



รายงานการวิจัย

เรื่อง

การแปลความหมายภาพด้วยแนวคิดพื้นฐานความสัมพันธ์ของกราฟแบบลำดับชั้น

Semantic Annotation Model of Hierarchical Relationships based on Conceptual
Graph Representation

โดย

นศพัชชาณัณ ชินปัญชธนะ

มหาวิทยาลัยธุรกิจบัณฑิตย์

รายงานผลการวิจัยนี้ได้รับทุนอุดหนุนจากมหาวิทยาลัยธุรกิจบัณฑิตย์

พ.ศ. 2556

ชื่อเรื่อง : การแปลความหมายภาพด้วยแนวคิดพื้นฐานความสัมพันธ์ของกราฟแบบลำดับชั้น
ผู้วิจัย : นศัพชาณณ ชินปัญชระนะ สถาบัน : มหาวิทยาลัยธุรกิจบัณฑิต
ปีที่พิมพ์ : พุทธศักราช 2557 สถานที่พิมพ์ : มหาวิทยาลัยธุรกิจบัณฑิต
แหล่งที่เก็บรายงานการวิจัยฉบับสมบูรณ์ : ศูนย์วิจัยมหาวิทยาลัยธุรกิจบัณฑิต
จำนวนหน้าวิจัย : 69 หน้า ลิขสิทธิ์ : มหาวิทยาลัยธุรกิจบัณฑิต
คำสำคัญ : การประมวลผลภาพ, กราฟ, ลำดับชั้น, การจำแนกภาพ

บทคัดย่อ

การจำแนกความหมายภาพเป็นหัวข้องานวิจัยที่ยังคงท้าทายอย่างมากในสาขาการประมวลผลภาพ มีนักวิจัยหลายกลุ่มพยายามปรับปรุงวิธีการเพื่อแก้ไขปัญหาของการแทนความหมายภาพด้วยการประมวลผลภาพระดับต่ำที่มีการใช้สถิติเข้ามาช่วยเพื่ออธิบายภาพ แต่อย่างไรวิธีการเหล่านั้นไม่สามารถที่จะนำมาใช้แทนความหมายของภาพได้อย่างแท้จริง

งานวิจัยนี้ได้นำเสนอวิธีการใหม่สำหรับการแปลความหมายภาพด้วยแนวคิดพื้นฐานความสัมพันธ์ของกราฟแบบลำดับชั้น เพื่อใช้ในการจำแนกความหมายของภาพ สำหรับแนวคิดนี้จะใช้กราฟในการแทนความสัมพันธ์ของวัตถุในภาพ วิธีการนี้มีทั้งหมด 3 ส่วนคือ (1) การใช้เครื่องในการเท็กภาพ (2) การประมวลผลด้วยกราฟแบบลำดับชั้น (3) การจำแนกความหมาย ผลที่ได้จากการจำแนกความหมายภาพด้วยวิธีที่นำเสนอใหม่นี้สามารถจำแนกความหมายของภาพได้ดีกว่าวิธีอื่นๆ และได้ค่าความถูกต้องสูงถึง 84.82%

Title : Semantic Annotation Model of Hierarchical Relationships based on Conceptual Graph Representation

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

Year of Publication : 2014 **Publisher** : Dhurakijpundit University.

Sources : Dhurakijpundit University Research Center.

Number of Pages : 69 Pages **Copyright** : Dhurakijpundit University.

Keyword : Image processing, Graph, Hierarchical, Image Classification

Abstract

Semantic classification is a challenging task in the field of image processing. Many researchers have attempted to improve semantic models such as developing more sophisticated models, or generating intermediate representations by making statistics on low level descriptions. However, the methods are rather rudimentary and are not specific enough for representing the actual meaning.

In this paper, we present a novel concept called the hierarchical representation graph for producing more semantic classification. This concept is formulated on a graph which captures the relationships among objects in the images. The approach is composed of three main phases: (1) image annotation tool, (2) hierarchical relationship graph, and (3) image classification. The experimental results indicate that our proposed approach offers significant performance improvements in the interpretation of semantic images, compared, with the maximum of 84.82% accuracy.

กิตติกรรมประกาศ

งานวิจัยชิ้นนี้สำเร็จล่วงได้ทั้งนี้เพราะได้รับความอนุเคราะห์ การสนับสนุน และแรงผลักดัน และอีกหลายฝ่ายที่ได้ให้ความช่วยเหลือเพื่อข้อมูลที่เป็นประโยชน์ในการทดลองโปรแกรม บันทึกผลการทดลอง และงานวิจัยนี้จะไม่สมบูรณ์ได้ หากไม่ได้ข้อเสนอแนะที่เป็นประโยชน์จากท่านอาจารย์ ผู้ทรงคุณวุฒิ ผู้ซึ่งให้คำแนะนำที่ดีและมุมมองที่ผู้วิจัยได้นำมาปรับปรุงให้สมบูรณ์ยิ่งขึ้น

นอกจากนี้ใคร่ขอขอบคุณมหาวิทยาลัยธุรกิจบัณฑิต ผู้ให้ทุนสนับสนุนงานวิจัยชิ้นนี้ หากงานวิจัยเล่มนี้มีข้อผิดพลาด ประการใดขออภัยไว้ ณ ที่นี้ด้วย

นศัพชาณณ ชินปัญชณะ

พฤษภาคม 2557

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	(1)
สารบัญตาราง	(3)
สารบัญรูปภาพ	(4)
บทที่ 1 บทนำ	1
1.1 ความเป็นมาของปัญหา	1
1.2 วัตถุประสงค์	3
1.3 สมมติฐาน	4
1.4 ขอบเขตของการวิจัย	4
1.5 ประโยชน์ที่คาดว่าจะได้รับ	4
1.6 นิยามคำศัพท์	5
บทที่ 2 ทฤษฎีและงานวิจัย	7
2.1 ทฤษฎีเบื้องต้น	7
2.2 การประมวลผลภาพดิจิทัล	18
2.3 การจำแนกประเภทข้อมูลภาพ	23
2.4 การวัดประสิทธิภาพ	32
บทที่ 3 ขั้นตอนวิธีการดำเนินงานวิจัย	34
3.1 ขั้นตอนการเตรียมข้อมูล	34
3.2 ขั้นตอนการประมวลผล	46
3.3 ขั้นตอนการวัดประสิทธิภาพ	48
บทที่ 4 ผลการทดลอง	49
4.1 การกำหนดข้อมูลภาพ	49
4.2 ผลการจำแนกความหมายของข้อมูลภาพ	51

บทที่ 5 สรุปผลการทดลอง

5.1	สรุปและอภิปรายผลการวิจัย	59
5.2	ข้อเสนอแนะ	59
		60
บรรณานุกรม		62
ประวัติผู้วิจัย		69



สารบัญตาราง

ตารางที่	หน้า
2.1 การวัดประสิทธิภาพ	34
3.1 ตัวอย่างงานวิจัยที่ใช้ฐานข้อมูลคำหลักในการให้ความหมายภาพ	38
3.2 แหล่งฐานข้อมูลของคำหลัก	41
4.1 ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี naïve-Bayes	54
4.2 ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี Multilayer	55
4.3 ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี SOM	55
4.4 ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี Similarity Matching	55
4.5 ตารางเปรียบเทียบผลลัพธ์การจำแนกข้อมูล	56

สารบัญรูปภาพ

รูปภาพที่	หน้า
2.1 การประมวลผลแบบคั่นคั่นด้วยคุณลักษณะพีเจอร์ระดับต่ำ	8
2.2 ตัวอย่างขององค์ประกอบของภาพ	9
2.3 ตารางแสดงการวัดประสิทธิภาพ	10
2.4 การแท็กองค์ประกอบของภาพ	11
2.5 การหาตำแหน่งวัตถุของภาพด้วยโครงสร้างสเกตตรีตรอน	13
2.6 การแสดงตำแหน่งของฟิกเชลบนภาพดิจิทัลด้วยเมทริกซ์	15
2.7 แบบจำลองของภาพด้วยระนาบ	15
2.8 การจัดเก็บและเรียกใช้ข้อมูลภาพดิจิทัลลงในเมทริกซ์	17
2.9 เซลระบบประสาทของสิ่งมีชีวิตและโครงข่ายประสาทเทียม	20
2.10 ตัวแบบของนิวรอนในคอมพิวเตอร์	21
2.11 สถาปัตยกรรมแบบป้อนไปข้างหน้า	22
2.12 สถาปัตยกรรมแบบมีการป้อนไปเวียนกลับ	22
2.13 โครงสร้างแบบป้อนไปข้างหน้าหลายชั้น	23
2.14 การเรียนรู้แบบมีการสอน	24
2.15 การเรียนรู้แบบไม่มีการสอน	25
2.16 โหนดการเรียนรู้ของแผนผังการจัดระเบียบตัวเอง	28
2.17 การหาผู้ชนะสำหรับการกระจายตัวของข้อมูล x_i หรือโหนดเพื่อนบ้าน	29
3.1 ขั้นตอนการจำแนกกลุ่มความหมายภาพ	36
3.2 ภาพตัวอย่างที่ถูกแท็กด้วยคำหลัก	37
3.3 ส่วนย่อยของโครงสร้างลำดับชั้นของคำหลัก“dog” บน WordNet	39

3.4	ความสัมพันธ์ของโครงสร้างคำหลักบน WordNet	49
3.5	โปรแกรม LabelMe บนเบราว์เซอร์	42
3.6	การเลือกสัดส่วนของวัตถุบนภาพ	43
3.7	ตัวอย่างวัตถุที่ถูกแท็กด้วยโปรแกรม LabelMe	43
3.8	ตัวอย่างภาพที่ถูกแท็กคำหลักบนภาพ ด้วยโปรแกรม LabelMe	44
3.9	ตัวอย่างของคำหลักที่พบบ่อยในการให้ความหมายวัตถุ	45
3.10	ตัวอย่างของคำหลักในการให้ความหมายวัตถุบนภาพ	45
3.11	ตัวอย่างความสัมพันธ์ของวัตถุด้วยแนวคิดกราฟ	47
3.12	ความสัมพันธ์ของวัตถุแทนด้วยแนวคิดกราฟ	48
4.1	แสดงความผิดพลาดของการเคลื่อนที่ของถุงปูนซีเมนต์บนสายพาน	52
4.2	แสดงการเปรียบเทียบผลการทดลองการตรวจนับการเคลื่อนถุงปูนซีเมนต์บนสายพานด้วย โปรแกรมประมวลผลภาพแบบแทมเพลตแมชชีง	54
4.3	ตัวอย่างผลลัพธ์ของการจำแนกความหมายภาพ	57

บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

สมัยก่อนการประมวลผลภาพ จะอยู่ในรูปของการประมวลผลสัญญาณแอนะล็อก (analog) โดยใช้อุปกรณ์จับภาพแต่งภาพที่ถ่ายออกมายังคงเป็นภาพขาวดำและไม่ชัดเจนเท่าที่ควร ทำให้ส่วนใหญ่ภาพถ่ายมักจะถูกถ่ายเฉพาะภาพเหตุการณ์ที่สำคัญหรือบุคคลสำคัญเท่านั้น เนื่องจาก อุปกรณ์ที่นำมาใช้ในการถ่ายภาพ หรือบันทึกภาพนั้นซับซ้อนและมีความยุ่งยากในการใช้งาน รวมทั้งราคาสูง แต่อย่างไรก็ตามได้มีการพัฒนาปรับปรุงเทคโนโลยีเพื่อตอบรับกับความต้องการที่เพิ่มขึ้น ทำให้ อุปกรณ์ถ่ายภาพในปัจจุบันมีราคาถูกลง และสะดวกต่อการใช้งานมาก และพัฒนาต่อเนื่องไปเป็น อุปกรณ์ถ่ายภาพแบบดิจิทัล (digital) ทำให้ภาพถ่ายดิจิทัล มีจำนวนเพิ่มมากขึ้นอย่างต่อเนื่อง ด้วยเหตุนี้ทำให้เกิดปัญหาในการจัดเก็บข้อมูลภาพรวมถึงการค้นหาข้อมูลภาพ (image retrieval) ที่เพิ่มมากขึ้นอย่างต่อเนื่อง จนทำให้เกิดคำถามเดิม ๆ ว่าควรจะใช้วิธีการใดจึงจะสามารถจัดเก็บอย่างมีระบบ และสามารถที่จะค้นหาภาพที่ต้องการได้อย่างถูกต้อง

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

ปัจจุบันมีงานวิจัยหลายกลุ่ม [Qian Huang,1995];[Vailaya A.,2001];[W. Ma,1997];[C. Carson,1999] พยายามนำเทคนิคต่าง ๆ เข้ามาใช้ในการค้นหาภาพเพื่อให้ผลลัพธ์ตรงกับความต้องการของผู้ใช้ โดยส่วนใหญ่แล้วจะมีการแบ่งประเภทของการนำเสนอรูปแบบงานวิจัยออกเป็น 3 ส่วน [Hafidha Bouyerbou,2011] ดังนี้

- (1) กระบวนการค้นคืนภาพตามเนื้อหาสาระ (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
- (2) การแทนคำอธิบายประกอบภาพ (Annotation) เป็นงานวิจัยอีกกลุ่ม ที่พยายามจะใช้เทคนิคของการเข้าใจความหมายของภาพแทน การสืบค้นแบบข้างต้นเรียกว่า การประมวลผลภาพระดับสูง (high level image processing) วิธีที่สร้างคำอธิบายประกอบภาพแบบเป็น คำหลัก (keyword) หรือคีย์ (key) ลงบนวัตถุในภาพ เช่น “building”, “car” เป็นต้น 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) เมื่อเปรียบเทียบผลลัพธ์ของการค้นคืนที่ค่อนข้างแม่นยำมากกว่าการใช้พีเจอร์ระดับต่ำ แต่อย่างไรก็ตามยังขึ้นอยู่กับว่าอัลกอริทึมที่ถูกนำมาใช้ร่วมด้วยนั้นจะเป็นลักษณะใด เช่น การใช้เทคนิคหาความสัมพันธ์ของแท็กชื่อวัตถุบนภาพ [Benitez A.B, 2001];[R. Zhao, 2002];[Philippe Mulhem,2002] ซึ่งเป็นการสอดคล้องกันด้วยความหมายตามพจนานุกรม การใช้ความสัมพันธ์ของความหมายที่เหมือนกันของคำ (synonym) การค้นหาภาพด้วยเทคนิคนี้จะได้ผลลัพธ์ที่ค่อนข้างขึ้นกับคำหลักที่ถูกให้ความหมายไว้ในภาพ หรือวัตถุบนภาพส่วนใหญ่ แต่อย่างไรก็ตามการสืบค้นภาพที่ถูกบรรยายนั้นส่วนใหญ่จะเป็นจะภาพส่วนบุคคล (personal images) ในบางงานวิจัยพยายามที่จะใส่ข้อมูลเหตุการณ์ต่างๆ จนครบถ้วนในรูปแบบของ ใคร ทำอะไร ที่ไหน เมื่อไหร่ (who, what, when, where) ในการใส่ข้อมูลบางครั้ง เป็นข้อมูลที่นอกเหนือ หรือเกินความจำเป็นโดยใช้เหตุ ข้อมูลเหล่านี้อาจจะไม่มีความจำเป็นเลยสำหรับการสืบค้นข้อมูลหรือ แทนความหมายของภาพจริงหรือบางครั้งอาจจะเป็นข้อมูลที่มีความเป็นส่วนตัวจนเกินไป และใช้คำศัพท์ซ้ำซ้อน ฟุ่มเฟือย จึงทำให้ความหมายภาพที่แท้จริงจะไม่สามารถเกิดขึ้นได้และมีนักวิจัยบางกลุ่มพยายามที่แก้ไขถึงความยุ่งยากลำบากในการใส่ข้อมูลบนภาพ (image annotation) และจำกัดขอบเขตคำศัพท์ของภาพให้รัดกุมยิ่งขึ้น

(3) การใส่คำอธิบายภาพแบบอัตโนมัติ (Automatic Image Annotation: AIA) [N. Serrano, 2002];[L. Zhang, 2010] เป็นการใช้รูปแบบแนวความคิดของการแปลความหมายภาพที่ได้มาจากการเรียนรู้อัตโนมัติจากจำนวนภาพตัวอย่าง โดยทั่วไปแล้ว AIA จะมีลักษณะแบบการใช้ข้อมูลรวมกันของส่วนประกอบ (component) บนภาพ เช่น การแบ่งส่วนภาพออกเป็น ส่วนย่อย (unit) เซกเมนต์ (segment) หรือ รีเจนท์ (region) บล็อก (block) และรูปแบบของการเก็บค่านั้นจะถูกแทน (representation) เป็นลักษณะเฉพาะของแต่ละส่วนด้วย และลักษณะเฉพาะที่ถูกแทนบนภาพทั้งหมดที่ได้ จะถูกนำมาใช้เพื่อการเรียนรู้เป็นลักษณะเฉพาะโดยรวมแทนภาพ (global features) [Z. Guo, 2007];[A. Yavlinsky, 2005]

มีงานวิจัยมากมายที่สนใจด้าน AIA โดยเน้นไปที่การเรียนรู้คำอธิบายความหมาย (Semantic learning/ annotation) และการจำแนกความหมายภาพ แต่อย่างไรก็ตามซอฟต์แวร์ส่วนใหญ่ๆนั้น จะทำการสืบค้นภาพในรูปแบบของการเปรียบเทียบคำหลักเป็นคำต่อคำตามที่มีการเก็บข้อมูลไว้เท่านั้น เพราะฉะนั้นผลลัพธ์ของกลุ่มภาพที่ได้จะไม่ได้ขึ้นกับความหมายของภาพอย่างแท้จริง แต่ขึ้นกับคำหลักที่มีการเก็บข้อมูลลงไปบนภาพเท่านั้น จึงทำให้ความหมายของภาพโดยรวมนั้นยังไม่ได้สื่อออกมาให้เห็นอย่างเด่นชัด สำหรับความหมายของภาพจะเกิดจากการแปลความหมายรวมของทุกวัตถุที่ปรากฏบนภาพ และความสัมพันธ์ของที่เกิดขึ้นของคำหลักวัตถุ ดังนั้นในงานวิจัยนี้ได้นำเสนอในรูปแบบของการแทนคำอธิบายข้อมูลภาพ ด้วยลำดับชั้นความสัมพันธ์ของกราฟ (Hierarchical Relationship Graph) ข้อมูลวัตถุภายในภาพถูกแทนความสัมพันธ์กันแบบลำดับชั้นด้วย แนวคิดกราฟ (Conceptual Graph) ข้อมูลกราฟจะประกอบด้วยวัตถุและจะมีแสดงความสัมพันธ์ระหว่างวัตถุที่เกิดขึ้นภายในภาพ เป็นความสัมพันธ์แบบลำดับชั้น ทำให้ข้อมูลทั้งหมดนั้นสามารถแทนเป็นคำความหมายของภาพได้

จากที่กล่าวมาข้างต้นในงานวิจัยนี้จึงได้นำเสนอการปรับปรุงการสืบค้นข้อมูลภาพด้วยการแทนความสัมพันธ์ของกราฟแบบลำดับชั้น เพื่อให้ได้ผลลัพธ์ของความหมายภาพอย่างแท้จริง โดยใช้วิธีการแทนข้อมูลภาพแบบลำดับชั้นความสัมพันธ์ของกราฟ โดยทำเปรียบเทียบความเหมือนกันของความหมายภาพด้วยการหาความเหมือนด้วยทฤษฎีความเชื่อเบย์ (naïve-Bayes) และ วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) แบบ One-against-One และแบบ One-against-all ด้วยฟังก์ชันเคอร์เนล แบบเชิงเส้น (Linear Kernel) และ เรเดียลเบสิสฟังก์ชัน (Radial Basis Function)

1.2 วัตถุประสงค์

1. เพื่อค้นหาเทคนิคใหม่ที่สามารถนำมาใช้ในการค้นหาความเหมือนกันของภาพ
2. เพื่อสร้างรูปแบบใหม่ในการแทนข้อมูลภาพ
3. เพื่อสร้างแบบจำลองวิเคราะห์การค้นหาภาพรูปแบบใหม่
4. เพื่อพัฒนาเทคนิคในการหาความเหมือนกันของภาพ
5. สามารถนำไปประยุกต์ใช้ในการค้นคืนภาพได้

1.3 สมมติฐาน

งานวิจัยนี้สกัดพีเจอร์จากวัตถุภายในภาพตามมาตรฐานของฐานข้อมูลคำหลัก [Miller, George A.,1990];[Gruber, T.R.,1993] เพื่อเป็นฐานข้อมูลสำหรับการทดลอง โดยที่ข้อมูลคำหลักมีการแทนคำอธิบายข้อมูลด้วยลำดับชั้นความสัมพันธ์ล่วงหน้าบนฐานข้อมูล ทำให้วิธีการที่นำเสนอมีการแสดงถึงความสัมพันธ์ของคำหลักในภาพที่เพิ่มขึ้นและสามารถแยกแยะถึงความหมายของภาพได้อย่างรัดกุม

1.4 ขอบเขตของการวิจัย

1. ข้อมูลภาพที่นำมาวิเคราะห์เป็นข้อมูลภาพดิจิทัลที่เห็นวัตถุชัดเจน
2. ภาพที่มีวัตถุเกิดปรากฏขึ้น สำหรับภาพที่ไม่มี ความหมายหรือเป็นภาพที่มีความหมายกำกวม หรือภาพที่สามารถแปลได้หลายความหมาย หรือภาพที่โฟกัสระยะใกล้ จะไม่นำภาพนั้นเข้ามาวิเคราะห์
3. ข้อมูลหรือวัตถุนภาพที่เป็นอินพุตบนภาพจะถูกแท็ก เป็นคำศัพท์เข้ามาก่อน
4. ในการสร้างแบบจำลองภาพ (image representation) ข้อมูลจะถูกจัดเก็บมาในรูปแบบของพีเจอร์ข้อมูลไว้แล้ว
5. ภาพจากคลังภาพจะถูกตั้งค่าเริ่มต้นโดยการใส่ข้อมูลวัตถุนภาพพร้อมความสัมพันธ์ภาพไว้ก่อนแล้ว
6. ภาพที่ถูกนำมาใช้เป็นข้อมูลทดลองจากคลังภาพ
7. การเก็บผลข้อมูลเบื้องต้นของความหมายของภาพ ใช้กลุ่มนักศึกษา และบุคคลทั่วไป เป็นกลุ่มบุคคลที่ตัดสิน 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 บทความวิจัยที่น่าเสนอในการประชุมระดับชาติ และนานาชาติ

1.5.3 ผลต่อกลุ่มผู้วิจัย

1. พบแนวทางในการคิดค้นสิ่งใหม่ ๆ ในการจัดกลุ่มและกระบวนการแปลความหมายของภาพ
2. สามารถนำผลวิจัยมาเขียนบทความลงวารสารนานาชาติ และร่วมประชุมวิชาการระดับชาติและนานาชาติ
3. พัฒนาความรู้ใหม่ ๆ หลักการแนวคิดใหม่ ให้เกิดขึ้นในกลุ่มของนักวิจัย

1.6 นิยามคำศัพท์

1. การประมวลผลภาพ (Image processing) หมายถึง การนำภาพมาผ่านกระบวนการเพื่อประมวลผลสัญญาณบนสัญญาณ 2 มิติ เช่น ภาพนิ่ง (ภาพถ่าย) หรือภาพวีดิทัศน์ (วิดีโอ) และนำมาใช้งาน
2. การจำแนกประเภทข้อมูลภาพ (Image Classification) หมายถึง การแยกแยะภาพที่มีคุณลักษณะเดียวกันออกเป็นกลุ่ม ๆ
3. การจำแนกความหมายของภาพ (Semantic Image Classification) หมายถึง การนำภาพมาแบ่งเป็นกลุ่มโดยพิจารณาจากความหมายโดยรวมของภาพ

4. การค้นคืนข้อมูล (Image Retrieval) หมายถึง การค้นหาข้อมูลของภาพจากระบบฐานข้อมูล
5. วัตถุ (Object) หมายถึง ส่วนของวัตถุบนภาพ ยกตัวอย่างเช่น คน, ต้นไม้ เป็นต้น
6. คำหลัก (Keyword) หมายถึง คำที่มีความหมายได้ใจความใช้แทนวัตถุบนภาพ
7. แท็ก (Tag) หมายถึง การกำหนดคำ หรือ คำศัพท์บนภาพ หรือเรียกว่าการ labeled
8. การแทนที่ข้อมูลภาพด้วยกราฟ (Graph representation) คือการแทนความหมายของภาพโดยใช้กราฟ
9. คุณลักษณะข้อมูล หรือ ฟีเจอร์ (Feature) หรือตัวแปร ที่ถูกสกัดออกมาจากภาพ เช่น สี (color) ลวดลาย (texture) หรือ รูปทรง (shape) รวมทั้ง วัตถุ ที่ปรากฏบนภาพเพื่อนำมาใช้ในการสืบค้นข้อมูลต่อไป
10. โครงข่ายประสาทเทียม (Artificial neural network) หมายถึง เทคนิคการพยากรณ์เหมาะสำหรับลักษณะข้อมูลซึ่งไม่ใช่เชิงเส้น

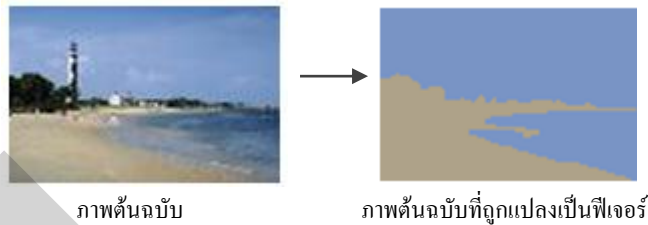
บทที่ 2

ทฤษฎีและงานวิจัย

2.1 ทฤษฎีเบื้องต้น

วิวัฒนาการของเครื่องมือและอุปกรณ์ถ่ายภาพดิจิทัลได้พัฒนาอย่างรวดเร็วจนทำให้ภาพถ่ายภาพดิจิทัล มีจำนวนเพิ่มมากขึ้น ปัญหาที่ตามมาก็คือการจัดเก็บข้อมูลภาพที่เพิ่มมากขึ้นอย่างไร ซึ่ดจำกัดนี้จะทำอย่างไรจึงจะสามารถจัดเก็บอย่างมีระบบและสามารถสืบค้นข้อมูลภาพ และจำแนกข้อมูลภาพให้ตรงตามความหมายของภาพที่ต้องการของผู้ใช้มากที่สุด ทำให้งานวิจัยในปัจจุบันที่เกี่ยวข้องกับการค้นคืนรวมทั้งการจัดกลุ่มภาพให้ตรงกับความต้องการเพิ่มมากขึ้น รวมไปถึงระยะเวลาในการสืบค้นที่น้อยลงกับปริมาณของภาพที่เพิ่มทวีคูณ ดังนั้นปัญหาดังกล่าวมาข้างต้นนั้นจึงได้รับความสนใจจากนักวิจัยหลายกลุ่ม ซึ่งเป็นงานด้านการประมวลผลภาพ (image processing) ด้านการค้นคืนสารสนเทศ (image retrieval) เป็นอีกหนทางหนึ่งในการแก้ปัญหาดังกล่าว เนื่องจากสามารถช่วยให้การค้นหาข้อมูลกระทำได้โดยสะดวกยิ่งขึ้นรวมถึง การค้นหาภาพและการจำแนกประเภทข้อมูลภาพ (image classification) เพื่อคัดเลือกภาพ เพื่อให้ตรงตามความต้องการของผู้ใช้ ยิ่งมีความจำเป็นอย่างยิ่ง

สำหรับงานวิจัยทางการประมวลผลภาพในการค้นคืนสารสนเทศกลุ่มแรก ๆ จะมีการค้นคืนตามคุณลักษณะพื้นฐานของภาพที่ถูกสกัดคุณลักษณะด้วยอัลกอริทึมต่าง 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] ด้วยคุณลักษณะพีเจอร์ระดับต่ำด้วยสี ลวดลาย และตำแหน่งของพื้นที่ของภาพ จากผลลัพธ์จะสังเกตว่าผลลัพธ์ของภาพเป็นภาพที่มีโทนสีคล้ายกันเป็นหลัก แต่มีลักษณะวัตถุที่แตกต่างกันอย่างสิ้นเชิงที่อยู่ในหมวดหมู่เดียวกัน คือ ภาพชายหาด หรือ ชายทะเล แต่มีคุณลักษณะของโทนสี และลวดลาย ที่แตกต่างกันอย่างเห็นได้ชัดเจน เมื่อนำมาจำแนกด้วยการประมวลผลภาพระดับต่ำ (low-level feature) แล้วนั้นค่อนข้างยากที่จะจัดให้หมวดหมู่เดียวกัน



ภาพต้นฉบับ

ภาพต้นฉบับที่ถูกแปลงเป็นฟีเจอร์

ก. ภาพต้นฉบับสำหรับการค้นคืนที่ถูกแปลงเป็นฟีเจอร์ระดับต่ำ

S-I-M-P-L-I-c-i-t-y

Semantics-sensitive Integrated Matching for Picture Libraries

Option 1 --> Image ID or URL

Option 2 --> **Random**

Option 3 --> Click an image to find similar images

22579 0.00 2	47908 4.00 2	9439 4.96 2	53001 5.25 2	52613 5.58 2	28435 6.06 2	13827 6.17 2	409 6.47 2
46631 6.49 2	13984 6.51 2	49297 6.91 2	11731 7.15 2	52398 7.16 2	821 7.17 2	48892 7.40 2	33226 7.49 2

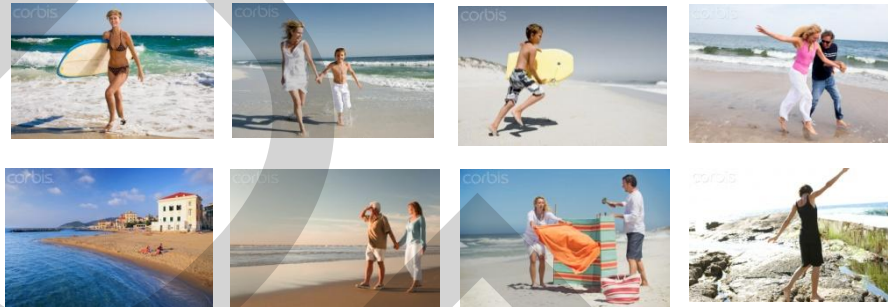
ข. ผลลัพธ์ของการค้นคืนด้วยภาพต้นฉบับ ก.

ภาพที่ 2.1 การประมวลผลแบบค้นคืนด้วยคุณลักษณะฟีเจอร์ระดับต่ำ¹

แต่อย่างไรก็ตามได้มีกลุ่มนักวิจัยที่พยายามปรับปรุงแปลงอัลกอริทึมด้วยการประมวลผลภาพระดับต่ำ เพื่อทำการค้นคืนภาพที่มีลักษณะฟีเจอร์ที่ใกล้เคียงกับภาพที่ต้องการมากที่สุด [M. Flickner, 1995];[W. Ma, 1997] การปรับปรุงเทคนิควิธีการเพื่อให้กระบวนการค้นคืนภาพได้ผลลัพธ์ที่ถูกต้องเพิ่มมากขึ้น ด้วยการนำวิธีการมาผสมผสานกันระหว่างคุณลักษณะเพื่อให้สามารถวิเคราะห์ในรูปแบบที่ซับซ้อนได้มากขึ้น เช่น การรวมเทคนิคด้วยคุณลักษณะสีและรูปร่างของภาพ เพื่อทำการค้นคืนภาพ [P.S. Hiremath, 2007] หรือมีการใช้อัลกอริทึมเพื่อทำการสกัดข้อมูลภาพเป็นฟีเจอร์ และนำฟีเจอร์นำมาใช้ในการค้นคืนภาพในรูปแบบที่แตกต่างกันออกได้ แต่ในความเป็นจริงแล้วนั้นลักษณะการมองภาพของคนโดยทั่วไปเป็นการมองจากความหมายของภาพ หรือมองจากชนิดของวัตถุของภาพ ดังแสดงในภาพที่ 2.2 แสดงภาพชายทะเล ภายในภาพประกอบด้วยวัตถุหลาย

¹ http://alipr.com/cgi-bin/zwang/regionsearch_show.cgi

ชนิดได้แก่ ทะเล และ ท้องฟ้า เพียงอย่างเดียว แต่บางภาพอาจจะประกอบด้วยวัตถุอื่น ๆ มนุษย์ สุนัข หรืออาคาร เป็นภาพประเภทเดียวกันหรือภาพที่สื่อความหมายอย่างเดียวกัน โดยที่ไม่จำเป็นต้องมีคุณลักษณะสีหรือรูปทรงแบบเดียวกันก็สามารถเป็นภาพชนิดเดียวกันได้ ดังนั้นในการค้นคืนที่ใช้คุณลักษณะพีเจอร์ระดับต่ำเพียงอย่างเดียว ทำให้ได้ผลลัพธ์ส่วนใหญ่ตรงกับคุณลักษณะของพีเจอร์ที่สกัดมาแต่ไม่ได้ตรงกับความหมายที่เกิดภายในภาพที่ต้องการอย่างแท้จริง



ภาพที่ 2.2 ตัวอย่างขององค์ประกอบของภาพ²

การวิจัยในยุคแรกเป็นความพยายามในการค้นหาข้อมูลภาพเริ่มมาจากการสกัดข้อมูลภายในภาพ (feature extraction) ออกมาเป็นส่วนๆแล้วใช้ข้อมูลระดับต่ำ เช่น สี (colour) รูปทรง (shape) และ พื้นผิว (texture) เป็นต้น เพื่อทำการสืบค้นข้อมูลภาพ ผลลัพธ์ที่ได้เป็นชุดของภาพที่เกิดจากการเปรียบเทียบจากข้อมูลจากพีเจอร์ต่างๆตามคุณลักษณะของพีเจอร์ที่สกัดได้เท่านั้น ทำให้ต้องมีการปรับปรุงพัฒนาอย่างต่อเนื่อง ด้วยเหตุนี้เองทำให้มีกลุ่มนักวิจัยที่สนใจในเรื่องของการจัดกลุ่มภาพตามความหมายเรียกว่า semantic image เป็นการจัดกลุ่มภาพตามความหมายของภาพโดยแปลความหมายของภาพจากองค์ประกอบ หรือวัตถุ (object) ที่ปรากฏในภาพนั้น ซึ่งมีวิธีการค้นคืนที่แตกต่างกันเพื่อให้ได้ผลลัพธ์ที่ตรงตามความต้องการมากที่สุด



ภาพที่ 2.3 ตัวอย่างการแท็กองค์ประกอบของภาพ³

² <http://www.corbisimages.com/>

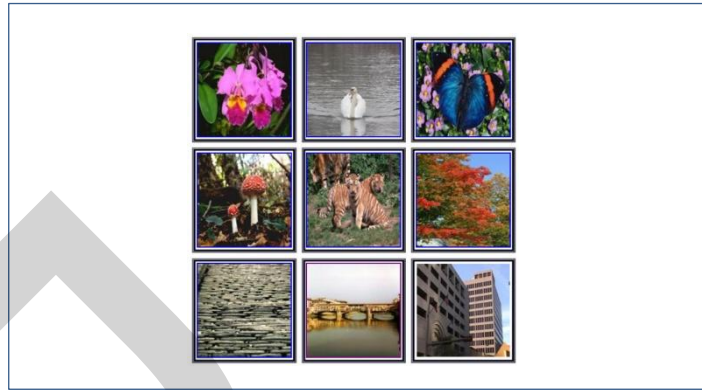
³ <http://vision.ucsd.edu/project/context-based-object-categorization>

แต่อย่างไรก็ตามได้มีงานวิจัยอีกกลุ่ม ที่พยายามจะใช้เทคนิคของการเข้าใจความหมายของภาพแทน การสืบค้นแบบข้างต้น เรียกว่า การประมวลผลภาพระดับสูง (high level image processing) งานวิจัยในกลุ่มนี้พยายามที่จะมองข้อมูลบนภาพเป็นวัตถุ (object) ที่มีความหมาย [Benitez A.B, 2002];[Galleguillos C., 2010] และแทนวัตถุนั้นๆ ด้วยคำหลัก (keyword) บนภาพ เรียกว่า การแท็ก (tag) หรือการให้ความหมายของวัตถุนั้นๆ เป็นชื่อวัตถุ หรือคำศัพท์ที่สอดคล้องกันเช่น “grass”, “plant”, “boat”, “sky” เป็นต้น ดังแสดงในภาพที่ 2.3 [Galleguillos C., 2010] และใช้ความหมายหรือคำศัพท์นั้นเพื่อทำการสืบค้นข้อมูลแทน ซึ่งเป็นการใช้ความหมายของคำศัพท์ที่มีความสอดคล้องกันด้วยความหมายตามพจนานุกรม หรือในลักษณะใช้ความสัมพันธ์ของความหมายที่เหมือนกันของคำหลัก (synonym) [Zhao T., 2001];[Benitez, A.B., 2002];[Kobus B., 2001];[Philippe M., 2002] เข้ามาใช้ในการค้นคืนข้อมูลภาพเพื่อให้ผลลัพธ์ตรงกับความต้องการของผู้ใช้ ยกตัวอย่างเช่น “stone” มีความหมายสอดคล้องกันกับ “rock” เป็นต้น จะได้ผลที่ค่อนข้างดีกว่า แต่ขึ้นอยู่กับว่าอัลกอริทึมที่ถูกนำมาใช้นั้นจะเป็นลักษณะใด

Sample queries for individual classes:

	Birds	Flowers	Textures
	Bridges and Buildings	Landscapes	Transport
	Bugs	Mammals	
Manmade	Buildings		
	Bridges		
	Misc	Aircraft	
	Transport	Ships	Cars
		Land	Buses
			Bikes
Textures	Clouds		
	Misc		
	Water		

ก. การค้นคืนด้วยการเลือกคำหลักที่เป็นหมวดหมู่



ข. การค้นคืนด้วยการเลือกภาพต้นฉบับ



ค. ผลลัพธ์ของการค้นคืนด้วยภาพต้นฉบับ

ภาพที่ 2.4 ตัวอย่างการแท็กองค์ประกอบของภาพ⁴

นักวิจัยบางกลุ่มใช้เทคนิคเพื่อสร้างความสัมพันธ์ (relationship) ของแท็กชื่อวัตถุบนภาพ [Benitez A.B, 2001];[R. Zhao, 2002];[Philippe M., 2002] ด้วยโมเดลเหตุการณ์ (event model) [Joo-Hwee L., 2003] เช่น การเชื่อมวัตถุด้วยคำต่างๆ เช่น “touch” “on” “top” เป็นต้น บางงานวิจัยได้พยายามใส่ข้อความ (context) เพื่อบรรยายความหมายของภาพ [Mathias Lux, 2003];[Mathias Lux, 2009] ในหัวข้อที่สอดคล้องกับภาพนั้นๆ เช่น “birthday party of uncle

⁴ http://amazon.ece.utexas.edu/~qasim/sample_queries.htm

Adam” หรือ “a picture showing a barking dog” แต่ภาพที่นำมาใช้จัดเก็บนั้นมักจะเป็นภาพส่วนตัว (personal images) ทำให้การค้นคืนจำกัดเพราะคำบรรยายส่วนใหญ่จะเป็นคำเฉพาะเจาะจง จึงไม่นิยมเท่าที่ควร บางงานวิจัยพยายามที่จะใส่ข้อมูลเหตุการณ์ต่าง ๆ จนครบถ้วนในรูปแบบของ ใครทำอะไร ที่ไหน เมื่อไหร่ (who, what, when, where) ดังนั้นในการใส่ข้อมูลบางครั้ง เป็นข้อมูลที่นอกเหนือ หรือเกินความจำเป็นโดยใช้เหตุ ข้อมูลเหล่านี้อาจจะไม่มีความจำเป็นเลยสำหรับการสืบค้นข้อมูล อาจจะเป็นข้อมูลที่มีความเป็นส่วนตัวจนเกินไป และใช้คำศัพท์ซ้ำซ้อน ฟุ่มเฟือย ทำให้การค้นคืนยิ่งเกิดความสับสนปนเปกันของภาพผลลัพธ์ที่ได้มา ดังแสดงในภาพที่ 2.4 ก. [Q. Iqbal, 2002] แสดงตัวอย่างการค้นคืนภาพที่มีการจัดเป็นหัวข้อ (topic) จากภาพที่ 2.4 ข. แสดงภาพต้นฉบับเพื่อสร้างคุณลักษณะหรือฟีเจอร์ที่คล้ายกันเพื่อใช้ในการค้นคืนในที่เก็บข้อมูลภาพ จากภาพที่ 2.4 ค. แสดงภาพต้นฉบับมีคำหลักเป็น “bridges” และกลุ่มผลลัพธ์ที่ได้หลังจากการใช้ฟีเจอร์ในการค้นคืนภาพ แต่ผลลัพธ์ที่ได้ออกมานั้นจะมีการผสมกับภาพอื่นที่ไม่ตรงกับคำหลักด้วย



ภาพที่ 2.5 การหาตำแหน่งวัตถุของภาพด้วยโครงสร้างสเกตตรีตรอน

การค้นหาภาพด้วยเทคนิคนี้จะได้ผลลัพธ์ที่ขึ้นกับคำศัพท์ที่ถูกแท็กไว้บนภาพยังมีการแท็กข้อมูลบนภาพมากยิ่งขึ้นสามารถหาความเหมือนกันบนภาพมากขึ้นเท่านั้น แต่ในความเป็นจริงแล้ว การแท็กข้อมูลบนภาพในปัจจุบันนั้นเป็นเพียงการหาคำศัพท์ที่ต้องการบนภาพ แต่ไม่ได้ให้ความหมายภาพโดยรวม ความหมายของภาพคือการนำวัตถุที่ปรากฏบนภาพมารวมกันเพื่อวิเคราะห์จากความคิดของมนุษย์เพื่อให้ได้ผลลัพธ์คือคำศัพท์ใหม่ที่แทนความหมายของภาพทั้งภาพ สำหรับการวิเคราะห์ความหมายของภาพจะได้รับการรับรู้ของมนุษย์ที่มองภาพนั้น จะมีกฎเกณฑ์ของการรับรู้จากการมองภาพได้กล่าวไว้เป็นทฤษฎี โครงสร้างสเกตตรีตรอน ดังแสดงในภาพที่ 2.5 ก. [Rudolph A., 1974] ได้มีการนำมาประยุกต์ใช้วิเคราะห์ในส่วนของ การจำแนกความหมายของภาพโดยรวม เพื่อให้ได้ความหมายของภาพอย่างแท้จริง โดยใช้ทฤษฎีการมองของมนุษย์ที่มีการพิจารณาจากตำแหน่งและ

ขนาดของวัตถุที่เกิดขึ้นเพื่อนำมาใช้ในการแปลความหมายของภาพ ด้วยการวัดโครงสร้างกับภาพต้นฉบับที่มีการแท็ก ดังแสดงในภาพที่ 2.5 ข. และ ค. [N. Chinpanthana., 2010];[นัศพ์ชาณัณ, 2552] แต่อย่างไรก็ตามค่าตำแหน่งในส่วนของภาพนั้นยังไม่เพียงพอต่อการนำมาใช้สำหรับการแปลความหมาย ภาพยังมีวิธีการที่ผสมผสานระหว่างการใช้แท็กคำหลักบนภาพกับการสร้างความสัมพันธ์ระหว่างวัตถุด้วยมัลติเคิลเคอร์เนล (multiple kernel) ของพื้นที่วัตถุบนภาพ [Galleguillos C., 2011];[Galleguillos C., 2010] ผลลัพธ์ที่ได้จะเป็นกลุ่มของคำหลักที่เกิดขึ้นบนภาพเสียส่วนใหญ่ และมีนักวิจัยบางกลุ่มพยายามที่จะแก้ไขถึงความยุ่งยากลำบากในการใส่ข้อมูลบนภาพ (image annotation) และจำกัดขอบเขตคำศัพท์ของภาพให้รัดกุมยิ่งขึ้น

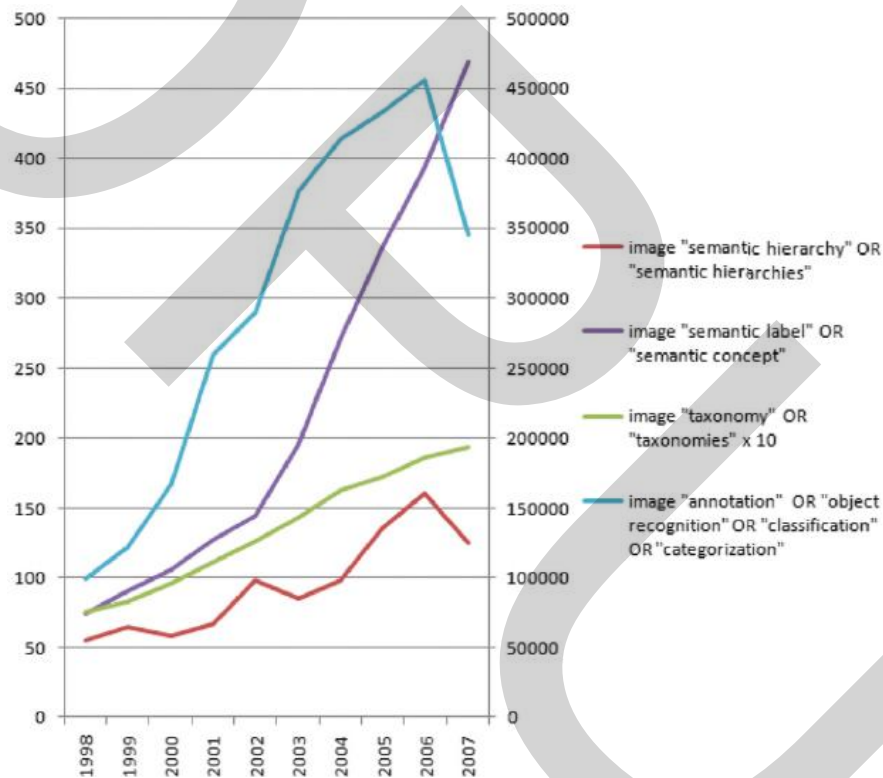


Two-time Formula One champion Mika Hakkinen drives a McLaren Mercedes F1 car down a section of the proposed F1 street circuit in Singapore March 30, 2008. Hakkinen says the first night race on a Singapore street circuit will pose unique challenges to drivers but safety concerns can be allayed by organization and preparation. Hakkinen drove on the street as part of an anti-drink driving campaign.

ภาพที่ 2.6 ตัวอย่างการใส่แท็กคำศัพท์ลงบนภาพ [A.-M. Tousch, 2011]

ปัจจุบันการพัฒนาซอฟต์แวร์เพื่อให้เกิดความสะดวกสบายในการแท็กข้อมูล และข้อมูลที่ถูกนำมาเป็นคำศัพท์นั้นแบบแผนและโครงสร้างที่แน่นอน ยกตัวอย่างเช่น Caliph & Emir [Mathias Lux, 2009], Annosearch [Xin-Jing, 2008], CAMEL [Apostol Paul N., 2001] เป็นต้น ส่วนใหญ่จะมีความสนใจในการแทนค่าความหมาย (Semantic representation) โดยขึ้นกับภาษาและโครงสร้างการแทนค่าเพื่อทำให้มนุษย์ และเครื่องคอมพิวเตอร์เกิดความเข้าใจในการประมวลผล ยกตัวอย่างจากภาพที่ 2.6 แสดงภาพรถแข่งในสนามแข่งขัน โดยมีคำอธิบายรายละเอียดของภาพด้านขวา คำศัพท์ที่จะมีเพิ่มลงในภาพ ขึ้นอยู่กับการผสมระดับคำอธิบาย สามารถแทนค่าอธิบายภาพด้วย “Formula One”, “event”, หรือ “a kind of sport” เป็นต้น แต่อย่างไรก็ตามการเพิ่มระดับการแท็กลงบนภาพเพื่อให้ได้ความหมาย เช่น “a Formula-1 car in a street”, “buildings in the

background”, “barriers delineating the circuit” และ “people” เป็นต้น บางข้อมูลสามารถได้มาจากภาพ แต่ข้อมูลบางอย่างไม่สามารถหาได้จากภาพ เช่น “Mika Hakkinen has been F1 world champion twice”, “the race he takes part in is part of an anti-drink driving campaign” หรือ “the photo has been taken on March 30, 2008” ซึ่งในภาพจะไม่สามารถปรากฏข้อมูลดังกล่าวบนภาพ S. Shatford Layne [S. Shatford Layne, 1994] ได้กำหนดวิธีการเก็บคำอธิบายรูปภาพ โดยแบ่งออกเป็น 4 ส่วนดังนี้ (1) แอทริบิวต์ชีวประวัติ เพื่อบอกสถานที่ เวลาของภาพ (2) แอทริบิวต์หัวเรื่อง เพื่อบอกเนื้อหาของภาพ (3) แอทริบิวต์ Exemplified เพื่อบอกประเภทของภาพถ่าย ภาพการ์ตูน หรือภาพร่าง เป็นต้น (4) แอทริบิวต์ความสัมพันธ์ เพื่อแสดงความสัมพันธ์ของภาพ



ภาพที่ 2.7 แนวโน้มงานวิจัยทางการแปลความหมายภาพด้วยการแท็ก จาก Google Scholar

จากภาพที่ 2.7 แสดงแนวโน้มของงานวิจัยของการแท็กความหมายภาพ (Semantic image annotations) ผลจาก Google Scholar [A.-M. Tousch, 2011] ซึ่งกล่าวถึงการใช้นวัตกรรมของคำศัพท์ในวิธีของ semantic hierarchy, semantic label (semantic concept), taxonomy และ

annotation (object recognition) จะเห็นว่าจำนวนผลลัพธ์ของการค้นหาจาก Google Scholar ที่ใช้วิธีการแท็กคำศัพท์ในรูปแบบของการรู้จำวัตถุ และการให้ความหมายภาพแบบการใช้ป้ายกำกับนั้นจะมีจำนวนที่เพิ่มขึ้นอย่างรวดเร็ว ซึ่งแตกต่างจากการใช้ taxonomy และ semantic hierarchy

The description levels suggested by Jaimes and Chang

1	Type, technique
2	Global distribution
3	Local structure
4	Global composition
5	Generic objects
6	Generic scene
7	Specific objects
8	Specific scene
9	Abstract objects
10	Abstract scene



1. photo
2. (histogram)
4. (segmentation)
5. flower, leaf, water
6. nature
7. water lily and its leaves
8. pond
10. stillness, coldness

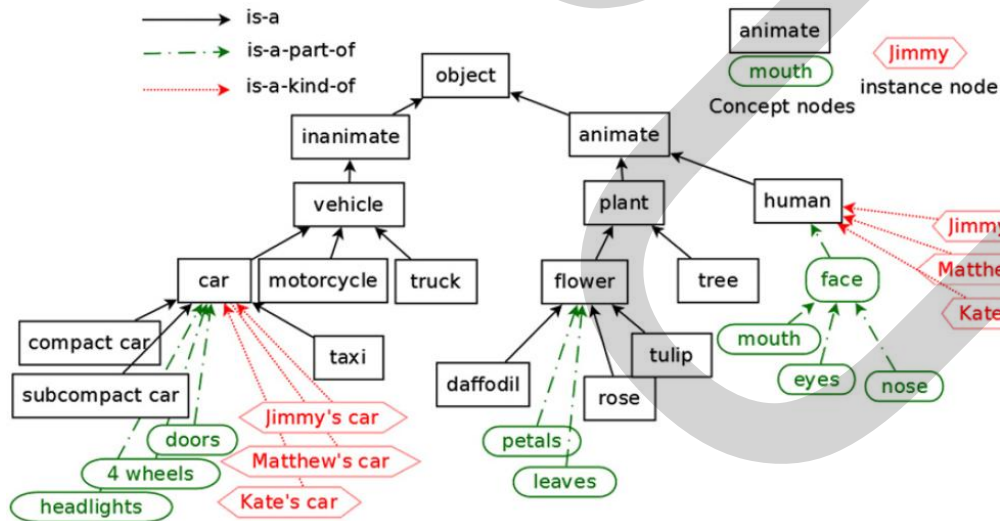
ภาพที่ 2.8 ตัวอย่างระดับคำอธิบายภาพ [A. Jaimes, 2000]

กลุ่มนักวิจัย [A. Jaimes, 2000];[L. Hollink, 2004] ได้สร้างโครงสร้างคำศัพท์ภาพแบบ 10 ระดับ เพื่อจัดเก็บข้อมูลภาพ ดังภาพที่ 2.8 แสดงตารางระดับคำอธิบายภาพโดยที่ 4 ลำดับแรกจะอ้างอิงถึง perceptual aspect หรือ syntax และ 6 ลำดับท้ายจะสัมพันธ์กับความหมายและแนวคิด ดังนั้นเมื่อทำการอธิบายความหมายของภาพตามระดับแล้วสามารถที่จะเปรียบเทียบภาพได้จากข้อมูลแต่ละระดับ ดังแสดงตัวอย่างการใส่คำอธิบายภาพตามลำดับ L. Hollink [L. Hollink, 2004] ได้มีการใช้ Unified Modeling Language (UML) ในการอธิบายลำดับการแทนคำอธิบายภาพเพื่อให้เกิดความเข้าใจวัตถุที่เกิดขึ้นบนภาพถ่ายขึ้น Hare et al. [J.S. Hare, 2006] ได้นำเสนอโครงสร้างคำศัพท์เริ่มต้นจากข้อมูลภาพและมีการแปลในรูปแบบของคำอธิบายวัตถุเป็นส่วนตามชื่อของวัตถุซึ่งมีความหมายในระดับสูง คล้ายวิธีการที่ Jaimes and Chang [A. Jaimes, 2000] ได้นำเสนอในส่วนต้น แต่ Eakins et al. [J.P. Eakins, 2004] ได้แสดงถึงความแตกต่างของนามธรรมออกมาแต่อย่างไรก็ตามพื้นฐานทั่วไปของภาพนั้นยังคงมีระดับเช่นเดียวกันกับ Jaimes and Chang [A. Jaimes, 2000]

จากตัวอย่างภาพที่ 2.6 สามารถอธิบายคำศัพท์หรือข้อมูลบนภาพได้หลายระดับขึ้นอยู่กับพื้นฐานของผู้มองภาพ เช่น an outdoor scene; a car, a street, buildings, people, sky; a town; a car race, a business district; a Formula One เป็นต้น ความสัมพันธ์ระดับชั้นของคำศัพท์สามารถถูกสร้างขึ้นเป็นคำอธิบายภาพได้ เช่น การอธิบายด้วยความสัมพันธ์แบบ Is-A และการใช้ความสัมพันธ์แบบ Part-Of รวมทั้งการใช้ความสัมพันธ์แบบ Co-occurrence เป็นต้น ดังนั้นสามารถเขียนเป็นคำอธิบายได้

McLaren MercedesFormulaOne < F1 < sports car < car
 Modern town with sky scrapers < town
 Sport car \cap spectators \sqsubset car race
 Buildings \cap street \sqsubset town

จะเห็นว่าเนื้อหาที่จะใช้เป็นคำอธิบายภาพภายในขึ้นอยู่กับวัตถุที่ผู้ใช้สนใจในการใส่คำอธิบายลงบนภาพของผู้ใช้ว่ามีลักษณะเป็นแบบใด ดังนั้นแต่ละระดับของการให้คำอธิบายลงบนภาพนั้นมีความสำคัญ ทำให้เป็นปัญหาของการใส่คำอธิบายภาพในระดับของผู้ใช้ที่มีความแตกต่างกันอย่างชัดเจน และทำให้มีนักวิจัยได้นำวิธีการปัญญาประดิษฐ์มาใช้แก้ปัญหาด้วยโครงสร้างคำศัพท์แบบออนโทโลยี (ontology) [L.M. Garshol, 2004]



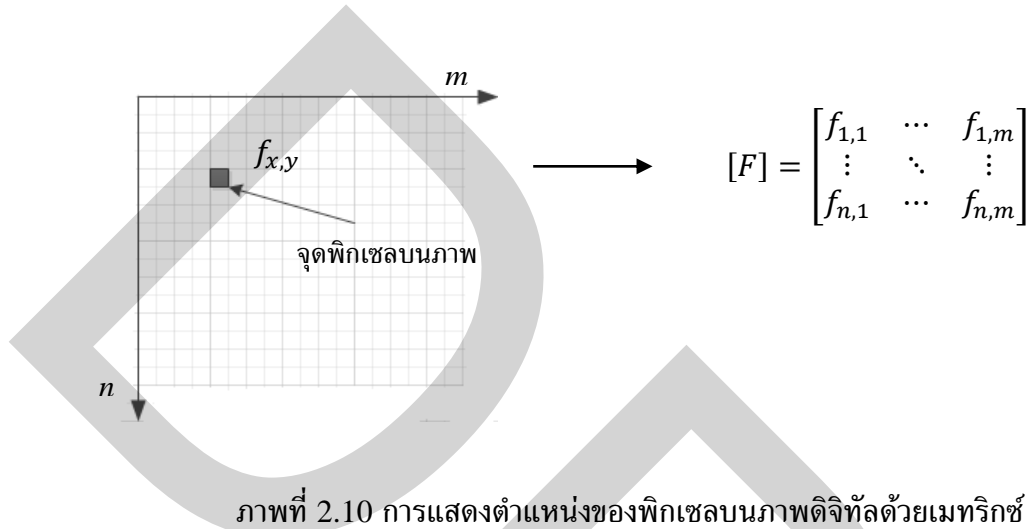
ภาพที่ 2.9 ตัวอย่างแนวคิดเครือข่ายความสัมพันธ์ระหว่างคำอธิบายความหมายบนภาพ

ออนโทโลยีเป็นแนวคิดแบบจำลองสำหรับรูปแบบคำอธิบาย ซึ่งจะมีการกำหนดประเภทวัตถุ ความสัมพันธ์ระหว่างวัตถุ รวมทั้งความสัมพันธ์ระหว่างประเภท ซึ่งจะมีการนำส่วนของเครือข่าย ความสัมพันธ์เข้ามาใช้นอนโทโลยีดังแสดงในภาพที่ 2.9 ได้แสดงเครือข่ายความหมายที่ถูกแทน ด้วยประเภทของความสัมพันธ์ โดยที่มีโหนด 2 ประเภท ที่เรียกว่า โหนดคอนเซปต์ (concept) จะ แทนกลุ่มของวัตถุ เช่น car เป็นต้น และ โหนดอินสแตน (instance) แทนวัตถุ เช่น Jimmy's car เป็นต้น จะเห็นว่า การใช้โครงสร้างความหมาย (semantic structure) จะแทนความสัมพันธ์แบบ ระดับชั้นระหว่างคอนเซปต์ แต่อย่างไรก็ตามการวิเคราะห์คำบรรยายเพื่อให้มาถึงความหมายของภาพ นั้น อาจจะไม่สามารถควบคุมการรับรู้วัตถุในบางกรณีได้เนื่องจาก ขนาดของวัตถุเล็กเกินไป และ คำศัพท์ที่แทนนั้นยากในการควบคุม ทุกปัญหาที่เกิดขึ้นจะมีความสัมพันธ์ที่เกี่ยวข้องกันกับ อินสแตน และกลุ่มของวัตถุสำหรับการใส่คำอธิบาย

มีนักวิจัยหลายกลุ่มพยายามที่จะแก้ไข และใช้หลากหลายวิธีเพื่อที่จะเชื่อมโยงให้ได้ ความหมายของภาพอย่างแท้จริง แต่อย่างไรก็ตามซอฟต์แวร์ส่วนใหญ่ นั้น จะทำการสืบค้นภาพใน รูปแบบของการเทียบคำหลักเป็นคำต่อคำตามการเรียนรู้ของเครื่องจักร (machine learning) ที่เข้ามาช่วยในการหาความหมายของภาพ หรือตามระดับชั้นของการเรียนรู้ตามที่มีการเก็บข้อมูลไว้เท่านั้น เพราะฉะนั้นคำตอบที่ได้จะไม่ได้ขึ้นกับความหมายของภาพ แต่ขึ้นกับคำศัพท์ที่มีการเก็บข้อมูลลงไป เท่านั้น เพราะฉะนั้นผลลัพธ์ของกลุ่มภาพที่ได้จะไม่ได้ขึ้นกับความหมายของภาพอย่างแท้จริง

ดังนั้นในงานวิจัยนี้ได้นำเสนอในรูปแบบของการแทนข้อมูลภาพ ด้วยความสัมพันธ์ของข้อมูล วัตถุภายในภาพ หรือเรียกว่า แนวคิดกราฟ (Conceptual Graph) ในลักษณะของกราฟที่นำเสนอ นั้น จะแสดงความสัมพันธ์ของข้อมูลภายในภาพของวัตถุที่เกิดขึ้นทั้งหมดในรูปแบบของกราฟรวมทั้ง ความสัมพันธ์ (Relationship) ระหว่างวัตถุที่เกิดขึ้น และจะมีการวัดค่าความสัมพันธ์ของความ เหมือนกันด้วยวิธีการที่เรียกว่า การจับคู่กราฟ (graph matching) ของข้อมูลภาพที่ถูกแทนค่า ความหมายไว้ ดังนั้นภาพที่มีความหมาย เหมือนกันจะมีค่าของความเหมือนกันมากที่สุด และทำ เปรียบเทียบความเหมือนกันของความหมายภาพด้วยการจำแนกภาพทั้งหมดด้วยทฤษฎีความเชื่อเบย์ (naïve-Bayes) และ วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) แบบ One-against-One และแบบ One-against-all ด้วยฟังก์ชันเคอร์เนล แบบเชิงเส้น (Linear Kernel) และ เรเดียลเบสิสฟังก์ชัน (Radial Basis Function: RBF) [R. O. Duda 1973];[T. Mitchell, 1997] ในงานวิจัยนี้ได้นำทฤษฎีมาใช้ในการพิจารณาดังนี้

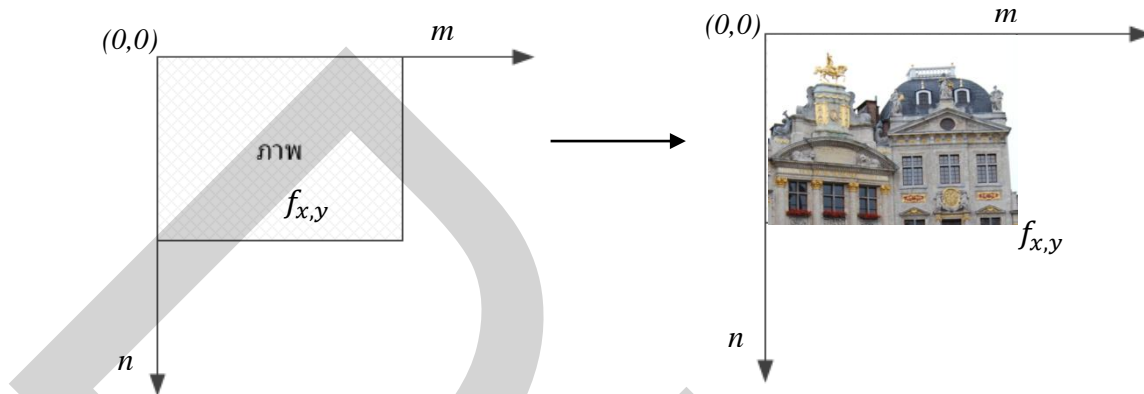
- การประมวลผลภาพดิจิทัล
- การแทนข้อมูลภาพแบบแนวคิดกราฟ
- การเปรียบเทียบความคล้ายของภาพ
- การวัดประสิทธิภาพ



2.2 การประมวลผลภาพดิจิทัล

ภาพจะถูกจัดเก็บเป็นสัญญาณดิจิทัล (digital) [นัศพ์ชาณัณ, 2553];[R.C. Gonzalez, 2002] และทำการจัดเก็บข้อมูลภาพในหน่วยความจำ ในรูปของตัวแปรแบบอะเรย์ (array) โดยค่าในแต่ละช่องของอะเรย์แสดงถึงค่าความเข้มของแสง (intensity) ของแต่ละจุดของภาพที่เรียกว่า พิกเซล (pixel) และตำแหน่งของช่องอะเรย์เป็นตัวกำหนดตำแหน่งของจุดภาพดังแสดงในภาพที่ 2.4 เมื่อนำภาพมาแสดงในรูปของ 2 มิติ ตามแนวแกน x และ y สามารถเขียนในรูปของฟังก์ชันภาพ สามารถแสดงในรูปของ $i = f(x, y)$ ในรูปแบบ 2 มิติ โดย i หมายถึง ความสว่างหรือความเข้มของแสง และ (x, y) เป็นพิกัดในแบบจำลองของภาพ (image model) ค่าของฟังก์ชัน $f(x, y)$ ดังกล่าวเป็นค่าความเข้ม ซึ่งเปลี่ยนไปตามตำแหน่งหรือพิกัดแบบจำลองภาพ ค่าความเข้มของแสงนี้เป็นค่าดิจิทัลหรือเชิงตัวเลข ดังนั้น เมื่อแสดงเป็นตัวเลขของค่าความเข้มในแต่ละพิกเซลเรียงกันทั้งภาพจะเห็นว่าเป็นเมทริกซ์ สิ่งที่แบบจำลองต่างจากแบบจำลองทางคณิตศาสตร์ คือ จุดกำเนิด (coordinate origin) ที่ตำแหน่ง $(0,0)$ ซึ่งอยู่ที่จุดซ้ายบนสุดของสเกลต่างจากจุดกำเนิดของ

แบบจำลองทางคณิตศาสตร์ที่อยู่จุดซ้ายล่างสุด ดังนั้นการนำเอาแบบจำลองทางคณิตศาสตร์มาประยุกต์ในการประมวลผลภาพ จำเป็นต้องเปลี่ยนจุดกำเนิดให้ถูกต้องด้วย ดังแสดงในภาพที่ 2.10

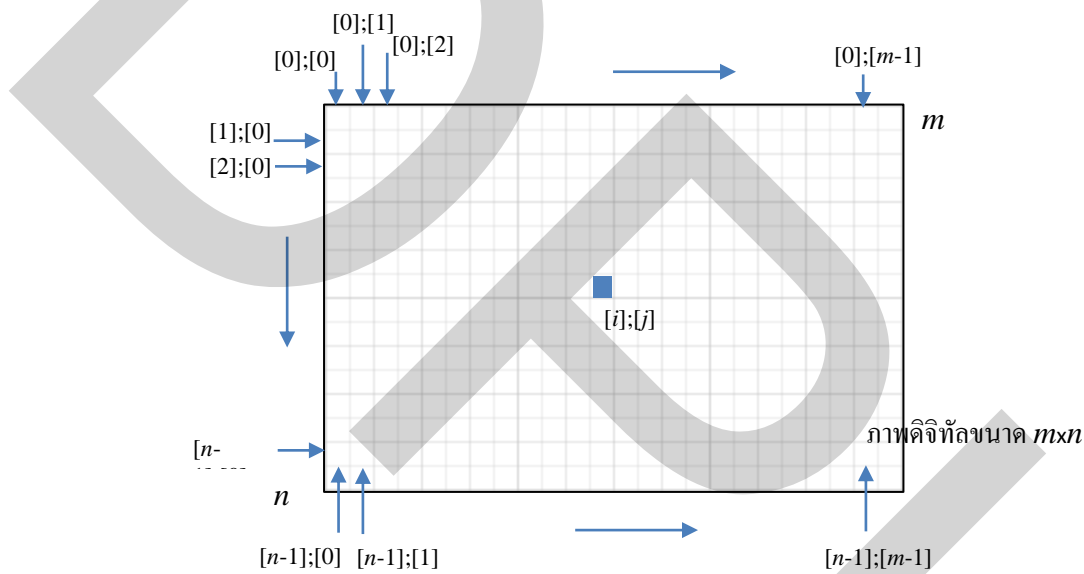


ภาพที่ 2.11 แบบจำลองของภาพด้วยระนาบ

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 2.11 สมมติให้ภาพแทนเป็นตัวแปรชื่อ x เป็นตัวแปรแบบอะเรย์ขนาด $m \times n$ (m แถว แถว และ n แถว คอลัมน์) ที่ใช้เก็บภาพขนาด $m \times n$ จุด และ ค่าของความเข้มของแสง (ค่าความสว่าง) ของจุดภาพในแถวที่ 5 คอลัมน์ที่ 4 จะตรงกับค่าของข้อมูล x เป็นรูปของ (5,4) จะเห็นว่าใช้ตำแหน่งของจุดภาพทั้งสองแกนเป็นตัวชี้ค่าข้อมูลในอะเรย์ ดังแสดงการเรียงตัวของข้อมูลบนอะเรย์ในภาพที่ 2.8 และการกำหนดความละเอียดของภาพ (image resolution) จากการกำหนดขนาดของพิกเซลตัวอย่างเช่น 1 ไมครอนต่อพิกเซล ($\mu\text{m}/\text{pix}$) 1 มิลลิเมตรต่อพิกเซล (mm/pix) เป็นต้น ในงานที่ต้องการทราบตำแหน่งหรือขนาดของวัตถุที่วัดเป็นค่าจริง เราสามารถที่จะคำนวณได้จาก

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

โดยปกติแล้วในการเก็บข้อมูลภาพโดยเครื่องมือต่างๆ 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



ภาพที่ 2.12 การจัดเก็บและเรียกใช้ข้อมูลภาพดิจิทัลลงในเมทริกซ์

จากคุณลักษณะของภาพดิจิทัลที่กล่าวมาข้างต้นนั้น เป็นการเก็บข้อมูลภาพเป็นแบบ เมทริกซ์ที่มีถึง 3 ระบาย (dimension) การประมวลผลภาพดิจิทัล (digital image processing) เป็นการเรียกใช้ขั้นตอนหรือกระบวนการที่มากกระทำบนภาพ โดยมีวัตถุประสงค์เพื่อปรับปรุงคุณภาพของภาพ ให้ได้ภาพใหม่ที่มีคุณสมบัติตามวัตถุประสงค์ที่นำไปใช้งาน เช่น การปรับให้ภาพมีความคมชัดมากขึ้น (enhancement) หรือการบีบอัดข้อมูลภาพ (compression) เพื่อประหยัดพื้นที่ในการจัดเก็บข้อมูล หรือการฝังลายน้ำ (watermark) เพื่อป้องกันการลักลอบการใช้ภาพที่ไม่ได้รับอนุญาต เป็นต้น สำหรับการประมวลผลภาพระดับสูงด้วยคอมพิวเตอร์ สามารถทำได้โดย นำภาพที่ได้มาจากกล้องหรือ image source ต่างๆ ซึ่งเป็นสัญญาณอนาล็อก แล้วนำมาแปลงเป็นสัญญาณดิจิทัลที่มีลักษณะเป็น

รหัสเชิงตัวเลขฐานสอง (binary) ประกอบด้วยตัวเลข 0 และ 1 ที่สามารถใช้รูปแบบทางคณิตศาสตร์เข้ามาช่วยในการคำนวณและการประมวลผลข้อมูลภาพ การดำเนินงานวิจัยสำหรับการจัดกลุ่มความหมายของภาพได้มีผู้วิจัยจำนวนมากที่ศึกษาและทำการทดลองโดยทั่วไป แบ่งเป็น 2 ระดับ [A. Gupta, 1997] ดังนี้

2.2.1 การประมวลภาพเบื้องต้น

สำหรับขั้นตอนการประมวลภาพเบื้องต้น (image preprocessing) ในรูปแบบของการหาบริเวณที่ต้องการในรูปแบบอัตโนมัติ ยังคงเป็นงานวิจัยที่ยังหาข้อยุติไม่ได้ โดยเฉพาะที่เป็นลักษณะของระบบเรียลไทม์ (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 และได้มีการนำเทคนิคของฮิสโตแกรม (Histogram matching) [F.Porikli, 2005]; [D.Comaniciu, 2003]; [D. Comaniciu, 2000] ที่ไม่ได้มีการนำส่วนของพื้นหลังเข้ามาเกี่ยวข้อง เพื่อทำการเปรียบเทียบแต่อย่างไรก็ตาม ด้วยคุณลักษณะของเทคนิคฮิสโตแกรมไม่สามารถที่จะรองรับตำแหน่งของพิกเซล (pixel location) ทำให้ยังคงมีปัญหาเมื่อมีวัตถุที่ทับซ้อนกัน จนไม่สามารถที่จะประมวลผลได้อย่างถูกต้อง เทคนิคสหสัมพันธ์ (Correlation) [A.J. Lipton, 1998]; [S. Wong, 2005] จะมีการทำงานที่ไม่สูญเสียในข้อมูลส่วนของสเปเชียลทำให้สามารถทดแทนวิธีการของฮิสโตแกรมได้ และได้มีการใช้เทคนิคสหสัมพันธ์เพื่อทำการหาขอบของวัตถุซึ่งวิธีการนี้ถูกเรียกว่า Edge-Enhanced Normalized Correlation (EENC) [Javed Ahmed, 2008] เพื่อทำการแก้ไขปัญหาต่างๆ ที่เกิดจากการหาวัตถุที่เกิดจากการทับซ้อนกัน หรือมีปัญหาจากสัญญาณรบกวน หรือวัตถุที่มีการเปลี่ยนแปลงทิศทางจากการหมุน เป็นต้น และการใช้เทคนิค EENC สามารถทำงานได้ดีสำหรับการแมชชีงซึ่งส่วนของพื้นที่ (matching region) พร้อมทั้งยังเป็นเทมเพลต (template) ที่ใช้ทำงานได้อย่างรวดเร็ว

2.2.2 การประมวลผลภาพระดับสูง

การประมวลผลภาพระดับสูง (high-level image processing) เป็นการใช้อัลกอริทึมหรือผลลัพธ์จากการประมวลผลข้างต้นเพื่อผ่านกระบวนการ หรืออัลกอริทึมทำให้คอมพิวเตอร์รู้จักและเข้าใจภาพ (image understanding) รวมทั้งการหาความหมายของภาพ (semantic image) ได้ เพราะฉะนั้นการประมวลผลภาพระดับสูงจำเป็นต้องใช้ข้อมูลที่ได้มาจากการประมวลผลภาพระดับต่ำ ดังนั้นจะเห็นได้ว่าการประมวลผลภาพระดับต่ำมีความสำคัญมากสำหรับการทำให้คอมพิวเตอร์รู้จักและเข้าใจภาพได้ โดยส่วนใหญ่กลุ่มนักวิจัยพยายามที่จะข้ามขั้นตอนการประมวลผลภาพระดับต่ำไป เนื่องจากยังไม่มีความสำเร็จเท่าที่ควรและพยายามที่จะเข้าใช้การประมวลผลภาพระดับสูงอย่างเดียว แต่อย่างไรก็ตามการวิจัยทั้งสองกลุ่มนี้ยังคงมีการพัฒนาการวิจัยอย่างต่อเนื่องเพื่อนำผลลัพธ์ หรือ ฟีเจอร์ต่างๆ เข้ามาทำการค้นคืนภาพ (image retrieval) หรือ การจำแนกข้อมูลภาพ (image classification) รวมทั้งการแยกแยะความหมายของภาพ (semantic image classification) ซึ่งเป็นหัวข้อหลักในการทำวิจัยครั้งนี้

2.3 การจำแนกประเภทข้อมูลภาพ

การจำแนกประเภทข้อมูลภาพ (image classification) เป็นการนำเอาข้อมูลทั้งหมดที่เก็บรวบรวมมาได้ มาผ่านกระบวนการขั้นตอนการแยกแยะข้อมูลลงในแต่ละกลุ่มที่จัดไว้ โดยในแต่ละกลุ่มของข้อมูลนั้นจะมีคุณลักษณะเด่นของแต่ละกลุ่มที่แตกต่างกัน ขึ้นกับข้อมูล หรือ ฟีเจอร์ ที่เก็บรวบรวมมาได้ รวมทั้งกระบวนการ หรือวิธีการที่ใช้แยกแยะข้อมูล ซึ่งในงานวิจัยนี้ได้ใช้ เครือข่ายแบบเบย์ เพื่อเปรียบเทียบข้อมูลวัตถุบนภาพเบื้องต้น และใช้วิธีการแยกประเภทข้อมูลภาพแบบลำดับชั้น ด้วยซอฟต์แวร์เวกเตอร์แมชชีน มีทฤษฎีดังนี้

2.3.1 ทฤษฎีความเชื่อเบย์

ทฤษฎีความเชื่อเบย์ (bayesian approach) [R. O. Duda, 1973];[T. Mitchell, 1997] เป็นการใช้อุปแบบของการประมาณค่าของความน่าจะเป็นที่จะเกิดเหตุการณ์ โดยอาศัยความสัมพันธ์ที่เกิดขึ้นของข้อมูลเพื่อมาคำนวณเหตุการณ์ที่เกิดขึ้น ดังนั้นทฤษฎีความเชื่อเบย์จึงอยู่บนพื้นฐานของการประมาณค่าความน่าจะเป็นที่จะเกิดขึ้น ในเบื้องต้นจะกล่าวถึงทฤษฎีพื้นฐานของความน่าจะเป็น (basic formulas for probability) ดังนี้

ถ้ากำหนดให้ A และ B เป็นเหตุการณ์ที่เป็นอิสระต่อกัน คือเหตุการณ์ A และ B ไม่มีความเกี่ยวข้องกัน สามารถเขียนสมการได้

$$p(A \cap B) = p(A) \cdot p(B)$$

โดยที่ $p(A \cap B)$ หมายถึงความน่าจะเป็นที่เหตุการณ์ A และ B เกิดขึ้นพร้อมกัน แต่ถ้ากำหนดให้เหตุการณ์ A และ B เป็นเหตุการณ์ที่ขึ้นต่อกัน

$$p(A \cap B) = p(A) \cdot p(B|A)$$

$$p(A \cap B) = p(A) \cdot p(A|B)$$

$$p(A) \cdot p(B|A) = p(A) \cdot p(A|B)$$

เมื่อแทนค่าความน่าจะเป็นของ A โดยที่เหตุการณ์ B ได้เกิดขึ้นแล้ว จึงสามารถเขียนสูตรของเบย์ ได้ดังนี้

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)}$$

- วิธีการเรียนรู้เบย์อย่างง่าย

วิธีการเรียนรู้เบย์อย่างง่าย (naïve-Bayes learning) [T. Mitchell, 1997] วิธีการเรียนรู้ที่ใช้หลักการของความน่าจะเป็น ซึ่งมีพื้นฐานมาจากทฤษฎีของเบย์ (bayes theorem) เข้ามาช่วยในการเรียนรู้ จุดมุ่งหมายก็เพื่อต้องการสร้างโมเดลที่อยู่ในรูปของความน่าจะเป็น ซึ่งเป็นค่าที่บันทึกได้จากการสังเกต จากนั้นนำโมเดลมาหาว่าสมมติฐานใดถูกต้องที่สุดโดยใช้ความน่าจะเป็นเข้ามาช่วยความรู้ก่อนหน้า หมายถึง ความรู้ที่เรามีเกี่ยวกับสมมติฐานแต่ละตัวก่อนที่จะเก็บข้อมูล เมื่อใช้งานเราจะนำความน่าจะเป็นของข้อมูลที่เก็บได้มาปรับสมมติฐานซ้ำอีกครั้ง เมื่อ C แทนกลุ่มข้อมูลที่จะถูกแบ่งประกอบด้วยกลุ่มที่ถูกสังเกต $c \in \{X_1, X_2, \dots, X_k\}$ แทนแอมทริบิวต์เวกเตอร์ข้อมูล ประกอบด้วยข้อมูลที่ถูกละสังเกต $x = \langle x_1, x_2, \dots, x_k \rangle$ สามารถเขียน $X = x$ โดยย่อได้ดังนี้

$$X_1 = x_1 \wedge X_2 = x_2 \wedge \dots \wedge X_k = x_k$$

สำหรับค่าผิดพลาดของการทำนายเป็น

$$\operatorname{argmax}_c (p(C = c | X = x))$$

สามารถเขียนสมการเบย์ได้ดังนี้

$$p(C = c|X = x) = \frac{p(C = c)p(X = x|C = c)}{p(X = x)},$$

$$p(C = c|X = x) \propto p(C = c)p(X = x|C = c),$$

โดยที่ค่าของ $p(C = c)$ และ $p(X = x|C = c)$ จะถูกประมาณค่าจากการเรียนรู้ซึ่งในความเป็นจริงแล้วไม่สามารถที่จะประมาณค่าของ $p(X = x|C = c)$ ได้โดยตรง ดังนั้นเมื่อค่าของ X_1, X_2, \dots, X_k เป็นเหตุการณ์ที่สามารถเกิดขึ้นต่อกัน แล้ว

$$p(X = x|C = c) = p(\bigwedge_{i=1}^k X_i = x_i|C = c)$$

$$p(X = x|C = c) = \prod_{i=1}^k p(X_i = x_i|C = c)$$

โดยทั่วไปนิยมเขียนเป็น

$$p(C = c|X = x) \propto p(C = c) \prod_{i=1}^k p(X_i = x_i|C = c)$$

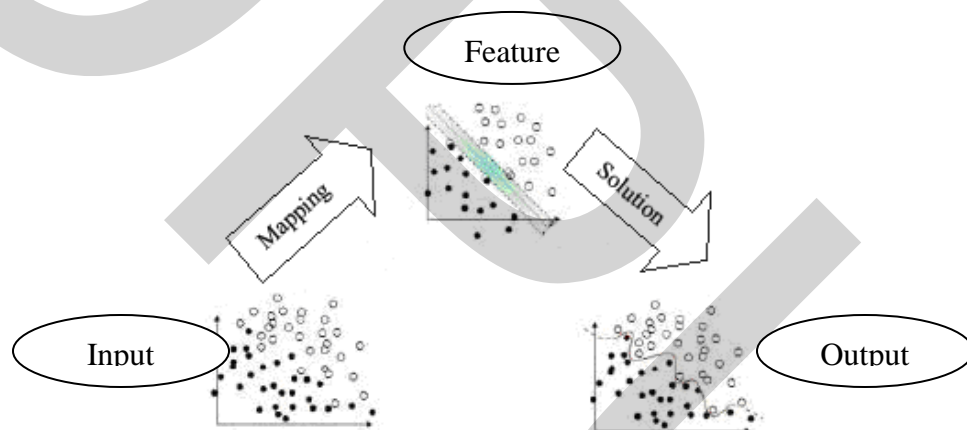
- ข้อดีของวิธีการเรียนรู้แบบเบย์

เบย์ เป็นวิธีการเรียนรู้ โดยใช้ ความรู้ก่อนหน้า (prior knowledge) เข้ามาช่วยในการเรียนรู้ได้ ซึ่งพบว่าวิธีนี้ให้ประสิทธิภาพในการเรียนรู้ได้ดีไม่ด้อยกว่าวิธีการเรียนรู้ประเภทอื่น สามารถลดข้อจำกัดอย่างง่ายในสมมติฐานของความไม่ขึ้นต่อกันระหว่างคุณสมบัติ แต่ในความเป็นจริงคุณสมบัติบางตัวจะขึ้นต่อกัน และควรนำค่าความขึ้นต่อกันนี้เข้ามาใส่ไว้ในโมเดล จึงใช้ข่างานเบย์ในการอธิบายความไม่ขึ้นต่อกันอย่างมีเงื่อนไข (condition independent) ระหว่างตัวแปร เพื่อให้กระบวนการเรียนรู้มีประสิทธิภาพ โดยสามารถใส่ความรู้ก่อนในข่างานความเชื่อเบย์ให้อยู่ในรูปแบบโครงสร้างข่างานและตารางความน่าจะเป็นแบบมีเงื่อนไข

2.3.2 ซัพพอร์ตเวกเตอร์แมชชีน

เป้าหมายของวิธีซัพพอร์ตเวกเตอร์แมชชีนจะเป็นกระบวนการสอนเครื่องแบบมีผู้สอน (supervise learning) เพื่อให้สามารถสร้างตัวจัดประเภทข้อมูล (Classifier) ที่มีความหลากหลายมาก ดังนั้น ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines : SVM) [R. O. Duda, 1973] เป็นตัวแบบที่ใช้ในการแบ่งข้อมูล สามารถทำงานได้ดีกับตัวอย่างที่ไม่รู้จัก (Unknown dataset) ด้วยกระบวนการปรับรูปแบบข้อมูลจากข้อมูลที่มีมิติต่ำ (Low dimension dataset) บนพื้นที่ข้อมูลนำเข้า

(input space) ให้อยู่ในรูปแบบของข้อมูลที่มีมิติสูง (high dimension dataset) บนพื้นที่ข้อมูลคุณลักษณะ (feature space) โดยใช้ฟังก์ชันในการปรับรูปแบบข้อมูลที่เรียกว่าฟังก์ชัน เคอร์เนล (kernel function) ซึ่งความสามารถดังกล่าวช่วยให้การสร้างตัวจัดประเภทข้อมูลด้วยสมการกำลังสอง (quadratic equation) บนพื้นที่ข้อมูลคุณลักษณะเป็นไปได้ง่ายขึ้นและ มีความชัดเจนในการจัดประเภทมากยิ่งขึ้นด้วย นอกจากนี้ ตัวจัดประเภทข้อมูลที่ตีควรมีโครงสร้างแบบเส้นตรง (linear classifier) และสามารถสร้างพื้นที่ระยะห่างระหว่างตัวจัดประเภทข้อมูลเองกับค่าที่ใกล้ที่สุดของแต่ละกลุ่มข้อมูลได้มากที่สุดเพื่อประสิทธิภาพในการแบ่งประเภทของชุดข้อมูลแต่ละประเภทออกจากกันอย่างชัดเจน ซึ่งเส้นที่เหมาะสมดังกล่าว ถูกเรียกว่า ระนาบแบ่งเขตข้อมูลที่เหมาะสม (optimal separating hyperplane) โดยหลักการในการทำงานเพื่อจำแนกประเภทข้อมูลของวิธีซัพพอร์ตเวกเตอร์แมชชีนสามารถแสดงตามภาพ 2.13



ภาพ 2.13 SVM algorithm

SVM จะทำการแบ่งชั้นของข้อมูลด้วยระนาบหลายมิติ จากข้อมูล 2 กลุ่มชุดข้อมูล โดยตัวแบบของ SVM มีเกี่ยวข้องกับโครงข่ายประสาทเทียม ตัวแบบของ SVM ใช้ sigmoid kernel function ซึ่งมีค่าเท่ากันทั้ง 2 ชั้น (layer) ตัวแบบของ SVM มีความคล้ายคลึงกับเพอร์เซพตรอน (perceptron) ซึ่งเป็นข่ายงานประสาทเทียมแบบง่ายมีหน่วยเดียวที่จำลองลักษณะของเซลล์ประสาทด้วยการใช้ kernel function ซึ่ง SVM จะใช้ข้อมูลของ คุณสมบัติและตัวแปรที่เปลี่ยนแปลงใช้ในการกำหนดระนาบหลายมิติ เรียกว่า โครงสร้าง (feature) ส่วนการเลือกที่มีความเหมาะสมที่สุดเรียกว่า โครงสร้างในการคัดเลือก (feature selection) จำนวนเขตของโครงสร้างที่ใช้อธิบายในกรณี

หนึ่ง (เช่น แกนของการค่าการคาน) เรียกว่า เวกเตอร์ (vector) ดังนั้นจุดมุ่งหมายของตัวแบบ SVM คือการประโยชน์สูงสุดจากระนาบหลายมิติที่แบ่งแยกกลุ่มของเวกเตอร์ที่ถูกเรียกว่า การหาเส้นแบ่ง hyper-planes ซึ่งใช้แบ่งข้อมูลสองคลาสเพื่อให้ได้ผลลัพธ์ที่ดีที่สุดโดยพิจารณาจากสมการเส้นตรง hyper planes และ SVMs จะทำการค้นหาเวกเตอร์ที่อยู่ใกล้เส้นแบ่ง hyper planes เรียกว่า ซัพพอร์ตเวกเตอร์ (Support vectors)

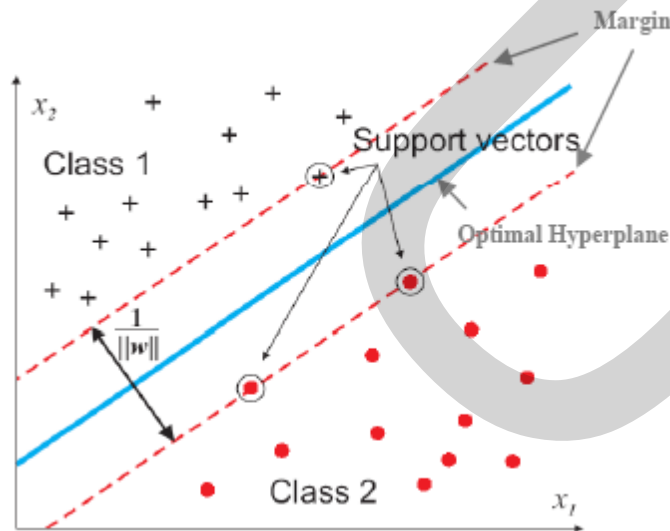
หลักการทํางานของ SVM

1. ข้อมูลที่อินพุตเข้าคำนวณหาค่า y ซึ่งค่าของ $y_i \in Y = \{-1, +1\}$ ได้จากสมการ $y = w^T \cdot x + b$ ถ้าค่าของ $w^T \cdot x + b > 0$ จะกำหนดให้ค่า $y = 1$ ซึ่งจะจัดอยู่ใน คลาส (class) ที่ 1 ถ้าค่าของ $w^T \cdot x + b < 0$ จะกำหนดให้ค่า $y = -1$ ซึ่งจะจัดอยู่ในคลาที่ 2

2. คำนวณหาเส้นตรงที่แบ่งเอกสารซึ่งเรียกว่า เส้น optimal hyperplane จากสมการ

$$w^T \cdot x + b = 0$$

3. นำค่าที่ได้จากข้อที่ 1 และ 2 ไปเขียนบนเส้นตรงตามแนวแกนตั้งและแกนนอนจะได้ดังภาพที่ 2.14 โดยระยะทาง (d) หรือ maximum margin จากเส้นขอบ ณ จุด x_i ไปยัง hyperplane สามารถแสดงได้ดังสมการ



ภาพ 2.14 การแบ่งข้อมูลโดย SVM

$$d = \frac{|w^T \cdot x + b|}{\|w\|}$$

โดยกำหนดให้

w คือ เวกเตอร์น้ำหนัก (weight vector)

x_i คือ input vector ของเอกสาร

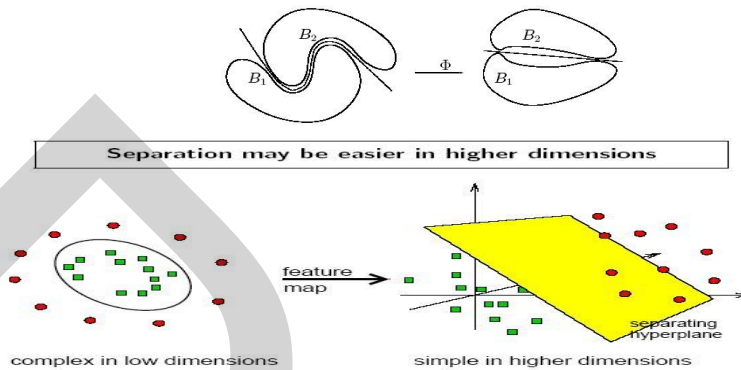
b คือ ค่าคงที่ที่กำหนดขึ้นเพื่อให้เหมาะสมกับการจัดกลุ่มเอกสาร

4. เลือกจุดที่อยู่ใกล้เส้นตรง optimal hyperplane ทั้งเหนือเส้นซึ่ง เรียกว่า “ขอบล่าง” ซึ่งเป็นขอบล่างสุดของคลาสเอกสารที่อยู่เหนือเส้นตรง optimal hyperplane และใต้เส้นเรียกว่า “ขอบบน” ซึ่งเป็นขอบบนสุดของคลาสเอกสารที่อยู่ใต้เส้นตรง optimal hyperplane เพื่อที่จะหาระยะทางระหว่างเส้นขอบทั้งสองโดยจะเลือกเอาค่าระยะทางที่ห่างจากเส้นตรง optimal hyper plane ที่น้อยที่สุดเป็นตัวเลือกในการจัดกลุ่มเอกสาร อย่างไรก็ตามโดยพื้นฐานของ SVMs นั้น จะสามารถแบ่งกลุ่มข้อมูลได้เพียง 2 กลุ่ม ดังแสดงในภาพที่ 2.14 ดังนั้นการปรับเทคนิคของการเรียนรู้ด้วย SVMs เพื่อให้ได้เป็นการจัดแบบหลายกลุ่มจึงเป็นสิ่งจำเป็น ทั้งนี้หากข้อมูลไม่สามารถแบ่งด้วย Hyperplane ออกได้เป็นสองคลาส ต้องใช้ Kernel Function ในการ Map ไปยังมิติ (Dimension) ที่สูงกว่า เพื่อให้ข้อมูลสามารถแบ่งออกจากกันได้

▪ ฟังก์ชันเคอร์เนล

ฟังก์ชันเคอร์เนล (Kernel Function) ถูกเลือกใช้ในกรณีการจัดแบ่งกลุ่มข้อมูลโดยใช้ระนาบแบบไม่เป็นเส้นตรง ซัพพอร์ตเวกเตอร์แมชชีนจะอาศัยหลักการของการแปลงข้อมูลจากพื้นที่ข้อมูลนำเข้า (Input Space) ให้เป็นพื้นที่คุณลักษณะ (Feature Space) ที่มีมิติสูงขึ้น จากภาพที่ 2.15 แสดงให้เห็นถึงแนวคิดของซัพพอร์ตเวกเตอร์แมชชีนซึ่งทำการแปลงข้อมูลที่ใช้ในการเรียนรู้แบบไม่เชิงเส้นไปเป็นขนาดพื้นที่คุณลักษณะที่ใหญ่ขึ้นผ่านฟังก์ชันเคอร์เนล (Kernel Function : Φ) และสร้างระนาบซึ่งแบ่งข้อมูลสองกลุ่มได้ดีที่สุด ทำให้เกิดเป็นขอบเขตการตัดสินใจ (Decision Surface) แบบไม่เชิงเส้นในพื้นที่ข้อมูลนำเข้า ในขณะที่ซัพพอร์ตเวกเตอร์แมชชีนแบบเชิงเส้นจะสร้างระนาบในพื้นที่คุณลักษณะที่ใหญ่ขึ้นภายใต้ทฤษฎีของ Mercer [Courant and Hilbert, 1953] ซึ่งต้องการการคำนวณที่สิ้นเปลืองในส่วนของตัวอย่างเพื่อให้ได้ขนาดพื้นที่คุณลักษณะที่ใหญ่ขึ้น ปัญหาดังกล่าว

สามารถแก้ไขได้โดยการใช้ฟังก์ชันเคอร์เนลเพื่อให้ได้ผลลัพธ์ที่น่าพึงพอใจเช่นเดียวกัน ซึ่งการใช้ฟังก์ชันเคอร์เนลจะทำให้สามารถคำนวณระนาบได้โดยไม่ต้องอาศัยการแปลงไปเป็นพื้นที่คุณลักษณะ



ภาพที่ 2.15 แนวความคิดการจำแนกประเภทข้อมูลของวิธีซัพพอร์ตเวกเตอร์แมชชีน⁵

ฟังก์ชันเคอร์เนล $K(x_i, x_j)$ เป็นฟังก์ชันที่แก้ปัญหาภายใต้เงื่อนไขของ Mercer's ซึ่งมีค่าเท่ากับการคูณกันของสองเวกเตอร์ x_i, x_j ในพื้นที่คุณลักษณะ $\phi(x_i)$ และ $\phi(x_j)$ ดังนั้น เคอร์เนลเชิงเส้น (Linear Kernel) สามารถเขียนได้ด้วยสมการดังนี้

$$K(x_i, x_j) = \phi(x_i) * \phi(x_j)$$

โดยที่ ϕ คือ ฟังก์ชันการแปลงแบบไม่เป็นเชิงเส้น (Nonlinear Projection Function) ซึ่งฟังก์ชันเคอร์เนลหลายตัวได้ถูกนำมาใช้กับซัพพอร์ตเวกเตอร์แมชชีนแบบไม่เป็นเชิงเส้น อย่างประสบความสำเร็จ การฟังก์ชันเคอร์เนลที่แตกต่างกันของซัพพอร์ตเวกเตอร์แมชชีน สามารถนำมาซึ่งวิธีการเรียนรู้ที่หลากหลาย ซึ่งตัวอย่างของฟังก์ชันเคอร์เนล มีดังนี้

- โพลีโนเมียลเคอร์เนลดีกรี d (Polynomial Kernel)

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^d$$

- เรเดียลเบสิสฟังก์ชัน (Radial Basis Function)

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / (2\rho^2))$$

- ซิกมอล (Sigmoid) ด้วย Parameter k และ θ

$$K(x_i, x_j) = \tanh(kx_i \cdot x_j + \theta)$$

⁵ ที่มา : <http://www.detrreg.com>

- แทนเจนต์เคอร์เนล (Tangent Kernel)

$$K(x_i, x_j) = \tanh(v(x_i * x_j) + c)$$

อย่างไรก็ตาม การเลือกฟังก์ชันเคอร์เนลและ ปัจจัยที่เกี่ยวข้องที่มีความเหมาะสมยังคงเป็นปัญหาที่ต้องอาศัยการทดสอบและแก้ไขสำหรับการเลือกแบบจำลอง

1. SVM แบบ One Class

เป็นเทคนิคหนึ่งของ SVM ที่มี Class Label เพียง 1 คลาส จะมีการคำนวณหาเส้นแบ่งคลาส (Decision Hyperplane : DHP) ของคลาสที่ปกติ หรือ normal ซึ่งเป็นขั้นตอนของการฝึกฝน (Training) เมื่อได้ DHP แล้ว จึงจะคำนวณหาจุดกำเนิด (origin) จากกลุ่ม normal เพื่อสำหรับให้ข้อมูลที่เหลือ ที่ต้องการตัดสินใจ มาเทียบกับจุด origin แล้วดูว่าข้อมูลนั้น อยู่ในเขตของ normal หรือไม่ ถ้าหากอยู่ในกลุ่ม normal ก็จะจัดให้ข้อมูลนั้นอยู่ในกลุ่ม ปกติ แต่หากอยู่นอกเหนือจาก normal ก็จะจัดให้อยู่ในกลุ่มที่ผิดปกติ

กำหนดให้ $v \in (0,1)$ คือ ข้อมูลใน Region และ i คือ data ใน region ถึงตัวที่ 1 ถึง l , ξ คือ ค่าความผิดพลาด (Error)

$$x_i \in \mathcal{H}^n, i = 1, \dots, l$$

ในการแยกชุดข้อมูล (data set) ออกจากจุด origin วิธีการของ One Class จะต้อง แก้ปัญหา Quadratic Programming ดังสมการต่อไปนี้

$$\min_{w, \xi, \rho} \frac{1}{2} w^T w - \rho + \frac{1}{vl} \sum_{i=1}^l \zeta_i$$

โดยที่ $w^T \phi(x_i) \geq \rho - \zeta_i$ และ $\zeta_i \geq 0, i = 1, \dots, l$

เมื่อ θ คือการ map function, w และ ρ หากได้จาก Linear Decision Function ดังสมการ

$$f(x) = \text{sign}((w^T \phi(s)) - n\rho)$$

n คือ threshold ของ decision function, α คือค่าที่ต้องปรับให้กับ w ดังนั้น จะได้ดังสมการ

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha$$

โดยที่ $0 \leq \alpha_i \leq \frac{1}{vl}, i = 1, \dots, l$ และ $e^T \cdot \alpha = 1$

2. SVM แบบ Multi Class

เป็นเทคนิคหนึ่งของ SVM ในการแบ่งคลาส ออกเป็นหลายคลาสได้ ซึ่งแต่เดิม SVM จะแบ่งได้เป็น 2 คลาสเท่านั้น มีประโยชน์สำหรับงานที่ต้องการจำแนกแบบแบ่งหลายกลุ่ม วิธีการที่นิยมใช้กันคือ One-against-All SVM และ One-against-One [C.-C. Chang and C.-J. Lin, 2001];[J.C. Platt, 1999]

– One-against-All SVM

One-against-All เป็น SVM ที่มีการจำแนกแบบหลายคลาส (Multi Class) โดยใช้วิธีการแบบ One-against-All สมมติให้มีการสร้างรูปแบบ SVM เมื่อกำหนดให้ จำนวนของคลาสเป็น k จะมีการฝึกฝนข้อมูล i คลาสด้วยข้อมูลที่เป็นทั้ง positive label และ ทุกข้อมูลที่เป็น negative label ดังนั้นจะมีข้อมูลที่ถูกฝึกฝนทั้งหมด l : $(x_1, y_1), \dots, (x_l, y_l)$ เมื่อ $x_i \in R^n, i=1, \dots, l$ และ $y_i \in \{1, \dots, k\}$ เป็นคลาสของ x_i เป็นค่า SVM ที่ i แก่สมการดังนี้

$$\min_{w^i, b^i, \xi^i} \frac{1}{2} (w^i)^T w^i + C \sum_{j=1}^l \xi_j^i (w^i)^T \phi(x_j) + b^i \geq 1 - \xi_j^i$$

โดยที่ $\text{if } y_j = i$

$$(w^i)^T \phi(x_j) + b^i \leq -1 + \xi_j^i$$

โดยที่ $\text{if } y_j \neq i$ และ $\xi_j^i \geq 0, j=1, \dots, l$

เมื่อข้อมูลฝึกฝน x_i เป็นข้อมูล map ไปยังข้อมูลที่มีพื้นที่ขนาดใหญ่กว่าด้วยฟังก์ชัน ϕ และ C การลดค่าต่ำสุดของ $(1/2)(w^i)^T w^i$ หมายถึงการทำค่าให้มากที่สุด $2/\|w^i\|$ เป็นค่าขอบระหว่างสองกลุ่มข้อมูล เมื่อข้อมูลไม่ได้มีการแยกข้อมูลออกเป็นเชิงเส้น ดังนั้น $C \sum_{j=1}^l \xi_j^i$ ทำให้สามารถลดจำนวนผิดพลาดของการฝึกฝนข้อมูลได้ แนวคิดของ SVM เป็นการค้นหาห้วงเทอมของ $(1/2)(w^i)^T w^i$ และข้อมูลฝึกฝนที่ผิดพลาด หลังจากที่มีการฝึกฝนข้อมูล ดังนั้นฟังก์ชันที่ใช้ในการตัดสินใจ k คือ

$$(w^1)^T \phi(x) + b^1$$

...

$$(w^k)^T \phi(x) + b^k.$$

ดังนั้น x เป็นคลาสที่มีคุณค่ามากที่สุดสำหรับฟังก์ชันในการตัดสินใจ

$$\text{class_of_} x \equiv \arg \max_{i=1, \dots, k} ((w^i)^T \phi(x) + b^i)$$

– One-against-One SVM

One-against-One SVM เป็น SVM ที่มีการจำแนกแบบหลายคลาสโดยใช้วิธีการแบบ One-against-One วิธีการนี้จะมีการสร้างการจำแนก $k(k-1)/2$ แต่ละข้อมูลฝึกฝนจะมาจาก 2 คลาส ดังนั้นกำหนดให้ i และ j เป็นคลาส สามารถแก้สมการดังนี้

$$\min_{w^{ij}, b^{ij}, \xi^{ij}} \frac{1}{2} (w^{ij})^T w^{ij} + C \sum_i \xi^{ij} (w^{ij})^T \phi(x_i) + b^{ij} \geq 1 - \xi^{ij}$$

โดยที่ $if y_i = i$

$$(w^{ij})^T \phi(x_i) + b^{ij} \leq -1 + \xi^{ij}$$

โดยที่ $if y_i = j, \xi^{ij} \geq 0$.

มีหลายวิธีการที่ใช้ในการทดสอบค่าจากการจำแนกทั้งหมด $k(k-1)/2$

ถ้า $sign((w^{ij})^T \phi(x_i) + b^{ij})$ เมื่อ x เป็นคลาสที่ i แล้วมีการโหวตให้คลาสที่ i หนึ่งครั้ง นอกนั้นจะเป็นการโหวตให้คลาส j ดังนั้นเมื่อ x เป็นคลาสที่มีการโหวตมากที่สุดแล้ว สามารถเรียกเหตุการณ์ดังกล่าวว่าเป็นกลยุทธ์แบบ Max Wins แต่ในกรณีที่เป็น 2 คลาสและมีการโหวตเท่ากันแล้วจะไม่ใช่วิธีการที่ดี ดังนั้นจึงควรเลือกจำนวนดัชนีที่น้อยกว่าขึ้นมา

การแก้ไขสมการ ดังกล่าวข้างต้น ถ้ามีจำนวนตัวแปรที่มีค่าเท่ากับข้อมูลใน 2 คลาส ดังนั้นควรจะมีการกำหนดคลาสข้อมูลเป็น l/k แล้วมี $k(k-2)/2$ สามารถแก้ปัญหาด้วย Quadratic programming เมื่อมีตัวแปร $2l/k$

2.4 การวัดประสิทธิภาพ

การวัดประสิทธิภาพ (evaluation) เป็นขั้นตอนสุดท้าย เพื่อทำการตรวจสอบวิธีการที่ทำการทดลองมาข้างต้นว่ามีประสิทธิภาพมากหรือน้อยเพียงใดเมื่อนำมาใช้งานจริง จะเป็นขั้นตอนที่สำคัญ เพราะการนำวิธีการที่นำเสนอไปข้างต้นมาใช้งานได้นั้นจะต้องสอดคล้องกับความต้องการ จึงต้องมีการทดสอบศักยภาพการนำไปใช้ สถาปัตยกรรมที่ใช้ตัววัดความสำเร็จหลังการนำไปใช้หากนำไปใช้แล้วไม่ประสบผลสำเร็จต้องย้อนกลับไปเริ่มกระบวนการแรกใหม่ จึงต้องมีการประเมินผลก่อนการใช้งาน ในการประเมินนั้นกระทำได้โดยการวัดประสิทธิภาพของการจัดกลุ่มภาพมักจะถูกพิจารณาเป็นค่าของความถูกต้องของแต่ละกลุ่มข้อมูลซึ่งจะประกอบด้วย การวัดค่าความแม่นยำ ค่าความระลึก ค่าความถูกต้อง และ F-measure 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: Pr) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องจากจำนวนภาพทั้งหมดที่ทำการค้นหาได้

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

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

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

- ค่าความถูกต้อง (accuracy: Acc) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องทั้งหมดจากจำนวนภาพที่มีอยู่

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

- ค่า F-measure เป็นการวัดค่าความสัมพันธ์ระหว่างค่าความระลึกและค่าความแม่นยำในเชิงฮาร์โมนิก (harmonic) เหมาะสำหรับฐานข้อมูลสารสนเทศที่มีขนาดใหญ่มาก และมักจะไม่สามารถหาข้อมูลภาพที่ถูกต้องทั้งหมดมีอยู่เท่าใด ทำให้ต้องทำการประมาณโดยใช้การสุ่มตัวอย่าง (sampling) ตามหลักทางสถิติหรือด้วยวิธีอื่นด้วย โดยทั่วไปจะเป็นการหาค่า F-measure ซึ่งแสดงสูตรได้ดังนี้

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

บทที่ 3

ขั้นตอนวิธีการดำเนินงานวิจัย

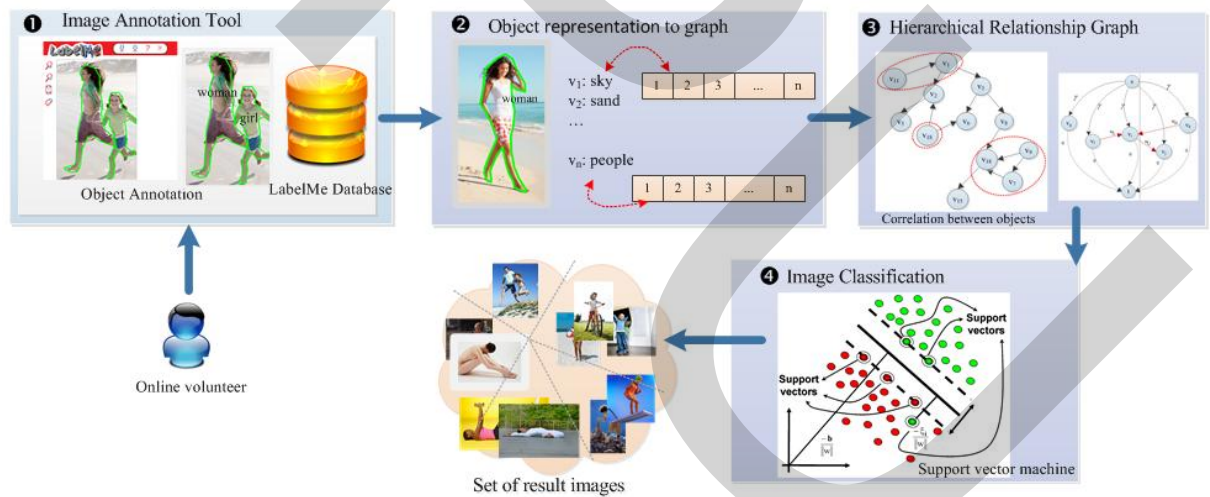
ในการศึกษางานวิจัยในครั้งนี้ได้มีการเก็บรวบรวมข้อมูลภาพและคัดเลือกภาพที่เหมาะสมเพื่อเตรียมเป็นข้อมูลภาพเบื้องต้น ดังนั้นข้อมูลภาพที่เตรียมพร้อมจะสามารถเข้าสู่กระบวนการประมวลผลภาพ โดยในงานวิจัยจะมีการแบ่งขั้นตอนวิธีการดำเนินงานวิจัยโดยทั่วไปจะแบ่งออกเป็น 3 ส่วนหลักดังนี้

3.1 ขั้นตอนการเตรียมข้อมูล (data preprocessing) เป็นการทำงานในส่วนของการนำข้อมูลเข้าด้วยเครื่องมือ (image annotation tool) และการแทนวัตถุลงในกราฟ (object representation into graph)

3.2 ขั้นตอนการประมวลผล (data processing) ทำการนำข้อมูลที่ได้ทั้งหมดมาทำงานบนกราฟแบบลำดับชั้น

3.3 ขั้นตอนการวัดประสิทธิภาพ (evaluation) เป็นการเปรียบเทียบการทำงานของวิธีการที่นำเสนอ

โดยขั้นตอนทั้งหมดจะสามารถอธิบายได้ดังภาพที่ 3.1



ภาพที่ 3.1 ขั้นตอนการจำแนกกลุ่มความหมายภาพ

3.1 ขั้นตอนการเตรียมข้อมูล

ขั้นตอนการเตรียมข้อมูล (data preprocessing) โดยการทำการแยกและคัดเลือกข้อมูลภาพดิจิทัลที่มีวัตถุบนภาพที่เด่นชัด มีวัตถุภาพพื้นหลัง และภาพที่คัดเลือกเข้ามานั้น

สามารถให้มนุษย์แปลความหมายภาพนั้นได้อย่างสมบูรณ์ สำหรับภาพบางภาพจะไม่นำเข้ามาทำการทดลองนั้นจะเป็นภาพที่มีความหมายกำกวม ภาพไม่มีความหมาย แปลความหมายไม่ได้ หรือภาพที่มนุษย์แปลได้หลายความหมาย ภาพที่มีการโฟกัสระยะใกล้ ข้อมูลภาพที่มีความซ้ำซ้อน หรือไม่สอดคล้องกันจะถูกคัดเลือกภาพนั้นออกไป และทำการรวบรวมข้อมูลภาพที่ต้องการที่มาจากหลายฐานข้อมูลจุดประสงค์ก็เพื่อให้มั่นใจว่าคุณภาพของข้อมูลที่ถูกเลือกนั้นเหมาะสม ดังนั้นกระบวนการทั้งหมดนี้จะประกอบด้วย 2 กระบวนการดังนี้

3.1.1 ฐานข้อมูลรูปภาพ

แหล่งข้อมูลภาพมีหลายแหล่งข้อมูลที่ได้รับการยอมรับและสามารถนำมาใช้เป็นฐานข้อมูลรูปภาพได้ เช่น Fotosearch stock² The Cobis Stock³ The Corel Corporation⁴ เป็นต้น เป็นแหล่งข้อมูลภาพที่หลากหลาย อาจจะมีภาพที่ไม่เหมาะสมกับการทดลองที่นำเสนอ เนื่องจาก ภาพบางภาพมีลักษณะผิดปกติ (outlier) หรือ คุณลักษณะวัตถุ (object characteristic) ไม่ชัดเจนคลุมเครือ มีขนาดวัตถุขนาดเล็กเกินไปไม่สามารถ บ่งชี้ชื่อวัตถุได้ ภาพถ่ายระยะใกล้ (close up) ภาพบางภาพอาจจะไม่สามารถแปลความหมาย หรือภาพมีความหมายกำกวมจนทำให้ไม่สามารถหาความหมายภาพได้ ทำให้ต้องมีการคัดเลือกภาพออกไป ไม่นำมาใช้ในการทดลอง ดังนั้นในการหาแหล่งข้อมูลของการนำภาพเข้ามาใช้จึงจำเป็นต้องสมบูรณ์ที่สุด ภาพจะต้องมีความเหมาะสมกับงานที่จะนำมาใช้เพื่อตอบสนองกับความต้องการของการทดลองมากที่สุด



ภาพที่ 3.2 ตัวอย่างที่ถูกแท็กด้วยคำหลัก⁵

² Fotosearch Stock: <http://www.fotosearch.com>

³ The corbis Stock: <http://pro.corbis.com>

⁴ The Corel Corporation: <http://www.corel.com/>

⁵ รูปภาพ อ้างอิง <http://www.corbisimages.com/> ค้นคืนเมื่อวันที่ 9 เมษายน 2557

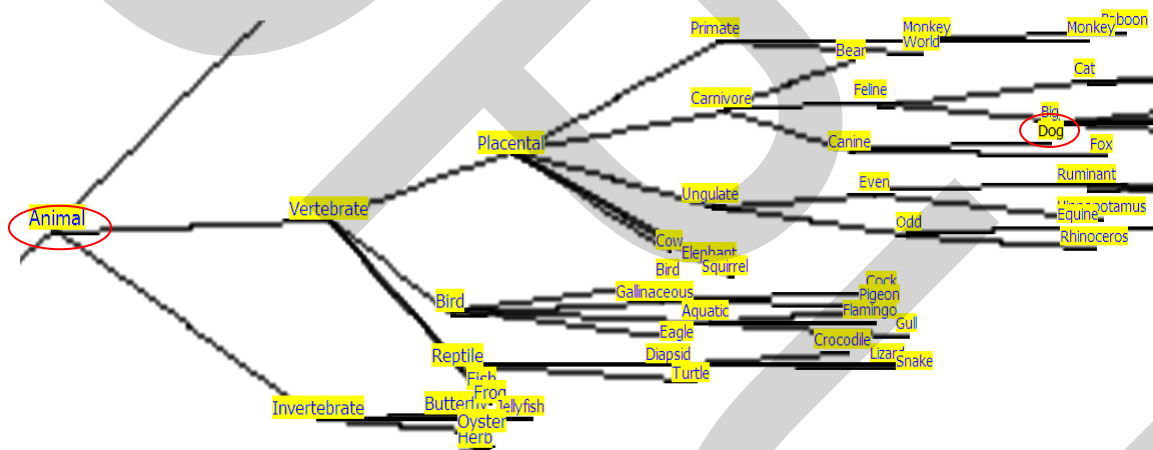
สำหรับการทดลองโดยทั่วไปในสาขาคอมพิวเตอร์วิชัน (computer vision) ในส่วนของการประมวลผลภาพระดับสูง รูปภาพที่นำเข้ามาทดลองจะเป็นภาพที่วัตถุถูกแท็ก หรือกระทำการให้ความหมายมาล่วงหน้าก่อน จะเรียกภาพจำพวกนี้ว่า annotated images [Carbonetto P., 2004];[Jia Li, 2003];[Winn J., 2005] ดังแสดงในภาพที่ 3.2 แสดงรูปภาพถูกแท็กด้วยคำหลัก ดังนั้นในการทดลองจะต้องมีฐานข้อมูลภาพที่สมบูรณ์เพียงพอที่จะสามารถนำส่วนของคำหลักที่ถูกแท็กมาใช้งานได้โดยไม่มีผลข้างเคียงต่อกระบวนการที่นำเสนอ ส่วนใหญ่แหล่งข้อมูลภาพจะทำการคัดเลือกบริเวณ (region) ที่เหมาะสมสำหรับการแท็กเป็นคำหลักหรือคำสำคัญ (keyword) เพื่อใช้สำหรับการสืบค้นข้อมูล ดังนั้นบางแหล่งข้อมูลจะทำการแบ่งหมวดหมู่วัตถุ (object categories) เข้ามาช่วยในการจัดกลุ่มของคำที่จะนำมาทดแทนส่วน บริเวณนั้นๆ [Everingham M., 2006];[Winn J., 2005] ซึ่งคำที่จะใช้จะมาจากพจนานุกรม (dictionary) จากตารางที่ 3.1 [A. Hanbury, 2007] แสดงตัวอย่างงานวิจัยที่ใช้ฐานข้อมูลคำหลัก ในการทดลองสำหรับการประมวลผลแบบการให้ความหมายภาพ (annotation)

แหล่งข้อมูลภาพ	จำนวนคำหลัก
EU LAVA Project [Perronnin F., 2006]	10
Chen and Wang [Chen Y., 2004]	20
Microsoft Research Cambridge Database [Winn J., 2005]	35
Carbonetto et al., [Carbonetto P., 2004]	55
PASCAL VOC Challenge 2005 Database [Everingham M., 2006]	101
Fei Fei et al., [L. Fei-Fei, 2004]	101
Li and Wang [Jia Li, 2003]	433
Barnard et al., [Barnard K., 2003]	323
University of Washington Ground Truth Image Database ⁶	392

ตารางที่ 3.1 ตัวอย่างงานวิจัยที่ใช้ฐานข้อมูลคำหลักในการให้ความหมายภาพ

⁶ <http://www.cs.washington.edu/research/imagedatabase/> ค้นคืนเมื่อวันที่ 9 เมษายน 2557

WordNet [Miller G.A., 1990];[Zinger S., 2005] เป็นงานวิจัยที่ได้รับความนิยมมาก มีนักวิจัยหลายกลุ่มที่อ้างอิงการใช้อ้างอิงข้อมูลของคำหลักที่แท็กบนภาพ [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 ดังแสดงโครงสร้างคำหลักของ WordNet ในภาพที่ 3.3 แสดงกิ่งก้านลำดับชั้นของคำว่า “dog” จะอยู่ในกิ่งลำดับชั้นหรือรากฐานมาจากคำหลัก “animal” และจากภาพที่ 3.4 แสดงถึงความสัมพันธ์ของคำหลักกับโครงสร้างใน WordNet ของคำว่า “dog” เพื่อให้ส่วนอื่นไปเรียกใช้ โดย WordNet จะถูกใช้คู่กับโปรแกรมเพื่อทำการสืบค้นข้อมูล และยังสามารถกำหนดความสัมพันธ์ของคำด้วยโครงสร้างของ WordNet เองและสามารถสร้างความหมายเหมือนกัน (synonymy) ได้เช่นกัน เป็นข้อดีของการนำ WordNet เข้ามาใช้จึงจะกล่าวในหัวข้อถัดไป



ภาพที่ 3.3 ส่วนย่อยของโครงสร้างลำดับชั้นของคำหลัก“dog” บน WordNet⁷

3.1.2 การแท็กคำหลักบนภาพ

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

⁷ <http://people.csail.mit.edu/torralba/research/LabelMe/wordnet/test.html> 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 และการรู้จำรูปแบบวัตถุ ดังนั้นในส่วนที่กล่าวต่อไปสำหรับการประมวลผลภาพระดับสูงคือการให้คำอธิบายภาพด้วยการแท็ก (tag) [Ismail Haritaoglu, 1998]; [Tele Tan, 2002]; [Vasileios Mezaris, 2003]; [R. Zhao, 2002] ข้อมูลภาพให้อยู่ในรูปของชื่อวัตถุ หรือคำหลักจากฐานข้อมูล

WordNet Search - 3.1
 - [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
 Display options for sense: "an example sentence"

Noun

- [S:](#) (n) [dog](#), [domestic dog](#), [Canis familiaris](#) "the dog barked all night"
- [S:](#) (n) [frump](#), [dog](#) "she got a reputation as a frump"; "she's a real dog"
- [S:](#) (n) [dog](#) "you lucky dog"
- [S:](#) (n) [cad](#), [bounder](#), [blackguard](#), [dog](#), [hound](#), [heel](#) "you dirty dog"
- [S:](#) (n) [frank](#), [frankfurter](#), [hotdog](#), [hot dog](#), [dog](#), [wiener](#), [wienerwurst](#), [weenie](#)
- [S:](#) (n) [paw](#), [detent](#), [click](#), [dog](#)
- [S:](#) (n) [andiron](#), [firedog](#), [dog](#), [dog-iron](#) "the andirons were too hot to touch"

Verb

- [S:](#) (v) [chase](#), [chase after](#), [trail](#), [tail](#), [tag](#), [give chase](#), [dog](#), [go after](#), [track](#) "The policeman chased the mugger down the alley"; "the dog chased the rabbit"

ภาพที่ 3.4 ความสัมพันธ์ของโครงสร้างคำหลักบน WordNet⁸

⁸ WordNet: <http://wordnetweb.princeton.edu/> ค้นคืนเมื่อวันที่ 9 เมษายน 2557

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];[website:flickr] รูปแบบนี้จะสามารถให้ความหมายวัตถุบนภาพได้ทางออนไลน์ผ่านทางเว็บไซต์ เป็นข้อดีสำหรับรูปแบบนี้เพราะทุกคนสามารถเข้าถึงได้อย่างง่าย ข้อมูลภาพและกลุ่มข้อมูลคำหลักจะถูกจัดเก็บลงบนเซิร์ฟเวอร์ เดียวกัน ฐานข้อมูลภาพหลากหลายและกลุ่มคนที่ให้ความหมายมีหลายกลุ่ม แต่อย่างไรก็ตามการที่มีฐานข้อมูล และกลุ่มคำศัพท์ที่เปิดกว้างมากเกินไปนี้อาจจะเป็นข้อเสียสำหรับการทดลองที่มีขีดจำกัดเช่นเดียวกัน ส่วนกลุ่มบุคคลที่เข้ามา จะไม่สามารถควบคุมได้การให้ความหมายภาพให้เป็นไปตามที่กำหนดของส่วนโปรแกรมที่ทำการทดลอง ตัวอย่างโปรแกรม เช่น Flickr [flickr] LabelMe [Russell, B.C.,2008] IBM EVA [Volkmer, T.,2005] ดังแสดงในตารางที่ 3.1
3. แบบออนไลน์เกมส์ (online game) [Von Ahn,2004];[Von Ahn L.,2006] เป็นการให้ความหมายวัตถุจะอยู่ในรูปแบบของการเล่นเกมผ่านทางออนไลน์บนเครือข่ายของอินเทอร์เน็ต คุณลักษณะพื้นฐานจะคล้ายกับแบบที่สอง แต่การให้ความหมายในลักษณะนี้อยู่บนพื้นฐานของการเล่นเกม ซึ่งต้องอาศัยความเร็ว ควบคู่กับจำนวนคำหลักของการแท็กบนวัตถุ ดังนั้นการให้ความหมายในรูปแบบค่อนข้างมีข้อจำกัด และ ขึ้นกับความสามารถของผู้

เล่นเกมส์เป็นส่วนใหญ่ ไม่ใช่ความถูกต้อง ดังนั้นอาจทำให้การแท็กวัตถุผิดพลาดหรือไม่ได้
 ใตรตรงให้ความหมายที่ถูกต้องอย่างแท้จริง แต่อย่างไรก็ตามรูปแบบออนไลน์เกมส์ยังเป็น
 เครื่องมือที่นิยมกันในปัจจุบัน ยกตัวอย่างเช่น ESP game [Von Ahn, 2004] Peekaboom
 [Von Ahn, 2006] Squigl [Squigl] ยังมีหลายหน่วยงานที่สร้างโปรแกรมเพื่อมารองรับการ
 ทำงานในรูปแบบทั้ง 3 ดังแสดงในตารางที่ 3.2

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				✓

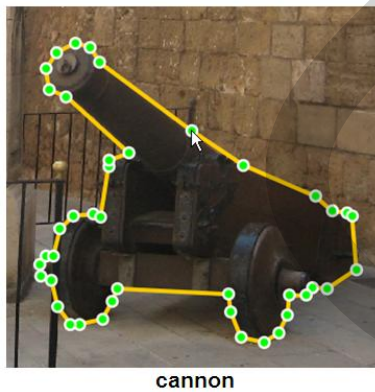
ตารางที่ 3.2 แหล่งฐานข้อมูลของคำหลัก

จากเครื่องมือให้ความหมายภาพ (image annotation tool) ในตารางที่ 3.2 ที่แบ่งเป็น 3
 รูปแบบ แต่ละรูปแบบมีความสามารถทำงานได้แตกต่างกันมีทั้งข้อดีและข้อจำกัด ดังนั้นใน
 งานวิจัยนี้ได้เลือกเครื่องมือแบบออนไลน์ โดยใช้โปรแกรม LabelMe⁹ [B. C.
 Russell,2008];[A. Torralba,2010] เป็นเครื่องมือที่ได้รับการยอมรับอย่างกว้างขวาง สำหรับ
 งานวิจัยทางด้าน Computer Vision โดยแอปพลิเคชันนี้สามารถทำงานได้อย่างเต็มรูปแบบบน
 เว็บในลักษณะของเครื่องมือให้ความหมาย (Web-based annotation tools) เริ่มตั้งแต่ปี 2005
 ปัจจุบันมีวัตถุรูปภาพที่ถูกให้ความหมายรวมทั้งสิ้น 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] ผู้ใช้สามารถเข้าถึงโปรแกรมผ่านทางเครือข่ายออนไลน์ได้
 สามารถแท็กวัตถุผ่านทางออนไลน์ได้ดังแสดงในภาพที่ 3.5 แสดงหน้าเว็บไซต์ของโปรแกรม

LabelMe โปรแกรมสามารถทำงานร่วมกันได้หลายแพลตฟอร์ม ซึ่งเป็นวัตถุประสงค์หลักของ
 คณะผู้จัดทำโปรแกรมนี้ ทำให้ผู้ใช้งานที่เข้ามาให้ความหมายภาพมาได้มากมาย และมีพื้นฐานของ
 การให้ความหมายที่แตกต่างกันตามความสามารถ ของแต่ละบุคคล

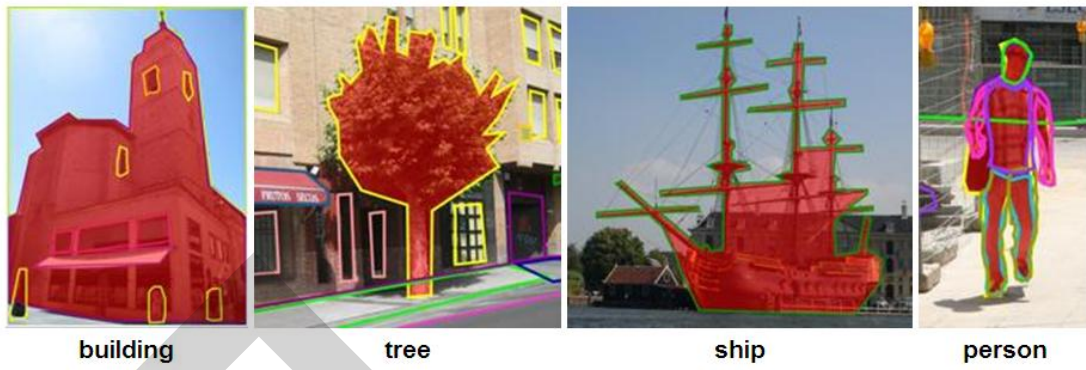


ภาพที่ 3.5 โปรแกรม LabelMe บนเบราว์เซอร์⁹



ภาพที่ 3.6 การเลือกสัดส่วนของวัตถุบนภาพ⁹

⁹ LabelMe: <http://labelme.csail.mit.edu/> ค้นคืนเมื่อวันที่ 9 เมษายน 2557



ภาพที่ 3.7 ตัวอย่างวัตถุที่ถูกแท็กด้วยโปรแกรม LabelMe¹⁰



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 3.8 ตัวอย่างภาพที่ถูกแท็กคำหลักบนภาพ ด้วยโปรแกรม LabelMe¹⁰

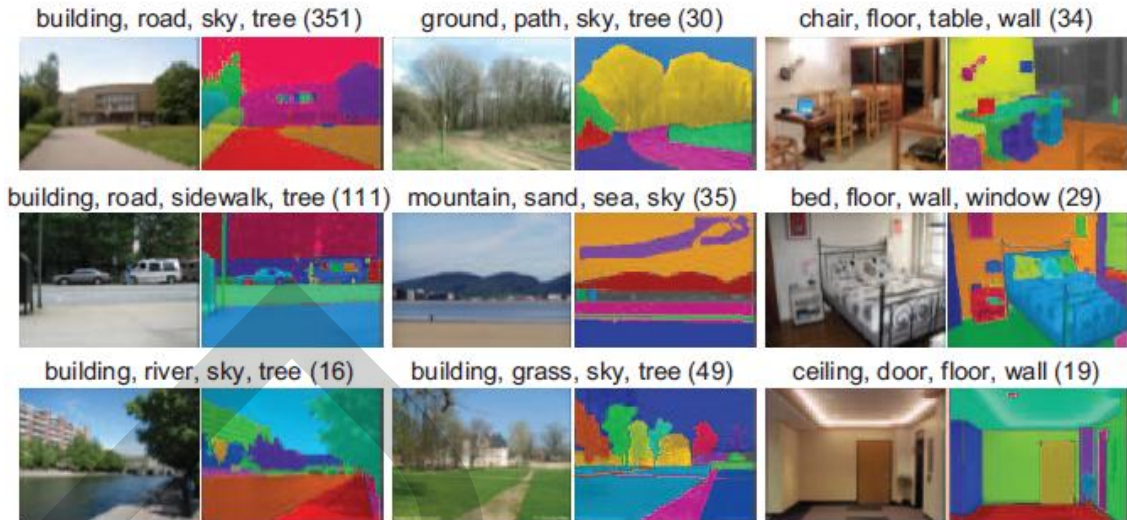
¹⁰ LabelMe: <http://labelme.csail.mit.edu/> ค้นคืนเมื่อวันที่ 9 เมษายน 2557

รูปแบบของการให้ความหมายภาพ หรือแท็กข้อมูลบนโปรแกรม LabelMe จากภาพที่ 3.5 ประกอบด้วยเครื่องมือช่วยการแท็ก (ด้านบน) รูปภาพสามารถคัดเลือกจากฐานข้อมูลภายในโปรแกรม หรือ โหลดรูปภาพที่ต้องการจากเครื่องคอมพิวเตอร์เข้ามาในโปรแกรม LabelMe ได้ วิธีการแท็กวัตถุ แสดงในภาพที่ 3.6 เพื่อทำการแท็กคำหลักบนวัตถุ ด้วยการใช้เมาส์ลากตามสัดส่วนของวัตถุแบบ freehand บนรูปภาพ เมื่อลากตามสัดส่วนของวัตถุเสร็จสิ้นโปรแกรมจะให้ใส่คำหลักเพื่อให้ความหมายของวัตถุ ข้อมูลคำหลักจะถูกจัดเก็บลงบนฐานข้อมูลพร้อมกับรูปภาพ และ ในภาพที่ 3.7 แสดงตัวอย่างของภาพที่ถูกแท็ก จากภาพที่ 3.5 ข้อมูลคำหลักที่ถูกแท็กแล้วจะแสดงไว้ทางขวามือ จะได้ข้อมูลคำหลักของวัตถุบนภาพประกอบด้วย grass, snorkel, snorkel, kid, kid, ball, ball, flipper และ flipper

หลังจากที่ผู้ใช้งานแท็กภาพในโปรแกรม LabelMe หรือทำการค้นหารูปภาพที่ถูกแท็ก โปรแกรมจะมีการแสดงภาพที่ถูกแท็ก ดังแสดงในภาพที่ 3.8 แสดงภาพที่ถูกแท็ก เส้นแสดงขอบเขตของคำหลักแต่ละคำที่แท็กไว้ ดังนั้นการใช้งานข้อมูลในโปรแกรม สามารถดาวน์โหลดมาใช้ร่วมกันได้ และข้อมูลคำหลักที่ถูกนำมาใช้มีมากมายหลายฐานดังแสดงในตารางที่ 3.1 ดังนั้นข้อมูลสำหรับการทดลอง จะมีการใช้ฐานข้อมูลที่เก็บรวบรวมมาจาก LabelMe [B. C. Russell, 2008] โดยจะทำการให้ความหมายในลักษณะของรูปทรงแบบ polygon ที่วาดลงบนวัตถุในภาพ ดังแสดงตัวอย่างในภาพที่ 3.9 [A. Torralba, 2010] จากโปรแกรม LabelMe เป็นตัวอย่างคำหลักที่พบบ่อยหรือเกิดขึ้นบ่อยในการให้ความหมายของวัตถุบนภาพ ตัวเลขด้านข้างแทนจำนวนของการให้ความหมาย สำหรับในภาพที่ 3.10 [A. Torralba, 2010] แสดงตัวอย่างของข้อมูลคำหลักที่ถูกแท็กบนภาพในโปรแกรม LabelMe เช่นกัน



ภาพที่ 3.9 ตัวอย่างของคำหลักที่พบบ่อยในการให้ความหมายวัตถุ



ภาพที่ 3.10 ตัวอย่างของคำหลักในการให้ความหมายวัตถุบนภาพ

3.1.2 การแทนวัตถุลงในกราฟ

หลังจากผ่านขั้นตอนของการเตรียมข้อมูลภาพ ฐานข้อมูลคำหลักและฐานข้อมูลภาพได้ถูกคัดเลือกเข้ามา จะทำการจัดเก็บข้อมูลคำหลักที่ได้ลงในกราฟ (object representation into graph) ดังนั้นข้อมูลทั้งหมดที่ได้จะมีการนำภาพและข้อมูลวัตถุเป็นพีเจอร์ลงในตัวแปรที่เป็นเมตริกซ์ เพื่อทำการประมวลผลตามทฤษฎีที่น่าเสนอ กำหนดให้ $X = [X_1^T, X_2^T, \dots, X_N^T] \in R^{N \times J_r}$ มีขนาดเมตริกซ์เป็น $N \times J_r$ เมื่อ N แทนจำนวนภาพ r แทนขนาดมิติของเมตริกซ์ และ J เป็นจำนวนของวัตถุบนภาพ กำหนดให้ $Y = [y_1, y_2, \dots, y_N]$ เป็นเวกเตอร์ของวัตถุที่เก็บค่าการจำแนกประเภทของ N ภาพ

ข้อมูลวัตถุภายในภาพที่ถูกจัดเก็บลงในเมตริกซ์จะถูกสร้างความสัมพันธ์ระหว่างวัตถุด้วยแนวคิดกราฟ (conceptual graph) กำหนดให้รูปภาพใดใด สามารถเขียนความสัมพันธ์ของแนวคิดกราฟ เป็นสเปเชียลเอ็นทิตี (spatial entities) ที่แทนด้วยเซตของ V คือจุด (vertex) หรือโหนด (node) เมื่อ $i \in \{1..N\}$ และความสัมพันธ์ของวัตถุภายในภาพใด เกิดจากโหนดสองโหนดที่เชื่อมต่อกันด้วย E เมื่อ $E \subset V \times V$, เมื่อ E คือความสัมพันธ์ระหว่างโหนดสองโหนด (edge) ความสัมพันธ์ในรูปแบบนี้ถูกเรียกว่า binary spatial relationship ดังนั้นสามารถแทนความสัมพันธ์ของสองจุดได้ด้วย $e_{ab} \equiv (v_a, v_b) \in E$ เมื่อกำหนดให้ $v_a, v_b \in V$ จากตัวอย่างภาพที่ 3.11 ก. แสดงภาพตัวอย่างจากโปรแกรม LabelMe ถูกให้ความหมายวัตถุตามคำหลัก ดังนี้ kid kid grass ball ball snorkel snorkel flipper ดังแสดงในภาพที่ 3.11 ข. และสามารถแทนด้วยความสัมพันธ์ของวัตถุ ด้วยแนวคิดกราฟโหนดแต่ละโหนดบนกราฟถูกแทนด้วยคำหลักดังนั้น

grass เป็น root node ที่มีความสัมพันธ์ร่วมกันกับ kid ทั้งสองและ โหนดของ kid จะเกิดความสัมพันธ์ต่อไปยังส่วนต่างๆ เช่น ball snorkel และ flipper ดังนั้นการแสดงถึงความสัมพันธ์จะถูกเชื่อมเข้าด้วยกันเป็นค่าของ E

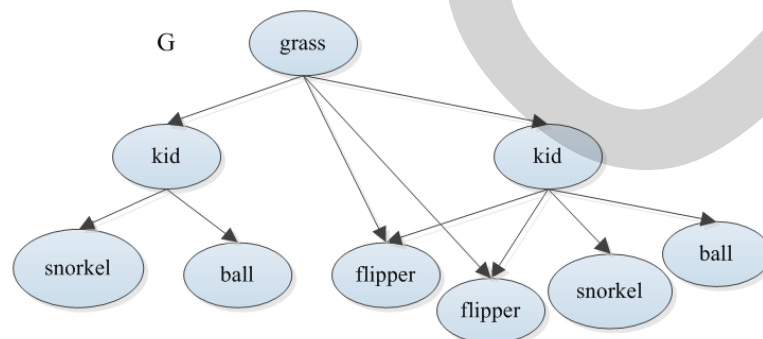


วัตถุ: kid kid grass ball ball snorkel snorkel flipper

ก. วัตถุที่ถูกแท็กบนภาพจากโปรแกรม LabelMe



ข. วัตถุที่ถูกแท็กบนภาพจากโปรแกรม LabelMe



ค. แทนความสัมพันธ์ของวัตถุที่ถูกแท็กด้วยแนวคิดกราฟ
ภาพที่ 3.11 ตัวอย่างความสัมพันธ์ของวัตถุด้วยแนวคิดกราฟ

จากภาพที่ 3.11 ค. แสดงความสัมพันธ์ ที่เกิดขึ้นดั่งนั้นตามแนวคิดกราฟคือโหนดทั้งหมด ประกอบด้วยคำหลักดังนี้ grass kid kid ball ball snorkel snorkel flipper และจุดเชื่อมต่อระหว่าง grass กับ kid ถูกแทนด้วย $e_{grasskid} \equiv (v_{grass}, v_{kid}) \in E$ ดั่งนั้นสามารถเขียนรูปแบบใหม่ เพื่อ่ายในการอ่าน สามารถเขียนเป็นสมการได้ดั่งนี้ $G \equiv (V, E)$ กำหนดให้ โหนดต่าง ๆ บนภาพ ด้วย $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8\}$, และแทนความสัมพันธ์ระหว่างโหนดต่าง ๆ ด้วยสมการ เมื่อ กำหนดให้ $e_{1,2} \equiv (v_1, v_2)$

3.2 ขั้นตอนการประมวลผล

สำหรับขั้นตอนการประมวลผลจะนำข้อมูลที่มีการจัดเก็บไว้แล้วมาประมวลผล ด้วยวิธีการที่นำเสนอคือการใช้แนวคิดกราฟ และความสัมพันธ์ภายในที่เป็นแบบลำดับชั้น รวมทั้งความสัมพันธ์ที่เกิดขึ้นระหว่างวัตถุภายในด้วยกันเองโดยจะแบ่งการประมวลผลเพื่อทดสอบค่าความถูกต้องและผลลัพธ์ที่ได้ออกเป็น 2 ส่วนดั่งนี้

3.2.1 การกำหนดค่าน้ำหนักวัตถุ

การกำหนดค่าน้ำหนักสำหรับข้อมูลวัตถุนบนภาพ โดยคิดค่าน้ำหนักจากความสัมพันธ์ข้อมูลวัตถุที่เกิดขึ้นทั้งหมดภายในภาพ X_N^T โดยจะเก็บเป็นค่า $\mathcal{G}(v_i)$ (Information Content) [Resnik et al., 1995] แทนรายละเอียดข้อมูลดั่งนั้นสามารถนับจำนวนคำหลักที่เกิดขึ้นเป็นค่าของความน่าจะเป็นของแต่ละคำหลักบนภาพได้ด้วยสมการดั่งนี้

$$P(v_i) = \frac{freq(v_i)}{\sum_i freq(v_i)}, i \in \{1, \dots, N\}$$

เมื่อกำหนดให้ $P(v_i)$ แทนความน่าจะเป็นของวัตถุ v_i และ $freq(v_i)$ แทนความถี่ของวัตถุ v_i ที่เกิดขึ้น

$$freq(v_i) = \sum_{n \in word(v)} count(n).$$

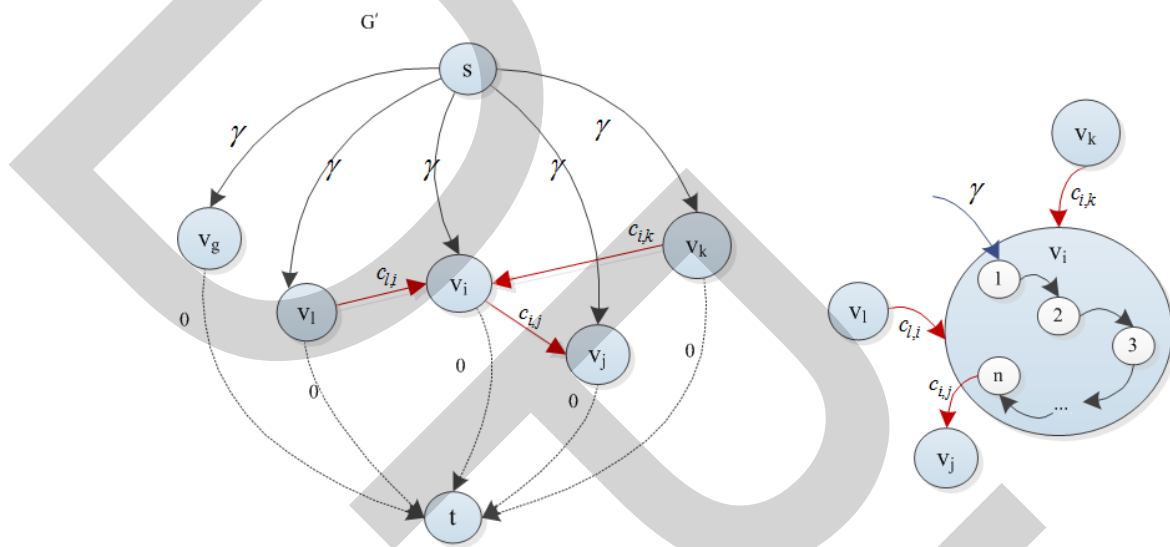
ดั่งนั้นเมื่อมีการคำนวณในแต่ละโหนดบนภาพสามารถเขียนเป็นสมการของ $\mathcal{G}(v_i)$ ใน v_i ได้ดั่งนี้

$$\mathcal{G}(v_i) = \log^{-1} P(v_i)$$

3.2.2 การหาความสัมพันธ์ของสายเชื่อมโยงข้อมูล

ภายในภาพหนึ่ง ๆ มีวัตถุที่ถูกแทนเป็นโหนดได้หลายโหนดแต่ละโหนดถูกเชื่อมโยงด้วย edge ระหว่างกันเป็นทอด ๆ สามารถแสดงลำดับความสัมพันธ์ระหว่างโหนดได้ดั่งแสดงในภาพที่ 3.12 เมื่อกำหนดให้ s เป็นโหนดที่มีความสัมพันธ์กับโหนดที่อยู่ด้านบนหรือโหนดพ่อแม่ (parent

node) มีโหนดที่อยู่ด้านล่างหรือโหนดลูก (children node) คือ v_i, v_j, v_k, v_l และ v_g ดังนั้นจะเห็นว่าความสัมพันธ์ที่เกิดขึ้นของ v_i นั้นเกิดขึ้นร่วมกันหลายโหนด ทำให้หาความสัมพันธ์ที่เกิดขึ้นของโหนดสายการเชื่อมโยง (link) โดยจะใช้การคำนวณในลักษณะของความน่าจะเป็นจากความสัมพันธ์ที่เกิดขึ้นเป็นค่าของ link strength (LS) ดังนั้นข้อมูลรูปภาพที่ได้มาจะถูกแทนที่ด้วยแนวคิดกราฟแบบลำดับชั้น (hierarchical graph) กระบวนการประมวลผลจะเป็นการเลือกจากจำนวนโหนดที่การติดต่อกันในกราฟย่อยที่มีจำนวนน้อยที่สุดจากกราฟ โดยเลือกจากโหนด s ที่เชื่อมโยงไปโหนด t ดังนั้น



ก. ความสัมพันธ์ของค่าน้ำหนักระหว่างโหนด s และ t ข. ความสัมพันธ์ภายในกราฟย่อย
 ภาพที่ 3.12 ตัวอย่างความสัมพันธ์ของค่าน้ำหนักภายในกราฟ

กำหนดให้ $G'=(V',E')$ เป็นกราฟใหม่ที่มีโหนด s เชื่อมโยงไปโหนด t ดังนั้น $V'=V\cup\{s,t\}$, และ $E'=E\cup\{(s,v):v\in V\}\cup\{(u,t):u\in V\}$ โดยที่ s และ t เป็นส่วนเชื่อมโยงไปยังวัตถุ ค่าน้ำหนักระหว่างโหนด s และโหนดอื่นๆ สามารถแทนค่าได้ด้วย γ เป็นพีเจอร์ที่ใช้เก็บเส้นทางดังแสดงในภาพที่ 3.12 การแก้สมการจะแก้ได้ด้วยการหาค่าต่ำสุดของความสัมพันธ์ภายในกราฟ G' จากความสัมพันธ์ ระหว่างโหนด s และ t สามารถเขียนสมการได้ดังนี้

$$\Omega(\beta) = \min_{f \in F} \left\{ \sum_{(u,v) \in E'} f_{uv} c_{uv} s t s_j(f) \geq |\beta_j|, \forall j \in \{1, \dots, J_r\} \right\},$$

$$s_j(f) = \sum_{u \in V': (u,j) \in E'} f_{uj}$$

สำหรับค่าน้ำหนักที่เกิดขึ้นของเวอร์เท็กซ์ j ใน $V = 1, 2, \dots, J_r$ ทุกเวอร์เท็กซ์จะมีการแทนค่าด้วย capacity/ค่าน้ำหนักสำหรับสองเวอร์เท็กซ์ เมื่อกำหนดให้ $[c_{u,v}]$ เป็นค่าน้ำหนักของเส้นเชื่อมระหว่าง u และ v เมื่อ $[c_{u,v}]_{(u,v) \in E'}$ และ F เป็นเซตของเส้นทางการไหลของการเชื่อมโยงบนกราฟ G' วัตถุที่ถูกเลือกจะสามารถเรียกดูค่าได้จากการไหลของ f_{uv} จากทุกๆ เส้นของ (u,v) บนกราฟ

$$f^* \in \arg \min_{f \in F} \left\{ \sum_{(u,v) \in E'} f_{uv} c_{uv} + \sum_{j=1}^{J_r} \frac{1}{2} \max(|u_j| - s_j(f), 0)^2 \right\},$$

เมื่อแทน u เป็นค่าใน \mathbb{R} และ F เป็นการไหลข้อมูลบนกราฟ G'

3.3 ขั้นตอนการวัดประสิทธิภาพ

ขั้นตอนการวัดประสิทธิภาพ (evaluation) เป็นขั้นตอนสุดท้ายของกระบวนการจำแนกความหมายภาพ กำหนดการจำแนกประเภทความหมายภาพด้วย $h_\beta(X) = \arg \max_y X\beta$, เมื่อ $\beta = (\beta_1, \beta_2, \dots, \beta_{J_r}) \in \mathbb{R}^{J_r}$ เป็นเวกเตอร์สำหรับการประมวลผลโดยที่ \mathbb{R}^{J_r} แทนค่าที่ได้จากการจำแนกประเภท นำผลการจำแนกกลุ่มของภาพที่จัดได้มาทำการวัดประสิทธิภาพของสิ่งที่น่าสนใจข้างต้น โดยจะตรวจสอบกลุ่มของภาพที่จัดได้ว่ามีค่าเป็นอย่างไร เมื่อเทียบกับกลุ่มของภาพที่ถูกต้อง ซึ่งต้องมีการวัดค่าความระลึก (recall) และค่าความแม่นยำ (precision) จะเป็นค่าที่แสดงว่า การค้นคืนข้อมูลได้ตรงกับความต้องการเพียงใด ส่วนค่าความระลึกจะเป็นค่าที่แสดงถึงความครอบคลุมในการจัดกลุ่มภาพ หลังจากนั้นจะนำค่ามาคำนวณในรูปของค่าความถูกต้อง (accuracy) และ F-measure ต่อไปและนำค่าทั้งหมดมาแปลผลและประเมินผลลัพธ์ที่ได้ว่ามีความเหมาะสม หรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ในรูปที่สามารถเข้าใจได้ง่าย สามารถอ่านเพิ่มเติมได้ในบทที่ 2

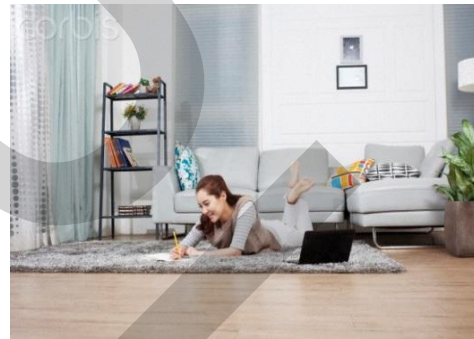
บทที่ 4

ผลการทดลอง

ในการทำวิจัยครั้งนี้เพื่อหาความหมายของภาพที่เกิดจากการรวมของทุกวัตถุที่ปรากฏบนภาพ จึงนำเสนอการปรับปรุงการจำแนกข้อมูลภาพด้วยการแทนความสัมพันธ์ของวัตถุด้วยแนวคิดกราฟแบบลำดับชั้น จะแทนข้อมูลภาพในลักษณะของความสัมพันธ์ ระหว่างวัตถุที่เกิดขึ้นภายในภาพทั้งหมดแบบลำดับชั้น เพื่อให้ได้ผลลัพธ์ของความหมายภาพอย่างแท้จริง โดยทำการเปรียบเทียบความเหมือนกันของความหมายภาพด้วยการหาความเหมือนด้วยทฤษฎีความเชื่อเบย์ (naïve-Bayes) และวิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) แบบ One-against-One และแบบ One-against-all ด้วยฟังก์ชันเคอร์เนล แบบเชิงเส้น (Linear Kernel) และเรเดียลเบสิสฟังก์ชัน (Radial Basis Function: RBF)



ก.ภาพสำนักงาน



ข.ภาพห้องนั่งเล่น



ค.ภาพห้องครัว

ภาพที่ 4.1 ตัวอย่างกลุ่มภาพภายใน

4.1 การกำหนดข้อมูลภาพ

ข้อมูลภาพสำหรับการทดลองได้ใช้ฐานข้อมูลที่เก็บรวบรวมและมาจากแอปพลิเคชัน LabelMe [B. C. Russell,2008];[A. Torralba,2010] โดยแอปพลิเคชันนี้สามารถทำงานได้อย่างเต็มรูปแบบบนเว็บในลักษณะของเครื่องมือให้ความหมาย และได้มีการคัดเลือกภาพสำหรับการทดลองให้อยู่ในหมวดหมู่ของภาพภายใน (indoor) และภาพภายนอก (outdoor) โดยได้กำหนดตามความหมายพื้นฐานจากการสุ่มและตัดสินใจจากการจำแนกความหมายภาพด้วยคนเป็นหลัก (human scenes classification) [Jianxiong Xiao, 2010] ดังนั้นได้มีการแบ่งกลุ่มภาพย่อยในวงจำกัดลงเพียง 7 กลุ่มเพื่อทดสอบการจำแนกความหมายของข้อมูลภาพ ประกอบด้วยกลุ่มของ ภาพสำนักงาน (office), ภาพห้องนั่งเล่น (living room), ภาพเมือง (city), ภาพห้องครัว (kitchen), ภาพชายฝั่ง (coast), ภาพภูเขา (mountain), ภาพป่า (forest) ดังแสดงภาพตัวอย่างในภาพที่ 4.1 และ ภาพที่ 4.2



ก.ภาพชายฝั่ง



ข.ภาพป่า



ค.ภาพภูเขา



ง.ภาพเมือง

ภาพที่ 4.2 ตัวอย่างกลุ่มภาพภายนอก

การกำหนดภาพที่ใช้ในการทดลอง จะใช้คำหลักที่ถูกแท็กไว้บนภาพจากผู้ใช้ใน LabelMe โดยกำหนดให้จำกัดขอบเขตของคำหลักที่ใช้ในการทดลองนี้รวมทั้งหมด 95 คำที่แตกต่าง ภาพรวมทั้งหมดที่ใช้ในการทดลอง 1,500 ภาพ ที่ถูกกำหนดมาจาก WordNet [Miller G.A., 1990];[Zinger S., 2005] และได้พยายามหลีกเลี่ยงคำหลักที่มีความหมายกำกวมและคำหลักที่เป็นทั้งคำเหมือน (Synonym) เช่น “abbey”, “church”, “cathedral” เป็นต้น เพราะอาจเป็นส่วนหนึ่ง ที่อาจทำให้ผลการทดลองนั้นเกิดความผิดพลาดได้ การทดลองจำแนกด้วย เครือข่ายแบบเบย์ (Bayesian Network) และ วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) ดังนั้นในการทดลองจะมีการแสดงผลดังนี้

4.2 ผลการจำแนกความหมายของข้อมูลภาพ

4.2.1 การจำแนกความหมายภาพด้วยค่าน้ำหนักวัตถุ

การทดลองเบื้องต้นเป็นข้อมูลพื้นฐานเพื่อทำให้ทราบว่า ข้อมูลของคำหลักในกลุ่มใดมีผลต่อการจำแนกภาพบ้าง และจำนวนของวัตถุบนภาพจะมีความสัมพันธ์ต่อกลุ่มภาพหรือไม่ ดังนั้นจึงได้ทำการทดลองเพื่อทำการจำแนกกลุ่มภาพด้วยพีเจอร์ที่เกิดจากค่าน้ำหนักวัตถุเด่นบนภาพ $\vartheta(v_i)$ จากคำหลัก โดยแบ่งการเก็บพีเจอร์ออกเป็น 3 เซตแต่ละเซตจะถูกจัดเก็บตามความสำคัญเรียงตามขนาดของวัตถุที่โดดเด่นบนภาพ (dominant objects) ดังนั้นจึงแบ่งการทดลองค่าน้ำหนักวัตถุเด่น $\vartheta(v_i)$ ออกเป็น 3 พีเจอร์ 5 พีเจอร์ และ 7 พีเจอร์ เพื่อจำแนกกลุ่มภาพเดียวกันด้วยวิธี naïve-Bayes และ SVM แสดงผลลัพธ์ในตารางที่ 4.1 ถึง 4.3

Scene/Categories		Performance (%)					
		naïve_Bayes			SVM		
		Precision	Recall	F ₁	Precision	Recall	F ₁
Indoor	Office	40.6	41.8	41.2	42.6	46.2	44.3
	Living room	47.1	44.0	45.5	46.1	43.5	44.8
	Kitchen	43.0	45.7	44.3	44.0	46.3	45.1
Outdoor	Coast	49.5	42.5	45.7	50.5	45.6	47.9
	Forest	47.5	50.5	49.0	49.5	51.6	50.5
	Mountain	42.6	46.7	44.6	43.6	44.9	44.2
	City	47.5	47.5	47.5	56.6	54.9	55.7
Accuracy		45.39			47.52		

ตารางที่ 4.1 ผลลัพธ์การจำแนกความหมายภาพด้วย $\vartheta(v_i)$ 3 พีเจอร์

Scene/Categories		Performance (%)					
		naïve_Bayes			SVM		
		Precision	Recall	F ₁	Precision	Recall	F ₁
Indoor	Office	50.5	47.7	49.0	51.5	49.5	50.5
	Living room	49.0	49.0	49.0	48.5	46.2	47.3
	Kitchen	47.0	46.5	46.8	49.0	48.5	48.8
Outdoor	Coast	51.5	49.1	50.2	51.0	50.0	50.5
	Forest	54.5	54.5	54.5	53.5	55.8	54.6
	Mountain	50.5	55.4	52.8	49.5	53.2	51.3
	City	52.5	54.2	53.3	54.5	55.1	54.8
Accuracy		50.78			51.07		

ตารางที่ 4.2 ผลลัพธ์การจำแนกความหมายภาพ 9(v_i) 5 พีเจอร์

Scene/Categories		Performance (%)					
		naïve_Bayes			SVM		
		Precision	Recall	F ₁	Precision	Recall	F ₁
Indoor	Office	57.4	56.9	57.1	59.4	60.0	59.7
	Living room	53.9	52.4	53.1	56.4	53.8	55.1
	Kitchen	52.0	52.5	52.3	56.0	53.3	54.6
Outdoor	Coast	53.4	53.9	53.7	58.4	56.7	57.6
	Forest	57.6	57.0	57.3	57.6	62.0	59.7
	Mountain	53.5	57.4	55.4	53.5	58.7	56.0
	City	59.6	57.3	58.4	60.4	58.1	59.2
Accuracy		55.32			57.39		

ตารางที่ 4.3 ผลลัพธ์การจำแนกความหมายภาพ 9(v_i) 7 พีเจอร์

จากผลการทดลองการจำแนกความหมายภาพเบื้องต้นที่มีการใช้เพียงค่าน้ำหนักวัตถุเด่นบนภาพนั้นจะได้ค่าความถูกต้องโดยเฉลี่ยเพียงครึ่งหนึ่งเท่านั้น ดังแสดงในตารางที่ 4.1 - 4.3 จะเห็นว่าเมื่อมีการใช้เพียง 3 พีเจอร์ในการจำแนก กลุ่มภาพ forest จะได้ค่าความถูกต้อง 49.0% ด้วยวิธีการจำแนกแบบ naïve-Bayes และ 50.5% ด้วย SVM กลุ่มภาพ city จะได้ค่าความถูกต้อง 47.5% ด้วยวิธีการจำแนกแบบ naïve-Bayes และ 55.7% ด้วย SVM แต่เมื่อทดลองจำแนกภาพด้วยข้อมูลค่าน้ำหนักวัตถุเด่นด้วยการเพิ่มจำนวนวัตถุเด่นเป็น 7 พีเจอร์ ผลที่ได้คือในกลุ่มภาพ city จะได้ค่าความถูกต้องถึง 59.2% ด้วย SVM และกลุ่ม forest ได้ค่าความถูกต้องถึง 59.7% ด้วย SVM แต่ กลุ่มเดียวกันจำแนกด้วย naïve-Bayes จะได้ค่าความถูกต้องเป็น 57.3% จะเห็นว่าเมื่อมีการใช้จำนวนพีเจอร์เพิ่มมากขึ้นทำให้ค่าความถูกต้องเพิ่มขึ้น อาจจะเป็นเนื่องจาก การที่มีวัตถุที่โดดเด่นในบาง

ภาพอาจจะไม่ได้แทนถึงความหมายภาพทั้งหมด เช่น ท้องฟ้า ทะเล หรือ คน ที่เป็นวัตถุหลักบนภาพ จะไม่สามารถจำแนกได้ว่ามีความหมายใด แต่อย่างไรก็ตามเมื่อมีการเพิ่มวัตถุที่มีขนาดพื้นที่เล็กลงในตำแหน่งรอบ ๆ ของวัตถุเด่นทำให้เพิ่มค่าความถูกต้อง ดังนั้น เมื่อมีการจำแนกภาพด้วยจำนวนพีเจอร์ที่โดดเด่นถึง 7 พีเจอร์ จะสามารถจำแนกได้ถึง 55.32% แบบ naïve-Bayes และ 57.39% ด้วย SVM

สรุปได้ว่าการจำแนกด้วยค่าน้ำหนักวัตถุที่โดดเด่นบนภาพ ในบางครั้งไม่สามารถจำแนกกลุ่มความหมายภาพโดยรวมได้อย่างแท้จริง ยังคงต้องมีการใช้วิธีการรูปแบบอื่นช่วยในการจำแนกความหมายของภาพเพิ่มขึ้น ดังที่จะนำเสนอผลการทดลองดังต่อไปนี้

4.2.2 การจำแนกด้วยแนวคิดกราฟแบบลำดับชั้น

การทดลองเพื่อทำการจำแนกกลุ่มภาพด้วย วิธีการใหม่ แนวคิดกราฟแบบลำดับชั้นที่นำเสนอ โดยกำหนดให้มีการใช้ข้อมูลภาพชุดเดียวกันกับการทดลองก่อนหน้า และใช้การเปรียบเทียบการจำแนกความหมายภาพด้วย SVM แบบเชิงเส้น (Linear) ด้วย Linear kernel และ แบบไม่เป็นเชิงเส้น (Nonlinear) ด้วย เรเดียลเบสิสฟังก์ชัน (RBF) กล่าวรายละเอียดไว้ในบทที่ 2 และได้เลือกแบบวิธีการเปรียบเทียบการจำแนกแบบ One-against-one และ One-against-all โดยจะมีการปรับค่าพารามิเตอร์ของ C และ ρ เพื่อให้ได้ค่าความถูกต้องที่สูงที่สุด ดังแสดงผลการทดลองในตารางที่ 4.4 - ตารางที่ 4.5

Scene/Categories		One-against-one			One-against-all		
		Precision	Recall	F ₁	Precision	Recall	F ₁
Indoor	Office	83.2	84.0	83.6	73.3	74.0	73.6
	Living room	79.4	78.6	79.0	70.6	73.5	72.0
	Kitchen	78.0	79.6	78.8	75.0	71.4	73.2
Outdoor	Coast	80.6	81.4	81.0	74.8	77.8	76.2
	Forest	73.7	73.7	73.7	78.8	77.2	78.0
	Mountain	78.2	80.6	79.4	80.2	81.0	80.6
	City	88.9	83.8	86.3	80.8	78.4	79.6
Accuracy		80.28			76.17		

ตารางที่ 4.4 ผลลัพธ์การเปรียบเทียบการจำแนกความหมายภาพด้วยวิธี แบบ Linear kernel

จากผลการทดลองได้แสดงในตารางที่ 4.4 เป็นตารางการเปรียบเทียบการจำแนกความหมายของกราฟแบบลำดับชั้นจะเห็นว่า การจำแนกด้วยวิธี Linear kernel จะได้ค่าความถูกต้องถึง

80.28% ด้วย One-against-one และ 76.17% ด้วย One-against-all จะเห็นว่าภาพกลุ่ม city ที่มีการจำแนกด้วย One-against-one ได้ค่า F_1 ถึง 86.3% มีค่า Precision 88.9% Recall 83.8% แต่ One-against-all จะได้ค่า F_1 เป็น 79.6% มีค่า Precision 80.8% Recall 78.4% กลุ่มภาพ forest ที่มีการจำแนกด้วย One-against-one มีค่า Precision 73.7% Recall 73.7% มีค่า F_1 73.7% แต่ One-against-all จะได้ค่า F_1 เป็น 78.0% มีค่า Precision 78.8% Recall 77.2% จะเห็นว่าการจำแนกในกลุ่มภาพ office และ city ได้ค่า F_1 มากที่สุดตามลำดับ 86.3% , 86.3% สำหรับการจำแนกด้วย One-against-one

Scene/Categories		One-against-one			One-against-all		
		Precision	Recall	F_1	Precision	Recall	F_1
Indoor	Office	88.1	91.8	89.9	84.2	89.5	86.7
	Living room	82.4	86.6	84.4	79.4	83.5	81.4
	Kitchen	81.0	84.4	82.7	78.0	84.8	81.3
Outdoor	Coast	79.6	85.4	82.4	78.6	87.1	82.7
	Forest	86.9	80.4	83.5	91.9	77.8	84.3
	Mountain	86.1	82.1	84.1	83.2	82.4	82.8
	City	89.9	84.0	86.8	89.9	81.7	85.6
Accuracy		84.82			83.55		

ตารางที่ 4.5 ผลลัพธ์การเปรียบเทียบการจำแนกความหมายภาพด้วยวิธี แบบ RBF

จากผลการทดลองได้แสดงในตารางที่ 4.5 สำหรับการจำแนกด้วยฟังก์ชันเคอร์เนล แบบเรเดียลเบสิสฟังก์ชัน (RBF) จะได้ค่าความถูกต้องเฉลี่ยถึง 84.82% ด้วย One-against-one และค่าความถูกต้องเฉลี่ยแบบ One-against-all 83.55% จะเห็นว่าภาพกลุ่ม city ที่มีการจำแนกด้วย One-against-one ได้ค่า F_1 ถึง 86.8% มีค่า Precision 89.9% Recall 84.0% แต่ One-against-all จะได้ค่า F_1 เป็น 85.6% มีค่า Precision 89.9% Recall 81.7% กลุ่มภาพ office ที่มีการจำแนกด้วย One-against-one มีค่า Precision 88.1% Recall 91.8% มีค่า F_1 ถึง 89.9% แต่ One-against-all จะได้ค่า F_1 เป็น 86.7% มีค่า Precision 84.2% Recall 89.5% จะเห็นว่าการจำแนกในกลุ่มภาพ office และ city ได้ค่า F_1 มากที่สุดตามลำดับ 86.8% , 89.9% สำหรับการจำแนกด้วย One-against-one



(ก) ภาพสำนักงาน



(ข) ภาพห้องนั่งเล่น



(ค) ภาพห้องครัว



(ง) ภาพชายฝั่ง



(จ) ภาพป่า



(ฉ) ภาพภูเขา



(ช) ภาพเมือง

ภาพที่ 4.3 ตัวอย่างผลลัพธ์ของการจำแนกความหมายภาพ

จากการทดลองที่มีการใช้ในส่วนของแนวคิดกราฟและความสัมพันธ์ของวัตถุภายในภาพจะ
ช่วยในการจำแนกความหมายภาพเพิ่มมากขึ้นจากที่มีการใช้เพียงวัตถุอย่างเดียว ทำให้ค่าความ
ถูกต้องโดยเฉลี่ยเมื่อเปรียบเทียบกับประสิทธิภาพที่จำแนกด้วย SVM แบบข้อมูลวัตถุ 7 ฟีเจอร์ กับ
การจำแนกความหมายภาพด้วยวิธี แบบ RBF One-against-one จะเห็นว่าค่าความถูกต้องโดยเฉลี่ย
เพิ่มขึ้นถึงประมาณ 27% ดังแสดงตัวอย่างผลการจำแนกความหมายภาพตามกลุ่มไว้ในภาพที่ 4.3



บทที่ 5

สรุปผลการทดลอง

ในบทนี้ได้ทำการสรุปผลการทดลอง และข้อเสนอแนะ รวมทั้งข้อผิดพลาดที่เกิดขึ้นจากการทดลอง รวมไปถึงสิ่งที่ควรปรับปรุงเพิ่มเติม เพื่อแก้ไขข้อผิดพลาดและแนวทางการทำวิจัยต่อในเรื่องของ semantic image ซึ่งเป็นงานวิจัยที่ปัจจุบันได้มีนักวิจัยให้ความสนใจอย่างแพร่หลายสามารถนำไปประยุกต์ใช้ได้หลายด้าน เช่น การแพทย์ การขนส่งคมนาคม การรักษาความปลอดภัย หรืออุตสาหกรรม เป็นต้น

5.1 สรุปและอภิปรายผลการวิจัย

งานวิจัยนี้ได้เสนอหัวข้อวิจัยทางการประมวลผลภาพ ในส่วนของการวัดความคล้ายกันของภาพ เพื่อให้ได้ความหมายของภาพที่อยู่ในหมวดหมู่เดียวกัน โดยปกติทั่วไปนั้นการใช้อัลกอริทึมที่มาสกัดข้อมูลภาพนั้นมักจะใช้สกัดเพียงข้อมูลที่เกิดขึ้นภายในภาพ แล้วนำมาประมวลผลเพื่อใช้ในการสืบค้นข้อมูลภาพ แต่ปัจจุบันได้มีการนำคำหลักที่ได้จากการให้ความหมายของการแท็กวัตถุบนภาพมาหาความสัมพันธ์ภายใน โดยพยายามหาความสัมพันธ์ที่คล้ายกันของวัตถุในหมวดหมู่เดียวกัน ดังนั้นในงานวิจัยนี้จึงได้นำเสนอ ในรูปแบบของการแทนข้อมูลภาพ ด้วยความสัมพันธ์ของข้อมูลวัตถุภายในภาพหรือเรียกว่า แนวคิดกราฟ ในลักษณะของกราฟที่นำเสนอ นั้นจะมีการใช้ความสัมพันธ์ระหว่างวัตถุที่เกิดขึ้นของข้อมูลภายในภาพในรูปแบบของลำดับชั้นเพื่อใช้ในการแสดงความสัมพันธ์ของความหมายภาพ ดังนั้นจากการทดลองจะสามารถพิสูจน์ได้ว่าทฤษฎีที่นำเสนอนี้เหมาะสมที่จะนำมาใช้ร่วมกันเพื่อการจำแนกความหมายของภาพ เพราะฉะนั้นในงานวิจัยนี้ยังคงเป็นอีกแนวทางหนึ่งที่สามารถนำวิธีการที่นำเสนอเข้ามาประยุกต์เพื่อให้สามารถนำวิธีการมาใช้ช่วยในแปลความหมายของภาพได้

จากการทดลองได้ทำการแบ่งกลุ่มของภาพไว้ทั้งหมด 7 กลุ่ม เพื่อทดสอบการจำแนกความหมายของข้อมูลภาพ ประกอบด้วยกลุ่มของ ภาพสำนักงาน (office), ภาพห้องนั่งเล่น (living room), ภาพเมือง (city), ภาพห้องครัว (kitchen), ภาพชายฝั่ง (coast), ภาพภูเขา (mountain), ภาพป่า (forest) โดยใช้ข้อมูลคำหลักที่ได้มาจาก WordNet [Miller, George A.,1990] ซึ่งภายใน

คำหลักแต่ละคำมีความสัมพันธ์กันตามความหมายของภาพและทำการจำแนกความเหมือนของภาพทั้งหมดด้วย เครือข่ายแบบเบย์ (Bayesian Network) และ วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) ทั้งแบบเชิงเส้นและไม่เป็นเชิงเส้น จากการทดลองในบทที่ 4 สามารถสรุปได้ว่าการจำแนกภาพด้วยฟังก์ชันเคอร์เนล แบบเรเดียลเบสิสฟังก์ชัน (RBF) ด้วย One-against-one จะได้ค่าความถูกต้องเฉลี่ยถึง 84.82% ด้วย และเมื่อมีการใช้การจำแนกด้วย One-against-all จะได้ค่าความถูกต้องเฉลี่ยลดลงเล็กน้อย 83.55% เพราะฉะนั้นจากการทดลองสามารถสรุปได้ว่า การที่นำทฤษฎีแนวคิดกราฟแบบลำดับชั้นเข้ามาช่วยในการจำแนกความเหมือนกันของกลุ่มภาพ สามารถช่วยในการจำแนกได้เป็นอย่างดีขึ้นเมื่อมีการเปรียบเทียบกับวิธีการจำแนกโดยใช้เพียงวัตถุเด่นบนภาพ

5.2 ข้อเสนอแนะ

ในงานวิจัยส่วนใหญ่เน้นไปที่การสกัดข้อมูลภาพในรูปแบบของการสกัดด้วยอัลกอริทึมที่แตกต่างกัน และค้นหาภาพเพียงความเหมือนกันของรูปทรงหรือลักษณะเฉพาะ หรือเพียงแต่วัตถุบนภาพ เท่านั้น ทั้งที่มีความหมายและไม่มีความหมาย แต่อย่างไรก็ตาม เบื้องต้นของผลลัพธ์ที่ได้จากการค้นคืนหรือจำแนกภาพ คือความเหมือนกันทางกายภาพ เช่น รูปทรง สี หรือชนิดของวัตถุ แต่ลักษณะการวิเคราะห์และพิจารณาของการเหมือนกันทางความหมายภาพ นั้นจะมีลักษณะการวิเคราะห์ที่ต่างออกไป ในอีกรูปแบบหนึ่งซึ่งเป็นรูปแบบที่เกิดจากความคิดของมนุษย์ที่มีการแปลงความจากภาพ แต่อย่างไรก็ตามสิ่งที่ได้จากการทดลองนั้นยังคงมีข้อที่ต้องปรับปรุงเพิ่มอยู่อีกหลายส่วนด้วยกัน

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

เพิ่มขึ้นคือการใช้หลักการของวัตถุที่โดดเด่น รวมทั้งขนาดและตำแหน่งของวัตถุที่เด่นบนภาพ เพื่อแสดงให้เห็นถึงความหมายหลักที่ควรจะนำไปพิจารณา

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

บรรณานุกรม

- นัศพ์ชาณัณ ชินปัญชัณนะ. เอกสารประกอบการสอนวิชา การประมวลผลภาพ (Image processing), มหาวิทยาลัยธุรกิจบัณฑิตย, 2552.
- A. Gupta, S.D. Larson, C. Condit, S. Gupta, L. Fong, L. Chen, and M.E. Martone. “Toward An Ontological Database for Subcellular Neuroanatomy”, 1st Int. Workshop on Conceptual Modeling for Life Sciences Applications (CMLSA), collocated with ER, Auckland New Zealand, 2007.
- 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. Jaimes, S.-F. Chang. “Conceptual framework for indexing visual information at multiple levels”, *SPIE*, vol. 3964, 2000, pp. 2–15.
- A.-M. Tousch, S. Herbin, and J.-Y. Audibert. “Semantic hierarchies for image annotation: A survey”, *Journal Pattern Recognition*, Elsevier Ltd., 2011.
- 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.
- A. Yavlinsky, E. Schoeld, S. Ruger. “Automated Image Annotation Using Global Features and Robust Nonparametric Density Estimation”, in *Proceedings of the International Conference on Image and Video Retrieval*, 2005.
- 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.
- 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.
- C.-W. Hsu and C.-J. Lin. "A comparison of methods for multi-class support vector machines", *IEEE transactions on Neural Networks*, vol. 13, 2002, pp. 415-425.
- Carson, C., et al.. "Blobworld: A system for regionbased image indexing and retrieval," *In Proceedings International Conference Visual Information System*, 1999.
- 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.
- 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.

- 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.S. Hare, P.H. Lewis, P.G.B. Enser, and C.J. Sandom. "Mind the gap: another look at the problem of the semantic gap in image retrieval", *Multimedia Content Analysis, Management and Retrieval*, SPIE, vol. 6073, 2006, pp. 607309. 1–607309.12.
- 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.
- 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.

- L.M. Garshol. "Metadata Thesauri Taxonomies Topic Maps! Making Sense of it all", *Journal of Information Science*, vol. 30, no.4, 2004, pp. 378–391.
- L. Zhang, J. Ma. "Image annotation by incorporating word correlations into multi-class SVM," *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, vol. 15, no. 5, pp. 917-927, 2010.
- M. Spain and P. Perona. "Measuring and predicting importance of objects in our visual world," Technical report, California Institute of Technology, 2007.
- Ma, W. 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.
- 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. Serrano, A. Savakis, J. Luo. "A computationally efficient approach to indoor/outdoor scene classification", In *Proceedings of the IEEE International Conference on Pattern Recognition*, Quebec, Canada, pp. 146-149, 2002.
- Kompatsiaris, Y., and Staab, S.. "M-ontomat-annotizer: Image Annotation Linking Ontologies and Multimedia Low-Level Feature," In *Proceedings of KES (3)*, vol. 4253, 2006.

- 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.
- Q. Iqbal and J. K. Aggarwal. "CIRES: A System for Content-based Retrieval in Digital Image Libraries", *International Conference on Control, Automation, Robotics and Vision*, 2002, pp. 205-210.
- 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. *Digital 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. Courant and D. Hilbert. *Methods of Mathematical Physics*, vol. 1 Interscience Publishers Inc., 1953.
- R. Rifkin and A. Klautau. "In defence of one-vs-all classification", *Journal of Machine Learning Research*, vol. 5, pp. 101-141, 2004.
- 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.
- S. Shatford Layne. "Some issues in the indexing of images", *Journal of the American Society for Information Science*, vol. 45, no. 8, 1994, pp. 583–588.

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.

Squigl: <http://www.gwap.com> ค้นคืนเมื่อวันที่ 9 ตุลาคม 2554.

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.

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.

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

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.

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.

Z. Guo, Z.M. Zhang, E.P. Xing, and C. Faloutsos. “A Max Margin Framework on image annotation and multimodal image retrieval”, *IEEE International Conference on Multimedia and Expo*, 2007, pp. 504–507.



ประวัติผู้วิจัย

ผู้ช่วยศาสตราจารย์นักศัพทศาสตร์ ชินปัญชธนะ

สถิติประยุกต์ (คอมพิวเตอร์) สถาบันบัณฑิตพัฒนบริหารศาสตร์
วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์) มหาวิทยาลัยหอการค้าไทย

ปัจจุบัน

อาจารย์ประจำภาควิชาเทคโนโลยีสารสนเทศธุรกิจ คณะเทคโนโลยีสารสนเทศ
มหาวิทยาลัยธุรกิจบัณฑิต

คณะกรรมการกลุ่มผลิตชุดวิชาการระบบสำนักงานอัตโนมัติและพาณิชย์อิเล็กทรอนิกส์

คณะกรรมการกลุ่มผลิตชุดวิชาการจัดการเว็บไซต์

คณะกรรมการกลุ่มผลิตชุดวิชาการสื่อสารข้อมูลและระบบเครือข่ายคอมพิวเตอร์

สาขาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

รางวัลและเกียรติประวัติ

ได้รับทุนการศึกษาประเภทเรียนดี มหาวิทยาลัยหอการค้าไทย พ.ศ. 2534 – 2537

ได้รับปริญญาตรีเกียรตินิยมอันดับ 1 จาก มหาวิทยาลัยหอการค้าไทย

งานวิจัยที่สนใจ

สนใจงานวิจัยทางการค้นคืนความหมายภาพ (Image retrieval) ทั้งภาพธรรมชาติทั่วไป (natural images) และ ภาพที่เป็นส่วนบุคคล (personal images) ในหัวข้อเกี่ยวกับการค้นคืน human activity หรือในหัวข้อการแปลความหมายภาพ (semantic human image) โดยเจาะจงทางการใช้ท่าทางของมนุษย์เพื่อแสดงถึงความหมายของภาพ รวมทั้งการจำแนกภาพต่างๆ เป็นกลุ่ม (Image classification) และการประยุกต์หลักการประมวลผลภาพ เพื่อนำมาใช้ในทางอุตสาหกรรมได้จริง

หนังสือ

นศัพนธ์ชาณัณ ชนป้ญช้ธนะ. หน่วยที่ 13 การจัตแฟ้มข้อมูลและการป้อกันระบบคอมพิวเตอร์, สถาปัตยกรรมคอมพิวเตอร์และระบบปฏิบัติการ (*Computer Architecture and Operation Systems*) (99315), มหาวิทยาลัยสุโขทัยธรรมาธิราช, สาขาวิทยาศาสตร์และเทคโนโลยี 2557.

นศัพนธ์ชาณัณ ชนป้ญช้ธนะ. หน่วยที่ 1 ความรู้เบื้องต้นเกี่ยวกับเครือข่ายคอมพิวเตอร์ และ หน่วยที่ 3 ชุมชนเครือข่าย, หลักการบริหารและจัตการเครือข่าย (*Network Management*) (99412) , มหาวิทยาลัยสุโขทัยธรรมาธิราช, สาขาวิทยาศาสตร์และเทคโนโลยี 2556.

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

นศัพนธ์ชาณัณ ชนป้ญช้ธนะ. หน่วยที่ 12 โลจิสติกส์และโซ่อุปทานในระบบพาณิชย์อิเล็กทรอนิกส์ และ หน่วยที่ 13 การชำระเงินในระบบพาณิชย์อิเล็กทรอนิกส์, ระบบสำนักงานอัตโนมัติและพาณิชย์อิเล็กทรอนิกส์ (*Office Automation System and Electronic Commerce*) (99311), มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขาวิทยาศาสตร์และเทคโนโลยี 2555.

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

นศัพนธ์ชาณัณ ชนป้ญช้ธนะ. หน่วยที่ 3 ตัวกลางในการสื่อสารข้อมูล และ หน่วยที่ 10 โพรโทคอล, การสื่อสารข้อมูลและระบบเครือข่ายคอมพิวเตอร์ (*Data Communications and Networking*), มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขา วิทยาศาสตร์และเทคโนโลยี, 2553.

งานวิจัย

นศัพนธ์ชาณัณ ชนป้ญช้ธนะ. การรู้จำท่าทางมนุษย์จากภาพส่วนบุคคลผ่านการวิเคราะห์การใช้พลังงานร่างกาย, มหาวิทยาลัยธุรกิจบัณฑิตย์, 2555.

นัศพ์ชาณัณ ชินปัญชัณนะ. การแปลความหมายภาพด้วยวิธีการวัดความคล้ายกันของกราฟแบบจับคู่, มหาวิทยาลัยธุรกิจบัณฑิตยั, 2554.

นัศพ์ชาณัณ ชินปัญชัณนะ. ระบบตรวจนับวัตถุแบบอัตโนมัติด้วยเทมเพลตแมชชิ่ง, มหาวิทยาลัยธุรกิจบัณฑิตยั, 2553.

นัศพ์ชาณัณ ชินปัญชัณนะ. การจำแนกความหมายของภาพจากวัตถุโดยใช้หลักการโครงสร้างสเกตริตรอน, มหาวิทยาลัยธุรกิจบัณฑิตยั, 2552.

ผลงานทางวิชาการ

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)*, Kuala Lumpur, 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 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.
