การจำแนกความหมายภาพด้วยการสกัดฟีเจอร์จากโครงสร้างสเกตรีตรอนบนพื้นฐานแนวคิดกราฟลำดับชั้น

Semantic Image Classification using Structure Skeleton with Hierarchical Graph Concept

นัศพ์ชาณัณ ชินปัญช์ธนะ
คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณฑิตย์

Nutchanun Chinpanthana
Faculty of Information Technology, Dhurakij Pundit University
nutchanun.cha@dpu.ac.th

Abstract

Semantic classification is challenging task in the field of image processing. Many researchers have attempted to improve semantic models such as developing more sophisticated models, or generating intermediate representations by making statistic on low level description. However, the methods are rather rudimentary and it does not specific enough for representing the actual meaning. In this paper, we present a technique of the semantic image classification by using the human perception. The structure skeleton is used to combine the object components and image meaning. The feature selection methods are introduced to select the essential features from existing features. We combine a novel concept called the hierarchical representation graph for producing more semantic classification. This concept is formulates on a graph which is captured the relationships among objects in the images. The experimental results indicate that our proposed approach offers significant performance improvements in the interpretation of semantic images, compared, with the maximum of 80.28% accuracy.

Keywords: Digital images , Image processing , Semantics , Hierarchical graph , Feature selection

บทคัดย่อ

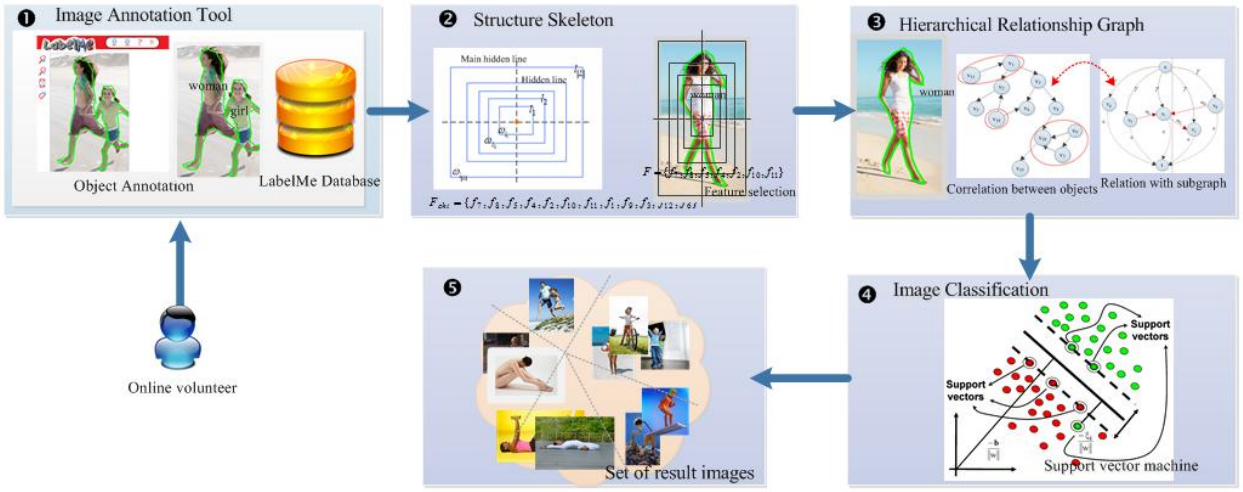
การจำแนกความหมายภาพเป็นงานวิจัยที่ท้าทายอย่างมากในสาขาการประมวลผลภาพ มีนักวิจัยหลายกลุ่มพยายามปรับปรุงวิธีการ เพื่อแก้ไขปัญหาของการแทนความหมายภาพด้วยการประมวลผลภาพระดับต่ำที่มีการใช้สถิติเข้ามาช่วยเพื่ออธิบายภาพ แต่อย่างไรวิธีการเหล่านั้นไม่สามารถที่จะนำมาใช้แทนความหมายของภาพได้อย่างแท้จริง งานวิจัยนี้ได้นำเสนอวิธีการใหม่สำหรับการแปลความหมายภาพจากการรับรู้ของมนุษย์ ด้วยการใช้โครงสร้างสเกลตรีตรอนร่วมกับการคัดเลือกฟีเจอร์ที่สำคัญเพื่อเป็นข้อมูลที่ใช้ช่วยในการจำแนกภาพและนำแนวคิดพื้นฐานความสัมพันธ์ของกราฟแบบลำดับชั้นเข้ามาเพื่อใช้ในการจำแนกความหมายของภาพ สำหรับแนวคิดนี้จะใช้กราฟในการแทนความสัมพันธ์ของวัตถุในภาพ ผลที่ได้จากการจำแนกความหมายภาพด้วยวิธีที่นำเสนอใหม่นี้สามารถจำแนกความหมายของภาพได้ดีกว่าวิธีอื่นๆ และได้ค่าความถูกต้องสูงถึง 80.28%

คำสำคัญ: ภาพดิจิทัล , การประมวลผลภาพ , ความหมาย , กราฟลำดับขั้น , การคัดเลือกฟีเจอร์

บทนำ

วิวัฒนาการของเครื่องมือและอุปกรณ์ถ่ายภาพดิจิทัลได้พัฒนาอย่างรวดเร็ว จนทำให้ภาพถ่ายภาพดิจิทัลมีจำนวนมากมาย ปัญหาที่ตามมาก็คือการจัดเก็บข้อมูลภาพที่เพิ่มขึ้น และจะจัดเก็บอย่างไรให้มีระบบที่ดีสามารถสืบค้นข้อมูลภาพได้อย่างง่ายและจำแนกข้อมูลภาพให้ตรงตามความหมายของภาพที่ต้องการของผู้ใช้มากที่สุด ทำให้งานวิจัยในปัจจุบันที่เกี่ยวข้องกับการค้นคืนรวมทั้งการจัดกลุ่มภาพให้ตรงกับความต้องการได้รับความสนใจจากนักวิจัยหลายกลุ่ม เป็นงานด้านการประมวลผลภาพ (image processing) ด้านการค้นคืนสารสนเทศ (image retrieval) และการจำแนกประเภทข้อมูลภาพ (image classification) เพื่อคัดเลือกภาพ ให้ตรงตามความต้องการของผู้ใช้งาน สำหรับงานวิจัยทางการประมวลผลภาพในการค้นคืนสารสนเทศ จะมีการค้นคืนตามคุณลักษณะพื้นฐานของภาพที่ถูกสกัดคุณลักษณะด้วยอัลกอริทึมต่างๆ เช่น สี (color) ลวดลาย (texture) รูปทรง (shape) เป็นต้น (Smeulders, et al., 2000) งานวิจัยของ SIMPLIcity (Li and Wang, 2003) เป็นการใช้คุณลักษณะฟีเจอร์ระดับต่ำด้วยสี ลวดลาย และตำแหน่งของพื้นที่ของภาพ จากผลลัพธ์จะสังเกตว่าผลลัพธ์ของภาพเป็นภาพที่มีโทนสีคล้ายกันเป็นหลัก แต่มีลักษณะวัตถุที่แตกต่างกันอย่างเห็นได้ชัดเจนการประมวลผลภาพระดับต่ำนั้นค่อนข้างยากที่จะจัดให้หมวดหมู่เดียวกัน แต่มีกลุ่มนักวิจัยที่พยายามปรับปรุงอัลกอริทึม เพื่อทำการค้นคืนภาพที่มีลักษณะฟีเจอร์ที่ใกล้เคียงกับภาพที่ต้องการมาก การปรับปรุงเทคนิควิธีการด้วยการนำวิธีการมาผสมผสานกันระหว่างคุณลักษณะเพื่อให้สามารถวิเคราะห์ในรูปแบบที่ซับซ้อน เช่นการรวมเทคนิคด้วยคุณลักษณะสีและรูปทรงของภาพเพื่อทำการค้นคืนภาพ (Hiremath, et al., 2007) แต่ในความเป็นจริงแล้วนั้น

ลักษณะการมองภาพของมนุษย์โดยทั่วไปเป็นการมองจากความหมายของภาพ โดยที่ไม่จำเป็นต้องมีคุณลักษณะสีหรือรูปทรงแบบเดียวกันก็สามารถเป็นภาพชนิดเดียวกันได้ ดังนั้นในการค้นคืนที่ใช้คุณลักษณะฟีเจอร์ระดับต่ำเพียงอย่างเดียว ทำให้ได้ผลลัพธ์ส่วนใหญ่ตรงกับคุณลักษณะของฟีเจอร์ที่สกัดมาแต่ไม่ได้ตรงกับความหมายที่เกิดภายในภาพที่ต้องการอย่างแท้จริง แต่อย่างไรก็ตามได้มีงานวิจัยอีกกลุ่ม ที่พยายามจะใช้เทคนิคของการเข้าใจความหมายของภาพแทน การสืบค้นแบบข้างต้น งานวิจัยในกลุ่มนี้พยายามที่จะมองข้อมูลบนภาพเป็นวัตถุ (object) ที่มีความหมายและแทนวัตถุนั้นๆ ด้วยคำหลัก (keyword) บนภาพ เรียกว่า การแท็ก (tag) หรือการให้ความหมายของวัตถุบนภาพเป็น ชื่อวัตถุ หรือคำศัพท์ ที่สอดคล้องกันเช่น “grass” , “plant” , “boat”, “sky” เป็นต้น (Galleguillos and Belongie, 2010) และใช้ความหมายหรือคำศัพท์นั้นเพื่อทำการสืบค้นข้อมูลแทนจะได้ผลที่ค่อนข้างดีกว่า แต่ขึ้นอยู่กับว่าอัลกอริทึมที่ถูกนำมาใช้นั้นจะเป็นลักษณะใด การค้นหาภาพด้วยเทคนิคนี้จะได้ผลลัพธ์ที่ขึ้นกับคำศัพท์ที่ถูกแท็กไว้บนภาพยิ่งมีการแท็กข้อมูลบนภาพมากยิ่งสามารถหาความเหมือนกันบนภาพมากขึ้นเท่านั้น แต่ในความเป็นจริงแล้ว การแท็กข้อมูลบนภาพในปัจจุบันนั้นเป็นเพียงการหาคำศัพท์ที่ต้องการบนภาพ แต่ไม่ได้ให้ความหมายภาพโดยรวม ความหมายของภาพเป็นการนำวัตถุที่ปรากฏบนภาพมารวมกันเพื่อวิเคราะห์จากความคิดของมนุษย์เพื่อให้ได้ผลลัพธ์คือคำศัพท์ใหม่ที่แทนความหมายของภาพทั้งภาพ มีนักวิจัยหลายกลุ่มพยายามที่จะแก้ไข และใช้หลากหลายวิธีเพื่อที่จะเชื่อมโยงให้ได้ความหมายของภาพอย่างแท้จริง แต่อย่างไรก็ตามซอฟต์แวร์ส่วนใหญ่นั้น จะทำการสืบค้นภาพในรูปแบบของการเทียบคำหลักเป็นคำต่อคำตามการเรียนรู้ของเครื่องจักร (machine learning) ที่เข้ามาช่วยในการหาความหมายของภาพ หรือตามระดับขั้นของการเรียนรู้ตามที่มีการเก็บข้อมูลไว้เท่านั้น เพราะฉะนั้นผลลัพธ์ของกลุ่มภาพที่ได้จะไม่ได้ขึ้นกับความหมายของภาพอย่างแท้จริง แต่ขึ้นกับคำหลักที่มีการเก็บข้อมูลลงไปบนภาพเท่านั้น



ภาพที่ 1 ขั้นตอนการจำแนกความหมายภาพ

วัตถุประสงค์

งานวิจัยนี้จึงได้นำเสนอการปรับปรุงการจำแนกความหมายภาพด้วยการแทนใช้ หลักทฤษฎีการรับรู้ภาพของมนุษย์ที่เรียกว่า โครงสร้างสเกลเลตอน เพื่อทำการเชื่อมโยงประสานกับวัตถุบนภาพให้สอดคล้องตามความหมายของกระบวนการคิดแปลความหมายภาพของมนุษย์ โดยข้อมูลทั้งหมดถูกคัดเลือกโดยใช้กระบวนการคัดเลือกข้อมูลเพื่อให้ได้ฟีเจอร์ที่มีคุณสมบัติที่ดีที่สุดเข้ามาทำการทดลอง และใช้ทฤษฎีลำดับชั้นความสัมพันธ์ของกราฟ (Hierarchical Relationship Graph) ในรูปแบบของการแทนข้อมูลภาพ ด้วยความสัมพันธ์ของข้อมูลวัตถุภายในภาพ กราฟจะแสดงความสัมพันธ์ของวัตถุในรูปแบบของกราฟรวมทั้งความสัมพันธ์ (Relationship) ระหว่างวัตถุที่เกิดขึ้น และทำเปรียบเทียบความเหมือนกันของความหมายภาพด้วยการจำแนกภาพทั้งหมดด้วย วิธีซัพพอร์ตเวคเตอร์แมชชีน (Support Vector Machines) แบบ One-against-One และแบบ One-against-all ด้วยฟังก์ชันเคอร์เนล แบบเชิงเส้น (Linear Kernel) การเก็บรวบรวมข้อมูลภาพและคัดเลือกภาพดิจิทัลที่มีวัตถุบนภาพเด่นชัด และภาพที่คัดเลือกเข้ามามนุษย์สามารถแปลความหมายภาพนั้นได้อย่างสมบูรณ์ ดังนั้นข้อมูลภาพที่เตรียมพร้อมจะสามารถเข้าสู่กระบวนการประมวลผลภาพ โดยในงานวิจัยจะมีการแบ่งขั้นตอนวิธีการดำเนินงานวิจัยโดยทั่วไปจะแบ่งออกเป็น 3 ส่วนดังนี้ 1) ขั้นตอนการเตรียมข้อมูล 2) ขั้นตอนการประมวลผล และ 3) ขั้นตอนการวัดประสิทธิภาพ ดังแสดงในภาพที่ 1

ขั้นตอนการประมวลผล

การประมวลผล (data processing) เป็นการสกัดข้อมูลจากภาพเพื่อนำมาเป็นข้อมูลฟีเจอร์เวกเตอร์ ประกอบด้วยข้อมูลวัตถุ ขนาดของแต่ละวัตถุ (object size) ตำแหน่งของแต่ละวัตถุ (object position) เข้ามาประกอบในการพิจารณาร่วมกับโครงสร้างสเกตตรีตรอน เมื่อได้ข้อมูลครบถ้วนจะนำเข้าสู่กระบวนการคัดเลือกฟีเจอร์

1. การให้ความหมายวัตถุหรือแท็กข้อมูล ได้ใช้ LabelMe (Russell, et al., 2008 ; Torralba, et al., 2010) เป็นเครื่องมือที่ได้รับการยอมรับอย่างกว้างขวาง สำหรับงานวิจัยทางด้าน Computer Vision โดยแอพพลิเคชันสามารถทำงานได้อย่างเต็มรูปแบบบนเว็บในลักษณะของเครื่องมือให้ความหมาย (Web-based annotation tools) ดังแสดงในภาพที่ 2 (ก) ปัจจุบันมีวัตถุบนภาพที่ถูกให้ความหมายรวมทั้งสิ้น 400,000 ผู้ใช้สามารถเข้าถึงโปรแกรมผ่านทางเครือข่ายออนไลน์ได้ ทำให้ผู้ใช้งานที่เข้ามาให้ความหมายภาพมาได้มากมาย และมีพื้นฐานของการให้ความหมายที่แตกต่างกันตามความสามารถ ของแต่ละบุคคล ข้อมูลคำหลักจะถูกจัดเก็บลงบนฐานข้อมูลพร้อมกับรูปภาพดังแสดงตัวอย่างของภาพที่ถูกแท็ก จากภาพที่ 2 (ข) ข้อมูลคำหลักที่ถูกแท็กแล้วจะแสดงไว้ทางขวามือ จะได้ข้อมูลคำหลักของวัตถุบนภาพประกอบด้วย grass, snorkel, snorkel, kid, kid, ball, ball, flipper และ flipper ดังนั้นวัตถุบนภาพถูกให้ความหมายด้วยวิธีการ

ที่เรียกว่า labeled object หรือ แท็ก (tag) (Mezaris, et al, 2003) และจัดเก็บข้อมูลในตัวแปรชื่อ $O_i = \{o_1, o_2, ...o_{12}\}$และการจำแนกข้อมูลภาพเป็นกระบวนการสุดท้าย มีขั้นตอน 3 ขั้นตอนดังนี้

(ก) (ข)

ภาพที่ 2 แสดงแอพพลิเคชัน LabelMe ก. การแท็กบริเวณของวัตถุ ข .คำศัพท์ของวัตถุที่ถูกแท็กบนภาพ

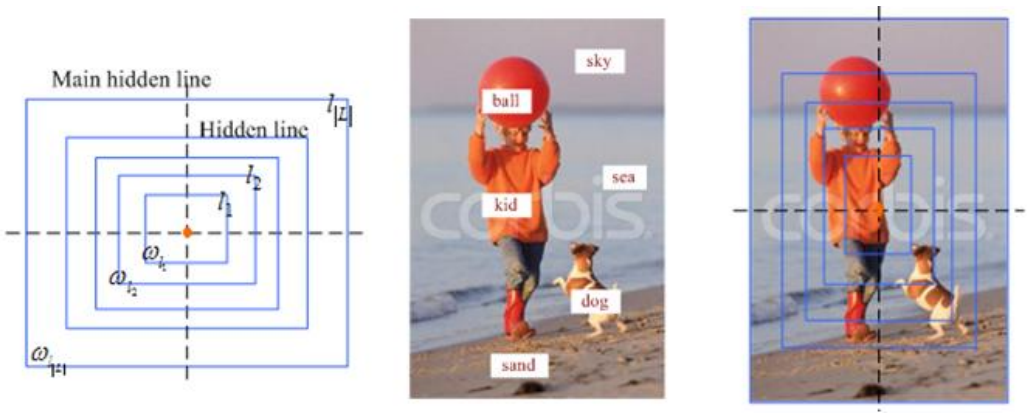
2. การสกัดข้อมูลจากวัตถุ (feature extraction) โดยใช้โครงสร้างสเกตรีตรอน (Structure Skeleton) เข้ามาประยุกต์เพื่อใช้ในการแปลความหมายภาพ ดังแสดงในภาพที่ 3 (ก) จากทฤษฎีของ Rudolph Arnheim (Arnheim, 1974 : 11-15) กล่าวไว้ว่า การรับรู้รวมถึงการสนใจภาพครั้งแรกของมนุษย์ นั้นจะรับรู้วัตถุที่เด่นก่อน โดยวัตถุที่เด่นนั้นจะเป็นวัตถุในแนวกึ่งกลางภาพ และจะต้องมีขนาดของวัตถุใหญ่เพียงพอ เพราะฉะนั้นวัตถุที่อยู่ด้านริมขอบภาพ หรือด้านข้างภาพจะรับรู้หรือสนใจเป็นส่วนถัดไปโดยคิดเป็นสัดส่วนลดหลั่นกันตามขนาดและตำแหน่งบนภาพ ดังแสดงในภาพที่ 3 (ข) ประกอบด้วย sea, sky, sand, kid, ball และ dog โดยที่ขนาดของวัตถุที่เป็น sea, sky และ sand จะมีพื้นที่มากที่สุด ตามลำดับ แต่เมื่อพิจารณาโดยใช้หลักการของ Rudolph Arnheim จะเห็นว่าเมื่อมนุษย์รับรู้ภาพ (human perception) ครั้งแรก จะเป็นวัตถุที่มนุษย์ให้ความสนใจและจะอยู่ตำแหน่งกึ่งกลางภาพและจะต้องมีขนาดที่เด่นพอ เพราะฉะนั้นจากภาพที่ 3 (ค) สิ่งที่มนุษย์จะสนใจมากที่สุดคือ kid เป็นลำดับแรกมากกว่าที่จะสนใจ sea ทั้งที่จะมีขนาดที่ใหญ่กว่ามาก เพราะฉะนั้น จากทฤษฎีข้างต้นจึงนำมาประยุกต์เพื่อที่จะคำนวณเป็นสมการเพื่อหาค่าของวัตถุที่ปรากฏบนภาพ และให้ค่าของวัตถุที่มีคุณลักษณะแตกต่างกันมีค่าของวัตถุแตกต่างกันด้วย แสดงสมการของการประมวลผลดังนี้

2.1.1 การคำนวณขนาดของวัตถุ (calculated object size) สำหรับการกำหนดขนาดของวัตถุ (object size: s_i) เป็นการนับจำนวนจุดพิกเซลทั้งหมดของแต่ละวัตถุที่ถูกแท็กไว้แล้วโดยที่ขนาดของวัตถุใหญ่ หรือ พื้นที่มากแสดงว่ามีจำนวนจุดพิกเซลมากกว่าวัตถุที่มีจำนวนพิกเซลน้อย

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

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

2.1.2 การคำนวณตำแหน่งวัตถุ (calculated object position) สำหรับการตำแหน่งวัตถุ (object position: p_i) จะใช้โครงสร้างสเกตตรีตรอนเพื่อทำการ mapping บนภาพเพื่อทำการคำนวณหาตำแหน่งของวัตถุและค่าน้ำหนักบนภาพ จากโครงสร้างที่กำหนดไว้ ดังแสดงในภาพที่ 3 (ข) เนื่องจากตำแหน่งของวัตถุที่อยู่จุดศูนย์กลางจะมีความสำคัญมากกว่า วัตถุที่อยู่ด้านริมหรือขอบภาพดังที่กล่าวมาแล้วข้างต้น ดังนั้นจึงได้กำหนดค่าของน้ำหนักบนโครงสร้างสเกตตรีตรอนของแต่ละเส้นถูกกำหนดเป็นค่าของ $\omega \in \{\omega_{l_1}, \omega_{l_2}, ..., \omega_{l_{|L|}}\}$ ที่สัมพันธ์กับเส้นบนโครงสร้าง $l \in \{l_1, l_2, ..., l_{|L|}\}$ สามารถเขียนสูตรได้ดังนี้



(ก) **(ข)** **(ค)**

ภาพที่ 3 การ Mapping ภาพด้วยโครงสร้างสเกตตรีตรอน (ก) โครงสร้างสเกตตรีตรอน (ข) ภาพตัวอย่างถูกแท็กด้วยคำศัพท์ (ค) ภาพถูก Mapping ด้วยโครงสร้างสเกตรอน

$$p_i = \sum_{region(x_i, y_i) \in l_j, j=1}^{o_n | L|} \omega_{l_j} \tag{2}$$

3. การคัดเลือกฟีเจอร์ (feature selection) เป็นการคัดเลือกข้อมูลหรือฟีเจอร์ที่มีประสิทธิภาพมาใช้ในการประมวลผล เนื่องจากข้อมูลมีจำนวนมากซึ่งคุณลักษณะบางตัวของข้อมูลไม่จำเป็นต้องใช้ เพราะฉะนั้นจะต้องมีการคัดเลือกข้อมูลที่ดีเพื่อ เพิ่มประสิทธิภาพการทำนาย รวมถึงการสังเคราะห์โมเดลได้รวดเร็ว และเพื่อลดความซับซ้อนของข้อมูล ได้ทำการเลือกอัลกอริทึม chi-

square และ information gain ซึ่งเป็นอัลกอริทึมที่วัดความสามารถของฟีเจอร์ทุกตัวเปรียบเทียบกัน

3.1 การคัดเลือกฟีเจอร์แบบ ไคส์-สแควร์

การคัดเลือกฟีเจอร์แบบ ไคส์-สแควร์ (chi-squared: χ^2) เป็นการคัดเลือกข้อมูลหรือฟีเจอร์โดยใช้การเปรียบค่าของฟีเจอร์ที่ได้จากการคำนวณของกลุ่มของข้อมูลทั้งหมด เพื่อทำการหาค่าคุณสมบัติของฟีเจอร์ ในแต่ละตัวว่ามีความสำคัญมากหรือน้อยกว่ากัน ค่าสุดท้ายที่ได้จะสามารถบอกลำดับความสำคัญของฟีเจอร์ได้ มีสูตรการคำนวณดังนี้

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{k}\frac{(A_{ij} - X_{ij})^2}{X_{ij}} \ , \tag{3}$$

เมื่อ $X_{ij} = \frac{R_i \times C_j}{N}$, โดยที่ m คือ จำนวนช่วงข้อมูล, k คือ จำนวนกลุ่มข้อมูล, A_{ij} คือ จำนวนกลุ่มข้อมูลตัวอย่างที่ i กลุ่มที่ j , R_i คือ จำนวนของข้อมูลในช่วงที่ i , C_j คือ จำนวนของข้อมูลในช่วงที่ j, N คือจำนวนรวมของข้อมูลทั้งหมด, ดังนั้น X_{ij} the expected frequency ของ A_{ij} ผลลัพธ์ที่ได้ออกมาค่า χ^2 ที่สูงสุดแสดงว่าฟีเจอร์ ตัวนั้นจะมีความสำคัญมากที่สุดในกลุ่มของข้อมูล

3.2 การคัดเลือกฟีเจอร์แบบ Information Gain ratio

Information Gain ratio (IGR) เป็นการคัดเลือกข้อมูลหรือฟีเจอร์โดยใช้การลดค่าของ entropy จากกลุ่มย่อยของข้อมูล (cluster) IGR ถูกนำมาใช้เป็นเครื่องมือเพื่อช่วยในการคัดเลือกฟีเจอร์ที่ดีที่สุดใน อัลกอริทึมแบบ C4.5 เป็นรูปแบบของ decision tree IGR แทนด้วย $gain\_r(x,C)$ คือค่าของ IGR ที่มีข้อมูลแอทริบิวต์ x ในกลุ่มย่อย C สามารถกำหนดสูตรของ IGR ได้ดังนี้

$$gain\_r(x,C) = \frac{gain(x,C)}{Split(C)}, \tag{4}$$

เมื่อให้ค่าของ $gain(x,C) = entropy(x,C) -$, $entropy_p(x,C)$ โดยที่

$$entropy(x,C) = \frac{-p(x|C)\log_2 p(x|C)}{-(1-p(x|C))\log_2(1-p(x|C))}, \tag{5}$$

$p(x|C) = \frac{freq(x,C)}{|C|}$, เมื่อ $freq(x,C)$ คือความถี่ที่เกิดขึ้นของข้อมูล x ใน C และ

$$entropy_p(x,C) = \sum_i \frac{|c_i|}{|C|} entropy(x,c_i), \tag{6}$$

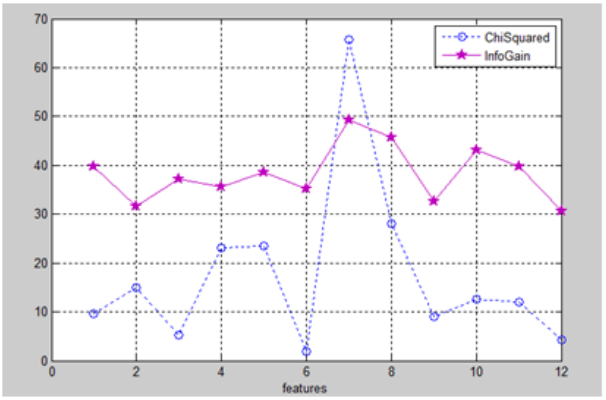
กลุ่มย่อย c_i คือเซตย่อยที่ i ใน C และ $|c_i|$ คือจำนวนของข้อมูลทั้งหมดใน c_i เมื่อ C คือกลุ่มข้อมูล

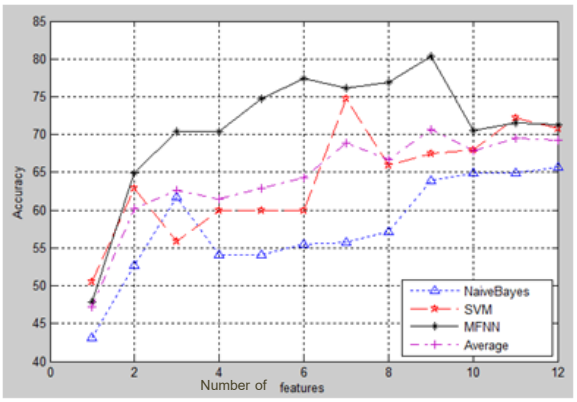
$$Split(C) = -\sum_{i=1} \frac{|c_i|}{|C|}\log_2 \frac{|c_i|}{|C|}, \tag{7}$$

ผลลัพธ์ที่ได้ออกมาค่าของ IGR ที่สูงสุดแสดงว่า ฟีเจอร์นั้นจะมีความสำคัญมากที่สุดในกลุ่มของข้อมูล เพราะฉะนั้นการที่นำฟีเจอร์ใดเข้ามาใช้ในการทดลองจึงควรที่จะหาความสำคัญของฟีเจอร์ก่อนเสมอ ผลที่ได้จากการใช้งานจะช่วยลดความซับซ้อนได้และไม่เปลืองค่าที่ใช้งานจริง

การคัดเลือกข้อมูล เป็นการคัดเลือกความสามารถของฟีเจอร์ ที่มีความสำคัญที่สุดมาใช้ในการทดลองต่อไป ซึ่งข้อมูลวัตถุ (O_i) ที่ได้จากการแท็กเบื้องต้นนั้นมีจำนวนถึง 12 ฟีเจอร์ ประกอบด้วย $\{f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}, f_{11}, f_{12}\}$ จึงต้องมีการคัดเลือกเฉพาะฟีเจอร์ที่มีความสำคัญโดยจะทำการทดลองเพื่อเปรียบเทียบโดยใช้ 2 วิธีการคือ chi-square และ information gain ผลการทดลองแสดงในภาพที่ 4 ก. แสดงข้อมูลแต่ละฟีเจอร์ต่างมีประสิทธิภาพในการจำแนกประเภทข้อมูลภาพที่ต่างกัน เมื่อนำค่า chi-square และ information gain มาเรียงลำดับความสำคัญจากมากไปน้อยจะเห็นว่า ลำดับของการเรียงตัวของการใช้ฟีเจอร์ มีลำดับที่เหมือนกันดังนี้ chi-square $F_{chi} = \{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}, f_1, f_9, f_3, f_{12}, f_6\}$ และinformation gain $F_{Info} = \{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}, f_1, f_9, f_3, f_{12}, f_6\}$

จากผลการคัดเลือกข้อมูลในภาพที่ 4 (ก) แสดงให้เห็นว่าข้อมูลในลำดับที่ 7 จะมีค่าของ chi-square และ information gain มีค่าสูงสุดเป็น 455.54 และ 65.7 $x10^2$ ตามลำดับ แสดงให้เห็นว่าข้อมูล เป็นข้อมูลที่มีผลต่อการจำแนกข้อมูลมากที่สุดจากจำนวนฟีเจอร์ทั้งหมด 12 ตัว และ ในลำดับถัดมาข้อมูลที่ 8 จะมีค่าของ chi-square 213.99 และ information gain 28.01x10^2 และถัดมาข้อมูลลำดับที่ 5 162.19 และ 23.54 x10^2 เป็นค่า chi-square และ information gain ตามลำดับ เมื่อได้ลำดับความสามารถของข้อมูลทั้งหมดแล้วนำฟีเจอร์ทั้งหมดทำการจำแนกข้อมูลภาพ ตามผลการทดลองข้างต้น โดยใช้ลักษณะการเรียงตัวของการคัดเลือก มาทดลองเพื่อเปรียบเทียบการจำแนกโดยใช้การตัวจำแนก (classifier) ทั้งหมด 3 วิธี คือ naïve-Bayes, Multiple Feedforward Neural Network (MFNN), Supporting Vector Machine (SVM) ได้ผลการทดลองดังภาพที่ 4 (ข) เป็นการจำแนกข้อมูลภาพ ตามจำนวนของข้อมูล (number of features) โดยเริ่มต้นที่จำนวน 1 ฟีเจอร์ คือ ข้อมูลในลำดับที่ 7 ถูกจำแนกข้อมูลด้วยวิธี naïve-Bayes ได้ค่าความถูกต้อง (accuracy) เพียง 43% จำแนกข้อมูลด้วย SVM ได้ค่าความถูกต้อง 50% และจำแนกข้อมูลด้วย MFNN ได้ค่าความถูกต้อง 47.9% เมื่อพิจารณาค่าเฉลี่ยของความถูกต้องเป็น 47% แต่เมื่อมีการรวมกันของ ข้อมูลทั้งหมดจนกระทั่งครบทั้ง 12 ฟีเจอร์ ได้จำแนกข้อมูลด้วยวิธี naïve-Bayes ได้ค่าความถูกต้องเพียง 65.7% จำแนกข้อมูลด้วย SVM ได้ค่าความถูกต้อง 70.7% และจำแนกข้อมูลด้วย MFNN ได้ค่าความถูกต้อง 71.3% และค่าเฉลี่ยของความถูกต้องเพียงแค่ 69% ซึ่งนับว่ามีค่าที่น้อย เพิ่มขึ้นถึง 11 ฟีเจอร์ได้รับค่าความถูกต้องเพิ่มขึ้นเพียง 22% เท่านั้น ได้ทำการทดลองเพื่อพิจารณาค่าความถูกต้อง โดยใช้การเรียงลำดับฟีเจอร์ของวัตถุ (O_i) ฟีเจอร์ในลำดับที่ 7 (f_7) เป็นฟีเจอร์ที่ได้ถูกคัดเลือกมาให้เป็นฟีเจอร์ตัวแรกที่ใช้ในการทดลองเนื่องจากเป็นตัวที่สามารถแยกแยะกลุ่มของภาพได้มากที่สุด โดยที่ค่าความถูกต้องเฉลี่ยที่ได้เป็น 47.2% ซึ่งเป็นเกือบครึ่งหนึ่งของการจำแนกข้อมูลทั้งหมด





(ก) (ข)

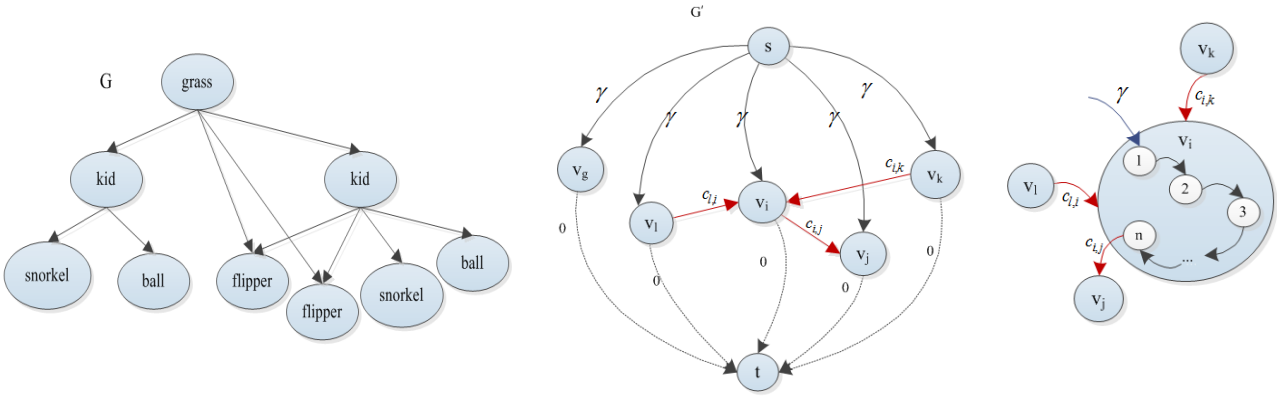
ภาพที่ 4 ผลลัพธ์การคัดเลือกฟีเจอร์ (ก) ผลลัพธ์การคัดเลือก 12 ฟีเจอร์ (ข) ผลลัพธ์ของการจำแนกข้อมูลตามจำนวนฟีเจอร์

เมื่อพิจารณาการรวมตัวของฟีเจอร์ที่ 7 9 และ 11 ฟีเจอร์จะได้ค่าความถูกต้องเป็น 68.9% 70.6% และ 69.5% ตามลำดับ ดังนั้นจึงเลือกใช้จำนวนฟีเจอร์เพียง 7 ฟีเจอร์ที่ประกอบด้วย $\{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}\}$ เพราะการใช้ข้อมูลถึง 11 ฟีเจอร์เป็นการใช้จำนวนของข้อมูลที่มากเกินไปจาก 12 ฟีเจอร์ และเมื่อเปรียบเทียบแล้วจะมีค่าความถูกต้องของการจำแนกด้วย SVM และ MFNN ที่น้อยกว่า เมื่อใช้เครื่องมือจำแนกด้วย naïve-Bayes จะได้ค่าความถูกต้องเพียง 55.7% แต่ใช้เครื่องมือจำแนกข้อมูลด้วย SVM ได้ค่าความถูกต้อง 74.7% ส่วนการจำแนกข้อมูลด้วย MFNN ได้ค่าความถูกต้องเป็น 76.2% แต่เมื่อพิจารณาถึงค่าเฉลี่ยของความถูกต้องจะได้เพียง 68.9% สำหรับการทดลองแบบที่มีการรวมกันของข้อมูลถึง 9 ฟีเจอร์ที่ประกอบด้วย $\{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}, f_1, f_9\}$ จะได้ค่าความถูกต้องของ 63.9% 67.5% และ 80.4% โดยที่มีการใช้เครื่องมือจำแนกข้อมูลด้วย naïve-Bayes SVM และ MFNN ตามลำดับซึ่งมีการคิดเป็นค่าเฉลี่ยของความถูกต้องได้ 70.6% ซึ่งเป็นค่าที่มากกว่าการรวมกันโดยใช้ฟีเจอร์แบบ 7 ตัวดังนั้นเมื่อพิจารณาอย่างละเอียดการรวมกันของข้อมูลทั้งหมด 7 ฟีเจอร์จะได้ค่าเฉลี่ยของความถูกต้องอยู่ที่68.9% และ 9 ฟีเจอร์ค่าเฉลี่ยของความถูกต้องอยู่ที่ 70.6% ซึ่งมีค่าเฉลี่ยความถูกต้องห่างกันอยู่เพียง 1.8% แต่จะมีความต่างกันอยู่ถึง 2 ฟีเจอร์ดังนั้นจึงเลือกใช้จำนวนเพียง 7 ฟีเจอร์ ประกอบด้วย $\{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}\}$

การแทนวัตถุลงในกราฟลำดับชั้น

ข้อมูลภาพที่ได้ถูกแปลงเป็นข้อมูลวัตถุเป็น ฟีเจอร์ลงในตัวแปรที่เป็นเมตริกซ์ กำหนดให้ $X = [X_1^T, X_2^T, ..., X_N^T] \in R^{N \times J_r}$ มีขนาดเมตริกซ์เป็น $N \times J_r$ เมื่อ N แทนจำนวนภาพ r แทนขนาดมิติของเมตริกซ์ และ J เป็นจำนวนของวัตถุบนภาพ กำหนดให้ $Y = [y_1, y_2, ..., y_N]$ เป็นเวกเตอร์ของวัตถุ

ที่เก็บค่าการจำแนกประเภทของ N ภาพข้อมูลวัตถุภายในภาพที่ถูกจัดเก็บลงในเมตริกซ์จะถูกสร้างความสัมพันธ์ระหว่างวัตถุด้วยแนวคิดกราฟ (conceptual graph) กำหนดให้รูปภาพใดใด สามารถเขียนความสัมพันธ์ของแนวคิดกราฟ เป็นสเปเชียลเอ็นทิตี้ (spatial entities) ที่แทนด้วยเซตของ V คือจุด (vertex) หรือโหนด (node) เมื่อ $i \in \{1...N\}$ และความสัมพันธ์ของวัตถุภายในภาพเชื่อมต่อกันด้วย E : $E \subset V \times V$, เมื่อ E คือความสัมพันธ์ระหว่างโหนดสองโหนด (edge) ดังนั้นสามารถแทนความสัมพันธ์ของสองจุดได้ด้วย $e_{ab} \equiv (v_a, v_b) \in E$ เมื่อกำหนดให้ $v_a, v_b \in V$ จากตัวอย่างภาพที่ 2 แสดงภาพที่ได้จาก LabelMe (Torralba, et al., 2010) วัตถุคำหลักมีดังนี้ kid kid grass ball ball snorkel snorkel flipper ดังแสดงในภาพที่ 4 (ก) และ (ข) เป็นการแทนความสัมพันธ์วัตถุด้วยโหนดบนกราฟดังนั้น grass เป็น root node ที่มีความสัมพันธ์ร่วมกันกับ kid ทั้งสองและ โหนดของ kid จะเกิดความสัมพันธ์ต่อไปยังส่วนต่างๆจะถูกเชื่อมเข้าด้วยกันเป็นค่าของ E ขั้นตอนการประมวลผลจะนำข้อมูลที่มีการจัดเก็บไว้ประมวลผล ด้วยแนวคิดกราฟ และความสัมพันธ์ภายในแบบลำดับชั้น



(ก) **(ข)** **(ค)**

ภาพที่ 5 ตัวอย่างความสัมพันธ์ของค่าน้ำหนักภายในกราฟ (ก) ความสัมพันธ์ของวัตถุด้วยแนวคิดกราฟ (ข) ค่าน้ำหนักระหว่างโหนด s และ t (ค) ความสัมพันธ์ภายในกราฟย่อย

แบ่งการประมวลผลเพื่อทดสอบค่าความถูกต้องและผลลัพธ์ที่ได้ออกเป็น 2 ส่วนดังนี้

1. การกำหนดค่าน้ำหนักวัตถุ สำหรับข้อมูลวัตถุบนภาพ โดยคิดค่าน้ำหนักจากความสัมพันธ์ข้อมูลวัตถุที่เกิดขึ้นทั้งหมดภายในภาพ X_N^T โดยจะเก็บเป็นค่า $\mathcal{I}(v_i)$ (Information Content) แทนรายละเอียดข้อมูลดังนั้นสามารถนับจำนวนคำหลักที่เกิดขึ้นเป็นค่าของความน่าจะเป็นของแต่ละคำหลักบนภาพได้ด้วยสมการดังนี้

$$P(v_i) = freq(v_i) \Big/ \sum_i freq(v_i), i \in \{1,..., N\} \quad (8)$$

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

$$freq(v_i) = \sum_{n \in word(v)} count(n) \tag{9}$$

ดังนั้นเมื่อมีการคำนวณในแต่ละโหนดบนภาพสามารถเขียนเป็นสมการของ $\vartheta(v_i)$ ใน v_i ได้ดังนี้

$$\vartheta(v_i) = \log^{-1} P(v_i) \tag{10}$$

2. การหาความสัมพันธ์ของสายเชื่อมโยงข้อมูล ภายในภาพจะมีวัตถุที่ถูกแทนเป็นโหนดได้หลายโหนดแต่ละโหนดถูกเชื่อมโยงด้วย edge ระหว่างกันเป็นทอดๆดังแสดงในภาพที่ 5 (ข) เมื่อกำหนดให้ s เป็นโหนดที่มีความสัมพันธ์กับโหนดพ่อแม่ (parent node) มีโหนดลูก (children node) คือ v_i, v_j, v_k, v_l และ v_g ดังนั้นความสัมพันธ์ที่เกิดขึ้นของ v_i นั้นเกิดขึ้นร่วมกันหลายโหนด สามารถหาความสัมพันธ์ที่เกิดขึ้นด้วยการคำนวณเป็นค่าของ link strength (LS) การประมวลผลจะเลือกจากจำนวนโหนดที่การติดต่อกันในกราฟย่อยที่มีจำนวนน้อยที่สุดจากกราฟ โดยเลือกจากโหนด s ที่เชื่อมโยงไปโหนด t ดังนั้น กำหนดให้

$$G' = (V', E') \tag{11}$$

เป็นกราฟใหม่ที่มีโหนด s เชื่อมโยงไปโหนด t ดังนั้น $V' = V \cup \{s,t\},$ และ $E' = E \cup \{(s,v) : v \in V\} \cup \{(u,t) : u \in V\}$ โดยที่ s และ t เป็นส่วนเชื่อมโยงไปยังวัตถุ ค่าน้ำหนักระหว่างโหนด s และโหนดอื่นๆ สามารถแทนค่าได้ด้วย γ เป็นฟีเจอร์ที่ใช้เก็บเส้นทางดังแสดงในภาพที่ 5 (ค) การแก้สมการจะแก้ได้ด้วยการหาค่าต่ำสุดของความสัมพันธ์ภายในกราฟ G' จากความสัมพันธ์ระหว่างโหนด s และ t สามารถเขียนสมการได้ดังนี้

$$\Omega(\beta) = \min_{f \in F} \{ \sum_{(u,v) \in E'} f_{uv} c_{uv} s.t s_j(f) \geq |\beta_j|, \ \forall_j \in \{1,..., J_r\} \}, \ s_j(f) = \sum_{u \in V' : (u,j) \in E'} f_{uj} \tag{12}$$

สำหรับค่าน้ำหนักที่เกิดขึ้นของเวอร์เท็กซ์ j ใน $V = 1,2,..., J_r$ ทุกเวอร์เท็กซ์จะมีการแทนค่าด้วยค่าน้ำหนัก เมื่อกำหนดให้ $[c_{u,v}]$ เป็นค่าน้ำหนักของเส้นเชื่อมระหว่าง u และ v เมื่อ $[c_{u,v}]_{(u,v) \in E'}$ และ F เป็นเซตของเส้นทางการไหลของการเชื่อมโยงบนกราฟ G' วัตถุที่ถูกเลือกจะสามารถเรียกดูค่าได้จากการไหลของ f_{uv} จากทุกๆเส้นของ (u,v) บนกราฟ

$$f^* \in \arg\min_{f \in F} \{ \sum_{(u,v) \in E'} f_{uv} c_{uv} + \sum_{j=1}^{J_r} \tfrac{1}{2} \max(|u_j| - s_j(f), 0)^2 \}, \tag{13}$$

เมื่อแทน u เป็นค่าใน R และ F เป็นการไหลข้อมูลบนกราฟ G'

การวัดประสิทธิภาพการจำแนกความหมายภาพ

ในงานวิจัยนี้ได้ใช้การวัดประสิทธิภาพการจำแนกภาพด้วย ระดับความแม่นยำ (Precision) และการเรียกคืน (Recall) และ F_1 (F-measure) โดยมีรายละเอียดสมการดังนี้

$$\Pr ecision = \frac{A}{A+B} \times 100\% \tag{14}$$

$$\mathrm{Re}call = \frac{A}{A+C} \times 100\% \tag{15}$$

$$F_1 = \frac{2 \times \Pr ecision \times \mathrm{Re}call}{(P\,\mathrm{Re}cision + \mathrm{Re}call)} \tag{16}$$

เมื่อ A แทนจำนวนภาพที่จำแนกได้ถูกต้อง B แทนจำนวนรูปภาพที่จำแนกได้แต่ไม่ถูกต้อง และ C แทนจำนวนภาพที่ถูกต้องแต่ไม่สามารถจำแนกได้

ผลการจำแนกความหมายของข้อมูลภาพ

ข้อมูลภาพในการทดลองได้ใช้ฐานข้อมูลที่เก็บรวบรวมและมาจากแอพพลิเคชัน LabelMe ได้มีการคัดเลือกภาพสำหรับการทดลองให้อยู่ในหมวดหมู่ของภาพภายในบ้าน (indoor) และภาพภายนอกบ้าน (outdoor) โดยได้กำหนดตามความหมายพื้นฐานจากการสุ่มและตัดสินใจจากการจำแนกความหมายภาพด้วยคนเป็นหลัก (human scenes classification) (Xiao, 2010) จึงได้มีการแบ่งกลุ่มภาพย่อยในเป็น 7 กลุ่มเพื่อทดสอบการจำแนกความหมายของข้อมูลภาพ ประกอบด้วยกลุ่มของ ภาพสำนักงาน (office), ภาพห้องนั่งเล่น (living room), ภาพเมือง (city), ภาพห้องครัว (kitchen), ภาพชายฝั่ง (coast), ภาพภูเขา (mountain), ภาพป่า (forest) คำหลักที่ถูกแท็กไว้บนภาพจากผู้ใช้ใน LabelMe โดยกำหนดให้จำกัดขอบเขตของคำหลักที่ใช้ในการทดลองนี้รวมทั้งหมด 95 คำที่แตกต่าง ภาพรวมทั้งหมดที่ใช้ในการทดลอง 1,500 ภาพ จำแนกด้วย เครือข่ายแบบเบย์ (Bayesian Network) และ วิธีซัพพอร์ตเวคเตอร์แมชชีน (Support Vector Machines) การทดลองแบ่งออกเป็น 3 ส่วนดังนี้

1. การจำแนกความหมายภาพด้วยข้อมูลจากการสกัดจากโครงสร้างสเกลตรีตรอน

การทดลองจะใช้ข้อมูลจากฟีเจอร์ที่ได้มาจากการคำนวณหาค่าน้ำหนักวัตถุเด่นบนภาพจากคำหลักที่ถูกสกัดมากจากโครงสร้างสเกลตรีตรอน แต่ละเซตจะถูกจัดเก็บตามความสำคัญเรียงตามขนาดของวัตถุที่โดดเด่นบนภาพ (o_i)และตำแหน่งของวัตถุ (p_i) เพื่อจำแนกกลุ่มภาพเดียวกันด้วยวิธี naïve-Bayes และ SVM แสดงผลลัพธ์ในตารางที่ 1

2. การจำแนกความหมายภาพด้วยฟีเจอร์ที่ถูกเลือก

การทดลองใช้ข้อมูลฟีเจอร์ที่ได้จากค่าน้ำหนักวัตถุเด่นบนภาพและตำแหน่งวัตถุจากคำหลักที่ถูกสกัดมากจากโครงสร้างสเกลตรีตรอน และจากฟีเจอร์ที่ถูกคัดเลือกมาทั้ง 7 ฟีเจอร์ประกอบด้วย $\{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}\}$ เพื่อทำการจำแนกกลุ่มภาพเดียวกันด้วยวิธี naïve-Bayes และ SVM แสดงผลลัพธ์ในตารางที่ 2

จากผลการทดลองการจำแนกความหมายภาพเบื้องต้นที่มีการใช้เพียงค่าน้ำหนักวัตถุเด่นบนภาพนั้นจะได้ค่าความถูกต้องโดยเฉลี่ยเพียงครึ่งหนึ่งเท่านั้น ดังแสดงในตารางที่ 1 จะเห็นว่า เมื่อมีการใช้เพียงฟีเจอร์วัตถุที่โดดเด่นบนภาพ (o_i)และตำแหน่งของวัตถุ (p_i) ในการจำแนก กลุ่มภาพ forest จะได้ค่าความถูกต้อง 49.0% ด้วยวิธีการจำแนกแบบ naïve-Bayes และ 50.5% ด้วย SVM กลุ่มภาพ city จะได้ค่าความถูกต้อง 47.5% ด้วยวิธีการ

ตารางที่ 1 ผลลัพธ์การจำแนกความหมายภาพจากการสกัดจากโครงสร้างสเกลตรีตรอน

Scene/Categories		Performance (%)					
		naïve_Bayes			SVM		
		Precision	Recall	F_1	Precision	Recall	F_1
Indoor	Office	40.6	41.8	41.2	42.6	46.2	44.3
	Living room	47.1	44.0	45.5	46.1	43.5	44.8
	Kitchen	43.0	45.7	44.3	44.0	46.3	45.1
Outdoor	Coast	49.5	42.5	45.7	50.5	45.6	47.9
	Forest	47.5	50.5	49.0	49.5	51.6	50.5
	Mountain	42.6	46.7	44.6	43.6	44.9	44.2
	City	47.5	47.5	47.5	56.6	54.9	55.7
Accuracy		45.39			47.52		

ตารางที่ 2 ผลลัพธ์การจำแนกความหมายภาพด้วยฟีเจอร์ที่ถูกเลือก $\{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}\}$

Scene/Categories		Performance (%)					
		naïve_Bayes			SVM		
		Precision	Recall	F_1	Precision	Recall	F_1
Indoor	Office	57.4	56.9	57.1	59.4	60.0	59.7
	Living room	53.9	52.4	53.1	56.4	53.8	55.1
	Kitchen	52.0	52.5	52.3	56.0	53.3	54.6
Outdoor	Coast	53.4	53.9	53.7	58.4	56.7	57.6
	Forest	57.6	57.0	57.3	57.6	62.0	59.7
	Mountain	53.5	57.4	55.4	53.5	58.7	56.0
	City	59.6	57.3	58.4	60.4	58.1	59.2
Accuracy		**55.32**			**57.39**		

จำแนกแบบ naïve-Bayes และ 55.7% ด้วย SVM แต่เมื่อทดลองจำแนกภาพด้วยข้อมูลค่าน้ำหนักวัตถุภาพเด่นด้วยการเพิ่มฟีเจอร์ที่ถูกคัดเลือกมาอีก 7 ฟีเจอร์ $\{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}\}$ ผลที่ได้คือในกลุ่มภาพ city จะได้ค่าความถูกต้องถึง 59.2% ด้วย SVM ในตารางที่ 2 และกลุ่ม office ได้ค่าความถูกต้องถึง 59.7% ด้วย SVM แต่กลุ่มเดียวกันจำแนกด้วย naïve-Bayes จะได้ค่าความถูกต้องเป็น 57.1% จะเห็นว่าเมื่อมีการใช้จำนวนฟีเจอร์เพิ่มมากขึ้นทำให้ค่าความถูกต้องเพิ่มขึ้น

3. การจำแนกความหมายภาพด้วยแนวคิดกราฟแบบลำดับชั้น

การทดลองใช้ข้อมูลฟีเจอร์ที่ได้จากค่าน้ำหนักวัตถุเด่นบนภาพและตำแหน่งวัตถุจากคำหลักที่ถูกสกัดมากจากโครงสร้างสเกลตรีตรอน รวมทั้งจากฟีเจอร์ที่ถูกคัดเลือกมาทั้ง 7 ฟีเจอร์ $\{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}\}$ และใช้การแทนค่าข้อมูลด้วยแนวคิดกราฟแบบลำดับชั้นเพื่อเพิ่มความสัมพันธ์ของโครงสร้างภายในภาพ และเพิ่ม การเปรียบเทียบการจำแนกความหมายภาพด้วย SVM แบบเชิงเส้น (Linear) ด้วย Linear kernel และได้เลือกแบบ วิธีการเปรียบเทียบการจำแนกแบบ One-against-one และ One-against-all ได้มีการใช้ข้อมูลฟีเจอร์ที่เกิด ซึ่งภาพแต่ละกลุ่มความหมายจะมีความเหมือนที่เกิดขึ้นภายในกลุ่มซึ่งจะถูกแทนค่าเหล่านั้นได้ด้วยกราฟแบบลำดับชั้น ดังแสดงผลการทดลองในตารางที่ 3 จะเห็นว่าการจำแนกด้วยวิธี Linear SVM จะได้ค่าความถูกต้องถึง 80.28% ด้วย One-against-one และ 76.17% ด้วย One-against-all ภาพกลุ่ม city ที่มีการจำแนกด้วย One-against-one ได้ค่า F1 ถึง 86.3% มีค่า Precision 88.9% Recall 83.8% แต่ One-against-all จะได้ค่า F1 เป็น 79.6% มีค่า Precision 80.8% Recall 78.4% ดังแสดงผลลัพธ์ในภาพที่5

ตารางที่ 3 ผลลัพธ์การจำแนกความหมายภาพแนวคิดกราฟแบบลำดับชั้น $\{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}\}$

Scene/Categories		Performance (%)								
		naïve_Bayes			SVM One-against-one			SVM One-against-all		
		Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1
Indoor	Office	76.8	81.7	79.2	83.2	84.0	83.6	73.3	74.0	73.6
	Living room	61.8	74.1	67.4	79.4	78.6	79.0	70.6	73.5	72.0
	Kitchen	68.7	61.3	64.8	78.0	79.6	78.8	75.0	71.4	73.2
Outdoor	Coast	60.9	55.4	58.0	80.6	81.4	81.0	74.8	77.8	76.2
	Forest	62.7	64.6	63.7	73.7	73.7	73.7	78.8	77.2	78.0
	Mountain	59.8	62.7	61.2	78.2	80.6	79.4	80.2	81.0	80.6
	City	72.8	66.4	69.4	88.9	83.8	86.3	80.8	78.4	79.6
Accuracy		66.2			80.28			76.17		

สรุปผลการทดลอง

จากการทดลองที่มีการใช้ในส่วนของการสกัดข้อมูลจากโครงสร้างสเกตรีตรอนและการคัดเลือกฟีเจอร์ที่มีความสำคัญเพื่อนำมาใช้เป็นข้อมูลพื้นฐานในการจำแนกภาพและยังใช้แนวคิดกราฟและความสัมพันธ์ของวัตถุภายในภาพเพื่อค้นหาความสัมพันธ์ที่มีการแฝงตัวอยู่ตามกลุ่มความหมายภาพ ข้อมูลที่ได้นั้นทำให้ช่วยการจำแนกความหมายภาพเพิ่มมากขึ้น ทำให้ค่าความถูกโดยเฉลี่ยเมื่อเปรียบเทียบกับประสิทธิภาพที่จำแนกด้วย SVM One-against-one จะเห็นว่าค่าความถูกต้องโดยเฉลี่ยเพิ่มขึ้นถึงประมาณ 80.28% ดังแสดงตัวอย่างผลการจำแนกความหมายภาพตามกลุ่มไว้ในภาพที่ 6 แต่อย่างไรก็ตามควรมีการปรับปรุงงานวิจัยนี้ให้สมบูรณ์มากขึ้นด้วยการแบ่งกลุ่มภาพให้หลากหลายและควรมีคำหลักเพิ่มขึ้น ถ้าเป็นกลุ่มภาพมนุษย์ควรมีคำหลักเฉพาะที่บออกถึงกิจกรรมของมนุษย์เพิ่มขึ้นด้วย

ภาพที่ 6 ตัวอย่างผลลัพธ์ของการจำแนกความหมายภาพ (ก) ภาพสำนักงาน (ข) ภาพห้องนั่งเล่น (ค) ภาพห้องครัว (ง) ภาพชายฝั่ง (จ) ภาพป่า (ฉ) ภาพภูเขา (ช) ภาพเมือง

เอกสารอ้างอิง

Arnheim, Rudolph. (1974). **Art and visual perception : a psychology of the creative eye**. Calif. : University of California Press**.**

Galleguillos, C. and Belongie, S. (2010). Context based object categorization: a critical survey. **Computer Vision and Image Understanding**, 114, 712-722.

Hiremath, P.S., Akkasaligar, Prema T. and Badiger, Sharan. (2007). Comparison of wavelet based despeckling of medical ultrasound images. In **International Conference on Advances in Computer Vision and Information Technology** (pp. 1026-1031).

Li, Jia and Wang, James Z. Wang. (2003). Automatic linguistic indexing of pictures by a statistical modeling approach. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 25(9), 1075-1088.

Mezaris, Vasileios, Kompatsiaris, Ioannis and Strintzis, Michael G. (2003). An ontology approach to object-based image. In **Proceedings of the IEEE International Conference on on Image Processing**.

Russell, B.C., Torralba, A., Murphy, K.P. and Freeman, W.T.. (2008). LabelMe : a database and web-based tool for image annotation. **International Journal of Computer Vision**, 77(1-3), 157–173.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Content-based image retrieval at the end of the early years. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 22(12).

Torralba, A., Russell, B.C. and Yuen, J. (2010). LabelMe : online image annotation and applications. In **Proceedings of the IEEE**, 98(8), 1467–1484.

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

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