



รายงานการวิจัย

เรื่อง

การแปลความหมายภาพด้วยวิธีการวัดความคล้ายกันของกราฟแบบจับคู่

Semantic Images Interpretation using a Similarity Measure with Graph Matching

โดย

นศัพนธ์ชาณัน ชินปัญญาชณะ

มหาวิทยาลัยธุรกิจบัณฑิตย์

รายงานผลการวิจัยนี้ได้รับทุนอุดหนุนจากมหาวิทยาลัยธุรกิจบัณฑิตย์

พ.ศ. 2554

## กิตติกรรมประกาศ

งานวิจัยชิ้นนี้สำเร็จล่วงได้ทั้งนี้เพราะได้รับความอนุเคราะห์ การสนับสนุน และแรงผลักดัน รวมถึงติดตามงานของผู้วิจัยอย่างต่อเนื่อง ได้แก่ รองศาสตราจารย์ ดร.สรชัย พิศาลบุตร รองศาสตราจารย์ ดร.นุชรี เปรมชัยสวัสดิ์ และอีกหลายฝ่ายที่ได้ให้ความช่วยเหลือเอื้อเฟื้อข้อมูลที่เป็นประโยชน์ในการทดลองโปรแกรม บันทึกผลการทดลอง และงานวิจัยนี้จะไม่สมบูรณ์ได้ หากไม่ได้ข้อเสนอแนะที่เป็นประโยชน์จากท่านอาจารย์ผู้ทรงคุณวุฒิ ผู้ซึ่งให้คำแนะนำที่ดีและมุมมองที่ผู้วิจัยได้นำมาปรับปรุงให้สมบูรณ์ยิ่งขึ้น

นอกจากนี้ใคร่ขอขอบคุณมหาวิทยาลัยธุรกิจบัณฑิต ผู้ให้ทุนสนับสนุนงานวิจัยชิ้นนี้ หากงานวิจัยเล่มนี้มีข้อผิดพลาด ประการใดขออภัยไว้ ณ ที่นี้ด้วย

นัศพัชฌัน ชินปัญชณะ

พฤศจิกายน 2554

## 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	4
1.3 สมมุติฐาน	4
1.4 ขอบเขตของการวิจัย	4
1.5 ประโยชน์ที่คาดว่าจะได้รับ	5
1.6 นิยามคำศัพท์	6
บทที่ 2 ทฤษฎีและงานวิจัย	7
2.1 ทฤษฎีเบื้องต้น	7
2.2 การประมวลผลภาพดิจิทัล	13
2.3 การเปรียบเทียบความคล้ายของภาพ	19
2.4 การใช้โมเดลในการสร้างการตัดสินใจ	32
2.5 การวัดประสิทธิภาพ	34
บทที่ 3 ขั้นตอนวิธีการดำเนินงานวิจัย	36
3.1 ขั้นตอนการเตรียมข้อมูล	36
3.2 ขั้นตอนการประมวลผล	46
3.3 ขั้นตอนการวัดประสิทธิภาพ	52
บทที่ 4 ผลการทดลอง	53
4.1 การกำหนดข้อมูลภาพ	53
4.2 ผลการทดลอง	54

บทที่ 5 สรุปผลการทดลอง	59
5.1 สรุปและอภิปรายผลการวิจัย	59
5.2 ข้อเสนอแนะ	60
บรรณานุกรม	62
ประวัติผู้วิจัย	69

DPU

## สารบัญตาราง

### ตารางที่หน้า

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 การประมวลผลแบบค้นคืนด้วยคุณลักษณะพีเจอร์ระดับต่ำ	8
2.2 ตัวอย่างขององค์ประกอบของภาพ	9
2.3 ตารางแสดงการวัดประสิทธิภาพ	10
2.4 การแก้กองค์ประกอบของภาพ	11
2.5 การหาตำแหน่งวัตถุของภาพด้วยโครงสร้างสเกตตรีตรอน	13
2.6 การแสดงตำแหน่งของพิกเซลบนภาพดิจิทัลด้วยเมทริกซ์	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 ภาพตัวอย่างที่ถูกแก้ด้วยคำหลัก	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	แสดงการเปรียบเทียบผลการทดลองการตรวจนับการเคลื่อนถุงปูนซีเมนต์บนสายพานด้วยโปรแกรมประมวลผลภาพแบบแทมเพลตแมชชีง	54
4.3	ตัวอย่างผลลัพธ์ของการจำแนกความหมายภาพ	57

- ชื่อเรื่อง : การแปลความหมายภาพด้วยวิธีการวัดความคล้ายกันของกราฟแบบจับคู่
- ผู้วิจัย : นิสพัชฌ์ณ ชินปัญชัรณะ                      สถาบัน : มหาวิทยาลัยธุรกิจบัณฑิตย์
- ปีที่พิมพ์ : พุทธศักราช 2554                              สถานที่พิมพ์ : มหาวิทยาลัยธุรกิจบัณฑิตย์
- แหล่งที่เก็บรายงานการวิจัยฉบับสมบูรณ์ : ศูนย์วิจัยมหาวิทยาลัยธุรกิจบัณฑิตย์
- จำนวนหน้าวิจัย : 72 หน้า                                      ลิขสิทธิ์ : มหาวิทยาลัยธุรกิจบัณฑิตย์
- คำสำคัญ : การประมวลผลภาพ, กราฟ, การวัดความคล้ายกันของกราฟ, การจำแนกภาพ

## บทคัดย่อ

วิวัฒนาการของเครื่องมือและอุปกรณ์ถ่ายภาพดิจิทัลได้พัฒนาอย่างรวดเร็วจนทำให้ภาพถ่ายภาพดิจิทัล มีจำนวนเพิ่มมากขึ้น ทำให้การจัดเก็บข้อมูลภาพเพิ่มมากขึ้น ทำให้การค้นคืนข้อมูลภาพและจำแนกข้อมูลภาพให้ตรงตามความหมายของภาพมีความยากลำบากเพิ่มขึ้นเช่นกัน ดังนั้นมีงานวิจัยมากมายพยายามจะคิดค้นกระบวนการ อัลกอริทึมต่าง ๆ เพื่อทำให้เกิดความพึงพอใจของผู้ใช้มากที่สุด แต่อย่างไรก็ตาม การสืบค้นหรือการจำแนกข้อมูลภาพส่วนใหญ่ยังคงคำนึงเพียงแต่ความเหมือนกันของวัตถุที่ปรากฏบนภาพเท่านั้น ซึ่งยังไม่เพียงพอต่อการแปลความหมายที่แท้จริง

งานวิจัยนี้ได้นำเสนอวิธีการใหม่สำหรับการแปลความหมายภาพด้วยวิธีการวัดความคล้ายกันของกราฟแบบจับคู่เพื่อใช้ในระบบการสืบค้น โดยมีทั้งหมด 3 ส่วนคือ (1) การแท็กภาพ (2) การกำหนดตัวแทนความหมาย (3) การกำหนดวิธีการจับคู่ความหมาย โดยที่ส่วนที่ 1 เป็นการให้ความหมายของวัตถุเด่นบนภาพลงบนฐานข้อมูล ส่วนที่ 2 เป็นการแทนที่ข้อมูลบนภาพด้วยแนวคิดกราฟ ส่วนสุดท้ายเป็นการวัดความคล้ายกันของกราฟแบบจับคู่ ผลที่ได้จากการจำแนกความหมายภาพด้วย วิธีที่นำเสนอใหม่นี้สามารถจำแนกความหมายของภาพได้ดีกว่าวิธีอื่นๆ และได้ค่าความถูกต้องสูงถึง 88.81%



**Title** : Semantic Images Interpretation using a Similarity Measure with Graph Matching  
**Researcher** : Nutchanun Chinpanthana      **Institution** : Dhurakijpundit University.  
**Year of Publication** : 2011      **Publisher** : Dhurakijpundit University.  
**Sources** : Dhurakijpundit University Research Center.  
**Number of Pages** : 72 Pages      **Copyright** : Dhurakijpundit University.  
**Keyword** : Image processing, Graph, Similarity Measure with Graph Correlation, Image Classification

## Abstract

Searching images their semantic is an active problem in multimedia image retrieval. Many researchers have attempted to improve semantic models by using high level concept based on keyword annotation. However, the annotation is tedious, inconsistent, and erroneous. The retrieval process of such approaches is done by keyword searching. This model is rather rudimentary and it does not specific enough for representing the actual meaning.

This paper presents a new approach to represent the semantic concepts that support the semantic image indexing and retrieval systems. The approach is composed of three main phases: (1) image annotation, (2) semantic representative, and (3) semantic schema matching. Phase 1 is labeling the salient contents into content archive. Phase 2 is representing an image into the conceptual graph. Last phase is finding the similarity matching between the conceptual graph and representative graph. The results are compared to the classification methods. The experimental results indicate that our proposed approach offers significant performance improvements in the interpretation of semantic images, compared, with the maximum of 88.81% accuracy.

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาของปัญหา

สมัยก่อนการประมวลผลภาพ จะอยู่ในรูปของการประมวลผลสัญญาณแอนะล็อก (analog) โดยใช้อุปกรณ์จับภาพแตงภาพที่ถ่ายออกมายังคงเป็นภาพขาวดำและไม่ชัดเจนเท่าที่ควร ทำให้ส่วนใหญ่ภาพถ่ายมักจะถูกถ่ายเฉพาะภาพเหตุการณ์ที่สำคัญหรือบุคคลสำคัญเท่านั้น เนื่องจาก อุปกรณ์ที่นำมาใช้ในการถ่ายภาพ หรือบันทึกภาพนั้นซับซ้อนและมีความยุ่งยากในการใช้งาน รวมทั้งราคาสูง แต่อย่างไรก็ตามได้มีการพัฒนาปรับปรุงเทคโนโลยีเพื่อตอบรับกับความต้องการที่เพิ่มขึ้น ทำให้อุปกรณ์ถ่ายภาพในปัจจุบันมีราคาถูกลง และสะดวกต่อการใช้งานมาก และพัฒนาต่อเนื่องไปเป็นอุปกรณ์ถ่ายภาพแบบดิจิทัล (digital) ทำให้ภาพถ่ายภาพดิจิทัล มีจำนวนเพิ่มมากขึ้นอย่างต่อเนื่อง ด้วยเหตุนี้ทำให้เกิดปัญหาในการจัดเก็บข้อมูลภาพรวมถึงการค้นหาข้อมูลภาพ (image retrieval) ที่เพิ่มมากขึ้นอย่างต่อเนื่อง จนทำให้เกิดคำถามเดิม ๆ ว่าควรจะใช้วิธีการใดจึงจะสามารถจัดเก็บอย่างมีระบบและสามารถที่จะค้นหาภาพที่ต้องการได้อย่างถูกต้อง

การประมวลผลภาพดิจิทัล (digital image processing) เป็นสาขาที่ได้รับความนิยมในการวิจัยค่อนข้างมาก ทั้งทางด้านภาพ (still image) และ วิดีโอ (video) เพราะสามารถที่จะนำไปประยุกต์ได้หลายด้าน เช่น ทางด้านการแพทย์ ทางด้านอุตสาหกรรม หรือนำมาประยุกต์ใช้ในส่วนของการจัดการจราจร ในการตรวจจับทะเบียนรถยนต์ เป็นต้น หัวข้อหลักในงานวิจัยแขนงนี้จะเกี่ยวเนื่องกันทั้งทางด้านการสร้างกระบวนการ (algorithm) เพื่อนำมาใช้ในการจำแนกหรือสืบค้นข้อมูลภาพ, การแทนความหมายของข้อมูลภาพ (image representation) และวิธีการใช้เครื่องมือในการจำแนกภาพ (classification method) [Pedro,2007] ซึ่งปัญหาทั้งหมดที่เกิดขึ้นเพียงเพื่อนำมาใช้ในการจำแนกหรือสืบค้นข้อมูลภาพ เพื่อให้ได้ภาพคำตอบที่ตรงตามความต้องการมากที่สุด สิ่งทีกล่าวมาข้างต้นเป็นเพียงส่วนหนึ่งที่นำการประมวลผลภาพดิจิทัลเข้ามาปรับปรุงและไปประยุกต์ใช้งาน ปัจจุบันมีงานวิจัยหลายกลุ่ม [Qian Huang,1995][Vailaya A.,2001][W. Ma,1997][C. Carson, 1999] พยายามนำเทคนิคต่าง ๆ เข้ามาใช้ในการค้นหาภาพเพื่อให้ผลลัพธ์ตรงกับความต้องการของผู้ใช้ ไม่ว่าจะเป็นการสืบค้นข้อมูลโดยใช้การประมวลผลภาพระดับต่ำ (low level image processing) เป็นการนำคุณลักษณะของภาพต่าง ๆ หรือฟีเจอร์ (feature) ที่ถูกสกัด (extraction) อย่างเช่น สี (color)

ลวดลาย (texture) หรือ รูปทรง (shape) ด้วยวิธีการ อัลกอริทึมต่างๆ เพื่อใช้สำหรับการสืบค้นข้อมูล

ความพยายามในการค้นหาข้อมูลภาพเริ่มมาจากการสกัดคุณลักษณะข้อมูลภายในภาพ (feature extraction) ออกมาเป็นข้อมูลที่สามารถคำนวณได้ ข้อมูลส่วนนี้มักจะถูกเรียกกันว่าข้อมูลภาพระดับต่ำหรือ low-level features โดยข้อมูลนี้จะถูกนำมาใช้สำหรับการสืบค้นข้อมูลภาพแต่อย่างไรก็ตาม การวิจัยในส่วนแรกนี้ยังคงมีการปรับปรุงพัฒนาอย่างต่อเนื่อง แต่อย่างไรก็ตามได้มีงานวิจัยอีกกลุ่ม ที่พยายามจะใช้เทคนิคของการเข้าใจความหมายของภาพแทน การสืบค้นแบบข้างต้นเรียกว่า การประมวลผลภาพระดับสูง (high level image processing) งานวิจัยในกลุ่มนี้พยายามที่จะพิจารณาข้อมูลบนภาพเป็นวัตถุ (object) ที่มีความหมาย [Benitez A.B,2002] และแทนวัตถุนั้นๆ ด้วยคำหลัก (keyword) ลงบนภาพ เช่น “building”, “car” , “ball” เป็นต้น และใช้ความหมายหรือคำหลักนั้นเพื่อทำการสืบค้นข้อมูลแทน เมื่อเปรียบเทียบผลลัพธ์ของการค้นคืนที่ค่อนข้างแม่นยำมากกว่าการใช้พีเจอร์ระดับต่ำ แต่อย่างไรก็ตามยังขึ้นอยู่กับว่าอัลกอริทึมที่ถูกนำมาใช้ร่วมด้วยนั้นจะเป็นลักษณะใด เช่น การใช้เทคนิคหาความสัมพันธ์ของแท็กชื่อวัตถุนบนภาพ [Benitez A.B,2001][ R. Zhao,2002][Philippe Mulhem,2002] ซึ่งเป็นการสอดคล้องกันด้วยความหมายตามพจนานุกรม การใช้ความสัมพันธ์ของความหมายที่เหมือนกันของคำ (synonym) การค้นหาภาพด้วยเทคนิคนี้จะได้ผลลัพธ์ ที่ค่อนข้างขึ้นกับคำหลักที่ถูกให้ความหมายไว้ในภาพหรือ วัตถุนบนภาพส่วนใหญ่ เพราะฉะนั้น ผลลัพธ์ของการค้นคืนภาพ คือการค้นหาคำหลักที่อยู่ในภาพนั่นเอง จึงทำให้ความหมายภาพที่แท้จริงจะไม่สามารถเกิดขึ้นได้เลย

นักวิจัยบางกลุ่มได้มีการสร้างความสัมพันธ์ของวัตถุนบนภาพด้วยโมเดลเหตุการณ์ (event model) [Joo-Hwee Lim 2003] เช่นการเชื่อมวัตถุด้วยคำต่างๆ เช่น “touch”, “ontop” เป็นต้น แต่ผลของการทดลองยังไม่ดีนัก บางงานวิจัยได้พยายามใส่ข้อความที่บรรยายความหมายของภาพ (context) [Mathias Lux 2003] [Mathias Lux 2009] ในหัวข้อที่สอดคล้องกับภาพ เช่น “birthday party of uncle Adam” หรือ “a picture showing a barking dog” แต่อย่างไรก็ตาม การสืบค้นภาพที่ถูกบรรยายนั้นส่วนใหญ่จะเป็นจะภาพส่วนตัว (personal images) บางงานวิจัยพยายามที่จะใส่ข้อมูลเหตุการณ์ต่างๆจนครบถ้วนในรูปแบบของ ใคร ทำอะไร ที่ไหน เมื่อไหร่ (who, what, when, where) ในการใส่ข้อมูลบางครั้ง เป็นข้อมูลที่นอกเหนือ หรือเกินความจำเป็นโดยใช้เหตุ ข้อมูลเหล่านี้อาจจะไม่มีความจำเป็นเลยสำหรับการสืบค้นข้อมูลหรือ แทนความหมายของภาพจริงๆ หรือบางครั้งอาจจะเป็นข้อมูลที่มีความเป็นส่วนตัวจนเกินไป และใช้คำศัพท์ซ้ำซ้อน ฟุ่มเฟือย

และมีนักวิจัยบางกลุ่มพยายามที่แก้ไขถึงความยุ่งยากลำบากในการใส่ข้อมูลบนภาพ (image annotation) และจำกัดขอบเขตคำศัพท์ของภาพให้รัดกุมยิ่งขึ้น จึงได้พัฒนาซอฟต์แวร์เพื่อให้เกิดความสะดวกสบายในการใส่ข้อมูล และการใส่ข้อมูลที่มีแบบแผนแน่นอน ยกตัวอย่างงานวิจัยเช่น Caliph & Emir [Mathias Lux 2009], Annosearch [Xin-Jing 2008], CAMEL [Apostol Paul Natsev 2001] เป็นต้น แต่อย่างไรซอฟต์แวร์ส่วนใหญ่ๆนั้น จะทำการสืบค้นภาพในรูปแบบของการเทียบคำศัพท์เป็นคำต่อคำตามที่มีการเก็บข้อมูลไว้เท่านั้น เพราะฉะนั้นคำตอบที่ได้จะไม่ได้ขึ้นกับความหมายของภาพ แต่ขึ้นกับคำหลักที่มีการเก็บข้อมูลลงไปเท่านั้น ความหมายของภาพยังไม่ได้ อย่างไรก็ตาม ในความเป็นจริงแล้วความหมายของภาพจะเกิดจากความสัมพันธ์ที่คล้ายกันของวัตถุในหมวดหมู่เดียวกัน เพราะฉะนั้นผลลัพธ์ของกลุ่มภาพที่ได้จะไม่ได้อันเนื่องมาจากความหมายของภาพอย่างแท้จริง แต่ขึ้นกับคำหลักที่มีการเก็บข้อมูลลงไปบนภาพเท่านั้น จึงทำให้ความหมายของภาพโดยรวมนั้นยังไม่ได้สื่อออกมาให้เห็นอย่างเด่นชัด ซึ่งในความเป็นจริงแล้วความหมายของภาพจะเกิดจากการแปลความหมายรวมของทุกวัตถุที่ปรากฏบนภาพ ดังนั้นในงานวิจัยนี้ได้นำเสนอในรูปแบบของการแทนข้อมูลภาพ ด้วยความสัมพันธ์ของข้อมูลวัตถุภายในภาพ หรือเรียกว่า แนวคิดกราฟ (Conceptual Graph) ในลักษณะของกราฟที่นำเสนอจะมีความสัมพันธ์ของข้อมูลภายในภาพของวัตถุที่เกิดขึ้นทั้งหมดในรูปแบบของกราฟรวมทั้งความสัมพันธ์ (Relationship) ระหว่างวัตถุที่เกิดขึ้น และจะมีการวัดค่าความสัมพันธ์ของความเหมือนกันด้วยวิธีการที่เรียกว่า การจับคู่กราฟ (Graph Matching) ของข้อมูลภาพที่ถูกแทนค่าความหมายไว้ ดังนั้นภาพที่มีความหมาย เหมือนกันจะมีได้ค่าของความเหมือนกันมากที่สุด

จากที่กล่าวมาข้างต้นในงานวิจัยนี้จึงได้นำเสนอการปรับปรุงการสืบค้นข้อมูลภาพด้วยความหมายภาพเพื่อให้ได้ผลลัพธ์ของภาพที่มีอย่างแท้จริง โดยใช้วิธีการแทนข้อมูลภาพแบบคอนเซ็ปชวลกราฟ (Conceptual Graph) และทำเปรียบเทียบความเหมือนกันของความหมายภาพด้วยการหาความเหมือนของภาพทั้งหมด 4 วิธีการ โครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network) แผนผังการจัดระบบตัวเอง (Self-organizing maps) เครือข่ายแบบเบย์ (Bayesian Network) และการวัดความคล้ายกันของกราฟแบบจับคู่ (Similarity Measure with Conceptual Graph Matching)

## 1.2 วัตถุประสงค์

1. เพื่อค้นหาเทคนิคใหม่ที่สามารถนำมาใช้วัดความคล้ายกันของกราฟแบบจับคู่
2. เพื่อสร้างโมเดลรูปแบบใหม่ในการแทนข้อมูลภาพ
3. เพื่อสร้างแบบจำลองวิเคราะห์การค้นหาภาพรูปแบบใหม่
4. เพื่อพัฒนาเทคนิคในการหาความเหมือนกันของภาพ
5. สามารถนำไปประยุกต์กับการสืบค้นภาพได้

## 1.3 สมมติฐาน

งานวิจัยนี้จึงเสนอวิธีการแปลความหมายภาพด้วย วิธีวัดความคล้ายกันของกราฟแบบจับคู่ โดยใช้หลักทฤษฎี การเชื่อมโยงประสานกับวัตถุบนภาพให้สอดคล้องตามความหมายของกระบวนการคิด แปลความหมายภาพของมนุษย์ และได้ทำการเปรียบเทียบความเหมือนกันของความหมายภาพด้วยการหาความเหมือนของภาพทั้งหมด 4 วิธีการ โครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network) แผนผังการจัดระบบตัวเอง (Self-organizing maps) เครือข่ายแบบเบย์ (Bayesian Network) และการวัดวิธีการวัดความคล้ายกันของกราฟแบบจับคู่ (Similarity Measure with Conceptual Graph Matching)

## 1.4 ขอบเขตของการวิจัย

1. ข้อมูลภาพที่นำมาวิเคราะห์เป็นข้อมูลภาพดิจิทัล เฉพาะภาพธรรมชาติ (natural images) เช่น ภาพทิวทัศน์ ชายทะเล ภูเขา หรือ สวนสาธารณะ ภาพภายในอาคาร ภายในบ้าน สำนักงาน หรือ โรงงาน เป็นต้น
2. ภาพที่มีวัตถุเกิดปรากฏขึ้น สำหรับภาพที่ไม่มีความหมายหรือเป็นภาพที่มีความหมายกำกวม หรือภาพที่สามารถแปลได้หลายความหมาย หรือภาพที่โฟกัสระยะใกล้ จะไม่นำภาพนั้นเข้ามาวิเคราะห์สำหรับการหาความหมายภาพข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลภาพดิจิทัลที่เห็นวัตถุชัดเจน
3. ข้อมูลหรือวัตถุบนภาพที่เป็นอินพุตบนภาพจะถูกแท็ก เป็นคำศัพท์และขนาดมาก่อนด้วย ขบวนการ รู้จำของการประมวลผลภาพในระดับต่ำ (low-level image)

4. ในการสร้างแบบจำลองภาพ (image representation) ข้อมูลจะถูกจัดเก็บมาในรูปแบบของพีเจอร์ข้อมูลไว้แล้ว
5. ภาพจากคลังภาพจะถูกตั้งค่าเริ่มต้นโดยการใส่ข้อมูลวัตถุบนภาพพร้อมความสัมพันธ์ภาพไว้ก่อนแล้ว
6. ภาพที่ถูกนำมาใช้เป็นข้อมูลทดลองจากคลังภาพ [11]-[13]
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. เพื่อเพิ่มตัวชีวิตให้กับองค์กร สถาบันในส่วนของงานวิจัยและพัฒนา
4. สร้างชื่อเสียงให้กับมหาวิทยาลัย เมื่อมีบทความวิจัยลงในวารสารต่างประเทศ บทความวิจัยที่นำเสนอในการประชุมระดับชาติ และนานาชาติ

### 1.5.3 ผลต่อกลุ่มผู้วิจัย

1. พบแนวทางในการคิดค้นสิ่งใหม่ๆ ในการจัดกลุ่มและกระบวนการแปลความหมายของภาพ
2. สามารถนำผลวิจัยมาเขียนบทความลงวารสารนานาชาติ และร่วมประชุมวิชาการระดับชาติและนานาชาติ
3. พัฒนาความรู้ใหม่ๆ 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 ส่วนของวัตถุนภาพ ยกตัวอย่างเช่น คน, ต้นไม้ เป็นต้น
6. คำหลัก (Keyword) หมายถึง คำที่มีความหมายได้ใจความใช้แทนวัตถุนภาพ
7. แท็ก (Tag) หมายถึง การกำหนดคำ หรือ คำศัพท์บนภาพ หรือเรียกว่าการ labeled
8. การแทนที่ข้อมูลภาพด้วยกราฟ (Graph representation) คือการแทนความหมายของภาพโดยใช้กราฟ
9. การเปรียบเทียบความเหมือนของกราฟ (Graph Similarity) คือการวัดค่าความคล้ายกันของกราฟ
10. คุณลักษณะข้อมูล หรือ ฟีเจอร์ (Feature) หรือตัวแปร ที่ถูกสกัดออกมาจากภาพ เช่น สี (color) ลวดลาย (texture) หรือ รูปทรง (shape) รวมทั้ง วัตถุ ที่ปรากฏบนภาพเพื่อนำมาใช้ในการสืบค้นข้อมูลต่อไป
11. โครงข่ายประสาทเทียม (Artificial neural network) หมายถึง เทคนิคการพยากรณ์เหมาะสำหรับลักษณะข้อมูลซึ่งไม่ใช่เชิงเส้น

## บทที่ 2

### ทฤษฎีและงานวิจัย

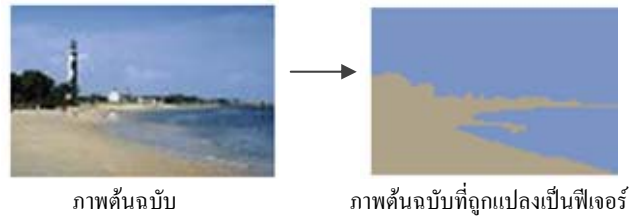
#### 2.1 ทฤษฎีเบื้องต้น

วิวัฒนาการของเครื่องมือและอุปกรณ์ถ่ายภาพดิจิทัลได้พัฒนาอย่างรวดเร็วจนทำให้ภาพถ่ายภาพดิจิทัล มีจำนวนเพิ่มมากขึ้น ปัญหาที่ตามมาก็คือการจัดเก็บข้อมูลภาพที่เพิ่มมากขึ้นอย่างไร ซึ่งจำกันี้จะต้องทำอย่างไรจึงจะสามารถจัดเก็บอย่างมีระบบและสามารถสืบค้นข้อมูลภาพ และจำแนกข้อมูลภาพให้ตรงตามความหมายของภาพที่ต้องการของผู้ใช้มากที่สุด ทำให้งานวิจัยในปัจจุบันที่เกี่ยวข้องกับการค้นคืนรวมทั้งการจัดกลุ่มภาพให้ตรงกับความถี่ความต้องการเพิ่มมากขึ้น รวมไปถึงระยะเวลาในการสืบค้นที่น้อยลงกับปริมาณของภาพที่เพิ่มทวีคูณ ดังนั้นปัญหาดังกล่าวข้างต้นนั้นจึงได้รับความสนใจจากนักวิจัยหลายกลุ่ม ซึ่งเป็นงานด้านการประมวลผลภาพ (image processing) ด้านการค้นคืนสารสนเทศ (image retrieval) เป็นอีกหนทางหนึ่งในการแก้ปัญหาดังกล่าว เนื่องจากสามารถช่วยให้การค้นหาข้อมูลกระทำได้อย่างสะดวกยิ่งขึ้นรวมถึง การค้นหาภาพและการจำแนกประเภทข้อมูลภาพ (image classification) เพื่อคัดเลือกภาพ เพื่อให้ตรงตามความต้องการของผู้ใช้ ยิ่งมีความจำเป็นอย่างยิ่ง

สำหรับงานวิจัยทางการประมวลผลภาพในการค้นคืนสารสนเทศกลุ่มแรก ๆ จะมีการค้นคืนตามคุณลักษณะพื้นฐานของภาพที่ถูกสกัดคุณลักษณะด้วยอัลกอริทึมต่างๆ ยกตัวอย่างเช่น สี (color) ลวดลาย (texture) รูปทรง (shape) เป็นต้น กระบวนการนี้ถูกเรียกว่า การประมวลผลภาพระดับต่ำ (low-level image processing) [Jain A.K.,1996][Cheng Y.C.,2003] กระบวนการนี้สามารถค้นหาภาพได้ตามคุณลักษณะพื้นฐานที่นำไปสืบค้น โดยภาพผลลัพธ์ส่วนใหญ่มักจะเป็นภาพที่มีคุณลักษณะไม่ซับซ้อนมากนัก เช่น โทนสี หรือ รูปทรงที่แตกต่างกันอย่างเด่นชัด ดังภาพที่ 2.1 แสดงการค้นคืนของ 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



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

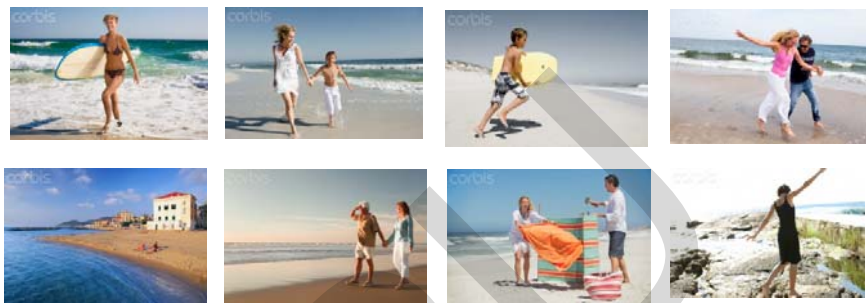
ข. ผลลัพธ์ของการค้นคืนด้วยภาพต้นฉบับ ก.

ภาพที่ 2.1 การประมวลผลแบบค้นคืนด้วยคุณลักษณะฟีเจอร์ระดับต่ำ<sup>1</sup>

แต่อย่างไรก็ตามได้มีกลุ่มนักวิจัยที่พยายามปรับปรุงแปลงอัลกอริทึมด้วยการประมวลผลภาพระดับต่ำ เพื่อทำการค้นคืนภาพที่มีลักษณะฟีเจอร์ที่ใกล้เคียงกับภาพที่ต้องการมากที่สุด [Andrew, 2005][M. Flickner, 1995][W. Ma, 1997] การปรับปรุงเทคนิควิธีการเพื่อให้กระบวนการค้นคืนภาพได้ผลลัพธ์ที่ถูกต้องเพิ่มมากขึ้น ด้วยการนำวิธีการมาผสมผสานกันระหว่างคุณลักษณะเพื่อให้สามารถวิเคราะห์ในรูปแบบที่ซับซ้อนได้มากขึ้น เช่นการรวมเทคนิคด้วยคุณลักษณะสีและรูปทรงของภาพเพื่อทำการค้นคืนภาพ [ P.S. Hiremath, 2007] หรือมีการใช้อัลกอริทึมเพื่อทำการสกัดข้อมูลภาพเป็นฟีเจอร์ และนำฟีเจอร์นำมาใช้ในการ ค้นคืนภาพในรูปแบบที่แตกต่างกันออกได้ แต่ในความเป็นจริงแล้วนั้นลักษณะการมองภาพของคนโดยทั่วไปเป็นการมองจากความหมายของภาพ หรือ

<sup>1</sup> [http://alipr.com/cgi-bin/zwang/regionsearch\\_show.cgi](http://alipr.com/cgi-bin/zwang/regionsearch_show.cgi)

มองจากชนิดของวัตถุของภาพ ดังแสดงในภาพที่ 2.2 แสดงภาพชายทะเล ภายในภาพประกอบด้วย วัตถุหลายชนิดได้แก่ ทะเล และ ท้องฟ้า เพียงอย่างเดียว แต่บางภาพอาจจะประกอบด้วยวัตถุอื่น ๆ มนุษย์ สุนัข หรืออาคาร เป็นภาพประเภทเดียวกันหรือภาพที่สื่อความหมายอย่างเดียวกัน โดยที่ไม่จำเป็นต้องมีคุณลักษณะสีหรือรูปร่างแบบเดียวกันก็สามารถเป็นภาพชนิดเดียวกันได้ ดังนั้นในการ คำนึงที่ใช้คุณลักษณะพีเจอร์ระดับต่ำเพียงอย่างเดียว อาจจะได้ผลลัพธ์ที่ไม่ตรงกับความต้องการ อย่างแท้จริง



ภาพที่ 2.2 ตัวอย่างขององค์ประกอบของภาพ<sup>2</sup>

การวิจัยในส่วนแรกเป็นความพยายามในการค้นหาข้อมูลภาพเริ่มมาจากการสกัดข้อมูล ภายในภาพ (feature extraction) ออกมาเป็นส่วนๆแล้วใช้ข้อมูลระดับต่ำเพียงแบบเดียว เพื่อทำการ สืบค้นข้อมูลภาพเท่านั้น วิธีการค้นคืนด้วยข้อมูลระดับต่ำเพียงแบบเดียวแล้วนั้นจะทำให้ได้ผลลัพธ์ ของภาพที่มีความหมายไม่ครบถ้วน ทำให้ต้องมีการปรับปรุงพัฒนาอย่างต่อเนื่อง ด้วยเหตุนี้เองทำให้ มีกลุ่มนักวิจัยที่สนใจในเรื่องของ การจัดกลุ่มภาพตามความหมายเรียกว่า semantic image เป็นการ จัดกลุ่มภาพตามความหมายของภาพโดยแปลความหมายของภาพจากองค์ประกอบ หรือวัตถุ (object) ที่ปรากฏในภาพนั้น ซึ่งมีวิธีการค้นคืนที่แตกต่างกันเพื่อให้ได้ผลลัพธ์ที่ตรงตามความ ต้องการมากที่สุด

แต่อย่างไรก็ตามได้มีงานวิจัยอีกกลุ่ม ที่พยายามจะใช้เทคนิคของการเข้าใจความหมายของ ภาพแทน การสืบค้นแบบข้างต้น เรียกว่า การประมวลผลภาพระดับสูง (high level image processing) งานวิจัยในกลุ่มนี้พยายามที่จะมองข้อมูลบนภาพเป็นวัตถุ (object) ที่มีความหมาย [Benitez A.B,2002] [Galleguillos C.,2010] และแทนวัตถุนั้นๆ ด้วยคำหลัก (keyword) บน

<sup>2</sup> <http://www.corbisimages.com/>

ภาพ เรียกว่า การแท็ก (tag) หรือการให้ความหมายของวัตถุบนภาพเป็น ชื่อวัตถุ หรือคำศัพท์ ที่ สอดคล้องกันเช่น “grass”, “plant”, “boat”, “sky” เป็นต้น ดังแสดงในภาพที่ 2.3 [Galleguillos C., 2010] และใช้ความหมายหรือคำศัพท์นั้นเพื่อทำการสืบค้นข้อมูลแทน ซึ่งเป็นการใช้ความหมาย ของคำศัพท์ที่มีสอดคล้องกันด้วยความหมายตามพจนานุกรม หรือในลักษณะใช้ความสัมพันธ์ของ ความหมายที่เหมือนกันของคำหลัก (synonym) [Zhao T., 2001] [Benitez, A.B., 2002] [Kobus B., 2001] [Philippe M., 2002] เข้ามาใช้ในการค้นคืนข้อมูลภาพเพื่อให้ผลลัพธ์ตรงกับ ความ ต้องการของผู้ใช้ ยกตัวอย่างเช่น “stone” มีความหมายสอดคล้องกันกับ “rock” เป็นต้น จะได้ผลที่ ค่อนข้างดีกว่า แต่ขึ้นอยู่กับว่าอัลกอริทึมที่ถูกนำมาใช้นั้นจะเป็นลักษณะใด

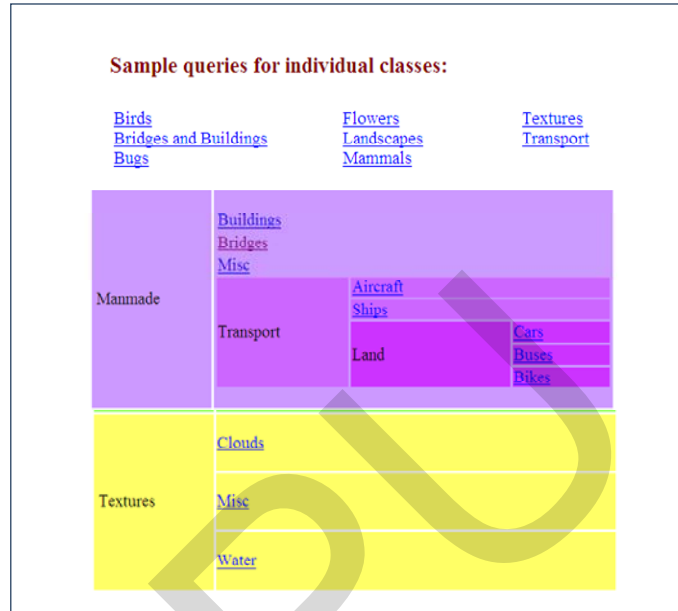


ภาพที่ 2.3 ตัวอย่างการแท็กองค์ประกอบของภาพ<sup>3</sup>

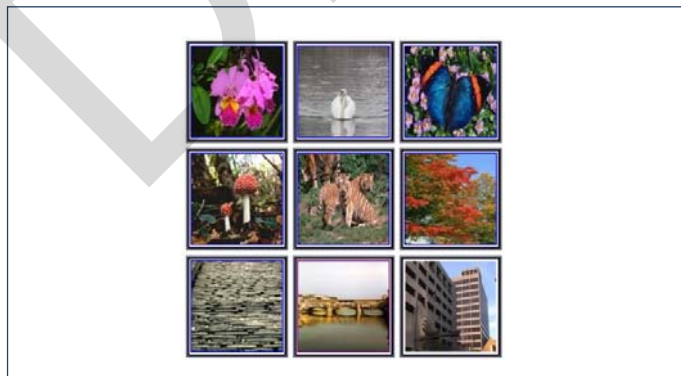
นักวิจัยบางกลุ่มใช้เทคนิคเพื่อสร้างความสัมพันธ์ (relationship) ของแท็กชื่อวัตถุบนภาพ [Benitez A.B, 2001] [R. Zhao, 2002] [Philippe M., 2002] ด้วยโมเดลเหตุการณ์ (event model) [Joo-Hwee L., 2003] เช่น การเชื่อมวัตถุด้วยคำต่างๆ เช่น “touch” “on” “top” เป็นต้น บางงานวิจัยได้พยายามใส่ข้อความ (context) เพื่อบรรยายความหมายของภาพ [Mathias Lux, 2003] [Mathias Lux, 2009] ในหัวข้อที่สอดคล้องกับภาพนั้นๆ เช่น “birthday party of uncle Adam” หรือ “a picture showing a barking dog” แต่ภาพที่นำมาใช้จัดเก็บนั้นมักจะเป็นจะภาพ ส่วนตัว (personal images) ทำให้การค้นคืนจำกัดเพราะคำบรรยายส่วนใหญ่จะเป็นคำเฉพาะเจาะจง จึงไม่นิยมเท่าที่ควร บางงานวิจัยพยายามที่จะใส่ข้อมูลเหตุการณ์ต่างๆจนครบถ้วนในรูปแบบของ ใคร ทำอะไร ที่ไหน เมื่อไหร่ (who, what, when, where) ดังนั้นในการใส่ข้อมูลบางครั้ง เป็นข้อมูลที่ นอกเหนือ หรือเกินความจำเป็นโดยใช้เหตุ ข้อมูลเหล่านี้อาจจะไม่มีความจำเป็นเลยสำหรับการสืบค้น ข้อมูล อาจจะเป็นข้อมูลที่มีความเป็นส่วนตัวจนเกินไป และใช้คำศัพท์ซ้ำซ้อน ฟุ่มเฟือย ทำให้การค้น คคืนยังเกิดความสับสนปนเป กันของภาพผลลัพธ์ที่ได้มา ดังแสดงในภาพที่ 2.4 ก. [Q. Iqbal, 2002] แสดงตัวอย่างการค้นคืนภาพที่มีการจัดเป็นหัวข้อ (topic) จากภาพที่ 2.4 ข. แสดงภาพต้นฉบับเพื่อ

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

สร้างคุณลักษณะหรือฟีเจอร์ที่คล้ายกันเพื่อใช้ในการค้นคืนในที่เก็บข้อมูลภาพ จากภาพที่ 2.4 ค. แสดงภาพต้นฉบับมีคำหลักเป็น “bridges” และกลุ่มผลลัพธ์ที่ได้หลังจากการใช้ฟีเจอร์ในการค้นคืนภาพ แต่ผลลัพธ์ที่ได้ออกมานั้นจะมีการผสมกับภาพอื่น ๆ ที่ไม่ตรงกับคำหลักด้วย



ก. การค้นคืนด้วยการเลือกคำหลักที่เป็นหมวดหมู่



ข. การค้นคืนด้วยการเลือกภาพต้นฉบับ



ค. ผลลัพธ์ของการค้นคืนด้วยภาพต้นฉบับ  
ภาพที่ 2.4 การแก้กองค์ประกอบของภาพ<sup>4</sup>

การค้นหาภาพด้วยเทคนิคนี้จะได้ผลลัพธ์ที่ขึ้นกับคำศัพท์ที่ถูกแก้ไขบนภาพยังมีการแก้ไขข้อมูลบนภาพมากยิ่งขึ้นสามารถหาความเหมือนกันบนภาพมากขึ้นเท่านั้น แต่ในความเป็นจริงแล้ว การแก้ไขข้อมูลบนภาพในปัจจุบันนั้นเป็นเพียงการหาคำศัพท์ที่ต้องการบนภาพ แต่ไม่ได้ให้ความหมายภาพโดยรวม ความหมายของภาพคือการนำวัตถุที่ปรากฏบนภาพมารวมกันเพื่อวิเคราะห์จากความคิดของมนุษย์เพื่อให้ได้ผลลัพธ์คือคำศัพท์ใหม่ที่แทนความหมายของภาพทั้งภาพ สำหรับการวิเคราะห์ความหมายของภาพจะได้จากการรับรู้ของมนุษย์ที่มองภาพนั้น จะมีกฎเกณฑ์ของการรับรู้จากการมองภาพได้กล่าวไว้เป็นทฤษฎี โครงสร้างสเกตตรีตรอน ดังแสดงในภาพที่ 2.5 ก. [Rudolph A., 1974] ได้มีการนำมาประยุกต์ใช้วิเคราะห์ในส่วนของการการจำแนกความหมายของภาพโดยรวมเพื่อให้ได้ความหมายของภาพอย่างแท้จริง โดยใช้ทฤษฎีการมองของมนุษย์ที่มีการพิจารณาจากตำแหน่งและขนาดของวัตถุที่เกิดขึ้นเพื่อนำมาใช้ในการแปลความหมายของภาพ ด้วยการวัดโครงสร้างกับภาพต้นฉบับที่มีการแก้ไข ดังแสดงในภาพที่ 2.5 ข. และ ค. [N. Chinpanthana., 2010] [นิตยสาร 2552] แต่อย่างไรก็ตามการคำตำแหน่งในส่วนของภาพนั้นยังไม่เพียงพอต่อการนำมาใช้สำหรับการแปลความหมายภาพ

<sup>4</sup> [http://amazon.ece.utexas.edu/~qasim/sample\\_queries.htm](http://amazon.ece.utexas.edu/~qasim/sample_queries.htm)



ภาพที่ 2.5 การหาตำแหน่งวัตถุของภาพด้วยโครงสร้างสเกตตริตรอน

ยังมีวิธีการที่ผสมผสานระหว่างการใช้แท็กคำหลักบนภาพกับการสร้างความสัมพันธ์ระหว่างวัตถุด้วยมัลติเคิลคอร์เนล (multiple kernel) ของพื้นที่วัตถุบนภาพ [Galleguillos C., 2011] [Galleguillos C., 2010] ผลลัพธ์ที่ได้จะเป็นกลุ่มของคำหลักที่เกิดขึ้นบนภาพเสียส่วนใหญ่ และมีนักวิจัยบางกลุ่มพยายามที่แก้ไขถึงความยุ่งยากลำบากในการใส่ข้อมูลบนภาพ (image annotation) และจำกัดขอบเขตคำศัพท์ของภาพให้รัดกุมยิ่งขึ้น จึงได้พัฒนาซอฟต์แวร์เพื่อให้เกิดความสะดวกสบายในการใส่ข้อมูล และการใส่ข้อมูลที่มีแบบแผนแน่นอน ยกตัวอย่างเช่น Caliph & Emir [Mathias Lux, 2009], Annosearch [Xin-Jing, 2008], CAMEL [Apostol Paul N., 2001] เป็นต้น แต่อย่างไรซอฟต์แวร์ส่วนใหญ่ นั้น จะทำการสืบค้นภาพในรูปแบบของการเทียบคำหลักเป็นคำต่อคำตามที่มีการเก็บข้อมูลไว้เท่านั้น เพราะฉะนั้นคำตอบที่ได้จะไม่ได้ขึ้นกับความหมายของภาพ แต่ขึ้นกับคำศัพท์ที่มีการเก็บข้อมูลลงไปเท่านั้น ความหมายของภาพยังไม่ได้อย่างแท้จริง เพราะฉะนั้นผลลัพธ์ของกลุ่มภาพที่ได้จะไม่ได้ขึ้นกับความหมายของภาพอย่างแท้จริง แต่ขึ้นกับคำหลักที่มีการเก็บข้อมูลลงไปบนภาพเท่านั้น จึงทำให้ความหมายของภาพโดยรวมนั้นยังไม่ได้สื่อออกมาให้เห็นอย่างเด่นชัด ซึ่งในความเป็นจริงแล้วความหมายของภาพจะเกิดจากการแปลความหมายรวมของทุกวัตถุที่ปรากฏบนภาพ ดังนั้นในงานวิจัยนี้ได้นำเสนอในรูปแบบของการแทนข้อมูลภาพ ด้วยความสัมพันธ์ของข้อมูลวัตถุภายในภาพ หรือเรียกว่า แนวคิดกราฟ (Conceptual Graph) ในลักษณะของกราฟที่นำเสนอ นั้นจะมีแสดงความสัมพันธ์ของข้อมูลภายในภาพของวัตถุที่เกิดขึ้นทั้งหมดในรูปแบบของกราฟรวมทั้งความสัมพันธ์ (Relationship) ระหว่างวัตถุที่เกิดขึ้น และจะมีการวัดค่าความสัมพันธ์ของความเหมือนกันด้วยวิธีการที่เรียกว่า การจับคู่กราฟ (Graph Matching) ของข้อมูลภาพที่ถูกแทนค่าความหมายไว้ ดังนั้นภาพที่มีความหมาย เหมือนกันจะมีได้ค่าของความเหมือนกันมากที่สุด

จากที่กล่าวมาข้างต้นในงานวิจัยนี้จึงได้นำเสนอ การปรับปรุงการสืบค้นข้อมูลภาพด้วยความหมายภาพเพื่อให้ได้ผลลัพธ์ของภาพที่มีอย่างแท้จริง โดยใช้วิธีการแทนข้อมูลภาพแบบแนวคิดกราฟ (Conceptual Graph) และทำเปรียบเทียบความเหมือนกันของความหมายภาพด้วยการหาความเหมือนของภาพทั้งหมด 4 วิธีการ โครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network) แผนผังการจัดระบบตัวเอง (Self-organizing maps) เครือข่ายแบบเบย์ (Bayesian Network) และการวัดความคล้ายของกราฟแบบจับคู่ (Similarity Measure with Conceptual Graph Matching)

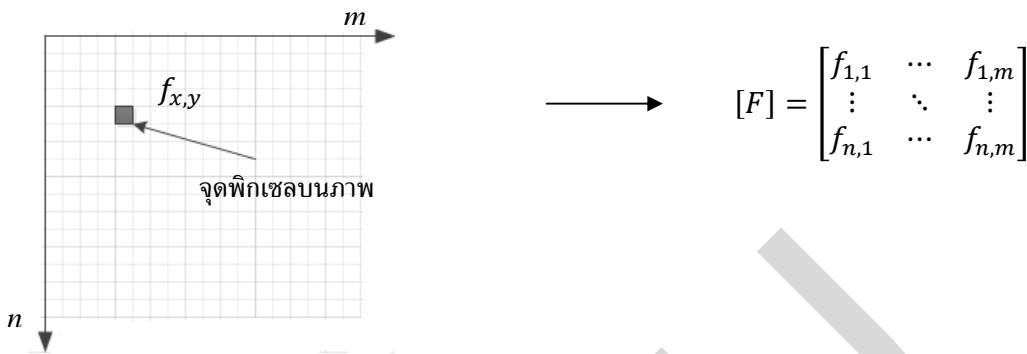
ในงานวิจัยนี้ได้นำทฤษฎีมาใช้ในการพิจารณา ดังนี้

- การประมวลผลภาพดิจิทัล
- การแทนข้อมูลภาพแบบแนวคิดกราฟ
- การเปรียบเทียบความคล้ายของภาพ
- การใช้โมเดลในการสร้างการตัดสินใจ
- การวัดประสิทธิภาพ

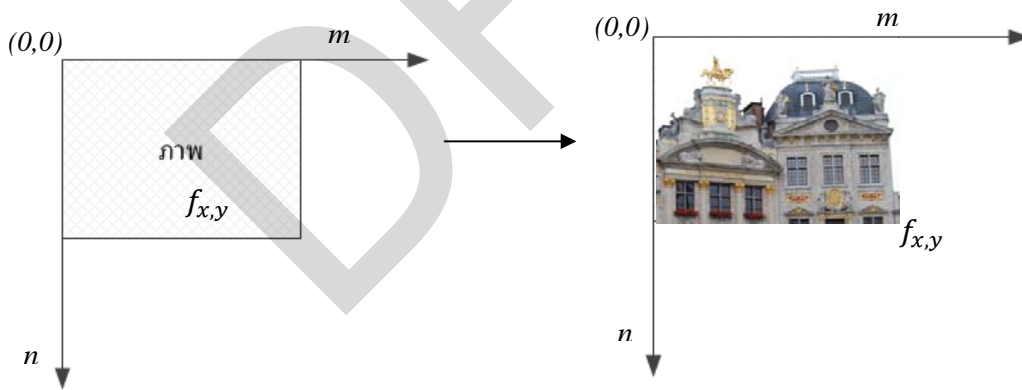
## 2.2 การประมวลผลภาพดิจิทัล

ภาพจะถูกจัดเก็บเป็นสัญญาณดิจิทัล (digital) [นัศพ์ชาณัณ 2553] [R.C. Gonzalez 2002] และทำการจัดเก็บข้อมูลภาพในหน่วยความจำ ในรูปของตัวแปรแบบอะเรย์ (array) โดยค่าในแต่ละช่องของอะเรย์แสดงถึงค่าความเข้มของแสง (intensity) ของแต่ละจุดของแต่ละจุดของภาพที่เรียกว่า พิกเซล (pixel) และตำแหน่งของช่องอะเรย์เป็นตัวกำหนดตำแหน่งของจุดภาพดังแสดงในภาพที่ 2.4 เมื่อนำภาพมาแสดงในรูปของ 2 มิติ ตามแนวแกน x และ y สามารถเขียนในรูปของฟังก์ชันของภาพ สามารถแสดงในรูปของ  $i = f(x, y)$  ในรูปของ 2 มิติ โดย i หมายถึง ความสว่างหรือความเข้มของแสง และ  $(x, y)$  เป็นพิกัดในแบบจำลองของภาพ (image model) ค่าของฟังก์ชัน  $f(x, y)$  ดังกล่าวเป็นค่าความเข้ม ซึ่งเปลี่ยนไปตามตำแหน่งหรือพิกัดแบบจำลองภาพ ค่าความเข้มของแสงนี้เป็นค่าดิจิทัลหรือเชิงตัวเลข ดังนั้น เมื่อแสดงเป็นตัวเลขของค่าความเข้มในแต่ละพิกเซลเรียงกันทั้งภาพจะเห็นว่าเป็นเมทริกซ์ สิ่งทีแบบจำลองต่างจากแบบจำลองทางคณิตศาสตร์ คือ จุดกำเนิด

(Coordinate Origin) ที่ตำแหน่ง (0,0) ซึ่งอยู่ที่จุดซ้ายบนสุดของสเกลต่างจากจุดกำเนิดของแบบจำลองทางคณิตศาสตร์ที่อยู่จุดซ้ายล่างสุด ดังนั้นการนำเอาแบบจำลองทางคณิตศาสตร์มาประยุกต์ในการประมวลผลภาพ จำเป็นต้องเปลี่ยนจุดกำเนิดให้ถูกต้องด้วย ดังแสดงในภาพที่ 2.6



ภาพที่ 2.6 การแสดงตำแหน่งของพิกเซลบนภาพดิจิทัลด้วยเมทริกซ์



ภาพที่ 2.7 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



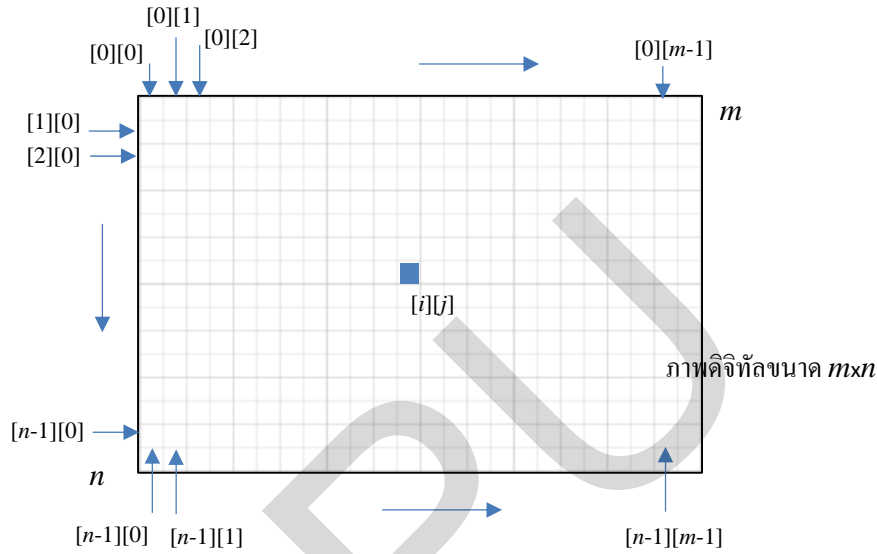
640 × 480, 800 × 600 และ 1024 × 768 จุด เป็นต้น การกำหนดความละเอียดจะขึ้นอยู่กับงานที่จะใช้ ในงานบางอย่างใช้ความละเอียดเพียง 30 × 50 จุด ก็พอซึ่งความละเอียดนั้นจะขึ้นกับงานที่จะใช้ ในบางงานจะใช้ความละเอียดถึง 1000 × 1000 จุด ก็ยังไม่พอ จากภาพที่ 2.7 สมมติให้ภาพแทนเป็นตัวแปรชื่อ  $x$  เป็นตัวแปรแบบอะเรย์ขนาด  $m \times n$  ( $m$  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 ( $\text{mm}/\text{pix}$ ) เป็นต้น ในงานที่ต้องการทราบตำแหน่งหรือขนาดของวัตถุที่วัดเป็นค่าจริง เราสามารถที่จะคำนวณได้จาก

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

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

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

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



ภาพที่ 2.8 การจัดเก็บและเรียกใช้ข้อมูลภาพดิจิทัลลงในเมทริกซ์

### 2.2.1 การประมวลผลเบื้องต้น

สำหรับขั้นตอนการประมวลผลเบื้องต้น (image preprocessing) ในรูปแบบของการหาบริเวณที่ต้องการในรูปแบบอัตโนมัติ ยังคงเป็นงานวิจัยที่ยังหาข้อยุติไม่ได้ โดยเฉพาะที่เป็นลักษณะของระบบเรียลไทม์ (real time) ด้วยแล้วนั้นจะต้องคำนึงถึงเวลาในการคำนวณของอัลกอริทึม (computational cost of algorithm) ที่เป็นสิ่งที่จำเป็นค่อนข้างมาก สำหรับระบบเรียลไทม์ ที่มีการแปรเปลี่ยนรูปแบบของพื้นหลัง (modeling of background) จะทำให้เกิดกระบวนการในการสกัดวัตถุที่ต้องการขึ้นมา ซึ่งเทคนิคที่เรียกว่า Gaussian Mixture Model [C. Stauffer, 2000] เป็นเทคนิคที่ค่อนข้างประสบความสำเร็จมากแต่อย่างไรก็ตามยังคงต้องมีส่วนภาพพื้นหลังที่ไม่มีการเปลี่ยนแปลงมากนัก เพราะฉะนั้นปัญหาที่เกิดขึ้นยังคงเกิดขึ้นเมื่อพื้นหลังมีการแปรเปลี่ยนและยังคงเป็นปัญหาที่ยังคงต้องมีการแก้ไข ฉะนั้นในงานวิจัยนี้จึงข้ามในส่วนการวิเคราะห์การแปรเปลี่ยนของ

พื้นหลังที่ไม่คงที่ 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] เพื่อทำการแก้ไขปัญหาดังกล่าวที่เกิดจากการหาวัตถุที่เกิดจากการทับซ้อนกัน หรือมีปัญหาจากสัญญาณรบกวน หรือวัตถุที่มีการเปลี่ยนแปลงทิศทางจากการหมุน เป็นต้น และการใช้เทคนิค EENC สามารถทำงานได้ดีสำหรับการแมชชีงส่วนของพื้นที่ (matching region) พร้อมทั้งยังเป็นเทมเพลต (template) ที่ใช้ทำงานได้อย่างรวดเร็ว

### 2.2.2 การประมวลผลภาพระดับสูง

การประมวลผลภาพระดับสูง (high-level image processing) เป็นการใช้ข้อมูลพีเจอร์หรือผลลัพธ์จากการประมวลผลข้างต้นเพื่อผ่านกระบวนการ หรืออัลกอริทึมทำให้คอมพิวเตอร์รู้จักและเข้าใจภาพ (image understanding) รวมทั้งการหาความหมายของภาพ (semantic image) ได้ เพราะฉะนั้นการประมวลผลภาพระดับสูงจำเป็นต้องใช้ข้อมูลที่ได้มาจากการประมวลผลภาพระดับต่ำ ดังนั้นจะเห็นได้ว่าการประมวลผลภาพระดับต่ำมีความสำคัญมากสำหรับการทำให้คอมพิวเตอร์รู้จักและเข้าใจภาพได้ โดยส่วนใหญ่กลุ่มนักวิจัยพยายามที่จะข้ามขั้นตอนการประมวลผลภาพระดับต่ำไปเนื่องจากยังไม่มีคุณสมบัติเท่าที่ควรและพยายามที่จะใช้การประมวลผลภาพระดับสูงอย่างเดียวแต่อย่างไรก็ตามการวิจัยทั้งสองกลุ่มนี้ยังคงมีการพัฒนาการวิจัยอย่างต่อเนื่องเพื่อนำผลลัพธ์ หรือพีเจอร์ต่างๆเข้ามาทำการค้นคืนภาพ (image retrieval) หรือ การจำแนกข้อมูลภาพ (image classification) รวมทั้งการแยกแยะความหมายของภาพ (semantic image classification) ซึ่งเป็นหัวข้อหลักในการท้าววิจัยครั้งนี้

## 2.3 การเปรียบเทียบความคล้ายของภาพ

การเปรียบเทียบความคล้ายของภาพ (image similarity measure) เป็นการนำเอาข้อมูลทั้งหมดที่เก็บรวบรวมมาได้ มาผ่านกระบวนการขั้นตอนการแยกแยะข้อมูลลงในแต่ละกลุ่มที่จัดไว้ โดยในแต่ละกลุ่มของข้อมูลนั้นจะมีคุณลักษณะเด่นของแต่ละกลุ่มที่แตกต่างกัน ขึ้นกับข้อมูล หรือ ฟีเจอร์ที่เก็บรวบรวมมาได้รวมทั้งกระบวนการหรือวิธีการที่ใช้จำแนกข้อมูล ซึ่งในงานวิจัยนี้ได้ใช้วิธีการแยกประเภทข้อมูลของภาพเพื่อเปรียบเทียบความสามารถ ทั้งหมด 3 วิธี

- โครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network),
- ทฤษฎีแผนผังการจัดระเบียบตัวเอง (Self-organizing maps)
- ทฤษฎีเครือข่ายแบบเบย์ (Bayesian Network)

### 2.3.1 โครงข่ายประสาทเทียม

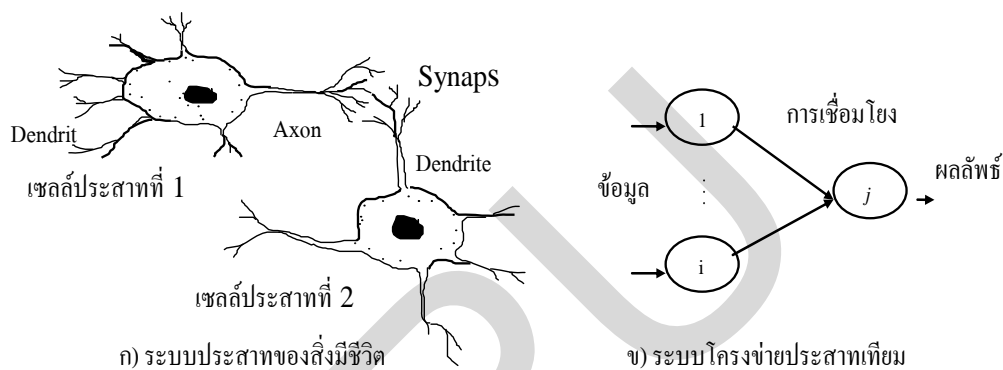
โครงข่ายประสาทเทียม (artificial neural network) [Flood, I.,1994] [Lippmann, R.P., 1987] หรือเรียกว่า ข่ายงานประสาท (neural network หรือ neural net) เป็นโมเดลทางคณิตศาสตร์สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนกชันนิสต์ (connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ ด้วยวัตถุประสงค์ที่จะสร้างเครื่องมือซึ่งมีความสามารถในการเรียนรู้การจดจำแบบรูป (pattern recognition) และการอุปมาความรูป (knowledge deduction) เช่นเดียวกับความสามารถที่มีในสมองมนุษย์ ข่ายงานประสาทเทียมถูกพัฒนามาจากวัตถุประสงค์ดังนี้

- การเลียนแบบการทำงานของเซลล์สมองของมนุษย์
- การหาความสัมพันธ์ของเหตุและผลซึ่งยังหาฟังก์ชันออกมาโดยตรงไม่ได้
- การทำให้คอมพิวเตอร์คิดและตัดสินใจได้ด้วยตัวเอง

แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาข่ายงานไฟฟ้าชีวภาพ (bioelectric network) ในสมอง ประกอบด้วย เซลล์ประสาท หรือ “นิวตรอน” (neurons) และ จุดประสานประสาท (synapses) แต่ละเซลล์ประสาทประกอบด้วยปลายในการรับกระแสประสาท เรียกว่า “เดนไดรท์” (dendrite) เป็นข้อมูลเข้า (input) และปลายในการส่งกระแสประสาทเรียกว่า “แอกซอน” (axon)

เป็น ผลลัพธ์ (output) ของเซลล์ เซลล์เหล่านี้ทำงานด้วยปฏิกิริยาไฟฟ้าเคมี เมื่อมีการกระตุ้นด้วยสิ่งเร้าภายนอกหรือกระตุ้นด้วยเซลล์ด้วยกัน กระแสประสาทจะวิ่งผ่านเดนไดรต์เข้าสู่นิวเคลียสซึ่งจะเป็นตัวตัดสินว่าต้องกระตุ้นเซลล์อื่น ๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอ นิวเคลียสก็จะกระตุ้นเซลล์อื่น ๆ ต่อไปผ่านทางแอกซอน ดังแสดงในภาพที่ 2.9

จากรูป เซลล์ร่างกาย (cell body) คือ ตัวของเซลล์ประสาท โดยสัญญาณไฟฟ้าจะเข้ามาทางเดนไดรต์ซึ่งจะมีลักษณะเป็นแขนงย่อยและมีจำนวนมาก จากนั้น เซลล์ร่างกายจะทำหน้าที่ประมวลผลสัญญาณ และยิงสัญญาณออกไปทางเดนไดรต์ให้ เซลล์ประสาทอื่นต่อไป



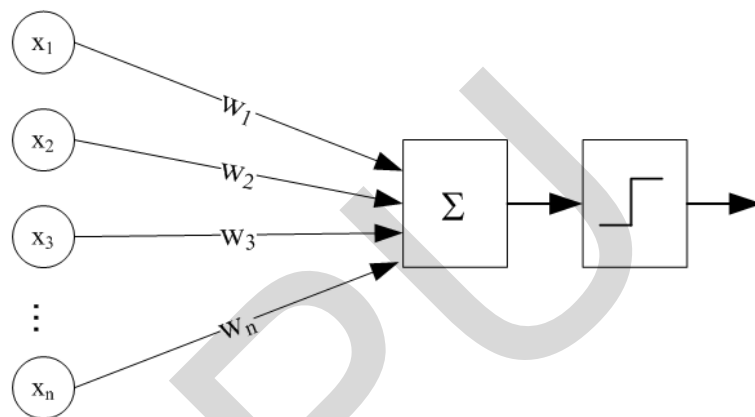
ภาพที่ 2.9 เซลล์ระบบประสาทของสิ่งมีชีวิตและโครงข่ายประสาทเทียม

#### ▪ หลักการทำงานข่ายงานประสาท

สำหรับในคอมพิวเตอร์นิวรอน [Chester M, 1993][ Smith M., 1993] ประกอบด้วย ข้อมูลเข้า และ ข้อมูลออกเหมือนกัน โดยจำลองให้ ข้อมูลเข้า แต่ละอันมีค่าน้ำหนัก (weight) เป็นตัวกำหนดน้ำหนักของ ข้อมูลเข้า โดย นิวรอนแต่ละหน่วยจะมีค่า threshold เป็นตัวกำหนดว่าน้ำหนักรวมของ ข้อมูลเข้าต้องมากขนาดไหนจึงจะสามารถส่งข้อมูลออกไปยัง นิวรอนตัวอื่นได้ เมื่อนำ นิวรอนแต่ละหน่วยมาต่อกันให้ทำงานร่วมกันการทำงานนี้ในทางตรรกจะเหมือนกับปฏิกิริยาเคมีที่เกิดในสมอง เพียงแต่ในคอมพิวเตอร์ทุกอย่างเป็นตัวเลข เพราะฉะนั้นเมื่อนำการทำงานเข้ามาใช้กับคอมพิวเตอร์สามารถเขียนได้ดังนี้

$$\text{if } (\text{sum}(\text{input} * \text{weight}) > \text{threshold}) \text{ then output}$$

เมื่อมี input แทนข้อมูลเข้า เข้ามายังโครงข่าย จะใช้ข้อมูลเข้า คูณกับ weight แทนน้ำหนักของ นิวรอนแต่ละขา ผลที่ได้จากข้อมูลเข้าทุกขาของนิวรอน จะรวมกันแล้วก็เอามาเทียบกับ threshold ที่ กำหนดไว้ ถ้าผลรวมมีค่ามากกว่า threshold แล้วนิวรอนจะส่งข้อมูลออกออกไปข้อมูลออกนี้ก็จะถูก ส่งไปยังข้อมูลเข้า ของนิวรอนอื่นๆ ที่เชื่อมกันในโครงข่ายถ้าค่าน้อยกว่า threshold ก็จะไม่เกิดข้อมูล ออกดังแสดงในภาพที่ 2.10



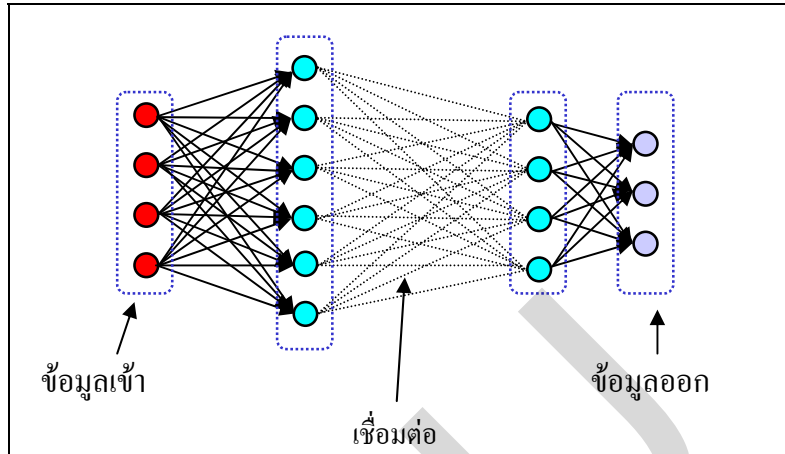
ภาพที่ 2.10 ตัวแบบของนิวรอนในคอมพิวเตอร์

สถาปัตยกรรมโครงข่ายประสาทเทียมที่พบทั่วไปจะมีลักษณะหลัก ๆ คือ มีการจัดเซลล์ ประสาทเทียมเป็นชั้น (layer) ชั้นที่รับข้อมูลเข้าเรียกว่า ชั้นข้อมูลเข้า (input layer) ชั้นที่ผลิตผลตอบ ของโครงข่ายเรียกว่า ชั้นข้อมูลออก (output layer) ส่วนชั้นอื่นๆ ที่มีส่วนในการช่วยทำการ ประมวลผลอยู่ภายในเรียกว่า ชั้นซ่อน (hidden layer) ในโครงข่ายประสาทเทียมอาจมีชั้นซ่อนได้ หลายชั้นโครงสร้างพื้นฐานจะมีลักษณะเป็นการประกอบกันของรูปแบบ ดังต่อไปนี้

### 1. แบบป้อนไปข้างหน้า

แบบป้อนไปข้างหน้า (feedforward network) ข้อมูลที่ประมวลผลในวงจรข่ายจะถูกส่งไป ในทิศทางเดียวจาก โหนดข้อมูลเข้า ส่งต่อมาเรื่อยๆ จนถึง โหนดข้อมูลออก โดยไม่มีการย้อนกลับ ของข้อมูล หรือแม้แต่ว่า โหนดในชั้นเดียวกันก็ไม่มีการเชื่อมต่อกัน สามารถจัดได้เป็น 2 แบบคือ แบบ มีชั้นของเซลล์ประสาทชั้นเดียว และแบบมีชั้นของเซลล์ประสาทหลายชั้น โดยปกติการเชื่อมโยงจะถูก กำหนดขึ้นระหว่างชั้นที่ติดกัน โดยจะมีการเชื่อมโยงระหว่างเซลล์ประสาทเทียมทุกตัว จากชั้นหนึ่งๆ

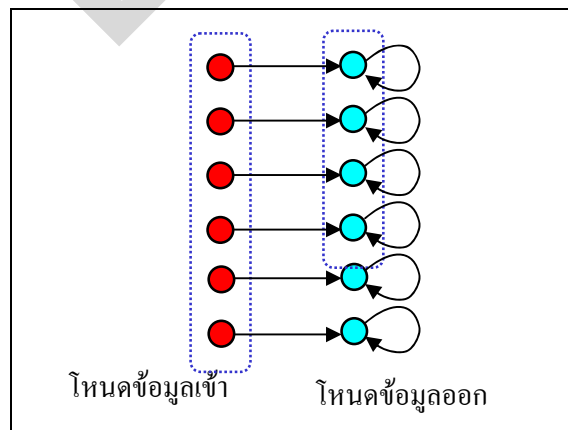
ไปยังเซลล์ประสาทเทียมทุกตัวในชั้นต่อไป ในบางสถาปัตยกรรมอาจมีการเชื่อมโยงข้ามชั้นก็ได้ ดังแสดงในภาพที่ 2.11



ภาพที่ 2.11 สถาปัตยกรรมแบบป้อนไปข้างหน้า

## 2. แบบมีการป้อนไปเวียนกลับ

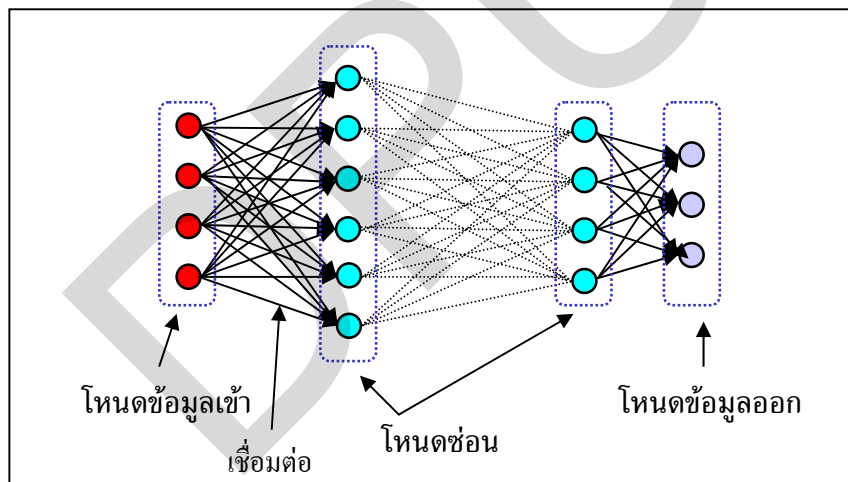
แบบมีการป้อนไปเวียนกลับ (feedback network) บางที่เรียกว่า recurrent network ข้อมูลที่ประมวลผลโครงข่ายประสาทเทียมจะมีการป้อนกลับเข้าไปยังวงจรข่ายหลายๆ ครั้ง จนกระทั่งได้คำตอบออกมา จะมีการเชื่อมโยงที่ถูกกำหนดขึ้นระหว่างเซลล์ประสาทเทียมในชั้นหนึ่งๆ ย้อนกลับไปยังชั้นอื่น ๆ ก่อนหน้านั้น หรือแม้แต่ภายในชั้นเดียวกันเอง ดังแสดงในภาพที่ 2.12



ภาพ 2.12 สถาปัตยกรรมแบบมีการป้อนไปเวียนกลับ

### 3. แบบป้อนไปข้างหน้าหลายชั้น

โครงข่ายประสาทเทียมแบบป้อนไปข้างหน้าหลายชั้น (multiple feedforward) ใช้การเชื่อมโยงแบบป้อนไปข้างหน้าประกอบด้วย 3 ชั้น ได้แก่ ชั้นโหนดข้อมูลเข้า ที่ถูกเชื่อมต่อกับชั้นของโหนดซ่อน และจะมีการเชื่อมต่อกับชั้นของโหนดข้อมูลเข้า เป็นการเพิ่มจำนวนชั้นใน ชั้นซ่อนดังแสดงในภาพที่ 2.11 การทำงานของโหนดข้อมูลเข้าจะทำหน้าที่แทนส่วนของข้อมูลดิบ ที่ถูกป้อนเข้าสู่เครือข่าย และการทำงานของแต่ละ โหนดซ่อน จะถูกกำหนด โดยการทำงานของโหนดข้อมูลเข้าและค่าน้ำหนักบนความสัมพันธ์ระหว่าง โหนดข้อมูลเข้า และ โหนดซ่อน พฤติกรรมการทำงานของ โหนดข้อมูลออก จะขึ้นอยู่กับการทำงานของ โหนดซ่อน และค่าน้ำหนักระหว่าง โหนดซ่อน และโหนดข้อมูลออก ดังแสดงในภาพที่ 2.13



ภาพ 2.13 โครงสร้างแบบป้อนไปข้างหน้าหลายชั้น

ลักษณะการทำงานแบบนี้สามารถกำหนดการแทนค่าให้แก่ โหนดข้อมูลเข้าได้อย่างอิสระ ค่า น้ำหนักระหว่าง โหนดข้อมูลเข้าและ โหนดซ่อน จะถูกกำหนดเมื่อ โหนดซ่อน กำลังทำงาน ฉะนั้นเวลาที่แก้ไขค่าน้ำหนัก โหนดซ่อน จะสามารถเลือกว่าอะไรคือค่าที่เราแทนเข้ามาในโมเดลทางคณิตศาสตร์ของโครงข่ายประสาทเทียม แบบป้อนไปข้างหน้าหลายชั้น แสดงในภาพ มีเวกเตอร์อินพุต  $X = \{x_i\}$  กำหนดให้  $x_i \in \mathfrak{R}^n$  และเวกเตอร์ผลลัพธ์  $Y = \{y_i\}$  กำหนดให้  $y_i \in \mathfrak{R}^m$  ซึ่งประกอบด้วยสมการการคำนวณภายในเซลล์ประสาทแต่ละชั้นดังนี้



ชั้นซ่อน

$$h_j = \sigma(u_j)$$

$$u_j = \sum_{i=1}^p w_{ji}^h x_i$$

กำหนดให้  $h_j$  คือผลลัพธ์ของเซลล์ใน ชั้นซ่อน ซึ่ง  $1 \leq j \leq L$

$\sigma(\cdot)$  คือ activation function

$w_{ji}^h$  คือ ค่าของน้ำหนักบนเส้น  $ji$  เชื่อมเซลล์ในชั้นซ่อน  $j$  กับเซลล์ในชั้นข้อมูลเข้า  $i$

$W^h$  คือ เวกเตอร์น้ำหนัก  $\{w_{ji}^h\}$

ชั้นข้อมูลออก

$$y_k = \sigma(v_k),$$

$$v_k = \sum_{j=1}^m w_{kj}^y h_j$$

โดยที่  $y_k$  คือ ผลลัพธ์ของเซลล์ใน ชั้นข้อมูลออก ซึ่ง  $1 \leq k \leq m$

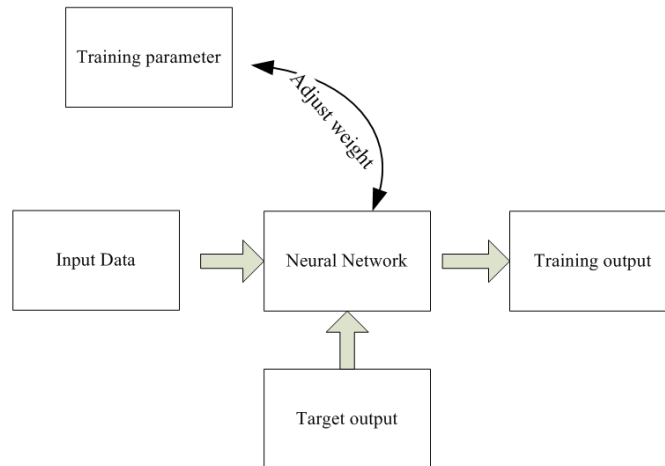
$\sigma(\cdot)$  คือ activation function

$w_{ki}^y$  คือ ค่าของน้ำหนักบนเส้น  $ki$  เชื่อมเซลล์ในชั้นซ่อน  $k$  กับเซลล์ในชั้นข้อมูลเข้า  $j$

$W^y$  คือ เวกเตอร์น้ำหนัก  $\{w_{kj}^y\}$

- รูปแบบการเรียนรู้ของโครงข่ายประสาทเทียม

การเรียนรู้เป็นกระบวนการที่สามารถทำให้เปลี่ยนแปลงพฤติกรรม ความคิด สามารถเรียนรู้ได้จากการ  
ได้ยีนการสัมผัส การอ่าน การใช้เทคโนโลยี สามารถแบ่งออกได้ 2 แบบ



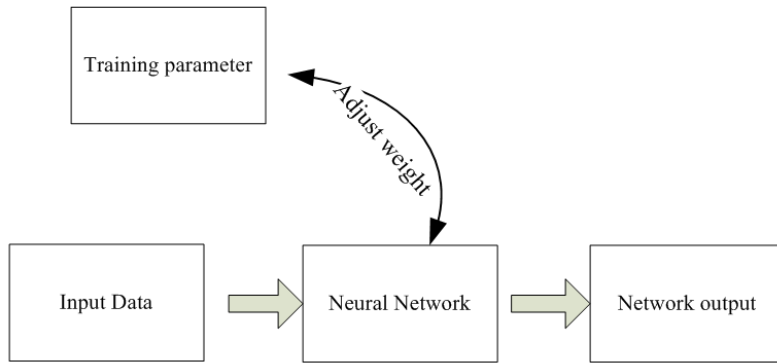
ภาพที่ 2.14 การเรียนรู้แบบมีการสอน

### 1. การเรียนรู้แบบมีการสอน

การเรียนรู้แบบมีการสอน (supervised learning) เป็นการเรียนรู้แบบที่มีการตรวจคำตอบ เพื่อให้วงจรรายปรับตัว ชุดข้อมูลที่ใช้สอนวงจรรายจะมีคำตอบไว้คอยตรวจดูว่าวงจรรายให้คำตอบที่ ถูกหรือไม่ ถ้าตอบไม่ถูก การเรียนรู้โดยให้ข่างานปรับตัวเข้ากับค่าจริง (target) หรือคำตอบที่มีอยู่ เพื่อให้ได้คำตอบที่เหมือนหรือใกล้เคียงกับคำตอบจริงดังแสดงในภาพที่ 2.14

### 2. การเรียนรู้แบบไม่มีผู้สอน

การเรียนรู้แบบไม่มีการสอน (unsupervised learning) เป็นการเรียนแบบไม่มีผู้แนะนำ ไม่มีการตรวจคำตอบว่าถูกหรือผิด จะไม่มีค่าจริงในการฝึกสอน เครือข่ายจะจัดเรียงโครงสร้างของตัวเอง ตามลักษณะของข้อมูล ผลลัพธ์ที่ได้เครือข่ายจะสามารถจัดหมวดหมู่ของข้อมูลได้ เช่น การแยก ลักษณะของคน พืช และสัตว์ เองได้โดยไม่มีผู้สอนจริงในการฝึกสอน เครือข่ายจะจัดเรียงโครงสร้าง ของตัวเองตามลักษณะของข้อมูล ผลลัพธ์ที่ได้เครือข่ายจะสามารถจัดหมวดหมู่ของข้อมูลได้ เช่นการ ที่เราสามารถแยกลักษณะของคน พืช และสัตว์ เองได้โดยไม่มีการสอนดังแสดงในภาพที่ 2.15



ภาพที่ 2.15 การเรียนรู้แบบไม่มีการสอน

- การเรียนรู้แบบแพร่ย้อนกลับ

การเรียนรู้แบบแพร่ย้อนกลับ (back-propagation learning) ถูกนำไปประยุกต์ใช้มากที่สุดในการใช้งานในโครงข่ายประสาทเทียม โดยลักษณะของการเรียนรู้แล้ว การเรียนรู้แบบแพร่ย้อนกลับไม่จำเป็นต้องใช้เฉพาะโครงข่ายประสาทเทียมที่มีสถาปัตยกรรมแบบป้อนไปข้างหน้าหลายชั้นเพียงแบบเดียว แต่การจับคู่ดังกล่าวได้รับความนิยมนำไปใช้งานมากที่สุด การเรียนรู้แบบแพร่ย้อนกลับนี้เป็นการเรียนรู้แบบมีการสอน (supervise learning) ขั้นตอนในการเรียนรู้แบบแพร่ย้อนกลับจะเริ่มด้วยการขนส่งข้อมูลที่จะใช้สอนเข้าไปในเครือข่ายทางชั้นข้อมูลเข้า ชั้นข้อมูลเข้าซึ่งโดยปกติจะไม่มีผลกระทบ จะทำหน้าที่ส่งกระจายข้อมูลนั้นไปยังเซลล์ประสาทเทียมต่างๆในชั้นซ่อน จากนั้นเซลล์ประสาทเทียมทุกตัวในชั้นซ่อน จะทำการประมวลผลตามวิธีการของแบบจำลองเซลล์ประสาทเทียม การประมวลผลเกิดจากชั้นซ่อนชั้นแรกผ่านไปตามชั้นซ่อนต่างๆจนสร้างผลตอบที่ชั้นข้อมูลออก

ผลที่ตอบออกมาจากชั้นเอาต์พุตจะถูกเปรียบเทียบกับค่าเป้าหมายที่กำหนด เพื่อคำนวณค่าความคลาดเคลื่อน (error) จากนั้นปริมาณการปรับแต่งค่าน้ำหนักของแต่ละข้อสัญญาณเข้าของแต่ละเซลล์ประสาทเทียมในชั้นข้อมูลออกก็จะถูกคำนวณขึ้น โดยดูจากความคลาดเคลื่อนและปริมาณการปรับค่าน้ำหนักที่ข้อสัญญาณของแต่ละเซลล์ประสาทเทียมในชั้นซ่อนต่างๆ แต่เนื่องจากไม่สามารถกำหนดค่าเป้าหมายของผลตอบจากเซลล์ประสาทเทียมในชั้นซ่อนจึงถูกมองว่า เป็นการไหลย้อนกลับของความคลาดเคลื่อนของเซลล์ประสาทเทียมต่างๆ ของชั้นข้อมูลออกหลังจากรวบรวมค่าความคลาดเคลื่อนที่ไหลย้อนกลับมาแล้ว หลักการเดียวกับชั้นข้อมูลออก คือ ดูจากผลรวมความคลาดเคลื่อนที่ไหลย้อนกลับมายังเซลล์ประสาทเทียมตัวนั้นๆ ประกอบกับค่าของสัญญาณเข้าที่วิ่งผ่านเข้ามาทางข้อสัญญาณเข้านั้นด้วย

เมื่อค่าน้ำหนักที่ขั้วสัญญาณเข้าของเซลล์ประสาทเทียมทุกตัวถูกปรับแต่งหมดแล้วโครงข่ายประสาทเทียมก็พร้อมจะเรียนรู้ข้อมูลตัวอย่างต่อไป ในทางปฏิบัติแล้ว จะต้องใช้ข้อมูลตัวอย่างจำนวนมาก และจะต้องสอนข้อมูลตัวอย่างเหล่านั้นซ้ำกันหลายรอบ จึงจะสามารถสอนโครงข่ายประสาทเทียมชนิดนี้จนเรียนรู้ได้อย่างถูกต้อง

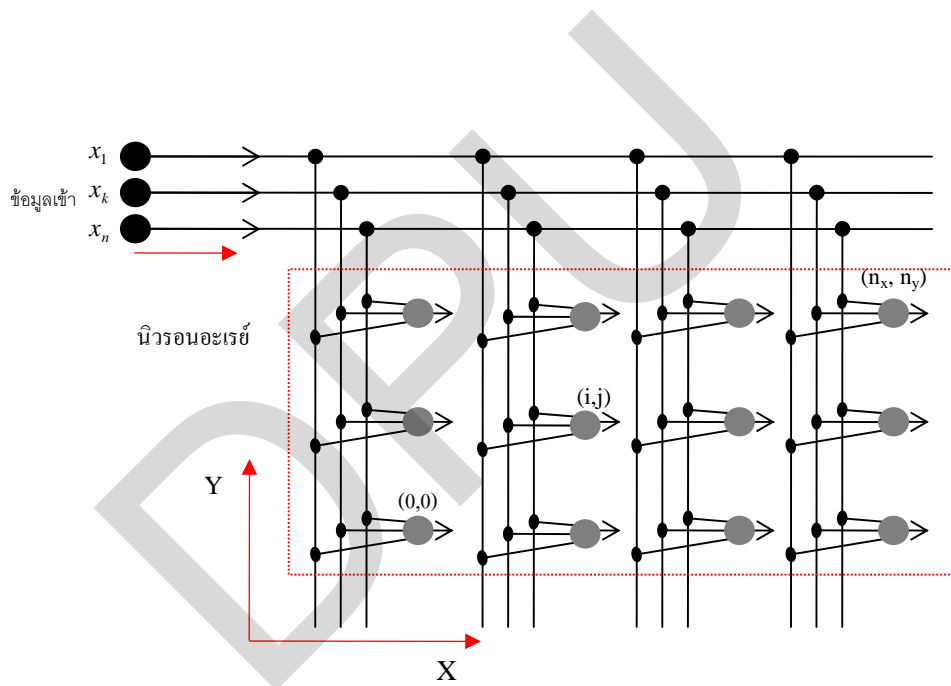
#### ▪ ข้อดีโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม เป็นแขนงหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence: AI) ซึ่งเลียนแบบการทำงานและคุณสมบัติเซลล์สมองหรือระบบประสาทของมนุษย์ เมื่อโครงข่ายประสาทเทียมผนวกกับความสามารถของวิทยาการคอมพิวเตอร์ในปัจจุบัน เช่น หน่วยความจำ การประมวลผลที่รวดเร็ว แม่นยำ และค่าใช้จ่ายที่ไม่สูงนัก ทำให้ได้ระบบที่มีศักยภาพในการทำงาน มีคุณลักษณะและคุณสมบัติที่น่าสนใจ เช่น สามารถจำลองปัญหาได้โดยไม่จำเป็นต้องทราบรูปแบบการกระจายของข้อมูล (distribution free) มีข้อผิดพลาดได้บ้าง (fault tolerance) เรียนรู้ด้วยตนเองได้ (self-organization) มีการทำงานแบบขนาน (parallel process) รวดเร็วระบบทำงานโดยใช้เพียงฟังก์ชันทางคณิตศาสตร์อย่างง่ายแทนที่จะใช้กลไกทางชีวเคมี และไม่ได้ทำงานตามชุดคำสั่งแต่อย่างใด เวทิงเช่นโปรแกรมคอมพิวเตอร์ทั่วไป คำตอบหรือผลลัพธ์น่าเชื่อถือ ด้วยเหตุผลดังกล่าวโครงข่ายประสาทเทียมจึงสามารถแก้ปัญหาใกล้เคียงกับเซลล์สมองหรือระบบประสาทของสิ่งมีชีวิตโดยเฉพาะมนุษย์ ระบบเรียนรู้หรือรู้จำจากตัวอย่างที่มีจำนวนมากและความหลากหลาย แหล่งที่มาของตัวอย่างอาจได้จากการสัมภาษณ์ตรง (direct interview) ข้อมูลในอดีต (historical record) หรือกระบวนการจำลอง (simulation)

#### 2.3.2 ทฤษฎีแผนผังการจัดระเบียบตัวเอง

ทฤษฎีแผนผังการจัดระเบียบตัวเอง (Self-Organizing Maps : SOM) [T. Kohonen, 1995] [T. Mitchell, 1997] เป็นอัลกอริทึมนิเวศเน็ตเวิร์คที่นิยมใช้มากที่สุด โดยแนวคิดของ SOM คือ การทำซ้ำข้อมูลเพื่อหาค่าของน้ำหนักของข้อมูลที่มีอยู่ทั้งหมดตามจำนวนกลุ่มที่ต้องการ เป็นวิธีเกี่ยวกับการจัดกลุ่มด้วยตัวเองโดยใช้โครงสร้างตาข่ายระบบประสาท เป็นการเรียนรู้แบบไม่มีผู้สอน มีลักษณะการเรียนรู้ แบบเครือข่าย 2 ชั้น เป็นการจัดข้อมูลนำเข้าหลายมิติ ให้อยู่ในรูปแบบของข้อมูลสองมิติคือการจัดกลุ่มของข้อมูลที่มีลักษณะคล้ายกันจะอยู่ในโหนดใกล้เคียงกัน โดยโครงสร้างการทำงานของ SOM ชั้นแรกทำหน้าที่นำเข้าข้อมูลและจัดส่งข้อมูลให้แก่นิเวศเน็ตเวิร์กชั้นที่สอง

ทุกโหนดในตาราง หรือ อะเรย์นิวรอน (array of neurons) ระหว่างชั้นจะเชื่อมต่อกันด้วยค่าน้ำหนัก (weight) จากนั้นข้อมูลจะถูกส่งไปยังนิวรอนในชั้นที่สอง เพื่อทำการเปรียบเทียบว่าใกล้เคียงกับค่ากลางกลุ่มใดมากที่สุด แต่ละโหนดในชั้นนี้จะมีความสัมพันธ์กันแบบ เพื่อนบ้าน (neighborhood relation) ทำให้เกิดเป็นรูปแบบ 2 มิติ จากนั้นจะทำการปรับค่า น้ำหนักของตัวที่เป็นผู้ชนะ (winner) ทำให้ข้อมูลที่มีข้อมูลเข้า เข้ามาเกิดการปรับเปลี่ยน และเมื่อผ่านการเรียนรู้ไปหลายๆรอบจะทำให้ ได้กลุ่มข้อมูลออกมาเป็นผลลัพธ์ จากนั้นจึงนำผลลัพธ์ที่ได้นั้นมาผ่านกระบวนการ Visualization เพื่อแสดงผลลัพธ์ที่ได้นั้นออกมาเป็นกราฟชนิดต่างๆต่อไป



ภาพที่ 2.16 โหนดการเรียนรู้ของแผนผังการจัดระเบียบตัวเอง

### ขั้นตอนการทำงานของทฤษฎีแผนผังการจัดระเบียบตัวเอง SOM

1. ทำการจัดกลุ่มของข้อมูลนำเข้าให้อยู่ในกลุ่มข้อมูลที่ใกล้เคียง ทำการ Normalize ค่าข้อมูลแต่ละตัว เพื่อให้ได้ค่ากลางเป็น  $x_i = \frac{x_i - \min}{\max - \min}$ , กำหนดให้  $x_1, x_2, \dots, x_n$  เป็นข้อมูลที่ถูกนำเข้า  $n$  คือจำนวนตัวอย่างข้อมูลนำเข้า  $\min$  คือ ค่าที่น้อยที่สุดของชุดข้อมูล  $\max$  คือ ค่าที่มากที่สุดของชุดข้อมูล

- กำหนดโครงสร้างโครงข่ายประสาทเทียมในลักษณะสองมิติ กำหนด แนวแกน x และแกน y กำหนดชุดของข้อมูล สุ่มค่าน้ำหนักเริ่มต้น ในโครงข่ายดังแสดงในภาพที่ 2.16
- สุ่มค่าเริ่มต้นให้กับค่ากลางของกลุ่ม (cluster center)  $W = \{w_1, w_2, \dots, w_p\}$ , โดยที่  $p$  คือ จำนวนกลุ่ม เวกเตอร์ค่ากลางคือ  $W_j = \{w_{j1}, w_{j2}, \dots, w_{jm}\}$ , เมื่อ  $1 \leq j \leq p$
- คำนวณหาระยะทางระหว่างโครงข่ายโดย ยูคลิเดียน (Euclidean distance) เพื่อหาผู้ชนะ (winner) ซึ่งจะหาได้จากโครงข่ายที่ใช้ระยะทางที่ได้จากการคำนวณ เพื่อหาค่าน้อยที่สุด

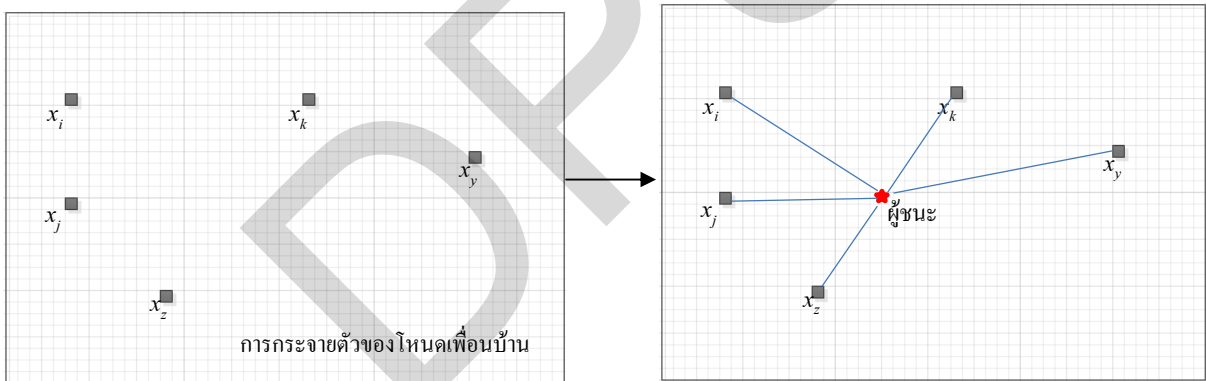
$$C(k_1, k_2) = \min_{i,j} C_{i,j}$$

เมื่อ  $k_1, k_2$  คือดัชนีของโหนดผู้ชนะ

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

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

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



ภาพที่ 2.17 การหาผู้ชนะสำหรับการกระจายตัวของข้อมูล  $x_i$  หรือโหนดเพื่อนบ้าน

- คำนวณค่ากลางสำหรับกลุ่มที่เป็น ค่าผู้ชนะใหม่

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

- ทำซ้ำขั้นตอนที่ 4-5 จนครบชุดข้อมูลนำเข้า หรือจนกว่าค่าน้ำหนักเริ่มจะคงที่
- โหนดที่เป็นผู้ชนะจะสอดคล้องตามสูตร  $C(k_1, k_2) = \min_{i,j} C_{i,j}$  เมื่อ  $k_1, k_2$  ดัชนีของโหนดที่เป็นผู้ชนะ

8. โหนดที่เป็นโหนดเพื่อนบ้านจะถูกกำหนดโดย  $h(\rho, t) = \exp\left(\frac{\rho^2}{2\sigma^2(t)}\right)$
9. ทหาระยะห่างระหว่างโหนดนั้น ๆ กับ โหนดที่เป็นผู้ชนะ  $\rho = \sqrt{(k_1 - i)^2 + (k_2 - j)^2}$
10. หาโหนดเพื่อนบ้านที่ใกล้เคียงที่สุด โดยขอบเขตจะลดลงตามเวลาตามรูปแบบดังนี้

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

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

11. ปรับค่าน้ำหนัก ของแต่ละโหนดด้วย

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

- ข้อเสียของทฤษฎีแผนผังการจัดระเบียบตัวเอง SOM

การเรียนรู้ที่ไม่ทราบค่าที่ต้องการจึงไม่สามารถจะหยุดการเรียนรู้เมื่อไร ต้องมีการปรับค่าน้ำหนักไปเรื่อย ๆ ปัญหานี้สามารถแก้ไขได้ โดยอาศัยเงื่อนไขหรือหลักเกณฑ์บางอย่างในการใช้เป็นเครื่องมือในการหยุดการเรียนรู้

ค่ากลางบางตัวซึ่งเกิดจากการสุ่ม อาจจะถูกในบริเวณที่ห่างจากกลุ่มค่ากลางตัวอื่นมาก ซึ่งไม่สามารถเป็น ตัวชนะได้เลย จึงต้องเพิ่มกระบวนการในการแก้ปัญหา

ไม่สามารถ กำหนดจำนวนชุดข้อมูลที่นำมาจัดกลุ่ม ควรมีที่กลุ่มถึงจะเหมาะสม จึงควรมีการศึกษาวิจัยต่อไป

### 2.3.3 ทฤษฎีความเชื่อเบย์

ทฤษฎีความเชื่อเบย์ (ทฤษฎีความเชื่อเบย์ (bayesian approach) [R. O. Duda, 1973] [T. Mitchell, 1997] เป็นการใช้รูปแบบของการประมาณค่าของความน่าจะเป็นที่จะเกิดเหตุการณ์ 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 จึงสามารถเขียนสูตรของเบย์ได้ดังนี้

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

- วิธีการเรียนรู้เบย์อย่างง่าย

วิธีการเรียนรู้เบย์อย่างง่าย (naïve-bayes learning) [T. Mitchell, 1997] วิธีการเรียนรู้ที่ใช้หลักการของความน่าจะเป็น ซึ่งมีพื้นฐานมาจากทฤษฎีของเบย์ (bayes theorem) เข้ามาช่วยในการเรียนรู้ จุดมุ่งหมายก็เพื่อต้องการสร้างโมเดลที่อยู่ในรูปของความน่าจะเป็น ซึ่งเป็นค่าที่บันทึกได้จากการสังเกต จากนั้นนำโมเดลมาหาว่าสมมติฐานใดถูกต้องที่สุดโดยใช้ความน่าจะเป็นเข้ามาช่วยความรู้ก่อนหน้า หมายถึง ความรู้ที่เรามีเกี่ยวกับสมมติฐานแต่ละตัวก่อนที่จะเก็บข้อมูล เมื่อใช้งานเราจะนำความน่าจะเป็นของข้อมูลที่เก็บได้มาปรับสมมติฐานซ้ำอีกครั้ง เมื่อ C แทนกลุ่มข้อมูลที่จะถูกแบ่งประกอบด้วยกลุ่มที่ถูกสังเกต  $c \in X \langle X_1, X_2, \dots, X_k \rangle$  แทนแอทริบิวต์เวกเตอร์ข้อมูล ประกอบด้วยข้อมูลที่สังเกต  $x \langle x_1, x_2, \dots, x_k \rangle$  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) เข้ามาช่วยในการเรียนรู้ได้ซึ่งพบว่าวิธีนี้ให้ประสิทธิภาพในการเรียนรู้ได้ดีไม่ด้อยกว่าวิธีการเรียนรู้ประเภทอื่น สามารถลดข้อจำกัดอย่างง่ายในสมมุติฐานของความไม่ขึ้นต่อกันระหว่างคุณสมบัต แต่ในความเป็นจริงคุณสมบัต บางตัวจะขึ้นต่อกัน และควรนำค่าความขึ้นต่อกันนี้เข้ามาใส่ไว้ในโมเดล จึงใช้ข่ายงานเบย์ในการอธิบายความไม่ขึ้นต่อกันอย่างมีเงื่อนไข (condition independent) ระหว่างตัวแปร เพื่อให้กระบวนการเรียนรู้มีประสิทธิภาพ โดยสามารถใส่ความรู้ก่อนในข่ายงานความเชื่อเบย์ให้อยู่ในรูปโครงสร้างข่ายงานและตารางความน่าจะเป็นแบบมีเงื่อนไข

## 2.4 การใช้โมเดลในการสร้างการตัดสินใจ

การใช้โมเดลในการสร้างการตัดสินใจ เป็นวิธีการที่จะทำให้ความผิดพลาดเฉลี่ยในการจำแนกเกิดขึ้นน้อยที่สุด ขึ้นอยู่กับความสามารถในการแยกข้อมูลประเภทให้ออกต้อง ซึ่งขึ้นอยู่กับความสามารถในการแยกความทับซ้อนกันของค่าการสะท้อนช่วงคลื่นระหว่างประเภท เพราะฉะนั้นการจำแนกจะรวมหลักการสำคัญหลายอย่างที่ต้องทำความเข้าใจในข้อกำหนด และปรัชญาของ

หลักการเหล่านั้น เพื่อให้การจำแนกได้ผลที่เชื่อถือได้และตรงกันทุกครั้ง (reliable and consistency)

การแบ่งข้อมูลให้เป็นแต่ละประเภท โดยทั่วไปจะใช้ขั้นตอนปฏิบัติ หรืออัลกอริทึมเพื่อทำการจำแนกข้อมูล เป็นเงื่อนไขที่แน่นอนที่กำหนดกระบวนการทางคอมพิวเตอร์ที่ใช้ข้อมูลเบื้องต้นที่หลากหลายไปสู่ผลลัพธ์ที่ต้องการ การใช้พีเจอร์หรือ พีเจอร์ที่ได้มาจากการสกัดข้อมูลภาพเป็นอีกสิ่งหนึ่งที่ผลอย่างมากต่อผลที่นำมาใช้ในการตัดสินใจ ซึ่งในการทำให้ค่าที่ได้เปลี่ยนแปลงจะทำให้การตัดสินใจเปลี่ยนแปลงไปและจะมีผลต่อความแม่นยำในการจำแนกข้อมูล

ในการคาดคะเนค่าพารามิเตอร์โดยทั่วไปจะใช้ข้อมูลตัวอย่างที่เลือกสำหรับแต่ละวัตถุประสงค์ หรือเรียกว่า ข้อมูลฝึกหัด (training data) ข้อมูลตัวอย่างดังกล่าวจะต้องเป็นตัวแทนของประเภทหรือกลุ่มนั้น ๆ (class) ในขั้นแรกของการจำแนกข้อมูลใดๆ จะเป็นการฝึกให้โปรแกรมคอมพิวเตอร์จดจำลักษณะเฉพาะของประเภทที่สนใจ ในลักษณะนี้ของการจำแนกจะเป็นจุดวิกฤตของความสำเร็จในกระบวนการจำแนกโดยรวม และกินเวลาส่วนใหญ่ของผู้วิเคราะห์ วิธีการฝึกคอมพิวเตอร์ จำเป็นต้องให้กลุ่มตัวอย่างแก่เครื่องคอมพิวเตอร์ที่จะนำไปพัฒนา สัญลักษณ์ประเภท (class signature) สัญลักษณ์ประเภทนี้หมายถึง ลักษณะทางสถิติของกลุ่มข้อมูลตัวอย่าง โดยทั่วไปสามารถพัฒนากลุ่มสัญลักษณ์ของแต่ละประเภทได้ด้วยเทคนิคการจำแนก 2 แบบ

- เทคนิคการจำแนกแบบไม่มีการสอน

เทคนิคการจำแนกแบบไม่มีการสอน (unsupervised classification) การจำแนกแบบไม่มีการสอน เป็นทางหนึ่งในการสร้างกลุ่มที่มีความคล้ายคลึงกันที่อยู่ในพื้นที่เดียวกัน โดยผู้ใช้ไม่ต้องรู้ถึงรายการที่จะถูกจำแนกมาก่อน ค่าการสะท้อนช่วงคลื่นทั้งหมดของทุกจุดจะถูกตรวจสอบเพื่อหาความสัมพันธ์ทางสถิติ และรวมกลุ่มกันตามความคล้ายคลึงกันของกลุ่ม หลังจากนั้นจึงเป็นหน้าที่ของผู้แปล หรือวิเคราะห์ข้อมูลที่จะให้คำจำกัดความของแต่ละกลุ่มข้อมูลที่ถูกจำแนกไว้ ซึ่งการจำแนกแบบที่ไม่มีผู้สอนบางครั้งอาจจะถูกเรียกว่า การรวมกลุ่ม (clustering) ข้อเสียของการจำแนกข้อมูลในลักษณะนี้คือ ผลลัพธ์ที่ได้อาจจะไม่ได้ตรงกับที่ผู้ใช้ต้องการ หรือคาดหวังในหัวข้อที่กำหนดไว้ บางรายการที่ถูกจำแนกออกมานั้นอาจจะแปลผลได้ยาก เนื่องจากจำนวนกลุ่มที่ใช้ในการจำแนกเป็นสิ่งที่ไม่รู้มาก่อนล่วงหน้า

- เทคนิคการจำแนกแบบมีการสอน

เทคนิคการจำแนกแบบมีการสอน (supervised classification) ผู้วิเคราะห์จะต้องมีความรู้เกี่ยวกับข้อมูลที่จะทำการจำแนกอยู่ก่อนแล้ว และจะต้องมีการหาความสัมพันธ์ระหว่างกลุ่มที่จำแนกและพารามิเตอร์ไว้ล่วงหน้า และจะต้องมีการใช้ข้อมูลฝึกหัด (training data) เพื่อเป็นกลุ่มตัวอย่างที่จะมีการนำค่าที่ถูกคำนวณมาใช้ในการประเมินต่อไป เพราะฉะนั้นในการเลือกข้อมูลฝึกหัด อาจใช้วิธีการสุ่มแบบธรรมดา การแบ่งสุ่ม (stratified random) หรือ การสุ่มเป็นระบบ (systematic random) วัตถุประสงค์เพื่อให้ได้ตัวอย่างข้อมูลฝึกหัดกระจายทั่วทุกกรณี ในแต่ละประเภทควรจะใช้ตัวอย่างประมาณ 30-40 ตัวอย่าง เพื่อให้สามารถวิเคราะห์ข้อมูลทางสถิติได้

## 2.5 การวัดประสิทธิภาพ

การวัดประสิทธิภาพ (evaluation) เป็นขั้นตอนสุดท้าย เพื่อทำการตรวจสอบวิธีการที่ทำการทดลองมาข้างต้นว่ามีประสิทธิภาพมากหรือน้อยเพียงใดเมื่อนำมาใช้งานจริง จะเป็นขั้นตอนที่สำคัญ เพราะการนำวิธีการที่นำเสนอไปข้างต้นมาใช้งานได้นั้นจะต้องสอดคล้องกับความต้องการ จึงต้องมีการทดสอบศักยภาพการนำไปใช้ สถาปัตยกรรมที่ใช้วัดความสำเร็จหลังการนำไปใช้หากนำไปใช้แล้วไม่ประสบผลสำเร็จต้องย้อนกลับไปเริ่มกระบวนการแรกใหม่ 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

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

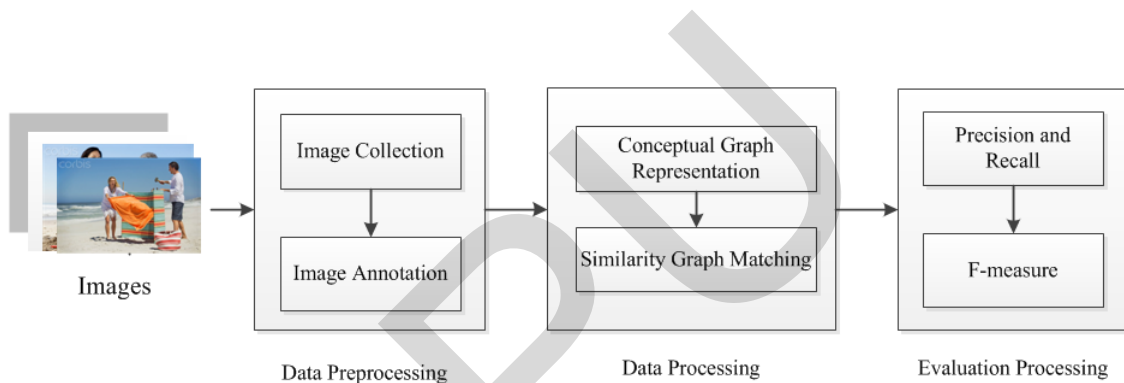
- ค่า F-measure เป็นการวัดค่าความสัมพันธ์ระหว่างค่าความระลึกและค่าความแม่นยำในเชิงฮาร์โมนิก (harmonic) เหมาะสำหรับฐานข้อมูลสารสนเทศที่มีขนาดใหญ่มาก และมักจะไม่ทราบว่ามีข้อมูลภาพที่ถูกต้องทั้งหมดมีอยู่เท่าใด ทำให้ต้องทำการประมาณโดยใช้การสุ่มตัวอย่าง (sampling) ตามหลักทางสถิติหรือด้วยวิธีอื่นด้วย โดยทั่วไปจะเป็นการหาค่า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

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

## บทที่ 3

### ขั้นตอนวิธีการดำเนินงานวิจัย

ในการศึกษางานวิจัยในครั้งนี้ได้มีการเก็บรวบรวมข้อมูลภาพและคัดเลือกภาพที่เหมาะสมเพื่อเตรียมเป็นข้อมูลภาพเบื้องต้น ดังนั้นข้อมูลภาพที่เตรียมพร้อมจะสามารถเข้าสู่กระบวนการประมวลผลภาพ โดยในงานวิจัยจะมีการแบ่งขั้นตอนวิธีการดำเนินงานวิจัยออกเป็น 3 ส่วนหลักดังนี้ ขั้นตอนการเตรียมข้อมูล (data preprocessing) ขั้นตอนการประมวลผล (data processing) และ ขั้นตอนการวัดประสิทธิภาพ (evaluation) โดยขั้นตอนทั้งหมดถูกแสดงไว้ในภาพที่ 3.1



ภาพที่ 3.1 ขั้นตอนการจำแนกกลุ่มความหมายภาพ

#### 3.1 ขั้นตอนการเตรียมข้อมูล

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

### 3.1.1 ฐานข้อมูลรูปภาพ

แหล่งข้อมูลภาพมีหลายแหล่งข้อมูลที่ได้รับการยอมรับและสามารถนำมาใช้เป็นฐานข้อมูลรูปภาพได้ เช่น Fotosearch stock<sup>2</sup> The Corbis Stock<sup>3</sup> The Corel Corporation<sup>4</sup> เป็นต้น เป็นแหล่งข้อมูลภาพที่หลากหลาย อาจจะมีภาพที่ไม่เหมาะสมกับการทดลองที่นำเสนอ เนื่องจาก ภาพบางภาพมีลักษณะผิดปกติ (outlier) หรือ คุณลักษณะวัตถุ (object characteristic) ไม่ชัดเจนหรือ มีขนาดวัตถุขนาดเล็กเกินไปไม่สามารถ บ่งชี้วัตถุได้ ภาพถ่ายระยะใกล้ (close up) ภาพบางภาพอาจจะไม่สามารถแปลความหมาย หรือภาพมีความหมายกำกวมจนทำให้ไม่สามารถหาความหมายภาพได้ ทำให้ต้องมีการคัดเลือกภาพออกไป ไม่นำมาใช้ในการทดลอง ดังนั้นในการหาแหล่งข้อมูลของการนำภาพเข้ามาใช้จึงจำเป็นต้องสมบูรณ์ที่สุด ภาพจะต้องความเหมาะสมกับงานที่จะนำมาใช้เพื่อตอบสนองกับความต้องการของการทดลองมากที่สุด



ภาพที่ 3.2 ภาพตัวอย่างที่ถูกแท็กด้วยคำหลัก<sup>5</sup>

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)

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

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

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

<sup>5</sup> รูปภาพ อ้างอิง <http://www.corbisimages.com/> ค้นคืนเมื่อวันที่ 9 กันยายน 2554

เพื่อใช้สำหรับการสืบค้นข้อมูล ดังนั้นบางแหล่งข้อมูลจะทำการแบ่งหมวดหมู่วัตถุ (object categories) เข้ามาช่วยในการจัดกลุ่มของคำที่จะนำมาทดแทนส่วน บริเวณนั้น ๆ [Everingham M., 2006] [Winn J., 2005] ซึ่งคำที่จะใช้จะมาจากพจนานุกรม (dictionary) จากตารางที่ 3.1 [A. Hanbury, 2007] แสดงตัวอย่างงานวิจัยที่ใช้ฐานข้อมูลคำหลัก ในการทดลองสำหรับการประมวลผลแบบการให้ความหมายภาพ (annotation)

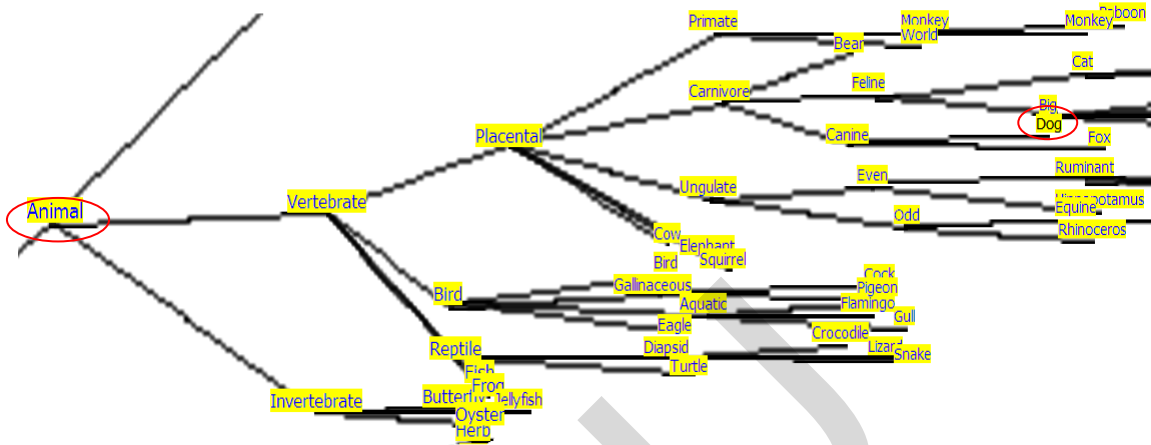
แหล่งข้อมูลภาพ	จำนวนคำหลัก
EU LAVA Project [Perronnin F., 2006]	10
Chen and Wang [Chen Y., 2004]	20
Microsoft Research Cambridge Database [Winn J., 2005]	35
Carbonetto et al., [Carbonetto P., 2004]	55
PASCAL VOC Challenge 2005 Database [Everingham M., 2006]	101
Fei Fei et al., [L. Fei-Fei, 2004]	101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

ตารางที่ 3.1 ตัวอย่างงานวิจัยที่ใช้ฐานข้อมูลคำหลักในการให้ความหมายภาพ

WordNet [Miller G.A., 1990] [Zinger S., 2005] เป็นงานวิจัยที่ได้รับความนิยมมากมีนักวิจัยหลายกลุ่มที่อ้างอิงการใช้ฐานข้อมูลของคำหลักที่แท้กบนภาพ [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”

<sup>6</sup> <http://www.cs.washington.edu/research/imagedatabase/> ค้นคืนเมื่อวันที่ 9 กันยายน 2554

เพื่อให้ส่วนอื่นไปเรียกใช้ โดย WordNet จะถูกใช้คู่กับโปรแกรมเพื่อทำการสืบค้นข้อมูล และยังสามารถกำหนดความสัมพันธ์ของคำด้วยโครงสร้างของ WordNet เองและสามารถสร้างความหมายเหมือนกัน (synonymy) ได้เช่นกัน เป็นข้อดีของการนำ WordNet เข้ามาใช้ดังจะกล่าวในหัวข้อถัดไป



ภาพที่ 3.3 ส่วนย่อยของโครงสร้างลำดับชั้นของคำหลัก“dog” บน WordNet<sup>7</sup>

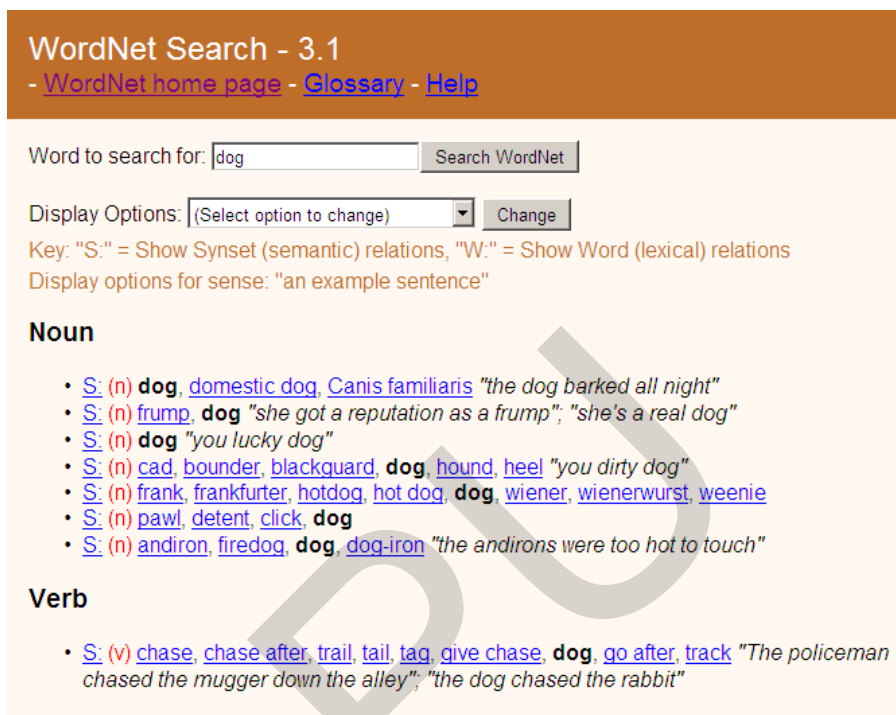
### 3.1.2 การแท็กคำหลักบนภาพ

สิ่งที่สำคัญถัดจากการเลือกใช้ฐานข้อมูลสำหรับกระบวนการของการประมวลผลภาพระดับสูง (high-level image processing) คือ การให้คำอธิบายภาพ (annotated images) โดยใช้ฐานข้อมูลคำหลักที่คัดเลือกมา การแท็กคำหลักบนภาพ โดยทั่วไป ภาพ (image) จะประกอบด้วย พื้นหลัง (background) และ พื้นหน้า (foreground) กล่าวคือภาพหนึ่งภาพจะประกอบด้วยวัตถุ (objects) หลายวัตถุ ทำให้ทุกวัตถุควรจะถูกแท็กด้วยคำหลักที่เหมาะสมจากฐานข้อมูล ดังนั้นในการแบ่งแยกวัตถุ (segmentation) เป็นหัวข้อที่มีการวิจัยอย่างต่อเนื่องในการประมวลผลภาพระดับต่ำ (low-level image processing) [Qian Huang, 1995][Vailaya A., 2001] ด้วยคุณลักษณะของภาพหลายคุณลักษณะ (features) เพื่อทำการแบ่งแยกให้ได้วัตถุที่สมบูรณ์แบบ แต่อย่างไรก็ตามการแบ่งแยกวัตถุที่มีรูปร่างลักษณะคล้ายคลึงกันหรือแตกต่างกันแต่มีความหมายเดียวกัน ยังคงเป็นเรื่องที่ค่อนข้างยากสำหรับการประมวลผลภาพระดับต่ำรวมถึงกระบวนการรู้จำรูปแบบวัตถุ (pattern recognition) เพื่อบอกความหมายของวัตถุหรือชื่อของวัตถุ แต่อย่างไรก็ตามในการวิเคราะห์ความหมายของภาพในงานวิจัยนี้ได้เฉพาะเจาะจงในส่วน

<sup>7</sup> <http://people.csail.mit.edu/torralba/research/LabelMe/wordnet/test.html> ค้นคืนเมื่อวันที่ 9 กันยายน 2554



การประมวลผลระดับสูง จึงได้ข้ามในส่วนของการแบ่งแยกวัตถุ และ การรู้จำรูปแบบวัตถุ ดังนั้นในส่วนที่กล่าวต่อไปสำหรับการประมวลผลภาพระดับสูงคือการให้คำอธิบายภาพด้วยการแท็ก (tag) [Ismail Haritaoglu, 1998] [Tele Tan, 2002] [Vasileios Mezaris, 2003] [R. Zhao, 2002] ข้อมูลภาพให้อยู่ในรูปของชื่อวัตถุ หรือคำหลักจากฐานข้อมูล



ภาพที่ 3.4 ความสัมพันธ์ของโครงสร้างคำหลักบน WordNet<sup>8</sup>

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) ข้อดีของโปรแกรมที่ถูกติดตั้งลงบนเครื่องคอมพิวเตอร์ส่วนตัวคือสามารถเพิ่มขีดความสามารถของการประมวลผลในส่วน of โปรแกรมได้มากกว่าวิธีอื่น ๆ ดังนั้นการโปรแกรมจะมีความสามารถได้มากกว่าวิธีการอื่น ๆ มีความยืดหยุ่นกว่า แต่อย่างไรก็ตามข้อจำกัดของวิธีการนี้ คือ เนื่องจากการให้ความหมายบนวัตถุนั้นสามารถทำได้ด้วยตนเองทำให้ ฐานข้อมูลภาพถูกจำกัด ขอบเขตของคำหลักถูกจำกัด หรืออาจไม่ครอบคลุมมาก

<sup>8</sup> WordNet: <http://wordnetweb.princeton.edu/> ค้นคืนเมื่อวันที่ 9 กันยายน 2554

พอการให้ความหมายภาพอาจจะถูกให้ความหมายด้วยตนเองหรือเพียงกลุ่มบุคคลใดเพียงกลุ่มเดียว 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] [flickr] รูปแบบนี้จะสามารถให้ความหมายวัตถุบนภาพได้ทางออนไลน์ผ่านทางเว็บไซต์ เป็นข้อดีสำหรับรูปแบบนี้ เพราะทุกคนสามารถเข้าถึงได้อย่างง่าย ข้อมูลภาพและกลุ่มข้อมูลคำหลักจะถูกจัดเก็บลงบนเซิร์ฟเวอร์ เดียวกัน ฐานข้อมูลภาพหลากหลายและกลุ่มคนที่ให้ความหมายมีหลายกลุ่ม แต่อย่างไรก็ตามการที่มีฐานข้อมูล และกลุ่มคำศัพท์ที่เปิดกว้างมากเกินไปนี้อาจจะเป็นข้อเสียสำหรับการทดลองที่มีขีดจำกัดเช่นเดียวกัน ส่วนกลุ่มบุคคลที่เข้ามา จะไม่สามารถควบคุมได้ การให้ความหมายภาพให้เป็นไปตามที่กำหนดของส่วนโปรแกรมที่ทำการทดลอง ตัวอย่างโปรแกรม เช่น Flickr [flickr] LabelMe [Russell, B.C.,2008] IBM EVA [ Volkmer, T.,2005] ดังแสดงในตารางที่ 3.1

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

ตารางที่ 3.2 แหล่งฐานข้อมูลของคำหลัก

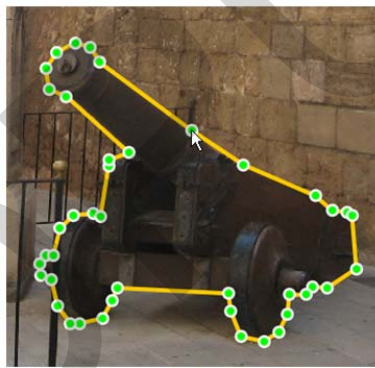


ภาพที่ 3.5 โปรแกรม LabelMe บนเบราว์เซอร์<sup>9</sup>

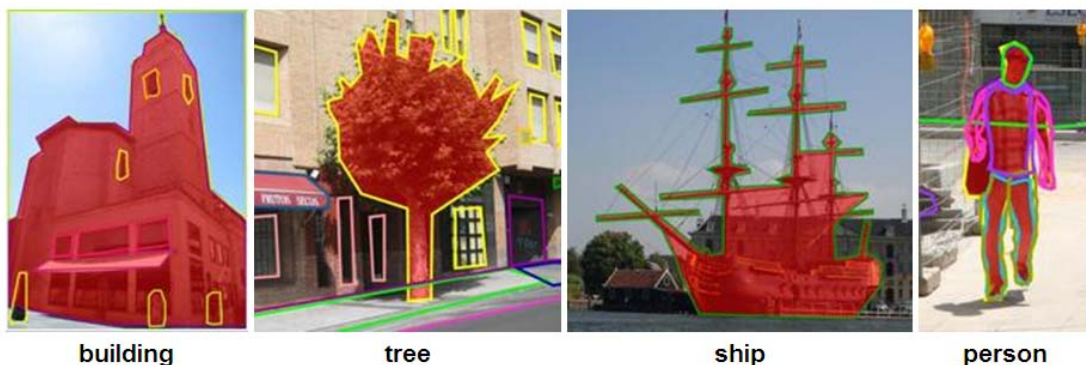
3. แบบออนไลน์เกมส์ (online game) [von Ahn,2004][ von Ahn, L.,2006] เป็นการให้ความหมายวัตถุจะอยู่ในรูปแบบของการเล่นเกมผ่านทางออนไลน์บนเครือข่ายของอินเทอร์เน็ต คุณลักษณะพื้นฐานจะคล้ายกับแบบที่สอง แต่การให้ความหมายในลักษณะนี้อยู่บนพื้นฐานของการเล่นเกม ซึ่งต้องอาศัยความเร็ว ควบคู่กับจำนวนคำหลักของการแท็กบนวัตถุ ดังนั้นการให้ความหมายในรูปแบบค่อนข้างมีข้อจำกัด และ ขึ้นกับความสามารถของผู้เล่นเกมเป็นส่วนใหญ่ ไม่ใช่ความถูกต้อง ดังนั้นอาจทำให้การแท็กวัตถุผิดพลาดหรือไม่ได้ไตร่ตรองให้ความหมายที่ถูกต้องอย่างแท้จริง แต่อย่างไรก็ตามรูปแบบออนไลน์เกมส์ยังเป็นเครื่องมือที่นิยมกันในปัจจุบัน ยกตัวอย่างเช่น ESP game [Von Ahn, 2004] Peekaboom [ Von Ahn, 2006] Squigl [Squigl] ยังมีหลายหน่วยงานที่สร้างโปรแกรมเพื่อมารองรับการทำงานในรูปแบบทั้ง 3 ดังแสดงในตารางที่ 3.2

<sup>9</sup> LabelMe: <http://labelme.csail.mit.edu/> ค้นคืนเมื่อวันที่ 9 กันยายน 2554

จากเครื่องมือให้ความหมายภาพ (image annotation tool) ในตารางที่ 3.2 ที่แบ่งเป็น 3 รูปแบบ แต่ละรูปแบบมีความสามารถทำงานได้แตกต่างกันมีทั้งข้อดีและข้อจำกัด ดังนั้นในงานวิจัยนี้ได้เลือกเครื่องมือแบบออนไลน์ โดยใช้โปรแกรม LabelMe<sup>9</sup> [B. C. Russell,2008][ A. Torralba,2010] เป็นเครื่องมือที่ได้รับการยอมรับอย่างกว้างขวาง สำหรับงานวิจัยทางด้าน Computer Vision โดยแอปพลิเคชันนี้สามารถทำงานได้อย่างเต็มรูปแบบบนเว็บในลักษณะของเครื่องมือให้ความหมาย (Web-based annotation tools) เริ่มตั้งแต่ปี 2005 ปัจจุบันมีวัตถุบนภาพที่ถูกให้ความหมายรวมทั้งสิ้น 400,000 วัตถุ [Von Ahn and L. Dabbish, 2004.] [B. C. Russell, 2008] [A. Sorokin and D. Forsyth, 2008] [M. Spain and P. Perona, 2007] [D. G. Stork, 1999] ผู้ใช้สามารถเข้าถึงโปรแกรมผ่านทางเครือข่ายออนไลน์ได้ สามารถแท็กวัตถุผ่านทางออนไลน์ได้ดังแสดงในภาพที่ 3.5 แสดงหน้าเว็บไซต์ของโปรแกรม LabelMe โปรแกรมสามารถทำงานร่วมกันได้หลายแพลตฟอร์ม ซึ่งเป็นวัตถุประสงค์หลักของคณะผู้จัดทำโปรแกรมนี้ ทำให้ผู้ใช้งานที่เข้ามาให้ความหมายภาพมาได้มากมาย และมีพื้นฐานของการให้ความหมายที่แตกต่างกันตามความสามารถ ของแต่ละบุคคล



cannon

ภาพที่ 3.6 การเลือกสัดส่วนของวัตถุบนภาพ<sup>10</sup>

building

tree

ship

person

ภาพที่ 3.7 ตัวอย่างวัตถุที่ถูกแท็กด้วยโปรแกรม LabelMe<sup>10</sup>

<sup>10</sup> LabelMe: <http://labelme.csail.mit.edu/> ค้นคืนเมื่อวันที่ 9 กันยายน 2554

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



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

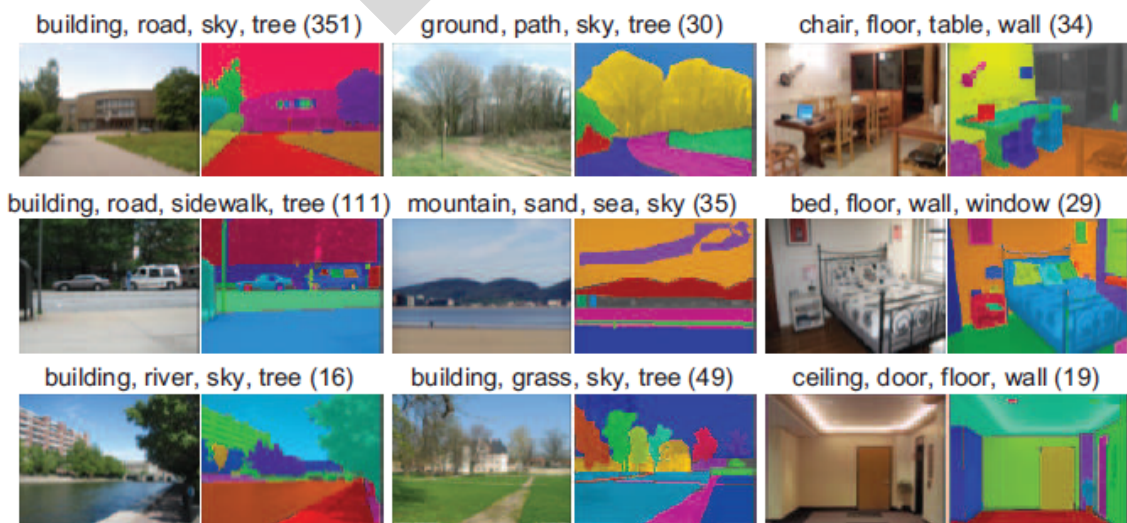
ภาพที่ 3.8 ตัวอย่างภาพที่ถูกแท็กคำหลักบนภาพ ด้วยโปรแกรม LabelMe<sup>11</sup>

<sup>11</sup> LabelMe: <http://labelme.csail.mit.edu/> ค้นคืนเมื่อวันที่ 9 กันยายน 2554

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



ภาพที่ 3.9 ตัวอย่างของคำหลักที่พบบ่อยในการให้ความหมายวัตถุ



ภาพที่ 3.10 ตัวอย่างของคำหลักในการให้ความหมายวัตถุนภาพ

เพื่อเพิ่มประสิทธิภาพของคำหลักที่มีอยู่จำนวนมากนั้นโปรแกรม LableMe ได้นำความสามารถของ WordNet ที่มีโครงสร้างคล้ายต้นไม้ในการเชื่อมโยงความสัมพันธ์ของคำหลักเป็นการจัดกลุ่มของคำตามกิ่งก้านสาขามาช่วยทำให้ลดความซ้ำซ้อนของคำที่ถูกแท็กลงบนภาพเพื่อจัดข้อมูลให้คงที่เพื่อพร้อมสำหรับการประมวลผลต่อไป [B. C. Russell,2008][ A. Torralba,2010]

## 3.2 ขั้นตอนการประมวลผล

หลังจากผ่านขั้นตอนของการเตรียมข้อมูลภาพ ฐานข้อมูลคำหลักและฐานข้อมูลภาพได้ถูกคัดเลือก ดังนั้นข้อมูลทั้งหมดที่ได้จะมีการนำภาพเข้ามาเพื่อทำการประมวลผลตามทฤษฎีที่นำเสนอ โดยแบ่งออกเป็น 2 ส่วนดังนี้

### 3.2.1 การแทนข้อมูลภาพด้วยแนวคิดกราฟ

ข้อมูลรูปภาพที่ได้มาจะถูกแทนที่ด้วยแนวคิดกราฟ (Conceptual Graph Representation) กำหนดให้รูปภาพใดใด สามารถเขียนความสัมพันธ์ของแนวคิดกราฟได้ดังนี้  $CG \equiv \langle V, E \rangle$  เมื่อ  $CG$  เป็นสเปเชียลเอ็นทิตี (spatial entities) ที่แทนด้วยเซตของ  $V$  คือจุด (vertex) หรือโหนด (node) เมื่อ  $i \in \{1 \dots N\}$  และความสัมพันธ์ของวัตถุภายในภาพใด เกิดจากโหนดสองโหนดที่เชื่อมต่อกันด้วย  $E$  เมื่อ  $E \subset V \times V$ , เมื่อ  $E$  คือความสัมพันธ์ระหว่างโหนดสองโหนด (edge) ความสัมพันธ์รูปแบบนี้ถูกเรียกว่า binary spatial relationship ดังนั้นสามารถแทนความสัมพันธ์ของสองจุดได้ด้วย  $e_{ab} \equiv \langle v_a, v_b \rangle \in E$  เมื่อกำหนดให้  $v_a, v_b \in V$  จากตัวอย่างภาพที่ 3.11 ก. แสดงภาพตัวอย่างจากโปรแกรม LabelMe ถูกให้ความหมายวัตถุตามคำหลักดังนี้ kid kid grass ball ball snorkel snorkel flipper ดังแสดงในภาพที่ 3.11 ข. และสามารถแทนด้วยความสัมพันธ์ของวัตถุ ด้วยแนวคิดกราฟโหนดแต่ละโหนดบนกราฟถูกแทนด้วยคำหลัก ดังนั้น grass เป็น root node ที่มีความสัมพันธ์ร่วมกันกับ kid ทั้งสองและ โหนดของ kid จะเกิดความสัมพันธ์ต่อไปยังส่วนต่างๆ เช่น ball snorkel และ flipper ดังนั้นการแสดงถึงความสัมพันธ์จะถูกเชื่อมเข้าด้วยกันเป็นค่าของ  $E$

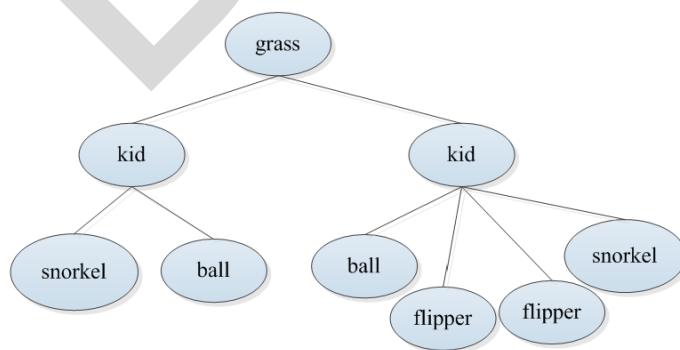


คำหลัก: kid kid grass ball ball snorkel snorkel flipper

ก. วัตถุที่ถูกแท็กบนภาพจากโปรแกรม LabelMe



ข. วัตถุที่ถูกแท็กบนภาพจากโปรแกรม LabelMe

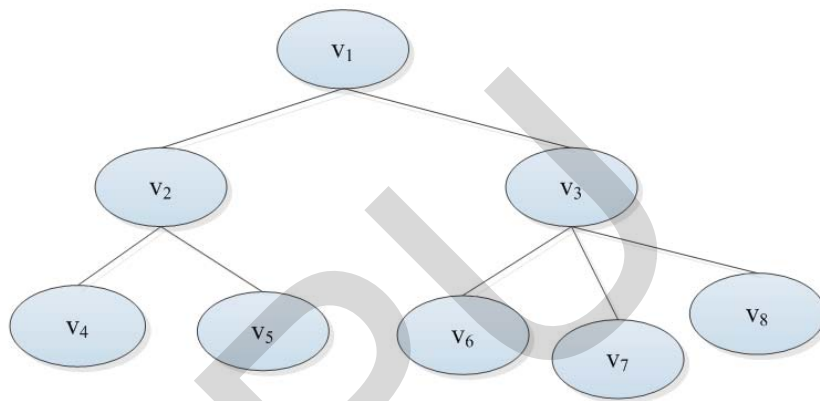


ค. แทนความสัมพันธ์ของวัตถุที่ถูกแท็กด้วยแนวคิดกราฟ

ภาพที่ 3.11 ตัวอย่างความสัมพันธ์ของวัตถุด้วยแนวคิดกราฟ



จากภาพที่ 3.11 ค. แสดงความสัมพันธ์ ที่เกิดขึ้นดังนั้นตามแนวคิดกราฟคือโหนดทั้งหมด ประกอบด้วยคำหลักดังนี้ grass kid kid ball ball snorkel snorkel flipper และจุดเชื่อมต่อระหว่าง grass กับ kid ถูกแทนด้วย  $e_{grass,kid} \equiv \langle v_{grass}, v_{kid} \rangle \in E$  ดังนั้นสามารถเขียนรูปแบบใหม่เพื่อง่ายในการอ่านดังแสดงในภาพที่ 3.12 สามารถเขียนเป็นสมการได้ดังนี้  $CG \equiv \langle V, E \rangle$  กำหนดให้ โหนดต่างๆบนภาพด้วย  $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8\}$ , และแทนความสัมพันธ์ระหว่างโหนดต่างๆด้วยสมการ เมื่อกำหนดให้  $e_{1,2} \equiv \langle v_1, v_2 \rangle$  ดังนั้นสามารถเขียน edge ทั้งหมดในกราฟได้ดังนี้  $E = \{e_{1,2}, e_{1,3}, e_{2,4}, e_{2,5}, e_{3,6}, e_{3,6}, e_{3,6}, e_{3,7}, e_{3,8}\}$



ภาพที่ 3.12 ความสัมพันธ์ของวัตถุแทนด้วยแนวคิดกราฟ

### 3.2.2 การกำหนดค่าน้ำหนักข้อมูล

การกำหนดค่าน้ำหนักสำหรับข้อมูลคำหลักบนภาพ โดยคิดค่าน้ำหนักจากความสัมพันธ์ ข้อมูลคำหลักที่เกิดขึ้นทั้งหมดระหว่างคำหลักภายในภาพโดยจะเก็บเป็นค่า IC (Information Content) [Resnik et al.,1995] แทนรายละเอียดข้อมูลดังนั้นสามารถนับจำนวนคำหลักที่เกิดขึ้นเป็นค่าของความน่าจะเป็นของแต่ละคำหลักบนภาพได้ด้วยสมการดังนี้

$$P(v_i) = \frac{freq(v_i)}{\sum_i freq(v_i)}$$

เมื่อกำหนดให้  $P(v_i)$  แทนความน่าจะเป็นของคำหลัก  $v_i$  และ  $freq(v_i)$  แทนความถี่ของคำหลัก  $v_i$  ที่เกิดขึ้น

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

เมื่อกำหนดให้  $word(v)$  แทนกลุ่มของคำหลักที่อยู่ในกลุ่มคำหลัก  $v$  ดังนั้นเมื่อมีการคำนวณในแต่ละโหนดสามารถเขียนเป็นสมการของ IC ใน  $v_i$  ได้ดังนี้

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

เมื่อคำหลักมีความสัมพันธ์มากกว่าหนึ่งกลุ่มใน WordNet สามารถเขียนสมการหาความสัมพันธ์ของกลุ่มใน WordNet สามารถที่จะเขียนสมการหาความสัมพันธ์ที่เกิดขึ้นสูงสุดได้ของกลุ่มได้จากการหาความเหมือนกันของข้อมูลมากที่สุดของคำหลักดังนี้

$$sim(v_1, v_2) = \max_{v_1, v_2} [sim(v_1, v_2)]$$

เมื่อกำหนดให้  $v_1$  และ  $v_2$  เป็นคำหลักที่เกิดความสัมพันธ์ต่อกันในกลุ่มของ  $v_1$  และ  $v_2$  ใน WordNet ตามลำดับ

### 3.2.3 การหาความสัมพันธ์ของสายเชื่อมโยงข้อมูล

ภายในภาพหนึ่ง ๆ มีวัตถุที่ถูกแทนเป็นโหนดได้หลายโหนดแต่ละโหนดถูกเชื่อมโยงด้วย edge ระหว่างกันเป็นทอด ๆ สามารถแสดงลำดับความสัมพันธ์ระหว่างโหนดได้ดังแสดงในภาพที่ 3.12  $v_2$  เป็นโหนดที่มีความสัมพันธ์กับโหนดที่อยู่ด้านบนหรือโหนดพ่อแม่ (parent node) คือ  $v_1$  และ  $v_2$  มีโหนดที่อยู่ด้านล่างหรือโหนดลูก (children node) คือ  $v_3$  และ  $v_4$  ดังนั้นจะเห็นว่าความสัมพันธ์ที่เกิดขึ้นของ  $v_2$  นั้นเกิดขึ้นร่วมกันหลายโหนด ดังนั้นจะหาความสัมพันธ์ที่เกิดขึ้นของโหนดสายการเชื่อมโยง(link) โดยจะใช้การคำนวณในลักษณะของความน่าจะเป็นจากความสัมพันธ์ที่เกิดขึ้นเป็นค่าของ link strength (LS) เป็นลอการิทึมแบบผกผันเพื่อหาค่า LS ของโหนด  $v_i$  ได้ดังนี้

$$LS(v_i) = -\log(P(v_i | parent(v_i)))$$

$$LS(v_i) = IC(v_i) - IC(parent(v_i))$$

โดยที่  $parent(v_i)$  แทนโหนดที่อยู่เหนือโหนด  $v_i$  เมื่อ  $i \in m$ ,  $m$  คือจำนวนของโหนดที่เกิดขึ้นทั้งหมดในภาพ และ  $IC(v_i)$  แทน Information Content ของโหนด  $v_i$  ดังนั้นสามารถเขียนสมการของความน่าจะเป็นที่จะเกิดขึ้นได้ดังนี้

$$P(v_i / parent(v_i)) = \frac{P(c_i \cap parent(v_i))}{P(parent(v_i))}$$

$$P(v_i / parent(v_i)) = \frac{P(v_i)}{P(parent(v_i))}$$

### 3.2.4 การหาความสัมพันธ์แพทเทิร์นระหว่างโหนด

สำหรับการหาความสัมพันธ์แพทเทิร์นระหว่างโหนด (link pattern: LP) จะเป็นการสื่อถึงความหมายของภาพที่สำคัญ เพราะในภาพจะมีโหนดหรือวัตถุที่มีการเชื่อมโยงต่อกันใน

รูปแบบของแพทเทิร์น ดังนั้นจากข้างต้นมีการกำหนดโหนดต่างๆบนภาพด้วย  $V = \{v_1, v_2, \dots, v_n\}$ , เมื่อ  $n$  แทนจำนวนโหนดบนภาพ  $v_1$  แทนโหนดที่ 1 และแทนความสัมพันธ์ระหว่างโหนดต่างๆด้วย  $e_{1,2}$  คือ edge ที่เชื่อมโยงระหว่างโหนด  $v_1$  และโหนด  $v_2$  เมื่อกำหนดให้  $e_{1,2} \equiv \langle v_1, v_2 \rangle$  ดังนั้น  $\delta_i = \{e_{1,2}, e_{2,3}, \dots, e_{m,n}\}$  เมื่อ  $\delta_i \in \delta$  สามารถเขียนสมการของการ LP ได้ดังนี้

$$LP(\delta_i) = \frac{freq(\delta_i)}{\sum_i freq(\delta_i)},$$

เมื่อกำหนดให้  $freq(\delta_i)$  แทนความถี่ที่เกิดขึ้นจากแพทเทิร์นของ  $\delta_i$  เมื่อ  $i \in n$ ,  $n$  คือจำนวนของแพทเทิร์นที่เกิดขึ้นทั้งหมดในภาพ

### 3.2.5 การวัดความคล้ายกันของกราฟแบบจับคู่

การหาความคล้ายกันของภาพต้องมีการเปรียบเทียบในแต่ละส่วนของกราฟตามค่าน้ำหนักของแต่ละกิ่งก้านที่เกิดขึ้นบนกราฟ โดยกำหนดให้  $G_1$  และ  $G_2$  แทนกราฟที่จะทำการเปรียบเทียบการเปรียบเทียบความคล้ายกันมีทั้งหมด 2 ขั้นตอน

— การเปรียบเทียบความคล้ายด้วยข้อมูล

การเปรียบเทียบความคล้ายด้วยข้อมูลภายในกราฟ  $G_1$  และ  $G_2$  แทนกราฟที่ 1 และ 2 โดยที่มีข้อมูลภายในเป็น  $v_i \in G_1$  และ  $v_j \in G_2$  โดยใช้วิธีการเปรียบเทียบข้อมูลภายในของแต่ละกราฟด้วยสมการดังนี้

$$sim\_content(G_1, G_2) = content\_match(G_1, G_2)$$

โดยที่อัลกอริทึมของ  $content\_match$  มีดังนี้

#### Algorithm 1: Content\_Matching.

1. input: Semantic Conceptual Graph  $G_1$ ,  
Representative Graph  $G_2$ ;
2. output: the content matching coefficient.
3. BEGIN
4. Initialize  $v_i \in G_1, v_j \in G_2$  where  $v_i, v_j$  are leaves,  
 $cm = 0, total\_cm = 0$ .
5. **While**  $G_1(v_i)$  is not NULL
6.     if  $find\_content(G_1(v_i) == G_2(v_j))$

7.  $cm = cm + LS(G_2(v_j))$ .
  8.  $increment(total\_cm, i)$ .
  9. else if  $find\_content(G_1(parent(v_i)) == G_2(parent(v_j))$
  10.  $cm = cm + LS(G_2(parent(v_j)))$ .
  11.  $increment(total\_cm, i)$ .
  12. else increment  $i$ .
  13. end if
  14. end if
  15. end while
  16. RETURN  $\left(\frac{cm}{total\_cm}\right)$ .
- 

— การเปรียบเทียบความคล้ายด้วยความสัมพันธ์ของข้อมูลกราฟ  
การเปรียบเทียบความคล้ายด้วยความสัมพันธ์ของข้อมูลภายในกราฟ จะใช้วิธีการเปรียบเทียบข้อมูลภายในของแต่ละกราฟด้วยสมการดังนี้

$$sim\_relation(G_1, G_2) = relationship\_match(G_1, G_2)$$

โดยที่อัลกอริทึมของ Relationship\_match มีดังนี้

---

#### Algorithm 2: Relationship\_Matching.

---

1. input: Conceptual Graph  $G_1$ ,  
Representative Graph  $G_2$ ;
2. output: the relationship matching coefficient.
3. BEGIN
4. Initialize  $\delta_i \in G_1, \delta_j \in G_2, rm = 0, total\_rm = 0$ .
5. **While**  $G_1(\delta_i)$  is not NULL
6. if  $find\_relationship(G_1(\delta_i) == G_2(\delta_j))$
7.  $rm = rm + LP(G_1(R\delta_j))$ .
8.  $increment(total\_rm, i)$ .
9. end if
10. end while

$$11. \text{ RETURN } \left( \frac{rm}{total\_rm} \right).$$


---

### 3.3 ขั้นตอนการวัดประสิทธิภาพ

จากภาพที่ 3.1 ขั้นตอนการวัดประสิทธิภาพ (evaluation) เป็นขั้นตอนสุดท้ายของกระบวนการจำแนกความหมายภาพ จะนำเอาผลการจำแนกกลุ่มของภาพที่จัดได้มาทำการวัดประสิทธิภาพของสิ่งที่นำเสนอข้างต้น โดยจะตรวจสอบกลุ่มของภาพที่จัดได้ว่ามีค่าเป็นอย่างไร เมื่อเทียบกับกลุ่มของภาพที่ถูกต้อง ซึ่งต้องมีการวัดค่าความระลึก (recall) และค่าความแม่นยำ (precision) จะเป็นค่าที่แสดงว่า การค้นคืนข้อมูลได้ตรงกับความต้องการเพียงใด ส่วนค่าความระลึกจะเป็นค่าที่แสดงถึงความครอบคลุมในการจัดกลุ่มภาพ หลังจากนั้นจะนำค่ามาคำนวณในรูปของค่าความถูกต้อง (accuracy) และ F-measure ต่อไปและนำค่าทั้งหมดมาแปลผลและประเมินผลลัพธ์ที่ได้ว่ามีความเหมาะสม หรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ในรูปที่สามารถเข้าใจได้ง่าย เช่น ตาราง หรือกราฟซึ่งในงานวิจัยนี้ได้แสดงในรูปของตารางเชื่อมโยงปฏิสัมพันธ์ความสับสน (confusion matrix) สามารถอ่านเพิ่มเติมได้ในบทที่ 2

## บทที่ 4

### ผลการทดลอง

ในการทำวิจัยครั้งนี้เพื่อหาความหมายของภาพที่เกิดจากการแปลความหมายรวมของทุกวัตถุที่ปรากฏบนภาพ ได้นำเสนอในรูปแบบของการแทนข้อมูลภาพ ด้วยความสัมพันธ์ของข้อมูลวัตถุภายในภาพ หรือเรียกว่า แนวคิดกราฟ (Conceptual Graph) ในลักษณะของกราฟที่นำเสนอขึ้นจะมีแสดงความสัมพันธ์ของข้อมูลภายในภาพของวัตถุที่เกิดขึ้นทั้งหมดในรูปแบบของกราฟรวมทั้งความสัมพันธ์ (Relationship) ระหว่างวัตถุที่เกิดขึ้น และจะมีการวัดค่าความสัมพันธ์ของความเหมือนกันด้วยวิธีการแทนข้อมูลภาพแบบคอนเซปชวลกราฟ (Conceptual Graph) และทำเปรียบเทียบความเหมือนกันของความหมายภาพด้วยการหาความเหมือนของภาพทั้งหมด 4 วิธีการ โครงข่ายประสาทเทียมแบบหลายชั้น แผนผังการจัดระบบตัวเอง เครือข่ายแบบเบย์ และการวัดวิธีการวัดความคล้ายกันของกราฟแบบจับคู่ (Similarity Measure with Conceptual Graph Matching) แสดงผลการทดลองดังนี้

#### 4.1 การกำหนดข้อมูลภาพ

ข้อมูลสำหรับการทดลองได้ใช้ฐานข้อมูลที่เก็บรวบรวมมาจาก LabelMe [B. C. Russell, 2008] โดยแอปพลิเคชันนี้สามารถทำงานได้อย่างเต็มรูปแบบบนเว็บในลักษณะของเครื่องมือให้ความหมาย ซึ่งได้มีการคัดเลือกภาพสำหรับการทดลองให้อยู่ในหมวดหมู่ในวงจำกัดเพียง 6 กลุ่มเพื่อทดสอบการจำแนกความหมายของข้อมูลภาพ ประกอบด้วยกลุ่มของ ภาพที่ทำงาน (office), ภาพสนามหญ้า (lawn), ภาพภายในห้อง (room), ภาพเมือง (city), ภาพทะเล (beach), ภาพสนามกีฬา (stadium) ซึ่งภาพที่นำมาถูกจำแนกความเหมือนกันของความหมายภาพด้วยการใช้คำหลักที่ถูกแท็กไว้บนภาพจากหลายผู้ใช้ใน LabelMe และได้ทำการทดลองจำแนกด้วยโครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network) แผนผังการจัดระบบตัวเอง (Self-organizing maps:SOM) เครือข่ายแบบเบย์ (Bayesian Network) และการวัดวิธีการวัดความคล้ายกันของกราฟแบบจับคู่ (Similarity Measure with Conceptual Graph Matching) ดังนั้นในการทดลองจะมีการแสดงผลลัพธ์ด้วยตารางเชื่อมโยงปฏิสัมพันธ์ความสับสน (confusion matrix) จุดประสงค์เพื่อทำการเปรียบเทียบการใช้ฟีเจอร์ในรูปแบบต่างๆกัน ได้แสดงผลการทดลองดังนี้

## 4.2 ผลการจำแนกความหมายของข้อมูลภาพ

จากการทำการทดลอง เพื่อจำแนกความหมายของข้อมูลภาพออกโดยในงานวิจัยนี้จะแบ่งออกเป็น 6 กลุ่มประกอบด้วยกลุ่มของ กลุ่มที่ 1 แสดงถึงภาพที่ทำงาน (office) เป็นสถานที่ภายใน (indoor) ที่บุคคลทำงาน อาจมีเครื่องใช้สำนักงานประกอบ, กลุ่มที่ 2 ภาพสนามหญ้า(lawn) เป็นภาพภายนอก (outdoor) ที่มีสนามหญ้า สนามเด็กเล่นหรือถนน เพื่อใช้ในการพักผ่อน ทำกิจกรรมอื่นๆ, กลุ่มที่ 3 ภาพภายในห้อง (room) เป็นภาพภายในลักษณะภายใน (indoor) บ้าน อาจมีโซฟา เฟอร์นิเจอร์ หรือห้องครัวประกอบ, กลุ่มที่ 4 ภาพเมือง (city) เป็นภาพในลักษณะ (outdoor) ที่ประกอบด้วยตึก อาคาร ถนน รถยนต์ เป็นต้น, กลุ่มที่ 5 ภาพทะเล (beach) เป็นภาพในลักษณะภายนอก (outdoor) ประกอบด้วยส่วนต่างๆ เช่น ท้องฟ้า ทะเล พื้นน้ำ พื้นทราย หรือ ต้นไม้ อาคาร สถานที่ร้านค้า กลุ่มที่ 6 ภาพสนามกีฬา (stadium) เป็นภาพในลักษณะภายนอก (outdoor) ที่ประกอบด้วย สนามแข่งขันต่างๆ ที่มีอุปกรณ์กีฬาประกอบ โดยที่เครื่องมือการจำแนกทั้งหมด 4 วิธีด้วยกันคือ naïve-Bayes, Multilayer, SOM, Similarity Matching โดยเป็นการจำแนกในส่วนของคำหลักที่ปรากฏบนภาพจะมีความสัมพันธ์กันในลักษณะของ WordNet ผลของการจำแนกความหมายของภาพใน 6 กลุ่ม ดังภาพแสดงในตารางที่ 4.1 ถึง 4.4 ด้วยตารางเชื่อมโยงปฏิสัมพันธ์ความสัมพันธ์ (%)

Class	Confusion Matrix (%)					
	Office	Lawn	Room	City	Beach	Stadium
Office	79	3	5	7	1	5
Lawn	3	65	8	5	14	5
Room	4	9	82	2	2	1
City	7	4	1	78	4	7
Beach	1	13	2	3	74	7
Stadium	0	2	0	3	0	96
Accuracy rate	78.7					

ตารางที่ 4.1 ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี naïve-Bayes

Class	Confusion Matrix (%)					
	Office	Lawn	Room	City	Beach	Stadium
Office	71	3	11	7	1	7
Lawn	8	65	4	5	11	7
Room	5	2	89	2	2	0
City	12	6	4	64	5	9
Beach	3	14	4	2	72	5
Stadium	2	3	1	3	5	86
Accuracy rate	74.5					

ตารางที่ 4.2 ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี Multilayer

Class	Confusion Matrix (%)					
	Office	Lawn	Room	City	Beach	Stadium
Office	79	3	5	7	1	5
Lawn	5	70	3	7	8	7
Room	3	3	92	1	1	0
City	4	5	2	77	5	7
Beach	2	11	3	3	74	7
Stadium	0	9	0	3	0	96
Accuracy rate	80.3					

ตารางที่ 4.3 ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี SOM

Class	Confusion Matrix (%)					
	Office	Lawn	Room	City	Beach	Stadium
Office	95	0	0	0	1	5
Lawn	4	80	1	4	5	6
Room	3	1	85	2	5	4
City	2	1	3	85	2	4
Beach	1	3	2	1	92	2
Stadium	0	5	0	0	0	95
Accuracy rate	88.8					

ตารางที่ 4.4 ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี Similarity Matching



Class	BayesNet.			Multilayer			SOM			Similarity Matching		
	Prec.	Recall	F1	Prec.	Recall	F1	Prec.	Recall	F1	Prec.	Recall	F1
Office	84.0	79.0	81.4	70.3	71.0	70.6	84.9	79.0	81.9	90.5	94.1	92.2
Lawn	67.7	65.0	66.3	69.9	65.0	67.4	69.3	70.0	69.7	88.9	80.0	84.2
Room	83.7	82.0	82.8	78.8	89.0	83.6	87.6	92.0	89.8	93.4	85.0	89.0
City	79.6	77.2	78.4	77.1	64.0	69.9	78.6	77.0	77.8	92.4	87.6	89.9
Beach	77.9	74.0	75.9	75.0	72.0	73.5	83.1	74.0	78.3	87.6	91.1	89.3
Stadium	79.3	95.0	86.5	75.4	86.0	80.4	78.7	88.9	83.5	81.9	95.0	88.0
Accuracy	78.74			74.50			80.26			88.81		

ตารางที่ 4.5 ตารางเปรียบเทียบผลลัพธ์การจำแนกข้อมูล

จากผลการทดลองได้แสดงในตารางที่ 4.1 - 4.4 เป็นตาราง จะเห็นว่า lawn ได้ค่าความถูกต้องเพียง 65% ด้วยวิธีการจำแนกแบบ naïve-Bayes และ Multilayer แต่สำหรับ กลุ่มภาพ stadium จะได้ค่าความถูกต้องถึง 96% สำหรับ naïve-Bayes และ SOM ซึ่งได้มากกว่าการจำแนกด้วย Similarity Matching ถึง 1% แต่อย่างไรก็ตามการจำแนกในกลุ่มภาพ Stadium ด้วยวิธี Multilayer จะได้ค่าความถูกต้อง 86% ซึ่งมีค่าน้อยกว่าวิธีอื่นๆ จากภาพในกลุ่ม room ผลการทดลองได้ค่าความถูกต้อง 82% 89% และ 92% สำหรับ naïve-Bayes, Multilayer และ SOM จะเห็นว่าวิธีการที่มีการจำแนกด้วย SOM ได้ค่าความถูกต้องมากที่สุดสำหรับกลุ่มภาพ room สำหรับการจำแนกด้วยวิธี Similarity Matching ได้ค่าความถูกต้องเพียง 85% เท่านั้น SOM จะสามารถจำแนกได้มีค่าความถูกต้องที่สูงกว่า แต่อย่างไรก็ตามเมื่อมีการเปรียบเทียบที่มีการจำแนกด้วย Similarity Matching จะเห็นว่ากลุ่มภาพที่เป็น lawn จะสามารถจำแนกความคล้ายกันของกราฟที่แทนความเหมือนกันของภาพได้ค่อนข้างสูงถึง 95% เช่นเดียวกันกับกลุ่มภาพที่เป็น stadium และในกลุ่มภาพที่เป็น beach มีค่าความถูกต้อง 92% ซึ่งในวิธีการจำแนก ของ naïve-Bayes, Multilayer และ SOM ในกลุ่มของ beach จะได้สูงสุดเพียง 72% เท่านั้น



(ก)



(ข)



(ค)



(ง)



(จ)



(ฉ)

ภาพที่ 4.1 ตัวอย่างผลลัพธ์ของการจำแนกความหมายภาพ

(ก) ภาพที่ทำงาน (office) (ข) ภาพสนามหญ้า (lawn) (ค) ภาพภายในห้อง (room)

(ง) ภาพเมือง (city) (จ) ภาพทะเล (beach) (ฉ) ภาพสนามกีฬา (stadium)

จากตารางที่ 4.5 แสดงการเปรียบเทียบวิธีการจำแนก 4 รูปแบบ naïve-Bayes, Multilayer, SOM และ Similarity Matching โดยที่นำเสนอวิธีการใช้ Similarity Matching โดยที่การจำแนกด้วยการวัดความคล้ายกันของกราฟจะได้ค่าความถูกต้องเฉลี่ยรวมที่มีค่ามากที่สุด ถึง 88.81% ในขณะที่ SOM naïve-Bayes และ Multilayer มีค่าความถูกต้องเฉลี่ยรวมเพียง 80.26% 78.74% และ 74.50 ตามลำดับ ดังนั้นจะเห็นว่าการใช้ความสัมพันธ์ข้อมูลที่ถูกแทนที่ด้วยกราฟ (Conceptual Graph Representation) และการวัดค่าความหมายภาพด้วยวิธีการวัดความคล้ายกันของกราฟแบบจับคู่ สามารถช่วยในการจำแนกความหมายของภาพได้มากกว่าวิธีอื่น ดังแสดงภาพผลลัพธ์ของการจำแนกในภาพที่ 4.1

DPU

## บทที่ 5

### สรุปผลการทดลอง

ในบทนี้ได้ทำการสรุปผลการทดลอง และข้อเสนอแนะ รวมทั้งข้อผิดพลาดที่เกิดขึ้นจากการทดลอง รวมไปถึงสิ่งที่ควรปรับปรุงเพิ่มเติม เพื่อแก้ไขข้อผิดพลาดและแนวทางการทำวิจัยต่อในเรื่องของ semantic image ซึ่งเป็นงานวิจัยที่ปัจจุบันได้มีนักวิจัยให้ความสนใจอย่างแพร่หลาย เพราะสามารถนำไปประยุกต์ใช้ได้หลายด้านทั้งทางการแพทย์ และการในอุตสาหกรรม เป็นต้น

#### 5.1 สรุปและอภิปรายผลการวิจัย

งานวิจัยนี้ได้เสนอหัวข้อวิจัยทางการประมวลผลภาพ ในส่วนของการวัดความคล้ายกันของภาพ เพื่อให้ได้ความหมายของภาพที่อยู่ในหมวดหมู่เดียวกัน โดยปกติทั่วไปนั้นการใช้อัลกอริทึมที่มาสกัดข้อมูลภาพนั้นมักจะใช้สกัดเพียงข้อมูลที่เกิดขึ้นภายในภาพ แล้วนำมาประมวลผลเพื่อใช้ในการสืบค้นข้อมูลภาพ แต่ปัจจุบันได้มีการนำคำหลักที่ได้จากการให้ความหมายของการแท็กวัตถุบนภาพมาหาความสัมพันธ์ภายใน โดยพยายามหาความสัมพันธ์ที่คล้ายกันของวัตถุในหมวดหมู่เดียวกัน ดังนั้นในงานวิจัยนี้จึงได้นำเสนอ ในรูปแบบของการแทนข้อมูลภาพ ด้วยความสัมพันธ์ของข้อมูลวัตถุภายในภาพ หรือเรียกว่า แนวคิดกราฟ (Conceptual Graph) ในลักษณะของกราฟที่นำเสนอ นั้นจะมีแสดงความสัมพันธ์ของข้อมูลภายในภาพของวัตถุที่เกิดขึ้นทั้งหมดในรูปแบบของกราฟรวมทั้งความสัมพันธ์ (Relationship) ระหว่างวัตถุที่เกิดขึ้น และจะมีการวัดค่าความสัมพันธ์ของความเหมือนกันด้วยวิธีการที่เรียกว่า การจับคู่กราฟ (Graph Matching) ของข้อมูลภาพที่ถูกแทนค่าความหมายไว้ ดังนั้นภาพที่มีความหมาย เหมือนกันจะมีได้ค่าของความเหมือนกันมากที่สุด จากการทดลองจะสามารถพิสูจน์ได้ว่าทฤษฎีนี้เหมาะสมที่จะนำมาใช้ร่วมกันเพื่อแปลความหมายของภาพ เพราะฉะนั้นในงานวิจัยนี้ยังคงเป็นอีกแนวทางหนึ่งที่สามารถนำวิธีการที่นำเสนอเข้ามาประยุกต์เพื่อให้สามารถนำวิธีการมาใช้ช่วยในแปลความหมายของภาพได้ ดังผลสรุปจากการทดลองดังนี้

จากขั้นตอนการทดลองในบทที่ 4 สามารถสรุปได้ว่า การใช้ข้อมูลคำหลักที่ได้มาจาก LabelMe ซึ่งภายในคำหลักแต่ละคำมีความสัมพันธ์ตามรูปแบบของ WordNet และได้แนะนำวิธีการแทนข้อมูลภาพแบบคอนเซ็ปชวลกราฟ (Conceptual Graph) และทำเปรียบเทียบความเหมือนกันของ

ความหมายภาพด้วยการหาความเหมือนของภาพทั้งหมด 4 วิธีการ โครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network) แผนผังการจัดระบบตัวเอง (Self-organizing maps) เครือข่ายแบบเบย์ (Bayesian Network) และการวัดความคล้ายกันของกราฟแบบจับคู่ (Similarity Measure with Conceptual Graph Matching)

จากการทดลองได้ทำการแบ่งกลุ่มของภาพไว้ทั้งหมด 6 กลุ่ม ในขั้นตอนนี้จะทำการทดลองเพื่อจำแนกความหมายของข้อมูลภาพภาพที่ทำงาน (Office), ภาพสนามหญ้า(Lawn), ภาพภายในห้อง (Room), ภาพเมือง (City), ภาพทะเล (Beach), ภาพสนามกีฬา (Stadium) จะเห็นว่าการจำแนกภาพ ด้วย SOM สามารถได้ค่าความถูกต้องเฉลี่ยถึง 80.26% แต่อย่างไรก็ตามเมื่อมีการใช้วิธีที่นำเสนอคือการวัดความคล้ายกันด้วยกราฟ จะได้ค่าความถูกต้องเฉลี่ยถึง 88.81% เพราะฉะนั้นจากการทดลองสามารถสรุปได้ว่า การที่นำทฤษฎีแนวคิดกราฟและการวัดความคล้ายกันของกราฟเข้ามาช่วยในการจำแนกความเหมือนกันของกลุ่มภาพ สามารถช่วยในการจำแนกได้เป็นอย่างดีประมาณ 8%

## 5.2 ข้อเสนอแนะ

ในกระบวนการจำแนกข้อมูลภาพออกเป็นกลุ่มย่อยที่มีความหมายในรูปแบบของความหมายภาพโดยรวมนั้นจะเป็นการพิจารณาภาพที่ซับซ้อนมากขึ้น นอกจากการค้นคืนข้อมูลภาพหรือการจำแนกภาพโดยทั่วไปที่ต้องการค้นคืนเพียง ความเหมือนกันของวัตถุที่เกิดขึ้นบนภาพเท่านั้น ในงานวิจัยส่วนใหญ่จึงเน้นไปที่การสกัดข้อมูลภาพในรูปแบบของการสกัดด้วยอัลกอริทึมที่แตกต่างกัน และค้นหาภาพเพียงความเหมือนกันของรูปทรงหรือลักษณะเฉพาะ หรือเพียงแต่วัตถุบนภาพ เท่านั้น ทั้งที่มีความหมายและไม่มี ความหมาย แต่อย่างไรก็ตาม เบื้องต้นของผลลัพธ์ที่ได้จากการค้นคืนหรือจำแนกภาพ คือความเหมือนกันทางกายภาพ เช่น รูปทรง สี หรือชนิดของวัตถุ แต่ลักษณะการวิเคราะห์และพิจารณาของการเหมือนกันทางความหมายภาพ หรือ semantic นั้นจะมีลักษณะการวิเคราะห์ที่แตกต่างกันออกไป ในอีกรูปแบบหนึ่งซึ่งเป็นรูปแบบที่เกิดจากความคิดของมนุษย์ที่มีการแปลงความจากภาพ แต่อย่างไรก็ตามสิ่งที่ได้จากการทดลองนั้นยังคงมีข้อที่ต้องปรับปรุงเพิ่มอยู่อีกหลายส่วนด้วยกัน

สิ่งควรจะมีการปรับปรุงเพิ่มเติมเพื่อให้การแปลความหมายของภาพได้ดียิ่งขึ้นอีก ก็คือการพิจารณารายละเอียดของวัตถุของภาพ ซึ่งคำหลักที่ถูกเลือกมาจากกลุ่มภาพใน LabelMe นั้นอาจจะถูกจำกัดเพียงบางกลุ่มคำเท่านั้น ยังไม่หลากหลายมากเท่าที่ควร เพราะฉะนั้นควรจะมีการจัดกลุ่มของวัตถุที่เข้ามาทำการทดลองให้มีความหมายที่รัดกุมมากขึ้น และเมื่อภาพที่มีส่วนของสภาพแวดล้อมที่คล้ายคลึงกันมาก จะถูกจัดอยู่ในกลุ่มเดียวกันทั้งที่ วัตถุภายในภาพ โดยเฉพาะ วัตถุที่เป็น *มนุษย์* ที่เป็นวัตถุเด่นในภาพ ถ้ามนุษย์มีรูปร่างลักษณะท่าทางที่แปลงเปลี่ยน เช่น ยืน นั่ง นอน หรือ กระโดด ทำให้การตีความหมายของภาพโดยรวมมีความหมายที่แตกต่างกันออกไป จึงทำให้ท่าทางของมนุษย์ (human action) สามารถบอกได้ถึงความหมายของภาพโดยรวมได้เช่นกัน อาจจะต้องมีการใช้การพิจารณา อารมณ์ จากใบหน้าและท่าทางของการแสดงออกสัมพันธ์ควบคู่กันไป ซึ่งงานวิจัยในส่วนของอารมณ์มนุษย์ที่แสดงออกทางใบหน้า และท่าทางก็ยังคงเป็นงานวิจัยยังคงวิจัยกันอยู่มาก เพราะฉะนั้นในงานวิจัยที่นำเสนอนี้เป็นอีกแนวทางหนึ่งที่พยายามจะคิดค้นวิธีการที่จะหาความหมายที่เกิดขึ้นจากภาพ ในอีกมุมมองหนึ่งซึ่งยังคงต้องมีการพิจารณาและวิเคราะห์ปรับปรุงการทดลองต่อไป.

## บรรณานุกรม

- นัศพ์ชาณัน ชินปัญชณะ, เอกสารประกอบการสอนวิชา การประมวลผลภาพ (Image processing), มหาวิทยาลัยธุรกิจบัณฑิต, 2552.
- 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. Hanbury, A Study of Vocabularies for Image Annotation, Proceedings of the second international conference on Semantics And digital Media Technologies (SAMT) 2007, Springer LNCS 4816, pages 284–287, Genoa, Italy. (© Springer-Verlag Berlin Heidelberg 2007 - Online on SpringerLink)
- A. Sorokin and D. Forsyth, “Utility Data Annotation with Amazon Mechanical Turk,” In First IEEE Workshop on Internet Vision CVPR, 2008.
- A. Torralba, B. C. Russell, J. Yuen, “LabelMe: Online Image Annotation and Applications,” Proceedings of the IEEE, Vol. 98, n. 8, August 2010, pp. 1467 – 1484.
- Adrian Popescu, Gregory Grefenstette, Pierre-Alain Moëllic, “Improving Image Retrieval Using Semantic Resources,” In Collection of Advances in Semantic Media Adaptation and Personalization, 2008, pp. 75-96.
- Apostol Paul Natsev , Atul Chadha , Basuki Soetarman , Jeffrey Scott Vitter, “CAMEL: Concept Annotated iMagE Libraries,” Storage and Retrieval for Media Databases Proceedings, vol. 4315, 2001.
- B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, “LabelMe: A Database and Web-based Tool for Image Annotation,” Intl. J. Computer Vision, vol 77, numbers 1-3, May, 2008, pp. 157–173.
- Barnard, K., Duygulu, P., de Freitas, N., Forsyth, D., Blei, D., Jordan, M.I.: Matching words and pictures. *Journal of Machine Learning Research* 3, 2003, pp. 1107–1135.
- Benitez, A.B., and S.-F. Chang, “Semantic Knowledge Construction From Annotated Image Collections”, *International Conference On Multimedia & Expo (ICME-2002)*, Lausanne, Switzerland, Aug 26-29, 2002.
- Carbonetto P., de Freitas, N., Barnard, K.: A statistical model for general contextual object recognition. In: Pajdla, T., Matas, J(G.) (eds.) *ECCV 2004*. LNCS, vol. 3021, Springer, Heidelberg, 2004, pp. 350–362.

- Carson, C., et al., "Blobworld: A system for regionbased image indexing and retrieval," In Proc. Int. Conf. Visual Information System, 1999.
- Chen, Y., Wang, J.Z., "Image categorization by learning and reasoning with regions," Journal of Machine Learning Research 5, 2004, pp. 913–939.
- Chester M, *Neural networks: a tutorial*, Prentice-Hall, Englewood Cliffs, 1993.
- Cox, I., Miller, M., Minka, T., and Papathomas, T., "The bayesian image retrieval\_ system, PicHunter," IEEE Trans. on Image Processing, Vol.9, No.1, pp. 20-37, 2000.
- D. Comaniciu, R. Visvanathan, P. Meer, "Kernel based object tracking". *IEEE Trans. Pattern Anal. Mach. Intell*, 25(5) , 2003, pp. 564–575.
- D. Comaniciu, V. Ramesh, 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.
- Di Sciascio E. and Mongiello, M., "Query by sketch and relevance feedback for content-based image retrieval over the web," *Journal of Visual Languages and Computing*, 1999.
- Everingham, M., et al., "The 2005 PASCAL visual object classes challenge," In: Selected Proceedings of the First PASCAL Challenges Workshop, 2006.
- F. Porikli, "Integral histogram a fast way to extract histograms in cartesian spaces," *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- Fei-Fei, L., Fergus, R., Perona, P., "Learning generative visual models from few training examples an incremental bayesian approach tested on 101 object categories," In: Proc. Workshop on Generative-Model Based Vision, June 2004.
- Flickner, M., et al., "Query by image and video content: The QBIC System," *IEEE Computer*, Vol. 28, No. 9, 1995.
- Flickr: <http://www.flickr.com> ค้นคืนเมื่อวันที่ 9 ตุลาคม 2554.
- Flood, I., and Kartam, N., "Neural networks in civil engineering. I: Principles and understanding." *J. Computing in Civil Engineering*, ASCE, Vol. 8, No. 2, 1994, pp.131-148.



- 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 S., Lanckriet G., "From Region Similarity to Category Discovery," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, 2011.
- Galleguillos C., McFee B., Belongie S., Lanckriet G.R G., "Multi-Class Object Localization by Combining Local Contextual Interactions", *IEEE Conference in Computer Vision and Pattern Recognition (CVPR)*, 2010.
- Gruber, T. R., "A Translation Approach to Portable Ontology Specifications," *Knowledge Acquisition*, Vol.5, No.2, 1993, pp. 199-220.
- 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., Worring, M., "Adding spatial semantics to image annotations," In: *Proceedings International Workshop on Knowledge Markup and Semantic Annotation at ISWC*, 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.
- James Z. Wang, Jia Li, 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, 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, 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, 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.
- Joo-Hwee Lim Mulhem, P. Qi Tian, "Event-based home photo retrieval," In *Proceedings of Multimedia and Expo (ICME)*, July 2003.
- Kobus Barnard and David Forsyth, "Exploiting Image Semantics for Picture Libraries", *The First ACM/IEEE-CS Joint Conference on Digital Libraries*, 2001.
- L. von Ahn and L. Dabbish, "Labeling images with a computer game," In *Proc. SIGCHI conference on Human factors in computing systems*, 2004.
- Li, J., Wang, J.Z., "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Trans. PAMI* 25(9), 2003, pp. 1075–1088.
- Lippmann, R.P., "An Introduction to Computing with Neural Nets," *IEEE Acoustical Speech and Signal Processing Magazine*, Vol. 4, 1987, pp.4–22.
- 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 G.A., Beckwith R., Fellbaum C., Gross D., Miller K., "Introduction to WordNet: An on-line lexical database," *International Journal of Lexicography*, Vol 3, 1990, pp. 235–244.
- Miller, George A., "WordNet: An on-line lexical database," *International Journal of Lexicography*, Vol. 3, 1990, pp. 235–312.
- Niblak, W., et al., "The QBIC project: Querying images by content using color, texture, and shape," In *Storage and Retrieval for Still Image and Video Databases*, vol. 1980, 1993, pp. 173-182.
- 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, 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.
- P. Resnik, “Using information content to evaluate semantic similarity in taxonomy,” In Proceeding of the 14th International Joint Conference on Artificial Intelligence, 1995.
- Perronnin, F., Dance, C., Csurka, G., Bressan, M., “Adapted vocabularies for generic visual categorization,” In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006.
- Petridis, K., Anastasopoulos, D., Saathoff, C., Timmermann, N.,Kompatsiaris, Y., Staab, S., “M-ontomat-annotizer: Image Annotation Linking Ontologies and Multimedia Low-Level Feature,” In: B. Gabrys, R.J. Howlett, and L.C. Jain, editors, KES (3), vol. 4253 of LNCS, Springer, 2006, pp. 633–640.
- Philippe Mulhem, 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.
- 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 (ICARCV), Singapore, December 2-5, 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, *Digital Image Processing*, 2nd Ed., Prentice-Hall, Inc., 2002.
- 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. 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., Freeman, W.T.: LabelMe: a database and web-based tool for image annotation. *Int. J. Comput. Vis.* vol.77, 2008.
- Smith, J., and Chang, S., “VisualSEEK: a fully automated content-based image query system,” In *Proc. Of the fourth ACM Int. Conf. on Multimedia*, 1996, pp. 87-98.
- Smith, M. *Neural networks for statistical modeling*, New York: Van Nostrand Reinhold, 1993.
- Squigl: <http://www.gwap.com> ค้นคืนเมื่อวันที่ 9 ตุลาคม 2554.
- T. Kohonen, *Self-organizing maps*, Springer-Verlag, Berlin, 1995.
- 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.
- Th. Gevers and A. W. M. Smeulders, “PicToSeek: Combining Colour and Shape Invariant Features for Image Retrieval,” *IEEE Trans. on Image Processing*, vol. 9, no. 1, 2000, pp. 102-119.
- The Corbis Corporation: <http://pro.corbis.com> ค้นคืนเมื่อวันที่ 9 ตุลาคม 2554.
- Vailaya, A. Figueiredo, M.A.T. Jain, 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 , Michael G. Strintzis, “An Ontology Approach to Object-Based Image,” In *Proc. IEEE International Conference on on Image Processing*, 2003.
- Volkmer, T., Smith, 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., 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, 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, A., Minka, T., “Object categorization by learned universal visual dictionary,” In: Proc. ICCV, 2005, pp. 1800–1807.
- Xin-Jing Wang, Lei Zhang, Xirong Li, Wei-Ying Ma, “Annotating Images by Mining Image Search Results,” IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), 2008.
- Yao, B., Yang, X., 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.
- Zhao T, Tang L H, Ip H, Qi F., “Visual Keyword Image Retrieval Based on Synergetic Neural Network for Web-Based Image Search”, *Journal of Real-Time Systems*, Vol.21, Issue 1/2 , 2001, pp.127 – 142.
- Zinger S., Millet, C., Mathieu, B., Grefenstette, G., H`ede, P., Mo`ellic, P.A., “Extracting an Ontology of Portrayable Objects from WordNet,” In: Proc. MUSCLE/ImageCLEF Workshop on Image & Video Retrieval Evaluation, 2005, pp. 17–23.
-

## ประวัติผู้วิจัย

### ผู้ช่วยศาสตราจารย์นศัพธ์ชาณัน 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

### ประสบการณ์งานสอน

อาจารย์ประจำ

มหาวิทยาลัยธุรกิจบัณฑิต

ภาควิชาเทคโนโลยีสารสนเทศธุรกิจ คณะเทคโนโลยีสารสนเทศ

- วิชาวิชาการพัฒนาและเขียนโปรแกรมบนเว็บ (2550-2554) (Tool: EditPlus, Html, java script)
- วิชาพื้นฐานสื่อประสมเชิงโต้ตอบ (2553-2554) (Tool: Adobe Photoshop, Illustrator , Flash CS3-5)
- วิชาธุรกิจอิเล็กทรอนิกส์ (1/2553) (Tool: Ms. Office, Dreamwaver)
- วิชาสำนักงานอิเล็กทรอนิกส์ (2550-2552) (Tool: Ms. Office)
- วิชาธุรกรรมอิเล็กทรอนิกส์เบื้องต้น (2/2552) (Tool: Ms. Office , Dreamwaver)

- วิชาการเขียนโปรแกรมเบื้องต้น (2/2550) (Tool: Java program)

#### ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ

- วิชาพื้นฐานสื่อประสม (1/2554) (Tool: Adobe Photoshop, Illustrator , Flash CS3-5)
- วิชาการประมวลผลภาพ (Image processing) (2/2552) (Tool: Matlab program)
- วิชาโครงการ (IT project)
  - ระบบวิเคราะห์เปรียบเทียบสเปคเครื่องคอมพิวเตอร์ที่เหมาะสม ( Specification and Comparative Analysis for Computer) 2553
  - ระบบวิเคราะห์งบรายได้-รายค่าใช้จ่ายส่วนบุคคล ( Analysis Personal Statement of income and Expenditure) 2553
  - โครงการ ระบบสนับสนุนการตัดสินใจในการเลือกอาชีพ (Online Career and Education Planning System) 2552
  - โครงการ ระบบสนับสนุนการตัดสินใจในการเลือกทัวร์ที่เหมาะสมกับลูกค้า (Intelligent Expert Tour System) 2552
  - โครงการ ระบบวิเคราะห์ดูแลสุขภาพ ( Health Care System Analysis) 2551
  - โครงการ ฐานข้อมูลเพื่อเรียกดูเส้นทางในเซตหลักสี่ (Route browsing database for Laksi) 2551

#### ภาควิชาคอมพิวเตอร์ คณะบริหารธุรกิจ

- วิชาการจัดการฐานข้อมูล (Database management system) (2542-2543)
- วิชาการเขียนโปรแกรมภาษา ซี (1/2542) (Tool: C program)
- วิชาการใช้โปรแกรมสำเร็จรูป (2/2541)
- วิชาคอมพิวเตอร์เบื้องต้น (Introduction to computer) (1/2541)

#### มหาวิทยาลัยหัวเฉียวเฉลิมพระเกียรติ

##### สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี

- วิชาโครงสร้างข้อมูล (Data Structure) (1/2541) (Tool: Borland Turbo C Programming language)

##### อาจารย์พิเศษ

- มหาวิทยาลัยศรีปทุม บรรยายและปฏิบัติการ วิชาโครงสร้างข้อมูล (Data Structure) (Tool: C++ program)
- มหาวิทยาลัยเกริก บรรยาย วิชาคลังข้อมูล (Data warehouse)
- วิทยาลัยภาคกลาง จ.นครสวรรค์ บรรยาย วิชาความปลอดภัยข้อมูล (Computer Security)

## งานวิจัยที่สนใจ

สนใจงานวิจัยทางการค้นคืนความหมายภาพ (Image retrieval) ทั้งภาพธรรมชาติทั่วไป (natural images) และ ภาพที่เป็นส่วนบุคคล (personal images) ในหัวข้อเกี่ยวกับการค้นคืน human activity หรือในหัวข้อการแปลความหมายภาพ (semantic human image) โดยเจาะจงทางการใช้ท่าทางของมนุษย์เพื่อแสดงถึงความหมายของภาพ รวมทั้งการจำแนกภาพต่าง ๆ เป็นกลุ่ม (Image classification) และการประยุกต์หลักการประมวลผลภาพ เพื่อนำมาใช้ในทางอุตสาหกรรม ได้จริง

## หนังสือ

การสื่อสารข้อมูลและระบบเครือข่ายคอมพิวเตอร์ (96304: Data Communications and Networking), มหาวิทยาลัยสุโขทัยธรรมาราช สาขาวิทยาศาสตร์และเทคโนโลยี, 2553. (หน่วยที่3 ตัวกลางในการสื่อสารข้อมูล, หน่วยที่ 10 โพรโทคอลประยุกต์)

## งานวิจัย

นศพีชานันธ ชินปัญชรณะ, ระบบตรวจนับวัตถุแบบอัตโนมัติด้วยเทมเพลตแมชชีง, มหาวิทยาลัยธุรกิจบัณฑิต, 2553.

นศพีชานันธ ชินปัญชรณะ, การจำแนกความหมายของภาพจากวัตถุโดยใช้หลักการโครงสร้างสเกตริตรอน, มหาวิทยาลัยธุรกิจบัณฑิต, 2552.

## ผลงานทางวิชาการ

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," Int. 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," 2007 International Workshop on Smart Info-Media Systems in Bangkok, November 1-2, 2007, Bangkok, Thailand.

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," Int. 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.

-----