สำนักงานคณะกรรมการวิจัยแห่งชาติ
ห้องสมุดงานวิจัย

บทที่ 1.
บทนำ

ปัจจุบันปริมาณภาพดิจิตอลมีปริมาณเพิ่มขึ้นเป็นจำนวนมากเนื่องจากการเปลี่ยนมาใช้กล้องดิจิตอล, ราคาของอุปกรณ์ข้อมูลดิจิตอลที่ถูกลง และการปริมาณภาพดิจิตอลในอินเตอร์เน็ตที่เพิ่มขึ้น ทำให้จำเป็นต้องพัฒนาระบบค้นหารูปภาพที่มีประสิทธิภาพและรวดเร็วเพื่อค้นหาภาพในฐานข้อมูล

การพัฒนาระบบค้นหาภาพสามารถย้อนหลังไปถึงช่วงหลังขอทศวรรษที่ 1970 ระบบดังกล่าวใช้ข้อความในการบรรยายความหมายของภาพและใช้ข้อความในการค้นหาภาพที่อยู่ในฐานข้อมูล [1-3] เพื่อให้ค้นหาภาพได้อย่างเที่ยงตรงจำเป็นต้องบรรยายความหมายของภาพอย่างถูกต้องและสมบูรณ์ แต่เทคโนโลยีที่มีในปัจจุบันยังไม่สามารถพัฒนาระบบอัตโนมัติสำหรับสร้างข้อความบรรยายความหมายของภาพได้ ดังนั้นจึงต้องใช้มนุษย์ในการกรอกข้อความบรรยายความหมายของแต่ละภาพ ซึ่งเป็นงานที่ต้องใช้แรงงานเป็นจำนวนมาก ใช้เวลานาน และสิ้นเปลืองเงิน และผลลัพธ์ที่ได้มักจะขึ้นกับความเห็นของผู้กรอกข้อความและไม่สมบูรณ์ ดังนั้นการค้นหาภาพโดยใช้ข้อความจึงไม่สามารถนำมาใช้งานได้อย่างกว้างขวาง

มันค่อนข้างจะชัดเจนว่าการค้นหาภาพที่มีประสิทธิภาพ ควรจะเป็นการค้นหาโดยพิจารณาจากคุณสมบัติที่อยู่ภายในภาพ นักวิจัยจากหลากหลายกลุ่มเช่น Computer Vision, Database Management, Human-Computer Interface และ Information Retrieval จึงได้หันมาพัฒนาระบบค้นหารูปภาพโดยรายละเอียดของภาพ [4-9]

1.1 ปัญหาของระบบค้นหารูปภาพโดยรายละเอียดของภาพ

ปัญหาของการค้นหารูปภาพโดยรายละเอียดของมีอยู่หลายอย่างที่จำเป็นจะต้องแก้ไขก่อนที่จะสามารถนำไปใช้งานจริงได้ ดังนี้

1.1.1 การทำความเข้าใจรายละเอียดภาพของมนุษย์

เนื่องจากมนุษย์เป็นผู้ใช้ระบบค้นหารูปภาพ ดังนั้นจึงมีความจำเป็นอย่างมากที่จะต้องศึกษาวิธีที่มนุษย์ทำความเข้าใจรายละเอียดของภาพ งานวิจัยในหัวข้อนี้ได้รับความสนใจ

สำนักงานคณะกรรมการวิจัยแห่งชาติ
ห้องสมุดงานวิจัย
วันที่ 2 8 ก.ย. 2555
เลขทะเบียน 246420

เป็นอย่างมากในปัจจุบัน โดยมุ่งไปที่การศึกษาวิธีการที่มนุษย์ทำความเข้าใจรายละเอียดของภาพและวิธีการรวมแบบจำลองของวิธีการทำความเข้าใจเข้ากับระบบค้นหารูปภาพ [13-20]

1.1.2 คุณลักษณะและการวัดความเหมือนของภาพ

รายละเอียดของภาพสามารถถูกบรรยายโดยสี, รูปร่าง, พื้นผิว และการจัดวางองค์ประกอบของภาพ ในระบบการค้นหารูปภาพการบรรยายดังกล่าวสามารถสร้างจากคุณลักษณะพื้นฐานของภาพเช่น Color Histogram, Tamura Feature หรือ Fourier Descriptor คุณลักษณะของภาพที่ดีจะต้องมีคุณสมบัติดังนี้ ถ้ามีภาพสองภาพที่คล้ายกัน คุณลักษณะของภาพทั้งสองจะต้องอยู่ใกล้กันด้วย

นอกจากนี้ความเปลี่ยนแปลงต่างๆในการถ่ายภาพยังทำให้เกิดปัญหาในระบบการค้นหาภาพ ความเปลี่ยนแปลงของการถ่ายภาพที่มีปัญหามีดังต่อไปนี้

1. การหมุนวัตถุ
2. อัตราส่วนภาพ
3. การเลื่อนตำแหน่งกล้อง
4. มุมกล้อง
5. ทิศทางของแสง
6. ความเข้มของแสงและเงา
7. ความเข้มของสี
8. การบดบังวัตถุในภาพ
9. ความสับสนระหว่างวัตถุกับพื้นหลัง
10. วัตถุอื่นๆที่อยู่ในกลุ่มเดียวกัน

ดังนั้นคุณลักษณะของภาพที่ดีจะต้องไม่ขึ้นกับความเปลี่ยนแปลงต่างๆดังกล่าว แต่อย่างไรก็ตามคุณลักษณะที่ดีจะต้องรักษาสมดุลระหว่างการไม่ขึ้นกับการเปลี่ยนแปลงกับการแยกแยะความแตกต่าง เนื่องจากถ้าคุณลักษณะไม่ขึ้นกับความเปลี่ยนแปลงมากเกินไปจะทำให้ไม่สามารถแยกแยะความแตกต่างที่มีความสำคัญได้

นอกจากนี้งานที่สำคัญมากอีกอย่างของระบบการค้นหารูปภาพคือการเปรียบเทียบความเหมือนระหว่างสองภาพ ดังนั้นปัญหาคือวิธีการวัดความเหมือนของภาพสองภาพ ผลลัพธ์ของการวัดความเหมือนจะต้องตรงกับวิธีการวัดความเหมือนของมนุษย์

1.1.3 ความเชื่อมโยงระหว่างคุณลักษณะพื้นฐานของภาพและแนวความคิดชั้นสูง

มนุษย์มักจะใช้มักนิยมใช้แนวความคิดชั้นสูงในทุกๆวัน ตัวอย่างแนวความคิดแบบง่ายๆคือ รถ บ้าน ต้นไม้ และเครื่องบิน จากแนวความคิดแบบง่ายๆก็พัฒนาเป็นแนวความคิดที่ซับซ้อนเช่น เมือง กีฬา สัตว์ และป่า

แต่อย่างไรก็ตามเทคนิคในงาน Computer Vision ทั่วไปในปัจจุบันสามารถทำได้แค่ดึงคุณลักษณะพื้นฐานของภาพได้เท่านั้น ในงานบางอย่างเช่นระบบจดจำหน้าและการตรวจลายนิ้วมือที่สามารถเชื่อมโยงระหว่างคุณลักษณะพื้นฐานกับแนวความคิดชั้นสูงได้ แต่ไม่สามารถทำได้ในกรณีทั่วไป มันเป็นการยากมากที่จะเชื่อมโยงคุณลักษณะพื้นฐานกับแนวความคิดชั้นสูง

เพื่อลดช่องว่างในการแปลความหมาย (Semantic Gap) จำเป็นต้องใช้เทคนิคที่ รวมคุณลักษณะหลายๆอย่างไว้ด้วยกัน ตัวอย่างเช่นควรจะรวมคุณลักษณะทางสีและพื้นผิวเข้ากับวงกลม เพื่อแยกความแตกต่างระหว่างลูกบอลและจาน หรือดวงอาทิตย์ ขั้นต่อไปจำเป็นต้องใช้การประมวลผลแบบ Off-line และ On-line การประมวลผลแบบ Off-line สามารถทำได้โดยใช้เทคนิค Supervised Learning, Unsupervised Learning หรือรวมทั้งสองวิธี เครื่องมือในการเรียนรู้เหล่านี้คือ Neural network, Genetic Algorithm และ Clustering สำหรับการประมวลผลแบบ On-line จำเป็นต้องใช้การติดต่อกับผู้ใช้ที่มีความฉลาดและเป็นมิตร มันจะต้องให้ผู้ใช้สามารถประเมินผลลัพธ์ของการค้นหารูปภาพ เทคนิคดังกล่าวคือ Relevance Feedback

1.1.4 ระบบติดต่อผู้ใช้

ความแตกต่างระหว่างระบบจดจำรูปแบบและระบบค้นหารูปภาพคือผู้ใช้มีส่วนสำคัญอย่างมากในระบบที่สอง ในระบบค้นหารูปภาพ การค้นหารูปภาพเป็นปัญหาที่ไม่ชัดเจน ขึ้นกับผู้ใช้แต่ละคน คนสองคนหรือคนเดียวกันแต่ต่างเวลาอาจจะมองภาพเดียวกันในมุมมองที่แตกต่างกันได้ เราเรียกปัญหานี้ว่า ความเห็นส่วนตัวของการรับรู้ของมนุษย์ (Human Perception Subjectivity) ความเห็นส่วนตัวนี้เกิดขึ้นในหลายระดับ ตัวอย่างเช่นคน

หนึ่งอาจจะสนใจเฉพาะสีของภาพ ในขณะที่อีกคนอาจจะสนใจรูปร่างเป็นต้น ถึงแม้ว่าทั้งสองคนอาจจะสนในรูปร่างแต่การรับรู้รูปร่างของแต่ละคนก็ค่อนข้างแตกต่างกัน

1.1.5 ตัวชี้ในมิติชั้นสูง

เพื่อให้ระบบค้นหารูปภาพสามารถใช้ได้กับฐานข้อมูลขนาดใหญ่ จำเป็นต้องใช้เทคนิคตัวชี้หลายมิติ เป้าหมายของการใช้ตัวชี้คือเพื่อให้การค้นหาข้อมูลทำได้รวดเร็ว มีปัญหา 2 อย่างในระบบการค้นหารูปภาพคือ

1. ปัญหามิติชั้นสูง : โดยปกติจำนวนมิติของเวคเตอร์คุณลักษณะจะมีค่าประมาณ 10^2 [11] แต่อย่างไรก็ตามโครงสร้างของตัวชี้ที่ถูกนำเสนอจะสามารถทำงานได้อย่างมีประสิทธิภาพเมื่อจำนวนมิติมีขนาดเล็กมาก [32]
2. การวัดความเหมือนในอวกาศที่ไม่ใช่แบบยูคลิเดียน : เนื่องจากการวัดในอวกาศแบบยูคลิเดียนอาจจะไม่ใช่การวัดที่มีประสิทธิภาพมากที่สุด จึงจำเป็นต้องพัฒนาวิธีการวัดในแบบอื่นๆเช่น Histogram Intersection, Cosine, Correlation

1.1.6 การประเมินประสิทธิภาพและตัวทดสอบมาตรฐาน

การพัฒนาเทคนิคต่างๆให้ก้าวหน้าขึ้นจำเป็นจะต้องมีการประเมินประสิทธิภาพที่มีประสิทธิภาพ ตัวอย่างเช่นค่า SNR ใช้ในการประเมินการลดข้อมูล และ Precision และ Recall ใช้ในการประเมินการค้นหาข้อความ วิธีการประเมินที่ดีจะทำให้เทคนิคต่างๆถูกพัฒนาไปในทิศทางที่ถูกต้อง ระบบการค้นหารูปภาพบางระบบในปัจจุบันใช้วิธีการประเมินแบบ Cost/Time ส่วนระบบอื่นๆที่เหลือใช้วิธีการประเมินแบบ Precision/Recall ที่ยืมมาจากระบบการค้นหาข้อความ

ถึงแม้ว่าวิธีการประเมินเหล่านี้จะเป็นวิธีที่ดี แต่ก็ยังห่างไกลจากความต้องการ เหตุผลหลักอันหนึ่งที่ทำให้วิธีการประเมินเหล่านี้ด้อยประสิทธิภาพลงคือปัญหาความคิดเห็นส่วนตัวของการรับรู้รายละเอียดของภาพ นั่นคือความคิดเห็นส่วนตัวทำให้ไม่สามารถประเมินผลลัพธ์ของการค้นหาได้อย่างถูกต้อง

ปัญหาที่สำคัญอีกอย่างคือการหาตัวทดสอบมาตรฐาน ในการลดขนาดภาพเราใช้ภาพเลนา ซึ่งเป็นภาพที่มีความสมดุลย์ระหว่างพื้นผิวหลายแบบ การลดขนาดภาพวีดีโอก็มีภาพวีดีโอที่เหมาะสมสำหรับการทดสอบ ในการค้นหาข้อความก็มีตัวทดสอบมาตรฐาน

สำหรับการค้นหาภาพ เริ่มมีการรวบรวมฐานข้อมูลภาพที่ใช้เป็นตัวทดสอบมาตรฐาน ตัวทดสอบที่ดีสำหรับการค้นหารูปภาพ จะต้องเป็นฐานข้อมูลภาพที่มีขนาดใหญ่ และต้องรักษาสมดุลย์ระหว่างรายละเอียดของภาพ เพื่อทดสอบประสิทธิภาพของคุณลักษณะของภาพและประสิทธิภาพโดยรวมของระบบ

1.2 คุณลักษณะของภาพ

รายละเอียดของภาพอาจจะประกอบด้วยรายละเอียดทางสายตาและรายละเอียดทางความหมาย รายละเอียดทางสายตาประกอบด้วย สี พื้นผิว รูปร่าง การวางองค์ประกอบ และอื่นๆ รายละเอียดทางความหมายอาจได้จากการกรอกข้อความโดยมนุษย์หรือใช้กระบวนการที่ซับซ้อนซึ่งอยู่บนพื้นฐานของรายละเอียดทางสายตา งานวิจัยนี้จะมุ่งเน้นเฉพาะการบรรยายรายละเอียดทางสายตาเท่านั้น

ในระบบการค้นหารูปภาพ การบรรยายรายละเอียดทางสายตาจะต้องถูกแทนที่ด้วยคุณลักษณะภาพที่เหมาะสม ในหัวข้อนี้จะแนะนำคุณลักษณะภาพอย่างกว้างๆที่ใช้ในการแสดงสี พื้นผิว รูปร่างและองค์ประกอบของภาพ

สี

สีเป็นรายละเอียดทางสายตาที่นิยมใช้มากที่สุดในการค้นหารูปภาพ [31, 50-58] ค่าสีใน 3 มิติทำให้มันมีศักยภาพในการแยกแยะความแตกต่างได้ดีกว่าค่าระดับสีเทาใน 1 มิติของภาพ ก่อนที่จะเลือกคุณลักษณะของสีที่เหมาะสม จะต้องเลือกปริภูมิสีก่อน

ปริภูมิสี (Color Space)

แต่ละจุดในภาพสามารถแสดงเป็นจุดในระบบ 3 มิติปริภูมิสี ปริภูมิสีที่นิยมใช้ในการค้นหารูปภาพคือ RGB, Munsell, CIE L*a*b*, CIE L*u*v*, HSV, และปริภูมิสีตรงข้าม แต่อย่างไรก็ตามจนถึงตอนนี้ยังคงไม่มีปริภูมิสีใดที่ดีที่สุด แต่ปริภูมิสีที่เหมาะสมสำหรับระบบการค้นหารูปภาพจะต้องมีคุณสมบัติสม่ำเสมอในการรับรู้ (Perceptually Uniform) ปริภูมิสีที่มีความสม่ำเสมอในการรับรู้หมายความว่า ระยะทางระหว่างสี 2 สีในปริภูมิสีจะต้องมีค่าเท่ากับความเหมือนที่วัดได้โดยการรับรู้ของมนุษย์

ปริภูมิ RGB เป็นปริภูมิสี ที่ใช้กันอย่างกว้างขวางในการแสดงภาพ มันประกอบด้วยองค์ประกอบของแสงสีแดง เขียวและน้ำเงิน แต่ปริภูมิ RGB เป็นปริภูมิสีที่รู้จักกันโดยดีว่าไม่มีคุณสมบัติความสม่ำเสมอในการรับรู้

ปริภูมิ CIE L*a*b* และ CIE L*u*v* เป็นปริภูมิสีที่มีความสม่ำเสมอในการรับรู้ มันประกอบด้วยองค์ประกอบทางความสว่าง (Lightness) และองค์ประกอบทางสี (a และ b หรือ u และ v)

ปริภูมิ HSV นิยมใช้กันอย่างกว้างขวางในงานทางด้าน Computer Graphic องค์ประกอบของมันคือ Hue, Saturation และ Value ค่าของ Hue จะไม่มีการเปลี่ยนแปลงตามความสว่างของแสงและทิศทางของกล้อง ดังนั้นจึงเหมาะสมสำหรับการค้นหารูปภาพ

ปริภูมิสีตรงข้าม ประกอบด้วยองค์ประกอบ R-G, 2B-R-G, R+G+B ข้อดีของมันคือการแยกข้อมูลความสว่างไว้ในองค์ประกอบที่ 3 และองค์ประกอบที่ 1 และ 2 เป็นองค์ประกอบทางสี ซึ่งสามารถที่จะลดขนาดลงมาได้ เนื่องจากการรับรู้ของมนุษย์มีความไวต่อการเปลี่ยนแปลงของความสว่างมากกว่าการเปลี่ยนแปลงของสี

คุณลักษณะของสีที่นิยมใช้กันมีดังต่อไปนี้

Color Histogram

Color Histogram เป็นวิธีการแสดงรายละเอียดของสีของภาพได้อย่างมีประสิทธิภาพ เราสามารถคำนวณ Color Histogram ของภาพได้อย่างมีประสิทธิภาพและรวดเร็ว นอกจากนี้มันยังทนทานต่อการเปลี่ยนแปลงทางการเลื่อนภาพ การหมุน และเปลี่ยนแบบช้าๆตามสัดส่วนของภาพ การบดบังและมุมมอง

Color Histogram คือการแจกแจงของจุดภาพในแต่ละ Bin ของ Color Space Color Histogram ที่มีจำนวน Bin มากจะทำให้มีศักยภาพสูงในการแยกแยะความแตกต่าง แต่ก็ทำให้สิ้นเปลืองเวลาในการคำนวณ และไม่เหมาะสำหรับการสร้างตัวชี้ของฐานข้อมูลภาพที่มีประสิทธิภาพ

1.3 การวัดความเหมือน

ระบบการค้นหารูปภาพใช้วิธีการวัดความเหมือนของรายละเอียดทางสายตาในการเปรียบเทียบภาพที่ต้องการค้นหากับภาพที่อยู่ในฐานข้อมูล และเรียงลำดับความเหมือนของ

ภาพในฐานข้อมูลกับภาพที่ต้องการค้นหา วิธีการวัดความเหมือนของภาพหลายวิธีถูกพัฒนาขึ้นมาเพื่อใช้กับระบบการค้นหารูปภาพ วิธีการวัดที่แตกต่างกันจะทำให้ประสิทธิภาพของระบบที่แตกต่างกัน ในหัวข้อนี้จะแนะนำวิธีการวัดความเหมือนที่นิยมใช้กันบางวิธี โดยกำหนดให้ $D(I,J)$ เป็นระยะทางระหว่างภาพ I ที่ต้องการค้นหา และภาพ J ที่อยู่ในฐานข้อมูล และ $f_i(I)$ เท่ากับจำนวนจุดภาพที่อยู่ใน Bin ที่ i ของภาพ I

Minkowski-Form Distance

Minkowski-Form Distance เหมาะสมสำหรับการคำนวณระยะทางระหว่างภาพ 2 ภาพถ้าแต่ละมิติของคุณลักษณะของภาพไม่ขึ้นต่อกันและมีความสำคัญเท่ากัน เราสามารถคำนวณ Minkowski-Form Distance โดย

$$D(I,J)=\left(\sum_i \left|f_i(I)-f_i(J)\right|^p\right)^{\frac{1}{p}} \qquad (1)$$

ถ้า $p=1,2$ และ ∞ เราสามารถ $D(I,J)$ แบบย่อๆเป็น L_1, L_2 และ L_∞ Mindkowski-Form Distance เป็นวิธีการวัดความเหมือนที่นิยมใช้ กันมากในระบบการค้นหารูปภาพ

Histogram Intersection

Histogram Intersection เป็นกรณีพิเศษของ L_1 ซึ่งถูกพัฒนาโดย Swain และ Ballard [57] เพื่อใช้คำนวณความเหมือนระหว่างภาพ 2 ภาพ Histogram Intersection สามารถคำนวณโดย

$$S(I,J)=\frac{\sum_{i=1}^{N}\min\left(f_i(I),f_i(J)\right)}{\sum_{i=1}^{N} f_i(J)} \qquad (2)$$

Histogram Intersection ได้แสดงให้เห็นว่ามันไม่ไวต่อการเปลี่ยนความละเอียดของภาพ, ขนาดของ Histogram, การบดบัง ความลึกและมุมมอง

Quadratic Form (QF) Distance

Minkowski Distance กำหนดให้ทุก Bin ของ Histogram มีความสำคัญเท่ากัน โดยไม่สนใจความจริงที่ว่า Bin บางอันมีความสำคัญมากกว่าอันอื่น Quadratic Form Distance ได้ถูกพัฒนาขึ้นมาเพื่อแก้ปัญหานี้

$$D(I,J)=\sqrt{\left(F_I-F_j\right)^T A\left(F_I-F_J\right)} \qquad (3)$$

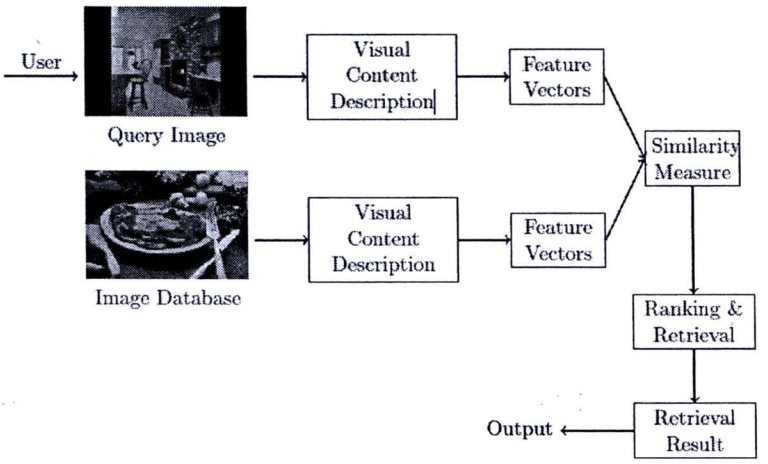
โดยที่ $A=[a_{ij}]$ เป็นเมทริกซ์ของความเหมือน และ a_{ij} หมายถึงความเหมือนระหว่าง Bin i และ Bin j

1.4 การประเมินประสิทธิภาพ

ในงานวิจัยนี้ใช้ค่า Precision ในการประเมินประสิทธิภาพของระบบการค้นหารูปภาพ สำหรับภาพที่ใช้ในการค้นหา q เซ็ทของภาพในฐานข้อมูลที่เกี่ยวข้องกับภาพ q กำหนดให้เป็น $R(q)$ และผลลัพธ์ของการค้นหาภาพ q กำหนดให้เป็น $Q(q)$ ค่า Precision คือสัดส่วนระหว่างจำนวนภาพที่เกี่ยวข้องกับภาพ q กับภาพผลลัพธ์ทั้งหมด

$$precision = \frac{Q(q) \cap R(q)}{|Q(q)|} \qquad (4)$$

1.5 ระบบการค้นหารูปภาพโดยรายละเอียดของภาพ



รูปที่ 3. แผนผังของระบบการค้นหารูปภาพโดยรายละเอียดของภาพ

แผนผังของระบบการค้นหารูปภาพโดยรายละเอียดของภาพที่ใช้ในงานวิจัยนี้