงานวิจัยโดยทั่วไปจะแบ่งออกเป็น 3 ส่วนหลักโดยขั้นตอนทั้งหมดจะสามารถอธิบายได้ดังภาพที่ 3.1 ดังนี้

3.1 ขั้นตอนการเตรียมข้อมูล (data preprocessing) เป็นการทำงานในส่วนของการนำข้อมูลเข้าด้วยเครื่องมือ (image annotation tool) หมายเลข ❶ และการแทนข้อมูลวัตถุจากภาพลงในตัวแปร (Image Content Representation) หมายเลข ❷ แสดงในภาพที่ 3.1

3.2 ขั้นตอนการประมวลผล (data processing) ทำการนำข้อมูลที่ได้ทั้งหมดมาผ่านกระบวนการจัดกลุ่มภาพด้วยกิจกรรม (Activity Events Clustering) หมายเลข ❸ แสดงในภาพที่ 3.1

3.3 ขั้นตอนการวัดประสิทธิภาพ (evaluation) เป็นการเปรียบเทียบการทำงานของวิธีการที่นำเสนอและแสดงผลลัพธ์ที่ได้จากการจัดกลุ่มกิจกรรม หมายเลข ❹ แสดงในภาพที่ 3.1



ภาพที่ 3.1 ขั้นตอนการจำแนกกลุ่มความหมายภาพ

3.1 ขั้นตอนการเตรียมข้อมูล

ขั้นตอนการเตรียมข้อมูล (data preprocessing) โดยการทำการแยกและคัดเลือกข้อมูลภาพดิจิทัลที่มีวัตถุบนภาพที่เด่นชัด มีวัตถุภาพพื้นหลัง และภาพที่คัดเลือกเข้ามานั้นสามารถให้มนุษย์แปลความหมายภาพนั้นได้อย่างสมบูรณ์ สำหรับภาพบางภาพจะไม่นำเข้ามาทำการทดลองนั้นจะเป็นภาพที่มีความหมายกำกวม ภาพไม่มีความหมาย แปลความหมายไม่ได้ หรือภาพที่มนุษย์แปลได้หลายความหมาย ภาพที่มีการโฟกัสระยะใกล้ ข้อมูลภาพที่มีความซ้ำซ้อน หรือไม่สอดคล้องกันจะถูกคัดเลือกภาพนั้นออกไป และทำการรวบรวมข้อมูลภาพที่ต้องการที่มาจากหลายฐานข้อมูลจุดประสงค์ก็เพื่อทำให้มั่นใจว่าคุณภาพของข้อมูลที่ถูกเลือกนั้นเหมาะสม ดังนั้นกระบวนการทั้งหมดนี้จะประกอบด้วย 2 กระบวนการดังนี้

3.1.1 ฐานข้อมูลรูปภาพ

แหล่งข้อมูลภาพมีหลายแหล่งข้อมูลที่ได้รับการยอมรับและสามารถนำมาใช้เป็นฐานข้อมูลรูปภาพได้ เช่น Fotosearch stock[2] The Cobis Stock[3] The Corel Corporation[4] เป็นต้น เป็นแหล่งข้อมูลภาพที่หลากหลาย อาจจะมีภาพที่ไม่เหมาะสมกับการทดลองที่นำเสนอ เนื่องจาก ภาพบางภาพมีลักษณะผิดปกติ (outlier) หรือ คุณลักษณะวัตถุ (object characteristic) ไม่ชัดคลุมเครือ มีขนาดวัตถุขนาดเล็กเกินไปไม่สามารถ บ่งชี้ชื่อวัตถุได้ ภาพถ่ายระยะใกล้ (close up) ภาพบางภาพอาจจะไม่สามารถแปลความหมาย หรือภาพมีความหมายกำกวมจนทำให้ไม่สามารถหาความหมายภาพได้ ทำให้ต้องมีการคัดเลือกภาพออกไป ไม่นำมาใช้ในการทดลอง ดังนั้นในการหาแหล่งข้อมูลของการนำภาพเข้ามาใช้จึงจำเป็นต้องสมบูรณ์ที่สุด ภาพจะต้องมีความเหมาะสมกับงานที่จะนำมาใช้เพื่อตอบสนองกับความต้องการของการทดลองมากที่สุด

สำหรับการทดลองโดยทั่วไปในสาขาคอมพิวเตอร์วิชัน (computer vision) ในส่วนของการประมวลผลภาพระดับสูง รูปภาพที่นำเข้ามาทดลองจะเป็นภาพที่วัตถุถูกแท็ก หรือกระทำการให้ความหมายมาล่วงหน้าก่อน จะเรียกภาพจำพวกนี้ว่า annotated [Jia Li, 2003];[Winn J., 2005] ดังแสดงในภาพที่ 3.2 แสดงรูปภาพถูกแท็กด้วยคำหลัก ดังนั้นในการทดลองจะต้องมีฐานข้อมูลภาพที่สมบูรณ์เพียงพอที่จะสามารถนำส่วนของคำหลักที่ถูกแท็กมาใช้งานได้โดยไม่มี

[2] Fotosearch Stock: http://www.fotosearch.com ค้นคืนเมื่อวันที่ 15 ส.ค. 2558

[3] The Corbis Stock: http://pro.corbis.com ค้นคืนเมื่อวันที่ 15 ส.ค. 2558

[4] The Corel Corporation: http://www.corel.com/ ค้นคืนเมื่อวันที่ 15 ส.ค. 2558

ผลข้างเคียงต่อกระบวนการที่นำเสนอ ส่วนใหญ่แหล่งข้อมูลภาพจะทำการคัดเลือกบริเวณ (region) ที่เหมาะสมสำหรับการแท็กเป็นคำหลักหรือคำสำคัญ (keyword) เพื่อใช้สำหรับการสืบค้นข้อมูล

ภาพที่ 3.2 ตัวอย่างที่ถูกแท็กด้วยคำหลัก[5]

WordNet [Miller G.A., 1990];[Zinger S., 2005] เป็นงานวิจัยที่ได้รับความนิยมและมีนักวิจัยหลายกลุ่มที่อ้างอิงการใช้ฐานข้อมูลของคำหลักที่แท็กบนภาพ [Adrian Popescu, 2008];[Javier Álvez, 2008] โดย WordNet จะใช้คำนาม (noun) ถูกเลือกจากพจนานุกรม (Dictionary) เป็นคำหลักและนำคำหลักเหล่านั้นมาจัดกลุ่ม (class) ตามลำดับชั้น (hierarchy) ถึง 73,733 กลุ่มซึ่งจะมีคำนามที่ถูกจัดกลุ่มอยู่ภายในกิ่งก้าน (leave) ของลำดับชั้นมากมายถึง 60,000 กิ่ง และมีคำนามที่ใช้ถึง 116,364 คำ WordNet จะถูกใช้คู่กับโปรแกรมเพื่อทำการสืบค้นข้อมูล และยังสามารถกำหนดความสัมพันธ์ของคำด้วยโครงสร้างของ WordNet เองและสามารถสร้างความหมายเหมือนกัน (synonymy) ได้เช่นกัน แต่อย่างไรก็ตามในปัจจุบัน การแท็กคำศัพท์ส่วนใหญ่นั้นจะขึ้นกับแอพพลิเคชันหรือโปรแกรมสำเร็จรูปที่ช่วยเอื้อประโยชน์ให้ผู้ใช้งานสามารถเลือกคำศัพท์ที่ใช้ในการแท็กภาพได้ทันที โดยที่ผู้ใช้งานไม่จำเป็นต้องทราบถึงการนำมาจากฐานข้อมูลใดดังนั้นที่จะกล่าวต่อไปนี้จะการเลือกพิจารณาถึงแอพพลิเคชันที่สะดวกในการแท็กคำศัพท์บนภาพที่มีความนิยมอย่างแพร่หลายในปัจจุบัน

3.1.2 การแท็กคำหลักบนภาพ

สิ่งที่สำคัญถัดจากการเลือกใช้ฐานข้อมูลสำหรับกระบวนการของการประมวลผลภาพระดับสูง (high- level image processing) คือ การให้คำอธิบายภาพ (annotated images) โดยใช้ฐานข้อมูลคำหลักที่คัดเลือกมา การแท็กคำหลักบนภาพ โดยทั่วไป ภาพ (image) จะประกอบด้วย พื้นหลัง (background) และ พื้นหน้า (foreground) กล่าวคือภาพหนึ่งภาพจะ

[5] รูปภาพ อ้างอิง http://www.corbisimages.com/ ค้นคืนเมื่อวันที่ 9 เมษายน 2557

ประกอบด้วยวัตถุ (objects) หลายวัตถุ ทำให้ทุกวัตถุควรจะถูกแท็กด้วยคำหลักที่เหมาะสมจากฐานข้อมูล ดังนั้นในการแบ่งแยกวัตถุ (segmentation) เป็นหัวข้อที่มีการวิจัยอย่างต่อเนื่องในการประมวลผลภาพระดับต่ำ(low-level image processing) [Qian Huang, 1995];[Vailaya A., 2001] ด้วยคุณลักษณะของภาพหลายคุณลักษณะ (features) เพื่อทำการแบ่งแยกให้ได้วัตถุที่สมบูรณ์แบบ แต่อย่างไรก็ตามการแบ่งแยกวัตถุที่มีรูปร่างลักษณะคล้ายคลึงกันหรือแตกต่างกันแต่มีความหมายเดียวกัน ยังคงเป็นเรื่องที่ค่อนข้างยากสำหรับการประมวลผลภาพระดับต่ำรวมถึงกระบวนการรู้จำรูปแบบวัตถุ (pattern recognition) เพื่อบอกความหมายของวัตถุหรือชื่อของวัตถุ แต่อย่างไรก็ตามในการวิเคราะห์ความหมายของภาพในงานวิจัยนี้ได้เฉพาะเจาะจงในส่วนการประมวลผลระดับสูง จึงได้ข้ามในส่วนของการแบ่งแยกวัตถุ และ การรู้จำรูปแบบวัตถุ ดังนั้นในส่วนที่กล่าวต่อไปสำหรับการประมวลผลภาพระดับสูงคือการให้คำอธิบายภาพด้วยการแท็ก (tag) [Ismail Haritaoglu, 1998];[Tele Tan, 2002];[Vasileios Mezaris, 2003];[R. Zhao, 2002] ข้อมูลภาพให้อยู่ในรูปของชื่อวัตถุ หรือคำหลักจากฐานข้อมูล มีหลายกลุ่มงานวิจัยที่คิดค้นกระบวนการให้ความหมาย หรือแท็กวัตถุบนภาพด้วยวิธีการแตกต่างกัน จากกลุ่มเครื่องมือที่ให้ความหมายภาพที่นิยม [Jeroen Steggink, 2011] สามารถจำแนกวิธีการให้ความหมายตามเครื่องมือได้ 3 รูปแบบ ดังนี้

1. แบบ desktop PC [Yao, B.,2007];[Petridis, K.,2006];[Hollink, L.,2004] เป็นรูปของการให้ความหมายบนโปรแกรมจะสามารถทำงานได้ด้วยการติดตั้งลงบนเครื่องคอมพิวเตอร์ส่วนตัว (personal computer) ข้อดีของโปรแกรมที่ถูกติดตั้งลงบนเครื่องคอมพิวเตอร์ส่วนตัวคือ สามารถเพิ่มขีดความสามารถของการประมวลผลในส่วนของโปรแกรมได้มากกว่าวิธีอื่นๆ ดังนั้นการโปรแกรมจะมีความสามารถได้มากกว่าวิธีการอื่นๆ มีความยืดหยุ่นกว่า แต่อย่างไรก็ตามข้อจำกัดของวิธีการนี้ คือ เนื่องจากการให้ความหมายบนวัตถุนั้นสามารถทำได้ด้วยตนเองทำให้ ฐานข้อมูลภาพถูกจำกัด ขอบเขตของคำหลักถูกจำกัด หรืออาจไม่ครอบคลุมมากพอการให้ความหมายภาพอาจจะถูกให้ความหมายด้วยตนเองหรือเพียงกลุ่มบุคคลใดเพียงกลุ่มเดียว ทำให้การทดลองไม่ครอบคลุมเท่าที่ควร ตัวอย่างโปรแกรม เช่น Image Parsing [Yao. B., 2007], M-OntoMat-Annotizer [Petridis K.,2006], Photostuff [Halaschek-Wiener,2005], Spatial Annotation [Hollink L.,2004] ดังแสดงในตารางที่ 3.2

2. แบบออนไลน์ (online) [Russell, B.C.,2008];[Volkmer, T.,2005]; รูปแบบนี้จะสามารถให้ความหมายวัตถุบนภาพได้ทางออนไลน์ผ่านทางเว็บไซต์ เป็นข้อดีสำหรับรูปแบบนี้เพราะทุกคนสามารถเข้าถึงได้อย่างง่าย ข้อมูลภาพและกลุ่มข้อมูลคำหลักจะถูกจัดเก็บลงบน

เซิร์ฟเวอร์ เดียวกัน ฐานข้อมูลภาพหลากหลายและกลุ่มคนที่ให้ความหมายมีหลายกลุ่ม แต่อย่างไรก็ตามการที่มีฐานข้อมูล และกลุ่มคำศัพท์ที่เปิดกว้างมากเกินไปนี้อาจจะเป็นข้อเสียสำหรับการทดลองที่มีขีดจำกัดเช่นเดียวกัน ส่วนกลุ่มบุคคลที่เข้ามา จะไม่สามารถควบคุมได้การให้ความหมายภาพให้เป็นไปตามที่กำหนดของส่วนโปรแกรมที่ทำการทดลอง ตัวอย่างโปรแกรม เช่น LabelMe [Russell, B.C.,2008], IBM EVA [Volkmer, T.,2005] ดังแสดงในตารางที่ 3.1

รูปแบบเครื่องมือ	เครื่องมือคำอธิบายภาพ	รูปแบบการเลือกวัตถุ			
		global	bounding box	polygon	free hand
Desktop PC	Image Parsing	✓		✓	
	M-OntoMat-Annotizer		✓		✓
	Photostuff		✓		
	Spatial Annotation	✓			
Online	Flickr	✓	✓		
	IBM EVA	✓			
	LabelMe			✓	✓
Online game	ESP game		✓		
	Peekaboom				✓
	Squigl				✓

ตารางที่ 3.1 ประเภทของเครื่องมือให้ความหมายภาพ

3. แบบออนไลน์เกมส์ (online game) [Von Ahn,2004];[Von Ahn L.,2006] เป็นการให้ความหมายวัตถุจะอยู่ในรูปแบบของการเล่นเกมส์ผ่านทางออนไลน์บนเครือข่ายของอินเทอร์เน็ต คุณลักษณะพื้นฐานจะคล้ายกับแบบที่สอง แต่การให้ความหมายในลักษณะนี้อยู่บนพื้นฐานของการเล่นเกมส์ ซึ่งต้องอาศัยความเร็ว ควบคู่กับจำนวนคำหลักของการแท็กบนวัตถุดังนั้นการให้ความหมายในรูปแบบค่อนข้างมีข้อจำกัด และ ขึ้นกับความสามารถของผู้เล่นเกมส์เป็นส่วนใหญ่ ไม่ใช่ความถูกต้อง ดังนั้นอาจทำให้การแท็กวัตถุผิดพลาดหรือไม่ได้ไตร่ตรองให้ความหมายที่ถูกต้องอย่างแท้จริง แต่อย่างไรก็ตามรูปแบบออนไลน์เกมส์ยังเป็นเครื่องมือที่นิยมกันในปัจจุบัน ยกตัวอย่างเช่น ESP game [Von Ahn, 2004] Peekaboom

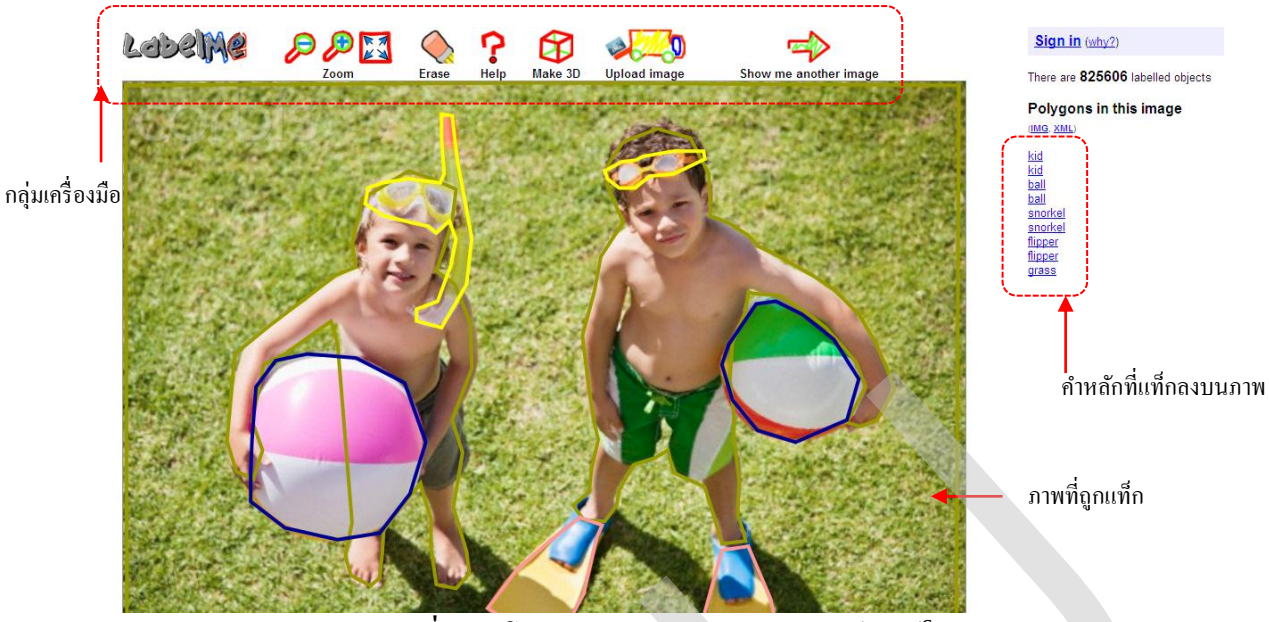
[Von Ahn, 2006] ยังมีหลายหน่วยงานที่สร้างโปรแกรมเพื่อมารองรับการทำงานในรูปแบบทั้ง 3 ดังแสดงในตารางที่ 3.1

ฐานข้อมูล	จำนวนกลุ่ม	จำนวนภาพ	จำนวนแท็ก	รูปแบบแท็ก
LabelMe	183	30,369	111,490	Polygons
Caltech-101	101	8,765	8,765	Polygons
MSRC	23	591	1,751	Polygons
CBCL-Streetscence	9	3,547	27,666	Region masks
Pascal2006	10	5,304	5,455	Bounding boxes

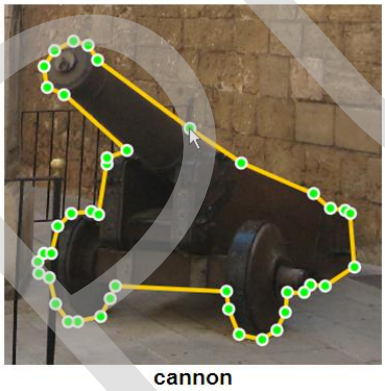
ตารางที่ 3.2 รายละเอียดเครื่องมือให้ความหมายภาพ [BC Russell et al.,2008]

จากเครื่องมือให้ความหมายภาพ (image annotation tool) ในตารางที่ 3.1 ที่แบ่งเป็น 3 รูปแบบ แต่ละรูปแบบมีความสามารถทำงานได้แตกต่างกันมีทั้งข้อดีและข้อจำกัด ดังนั้นในงานวิจัยนี้ได้เลือกเครื่องมือแบบออนไลน์ โดยใช้โปรแกรม LabelMe[9] [BC Russell et al.,2008] [A. Torralba,2010] เป็นเครื่องมือที่ได้รับการยอมรับอย่างกว้างขวาง สำหรับงานวิจัยทางด้าน Computer Vision โดยแอพพลิเคชันนี้สามารถทำงานได้อย่างเต็มรูปแบบบนเว็บในลักษณะของเครื่องมือให้ความหมาย (Web-based annotation tools) เริ่มตั้งแต่ปี 2005 ปัจจุบันมีวัตถุบนภาพที่ถูกให้ความหมายรวมทั้งสิ้น 400,000 วัตถุ [Von Ahn and L. Dabbish, 2004.];[B. C. Russell, 2008];[A. Sorokin and D. Forsyth, 2008];[M. Spain and P. Perona, 2007];[D. G. Stork, 1999] เมื่อทำการเปรียบเทียบจำนวนแท็กและจำนวนภาพที่เกิดขึ้นของ LabelMe กับเครื่องมืออื่น ๆแล้วจะเห็นว่าจำนวนของแท็กและภาพมีความหลากหลายมากกว่าเครื่องมืออื่น ๆอย่างชัดเจน ดังแสดงในตารางที่ 3.2

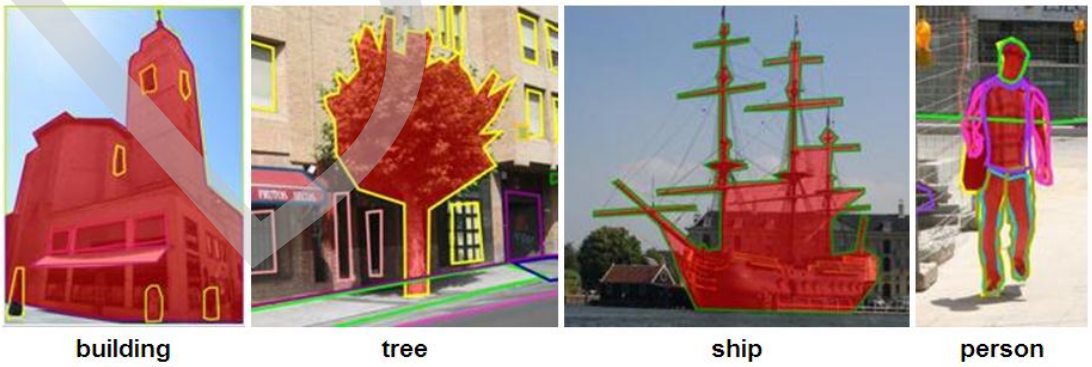
ผู้ใช้สามารถเข้าถึงโปรแกรมผ่านทางเครือข่ายออนไลน์ได้ สามารถแท็กวัตถุผ่านทางออนไลน์ได้ดังแสดงในภาพที่ 3.3 แสดงหน้าเว็บไซต์ของโปรแกรม LabelMe โปรแกรมสามารถทำงานร่วมกันได้หลายเพลตฟอร์ม ซึ่งเป็นวัตถุประสงค์หลักของคณะผู้จัดทำโปรแกรมนี้ ทำให้ผู้ใช้งานที่เข้ามาให้ความหมายภาพมาได้มากมาย และมีพื้นฐานของการให้ความหมายที่แตกต่างกันตามความสามารถ ของแต่ละบุคคล



ภาพที่ 3.3 โปรแกรม LabelMe บนเบราว์เซอร์[6]



ภาพที่ 3.4 การเลือกสัดส่วนของวัตถุบนภาพ[6]



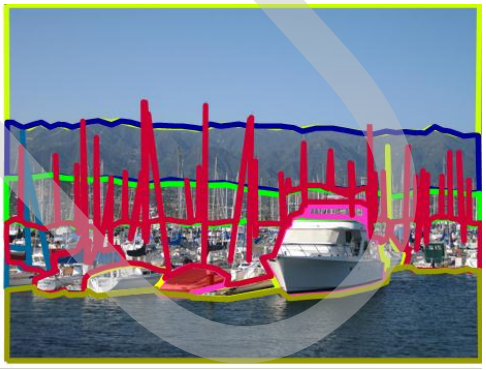
ภาพที่ 3.5 ตัวอย่างวัตถุที่ถูกแท็กด้วยโปรแกรม LabelMe[6]

[6] LabelMe: http://labelme.csail.mit.edu/ ค้นคืนเมื่อวันที่ 9 เมษายน 2557

คำหลัก: car traffic light car taxi building sky

คำหลัก: person woman sea person woman sky sea sand

คำหลัก: bus person wheel building

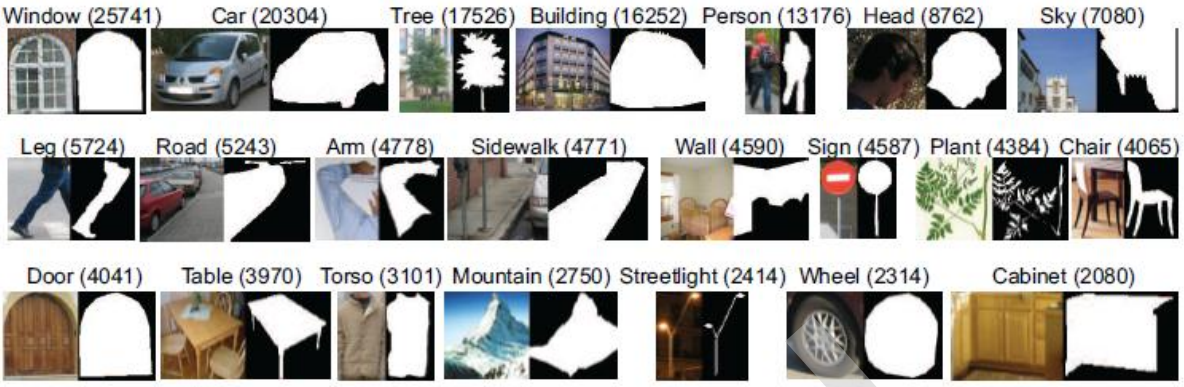


คำหลัก: boat sea water sky mountain wheel building

ภาพที่ 3.6 ตัวอย่างภาพที่ถูกแท็กคำหลักบนภาพ ด้วยโปรแกรม LabelMe[7]

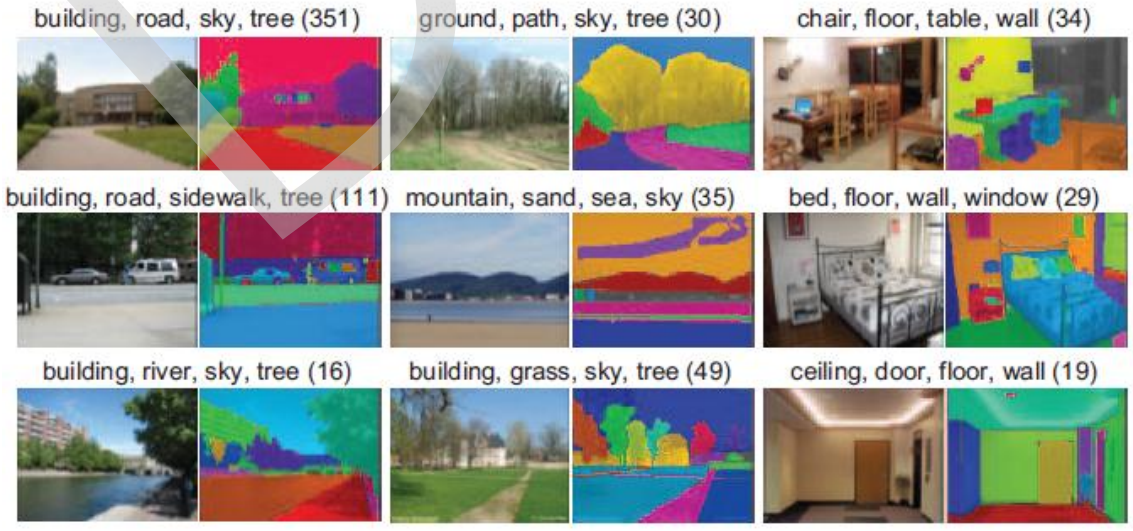
รูปแบบของการให้ความหมายภาพ หรือแท็กข้อมูลบนโปรแกรม LabelMe จากภาพที่ 3.3 ประกอบด้วยเครื่องมือช่วยการแท็ก (ด้านบน) รูปภาพสามารถคัดเลือกจากฐานข้อมูลภายในโปรแกรม หรือ โหลดรูปภาพที่ต้องการจากเครื่องคอมพิวเตอร์เข้ามาในโปรแกรม LabelMe ได้ วิธีการแท็กวัตถุ แสดงในภาพที่ 3.4 เพื่อทำการแท็กคำหลักบนวัตถุ ด้วยการใช้เมาส์ลากตามสัดส่วนของวัตถุแบบ freehand บนรูปภาพ เมื่อลากตามสัดส่วนของวัตถุเสร็จสิ้นโปรแกรมจะให้ใส่คำหลักเพื่อให้ความหมายของวัตถุ ข้อมูลคำหลักจะถูกจัดเก็บลงบนฐานข้อมูลพร้อมกับรูปภาพ ข้อมูลคำหลักที่ถูกแท็กแล้วจะแสดงไว้ทางขวามือ จะได้ข้อมูลคำหลักของวัตถุบนภาพประกอบด้วย grass, snorkel, snorkel, kid, kid, ball, ball, flipper และ flipper ภาพที่ 3.7 แสดงตัวอย่างของภาพที่ถูกแท็กเรียบร้อยแล้ว สามารถค้นหาได้จากฐานข้อมูลที่เก็บไว้ตามลักษณะภาพที่ต้องการ

[7] LabelMe: http://labelme.csail.mit.edu/ ค้นคืนเมื่อวันที่ 9 เมษายน 2557

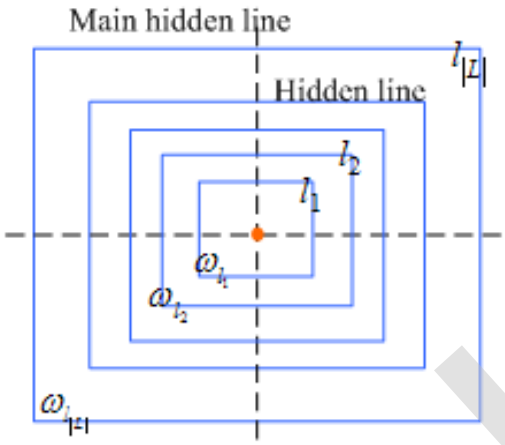


ภาพที่ 3.7 ตัวอย่างของคำหลักที่พบบ่อยในการให้ความหมายวัตถุ [A. Torralba, 2010]

หลังจากที่ผู้ใช้งานแท็กภาพในโปรแกรม LabelMe หรือทำการค้นหารูปภาพที่ถูกแท็ก โปรแกรมจะมีการแสดงภาพที่ถูกแท็ก ดังแสดงในภาพที่ 3.5 แสดงภาพที่ถูกแท็ก เส้นแสดงขอบเขตของคำหลักแต่ละคำที่แท็กไว้ ดังนั้นการใช้ฐานข้อมูลในโปรแกรม สามารถดาวน์โหลดมาใช้ร่วมกันได้ และข้อมูลคำหลักที่ถูกนำมาใช้มีมากมายหลายฐานดังแสดงในตารางที่ 3.3 ดังนั้นข้อมูลสำหรับการทดลอง จะมีการใช้ฐานข้อมูลที่เก็บรวบรวมมาจาก LabelMe [B. C. Russell, 2008] โดยจะทำการให้ความหมายในลักษณะของรูปทรงแบบ polygon ที่วาดลงบนวัตถุในภาพ ดังแสดงตัวอย่างในภาพที่ 3.7 [A. Torralba, 2010] จากโปรแกรม LabelMe เป็นตัวอย่างคำหลักที่พบบ่อยหรือเกิดขึ้นบ่อยในการให้ความหมายของวัตถุบนภาพ ตัวเลขด้านข้างแทนจำนวนของการให้ความหมาย สำหรับในภาพที่ 3.8 [A. Torralba, 2010] แสดงตัวอย่างของข้อมูลคำหลักที่ถูกแท็กบนภาพในโปรแกรม LabelMe เช่นกัน



ภาพที่ 3.8 ตัวอย่างของคำหลักในการให้ความหมายวัตถุบนภาพ [A. Torralba, 2010]



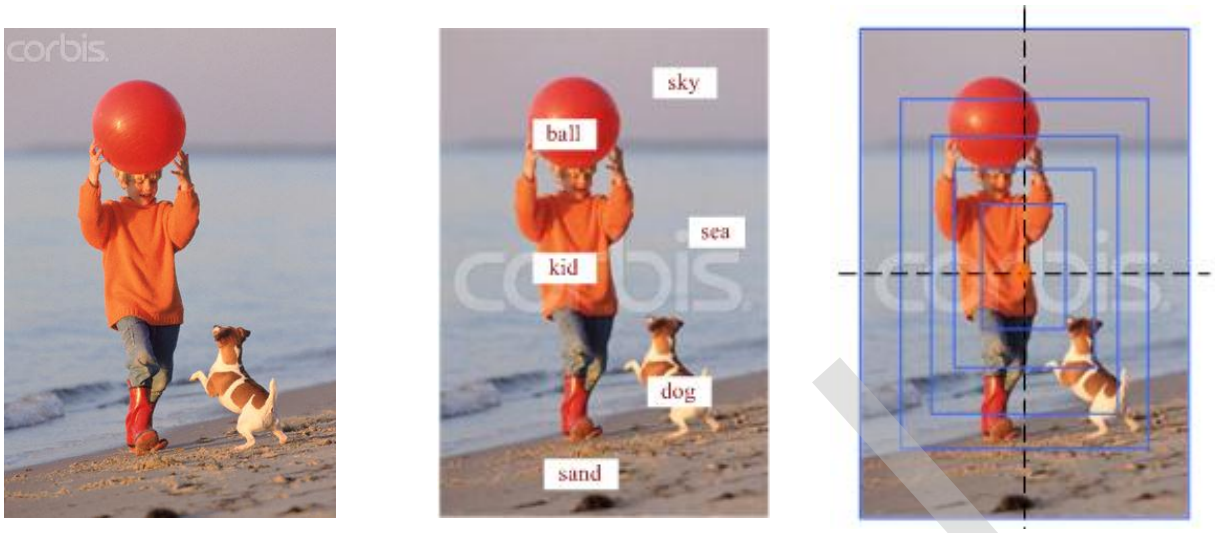
ภาพที่ 3.9 แสดงโครงสร้างสเกลเลตอน

3.2 ขั้นตอนการประมวลผล

สำหรับขั้นตอนการประมวลผลจะนำข้อมูลภาพให้อยู่ในรูปของชื่อวัตถุ หรือความหมายของวัตถุที่ปรากฏบนภาพ ที่ถูกอินพุตเข้ามาในระบบและถูกให้ความหมายเป็นคำศัพท์ หรือแท็กข้อมูลวัตถุทั้งหมดและถูกจัดเก็บลงในรูปแบบของฟีเจอร์เวกเตอร์ (feature vector)

3.2.1 การสกัดข้อมูลจากภาพ

การสกัดข้อมูลออกจากภาพ (feature extraction) [N. Chinpanthana, 2010] จะเป็นวิธีที่ดึงข้อมูลที่ต้องการออกจากภาพซึ่งในขั้นตอนนี้จะเป็นการนำเอา โครงสร้างสเกลเลตอนเข้ามาประยุกต์เพื่อใช้ในการแปลความหมายภาพ ดังแสดงในภาพที่ 3.9 โดยพิจารณาจากข้อมูลวัตถุ O_i เริ่มต้น จากทฤษฎีของ Rudolph Arnheim [Rudolp A., 1974] ที่กล่าวไว้ว่า การรับรู้รวมถึงการสนใจภาพครั้งแรกของมนุษย์ นั้นจะรับรู้วัตถุที่เด่นก่อน โดยวัตถุที่เด่นนั้นจะเป็นวัตถุในแนวกึ่งกลางภาพ และจะต้องมีขนาดของวัตถุใหญ่เพียงพอ เพราะฉะนั้นวัตถุที่อยู่ด้านริมขอบภาพหรือด้านข้างภาพจะรับรู้หรือสนใจเป็นส่วนถัดไปโดยคิดเป็นสัดส่วนลดหลั่นกันตามขนาดและตำแหน่งบนภาพ ดังแสดงในภาพที่ 3.10 (ข) ประกอบด้วยวัตถุ *sea*, *sky*, *sand*, *kid*, *ball* และ *dog* โดยที่ขนาดของวัตถุที่เป็น *sea*, *sky* และ *sand* จะมีพื้นที่มากที่สุด ตามลำดับ แต่เมื่อพิจารณาโดยใช้หลักการของ Rudolph Arnheim ดังแสดงในภาพที่ 3.9 จะเห็นว่าเมื่อมนุษย์รับรู้ภาพครั้งแรก วัตถุที่มนุษย์สนใจ จะอยู่ตำแหน่งกึ่งกลางภาพและมีขนาดที่เด่นพอ เพราะฉะนั้นสิ่งที่มนุษย์จะสนใจมากที่สุดคือ *kid* เป็นลำดับแรกมากกว่าที่จะสนใจ *sea* ทั้งที่จะมีขนาดที่ใหญ่กว่ามาก เพราะฉะนั้น จากทฤษฎีข้างต้นจึงนำมาดัดแปลงเพื่อที่จะคิดคำนวณเป็นสมการเพื่อหาค่าของวัตถุที่ปรากฏบนภาพ และให้ค่าของวัตถุที่มีคุณลักษณะแตกต่างกันมีค่าของวัตถุแตกต่างกันด้วย



(ก) ภาพตัวอย่าง (ข) ภาพถูกแท็กด้วยคำศัพท์ (ค) ภาพถูก Mapping ด้วยโครงสร้างสเกลเลตอน

ภาพที่ 3.10 แสดงภาพที่ถูก Mapping ด้วยโครงสร้างสเกลเลตอน[8] [N. Chinpanthana, 2010]

ขั้นตอนการประมวลผล (data processing) นำข้อมูลฟีเจอร์เวกเตอร์ ทั้งหมดที่ได้ซึ่งประกอบด้วยข้อมูลวัตถุ ที่มีการแท็กมาเป็นพารามิเตอร์, $O_i \in \{o_1, o_2, ..., o_n\}$ และข้อมูลที่เป็นขนาดของแต่ละวัตถุ (object size), ตำแหน่งของแต่ละวัตถุ (object position) เข้ามาประกอบในการพิจารณาร่วมกับโครงสร้างสเกลเลตอน เมื่อได้ข้อมูลครบถ้วนจะนำเข้าสู่กระบวนการคัดเลือกฟีเจอร์ และการจำแนกข้อมูลภาพเป็นกระบวนการสุดท้าย วิธีการมีดังนี้

- การหาขนาดของวัตถุ

การหาขนาดของวัตถุ (object size: s_i) เป็นการนับจำนวนจุดพิกเซลทั้งหมดของแต่ละวัตถุที่ถูกแท็กไว้แล้วโดยที่ขนาดของวัตถุใหญ่หรือพื้นที่มากแสดงว่ามีจำนวนจุดพิกเซลมากกว่าวัตถุที่มีจำนวนพิกเซลน้อยหรือพื้นที่น้อย ยกตัวอย่างเช่น ภาพที่ 3.10 (ข) ขนาดของวัตถุ *sea* มีขนาดที่ใหญ่กว่า วัตถุ *kid* เป็นต้น

เมื่อกำหนดให้ $O_i \in \{o_1, o_2, ..., o_n\}$ โดยที่แต่ละ o_i จะมีพื้นที่เป็น $region(o_i) \in (x_i, y_i)$ สามารถเขียนสูตรนับจำนวนพิกเซลบนพื้นที่ของวัตถุทั้งหมดได้ดังนี้

$$s_i = \sum_{i=1}^{|o_n|} (pixel(x_i, y_i))$$

[8] อ้างอิงรูปภาพ The corbis stock: http://pro.corbis.com

- การหาตำแหน่งวัตถุ

การหาตำแหน่งวัตถุ (object position: p_i) [N. Chinpanthana, 2010] จะใช้โครงสร้างสเกลเลตอนเพื่อทำการ mapping บนภาพเพื่อทำการคำนวณหาตำแหน่งของวัตถุและค่าน้ำหนักบนภาพ จากโครงสร้างที่กำหนดไว้ ดังแสดงในภาพที่ 3.10 (ค) เนื่องจากตำแหน่งของวัตถุที่อยู่จุดศูนย์กลางจะมีความสำคัญมากกว่า วัตถุที่อยู่ด้านริมหรือขอบภาพ ดังที่กล่าวมาแล้วข้างต้น

ดังนั้นจึงได้กำหนดค่าของน้ำหนักบนโครงสร้างสเกลเลตอนของแต่ละเส้นถูกกำหนดเป็นค่าของ $\omega \in \{\omega_{l_1}, \omega_{l_2}, ..., \omega_{l_{|L|}}\}$ ที่สัมพันธ์กับเส้นบนโครงสร้าง $l \in \{l_1, l_2, ..., l_{|L|}\}$ สามารถเขียนสูตรได้ดังนี้

$$p_i = \sum_{region(x_i, y_i) \in l_j, j=1}^{o_{n^*}|L|} \omega_{l_j}$$

3.2.2 การแทนค่าวัตถุลงในตัวแปร

หลังจากผ่านขั้นตอนของการเตรียมข้อมูลภาพ ฐานข้อมูลคำหลักและฐานข้อมูลภาพได้ถูกคัดเลือกเข้ามา จะทำการจัดเก็บข้อมูลคำหลักที่ได้ลงในกลุ่มตัวแปรดังนั้นข้อมูลทั้งหมดที่ได้จะมีการนำภาพและข้อมูลวัตถุเป็นฟีเจอร์ลงในตัวแปรที่เป็นเมตริกซ์ เพื่อทำการประมวลผลตามทฤษฎีที่นำเสนอ กำหนดให้ I แทนข้อมูลภาพ และในข้อมูลภาพประกอบด้วยข้อมูลวัตถุที่เก็บเป็นชุดฟีเจอร์ $\gamma = \{X_i, P_i, S_i\}, i = 1,2, \dots, I.$ เมื่อกำหนดให้ k แทนลำดับของวัตถุที่ถูกแท็กในภาพที่ i สามารถเขียนแทนด้วย$X_i = [x_i^1, \dots, x_i^{n_i}] \in \mathcal{R}$,ตำแหน่งวัตถุสามารถเขียนแทนด้วย $P_i = [p_i^1, \dots, p_i^{n_i}] \in \mathcal{R}$, ขนาดของวัตถุ $S_i = [s_i^1, \dots, s_i^{n_i}] \in \mathcal{R}$,เมื่อ n_i แทนจำนวนที่แท็กภาพลำดับที่ i ของ $X = [x_1, \dots, x_i, \dots, x_N]$, แล้วค่าของจำนวนแท็กบนภาพทั้งหมดสามารถเขียนแทนได้ด้วย N เมื่อ $N = \sum_{i=1}^{I} n_i$.

3.2.3 การจัดกลุ่มภาพด้วยเหตุการณ์กิจกรรม (Activity Events Clustering)

กระบวนการจัดกลุ่มภาพด้วยเหตุการณ์กิจกรรม (Activity Events Clustering) จะใช้การจัดกลุ่มภาพแบบฟัซซี่ซีมีน (Fuzzy C-Means) [T. Luis,2009] [L. Hsiang-Chuan, 2008] เริ่มจากการกำหนดกลุ่มข้อมูลที่ต้องการแบ่งกลุ่ม กำหนดเงื่อนไขในการให้ ข้อมูลหยุดให้เป็น ε และกำหนดค่าฟัซซี่พารามิเตอร์ m สำหรับการสร้างฟัซซี่พาทิชัน (Fuzzy partition) เริ่มจากกำหนดจุดศูนย์กลางเริ่มต้น ของ ข้อมูลจำนวน N ในกลุ่ม C โดยเลือกจากเมตริกซ์การเป็นสมาชิก (Membership Matrix) ได้จาก จำนวน N คูณด้วยกลุ่ม C สมาชิกตัวที่ μ_{ij} ของเมตริกซ์จะแสดงระดับการเป็น สมาชิกตัวที่ x_j ในกลุ่มข้อมูลที่ c_i และทำการหาค่าฟัซซี่ฟังก์ชัน (fuzzy objective function) ด้วยค่า μ_i เขียนเป็นสมการ Object Function ได้ดังนี้

$$\delta_{FCM} = \sum_{i=1}^{c}\sum_{j=1}^{n} d\mu_{ij}{}^{m} d(x_i, z_i)$$

กำหนดให้ δ_{FCM} แทน Objective Function ของขั้นตอนวิธีฟัซซี่ซีมีน และ ให้ $X = \{x_1, x_2, \ldots, x_n\}$ เมื่อ n แทนจำนวนข้อมูล C แทนจำนวนกลุ่มข้อมูล m แทน ฟัซซี่พารามิเตอร์ ต้องมีค่ามากกว่า 1 เมื่อ μ_{ij} เป็นสมาชิก (membership) ของข้อมูล ที่ j กลุ่มที่ i และ$d^2(x_j - z_i)$ เป็นสมการระยะทางระหว่างข้อมูล x ที่ j จากจุดศูนย์กลางของข้อมูล z กลุ่มที่ i ดังนั้น สามารถคำนวณค่าจุดศูนย์กลางของข้อมูล z กลุ่มที่ i ได้ดังนี้

$$z_i = \frac{\sum_{j=1}^{n}(\mu_{ij})^m x_j}{\sum_{j=1}^{n}(\mu_{ij})^m}$$

ทำการปรับกลุ่มเพื่อลดค่าฟัซซี่ฟังก์ชันและทาการคำนวณ μ ใหม่การหาค่าการเป็นสมาชิก μ_{ij} แสดงได้ดังสมการ

$$\mu_{ij} = \frac{[1/d^2(x_j - z_i)]^{1/(m-1)}}{\sum_{i=1}^{c}[1/d^2(x_j - z_i)]^{1/(m-1)}}$$

ผลรวมของค่าการเป็นสมาชิกของข้อมูลสามารถคำนวณได้ดังสมการ

$$\sum_{i=1}^{c} \mu_{ij} = 1$$

จะทำการคำนวณเพื่อหาค่า δ_{FCM} ใหม่จนกระทั่งข้อมูลใน μ_iไม่เปลี่ยนแปลง หรือน้อยกว่าค่า ε จะหยุดหาค่า

3.3 ขั้นตอนการวัดประสิทธิภาพ

ขั้นตอนการวัดประสิทธิภาพ (evaluation) เป็นขั้นตอนสุดท้ายของกระบวนการจัดกลุ่มกความหมายภาพ ได้นำผลที่ได้จากการจัดกลุ่มของภาพที่จัดได้มาทำการวัดประสิทธิภาพของสิ่งที่นำเสนอข้างต้น โดยจะตรวจสอบกลุ่มของภาพที่จัดได้ว่ามีค่าเป็นอย่างไร เมื่อเทียบกับกลุ่มของภาพที่ถูกต้อง ซึ่งต้องมีการวัดค่าความระลึก (recall) และค่าความแม่นยำ (precision) จะเป็นค่าที่แสดงว่า การค้นคืนข้อมูลได้ตรงกับความต้องการเพียงใด ส่วนค่าความระลึกจะเป็นค่าที่แสดงถึงความครอบคลุมในการจัดกลุ่มภาพ หลังจากนั้นจะนำค่ามาคำนวณในรูปของค่าความถูกต้อง (accuracy) และ F-measure (F_1) ต่อไปและนำค่าทั้งหมดมาแปลผลและประเมินผลลัพธ์ที่ได้ว่ามีความเหมาะสม หรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ในรูปที่สามารถเข้าใจได้ง่าย สามารถอ่านเพิ่มเติมได้ในบทที่ 2

บทที่ 4
ผลการทดลอง

ในการทำวิจัยครั้งนี้เพื่อ จัดกลุ่มภาพตามความหมายของภาพในรูปแบบของภาพส่วนบุคคลที่จัดกลุ่มตามกิจกรรม (Activity Events Clustering) โดยจะใช้ความสัมพันธ์ของขนาด รตำแหน่งของวัตถุพร้อมทั้งทำศัพท์ที่ถูกแท็กลงบนภาพ เพื่อทำให้ได้ผลลัพธ์ของความหมายภาพอย่างแท้จริง โดยทำการเปรียบเทียบความเหมือนกันของความหมายภาพด้วยการหาความเหมือนด้วยทฤษฎีการจัดกลุ่มแบบ K-means (K-means), การจัดกลุ่มข้อมูลแบบเคฮาร์โมนิกมีน (K-Harmonic Means), การจัดกลุ่มแบบการจัดระเบียบตัวเอง (Self-Organizing Maps) และ การจัดกลุ่มแบบฟัซซี่ซีมีน (Fuzzy C-Means)

4.1 การกำหนดข้อมูลภาพ

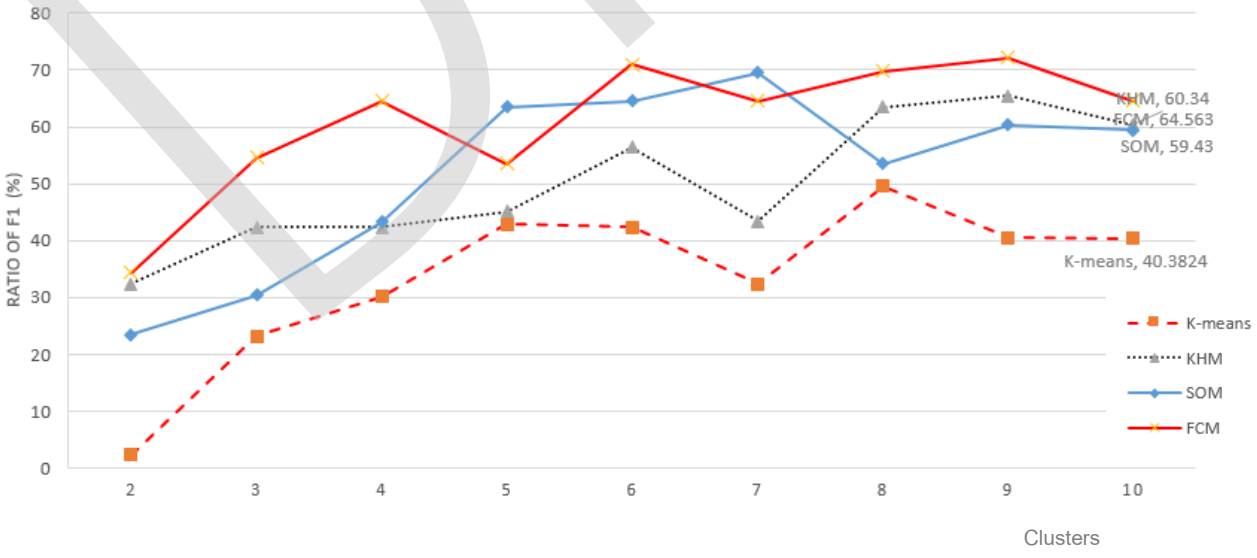
ข้อมูลภาพสำหรับการทดลองได้ใช้ฐานข้อมูลที่เก็บรวบรวมและมาจากแอพพลิเคชัน LabelMe [B. C. Russell,2008];[A. Torralba,2010] โดยแอพพลิเคชันนี้สามารถทำงานได้อย่างเต็มรูปแบบบนเว็บในลักษณะของเครื่องมือให้ความหมาย และได้มีการคัดเลือกภาพสำหรับการทดลองให้ครอบคลุมกับการทดลองโดยเป็นภาพส่วนบุคคลที่อยู่ในหมวดหมู่ของภาพภายใน (indoor) และภาพภายนอก (outdoor) มาจากฐานข้อมูล Corbis [Corbis] และคำศัพท์ได้อ้างอิงจากฐานข้อมูล NIST TRECVID 2015 [NIST] และWordNet [Miller G.A., 1990];[Zinger S., 2005] คัดเลือกคำศัพท์มาทั้งหมด 95 คำหลักที่เป็นภาษาอังกฤษเพื่อให้สอดคล้องกับแอพพลิเคชันที่ใช้งาน และฐานข้อมูลที่มีอยู่ และคำหลักที่ถูกสร้างขึ้นนั้นได้พยายามหลีกเลี่ยงคำหลักที่มีความหมายกำกวมและคำหลักที่เป็นทั้งคำเหมือน (Synonym) เช่น “abbey”, “church”, “cathedral” เป็นต้น เพราะอาจเป็นส่วนหนึ่งที่ทำให้ผลการทดลองนั้นเกิดความผิดพลาดได้ ดังนั้นในการทดลองนี้จะมีการแท็กคำหลักตามที่กำหนดไว้บนภาพที่ถูกคัดเลือกมาตามความหมายพื้นฐานจากการสุ่มและตัดสินใจจากการจำแนกความหมายภาพด้วยคนเป็นหลัก (human scenes classification) [Jianxiong Xiao, 2010] ข้อมูลภาพทั้งหมดจะถูกแท็กไว้บนภาพจากผู้ใช้ในแอพพลิเคชัน LabelMe โดยกำหนดให้จำกัดขอบเขตของคำหลักที่ใช้ในการทดลองนี้รวมทั้งหมด

ภาพรวมทั้งหมดที่ใช้ในการทดลอง 1,200 ภาพ และการจัดกลุ่มแบบ K-means (K-means), การจัดกลุ่มข้อมูลแบบเคฮาร์โมนิกมีน (K-Harmonic Means: KHM), การจัดกลุ่มแบบการจัดระเบียบตัวเอง (Self-Organizing Maps: SOM) และ การจัดกลุ่มแบบฟัซซี่ซีมีน (Fuzzy C-Means: FCM) ทั้งหมดจะถูกนำมาเปรียบเทียบความถูกต้องด้วยตาราง Confusion matrix และ แสดงผลเปรียบเทียบการจัดกลุ่มด้วยค่าความแม่นยำ, ค่าความระลึก และค่า F_1 ดังแสดงผลการทดลองได้ดังนี้

4.2 ผลการทดลองการจัดกลุ่มความหมายภาพส่วนบุคคล

4.2.1 การค้นหากลุ่มความหมายภาพส่วนบุคคล

กำหนดการทดลองเบื้องต้นเป็นการค้นหากลุ่มความหมายภาพส่วนบุคคลที่เหมาะสมโดยสุ่มภาพจากฐานข้อมูลที่กำหนดไว้ขึ้นมาและใช้ชุดข้อมูลวัตถุที่เก็บเป็นชุดฟีเจอร์ $\gamma = \{X_i, P_i, S_i\}, i = 1,2,\dots,I.$ โดยที่ $X_i = [x_i^1, \dots, x_i^{n_i}] \in \mathcal{R}$ แทนชุดข้อมูลของคำหลักในภาพ, $P_i = [p_i^1, \dots, p_i^{n_i}] \in \mathcal{R}$, แทนข้อมูลตำแหน่งวัตถุ $S_i = [s_i^1, \dots, s_i^{n_i}] \in \mathcal{R}$, แทนขนาดของวัตถุ เมื่อ n_i แทนจำนวนที่แท็กภาพลำดับที่ i ของ $X = [x_1, \dots, x_i, \dots, x_N]$ และทำการทดลองเพื่อค้นหากลุ่มภาพด้วยเทคนิคการจัดกลุ่มเหตุการณ์กิจกรรมแบบ K-means, KHM, SOM และ FCM รายละเอียดของสมการดูเพิ่มเติมได้ในบทที่ 2 และ 3 โดยการทดลองให้เริ่มจัดกลุ่มตั้งแต่ 2-10 กลุ่ม เพื่อค้นหาค่าความถูกต้องที่มีค่าสูงสุดดังแสดงในภาพที่ 4.1



ภาพที่ 4.1 เปรียบเทียบการจัดกลุ่มความหมายภาพส่วนบุคคลด้วยค่า F_1 (%)

จากภาพที่ 4.1 จะเห็นว่าการจัดกลุ่มภาพส่วนบุคคลเป็น 9 กลุ่มจะได้ค่า F_1 ที่สูงถึง 72.11% ด้วยวิธี FCM และค่า F1 ของวิธีจัดกลุ่มแบบ KHM ได้ 65.44% แต่เมื่อมีการจัดกลุ่มเหตุการณ์กิจกรรมเป็น 10 กลุ่มกลับได้ค่า F_1 ของ FCM ที่ต่ำลงมาอยู่ที่ 64.56% และค่า F_1 ของ SOM อยู่ที่ 59.43% ดังนั้นเมื่อทำการเปรียบเทียบและเฉลี่ยแล้วจะเห็นว่าการแบ่งกลุ่มเหตุการณ์กิจกรรมเป็น 9 กลุ่มจะได้ค่า F_1 สูงกว่าการแบ่งแบบอื่นดังนั้นจึงเลือกการแบ่งแบบ 9 กลุ่มและจะทำการทดลองเพื่อจัดกลุ่มและแสดงค่าความถูกต้องอย่างละเอียดในส่วนถัดไป

4.2.2 การจัดกลุ่มความหมายภาพส่วนบุคคล

หลังจากทำการทดลองเพื่อหาจำนวนกลุ่มเหตุการณ์กิจกรรมที่เหมาะสมสำหรับการจัดกลุ่มภาพส่วนบุคคลที่สัมพันธ์กับจำนวนฟีเจอร์และวิธีการจัดกลุ่มทั้งหมด ผลที่ได้จากการคัดเลือกจำนวนกลุ่มเหตุการณ์กิจกรรมในการทดลองขั้นต้นได้ทั้งหมด 9 กลุ่มประกอบด้วย beach fun, skiing, graduation, wedding, birthday party, Christmas, yard park, ball game และ family time จากการทดลองสามารถแสดงผลได้ในตาราง Confusion Matrix ที่ 4.1 ถึง 4.4 ด้วยวิธีการ K-means, SOM, KHM และ FCM ตามลำดับ

Confusion Matrix (%)									
Clusters	**beach fun**	**skiing**	**graduation**	**wedding**	**birthday**	**christmas**	**yard park**	**ball game**	**family time**
beach fun	**41**	13	6	6	8	4	7	3	12
skiing	11	**43**	7	9	5	6	5	3	11
graduation	5	7	**40**	5	9	6	8	11	9
wedding	3	4	5	**41**	10	7	6	9	15
birthday	5	5	8	7	**44**	8	8	7	8
christmas	8	9	7	9	8	**42**	5	4	8
yard park	11	9	8	9	5	7	**36**	5	10
ball game	6	4	7	7	8	8	6	**42**	12
family time	9	8	7	9	8	9	6	9	**35**
Accuracy rate	**40.44**								

ตารางที่ 4.1 ผลลัพธ์การจัดกลุ่มความหมายภาพส่วนบุคคลด้วย $K-means$

จากตารางที่ 4.1 แสดงการจัดกลุ่มความหมายภาพส่วนบุคคล 9 กลุ่มด้วยวิธีการ $K-means$ ได้ค่าความถูกต้องเฉลี่ย 40.44% สำหรับภาพกลุ่ม family time และ yard park มีค่าความถูกต้อง 35% และ 36% ซึ่งจะเห็นว่าการจัดกลุ่มด้วยวิธีแบบ $K-means$ ได้ค่าความถูกต้องที่ค่อนข้างน้อยกว่าเมื่อเปรียบเทียบด้วยค่าความถูกต้องกับวิธีการจัดกลุ่มแบบ SOM, KHM และ FCM ได้ค่าความถูกต้องเฉลี่ย 60.33% ,65.44% และ 72.11% ตามลำดับดังนั้น $K-means$ ก็พบว่ายังไม่เหมาะสมกับชุดของข้อมูลในลักษณะนี้

Confusion Matrix (%)

Clusters	beach fun	skiing	graduation	wedding	birthday	christmas	yard park	ball game	family time
beach fun	**63**	5	6	8		2	1	6	9
skiing	3	**57**	2	11	4	8	5	2	8
graduation	2	5	**62**	6	9	4			12
wedding	1	6	4	**61**	8	4	8		8
birthday	7	1	2	7	**59**	4	5		15
christmas	3	3	4	2	8	**57**	5	6	12
yard park	1	2	1	3	9	1	**64**	8	11
ball game	5	8	1	3		2	5	**67**	9
family time	8	9	3	4	5	5	4	9	**53**
Accuracy rate	**60.33**								

ตารางที่ 4.2 ผลลัพธ์การจัดกลุ่มความหมายภาพส่วนบุคคลด้วย SOM

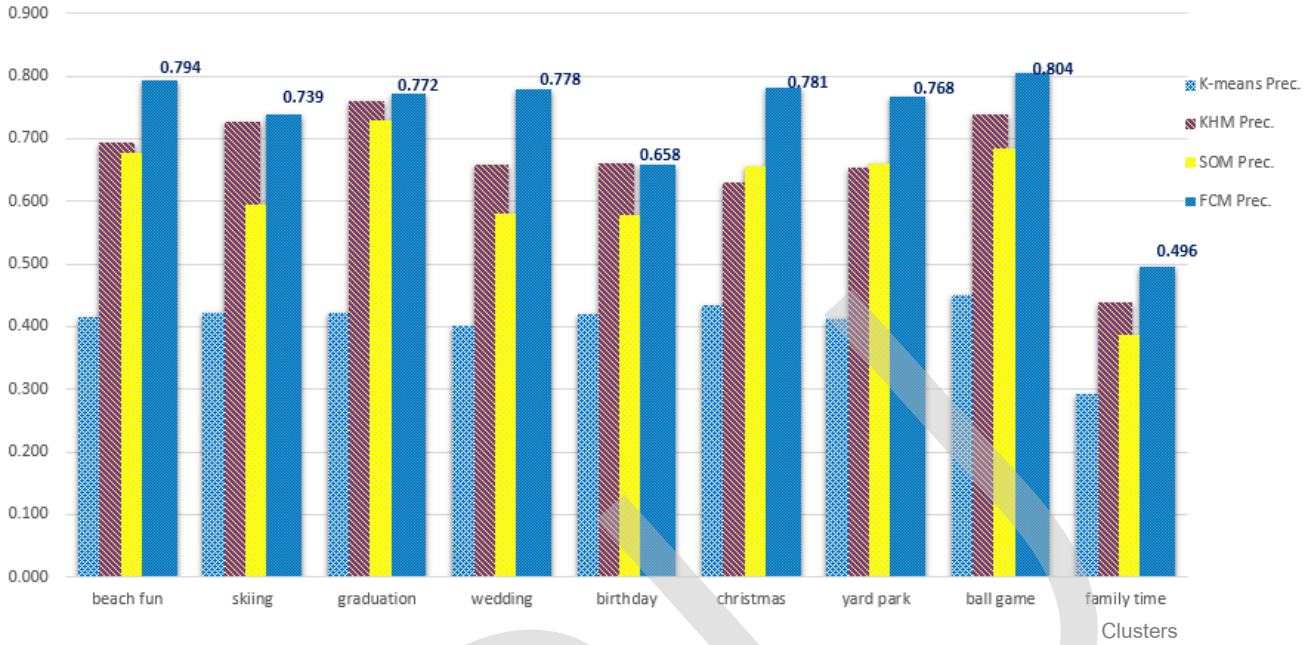
จากผลการทดลองได้แสดงในตารางที่ 4.2-4.3 สำหรับการจัดกลุ่มความหมายภาพส่วนบุคคลด้วย SOM, KHM และ FCM จะได้ค่าความถูกต้องเฉลี่ย 60.33%, 65.44% และมีการจัดกลุ่มได้ค่าเฉลี่ยความถูกต้องที่สูงสุดที่ 72.11% จากการจัดกลุ่มด้วยวิธี FCM ตามลำดับจากตารางแสดงผลการทดลองจะสังเกตได้ว่าการกลุ่มที่มีข้อผิดพลาดมากที่สุดคือกลุ่ม family time จะได้ค่าความถูกต้องสูงสุดเพียง 65% ด้วยการจัดกลุ่มแบบวิธี FCM และ 53% ด้วยวิธี SOM การจัดกลุ่ม beach fun ได้ค่าความถูกต้องสูงสุดถึง 77% ด้วยการจัดกลุ่มแบบวิธี FCM และ 75% ด้วยการจัดกลุ่มแบบวิธี KHM

Confusion Matrix (%)									
Clusters	**beach fun**	**skiing**	**graduation**	**wedding**	**birthday**	**christmas**	**yard park**	**ball game**	**family time**
beach fun	**75**	2	1	4		4	2	3	9
skiing	5	**64**	2	3	4	2	5	2	13
graduation	1	2	**70**	9	7	5	2	1	3
wedding	2	3	3	**69**	5	3	1	2	12
birthday	5	2	1	7	**62**	7	5	3	8
christmas	3	8	4	3	4	**63**	6	4	5
yard park	5	3	4	5	1	5	**64**	5	8
ball game	5	2	2	3	2	3	4	**68**	11
family time	7	2	5	2	9	8	9	4	**54**
Accuracy rate					**65.44**				

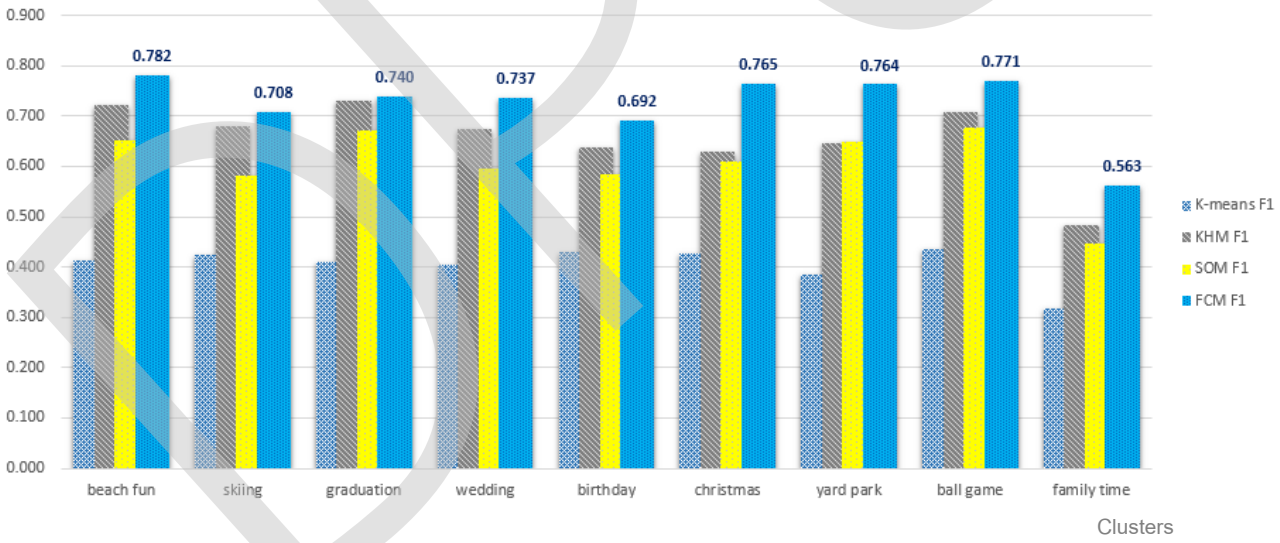
ตารางที่ 4.3 ผลลัพธ์การจัดกลุ่มความหมายภาพส่วนบุคคลด้วย KHM

Confusion Matrix (%)									
Clusters	**beach fun**	**skiing**	**graduation**	**wedding**	**birthday**	**christmas**	**yard park**	**ball game**	**family time**
beach fun	**77**	3		2	1	1	2	6	8
skiing	3	**68**	3	4	3	7	2	4	6
graduation	2	4	**71**	1	12		2	1	7
wedding		3	3	**70**	6	3	3	1	11
birthday	7		3	5	**73**	1	2		9
christmas	2	6			7	**75**	3		7
yard park			7	3	5		**76**	4	5
ball game	2	5			1	2	3	**74**	13
family time	4	3	5	5	3	7	6	2	**65**
Accuracy rate					**72.11**				

ตารางที่ 4.4 ผลลัพธ์การจัดกลุ่มความหมายภาพส่วนบุคคลด้วย FCM



ภาพที่ 4.2 กราฟแสดงการเปรียบเทียบค่าความแม่นยำของการจัดกลุ่มความหมายภาพส่วนบุคคล

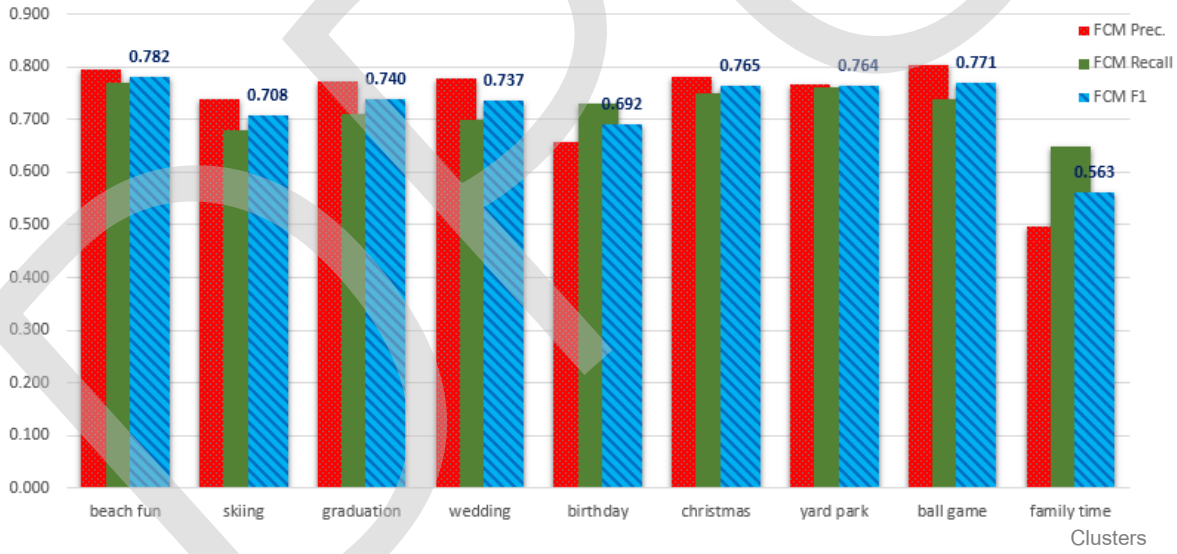


ภาพที่ 4.3 กราฟแสดงการเปรียบเทียบ F_1 ของการจัดกลุ่มความหมายภาพส่วนบุคคล

จากภาพที่ 4.2 แสดงการเปรียบเทียบค่าความแม่นยำของการจัดกลุ่มความหมายภาพส่วนบุคคลในแต่ละกลุ่มภาพเหตุการณ์ด้วยวิธีการจัดกลุ่มแบบ $K-means$, SOM, KHM และ FCM จะสังเกตได้ว่าในกลุ่มเหตุการณ์กิจกรรมของ graduation และ birthday ได้รับค่าความแม่นยำของ KHM และ FCM ที่ใกล้เคียงกันคือประมาณ 77.2% และ 65.8% ตามลำดับ แต่การจัดกลุ่มด้วย FCM ที่

ได้รับค่าความแม่นยำสูงสุดคือกลุ่มภาพเหตุการณ์กิจกรรม ball game ได้ค่าสูงถึง 80.4% แต่กลุ่มภาพเหตุการณ์กิจกรรม family time ได้เพียง 49.6% เท่านั้นแต่อย่างได้ก็ตามค่าความแม่นยำของการจัดกลุ่มด้วย FCM มีค่าสูงกว่าการจัดกลุ่มด้วยวิธีการอื่น ๆ

จากภาพที่ 4.3 กราฟแสดงการเปรียบเทียบ F_1 ของการจัดกลุ่มความหมายภาพส่วนบุคคล ในแต่ละกลุ่มภาพเหตุการณ์ด้วยวิธีการจัดกลุ่มแบบ $K-means$, SOM, KHM และ FCM จะเห็นว่าการจัดกลุ่มด้วย ในแต่ละกลุ่มภาพเหตุการณ์กิจกรรมด้วยวิธีการจัดกลุ่มแบบ $K-means$ มีค่า F_1 ที่ต่ำที่สุดซึ่งแปรผันตามค่าความถูกต้องที่ได้รับจากตารางข้างต้น แต่ค่า F_1 ของกลุ่มภาพเหตุการณ์กิจกรรม yark park ที่ถูกจัดกลุ่มด้วยวิธี SOM และ KHM ได้ค่าของ F_1 ที่มีค่าเท่ากันคือ 65% และกลุ่มเหตุการณ์กิจกรรมของ graduation ได้รับค่า F_1 ของ KHM และ FCM ที่ใกล้เคียงกันคือประมาณ 74%



ภาพที่ 4.4 กราฟแสดงการเปรียบเทียบผลการจัดกลุ่มความหมายภาพส่วนบุคคลด้วย FCM

จากภาพที่ 4.4 เป็นการแสดงผลการทดลองการจัดกลุ่มความหมายภาพส่วนบุคคลที่จะแสดงการเปรียบเทียบค่าความแม่นยำ ค่าความระลึก และค่า F_1 ของวิธีการ FCM กลุ่มเหตุการณ์กิจกรรมของ beach fun จะได้รับ ค่าความแม่นยำ ค่าความระลึก ค่าF_1 สูงถึง 79.4% 77% และ 78.2% ตามลำดับ ส่วนกลุ่มเหตุการณ์กิจกรรมของ ball game จะได้รับ ค่าความแม่นยำ ค่าความระลึก ค่าF_1 สูงถึง 80.4% 74% และ 77.1% ตามลำดับ แต่กลุ่มเหตุการณ์กิจกรรมของ family time จะได้รับ ค่า

ความแม่นยำ ค่าความระลึก ค่า F_1 น้อยที่สุดเมื่อเปรียบเทียบกับกลุ่มเหตุการณ์กิจกรรมด้วยกันเอง 49.6% 65% และ 56.3% ตามลำดับ และได้แสดงตัวอย่างภาพผลลัพธ์จากการการจัดกลุ่มเหตุการณ์กิจกรรมแสดงในภาพที่ 4.5 ประกอบด้วย 9 กลุ่มเหตุการณ์กิจกรรม ดังนี้ beach fun, skiing, graduation, wedding, birthday party, Christmas, yard park, ball game และ family time

(ก) กลุ่มภาพ beach fun

(ข) กลุ่มภาพ skiing

(ค) กลุ่มภาพ graduation

(ง) กลุ่มภาพ wedding

(จ) กลุ่มภาพ birthday party

(ฉ) กลุ่มภาพ Christmas

(ช) กลุ่มภาพ yard park

(ซ) กลุ่มภาพ ball game

(ฌ) กลุ่มภาพ family time

ภาพที่ 4.5 ตัวอย่างผลลัพธ์ของจัดกลุ่มความหมายภาพส่วนบุคคลด้วยข้อความกิจกรรม

บทที่ 5
สรุปผลการทดลอง

ในบทนี้ได้ทำการสรุปผลการทดลอง และข้อเสนอแนะ รวมทั้งข้อผิดพลาดที่เกิดขึ้นจากการทดลอง รวมไปถึงสิ่งที่ควรจะปรับปรุงเพิ่มเติม เพื่อแก้ไขข้อผิดพลาดและแนวทางการทำวิจัยต่อในเรื่องของการแปลความหมายภาพซึ่งเป็นงานวิจัยที่ปัจจุบันได้มีนักวิจัยให้ความสนใจอย่างแพร่หลาย สามารถนำไปประยุกต์ใช้ได้หลายด้าน เช่น การแพทย์ การขนส่งคมนาคม การรักษาความปลอดภัย หรือทางภาคอุตสาหกรรม เป็นต้น

5.1 สรุปและอภิปรายผลการวิจัย

งานวิจัยนี้ได้เสนอหัวข้อวิจัยทางด้านการประมวลผลภาพ ในส่วนของการวัดความคล้ายกันของภาพ เพื่อให้ได้ความหมายของภาพที่อยู่ในหมวดหมู่เดียวกัน โดยปกติทั่วไปนั้นการใช้อัลกอริทึมที่มาสกัดข้อมูลภาพนั้นมักจะใช้สกัดเพียงข้อมูลที่เกิดขึ้นภายในภาพ แล้วนำมาประมวลผลเพื่อใช้ในการสืบค้นข้อมูลภาพ แต่ปัจจุบันได้มีการนำคำหลักที่ได้จากการให้ความหมายของการแท็กวัตถุบนภาพมาหาความสัมพันธ์ภายใน โดยพยายามหาความสัมพันธ์ที่คล้ายกันของวัตถุในหมวดหมู่เดียวกัน ในงานวิจัยนี้ก็เช่นเดียวกันแต่จะเฉพาะเจาะจงลงในกลุ่มของภาพส่วนบุคคลเฉพาะเหตุการณ์กิจกรรม โดยใช้ความสัมพันธ์ที่เกิดจากขนาด ตำแหน่งและคำหลักของวัตถุแต่ละตัวที่ปรากฏบนภาพเป็นข้อมูล โดยข้อมูลคำหลักที่ได้มาจาก WordNet [Miller, George A.,1990] ซึ่งภายในคำหลักแต่ละคำมีความสัมพันธ์กันตามความหมายของภาพ ในการจัดกลุ่มภาพเหตุการณ์กิจกรรม โดยทำการทดลองเพื่อเปรียบเทียบการจัดกลุ่มทั้งหมด 4 วิธีการคือ ทฤษฎีการจัดกลุ่มแบบ K-means (K-means), การจัดกลุ่มข้อมูลแบบเคฮาร์โมนิกมีน (K-Harmonic Means), การจัดกลุ่มแบบการจัดระเบียบตัวเอง (Self-Organizing Maps) และ การจัดกลุ่มแบบฟัซซี่ซีมีน (Fuzzy C-Means) ลงในกลุ่มภาพส่วนบุคคลที่เป็น 9 กลุ่มเหตุการณ์กิจกรรม ประกอบด้วย beach fun, skiing, graduation, wedding, birthday party, Christmas, yard park, ball game และ family time ในงานวิจัยนี้ยังคงเป็นอีกแนวทางหนึ่งที่สามารถนำวิธีการที่นำเสนอเข้ามาประยุกต์เพื่อให้สามารถนำวิธีการมาใช้ช่วยในแปลความหมายของภาพได้

จากการทดลองในบทที่ 4 สามารถสรุปได้ว่า การจัดกลุ่มด้วย FCM จะได้ค่าความถูกต้องเฉลี่ยถึง 72.11% สามารถจัดกลุ่ม beach fun ได้ค่าความถูกต้องอยู่ที่ 77%, กลุ่ม skiing ได้ค่าความถูกต้องอยู่ที่68%, กลุ่ม graduation ได้ค่าความถูกต้องอยู่ที่ 71%,กลุ่ม wedding ได้ค่าความถูกต้องอยู่ที่ 70%, กลุ่ม birthday party ได้ค่าความถูกต้องอยู่ที่ 73%, กลุ่ม Christmas ได้ค่าความถูกต้องอยู่ที่ 75%, กลุ่ม yard park ได้ค่าความถูกต้องอยู่ที่ 76%, กลุ่ม ball game ได้ค่าความถูกต้องอยู่ที่ 74% และกลุ่ม family time ได้ค่าความถูกต้องที่ต่ำที่สุดจากทุกกลุ่มจะได้อยู่ที่ 65% แต่อย่างไรก็ตามค่าความถูกต้องที่ได้จากการจัดกลุ่มด้วย FCM ได้มากกว่าวิธีการอื่นๆ เพราะฉะนั้นจากการทดลองสามารถสรุปได้ว่า การที่นำทฤษฎีของการจัดกลุ่ม FCM และการใช้ข้อมูลที่เกิดขึ้นภายในภาพเป็นฟีเจอร์มาช่วยในการจัดกลุ่มภาพส่วนบุคคล สามารถช่วยในการจัดกลุ่มได้เป็นอย่างดีขึ้น

(ก)ภาพเด็กเป่าเค้กกับเพื่อน

(ข) ภาพเด็กเป่าเค้กกับครอบครัว

ภาพที่ 5.1 ตัวอย่างการเปรียบเทียบความหมายกลุ่มภาพเหตุการณ์กิจกรรม

5.2 ข้อเสนอแนะ

ในงานวิจัยส่วนใหญ่เน้นไปที่การสกัดข้อมูลภาพในรูปแบบของการสกัดด้วยอัลกอริทึมที่แตกต่างกัน และค้นหาภาพเพียงความเหมือนกันของรูปทรงหรือลักษณะเฉพาะ หรือเพียงแต่วัตถุบนภาพ เท่านั้น ทั้งที่มีความหมายและไม่มีความหมาย แต่อย่างไรก็ตาม เบื้องต้นของผลลัพธ์ที่ได้จากการค้นคืน จำแนกภาพ และการจัดกลุ่มภาพ คือความเหมือนกันทางกายภาพ เช่น รูปทรง สี หรือชนิดของวัตถุ แต่ลักษณะการวิเคราะห์และพิจารณาของการเหมือนกันทางความหมายภาพ นั้นจะมีลักษณะการวิเคราะห์ที่แตกต่างกันออกไป ในอีกรูปแบบหนึ่งซึ่งเป็นรูปแบบที่เกิดจากความคิดของมนุษย์ที่มีการแปลงความจากภาพ ตัวอย่างเช่นภาพบางภาพอาจจะมีความหมายที่คลุมเครือและการ

ให้ความหมายกลุ่มแต่ละบุคคลอาจจะมีความไม่เหมือนกัน เช่นภาพที่ 5.1 (ก) ภาพเด็กเป่าเค้กจะสามารถจัดลงในกลุ่มของ birthday ได้ทันทีแต่ ภาพที่ 5.1 (ข) ภาพเด็กเป่าเค้กกับครอบครัว นั้นสามารถจัดลงกลุ่มใน birthday ได้เช่นกันแต่เมื่อดูอีกครั้งสามารถจัดลงกลุ่ม family time ได้เช่นเดียวกัน

สิ่งควรจะมีการปรับปรุงเพิ่มเติมเพื่อให้การแปลความหมายของภาพได้ดียิ่งขึ้น ก็คือการพิจารณารายละเอียดของวัตถุของภาพ ซึ่งคำหลักที่ถูกเลือกมาจากกลุ่มภาพใน LabelMe [A. Torralba,2010];[B. C. Russell,2008] เพื่อเพิ่มความหลากหลายของกลุ่มคำศัพท์ควรจะมีการจัดกลุ่มของวัตถุที่เข้ามาทำการทดลองให้มีความหมายที่รัดกุมมากขึ้น และเมื่อภาพที่มีส่วนของสภาพแวดล้อมที่คล้ายคลึงกันมาก จะถูกจัดอยู่ในกลุ่มเดียวกันทั้งที่ วัตถุภายในภาพมีความแตกต่างกันของรายละเอียดภายในซึ่งบางครั้งการจัดคำหลักจะมีผลต่อการแปลความหมายโดยตรง แต่อย่างไรก็ตามการให้ความสัมพันธ์ของวัตถุภายในภาพนั้นสามารถช่วยทำการแปลความหมายที่ได้มีความสมบูรณ์ขึ้นอย่างเห็นได้ชัดเจน เพราะฉะนั้นในงานวิจัยที่นำเสนอนี้เป็นอีกแนวทางหนึ่งที่พยายามจะคิดค้นวิธีการที่จะหาความหมายที่เกิดขึ้นจากภาพ ในอีกมุมมองหนึ่งซึ่งยังคงต้องมีการพิจารณาและวิเคราะห์ปรับปรุงการทดลองต่อไป

สำหรับในกระบวนการจำแนกข้อมูลภาพในงานวิจัยนี้ได้ ทำการพิจารณาภาพ ประกอบด้วยวัตถุของภาพเป็นหลัก และมีการใช้ความสัมพันธ์ที่เกิดจากวัตถุเพื่อให้สามารถสื่อความหมายของภาพได้มากยิ่งขึ้น ผลที่ได้จากการจำแนกภาพโดยทั่วไปสามารถที่จะจำแนกได้อย่างไม่มีปัญหา แต่เมื่อวัตถุของภาพมีจำนวนมากขึ้นทำให้เกิดการแสดงความสัมพันธ์ที่ซับซ้อนมากขึ้นตามลำดับ จนในบางครั้งไม่สามารถทำการจำแนกได้อย่างถูกต้อง รวมทั้งจำนวนของวัตถุที่มีการจัดเก็บเพื่อเป็นข้อมูล อาจจะมีข้อผิดพลาดในการให้ข้อมูลของภาพและความสัมพันธ์ หรือ ควรจะมีการเพิ่มวิธีการที่นำมาใช้ในการช่วยการจำแนกเพิ่มขึ้น เช่น การแสดงออกด้วยท่าทางของมนุษย์นั้นสามารถสื่อความหมายของภาพได้อย่างสมบูรณ์ เช่น การกระโดด การกอด การกุมขมับ การเอนหลัง เป็นต้น และสิ่งที่ควรจะมีการเพิ่มขึ้นคือการใช้หลักการของวัตถุที่โดดเด่น รวมทั้งขนาดและตำแหน่งของวัตถุที่เด่นบนภาพ เพื่อแสดงให้เห็นถึงความหมายหลักที่ควรจะนำไปพิจารณา

****

บรรณานุกรม

A. J. Lipton, H. Fujiyoshi, R.S. Patil. "Moving Target Classification and Tracking from Real-time Video," *IEEE Workshop on Applications of Computer Vision*, 1998.

A. Sorokin and D. Forsyth. "Utility Data Annotation with Amazon Mechanical Turk," In First IEEE Workshop on Internet Vision (CVPR), 2008.

A. Torralba, B. C. Russell, J. Yuen, "LabelMe: Online Image Annotation and Applications," Proceedings of the IEEE, Vol. 98, n. 8, August 2010, pp. 1467 – 1484.

Adrian Popescu, Gregory Grefenstette, Pierre-Alain Moëllic. "Improving Image Retrieval Using Semantic Resources," In Collection of Advances in Semantic Media Adaptation and Personalization, 2008, pp. 75-96.

Apostol Natsev, Atul Chadha, Basuki Soetarman, and Jeffrey S. Vitter, "CAMEL: concept annotated image libraries," Proc. SPIE 4315, Storage and Retrieval for Media Databases 2001.

B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. "LabelMe: A Database and Web-based Tool for Image Annotation," Intl. J. Computer Vision, vol 77, numbers 1-3, May, 2008, pp. 157–173.

Barnard, K., Duygulu, P., de Freitas, N., Forsyth, D., Blei, D., Jordan. "M.I.: Matching words and pictures", Journal of Machine Learning Research 3, 2003, pp. 1107–1135.

Benitez, A.B., and S.-F. Chang. "Semantic Knowledge Construction From Annotated Image Collections", *International Conference On Multimedia & Expo (*ICME-2002), Lausanne, Switzerland, Aug 26-29, 2002.

Benjamin Yao, Xiong Yang and Tianfu Wu, Image Parsing with Stochastic Grammar: The Lotus Hill Dataset and Inference Scheme, First Workshop SIG-09: First International Workshop on Stochastic Image Grammars, Miami, Miami, Florida, June 21, 2009

Website: The Corbis Corporation home page: http://pro.corbis.com/

Carson, C., et al.. "Blobworld: A system for regionbased image indexing and retrieval," *In Proceedings International Conference Visual Information System,* 1999.

Chenliang Xu, Shao-Hang Hsieh, Caiming Xiong, Jason J. Corso , "Can Humans Fly? Action Understanding With Multiple Classes of Actors", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 2264-2273

Chris Stauffer and Eric Grimson. "Learning Patterns of Activity Using Real-Time Tracking", *IEEE Transactions on Pattern Recognition and Machine Intelligence (TPAMI)*, 22(8):747-757, 2000.

D. Comaniciu, R. Visvanathan, and P. Meer. "Kernel based object tracking", *IEEE Trans. Pattern Anal. Mach. Intell*, 25(5) , 2003, pp. 564–575.

D. Comaniciu, V. Ramesh, and P. Meer. "Real-time tracking of non rigid objects using mean shift", *Proceedings on Computer Vision and Pattern Recognition*, Hilton Head, vol. 1, 2000, pp. 142–149.

D. G. Stork. "The open mind initiative", *IEEE Intelligent Systems and Their Applications*, 14(3), 1999, pp. 19–20.

David Liu, Content-Based Image Retrieval ("Where's That Image?") Machine learning project to integrate color and texture features in image retrieval, 2007. Website: http://iceboundflame.com/projects/content-based-image-retrieval

F. Porikli. "Integral histogram a fast way to extract histograms in cartesian spaces", *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

Galleguillos C., Belongie S.. "Context Based Object Categorization: A Critical Survey," Computer Vision and Image Understanding (CVIU), Vol. 114, 2010, pp. 712-722.

Galleguillos C., McFee B., Belongie and S., Lanckriet G.. "From Region Similarity to Category Discovery," *IEEE Conference on Computer Vision and Pattern Recognition* , Colorado Springs, 2011.

Galleguillos C., McFee B., Belongie S., and Lanckriet G.R G.. "Multi-Class Object Localization by Combining Local Contextual Interactions", *IEEE Conference in Computer Vision and Pattern Recognition*, 2010.

Halaschek-Wiener, C., Golbeck, J., Schain, A., Grove, M., Parsia, B., Hendler, J.. "Photostuff an image annotation tool for the semantic web," *In Proceedings International Semantic Web Conference*, 2005.

Hollink, L., Nguyen, G., Schreiber, G., Wielemaker, J., Wielinga,B., and Worring, M.. "Adding spatial semantics to image annotations," *In Proceedings International Workshop on Knowledge Markup and Semantic Annotation*, 2004.

Huda Abdulaali Abdulbaqi,Ghazali Sulong and Soukaena Hassan Hashem, "A Sketch based Image Retrieval: A Review of Literature,Journal of Theoretical and Applied Information Technology, Vol. 63 No.1,May 2014.

I Simon, N Snavely, SM Seitz, "Scene summarization for online image collections," Computer VisionICCV 2007. IEEE 11th International Conference on, 1-8, 2007.

Ismail Haritaoglu. "A Real Time System for Detection and Tracking of People and Recognizing Their Activities", Phd. Proposal, University of Maryland at College Park, 1998.

Ismail Haritaoglu, D.Harwood, and L.Davis. "Ghost: A Human Body Part Labeling System Using Silhouettes", *In Proceedings of the 4th International Conference on Pattern Recognition*, Brisbane, August 1998.

J. Han, K.N. Ngan, M. Li and H. Zhang, "Learning semantic concepts from user feedback log for image retrieval," To appear in IEEE International Conference on Multimedia and Expo 2004, Taipei, Taiwan, 2004.

Jain, A. K. and Vilaya, A.. "Image Retrieval Using Color and Shape", Pattern Recognition, vol. 29, no. 8, 1996, pp. 1233-1244.

James Z. Wang, Jia Li, and Gio Wiederhold. "SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol 23, no.9, 2001, pp. 947-963.

Javed Ahmed, M. N. Jafri, and Mubarak Shah,Muhammad Akbar, "Real-time edge-enhanced dynamic correlation and predictive open-loop car-following control for robust tracking*," Machine Vision and Applications*, vol. 19 , Issue 1, January, 2008, pp. 1-25.

Javier Álvez, Jordi Atserias, Jordi Carrera, Salvador Climent, Egoitz Laparra, Antoni Oliver, and German Rigau. "Complete and Consistent Annotation of WordNet using the Top Concept Ontology," *In Proceedings of LREC*, 2008.

Jeroen Steggink and Cees G. M. Snoek. "Adding Semantics to Image-Region Annotations with the Name-It-Game," Multimedia Systems, 2011.

Jia Li, and James Z. Wang. "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, 2003, pp. 1075-1088.

Jianxiong Xiao. "SUN database: Large-scale scene recognition from abbey to zoo", Computer Vision and Pattern Recognition (CVPR), IEEE Conference, June, 2010, pp. 3485 – 3492.

Joo-Hwee Lim, Jun Li, Philippe Mulhem, Qi Tian, "Content-based summarization for personal image library," In: JCDL03: Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries , pp. 393, 2003.

L. Hollink, G. Schreiber, B. Wielinga, and M. Worring. "Classification of user image descriptions", International Journal of Human Computer Studies, vol. 61, no.5, 2004, pp. 601–626.

M. Johnson and R. Cipolla. Improved Image Annotation and Labelling Through Multi-Label Boosting. In Proc. of British Machine Vision Conf., September 2005.

M. Spain and P. Perona. "Measuring and predicting importance of objects in our visual world," Technical report, California Institute of Technology, 2007.

M.S. Ryoo, J.K. Aggarwal, "Semantic Representation and Recognition of Continued and Recursive Human Activities," Int J Comput Vis, 2009, 82, pp. 1–24.

Manolis Kamvysselis , Cognitive abstraction approach to sketch-based image retrieval Massachusetts Institute of Technology, 1999.

Mathias Lux , Jutta Becker , Harald Krottmaier. " Semantic Annotation and Retrieval of Digital Photos," In Proc. Of CAiSE Proceedings Information Systems for a Connected Society, 2003.

Mathias Lux. "Caliph & Emir: MPEG-7 Photo Annotation and Retrieval," Proceedings of the Seventeen ACM International Conferences on Multimedia, Beijing, China, 2009, pp. 925-926.

Miller, George A.. "WordNet: An on-line lexical database," International Journal of Lexicography, Vol. 3, 1990, pp. 235–312.

N. Chinpanthana. "Integrating Qualitative Features with Feature Selection for Semantic Image Classification", *International Conference on Management Technology and Applications*, Singapore, Sept., 2010.

N. Chinpanthana. "Extracting Features with Structural Skeleton Framework for Semantic Image Classification by using Supporting Vector Machine", *The 3rd International Conference on Advanced Computer Theory and Engineering*, Chengdu, China, 2010.

N. Chinpanthana and T. Phiasai., "Kernel-based on Data Fusion for Image Classification with Body Energy Action Model", International Journal of Signal Processing Systems, vol. 1, no. 2, December, 2013.

NIST TRECVID 2015 website: http://trecvid.nist.gov/

Philippe Mulhem, and Joo Hwee Lim. "Symbolic photograph content-based retrieval", *In Proceedings of the eleventh international conference on Information and knowledge management*, McLean, Virginia, USA, 2002, pp. 94 – 101.

P.S. Hiremath, Prema. T. Akkasaligar and Sharan. Badiger. "Comparison of Wavelet Based Despeckling Of Medical Ultrasound Images", *International Conference on Advances in Computer Vision and Information Technology 2007*, pp. 1026-1031.

Qian Huang, B. Dom, D.Steele, J. Ashley and W. Niblack. "Foreground/Background Segmentation of Color Images by Integration of Multiple Cues", *Proceedings of the IEEE International Conference on Image Processing (ICIP95),* 1995.

R. C. Gonzalez, R. E. Woods, and S. L. Eddins. *Digtal Image Processing Using MATLAB*, Pearson Education Pte. Ltd., 2004.

R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*, New York: Wiley, 1973.

R. Zhao and W. I. Grosky. "Narrowing the Semantic Gap–Improved Text-Based Web Document Retrieval Using Visual Features", *IEEE Transactions on Multimedia*, vol. 4, no. 2, 2002, pp. 189-200.

Rudolph Arnheim. *Art and Visual Perception A Psychology of the Creative Eye*, University of California Press, Ltd., 1974, pp. 11-15.

Russell, B.C., Torralba, A., Murphy, K.P., and Freeman, W.T.. "LabelMe: a database and web-based tool for image annotation," *International Journal Computer Vision*, vol.77, 2008.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R.. "Content-Based Image Retrieval at the End of the Early Years", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, 2000.

T. Luis, B. Chitta, and K. Seungchan, "Fuzzy c-means clustering with prior biological knowledge" Journal of Biomedical Informatics, Vol.42, No.1, pp.74-81, 2009.

T. Mitchell. *Machine Learning*, McGraw Hill , 1997.

Tele Tan, Jiayi Chen, Philippe Mulhem and Mohan Kankanhalli. "SmartAlbum – A Multi-Modal Photo Annotation System", *In Proceedings of the ACM Multimedia*, Juanles-Pins, France, Dec, 2002.

Teuvo Kohonen, “Self-Organizing Maps,” Springer Series in Information Sciences, Vol. 30, 1995; Second edition, 1997.

The Corbis Corporation: http://pro.corbis.com ค้นคืนเมื่อวันที่ 9 ตุลาคม 2556.

Vailaya, A. Figueiredo, M.A.T. Jain, and A.K. Hong-Jiang Zhang. “ Image classification for content-based indexing*,” IEEE Transactions on Image Processing,* Vol. 10, issue 1, 2001.

Vasileios Mezaris, Ioannis Kompatsiaris, and Michael G. Strintzis. “An Ontology Approach to Object-Based Image*,” In Proc. IEEE International Conference on on Image Processing*, 2003.

Volkmer, T., Smith, and J.R., Natsev. “A web-based system for collaborative annotation of large image and video collections: an evaluation and user study,” *In: Proceedings ACM international conference on Multimedia*, 2005, pp. 892–901.

Von Ahn, L., and Dabbish, L.. “Labeling images with a computer game. *In: Proceedings SIGCHI Conference on Human Factors in Computing Systems*,” 2004, pp. 319–326.

Von Ahn, L., Liu, and R. Blum, M.. “Peekaboom: a game for locating objects in images”, *In: Proceedings SIGCHI conference on Human Factors in Computing Systems*, 2006, pp. 55–64.

W. Ma, and Manjunath B.. “NETRA: A toolbox for navigating large image database,” In Proc. of the IEEE Int. Conf. on Image Processing, vol. 1, 1997, pp. 568-571.

Wei Yang, Ping Luo, Liang Lin, “Clothing Co-Parsing by Joint Image Segmentation and Labeling”, CVPR 2014.

Winn J., Criminisi, and A. Minka, T.. “Object categorization by learned universal visual dictionary,” *In Proceeding ICCV*, 2005, pp. 1800–1807.

Y. Wu, J.-Y. Bouguet, A. Nefian, and I. Kozintsev, “Learning concept templates from web images to query personal image databases,”in IEEE Int. Conf. Multimedia and Expo, 2007, pp. 1986–1989.

Y. Wu, I. Kozintsev, J.-Y. Bouguet, and C. Dulong, “Sampling strategies for active learning in personal photo retrieval,” in IEEE Int. Conf. Multimedia and Expo, 2006, pp. 529–532.

Xin Xu, Jinshan Tang, Xiaolong Zhang,Xiaoming Liu, Hong Zhang and Yimin Qiu, “Exploring Techniques for Vision Based Human Activity Recognition: Methods, Systems, and Evaluation,” Journal Sensors, 2013, 13, pp. 1635-1650.

Xin-Jing Wang, Lei Zhang, Xirong Li, Wei-Ying Ma, “Annotating images by mining image search results,” Pattern Analysis and Machine Intelligence, IEEE, 2008

Yao, B., Yang, X., and Zhu, S.-C. "Introduction to a Large-scale General Purpose Ground Truth Database: Methodology, Annotation Tool and Benchmarks," *In: Energy Minimization Methods in Computer Vision and Pattern Recognition*, vol. 4679, 2007, pp. 169–183.

Zhang, B., Hsu, M., and Dayal, "K-harmonic-means-a data clustering algorithm," Technical Report HPL, HP Laboratories, PaloAlto,1999.

Zhang, B., Hsu, M., and Dayal, U, "K-Harmonic Means," Proceedings of International Workshop on Temporal, Spatial and Spatio-Temporal Data Mining, TSDM2000, Lyon, 2000.

Zinger S., Millet, C., Mathieu, B., Grefenstette, G., H`ede, P., and Mo¨ellic, P.A.. "Extracting an Ontology of Portrayable Objects from WordNet," *In: Proceeding MUSCLE/ImageCLEF Workshop on Image & Video Retrieval Evaluation*, 2005, pp. 17–23.

ประวัติผู้วิจัย

ผู้ช่วยศาสตราจารย์นัศพ์ชาณัณ ชินปัญช์ธนะ

สถิติประยุกต์ (คอมพิวเตอร์) สถาบันบัณฑิตพัฒนบริหารศาสตร์
วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์) มหาวิทยาลัยหอการค้าไทย

ปัจจุบัน
อาจารย์ประจำภาควิชาเทคโนโลยีสารสนเทศธุรกิจ คณะเทคโนโลยีสารสนเทศ
มหาวิทยาลัยธุรกิจบัณฑิตย์
คณะกรรมการปรับปรุงกลุ่มผลิตชุดวิชาคอมพิวเตอร์เบื้องต้น
คณะกรรมการกลุ่มผลิตชุดวิชาการระบบสำนักงานอัตโนมัติและพาณิชย์อิเล็กทรอนิกส์
คณะกรรมการกลุ่มผลิตชุดวิชาการจัดการเว็บเซอร์วิส
คณะกรรมการกลุ่มผลิตชุดวิชาการสื่อสารข้อมูลและระบบเครือข่ายคอมพิวเตอร์
สาขาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

รางวัลและเกียรติประวัติ

- รับรางวัล The best paper เรื่อง ,"การคัดเลือกข้อมูลเพื่อใช้จำแนกโครงสร้างท่าทางมนุษย์ด้วยซัพพอร์ตเวคเตอร์แมชชีน" จาก The 8th National Conference on Applied Computer Technology and Information Systems 2015.
- รับรองคุณภาพการสอน ปีการศึกษา 2557 จาก มหาวิทยาลัยธุรกิจบัณฑิตย์
- รับรางวัลนักวิจัยดีเด่น สาขาเทคโนโลยีสารสนเทศ 2558 จาก มหาวิทยาลัยธุรกิจบัณฑิตย์
- Microsoft Office Specialist (MOS) Certification (PowerPoint and Word) 2015
- Certificate of Achievement for iOS Development June 2013
- ได้รับทุนการศึกษาประเภทเรียนดี มหาวิทยาลัยหอการค้าไทย พ.ศ. 2534 – 2537
- ได้รับปริญญาตรีเกียรตินิยมอันดับ 1 จาก มหาวิทยาลัยหอการค้าไทย

งานวิจัยที่สนใจ

สนใจงานวิจัยทางด้านการค้นคืนความหมายภาพ (Image retrieval) ทั้งภาพธรรมชาติทั่วไป (natural images) และ ภาพที่เป็นส่วนบุคคล (personal images) ในหัวข้อเกี่ยวกับ การค้นคืน human activity หรือในหัวข้อการแปลความหมายภาพ (semantic human image) โดยเจาะจงทางด้านการใช้ท่าทางของมนุษย์เพื่อแสดงถึงความหมายของภาพ รวมทั้งการจำแนกภาพต่าง ๆ เป็นกลุ่ม (Image classification) และการประยุกต์หลักการประมวลผลภาพ เพื่อนำมาใช้ในทางอุตสาหกรรมได้จริง

หนังสือ

นัศพ์ชาณัณ ชินปัญช์ธนะ. หน่วยที่ 13 การจัดแฟ้มข้อมูลและการป้องกันระบบคอมพิวเตอร์, *สถาปัตยกรรมคอมพิวเตอร์และระบบปฏิบัติการ (Computer Architecture and Operation Systems) (99315)*, มหาวิทยาลัยสุโขทัยธรรมาธิราช, สาขาวิทยาศาสตร์และเทคโนโลยี 2557.

นัศพ์ชาณัณ ชินปัญช์ธนะ. หน่วยที่ 1 ความรู้เบื้องต้นเกี่ยวกับเครือข่ายคอมพิวเตอร์ และ หน่วยที่ 3 ชุมชนเครือข่าย, *หลักการบริหารและจัดการเครือข่าย (Network Management) (99412)* ,มหาวิทยาลัยสุโขทัยธรรมาธิราช, สาขาวิทยาศาสตร์และเทคโนโลยี 2556.

นัศพ์ชาณัณ ชินปัญช์ธนะ. หน่วยที่ 10 โครงสร้างแบบต้นไม้ และ หน่วยที่ 11 โครงสร้างแบบกราฟ, *โครงสร้างข้อมูลและขั้นตอนวิธี (Data Structure and Algorithms) (993314)*, มหาวิทยาลัยสุโขทัยธรรมาธิราช, สาขาวิทยาศาสตร์และเทคโนโลยี 2556.

นัศพ์ชาณัณ ชินปัญช์ธนะ. หน่วยที่ 12 โลจิสติกส์และโซ่อุปทานในระบบพาณิชย์อิเล็กทรอนิกส์ และหน่วยที่ 13 การชำระเงินในระบบพาณิชย์อิเล็กทรอนิกส์, *ระบบสำนักงานอัตโนมัติและพาณิชย์อิเล็กทรอนิกส์ (Office Automation System and Electronic Commerce) (99311)*, มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขาวิทยาศาสตร์และเทคโนโลยี 2555.

นัศพ์ชาณัณ ชินปัญช์ธนะ. หน่วยที่ 3 แบบจำลองเครือข่ายและโพรโทคอลกับการบริการผ่านเว็บ หน่วยที่ 5 ความรู้พื้นฐานของภาษาเอ็กซ์เอ็มแอล, *เทคโนโลยีการบริการผ่านเว็บและการประยุกต์ (Web Service) (99301)*, มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขาวิทยาศาสตร์และเทคโนโลยี 2556.

นัศพ์ชาณัณ ชินปัญช์ธนะ. หน่วยที่ 3 ตัวกลางในการสื่อสารข้อมูล และ หน่วยที่ 10 โพรโทคอล, *การสื่อสารข้อมูล*

และระบบเครือข่ายคอมพิวเตอร์ (Data Communications and Networking),มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขา วิทยาศาสตร์และเทคโนโลยี, 2553.

งานวิจัย

นัศพ์ชาณัณ ชินปัญช์ธนะ, การแปลความหมายภาพด้วยแนวคิดพื้นฐานความสัมพันธ์ของกราฟแบบลำดับชั้น (Semantic Annotation Model of Hierarchical Relationships based on Conceptual Graph Representation), มหาวิทยาลัยธุรกิจบัณฑิตย์, 2556.

นัศพ์ชาณัณ ชินปัญช์ธนะ, การรู้จำท่าทางมนุษย์จากภาพส่วนบุคคลผ่านการวิเคราะห์การใช้พลังงานร่างกาย (Human Action Recognition from Personal Photos via Analysis of Body Energy Expenditure), มหาวิทยาลัยธุรกิจบัณฑิตย์, 2555.

นัศพ์ชาณัณ ชินปัญช์ธนะ. การแปลความหมายภาพด้วยวิธีการวัดความคล้ายกันของกราฟแบบจับคู่, มหาวิทยาลัยธุรกิจบัณฑิตย์, 2554.

นัศพ์ชาณัณ ชินปัญช์ธนะ, ระบบตรวจนับวัตถุอัตโนมัติด้วยเทมเพลตแมชชิ่งแบบนอร์มอลไลซ์คอรรีเลชั่น (Automatic Counting Objects System by using Template Matching with Normalized Correlation) 2553.

นัศพ์ชาณัณ ชินปัญช์ธนะ. การจำแนกความหมายของภาพจากวัตถุโดยใช้หลักการโครงสร้างสเกตรีตรอน, มหาวิทยาลัยธุรกิจบัณฑิตย์, 2552.

ผลงานทางวิชาการ

นัศพ์ชาณัณ ชินปัญช์ธนะ, “การจำแนกความหมายภาพด้วยการสกัดฟีเจอร์จากโครงสร้างสเกตรีตรอนบนพื้นฐานแนวคิดกราฟลำดับชั้น” , วารสารแม่โจ้เทคโนโลยีสารสนเทศและนวัตกรรม,1, 2558.

นัศพ์ชาณัณ ชินปัญช์ธนะ, “การคัดเลือกข้อมูลเพื่อใช้จำแนกโครงสร้างท่าทางมนุษย์ด้วยซัพพอร์ตเวคเตอร์แมชชีน” , “The 8th National Conference on Applied Computer Technology and Information Systems”, วันที่ ๓๐ – ๓๑ มกราคม ๒๕๕๘, มหาวิทยาลัยนครพนม,จังหวัดนครพนม. (The best paper award)

นัศพ์ชาณัณ ชินปัญช์ธนะ, “การจำแนกความหมายภาพด้วยเรเดียลเบสิสฟังก์ชันบนพื้นฐานของแนวคิดกราฟลำดับชั้น” ,นานาวิทยาการวิจัยและนวัตกรรมเพื่อการพัฒนาที่ยั่งยืนสู่เอเชีย” วันศุกร์ที่ 10 ตุลาคม 2557 , วิทยาลัยบัณฑิตเอเชีย จังหวัดขอนแก่น

N. Chinpanthana and T. Phiasai., “Kernel-based on Data Fusion for Image Classification with Body Energy Action Model”, *International Journal of Signal Processing Systems*, vol. 1, no. 2, December, 2013.

N. Chinpanthana and T. Phiasai., “Kernel-based on Data Fusion for Image Classification with Body Energy Action Model”, *The 2013 5th International Conference on Signal Processing Systems* (ICSPS 2013), Sydney, Australia, 2013.

N. Chinpanthana, “Semantic Similarity Measure with Conceptual Graph-Based Image Annotation”, *International Conference on Advanced Computer Science Application and Technologies* (ASCAT 2012), Kualalumpur,Malaysia, Nov., 2012.

N. Chinpanthana and T. Phiasai., “Automatic Counting System With Normalized Correlation Coefficient Template Matching”, *International Conference on Computer and Information Technology*, Amsterdam, Netherlands, July 13-15, 2011.

N. Chinpanthana, “Integrating Qualitative Features with Feature Selection for Semantic Image Classification”, *International Conference on Management technology and applications* (ICMTA2010), Singapore,10-12 Sept.,2010.

N. Chinpanthana and T. Phiasai., “Multi-Layer Perception Networks for Semantic Image Classification with Structural Skeleton Framework”, *International Technical Conference on Circuit/Systems Computers and Communications*, Pattaya, Thailand,July., 2010.

N. Chinpanthana, “Extracting Features with Structural Skeleton Framework for Semantic Image Classification by using Supporting Vector Machine”, *The 3rd International Conference on Advanced Computer Theory and Engineering*, Chengdu, China, 20-22 Aug.,2010.

N. Chinpanthana, “Semantic Semantic Salient Images Based on Similarity Matching with Conceptual Graph”, *International Technical Conference on Circuits/Systems, Computers and Communications* , Jeju, Korea 2009.

S. Chinpanchana, S. Maneewongvatana, and B. Thipakorn, “Semantic Human Image Classification Based on Energy Action Model with Essential Reference points”, *International Symposium on Communications and Information Technologies*, 16-19 Oct, Sydney, AUS, 2007.

S. Chinpanchana, S. Maneewongvatana, and B. Thipakorn, “High-Level Semantic Image Classification by Using Energy Expenditure”, *International Workshop on Smart Info-Media Systems*, Thailand, 2007.

S. Chinpanchana, "Semantic Human Action Classification Based on Energy-Action Model", *Tencon 2006 IEEE Region 10* , Hongkong, China, 2006.

S. Chinpanchana, S. Maneewongvatana, and B. Thipakorn, "Semantic Personal Image Classification by Energy Expenditure," *International Symposium on Communications and Information Technologies*, Beijing, China, 2005.

S. Chinpanchana, S. Maneewongvatana and B. Thipakorn, "Semantic Personal Image Pattern Classification Based on Human Body," *Asia Information Retrieval Symposium*, Beijing, China, Oct. 2004.

S. Chinpanchana and B. Thipakorn, "Semantic Classification of Personal Images Based on Human Action and Associate Bayesian Rule," *International Technical Conference on Circuits/Systems, Computers and Communications*, Sendai, Japan, 2004.

****