บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาและออกแบบระบบการค้นคืนภาพในงานวิจัยนี้ ผู้วิจัยได้ทำการศึกษาแนวคิด ทฤษฎี หลักการตลอดจนเอกสารงานวิจัยที่เกี่ยวข้อง เพื่อเป็นพื้นฐานความรู้สำหรับการทำวิจัย ดังต่อไปนี้

1. ระบบการค้นคืนภาพโดยยึดหลักเนื้อหา [Content-Based Image Retrieval (CBIR) System]
2. การตอบสนองความสัมพันธ์ [Relevance Feedback (RF)]
3. เครือข่ายฟังก์ชันพื้นฐานรัศมี [Radial Basis Function Network (RBFN)]
4. อัลกอริทึมเคมีน (K-Mean Algorithm)
5. ระบบการโต้ตอบของเครื่องแบบอัตโนมัติ (Automatic Machine Interactions System)
6. MPEG-7
7. เครื่องมือสรุปวีดีโอแบบอัตโนมัติ (Automatic Video Summarizing Tool using MPEG-7 Descriptor for Personal Video Recorder)
8. การใช้ตัวอธิบายเส้นแบบโลคอลที่มีประสิทธิภาพ (Efficient Use of Local Edge Histogram Descriptor)

การค้นคืนภาพโดยยึดหลักเนื้อหา (Content-Based Image Retrieval System)

เป็นระบบสืบค้นข้อมูลภาพที่ยึดหลักของเนื้อหา (Content) ของภาพโดยเนื้อหาองภาพจะแบ่งได้ 2 ระดับ คือ เนื้อหาของภาพในระดับต่ำเช่น สี (Color) รูปร่าง (Shape) พื้นผิว (Texture) เป็นต้น ซึ่งเป็นมุมมองระดับของคอมพิวเตอร์และเนื้อหาของภาพในระดับสูงเช่น วัตถุ (Object) เหตุการณ์ (Event) เป็นต้น ซึ่งเป็นมุมมองในระดับของมนุษย์สำหรับระบบสืบค้นข้อมูลภาพโดยยึดหลักเนื้อหาในปัจจุบันจะเน้นที่เนื้อหาของภาพในระดับต่ำเพราะระบบคอมพิวเตอร์สามารถที่จะประมวลผลได้ ส่วนเนื้อหาของภาพในระดับสูงนั้นยังมีน้อยเนื่องจากระบบคอมพิวเตอร์ในปัจจุบันยังมีความสามารถน้อยในการที่จะดึงเนื้อหาของภาพในระดับสูง

ดังกล่าวได้ ตัวอย่างของระบบที่นำหลักการของ CBIR นี้ไปประยุกต์ใช้เช่น การจัดทำระบบค้นหาภาพทางการแพทย์ การจัดทำระบบค้นหาภาพในพิพิธภัณฑ์ โลโก้สินค้าและตัวอธิบาย (Descriptor) เนื้อหาของมาตรฐาน MPEG-7 เป็นต้น

1. **องค์ประกอบพื้นฐานของระบบค้นคืนภาพ** (Basic Component of CBIR)

องค์ประกอบพื้นฐานของระบบค้นคืนภาพ (Basic Component of CBIR) ประกอบด้วย 3 ส่วนหลักคือ

1.1 ตัวดึงฟีเจอร์ (Feature Extractor) คือส่วนของการสร้างข้อมูล (Metadata) ของข้อมูลภาพขึ้นมา ซึ่งข้อมูล (Metadata) คือฟีเจอร์ของภาพ (Image Feature) เช่น สี รูปร่าง พื้นผิว และโครงสร้าง (Structure) เป็นต้น ซึ่งตัวอย่างของฟีเจอร์แต่ละประเภทสามารถอธิบายคร่าว ๆ ได้คือ

1.1.1 ส่วนของฟีเจอร์สี (Color Feature) จะมี Color Space ที่ใช้หลายชนิดเช่น RGB , YCrCb เป็นต้น

1.1.2 ส่วนของฟีเจอร์รูปร่าง (Shape Feature) จะมี 3 แบบหลัก ๆ คือ Region-Based Shape , Contour-Based Shape และ Edge-Based Shape

1.1.3 ส่วนของฟีเจอร์พื้นผิว (Texture Feature) จะมีที่นิยมใช้คือ Wavelete - Based Texture Feature เป็นต้น

สำหรับรายละเอียดของวิธีการสร้าง Metadata หรือฟีเจอร์ดังกล่าวจะอยู่นอกเหนือจากงานวิจัยนี้ซึ่งผู้อ่านสามารถดูรายละเอียดเพิ่มเติมได้ใน [4]

1.2 คิวรี่เอ็นจิ้น (Query Engine) คือส่วนที่ใช้ค้นหาความเหมือนกันของภาพ

ส่วนของความเหมือนในลักษณะของการมองเห็น (Visual Similarity) โดยทางความหมายแล้ว Similarity คือ Subjective , Context-Dependent และแนวความคิดในระดับสูง (High-Level Concept) แต่เนื่องจากฟีเจอร์ที่ใช้จะเป็นแบบฟีเจอร์ในระดับต่ำ (Low-Level Feature) จึงทำให้เกิดปัญหาที่เรียกว่า Semantic Gab ขึ้น

ในส่วนของฟีเจอร์นั้นไม่สามารถบอกได้ว่าฟีเจอร์ใดที่มีความสำคัญมากที่สุดในการค้นหาขึ้นอยู่กับลักษณะการใช้งานและข้อมูลดังนั้นจึงใช้วิธีการระบุน้ำหนัก (Weighting) ของฟีเจอร์แทนซึ่งจะมีอยู่ 2 แบบคือ

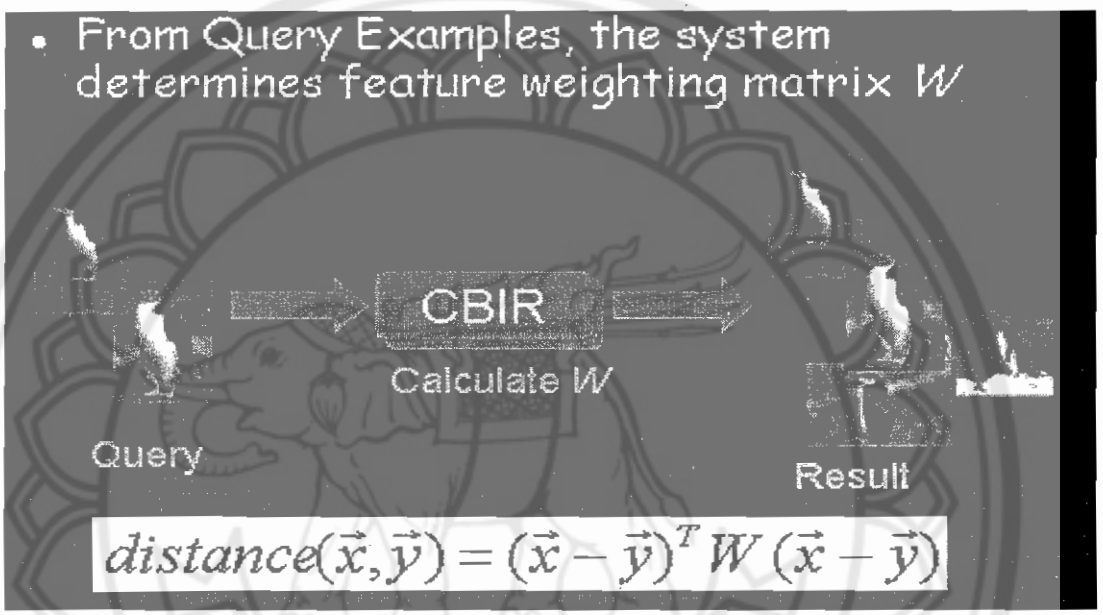
1.2.1 การ Weight ด้วยมือ (Manual)

เป็นลักษณะของการให้ผู้ใช้เป็นผู้กำหนดน้ำหนักเองว่าฟีเจอร์ประเภทใดจะมีน้ำหนักเท่าไร (%) ซึ่งวิธีนี้จะยากในการคำนวณ

1.2.2 การ Weight แบบอัตโนมัติ (Automatic)

เป็นลักษณะของการกำหนดน้ำหนักโดยเรียนรู้จากภาพที่ใช้ค้นหาโดยตัวระบบเองซึ่งวิธีนี้จะใช้หลักการของ Relevance Feedback

สำหรับตัวอย่างของการ Weight โดยพิจารณาจากภาพที่ใช้ค้นหาสามารถแสดงได้ดังภาพ 1 ข้างล่าง



ภาพ 1 แสดงตัวอย่างการ Weight โดยพิจารณาจากภาพที่ใช้ค้นหา [4]

สำหรับรายละเอียดของวิธีการดังกล่าวจะอยู่นอกเหนือจากงานวิจัยนี้ซึ่งผู้อ่านสามารถดูรายละเอียดเพิ่มเติมได้ใน [4]

1.3 ส่วนโต้ตอบกับผู้ใช้ (User Interface) คือส่วนที่ใช้โต้ตอบกับผู้ใช้ซึ่งจะมีส่วนที่ให้ผู้ใช้ทำการเลือกภาพว่าภาพใดบ้างที่ตนต้องการเพื่อระบบจะนำข้อมูลการเลือกนั้นไปทำการเรียนรู้และประมวลผลต่อไป

การทำงานของระบบ CBIR มีขั้นตอนดังนี้

1. การดึงฟีเจอร์ (Feature Extraction) จากภาพ
2. ให้ผู้ใช้ทำการ Query โดยทั่วไปจะมีอยู่ 3 แบบคือ
   2.1 การค้นหาโดยใช้ภาพ Sketch เป็น Query (Query by Sketch)
   2.2 การค้นหาโดยใช้ Keywords เป็น Query (Query by Keywords)
   2.3 การค้นหาโดยใช้ภาพ Examples เป็น Query (Query by Examples)

3. ทำการกลั่นกรองหรือเรียนรู้จากข้อมูลที่ผู้ใช้เลือกโดยใช้หลักของ Relevance Feedback และทำการประมวลผล

4. แสดงผลลัพธ์จากการประมวลผล

ตัวอย่างของการค้นหาโดยใช้ภาพตัวอย่าง (Query by Examples) แสดงได้ดังภาพ 2 ข้างล่าง

ภาพ 2 แสดงการค้นหาภาพของระบบ CBIR โดยใช้ภาพตัวอย่าง [4]

สำหรับรายละเอียดเพิ่มเติมของระบบ CBIR และวิธีการคำนวณต่าง ๆ นั้น สามารถดูเพิ่มเติมได้ใน [4]

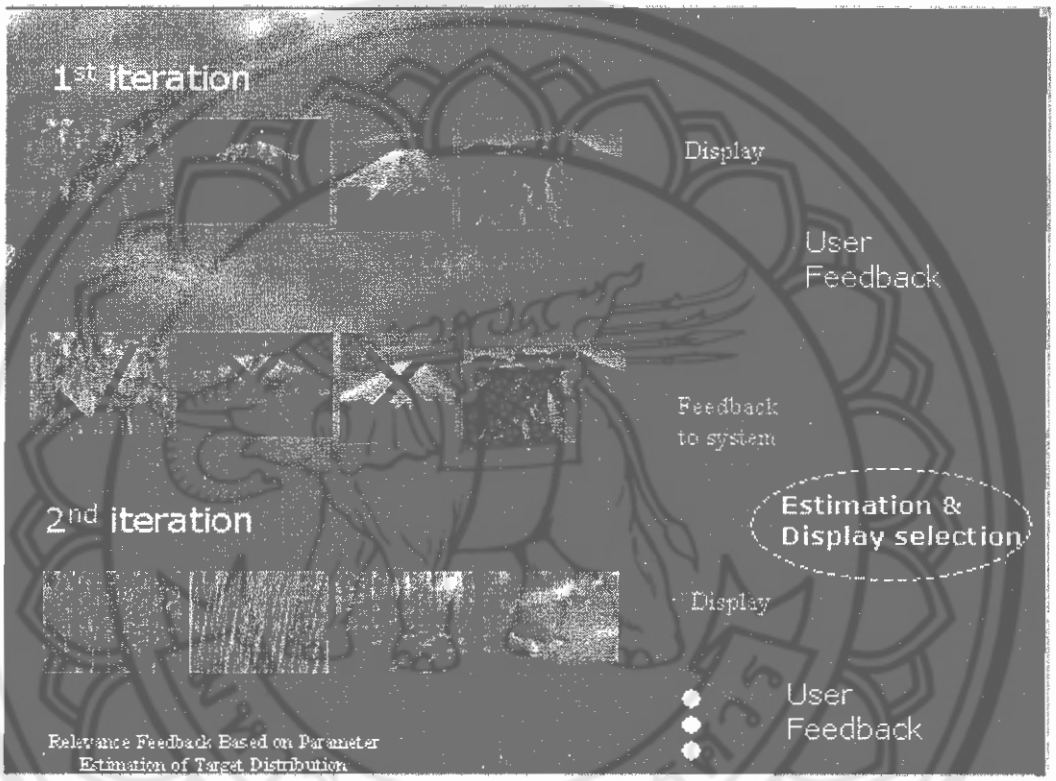
จากเนื้อหาของ Content-Based Image Retrieval ที่อธิบายข้างต้นเป็นหลักการพื้นฐานในการสืบค้นข้อมูลภาพที่ผู้วิจัยได้นำมาใช้เป็นพื้นฐานในการพัฒนาระบบในงานวิจัยนี้

การตอบสนองความสัมพันธ์ [Relevance Feedback (RF)]

เป็นสถาปัตยกรรมที่ใช้สำหรับจับ (Capture) เป้าหมาย (Target) ในการค้นหา (Search) ของผู้ใช้ซึ่งเป็นลักษณะของกระบวนการเรียนรู้ (Learning Process) โดยวิธีการนี้มีหลักการง่าย ๆ คือ การนำข้อมูลจากการค้นหาที่ผู้ใช้เลือกมาทำการเรียนรู้แล้วทำการนำข้อมูลจากการเรียนรู้ที่ได้ไปปรับปรุงการค้นหาในรอบถัดไปแล้วนำผลลัพธ์จากการค้นหามาแสดงให้ผู้ใช้เห็นผ่านหน้าจอจากนั้นก็ให้ผู้ใช้ทำการเลือกภาพชุดใหม่ที่ตนต้องการจากผลลัพธ์ที่แสดงจากนั้นระบบก็จะนำภาพดังกล่าวไปทำการเรียนรู้และปรับปรุงการค้นหาใหม่เป็นลักษณะวนซ้ำไปเรื่อย ๆ จนผู้ใช้พบภาพที่ต้องการ สำหรับกระบวนการทำงานจะแบ่งได้เป็น 2 ขั้นตอน คือ

1. การตอบสนอง (Feedback) คือขั้นตอนของการเรียนรู้จากการตอบสนองของผู้ใช้ในการเลือกภาพต่าง ๆ

2. การแสดงผล (Display) คือขั้นตอนของการเลือกข้อมูลชุดใหม่แล้วแสดงให้ผู้ใช้เห็น โดยกระบวนการทำงานดังกล่าวสามารถสรุปได้ดังภาพ 3 ข้างล่าง



ภาพ 3 แสดงลำดับขั้นตอนการทำงานของ RF [2]

จากเนื้อหาของ Relevance Feedback ที่อธิบายข้างต้น [2] เป็นหลักการพื้นฐานในการสืบค้นข้อมูลภาพที่ผู้วิจัยได้นำมาใช้เป็นพื้นฐานในการพัฒนาระบบในงานวิจัยนี้

เครือข่ายฟังก์ชันพื้นฐานรัศมี [Radial Basis Function Network (RBFN)]

เป็นนิวโรนเน็ตเวิกต์ (Neural Network) แบบหนึ่งซึ่งมีวิธีในการดำเนินการแมพ (Mapping) ค่าแบบ 2 ระยะ (2-Stage) ซึ่งแต่ละนิวโรน (Neuron) ที่ i ของเลเยอร์ (Layer) แรกจะประกอบด้วยศูนย์กลาง (Center) คือ μ_i และความกว้าง (Width) คือ σ_i สำหรับค่าความแตกต่างทางยูคลิด (Euclidean Distance) ระหว่างสัญญาณอินพุท (Input Signal) คือ S

และศูนย์กลางของนิวโรนจะถูกคำนวณเป็นอันดับแรก หลังจากนั้น Activation Function ที่รู้จักกันในชื่อ Basis Function (BF) อย่างเช่น Gaussian Function จะถูกใช้สำหรับประมวลผล ส่วนเอาท์พุท (Output) ของ Hidden Layer จะถูกรวมเข้าด้วยกันในลักษณะเชิงเส้น (Linear) โดยนิวโรนของเลเยอร์ที่ 2 ตามสมการ 1 ดังนี้

$$f_i(S) = \lambda_0 + \sum_{i=1}^{n} \lambda_i \phi(\| S - \mu_i \|) \quad \text{.....(1)}$$

จากหลักการของ RBFN และสมการที่อธิบายข้างต้นผู้วิจัยได้นำมาประยุกต์ใช้สำหรับการเรียนรู้ความเหมือนในระดับสูงภายในส่วนของการดำเนินการของ Relevance Feedback ในงานวิจัยนี้ สำหรับรายละเอียดของ RBF Network สามารถดูเพิ่มเติมได้ใน [3]

อัลกอริธึมเคมีน (K-Mean Algorithm)

ในงานวิจัยนี้ได้เสนอ K-Mean สำหรับการจัดกลุ่มข้อมูลเพื่อการเรียนรู้ความเหมือนในระดับต่ำภายในส่วนของการดำเนินการของ Relevance Feedback โดย K-Mean Clustering คืออัลกอริธึมในการจัดกลุ่มข้อมูลออกเป็นจำนวน k กลุ่มโดยมีพื้นฐานอยู่บน Attribute หรือฟีเจอร์ของข้อมูลเหล่านั้นโดยที่ k คือ ค่าจำนวนเต็มบวกซึ่งการจัดกลุ่มนั้นจะกระทำโดยการลดค่าของผลรวมของค่า Distance ระหว่างข้อมูลและศูนย์กลางของกลุ่มข้อมูล (Cluster Center) ดังนั้นจุดประสงค์ของ K-Mean Clustering ก็คือการจัดกลุ่มข้อมูล

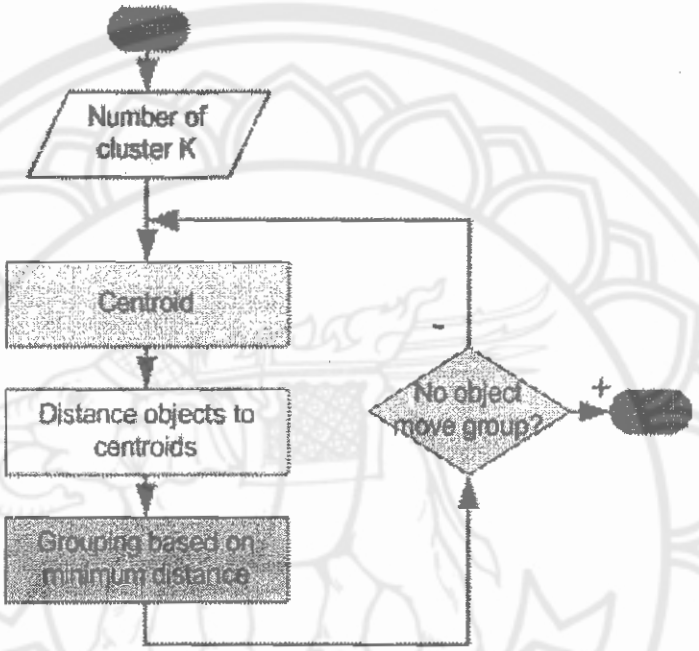
สำหรับขั้นตอนการทำงานพื้นฐานของ K-Mean Clustering จะมีรูปแบบง่าย ๆ โดยเริ่มต้นจากการกำหนดจำนวนของกลุ่ม (k Cluster) ที่ต้องการและกำหนดศูนย์กลางของกลุ่มเหล่านั้นโดยอาจกำหนดจากการสุ่มเลือก Object ภายในกลุ่มหรือ Object แรกของกลุ่มเป็นค่าของศูนย์กลางของกลุ่มเริ่มต้น (Initial Center)

ในส่วนของอัลกอริธึมของ K-Mean จะมีขั้นตอนอยู่ 3 ขั้นตอนคือ

1. **คำนวณคู่ลำดับของศูนย์กลางกลุ่ม** (Determine the Center Coordinate)
2. **คำนวณค่า** Distance **ของแต่ละ** Object **กับค่าของศูนย์กลางกลุ่ม**
3. **จัดกลุ่ม** Object **บนพื้นฐานของค่า** Distance **ที่น้อยที่สุด** (Minimum Distance)

*หมายเหตุ การทำงานจะวนซ้ำจนกระทั่งข้อมูลนั้นคงที่ (Stable) คือไม่มี Object ใดมีการเคลื่อนย้ายไปไว้ในกลุ่มใดกลุ่มหนึ่ง

สำหรับขั้นตอนการทำงานของอัลกอริธึมดังกล่าวสามารถแสดงเป็นผังการทำงานได้ดังภาพ 4



ภาพ 4 แสดงผังลำดับการทำงานของอัลกอริทึม K-Mean [5]

สำหรับเนื้อหาของ K-Mean Algorithm ที่อธิบายข้างต้นดังกล่าวผู้วิจัยได้นำเอาหลักการของ K-Mean มาประยุกต์ใช้กับระบบค้นคืนภาพที่พัฒนาโดยนำไปใช้สำหรับการจัดกลุ่มข้อมูลในมุมมองระดับต่ำซึ่งได้นำ Source Code ของ K-Mean ที่มีผู้พัฒนาไว้แล้วมาประยุกต์ใช้ โดยข้อมูลที่อธิบายและ Source Code นั้นถูกพัฒนาขึ้นโดย Kardi Teknomo ซึ่งผู้วิจัยได้นำมาจาก [5]

การโต้ตอบของเครื่องแบบอัตโนมัติ (Automatic Machine Interactions)

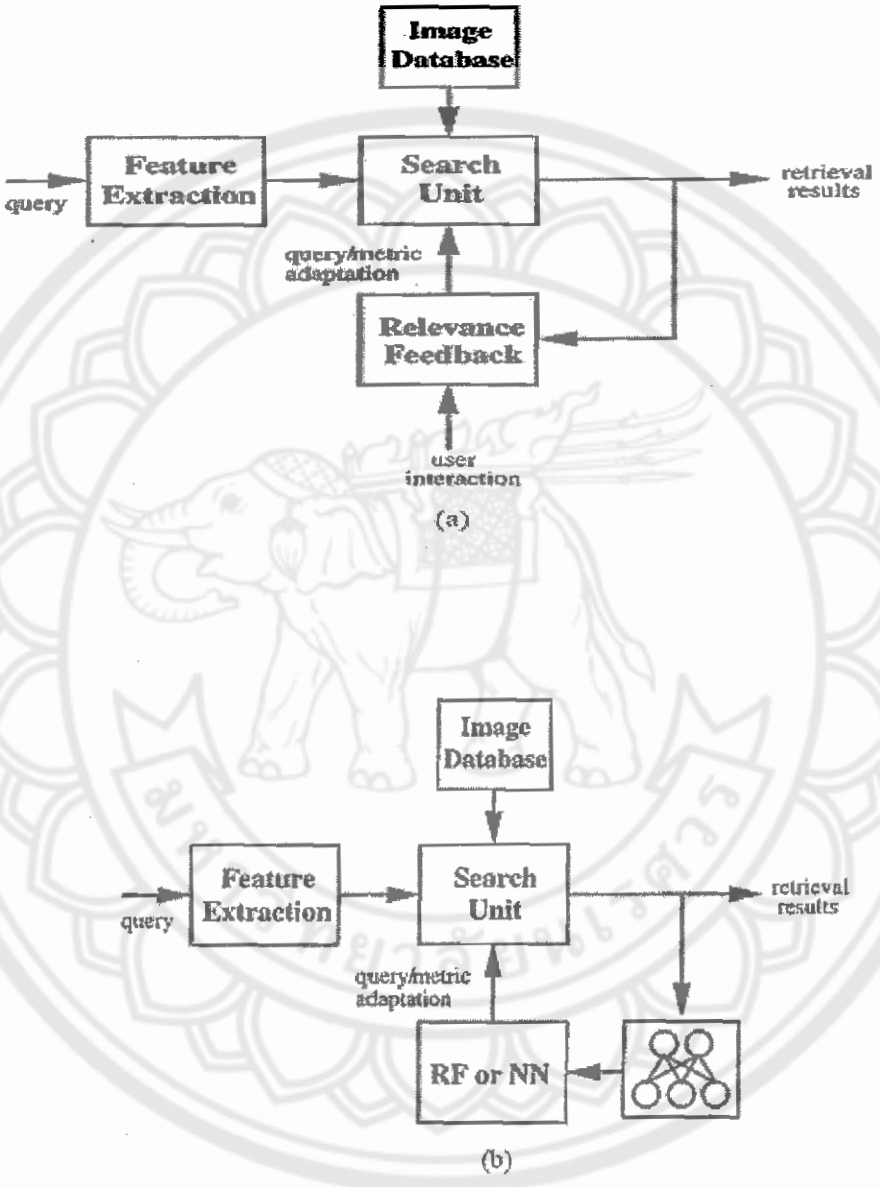
ในส่วนของวิธีการลดภาระการทำงานของ User นั้นได้มีการใช้วิธีการโต้ตอบแบบอัตโนมัติ [1] โดยมีรายละเอียดดังนี้

1. **ส่วนของแนวความคิดโดยรวมของระบบ**

แนวความคิดโดยรวมของงานวิจัยนี้คือการเสนอเทคนิคในการค้นคืนภาพแบบใหม่ที่มีวิธีการทำงานเป็นแบบวนซ้ำ (Recursive) ที่มีลักษณะเป็นแบบอัตโนมัติ (Automatic) โดยจะนำเอา Unsupervised Learning Network คือ SOTM มาใช้โดยเอาความสามารถในการเรียนรู้ด้วยตนเอง (Self-Learning Capability) ของ Network ดังกล่าวมาใช้ร่วมกับระบบการค้นคืนภาพโดยยึดหลักเนื้อหา (Content-Based Image Retrieval System) ที่มีอยู่ในปัจจุบันเพื่อลดภาระการทำงานของผู้ใช้ในการที่จะต้องคอยโต้ตอบกับระบบและได้เสนอวิธีการที่เรียกว่า Semiautomatic เพื่อใช้สนับสนุนในการค้นคืนภาพกับผู้ใช้ที่มีแนวความคิด (Subjectivity) ของความต้องการที่แตกต่างกันและในส่วนของการวิเคราะห์ (Evaluated) ความเหมือนของภาพ (Image Similarity) จะใช้วิธีของ Nonlinear Model ที่มีพื้นฐานอยู่บน Gaussian-Shaped Radial Basis Function (Gaussian-Shaped RBF) ในการทำงานซึ่งสามารถจะทำการแยกแยะข้อแตกต่าง (Discrimination) ของข้อมูลโดยมีพื้นฐานอยู่บนการวิเคราะห์ในระดับโลคอล (Local Analysis) โดยในการดำเนินการวิเคราะห์ดังกล่าวนี้จะใช้ Single-Pass RBFN ในการดำเนินการซึ่งข้อดีคือจะใช้ Parameter น้อย

2. **ส่วนของการดำเนินการ**

ในส่วนของการดำเนินการสามารถอธิบายได้ดังหัวข้อต่อไปนี้



ภาพ 5 (a) User Interactive CBIR System [1]

(b) CBIR System with Neural-Network Processing [1]

2.1 โครงสร้างของระบบ

ในส่วนของโครงสร้างของระบบค้นคืนภาพนี้จะแสดงได้ดังภาพ 5 (b) เปรียบเทียบกับระบบการค้นคืนภาพ (CBIR) ที่มีการโต้ตอบแบบเดิมดังแสดงในภาพ 5 (a) โดยในระบบที่เสนอดังกล่าวจะเป็นระบบการค้นคืนที่มีการโต้ตอบแบบวนซ้ำ (Recursive Interactive Retrieval System) โดยใช้วิธี Automatic Machine Learning Approach และมีการนำ Unsupervised Neural-Network Model คือ SOTM มาปรับปรุงเพื่อใช้เป็นผู้ช่วยสอนให้กับ RF/NN Module แทนคนดังในภาพ 5 (b) ทำให้ความสามารถในการค้นคืนภาพสามารถที่จะปรับปรุงได้อย่างอัตโนมัติโดยปราศจากผู้สอนและในส่วนของการเรียนรู้ความเหมือน (RF or NN Module) จะใช้ RBFN ที่มีการปรับปรุงใหม่คือ Single-Pass Radial Basis Function Network (RBFN) มาใช้สำหรับการวิเคราะห์ความเหมือนเพื่อใช้ในการปรับปรุง Query สำหรับการค้นหาในรอบถัดไป

2.2 วิธีการวัดค่าประสิทธิภาพในการค้นหาข้อมูลภาพ

ในส่วนของการวัดประสิทธิภาพจะแบ่งเป็น 2 ส่วนคือการวัดจากอัตราการค้นคืนเฉลี่ย [Average Retrieval Rate (AVR)] ของภาพคิวรี่จำนวน M ภาพ (Query Image) คือ

$$AVR = \frac{\left[\sum_{i=1}^{M} R_i\right]}{M} \quad \text{.....(2)}$$

โดยอัตราการค้นคืน Retrieval Rate (R_i) จะนิยามได้ดังนี้

$$\text{Retrieval Rate (Ri)} = \frac{\text{no.relevant image retrieval for query i}}{\text{N} \times 100\%} \quad \text{.....(3)}$$

โดย N คือจำนวนภาพที่แสดง

ในส่วนที่ 2 นั้นประสิทธิภาพจะวัดจากความแม่นยำเฉลี่ย (Average Precision) โดยสมการของความแม่นยำคือ

$$\text{Precision} = \frac{\text{no.Relevant image, Nr}}{\text{no.Retrievaled image, N}_{\text{RT}}} \times 100\% \quad \text{.....(4)}$$

โดย N_{RT} คือจำนวนภาพที่แสดงและ N_r คือจำนวนภาพที่เหมือนกับภาพที่ใช้เป็นคิวรี่

จากงานวิจัยที่อธิบายข้างต้นผู้วิจัยได้นำเอาวิธีการวัดค่าประสิทธิภาพการทำงานของระบบมาใช้เป็นพื้นฐานสำหรับการวัดประสิทธิภาพของระบบที่พัฒนา

MPEG-7

1. การแนะนำ (Introduction)

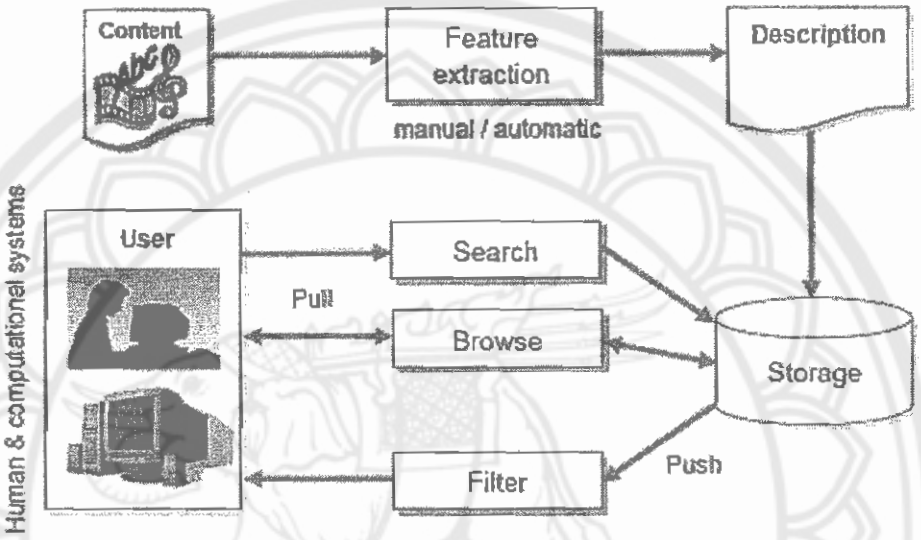
MPEG-7 เป็นมาตรฐาน ISO/IEC (ISO/IEC Standard) ที่ถูกพัฒนาโดย Moving Pictures Experts Group (MPEG) ซึ่งเป็นกลุ่มการทำงานที่อยู่ภายใน International Organization for Standardization (ISO) MPEG-7 ถูกกำหนดเป็นมาตรฐานในปี 2001 มีชื่ออย่างเป็นทางการคือ Multimedia Content Description Interface โดยมีจุดมุ่งหมายเพื่ออธิบายเนื้อหา (Content) ของข้อมูลมัลติมีเดีย (Multimedia Data) โดยการผูก Metadata เข้ากับ Multimedia Content โดย MPEG-7 จะกำหนดมาตรฐานของชุดของเครื่องมือที่ใช้ในการอธิบาย (Set of Description Tools) ซึ่งสามารถใช้ในการอธิบายสารสนเทศมัลติมีเดีย (Multimedia Information) ชนิดต่าง ๆ ได้ MPEG-7 ไม่ได้ถูกนำมาแทนที่มาตรฐานของ MPEG เดิม (MPEG-1 , MPEG-2 , MPEG-4) แต่มีจุดประสงค์เพื่อจัดเตรียมฟังก์ชันเพิ่มเติมให้กับมาตรฐาน MPEG อื่น ๆ โดยมาตรฐาน MPEG เดิมจะเป็นตัวสร้าง Content ที่ใช้งานได้ขณะที่ MPEG-7 จะใช้สำหรับการค้นหา Content ต่าง ๆ ที่ผู้ใช้ต้องการผ่านเครื่องมือต่าง ๆ เช่น Browser และ Retrieval Tool เป็นต้น

2. ลักษณะเฉพาะของ MPEG-7 (MPEG-7 Characteristics)

MPEG-7 เป็นมาตรฐานสำหรับการอธิบาย Features ของ Multimedia Content ซึ่งผู้ใช้สามารถที่จะทำการ Search , Browse และ Retrieve เนื้อหาเหล่านั้นได้อย่างมีประสิทธิภาพมากขึ้นเพื่อสร้างคำอธิบาย (Descriptions) MPEG-7 ได้เสนอชุด (Set) ของ Audio-Visual Metadata Elements และโครงสร้าง (Structure) และความสัมพันธ์ (Relationships) ของอีลีเม้น (Element) เหล่านั้นซึ่งสิ่งเหล่านี้จะถูกกำหนดมาตรฐานที่อยู่ในรูปแบบของ Descriptors และ Description Schemes ซึ่ง MPEG-7 มีความสามารถที่จะอธิบายฟีเจอร์ในระดับต่ำ (Low-Level Features) ลักษณะทางความหมาย (Semantics) และทางโครงสร้าง (Structural) ของไฟล์มัลติมีเดีย (Multimedia File) ต่าง ๆ ได้

2.1 ห่วงโซ่ของ MPEG-7 (MPEG7 Chain)

กระบวนการของ MPEG-7 ที่เป็นลักษณะลูกโซ่ (Chain) จะประกอบด้วย Feature Extraction , การอธิบายเนื้อหา และ Application ที่ใช้ในการอธิบายดังที่แสดงอยู่ใน ภาพ 6



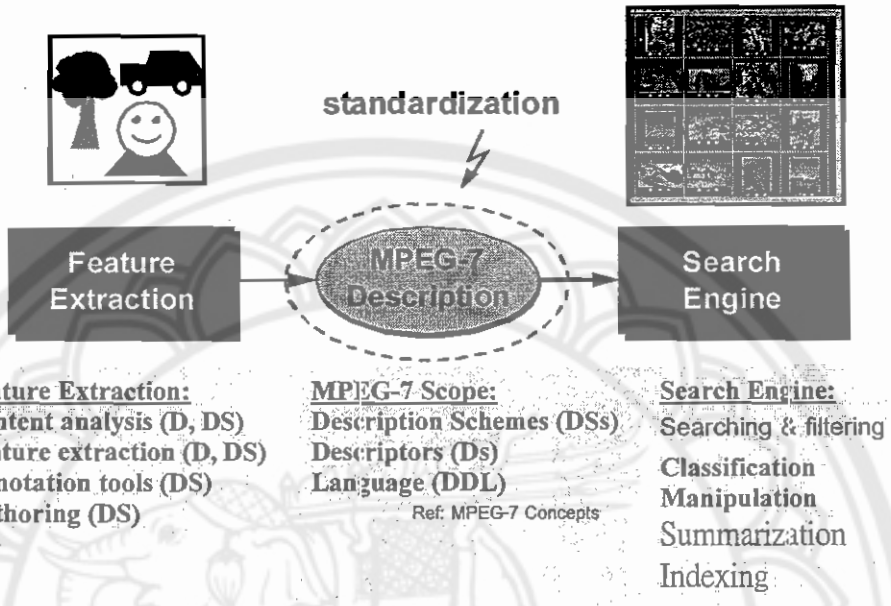
ภาพ 6 แสดงตัวอย่างการประยุกต์ใช้งาน MPEG-7 (MPEG7 Chain) [11]

จากภาพ 6 เป็นตัวอย่างของกระบวนการของ MPEG-7 ที่เป็นลักษณะลูกโซ่ โดย Description นั้นจะได้มาจาก Multimedia Content ด้วยการทำ Manual or Semi-Automatic Extraction แล้ว Description นั้นจะถูกจัดเก็บไว้ในอุปกรณ์จัดเก็บข้อมูล (Storage) เพื่อใช้งานในภายหลังต่อไป ในส่วนของ Pull Scenario นั้น Set ของ Descriptions ที่ตรง (Match) กับ Query ของผู้ใช้จะถูกส่งกลับมาสำหรับการทำการ Browsing ในส่วนของ Push Scenario นั้นตัว Agent Filter จะทำหน้าที่ในการอธิบายและกระทำการกระทำต่าง ๆ ที่ถูกโปรแกรมไว้ เช่น การสลับช่องสัญญาณแบบกระจาย (Switches a Broadcast Channel) หรือการบันทึกชุดข้อมูลที่ถูกอธิบายไว้

2.2 ของเขตของ MPEG-7 (Scope of MPEG-7)

สำหรับในส่วนของขอบเขตของ MPEG-7 นั้นไม่ได้กำหนดมาตรฐานในส่วนของการดึงฟีเจอร์ของออดิโอวิชวล (Audiovisual Feature Extraction) และโปรแกรมประยุกต์ (Application) ที่ใช้ Description เหล่านั้นซึ่งมันเป็นหน้าที่ของผู้ผลิต Software ที่จะรวม

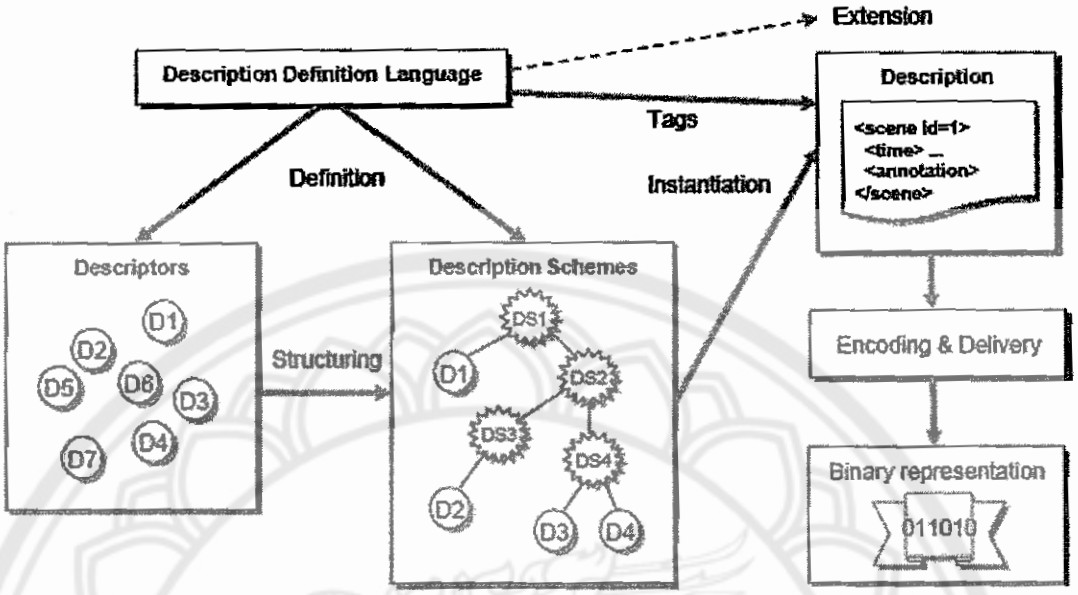
Content ที่ถูกอธิบายด้วย MPEG-7 นั้นไว้ในสินค้า (Product) ของเขาเองโดยขอบเขตของ MPEG-7 นั้นจะอยู่ในส่วนของการอธิบายเนื้อหาของข้อมูลมัลติมีเดียเท่านั้นดังภาพ 7



ภาพ 7 แสดง Scope ของ MPEG-7 [6]

3. มุมมองทางเทคนิค (MPEG-7 Technical Overview)

เครื่องมือหลัก ๆ ที่ใช้ในการอธิบายข้อมูลของ MPEG-7 คือตัวอธิบาย [Descriptors (D)] รูปแบบคำอธิบาย [Description Schemes (DS)] และภาษาการนิยามคำอธิบาย [Description Definition Language (DDL)] ซึ่งเครื่องมือทั้ง 3 นี้จะมีความสัมพันธ์เกี่ยวเนื่องกันดังแสดงเป็นโครงสร้างได้ดังภาพ 8 ซึ่งเป็นโครงสร้างของคำอธิบาย (Structure of Descriptions)



ภาพ 8 แสดงโครงสร้างความสัมพันธ์ของอีลีเม้นหลัก ๆ ของ MPEG-7 [11]

จากภาพ 8 แสดงโครงสร้างความสัมพันธ์ระหว่างอีลีเม้นต่าง ๆ [11] โดย DDL จะเป็นตัวที่ใช้ในการนิยาม (Definition) ส่วนของ Ds (Descriptor) และ DSs (Description Schemes) ซึ่งจะกำหนดหรือจัดเตรียมวิธีสำหรับการกำหนดโครงสร้าง (Structuring) ของ Ds ใน DSs และ DDL นั้นยังยินยอมที่จะให้เราสามารถทำการเพิ่มขยาย (Extension) ในส่วนของ DSs เพื่อใช้ในการทำงานเฉพาะด้านได้ ส่วนของ Ds และ DSs จะถูกกระทำ Instantiation เป็น Description ที่อยู่ในรูปแบบของ XML (XML Format) ส่วนรูปแบบข้อมูลแบบฐานสอง (Binary Format) จะถูกใช้สำหรับการอธิบาย (Description) สำหรับการทำการจัดส่ง (Delivery) และการจัดเก็บ (Storage) ที่มีประสิทธิภาพ

3.1 ตัวอธิบายและรูปแบบคำอธิบาย (Descriptors and Description Schemes)

ตัวอธิบาย (Descriptor) เป็นตัวนิยามไวยากรณ์ (Syntax) และความหมาย (Semantics) ของฟีเจอร์เช่นกราฟสี (Color Histogram) ส่วนรูปแบบคำอธิบาย (Description Scheme) เป็นตัวกำหนดโครงสร้างและ ความหมายของความสัมพันธ์ (Relationships) ระหว่างองค์ประกอบ (Components) ต่าง ๆ ของมัน ซึ่งอาจเป็นได้ทั้ง Descriptors และ Description Schemes ยกตัวอย่างเช่น ImageText DS ซึ่งเป็น Description Scheme ที่สามารถอธิบาย

ตัวอักษร (Characters) ภาษา ขนาดของตัวอักษร (Font Size) รูปแบบของตัวอักษร (Font Style) ของข้อมูลที่เป็นตัวอักษร (Textual) ได้ซึ่งแสดงได้ดังภาพ 9 , 10 และ 11

ภาพ 9 แสดงตัวอย่างภาพที่มีตัวอักษรภายในภาพ [11]

```
<!-- ################################################ -->
<!-- Definition of ImageText DS                       -->
<!-- ################################################ -->
<complexType name="ImageTextType">
                              <complexContent>
    <extension base="mpeg7:StillRegionType">
      <sequence>
        <element name="Text" type="mpeg7:TextualType" minOccurs="0"/>
      </sequence>
      <attribute name="textType" use="optional">
        <simpleType>
          <restriction base="string">
            <enumeration value="superimposed"/>
            <enumeration value="scene"/>
          </restriction>
        </simpleType>
                                 </attribute>
      <attribute name="fontSize" type="positiveInteger" use="optional"/>
      <attribute name="fontType" type="string" use="optional"/>
    </extension>
  </complexContent>
</complexType>
```

ภาพ 10 แสดงคำนิยามของ ImageText DS [11]

```
<ImageText id="ImageText1" textType="scene">
  <RegionLocator>...</RegionLocator>
  <Text xml:lang="fr">Cherie, je suis la</Text>
</ImageText>
<!-- Image Text 2  -->
<ImageText id="ImageText2" textType="superimposed"
    fontSize="40" fontType="Arial">
  <RegionLocator>...</RegionLocator>
  <Text xml:lang="fr">
   En direct du state DE-KUIP Rotterdam Jean-Francois DEVELEY</Text>
</ImageText>
<!-- Image Text 3  -->
<ImageText id="ImageText3" textType="superimposed"
    fontSize="30" fontType="Times New Roman">
  <RegionLocator>...</RegionLocator>
  <Text xml:lang="fr">
   Nous allons assister a un match fantastique</Text>
</ImageText>
```

ภาพ 11 แสดงคำอธิบาย (Description) ของ MPEG-7 [11]

จากภาพ 10 เป็นตัวอย่างของคำนิยามของ ImageText DS (Definition of ImageText Description Schemes) ที่ได้จากภาพ 9 ส่วนภาพ 11 เป็นตัวอย่างของ Description ของ MPEG-7 (MPEG-7 Description) ที่ได้จากภาพ 11

สำหรับตัวอย่างไฟล์ XML ที่แสดงค่าของตัวอธิบาย (Descriptor Value) ต่าง ๆ ซึ่งเป็นส่วนที่ถูกนำมาใช้ในการประมวลผลของระบบค้นคืนภาพสามารถแสดงเป็นตัวอย่างได้ดังภาพ 12 และ 13

ภาพ 12 แสดงภาพตัวอย่างที่ใช้ในการดึง Content

ป
TK
6680.5
ป6218
2550
c.2

สำนักหอสมุด
1.3781733 พ.ย. 2550

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <Mpeg7 xmlns="http://www.mpeg7.org/2001/MPEG-7_Schema"
    xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance">
  - <DescriptionUnit xsi:type="DescriptorCollectionType">
    - <Descriptor xsi:type="ColorLayoutType">
        <YDCCoeff>52</YDCCoeff>
        <CbDCCoeff>26</CbDCCoeff>
        <CrDCCoeff>46</CrDCCoeff>
        <YACCoeff5>19 17 22 16 15</YACCoeff5>
        <CbACCoeff2>16 16</CbACCoeff2>
        <CrACCoeff2>14 15</CrACCoeff2>
      </Descriptor>
    </DescriptionUnit>
  </Mpeg7>
```

ภาพ 13 แสดงตัวอย่างของ Descriptor Value ของ ColorLayout Descriptor ใน XML file

3.2 ภาษาการนิยามคำอธิบาย (Description Definition Language)

ภาษาการนิยามคำอธิบาย [Description Definition Language (DDL)] คือ ภาษาที่ยอมให้มีการสร้าง Description Schemes และ Descriptors ตามมาตรฐานของ MPEG-7 ขึ้นมาใหม่ได้และยอมให้มีการทำการขยายและการแก้ไขส่วนของ Description Schemes ที่มีอยู่เดิมได้ ส่วน DDL Schema ซึ่งคือไฟล์แบบ DDL (DDL File) จะเป็นตัวกำหนดเงื่อนไขหรือข้อบังคับ (Constraints) ของ Description ที่ถูกต้องตามมาตรฐานของ MPEG-7 โดยจะถูกเข้ารหัส (Encoded) อยู่ใน XML ที่ใช้นิยามโครงสร้างของเอกสาร (Document Structure) และมันจะถูกฝัง (Embed) ไว้ในเอกสารโดยตรงผ่านการใช้ Markup โดย XML เป็นภาษาที่คล้ายกับ HTML แต่ต่างกันที่ผู้ใช้สามารถกำหนด Tag ขึ้นเองได้ สำหรับการนิยามของโครงสร้างเอกสารที่ถูกต้อง (Valid Document Structure) จะถูกแสดงอยู่ในไวยากรณ์ (Gramma) ที่เรียกว่า DTD (Document Type Definition) สำหรับตัวอย่างของ DDL ที่อยู่ในรูปของไฟล์ XML สามารถแสดงเป็นตัวอย่างได้ดังภาพ 14

```
<?xml version="1.0" encoding="UTF-8" ?>
<!--  ##################################################################  -->
<!--  ISO/IEC 15938 Information Technology - Multimedia Content           -->
<!--  Description Interface                                              -->
<!--  Part 2: Description Definition Language (ISO/IEC 15938-2)          -->
<!--  ##################################################################  -->
- <schema targetNamespace="urn:mpeg:mpeg7:schema:2001"
    xmlns="http://www.w3.org/2001/XMLSchema" xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
    elementFormDefault="qualified" attributeFormDefault="unqualified">
  + <annotation>
    <!--  ################################################  -->
    <!--  Definition of 'mpeg7:dim" for Matrix Datatype     -->
    <!--  ################################################  -->
    <!--  definition of listOfPositiveIntegerForDim datatype  -->
  + <simpleType name="listOfPositiveIntegerForDim">
    <!--  definition of mpeg7:dim attribute  -->
  + <attribute name="dim">
    <!--  ################################################  -->
    <!--  Definition of MPEG-7 Datatype Extensions          -->
    <!--  ################################################  -->
    <!--  definition of basicTimePoint datatype  -->
  + <simpleType name="basicTimePointType">
    <!--  definition of basicDuration datatype  -->
  + <simpleType name="basicDurationType">
  </schema>
```

ภาพ 14 แสดงตัวอย่าง DDL ใน XML file

3.3 ส่วนของมาตรฐาน (Parts of the Standard)

มาตรฐานของ MPEG-7 ประกอบด้วย 8 ส่วนคือ

1. Part 1 ระบบ (Systems) : จะกล่าวถึงสถาปัตยกรรม (Architecture) ของมาตรฐาน Syntax ของ Binary Format และเครื่องมือต่าง ๆ สำหรับการเตรียม (Preparing) คำอธิบาย (Descriptions) สำหรับการทำ Transport , Storage , Compressing Descriptions และ การทำ Synchronization ระหว่าง Content และ Descriptions

2. Part 2 ภาษาการนิยามคำอธิบาย (Description Definition Language) : จะกล่าวถึงภาษาที่ใช้ในการนิยามการอธิบาย (Description Definition Language)

3. Part 3 วิชวล (Visual) : จะกล่าวถึง Descriptors และ Description Schemes ตามมาตรฐานที่ใช้กับข้อมูลประเภท Visual

4. Part 4 ออดิโอ (Audio) : จะกล่าวถึง Descriptors และ Description Schemes ตามมาตรฐานที่ใช้กับข้อมูลประเภท Audio

5. Part 5 รูปแบบการอธิบายมัลติมีเดีย (Multimedia Description Schemes) : จะกล่าวถึง Descriptors และ Description Schemes ที่ใช้กับข้อมูลประเภท Visual หรือ Audio

6. Part 6 ซอพต์แวร์ (Reference Software) : จะกล่าวถึงเอกสารที่ประกอบด้วยซอพต์แวร์ที่อ้างอิงต่าง ๆ (Reference Software) ที่เรียกว่า eXperimentation Model หรือ XM

7. Part 7 การทดสอบ (Conformance Testing) : จะกล่าวถึงการแนะนำ (Guideline) และ กระบวนการ (Procedure) สำหรับการทดสอบ (Testing) โครงสร้าง (Conformance) ของการดำเนินงาน (Implementation) และกล่าวถึงแผนการทำงานของกลุ่ม MPEG (MPEG Work Plan)

8. Part 8 การดึงและการใช้ (Extraction and Use of MPEG-7 Descriptions) : จะกล่าวถึงการจัดการเกี่ยวกับการทำ Extraction และการใช้คำอธิบาย (Descriptions) ของ MPEG-7

4. สภาและสมาคมที่เกี่ยวข้องกับ MPEG-7 (MPEG-7 Related Forums and Consortiums)

ประกอบด้วยองค์กรต่าง ๆ [11] ดังนี้

4.1 Moving Picture Experts Group (MPEG) : เป็นกลุ่มการทำงาน (Working Group) ในการดูแลในส่วนของการพัฒนา (Development) ในส่วนของมาตรฐาน MPEG

4.2 The MPEG Industry Forum (MPEGIF) : ก่อนหน้านั้นถูกเรียกว่า MPEG-4 Industry Forum เป็นกลุ่มที่ไม่ได้เกี่ยวข้องทางผลประโยชน์กับการสนับสนุนองค์กรของมาตรฐาน MPEG และการให้บริการข้อมูล (Information) ในเรื่องเทคโนโลยี (Technology) สินค้า (Products) และการบริการ (Services) ของ MPEG ซึ่งเป็นกลุ่มดำเนินการเกี่ยวกับการทดสอบ (Test) และการจัดตั้ง (Organizes) การแสดง (Exhibitions) และการแข่งขัน (Events) โดยกลุ่มหรือองค์กรนี้ได้มีการเปลี่ยนชื่อใหม่เพราะปัจจุบันได้มีมาตรฐาน MPEG-7 และ MPEG-21 เกิดขึ้นมาแต่ MPEG-4 ก็ยังคงรักษารูปแบบการทำงานเดิมไว้ซึ่งปัจจุบันมีบริษัทที่เป็นสมาชิกประมาณ 100 บริษัท เช่น Apple, Hitachi, IBM, Microsoft, Nokia, RealNetworks และ Sony

4.3 The MPEG-7 Consortium (MP7C) : เป็นกลุ่มที่ให้ข้อมูล (Information) ในเรื่องของการกำหนดมาตรฐาน MPEG-7 และการแนะนำ (Guidance) และตัวอย่างในเรื่องของ

การใช้เทคโนโลยีของ MPEG-7 หน้าที่ของกลุ่มนี้คือการนำเทคโนโลยีของ MPEG-7 เข้าสู่ตลาด โดยการโฆษณาความสามารถในการทำงานของสินค้าและสนับสนุน (Support) ในเรื่องของการใช้งานเนื้อหามัลติมีเดีย (Multimedia Content) ที่ง่าย (Easy-to Use) ของมาตรฐานให้กับโรงงานต่าง ๆ

4.4 The World Wide Web Consortium (W3C) : เป็นกลุ่มที่พัฒนาข้อกำหนด (Specification) คำแนะนำ (Guideline) ซอฟต์แวร์ (Software) เครื่องมือ (Tools) ต่าง ๆ ที่เกี่ยวกับเว็บ มาตรฐาน W3C ถูกเรียกว่า Recommendations และ Technical Report อื่น ๆ ซึ่งเป็นส่วนที่สามารถใช้ได้ฟรี โดย W3C นั้นมีผู้ดูแลอยู่ 3 องค์กรใน 3 แห่ง คือ ใน United State, Europe and Japan ซึ่งส่วนหนึ่งของเทคโนโลยีที่ W3C ได้พัฒนาคือ XML Schema ซึ่งเป็นภาษาพื้นฐานสำหรับ MPEG-7 DDL

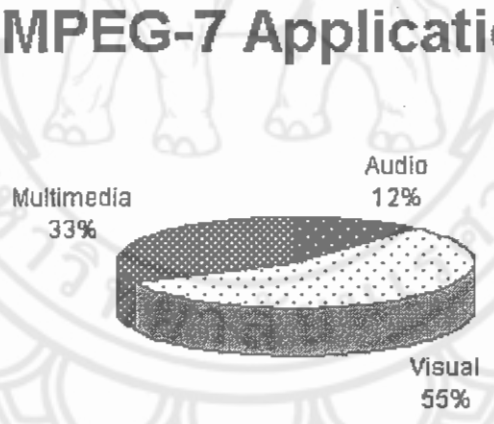
5. ตัวอย่างของข้อมูลด้านต่าง ๆ ที่สามารถนำ MPEG-7 ไปประยุกต์ใช้งานได้ ดังแสดงเป็นตัวอย่าง [7] ดังนี้

1. Digital Libraries (Image Catalogue, Musical Dictionary)
2. Multimedia Directory Services (เช่น Yellow Pages)
3. Broadcast Media Selection (Radio Channel, TV Channel)
4. Multimedia Editing (Personalised Electronic News Service, Media Authoring)
5. Education
6. Journalism (เช่น Searching Speeches of a Certain Politician Using his Name, His Voice or His Face)
7. Tourist Information
8. Cultural Services (History Museums, Art Galleries)
9. Entertainment (เช่น Searching a Game, Karaoke)
10. Investigation Services (Human Characteristics Recognition, Forensics)
12. Geographical Information Systems
13. Remote Sensing (Cartography, Ecology, Natural Resources Management)
14. Surveillance (Traffic Control, Surface Transportation, Non-Destructive Testing in Hostile Environments)

15. Bio-Medical Applications
16. Shopping (เช่น Searching for Clothes that you like)
17. Architecture, Real Estate, and Interior Design
18. Social (เช่น Dating Services)
19. Film, Video and Radio Archives

6. **มุมมองด้านสัดส่วนของโปรแกรมต่าง ๆ ของ** MPEG-7 (Overview of List of MPEG-7 Applications, Demos and Projects)

ในปัจจุบันได้มีการพัฒนาโปรแกรมที่อ้างอิงตามมาตรฐานของ MPEG-7 เพิ่มมากขึ้นทั้งในส่วนของโปรแกรมที่พัฒนาโดยบุคคลภายนอกและโปรแกรมที่พัฒนาโดยกลุ่มของ MPEG-7 เอง ซึ่งโปรแกรมต่าง ๆ ดังกล่าวก็จะมีการพัฒนาขึ้นมาเพื่อจัดการกับข้อมูลประเภทต่าง ๆ เช่น Video, Audio และ Image เป็นต้น ซึ่งสามารถที่จะแสดงเป็นสัดส่วนของเปอร์เซ็นต์การใช้งานตามประเภทข้อมูลได้ดังภาพ 15



ภาพ 15 แสดงเปอร์เซ็นต์การประยุกต์ใช้งานของ MPEG-7 กับข้อมูลประเภทต่าง ๆ [7]

สำหรับตัวอย่างของ Applications, Demos และ Projects ต่าง ๆ สามารถดูเพิ่มเติมได้ใน [7]

7. **ตัวอธิบายวิชวลของ** MPEG-7 (MPEG-7's Visual Descriptors)

สำหรับตัวอย่างของ Visual Descriptor และคุณสมบัติหรือลักษณะของ Visual Descriptors ต่าง ๆ ที่มีอยู่ในมาตรฐานของ MPEG-7 ซึ่งส่วนใหญ่จะเป็น Visual Feature

พื้นฐานต่าง ๆ เช่น Color, Shape และ Texture ซึ่งจะแสดงอยู่ใน ภาคผนวก ก ซึ่งผู้วิจัยได้สำเนาข้อมูลมาจาก [12] โดยในงานวิจัยนี้จะไม่ขอกล่าวถึงรายละเอียดของ Visual Descriptors ทั้งหมดเนื่องจากอยู่นอกเหนือจากขอบเขตงานวิจัยโดยจะขออธิบายรายละเอียดเฉพาะ Visual Descriptors ที่เกี่ยวข้องกับงานวิจัยนี้เท่านั้นซึ่งได้อธิบายไว้ในบทที่ 3

8. สรุปข้อดีของ MPEG-7 (Advantages of MPEG-7 – A Summary) [6]

1. MPEG ประสบความสำเร็จในตลาด (MPEG Standards have been Successful in the Marketplace)

มาตรฐาน MPEG ประสบความสำเร็จในตลาดเนื่องจากมีกระบวนการที่ไม่เหมือนใคร (Unique) ในการทำ Sharing Leading-Edge Technology ขณะที่มีการป้องกันเรื่องทรัพย์สินทางปัญญาด้วย

2. การใช้ข้อดีหรือความได้เปรียบด้านความรู้ความชำนาญของ MPEG-7 (Taking Advantage of MPEG-7 Expertise)

ผู้ที่มีส่วนร่วมกับ MPEG-7 ประกอบด้วยทุก ๆ ส่วนของ Content Value Chain คือ Production, Post-Production, Delivery และ Consumption ซึ่งกระบวนการทั้งหมดนี้ MPEG-7 ได้มีการกำหนดมาตรฐานในเรื่องรูปแบบของคำอธิบาย (Description Schemes) สำหรับการทำ Content Description, Management และ Organization เช่นเดียวกับการทำ Navigation, Access, User Preferences และ Usage History

3. ความสามารถในการทำงานร่วมกัน (Interoperability)

ความสามารถในการทำงานร่วมกันกับมาตรฐานอื่น ๆ ซึ่งเป็นการยกระดับของ MPEG-7 ที่รวดเร็วซึ่งถูกสร้างขึ้นบนความสามารถทางด้านเทคโนโลยีและมาตรฐาน MPEG-7 เป็นการประสานกัน (Harmonize) กับมาตรฐานอื่น ๆ ที่มีใช้อยู่ซึ่งแสดงให้เห็นถึงความสำเร็จและการยอมรับทั้งในด้านของธุรกิจสื่อแบบเดิม (Traditional Media Business) และธุรกิจสื่อใหม่ ๆ (New Media Business) เช่น W3C (XML, XML Schema), IETF (URI, URN, URL), Dublin Core, ISO/ANSI Thesaurus Guidelines, SMPTE Metadata Dictionary, TV-Anytime เป็นต้น ซึ่งความสามารถในการใช้งานร่วมกับระบบอื่น ๆ ดังกล่าวนี้ทำให้ MPEG-7 สามารถรวมเข้ากับสินค้าของบริษัทต่าง ๆ ได้อย่างรวดเร็ว

4. สามารถแลกเปลี่ยนข้อมูลระหว่างบริษัทสาขา (Data Exchange between Subsidiaries)

MPEG-7 ช่วยให้ระบบการจัดการ Content (Content Management System) ในสาขาหนึ่งสามารถที่จะจัดการกับ Content ในอีกสาขาหนึ่งได้ เช่น Broadcasters และ Content Producers สามารถที่จะทำการ Categorize, Exchange, Process และ Manage กับ Content ข้ามสาขาได้

5. ความสามารถด้านการตลาดของ MPEG-7 (Market Potential for MPEG-7 Applications)

ตามการคาดคะเนหรือการวางแผนของกลุ่ม Goldman Sachs (Goldman Sachs projection) ระบุว่าตลาดสำหรับเครื่องมือในการจัดการ Content (Content Management Tools) จะเพิ่มขึ้นจาก 378 ล้านดอลลาสหรัฐ (US$378Million) ในปี 2000 เป็น 4.5 ล้านล้านดอลลาสหรัฐ (US$4.5Billion) ในปี 2005 โดยเครื่องมือที่มีความสามารถในการใช้งานร่วมกับระบบอื่น (Interoperable Tools) จะขายดีกว่าเครื่องมือที่ไม่มีความสามารถในการทำงานร่วมกับระบบอื่น ๆ (Non-Interoperable Tools) ซึ่ง MPEG-7 เป็น Gold Standard สำหรับความสามารถในการใช้งานร่วมกันในการจัดการ Content (Content Management Interoperability) ซึ่งไม่ใช่เฉพาะบริษัทบันเทิง (Entertainment Companies) แต่รวมถึงทุกบริษัท (Every Company) ทุกโรงงาน (Every Industry) และ ทุกที่ (Everywhere)

6. MPEG-7 ช่วยให้เกิดโปรแกรมด้านมัลติมีเดียใหม่ ๆ (MPEG-7 will Enable a New Generation of Multimedia Applications)

MPEG-7 ได้กำหนดมาตรฐานของเครื่องมือในการอธิบายข้อมูลมัลติมีเดีย (Multimedia Description Tools) สำหรับ Content ที่ครอบคลุมซึ่ง Descriptions สำหรับ Catalogue Level (เช่น Title), Semantic Level (เช่น Who, What, When, Where) และ Structural Level (เช่น Spatio-Temporal Region, Color Histogram, Timbre, Texture) จะเป็นส่วนที่กำหนดหรือจัดเตรียมเครื่องมือสำหรับผู้พัฒนาในการที่จะสร้างโปรแกรมสำหรับข้อมูลมัลติมีเดียใหม่ ๆ เช่น Search Engine, Digital Libraries, Educational เป็นต้น

7. สามารถใช้งาน MPEG-4 และ MPEG-7 ร่วมกันได้ (MPEG-4 and MPEG-7 Tools for Killer Applications)

MPEG-4 สามารถนำมารวมเข้ากับ MPEG-7 เพื่อเพิ่มประสิทธิภาพในการจัดการ Content ในด้านต่าง ๆ เช่น การทำ Streaming, Manipulation, Indexing และ Retrieval

8. MPEG-7 มีการป้องกันเรื่องการจัดการและทรัพย์สินทางปัญญา (MPEG-7 Intellectual Property and Management Protection)

MPEG ได้ทำงานร่วมกับกลุ่มผู้ผลิตอย่างใกล้ชิดเพื่อให้แน่ใจว่ามีการป้องกันเรื่องทรัพย์สินของผู้ที่เป็นเจ้าของข้อมูลทั้งในส่วนของ Content และ Metadata

9. MPEG-7 ช่วยทำให้ Content ต่าง ๆ มีประโยชน์และคุณค่ามากขึ้น (MPEG-7 Makes Content More Valuable)

ด้วยเทคโนโลยีในการทำดัชนีของ MPEG-7 (MPEG-7 Indexing Technology) จะช่วยให้ Audio-Visual Content ที่เพิ่มมากขึ้นทุกปี เช่น Libraries นั้นมีประโยชน์มากขึ้น เนื่องจาก MPEG-7 มีวิธีการในการเข้าถึงข้อมูลที่ครอบคลุมสำหรับผู้ใช้ในการเข้าถึง (Access) และการค้นคืน (Retrieve) รายละเอียดของ Descriptions ของ Content

10. MPEG-7 จัดเตรียมระบบการจัดการ Content ที่ชาญฉลาด (MPEG-7 Provides a Seamless Path Towards Increasingly Intelligent Content Management Systems)

จากกระบวนการผลิตจนถึงการบริโภคข้อมูลสารสนเทศ อุปกรณ์ทางด้าน Technical Hardware และโครงสร้างพื้นฐานของการติดต่อสื่อสาร (Communication Infrastructure) จะมาถึงจุดที่การคำนวณ (Computing) และการติดต่อสื่อสาร (Communication) จะกลายเป็นสิ่งที่ถูกผูกติดกับ Objects และ สภาพแวดล้อม (Environment) ทุก ๆ วันของมนุษย์ สื่อ (Media) จะกลายเป็นสิ่งที่แพร่หลายและสื่อที่แพร่หลายจะทำให้เกิดความต้องการ (Demand) สำหรับ Content ใหม่ ๆ ซึ่งจะส่งผลถึงการเปลี่ยนแปลงพื้นฐานต่อกระบวนการผลิต (Production) การจัดการ (Management) และการจัดส่ง (Delivery) สื่อ และสื่อที่ได้รับเพิ่มมากขึ้นจะส่งผลการจัดการข้อมูลด้วยมือเป็นไปได้ยากขึ้นทำให้การจัดการข้อมูลจะถูกกระทำผ่านตัวโปรแกรมแทนส่งผลให้ผู้ใช้สามารถใช้ข้อมูลต่าง ๆ ได้ง่ายและสะดวกมากขึ้น

จากมาตรฐานของ MPEG-7 ดังที่อธิบายข้างต้นผู้วิจัยได้นำเอาหลักการของมาตรฐานดังกล่าวมาใช้เป็นพื้นฐานในการพัฒนางานวิจัยโดยส่วนของมาตรฐานของ MPEG-7 ที่ผู้วิจัยได้นำมาใช้ในงานวิจัยนั้นคือส่วนที่ 3 ของมาตรฐาน (Part 3 : Visual) ซึ่ง Visual Descriptor ที่ใช้ประกอบด้วย Color Layout Descriptor, Edge Histogram Descriptor และ Region-Based Shape Descriptor โดยค่าของแต่ละ Descriptor จะถูกเก็บอยู่ในรูปของ XML file ซึ่งถูกสร้างขึ้นโดยใช้ Application ที่พัฒนาขึ้นโดยอ้างอิงกับมาตรฐาน MPEG-7

เครื่องมือสรุปวีดีโอแบบอัตโนมัติ (Automatic Video Summarixing Tool using MPEG-7 Descriptor for Personal Video Recorder)

ในส่วนนี้จะเป็นงานวิจัยหนึ่งที่พัฒนาเครื่องมือที่ใช้สำหรับการสรุปเนื้อหาของข้อมูลใน Video โดยใช้ MPEG-7 Descriptor เพื่อใช้ในเครื่องบันทึก Video (Personal Video Recorder) ที่เสนอโดย Jae-Ho Lee, Gwang-Gook Lee และ Whoi-Yul Kim [8] ซึ่งได้มีการใช้ฟังก์ชันสำหรับการเปรียบเทียบความแตกต่างระหว่างภาพโดยเปรียบเทียบจากความแตกต่างของ Descriptor ประเภทสีคือ Color Layout Descriptor ซึ่งเป็นฟังก์ชันที่ถูกนำมาใช้ในงานวิจัยที่พัฒนาโดยจะขออธิบายเฉพาะส่วนของฟังก์ชันที่นำมาใช้คือ ฟังก์ชันเปรียบเทียบค่าความแตกต่าง (Distance) ระหว่าง 2 Color Layout Descriptor ซึ่งใช้ค่าของสัมประสิทธิ์ Discrete Cosine Transform (DCT) ที่ได้จาก YCbCr (Color Space) โดยมีค่าของสัมประสิทธิ์คือ 6 (สัมประสิทธิ์ของ Y), 3 (สัมประสิทธิ์ของ Cb และ Cr) โดยมีสมการในการเปรียบเทียบดังสมการ 5 [8]

$$D(i,j)=\sqrt{\sum_{k=0}^{6}\lambda_{yk}\left(Y_{coeff}(i,k)-Y_{coeff}(j,k)\right)}$$

$$+\sqrt{\sum_{k=0}^{3}\lambda_{Cb_k}\left(Cb_{coeff}(i,k)-Cb_{coeff}(j,k)\right)}$$

$$+\sqrt{\sum_{k=0}^{3}\lambda_{Cr_k}\left(Cr_{coeff}(i,k)-Cr_{coeff}(j,k)\right)} \quad \text{.....(5)}$$

โดยที่ i และ j คือ ดัชนีของหมายเลขภาพที่ใช้เปรียบเทียบ k คือจำนวนของสัมประสิทธิ์และ λ คือสัมประสิทธิ์ของน้ำหนัก (Weight Coefficient) ซึ่งกำหนดค่าได้ดังตาราง 1

ตาราง 1 แสดงค่าของ Weight Coefficient [8]

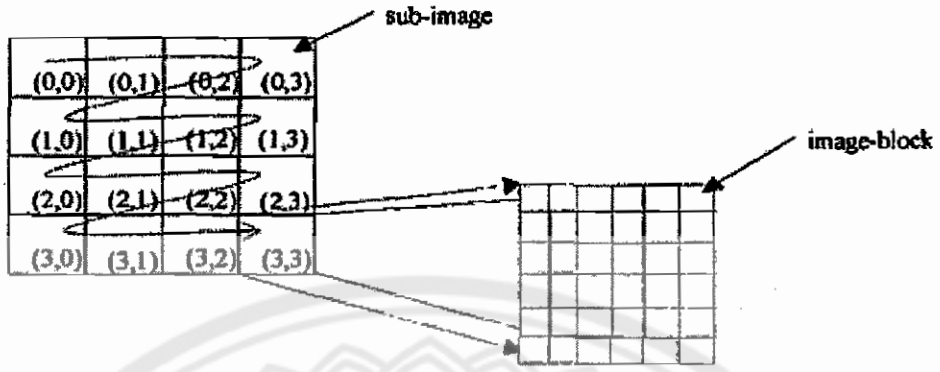
	λ					
	0	1	2	3	4	5
Y_{coeff}	2	2	2	1	1	1
Cb_{coeff}	2	1	1	-	-	-
Cr_{coeff}	2	1	1	-	-	-

การใช้ตัวอธิบายเส้นแบบโลคอลที่มีประสิทธิภาพ (Efficient Use of Local Edge Histogram Descriptor)

งานวิจัยนี้ถูกเสนอโดย Dong Kwon Park, Yoon Seok Jeon และ Chee Sun Won [9] ซึ่งเสนอวิธีการสร้าง Global และ Semi-Global Edge Histogram จาก Local Edge Histogram เพื่อเพิ่มประสิทธิภาพของการสืบค้นข้อมูลภาพโดยที่ Edge Histogram Descriptor นี้ตามมาตรฐานของ MPEG-7 นั้นจะกำหนดขนาด (Size) ของ Local Edge Histogram Bins ไว้ที่ 80 Bins ซึ่งวิธีการสร้าง Local, Global และ Semi-Global Edge Histogram นั้นจะอธิบายตามหัวข้อดังต่อไปนี้

1. **การแบ่งเนื้อที่ภาพสำหรับการระบุและกำหนดตำแหน่งของเส้น** (Partition of Image Space for Edge Identification and Localization)

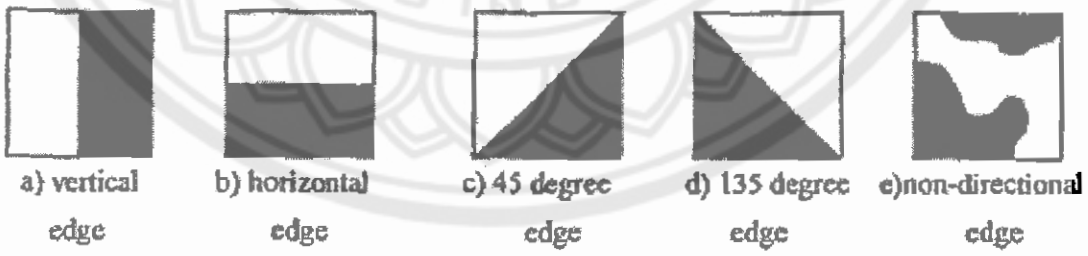
ในการกำหนดตำแหน่งทางโลคอล (Localize) ของการกระจายตัวของเส้น (Edge Distribution) บนภาพที่มีของเขตที่แน่นอนนั้นกระทำได้โดยแบ่งเนื้อที่ของภาพ (Image Space) ออกเป็นภาพย่อย ๆ (Sub-Image) ที่มีขนาด 4 x 4 Sub-Image และทำการสร้าง Edge Histogram ของการกระจายตัวของเส้น (Edge Distribution) ที่ปรากฏอยู่ใน Sub-Image เหล่านั้นโดยในการกำหนดชนิดของเส้น (Edge Type) ที่แตกต่างกันนั้นส่วนของ Sub-Image จะถูกแบ่งออกเป็นช่องสี่เหลี่ยม (Square Block) ย่อย ๆ อีกซึ่งเรียกว่า Image-Block ดังที่แสดงอยู่ในภาพ 16 และส่วนของขนาดและจำนวนของ Image-Block ของแต่ละ Sub-Image นั้นจะอธิบายอยู่ในหัวข้อที่ 4. วิธีการดึงเส้น (Edge Extraction Method)



ภาพ 16 คำนิยามของ Sub-Image และ Image-Block [9]

2. **ชนิดของเส้น** (Edge Types)

ชนิดของเส้นแบ่งได้เป็น 5 ชนิดซึ่งถูกกำหนดอยู่ใน Edge Histogram Descriptor โดยแบ่งได้เป็นเส้นที่มีทิศทาง (Directional Edge) จำนวน 4 เส้น และเส้นที่ไม่มีทิศทาง (Non-Directional Edge) จำนวน 1 เส้นสำหรับเส้นที่มีทิศทางจะประกอบด้วยเส้นในแนวตั้ง (Vertical) เส้นในแนวนอน (Horizontal) เส้นในแนว 45 องศา (45 Degree Diagonal) และเส้นในแนว 135 องศา (135 Degree Diagonal) ดังแสดงในภาพ 17 โดยเส้นที่มีทิศทางนี้จะถูกดึง (Extract) มาจาก Image-Block ถ้า Image-Block นั้นไม่มีเส้นที่มีทิศทางอยู่ก็จะถูกกำหนดให้อยู่ในกลุ่มของเส้นที่ไม่มีทิศทางแทนสำหรับการดึงข้อมูลของเส้น (Edge Information) นั้นจะอธิบายอยู่ในหัวข้อที่ 4. วิธีการดึงเส้น (Edge Extraction Method)



ภาพ 17 ชนิดของเส้น (Edge Type) [9]

3. **ความหมายของกราฟเส้นแบบโลคอล** (Semantic of Local Edge Histogram)

ภายหลังจากที่ดึงเส้น (Edge Extraction) จาก Image-Block แล้วเราจะนับจำนวนของเส้นทั้งหมดของแต่ละชนิด (Edge Type) ในแต่ละ Sub-Image เนื่องจากชนิดของเส้นมีทั้งหมด 5 ชนิดดังนั้นเราจะกำหนดจำนวนของ Histogram Bins ได้เป็น 5 Histogram Bins ในแต่ละ Sub-Image และจากการแบ่งภาพออกเป็น 4 x 4 Sub-Image ซึ่งมีจำนวนเท่ากับ 16 Sub-Image ดังนั้นเราจะได้จำนวนของ Histogram Bins เท่ากับ 80 Bins (16 x 5) สำหรับ Edge Histogram ซึ่งจากการสแกนภาพตามลำดับดังแสดงในภาพ 16 จะได้ความหมาย (Semantic) ของ Local Edge Bins ออกมาซึ่งสามารถกำหนดได้ดังตาราง 2

ตาราง 2 ความหมายของ Local Edge Bins [9]

Histogram bins	Semantics
Local_Edge [0]	Vertical edge of sub-image at (0,0)
Local_Edge [1]	Horizontal edge of sub-image at (0,0)
Local_Edge [2]	45degree edge of sub-image at (0,0)
Local_Edge [3]	135degree edge of sub-image at (0,0)
Local_Edge [4]	Non-directional edge of sub-image at (0,0)
Local_Edge [5]	Vertical edge of sub-image at (0,1)
:	:
Local_Edge [74]	Non-directional edge of sub-image at (3,2)
Local_Edge [75]	Vertical edge of sub-image at (3,3)
Local_Edge [76]	Horizontal edge of sub-image at (3,3)
Local_Edge [77]	45degree edge of sub-image at (3,3)
Local_Edge [78]	135degree edge of sub-image at (3,3)
Local_Edge [79]	Non-directional edge of sub-image at (3,3)

4. **วิธีการดึงเส้น** (Edge Extraction Method)

ในการดึง Feature ของเส้นทั้งแบบมีและไม่มีทิศทางนั้นจะกระทำโดยการกำหนดขนาดของ Image-Block ให้มีขนาดเล็กโดยการแบ่งเนื้อที่ของภาพ (Image Space) ออกเป็นช่องสี่เหลี่ยม (Square Block) ที่ไม่เหลื่อมล้ำกัน (Non-Overlapping) แล้วจึงทำการดึงข้อมูลของเส้น (Edge Information) จาก Block เหล่านั้นโดยไม่มีการพิจารณาในส่วนของขนาดของภาพซึ่งจะทำการแบ่ง Sub-Image ออกเป็น Image-Block ในจำนวนที่แน่นอน (Fixed) นั่นคือขนาดของ

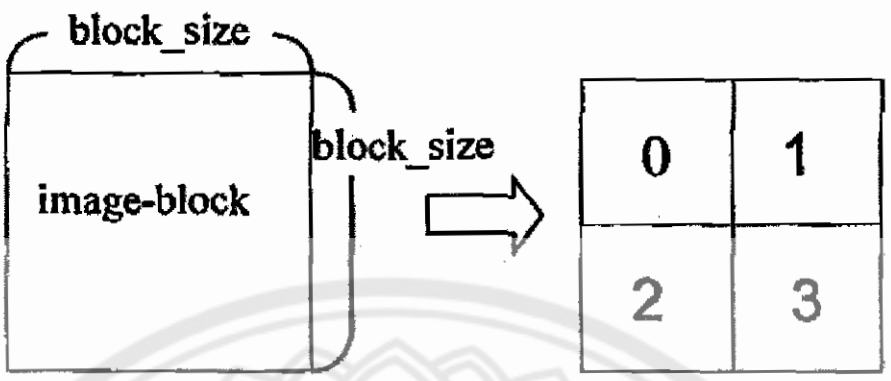
Image-Block นั้นจะเป็นสัดส่วนกับขนาดเริ่มต้นของภาพ (Original Image) สำหรับการคำนวณขนาดของ Image-Block นั้นสามารถคำนวณได้จากสมการ 6 และ 7 [9]

$$x = \sqrt{\frac{image_width \times image_height}{desired_num_block}} \quad \text{.....(6)}$$

$$block_size = \left| \frac{x}{2} \right| \times 2 \quad \text{.....(7)}$$

โดยที่ Image_Width และ Image_Height คือขนาดของภาพในแนวนอนและแนวตั้งตามลำดับ Desired_Num_Block คือจำนวนของ Block ที่ต้องการ

Feature ของเส้น (Edge Feature) จากถูกดึงมาจาก Image-Block ดังที่แสดงในภาพ 18 โดยที่ Image-Block นี้จะถูกแบ่งออกเป็น 4 Sub-Block หลังจากนั้นค่ากลางของลูมิแนน (Luminance Mean Value) สำหรับ 4 Sub-Block จะถูกใช้สำหรับค้นหาเส้น (Edge Detection) และค่ากลาง (Mean Value) ของ 4 Sub-Block ที่ได้นั้นจะเชื่อมโยงกับสัมประสิทธิ์ของตัวฟิวส์เตอร์ (Filter Coefficient) ที่แสดงอยู่ในภาพ 19 ซึ่งจะได้ค่าของ Edge Magnitude ออกมาโดยใช้สมการ 9 ถึงสมการ 13 เราจะได้ค่าความยาวของเส้นที่มีทิศทาง (Directional Edge Strength) ออกมาโดยในระหว่างคำนวณหาค่าของความยาวของเส้นที่มีทิศทางของเส้นทั้ง 5 ชนิดนั้นถ้าค่าสูงสุด (Maximum) ของค่าเหล่านั้นมีค่ามากกว่าค่าเริ่มต้น (Threshold Value) คือ Th_{Edge} จะกำหนดให้ Block นั้นมีชนิดของเส้นที่สอดคล้องกันนอกจากนี้จะกำหนดสัญลักษณ์ (Label) ของ Sub-Block จาก 0 ถึง 3 ดังแสดงในภาพ 18 ซึ่ง Sub-Block ที่ k (k=0, 1, 2, 3) ของ Image-Block ที่ (i, j) จะสามารถคำนวณหาค่าเฉลี่ยของระดับของสีเทาได้ (Average Gray Level) ซึ่งถูกกำหนดอยู่ในตัวแปล $A_k(i, j)$

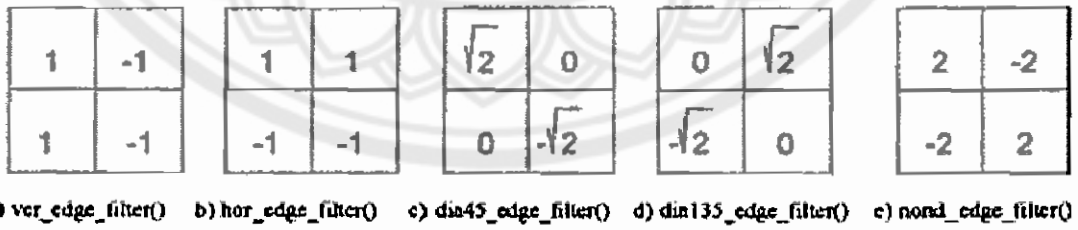


ภาพ 18 Sub-Block และการกำหนด Label [9]

จากการกำหนดสัญลักษณ์ (Label) ดังภาพ 18 จะได้สัมประสิทธ์ของตัวกรองเส้นในแนวตั้ง (Vertical Edge Filter) ดังแสดงในภาพ 19 – a ซึ่งสามารถกำหนดเป็นตัวแปล [9] ดังนี้

$$ver_edge_filter(0) = 1$$
$$ver_edge_filter(1) = -1$$
$$ver_edge_filter(2) = 1$$
$$ver_edge_filter(3) = -1$$

ในลักษณะเดียวกันเราสามารถกำหนดสัมประสิทธ์ของตัวกรอง (Filter Coefficient) ของตัวกรองเส้น (Edge Filter) ชนิดอื่น ๆ ได้ดังภาพ 19 – b ถึง 19 - e เช่นเดียวกัน



ภาพ 19 ตัวกรองสำหรับค้นหาเส้น (Filter for Edge Detection) [9]

จากการใช้ตัวกรองเส้น (Edge Filter) ทั้ง 5 ชนิดเราจะได้ความยาวเส้น (Edge Strength) ทั้ง 5 ชนิดสำหรับ Image-Block ที่ (i, j) ดังสมการ 9 – 13 [9] ดังนี้

$$ver_edge_stg(i,j)=\sum_{k=0}^{3}|A_k(i,j)\times ver_edge_filter(k)| \quad (9)$$

$$hor_edge_stg(i,j)=\sum_{k=0}^{3}|A_k(i,j)\times hor_edge_filter(k)| \quad (10)$$

$$dia45_edge_stg(i,j)=\sum_{k=0}^{3}|A_k(i,j)\times dia45_edge_filter(k)| \quad (11)$$

$$dia135_edge_stg(i,j)=\sum_{k=0}^{3}|A_k(i,j)\times dia135_edge_filter(k)| \quad (12)$$

$$non_edge_stg(i,j)=\sum_{k=0}^{3}|A_k(i,j)\times non_edge_filter(k)| \quad (13)$$

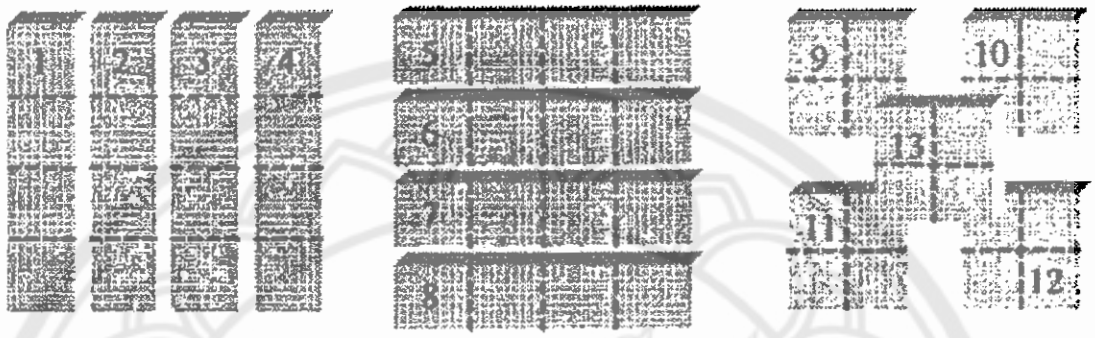
ถ้าค่าสูงสุด (Maximum) ของความยาวเส้น (Edge Strength) ทั้ง 5 ชนิดที่ได้จากสมการ 9 ถึง 13 มีค่ามากกว่าค่าเริ่มต้น (Threshold) คือ Th_{Edge} ดังที่แสดงอยู่ในสมการ 14 [9] แสดงว่า Image-Block นั้นมีเส้นที่มีลักษณะคล้ายกันกับเส้นชนิดนั้น ๆ

$$\max\begin{Bmatrix} ver_edge_stg(i,j), hor_edge_stg(i,j), dia45_edge_stg(i,j), \\ dia135_edge_stg(i,j), nond_edge_stg(i,j) \end{Bmatrix} > Th_{edge} \quad (14)$$

5. การสร้าง Global และ Semi-Global Edge Histogram

สำหรับกราฟเส้นแบบโกบอล (Global Edge Histogram) นั้นจะเป็นตัวที่แสดงถึงการกระจายตัวของเส้น (Edge Distribution) ภายในภาพ (Image Space) ทั้งหมดและเนื่องจากเส้นมีทั้งหมด 5 ชนิด ดังนั้น Global Edge Histogram ก็จะมีจำนวน 5 Bins ด้วย ส่วน Semi-Global Edge Histogram นั้นกระทำโดยการจับกลุ่มของภาพย่อย (Sub-Image) ที่เชื่อมต่อกัน (Connected) ออกเป็น 13 กลุ่มดังภาพ 20 ซึ่งในแต่ละกลุ่มจะทำการสร้างการกระจายตัวของเส้น (Edge Distribution) สำหรับชนิดของเส้นทั้ง 5 ชนิดซึ่ง Semi-Global Edge Histogram นี้จะใช้สำหรับแสดงถึงการกระจายตัวของเส้นภายในภาพที่มีพื้นที่ที่กว้าง (Larger Area) มากขึ้นกว่ากราฟเส้นแบบโลคอล (Local Edge Histogram) ที่มีอยู่เดิมยกตัวอย่างเช่นในกลุ่มที่ 1 ถึง 4 จะเป็นส่วนที่แสดงถึงการเชื่อมต่อกันของเส้น (Edge Connectivity) ในแนวตั้งและกลุ่มที่ 5 ถึง 8 จะเป็นส่วนที่แสดงถึงการเชื่อมต่อกันของเส้น (Edge Connectivity) ในแนวนอนเป็นต้นซึ่งหลังจากสร้าง Edge Histogram ทั้ง 3 ประเภท (Local, Global และ Semi-Global) แล้วก็จะ

ได้จำนวนของ Histogram Bins ทั้งหมดคือ 80 Bin (Local) + 5 Bin (Global) + 65 Bin (13 x 5, Semi-Global) = 150 Bin จากนั้นก็นำค่า Histogram Bin นี้ไปใช้สำหรับการดำเนินงานในขั้นตอนอื่น ๆ ต่อไป



ภาพ 20 การจับกลุ่มของ Sub-Image [9]

จากงานวิจัยดังกล่าวนี้ผู้วิจัยได้นำหลักการในการสร้าง Global และ Semi-Global Edge Histogram จาก Local Edge Histogram ไปประยุกต์ใช้กับงานวิจัยในส่วนของกระบวนการของการค้นหาภาพโดยใช้ตัวอธิบายเสมือน (Visual Descriptor) ประเภท Edge Histogram Descriptor (EHD) เพื่อเพิ่มประสิทธิภาพของระบบค้นคืนภาพที่พัฒนา