

ระบบการสืบค้นวิดีโอโดยใช้ภาพและเสียง

Video search engine using audiovisual content

นายชัชชา ชาวเหนือ รหัส 47380018
นายสุทธิจักร เครือสาร รหัส 47380039

ห้องสมุดคณะวิศวกรรมศาสตร์
วันที่รับ..... 7 / ส.ย. 2553 /

เลขทะเบียน..... 14942949

เลขเรียกหนังสือ..... ๗๕

มหาวิทยาลัยนเรศวร ๕๓๕1 ๕

2550

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร


ปีการศึกษา 2550

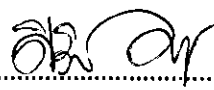



ใบรับรองโครงการวิศวกรรม

หัวข้อโครงการ ระบบการสืบค้นวิดีโอโดยใช้ภาพและเสียง
ผู้ดำเนินโครงการ นายชัชชา ชามเหนือ รหัส 47380018
นายฤทธิจักร เครือสาร รหัส 47380039
อาจารย์ที่ปรึกษา ดร. ไพศาล มณีสว่าง
สาขาวิชา วิศวกรรมคอมพิวเตอร์
ภาควิชา วิศวกรรมไฟฟ้าและคอมพิวเตอร์
ปีการศึกษา 2550

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น อนุมัติให้โครงการฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์
คณะกรรมการสอบโครงการวิศวกรรม


.....ประธานกรรมการ
(ดร.ไพศาล มณีสว่าง)


.....กรรมการ
(อาจารย์ศิริพร เดชะศิลารักษ์)


.....กรรมการ
(ดร. อัครพันธ์ วงศ์กังแห)

หัวข้อโครงการ	ระบบการสืบค้นวีดีโอโดยใช้ภาพและเสียง
ผู้ดำเนินโครงการ	นายชัชชา ชาวเหนือ รหัส 47380018
	นายฤทธิจักร เครือสาร รหัส 47380039
อาจารย์ที่ปรึกษา	ดร. ไพศาล มณีสว่าง
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
ภาควิชา	วิศวกรรมไฟฟ้าและคอมพิวเตอร์
ปีการศึกษา	2550

บทคัดย่อ

โครงการนี้ศึกษาและพัฒนาโปรแกรมการค้นหาไฟล์วีดีโอคลิปที่มีอยู่แล้ว โดยใช้หลักการของ Adaptive Video Indexing (AVI) ในการคำนวณหาเวกเตอร์ของภาพและใช้หลักการของการแปลงเวฟเลต (Wavelet Transform) ในการคำนวณหาเวกเตอร์ของเสียง โดยที่โครงการนี้จะใช้การเปรียบเทียบของภาพ, เสียง และการรวมเวกเตอร์ของภาพและเสียงเข้าด้วยกันเพื่อใช้ในการสืบค้น จากนั้นนำไปเปรียบเทียบกับฐานข้อมูล ที่มีอยู่โดยการเปรียบเทียบของภาพจะใช้ Index ของ Video มาเก็บไว้ในรูปแบบของ String มาเทียบกัน ส่วนการเปรียบเทียบแบบเสียงจะใช้การเปรียบเทียบของ Wavelet Transform และการเปรียบเทียบของภาพและเสียงจะใช้การเปรียบเทียบของภาพกับเสียงรวมกัน แล้วนำไฟล์วีดีโอคลิป 16 ไฟล์ที่เหมือนกันมาแสดง โดยโปรแกรมที่พัฒนานี้จะใช้วิชวลเบสิกคอตเน็ต (Visual Basic.NET) ของไมโครซอฟท์วิชวลสตูดิโอคอตเน็ต (Microsoft Visual Studio. NET) และมีการจัดการฐานข้อมูลโดยใช้ไมโครซอฟท์แอคเซส (Microsoft Access) เป็นเครื่องมือในการจัดการ

โดยผลการทดลอง โปรแกรมสามารถค้นหาไฟล์วีดีโอคลิปจากฐานข้อมูลได้ใกล้เคียงมากกับไฟล์วีดีโอคลิปที่เลือก

Project title Video search engine using audiovisual content.

Name Mr.Chatcha chawnua ID 47380018

Mr. Ridthichak Kruasarn ID 47380039

Project advisor Paisarn Munecsawang , Ph.D.

Major Computer Engineering.

Department Electrical and Computer Engineering.

Academic year 2007

.....

Abstract

This project studies and develops a program that searches for video clip files through the Adaptive Video Indexing (AVI) to calculate the picture vectors and Wavelet Transform to calculate the sound vectors.

The project collects the pictures, sounds, and the vectors of both to search for the clip files and compare the data with the databases. The clip files will be put into a Video Index kept in the String to be able to compare the pictures. The Wavelet Transform will be used to compare the sounds. The program developed by Visual Basic.net, Microsoft Visual Studio, will be used to compare the video clip files with picture and sound with the same 16 video clip files, managed and arranged by Microsoft Access.

As a result of the experiment, the program will be able to search for the video clip files from the databases and chose the appropriate video clip file.

กิตติกรรมประกาศ

ในการทำโครงการวิศวกรรมครั้งนี้ ผู้จัดทำขอกราบขอบพระคุณ ดร.ไพศาล มณีสว่าง ที่ได้
ให้คำปรึกษาโครงการนี้ ทั้งทฤษฎีและขั้นตอนการปฏิบัติงาน
และขอกราบขอบพระคุณ ดร. อัครพันธ์ วงศ์กั้งแห และ อาจารย์ศิริพร เดชะศิริรักษ์ ที่
ได้เสียสละเวลาเพื่อทำการตรวจสอบการทำงานและชี้แนวทางในการแก้ไขปัญหาโครงการนี้

นายชัชชา ชาวเหนือ
นายฤทธิจักร เกรือสาร



สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญรูป	ช
สารบัญตาราง	ซ

บทที่ 1 บทนำ

1.1 ความเป็นมาและความสำคัญของเรื่อง	1
1.2 วัตถุประสงค์ของโครงการ	3
1.3 ขอบข่ายของโครงการ	3
1.4 แผนการทำงาน	4
1.5 ผลที่คาดว่าจะได้รับ	4
1.6 งบประมาณที่ใช้	4

บทที่ 2 ทฤษฎีเบื้องต้น

2.1 บทนำ	5
2.1.1 รูปแบบของโครงการ	5
2.1.2 เครื่องมือค้นหา	8
2.1.3 เข้าถึงเนื้อหาวิดีโอ	8
2.2 ภาพการมองเห็น (Visual)	9
2.2.1-ทฤษฎีภาพ	10
2.2.1.1 แบบจำลองภาพการมองเห็น โดยใช้ดัชนีของวิดีโอ ปรับตัวได้ (Visual Modeling by Adaptive Video Indexing)	10
2.3 เสียง (Audio)	13
2.3.1 ทฤษฎีเสียง	13
2.3.1.1 อัลกอริทึมสำหรับการขยายคุณลักษณะเด่นของเสียง (Feature Extraction Algorithm of Audio)	13

สารบัญ (ต่อ)

	หน้า
2.4 หลักการจัดการของโปรแกรม	16
2.4.1 การจัดการกับฐานข้อมูล	16
2.4.2 การเปรียบเทียบตัวชี้ของภาพ (Compare Picture)	18
2.4.3 การเปรียบเทียบตัวชี้ของเสียง (Compare Audio)	18
2.4.4 การจัดเรียงค่าลำดับโดยใช้ Ranking	19
2.5 การรวมภาพและเสียง	20
2.5.1 ทฤษฎีการรวมภาพและเสียง	20
2.5.1.1 ลำดับค่าของการรวมภาพและเสียง(Audio and Visual Score Ranking)	20
บทที่ 3 วิธีการดำเนินโครงการ	
3.1 การจัดการกับฐานข้อมูล	22
3.2 โปรแกรมสืบค้นวีดีโอคลิป	23
บทที่ 4 ผลการทดลอง	
4.1 สิ่งที่ต้องเตรียมก่อนทำการทดลอง	35
4.2 ผลการเปรียบเทียบความเร็วในการสืบค้นวีดีโอคลิปด้วยโปรแกรม	35
4.3 การทดลองค้นหาวีดีโอคลิปโดยใช้ภาพ, เสียง และการรวมภาพและเสียง	36
4.4 เปรียบเทียบผลการทดลอง	40
4.5 การเปรียบเทียบอัลกอริทึมของระบบที่พัฒนาจากนายฉัตรชัย สุขสอน [3] และระบบที่พัฒนาขึ้น	42
4.5.1- ระบบที่พัฒนาจากนายฉัตรชัย-สุขสอน[3]-ในการสืบค้นภาพและ เสียง	42
4.5.2 ระบบที่ได้พัฒนาการสืบค้นภาพและเสียง	43
4.5.3 ผลการเปรียบเทียบในการดึงข้อมูลจากฐานข้อมูล	44
4.5.4 ความเร็วในการสืบค้นวีดีโอคลิป	44

สารบัญ (ต่อ)

หน้า

บทที่ 5 บทสรุป

5.1 สรุปผลการทดลอง	45
5.2 ปัญหาที่พบ	46
5.3 ข้อเสนอแนะ	46
เอกสารอ้างอิง	47
ประวัติผู้เขียนโครงการ	48



สารบัญรูป

รูปที่	หน้า
1.1 แสดงการทำงานของโปรแกรม (Software)	2
2.1 แสดงการ Search Video ของ Google	5
2.2 แสดงการ Search Video ของโครงการที่ทำ	6
2.3 แสดงวิธีการทำงานของการค้นหาของโปรแกรม	7
2.4 ความสัมพันธ์ระหว่างกลุ่มของงานวิจัยทั้ง 4	8
2.5 แสดงการอ้างอิงของวีดีโอ	10
2.6 แสดงวงจร XOR เกท	18
2.7 การเรียงค่าโดยใช้ Bubble Sort	19
3.1 ขั้นตอนการค้นหาวีดีโอคลิป	22
3.2 แสดงภาพรวมการค้นหาวีดีโอคลิป.....	23
3.3 แสดงการรวมภาพการค้นหาวีดีโอคลิปโดยโปรแกรมของ Class Form 1	24
3.4 แสดงภาพรวมการค้นหาวีดีโอคลิปโดยโปรแกรมของ Class Search	24
3.5 แสดงการทำงานใน Exclusive OR	26
3.6 แสดงการทำงาน Euclidean distance ของเสียง	27
3.7 แสดงการหาค่าความคล้ายคลึงของภาพและเสียง.....	28
3.8 แสดงการเรียงลำดับของภาพและเสียง	28
3.9 แสดงการทำงานการเรียงลำดับ (Rank) ภาพ	29
3.10 แสดงการทำงานการเรียงลำดับ (Rank) เสียง	30
3.11 แสดงการเรียงระยะของภาพ (Distance picture)	31
3.12 แสดงการเรียงระยะของเสียง (Distance audio)	32
3.13 แสดงการทำงานการรวมภาพและเสียง	33
4.1 แสดงผลการค้นหาวีดีโอคลิปด้วยภาพ	37
4.2 แสดงผลการค้นหาวีดีโอคลิปด้วยเสียง	38
4.3 แสดงผลการค้นหาวีดีโอคลิปด้วยภาพและเสียง	39
4.4 แสดงกราฟเปรียบเทียบการสืบค้นวีดีโอแบบภาพ, เสียง และรวมภาพกับเสียง	40
4.5 แสดงกราฟแห่งเปรียบเทียบการสืบค้นวีดีโอจากกราฟแห่ง	41

สารบัญตาราง

ตารางที่	หน้า
1.1 แสดง Gantt chart	4
2.1 การจัดเก็บฐานข้อมูลในตาต้าเบสเดิม	16
2.2 แสดงภาพการจัดเก็บของตัวชี้ของภาพให้เป็นชุดข้อมูล	17
4.1 บันทึกผลการทดลองเปรียบเทียบเวลาที่ใช้ในการค้นไฟล์วิดีโอ กับโปรเจกนวย ฉัตรชัย สุขสอน[3]	35
4.2 บันทึกผลการทดลองการสืบค้นด้วยภาพในวิดีโอคลิปเจ็ลี่ย โดยผู้ใ้ใช้	37
4.3 บันทึกผลการทดลองการสืบค้นด้วยเสียงในวิดีโอคลิปเจ็ลี่ย โดยผู้ใ้ใช้	38
4.4 บันทึกผลการทดลองการสืบค้นด้วยภาพและเสียงในวิดีโอคลิปเจ็ลี่ย โดยผู้ใ้ใช้	39



บทที่ 1

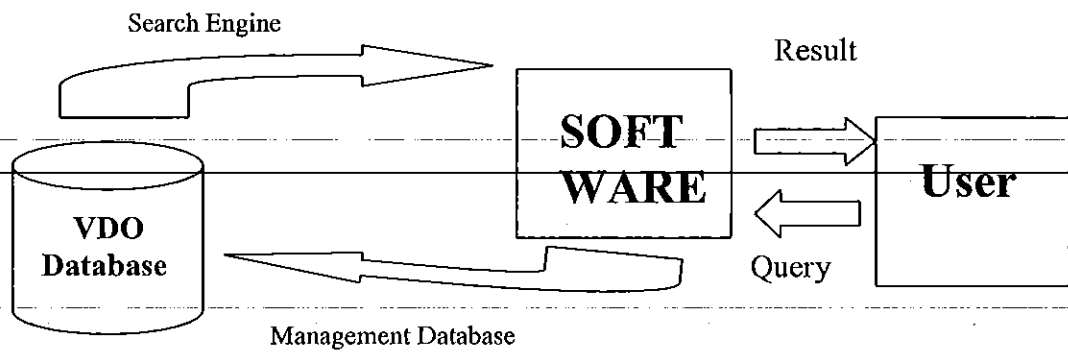
บทนำ

1.1 ความเป็นมาและความสำคัญของเรื่อง

ปัจจุบันการค้นหาไฟล์วิดีโอเป็นสิ่งที่น่าสนใจและยังเริ่มเข้ามามีบทบาทในการค้นหาไฟล์วิดีโอที่อยู่เป็นจำนวนมากซึ่งก็เคยเห็นกันมาบ้างแล้ว เช่น Google Video ในการที่จะค้นหาของ Google จะต้องทราบตัวบ่งชี้ (Keyframe) ของไฟล์ที่จะค้นหา ซึ่งจะทำให้การค้นหาอาจจะไม่ตรงตามที่ต้องการ อีกทั้งถ้าไม่ทราบชื่อไฟล์ก็จะไม่สามารถค้นหาไฟล์นั้นได้เลย เพราะฉะนั้นจึงคิดที่จะใช้การค้นหาไฟล์วิดีโอ ด้วยไฟล์วิดีโอในการค้นหาแทน โดยจะใช้ไฟล์วิดีโอที่ใกล้เคียงกันในการค้นหา ซึ่งจะช่วยให้ค้นหาไฟล์ได้ตรงตามความต้องการตรงประเภทของคลิป และค้นหาไฟล์วิดีโอได้ง่ายขึ้นกว่าการค้นหาไฟล์แบบใช้คำสำคัญ (Key Word)

จากที่ผู้จัดทำโครงการนี้ได้ศึกษามาแล้วว่า ได้มีนายฉัตรชัย สุขสอน[3] ทำโครงการลักษณะนี้คือ เป็นการใช้ระบบการค้นหา (Search Engine) จากเพิ่มวิดีโอ ภาพและเสียงรวมกัน ทางกลุ่มผู้จัดทำได้เห็นถึงปัญหาของโครงการนั้นว่าการหาไฟล์วิดีโอแบบเดิมนั้นมีการเก็บฐานข้อมูลที่ใหญ่มากจึงทำให้การดึงข้อมูลจากฐานข้อมูลช้า และใช้ อัลกอริทึมในการเปรียบเทียบ ที่ต้องสร้างเวกเตอร์ของการสอบถาม (Query) ให้มีขนาดเท่าตัวฐานข้อมูลก่อนทุกครั้งที่จะนำมาคำนวณเปรียบเทียบ เช่น เวกเตอร์ในฐานข้อมูลมี 6000x2000 จาก ตัวร้องขอที่มีเวกเตอร์ 1x2000 จะต้องทำให้เป็น 6000x2000 เพื่อที่จะนำข้อมูลนั้นมา ดอทกันในแต่ละตัว ถ้าฐานข้อมูลยังมีขนาดใหญ่มาก ก็จะทำให้ยังช้าในการค้นหาแต่ละครั้ง และสามารถทำการค้นหาได้แบบเดียว คือการค้นหาแบบภาพและเสียง ดังนั้นจึงได้มีการจัดทำโครงการนี้ขึ้นมา

โครงการที่ทางผู้จัดทำได้ทำขึ้นมาจะเข้าไปทำในส่วนของการการเก็บฐานข้อมูลใหม่ให้มีขนาดฐานข้อมูลให้เล็กลงกว่าเดิม และสร้างโปรแกรม (Software) ขึ้นมา เพื่อที่จะทำระบบการค้นหาไฟล์ที่เร็วกว่าแบบเดิมโดยใช้ อัลกอริทึมใหม่ และทำโปรแกรม (Software) ให้สามารถ เลือกแยกประเภทการค้นหา (Search) เพิ่มวิดีโอ (File Video) สามแบบ คือ ภาพอย่างเดียวเสียงอย่างเดียว และการรวมภาพและเสียง



รูปที่ 1.1 แสดงการทำงานของโปรแกรม (Software)

การที่ผู้จัดทำเลือกทำโครงการ (Project) นี้ก็เพื่อที่จะช่วยให้การค้นหาไฟล์วิดีโอง่ายขึ้นกว่าเดิม เพราะในอดีตการค้นหา (Search) ไฟล์ต่างๆ จำเป็นที่จะต้องทราบตัวบ่งชี้ (Keyframe) ของไฟล์นั้นถึงจะสามารถดูเพิ่มวิดีโอ (File Video) นั้นได้ ถ้าไม่ทราบก็ไม่สามารถค้นหาไฟล์นั้นได้ ดังนั้นทางผู้จัดทำจึงคิดที่จะหาทางให้การสืบค้นหาไฟล์วิดีโอให้เร็วขึ้นและถูกต้องมากขึ้น โครงการที่ทำนี้สามารถที่จะช่วยหาไฟล์ที่ต้องการได้ โดยใช้เพิ่มวิดีโอ (File Video) อีกเพิ่มหาได้โดยตรงโดยมีความต่างกับการหาแบบอื่นที่ใช้อยู่ เป็นอีกทางเลือกในการนำไปใช้งาน และอาจเป็นประโยชน์ในการค้นหาไฟล์ในอนาคตซึ่งจะช่วยลดปัญหาในการค้นหาเพิ่มวิดีโอ (File Video) โดยไม่ทราบชื่อได้ และได้ใช้หลักการการเปรียบเทียบความเหมือนของไฟล์ และดึงเวกเตอร์ที่เหมือนกันออกมา

ความแตกต่างจากโครงการ ของ นายจักรชัย สุขสอน[3] ที่ได้ทำมาแล้ว คือ

- ได้แก้ไข ส่วนของการจัดเก็บฐานข้อมูลของภาพให้มีขนาดเล็กลง จาก 2000 คอลัมน์ เป็นชุดข้อมูลใน คอลัมน์เดียว
- ได้ใช้อัลกอริทึมใหม่ ในการค้นหาภาพ, เสียง และการรวมภาพและเสียง และมีผลลัพธ์ในการค้นหาเร็วกว่า เมื่อเทียบกับการทำงานของอัลกอริทึมเดิม
- ได้เพิ่มความสะดวกในการค้นหาไฟล์แบบแยกประเภทได้ คือ ค้นหาแบบภาพอย่างเดียว ค้นหาแบบเสียงอย่างเดียว และการค้นหาแบบการรวมภาพและเสียง

1.2 วัตถุประสงค์ของโครงการ

- เพื่อที่จะทำให้การค้นหาไฟล์วีดีโอง่ายขึ้นกว่าเดิม โดยสามารถประเภทของการค้นหาได้ตามประเภทของคลิปวีดีโอ

- เพื่อเป็นตัวเลือกในการค้นหาไฟล์วีดีโอ อีกตัวเลือกหนึ่งให้ผู้ที่สนใจจะใช้

- เพื่อเพิ่มความเร็วในการค้นหาจากโครงการที่ นายจักรชัย สุขสอน [3] ได้ทำมาแล้วโดยใช้ อัลกอริทึมใหม่

- เพื่อประหยัดพื้นที่จัดเก็บข้อมูลของเวกเตอร์ของภาพให้เล็กกว่าโครงการที่รุ่นพี่ได้ทำมา

- เพื่อจะได้ศึกษาค้นคว้าเกี่ยวกับการค้นหาไฟล์วีดีโอ โดยใช้โปรแกรมต่างๆที่เกี่ยวข้องในการทำโครงการ และจะทำให้สามารถนำไปใช้ในชีวิตประจำวันและแก้ปัญหาต่างๆได้

1.3 ขอบข่ายของงาน

ในการพัฒนาโปรแกรม (Software) จะใช้โปรแกรม (Software) ต่าง ๆ ในการพัฒนาเพื่อให้โปรแกรม (Software) มีความถูกต้องและเป็นมาตรฐานในการใช้งานมากที่สุด โดยจะค้นหาจากไฟล์วีดีโอ ที่แปลงเป็นตัวชี้ (Index) ในฐานข้อมูล (Database) ที่มีอยู่แล้ว นำมาใช้ การค้นหาไฟล์วีดีโอ จะใช้หลักการของการค้นหาไฟล์วีดีโอด้วยไฟล์วีดีโออีกไฟล์ในการค้นหา กล่าวคือในการค้นหาไฟล์นั้นถ้ามีไฟล์วีดีโอก็จะสามารถค้นหาไฟล์วีดีโอที่ใกล้เคียงกันได้ ในฐานข้อมูลที่มีอยู่

- ทำการค้นหา (Search) แฟ้มวีดีโอ (File Video) โดยใช้ วีดิโอ ค้นหา (Search) ซึ่งแตกต่างกับการค้นหา (Search) ไฟล์วีดีโอในปัจจุบันคือใช้คำสำคัญ (Key Word) ในการค้นหาเป็นส่วนใหญ่ อย่างเช่น Google Video

- ได้ทำการพัฒนาเครื่องมือค้นหา (Search Engine) เพื่อทำการค้นหาแฟ้มวีดีโอ (File Video) ให้เร็วขึ้นกว่าเดิมที่มีอยู่

- จะทำการเปรียบเทียบข้อมูลวีดีโอ (File Video) ที่แปลงเป็นเวกเตอร์ไว้แล้ว

- จะเข้าไปจัดการข้อมูลในฐานข้อมูลให้มีขนาดเล็กลง ซึ่งจะทำให้การค้นหาไฟล์มีความเร็วขึ้น

1.4 แผนการทำงาน

ตารางที่ 1.1 แสดง Gantt chart

กิจกรรม	ระยะเวลาดำเนินงาน(เดือน)								
	1	2	3	4	5	6	7	8	9
1. ศึกษาข้อมูลของโปรแกรมต่างๆที่เกี่ยวข้อง ในการเขียนโครงงาน	←————→								
2. ศึกษาข้อมูลการแยกไฟล์จากไฟล์วีดีโอ		←————→							
3. เริ่มลองเขียนโปรแกรมเพื่อเชื่อมต่อกับฐานข้อมูลใน Database					←————→				
4. สร้าง software ตามแบบขึ้นมา									
5. ทดสอบsoftware ที่สร้างขึ้น					←————→				
6. หากพบบกพร่องที่ไม่สมบูรณ์							←————→		
7. ประเมินประสิทธิภาพของชิ้นงาน							←————→		
8. จัดทำรายงานและเตรียมนำเสนอ					←				→

1.5 ผลที่คาดว่าจะได้รับ

1. สามารถค้นหาไฟล์วีดีโอได้ง่ายกว่าเดิม
2. เป็นทางเลือกใหม่ในการค้นหาไฟล์วีดีโอ โดยไม่จำเป็นต้องทราบชื่อไฟล์ก็สามารถค้นหาไฟล์ได้
3. สามารถนำไปใช้ได้จริงในชีวิตประจำวัน และเป็นแนวทางในการพัฒนาต่อไป
4. สามารถลดขนาดไฟล์ในฐานข้อมูลให้มีขนาดเล็กลง
5. สามารถเปรียบเทียบไฟล์ได้อย่างมีประสิทธิภาพและถูกต้อง

1.6 งบประมาณทั้งโครงการ

หมวดค่าวัสดุและใช้สอย

งบที่จะได้ 2000 บาท

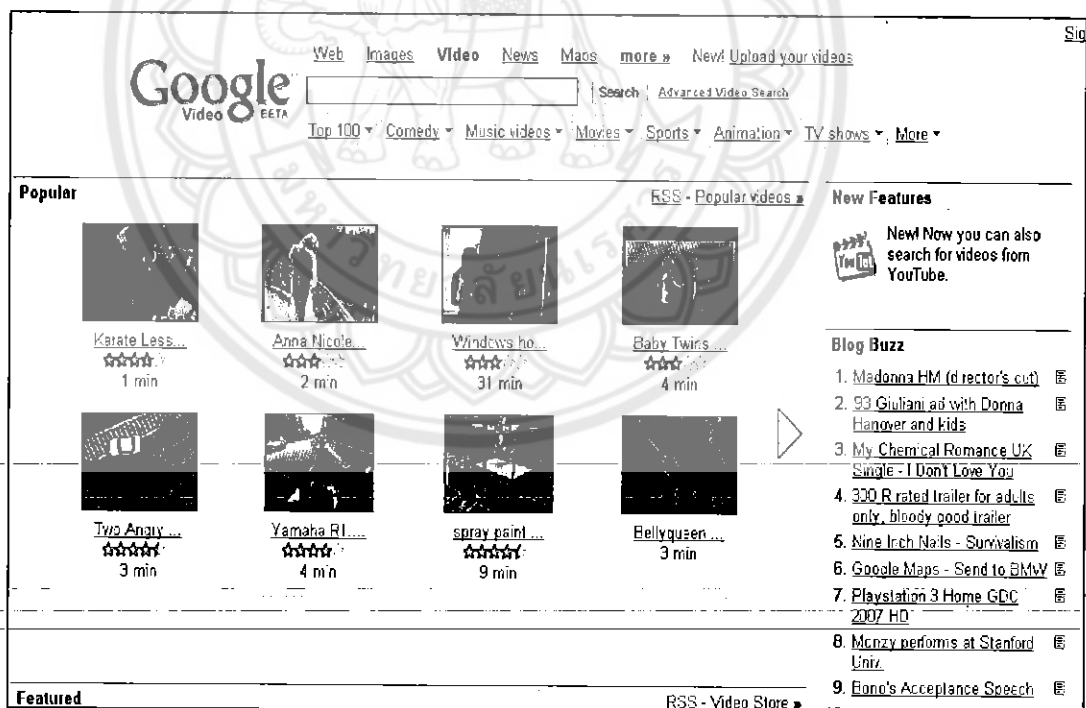
บทที่ 2 ทฤษฎีเบื้องต้น

2.1 บทนำ

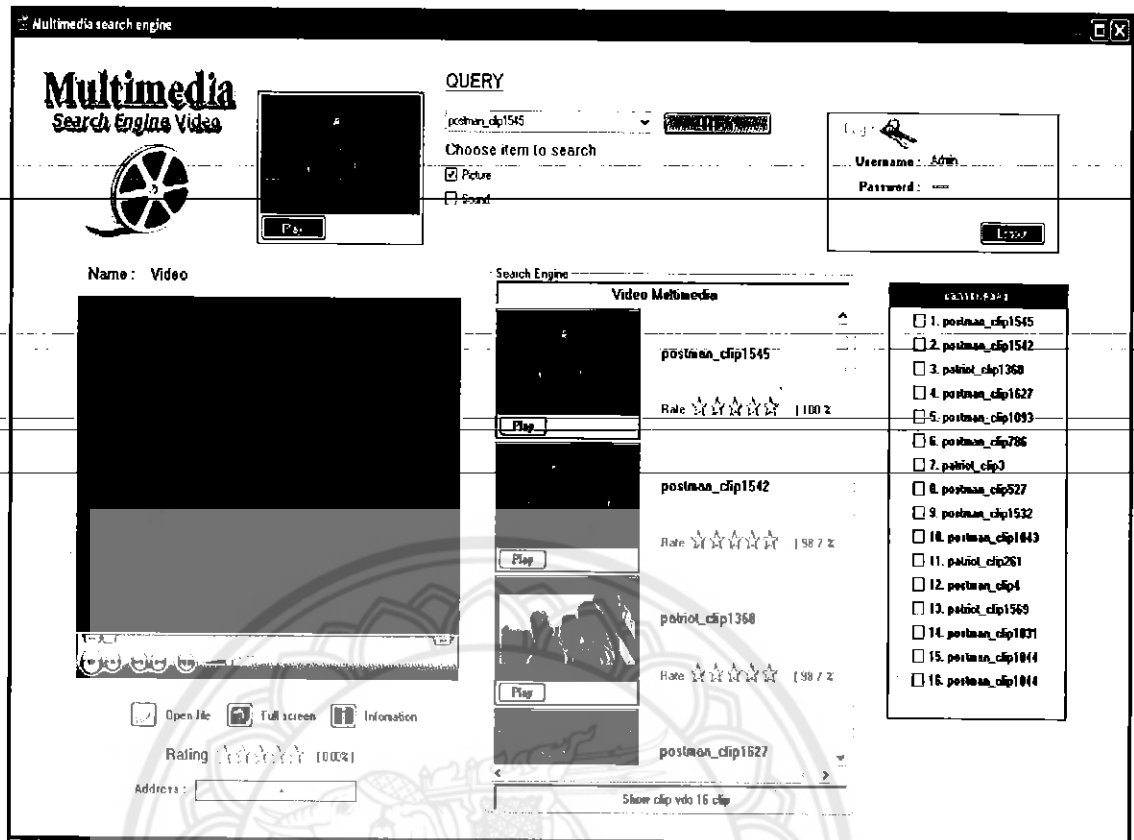
โครงการนี้มีวัตถุประสงค์ เพื่อพัฒนาความสามารถในการค้นหา(Search) เพิ่มวิดีโอ (File Video) ที่มีประสิทธิภาพมากขึ้นกว่าเดิมที่มีอยู่-เพราะว่าการค้นหาไฟล์แบบเดิมมีความล่าช้าและจัดการกับไฟล์ข้อมูลขนาดใหญ่ได้ไม่ดีมากนัก เพราะฉะนั้นจึงคิดวิธีการค้นหาและเปรียบเทียบที่ดี และเร็วกว่าเดิมอย่างมีประสิทธิภาพสูงสุด

2.1.1 รูปแบบของโครงการ

ความสามารถของ Google จะค้นหาวิดีโอที่ออกมามีข้อจำกัดคือคลิปจะไม่มี ความคล้ายกันเลยเพราะ Google แค่ทำการค้นหา ตาม คำสำคัญ (Key Word) เท่านั้นซึ่งได้ไปศึกษา รูปแบบของ Google Video รูปแบบส่วนใหญ่เป็น ดังนี้



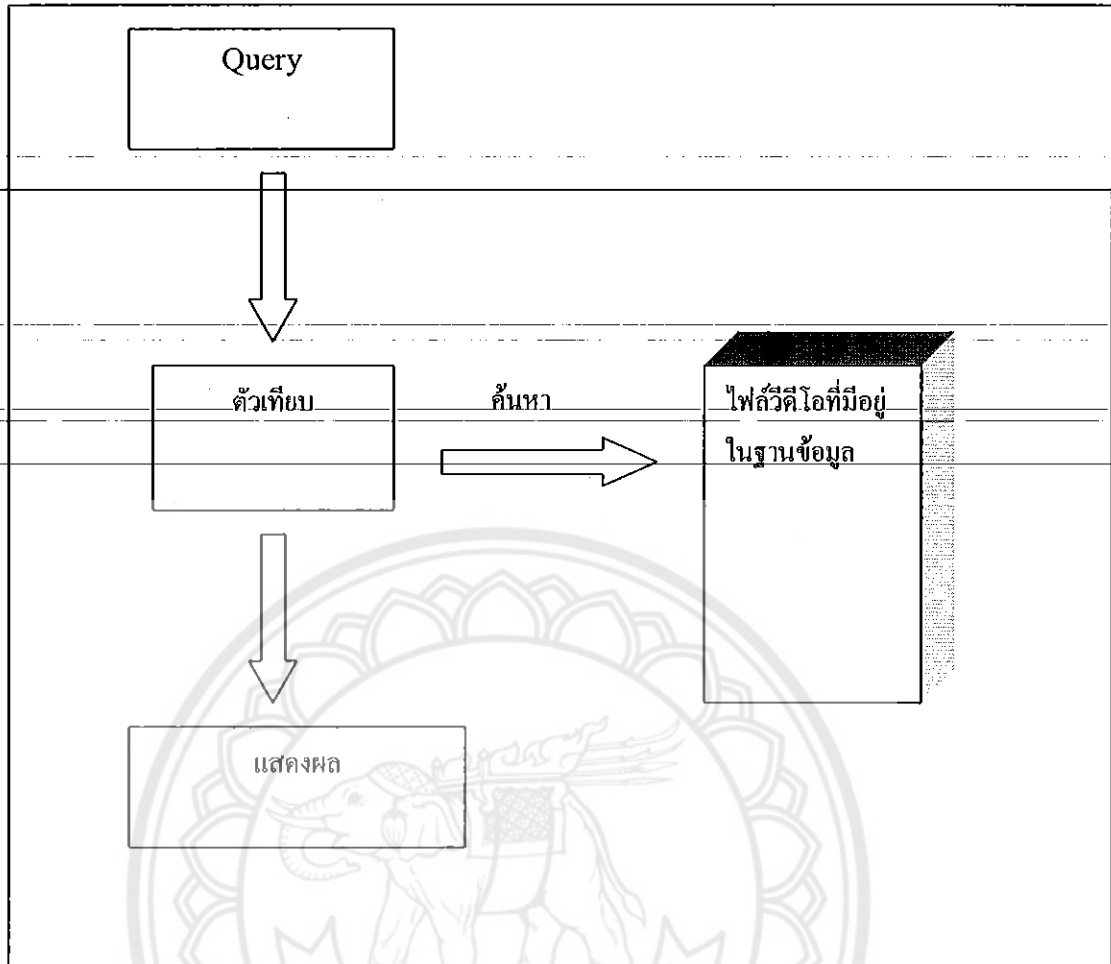
รูปที่ 2.1 แสดงการ Search Video ของ Google



รูปที่ 2.2 แสดงการ Search Video ของโครงการที่ทำ

โครงการที่ได้จัดทำขึ้นมาจะเป็นดังรูป 2.2 โดยจะมีความแตกต่างจาก Google Video ดังนี้

1. การค้นหาไฟล์ Google Video ได้ทำการค้นหาไฟล์หรือคลิปวิดีโอ โดยใช้คำสำคัญ (Key word) เป็นตัวค้นหาเป็นหลัก ส่วนได้ทำการค้นหาโดยใช้ ไฟล์วิดีโอในการค้นหาเป็นสำคัญ ซึ่งต่างกับ Google Video
2. รายละเอียดของ Google Video เช่น จะแสดงเวลา ชื่อของ คลิป โครงการนี้จะระบุรายละเอียดเพิ่มเติมซึ่งจะแตกต่างจาก Google Video เล็กน้อย ตรงรายละเอียดของ ไฟล์วิดีโอต่างๆ ใน ข้างๆคลิปที่ ค้นหา
3. รูปแบบส่วนใหญ่ยังคงคล้ายคลึงกับ Google Video แต่จะมีบล็อก ของไฟล์มากกว่าในการค้นหาและจะคัดเลือกไฟล์ที่มีความใกล้เคียงกับไฟล์อีกอันที่เอามาเปรียบเทียบขึ้นมา
4. โครงการนี้สามารถเพิ่มประสิทธิภาพของการค้นหาสามารถแยกคลิปได้ตามประเภท ได้การค้นหา จากเสียงการค้นหา จากภาพ หรือ การค้นหาภาพและเสียง
5. เพิ่มระดับความปลอดภัยขึ้น โดยมีการ ล็อกอิน ก่อน



รูปที่ 2.3 แสดงวิธีการทำงานของการค้นหาของโปรแกรม

จากรูปแสดง 2.3 แสดงการค้นหา คือ เลือกการสอบถาม (Query) จากนั้นโปรแกรมจะทำการค้นหา (Search) โดยผู้จัดทำจะเขียน รูปแบบการค้นหาในการเปรียบเทียบ โดยใช้ข้อมูลในฐานข้อมูลเดิมที่มีค่าเป็นเวกเตอร์ จากนั้นจะจัดการกับฐานข้อมูลใหม่ในการจัดการ ซึ่งจะลดการเปรียบเทียบกับข้อมูลทั้งตาราง ซึ่งจะช่วยให้การเปรียบเทียบเร็วขึ้นด้วย

การทำงานของระบบ โครงสร้างจะเป็นการทำงานการเชื่อมต่อระหว่างไฟล์วีดีโอกับฐานข้อมูล โดยจำทำการเปรียบเทียบไฟล์วีดีโอที่อัปโหลด ขึ้นมาจากนั้นจะเอาไฟล์ที่ได้ไปเปรียบเทียบกับไฟล์ที่มีอยู่แล้วในฐานข้อมูล โดยอาจจะใช้ Vb.net เป็นตัวเชื่อมต่อระหว่างไฟล์ที่ต้องการกับข้อมูลในฐานข้อมูล โดยโปรแกรมนี้เมื่อได้ไฟล์ที่มีความใกล้เคียงกันก็จะดึงไฟล์นั้นออกมาให้ผู้ใช้เลือกใช้งาน หรือเลือกที่จะนำไปวิเคราะห์ต่อไป

2.1.2 เครื่องมือค้นหา (Search engine)

เครื่องมือค้นหา (Search engine) คือ การค้นหาข้อมูลบนเครือข่ายอินเทอร์เน็ตที่มีอยู่เป็นจำนวนมาก ถ้าเปิดไปที่หน้าจอลงมาจะต้องเสียเวลาในการค้นหาและอาจหาข้อมูลที่ต้องการไม่พบ การที่จะค้นหาข้อมูลให้พบอย่างรวดเร็วจะต้องใช้เว็บไซต์สำหรับการค้นหาข้อมูล ซึ่งจะทำหน้าที่รวบรวมรายชื่อเว็บไซต์ต่าง ๆ เอาไว้โดยจัดแยกเป็นหมวดหมู่ ผู้ใช้งานเพียงแต่ทราบหัวข้อที่ต้องการค้นหาแล้วป้อน คำหรือข้อความของหัวข้อนั้น ๆ ลงไปในช่องที่กำหนด

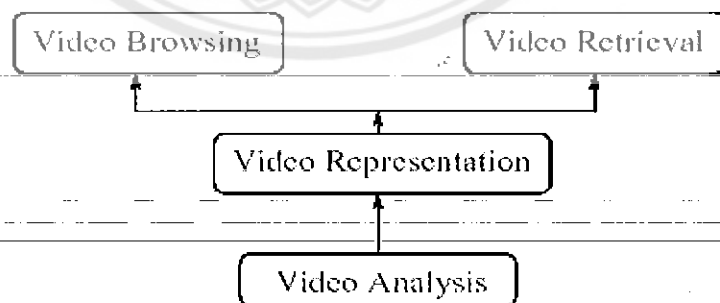
2.1.3 เข้าถึงเนื้อหาของวิดีโอ (Video content)

การวิจัยเกี่ยวกับการจะเข้าถึงเนื้อหาของวิดีโอ (Video content) อย่างมีประสิทธิภาพเพียงจะมี การค้นคว้ามากขึ้นเมื่อไม่กี่ปีมานี้ ความก้าวหน้าอื่นมากมายได้กระทำอยู่ในเรื่องของการวิเคราะห์ วิดีโอ (Video analysis), การนำเสนอตัวแทนวิดีโอ (Video representation), การค้นหาวิดีโอ (Video browsing), และการเรียกวิดีโอคืนกลับ (Video retrieval) ซึ่งเป็นเนื้อหาหลักสำหรับการเข้าถึงเนื้อหาของวิดีโอ (Video content)

การวิเคราะห์วิดีโอ (Video analysis) เป็นเรื่องเกี่ยวกับการจัดการในส่วนของกระบวนการ สัญญาณของระบบวิดีโอ รวมไปถึงการสืบหาขอบเขตของ shot (shot boundary) การดึงเฟรมหลัก (key frame) ออกมา และอื่น ๆ

การนำเสนอตัวแทนวิดีโอ (Video representation) เป็นเรื่องเกี่ยวกับ โครงสร้างของวิดีโอ เช่น ลำดับชั้นของเฟรมหลัก (key frame) ที่มีโครงสร้างแบบต้นไม้ (tree)

การค้นหาวิดีโอ (Video browsing) จะเกี่ยวข้องกับเรื่องทำอย่างไรถึงจะใช้โครงสร้างของการนำเสนอตัวแทน (Representation) ให้ช่วยผู้ชมค้นหาเนื้อหาของวิดีโอ (Video content) และ เกี่ยวข้องกับการนำตัววิดีโอที่น่าสนใจคืนกลับ



รูปที่ 2.4 ความสัมพันธ์ระหว่างกลุ่มของงานวิจัยทั้ง 4

จากรูปที่ 2.4 เห็นได้ว่าการค้นหาวิดีโอ (Video browsing) และการเรียกวิดีโอคืนกลับ (Video retrieval) จะอยู่ด้านบนของแผนภาพ (diagram) ซึ่งช่วยสนับสนุนโดยตรงในการช่วยให้

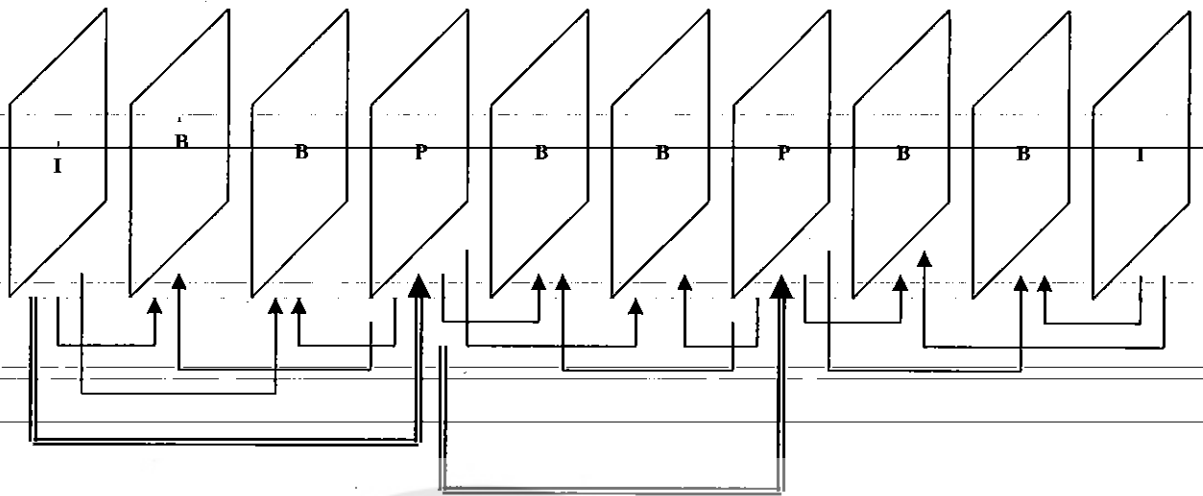
เข้าถึง เนื้อหาของวิดีโอ (Video content) สำหรับการเข้าถึงชั่วระยะเวลาหนึ่ง เช่น วิดีโอคลิป (Video clip) ในการค้นหา (browsing) และการเรียกกลับคืน (retrieval) จะให้ความสำคัญเท่า ๆ กัน

ในขณะที่การค้นหา (browsing) จะช่วยให้ผู้ใช้จับภาพกว้าง ๆ (global picture) ได้อย่างรวดเร็ว แต่การเรียกกลับคืน (retrieval) จะช่วยให้ผู้ใช้ค้นพบผลลัพธ์ของการเลือกแบบเจาะจง (specific query) ซึ่งสามารถอธิบายได้ดังนี้ นักอ่านจะเข้าถึงเนื้อหาของหนังสือ 1000 หน้า ได้อย่างมีประสิทธิภาพได้อย่างไร โดยปราศจากการอ่านหนังสือหมดทั้งเล่ม ประการแรกเขาต้องไปดูที่สารบัญ (Table-of-Contents : ToC) เพื่อหาบทหรือส่วนที่ต้องการ ถ้าอยากรู้บางที่ เช่น คำเฉลย สามารถไปหาที่ดัชนีท้ายเล่มได้ และพบเจอส่วนที่สอดคล้องกับสิ่งที่ต้องการ สรุปก็คือ สารบัญ (ToC) ของหนังสือจะช่วยผู้อ่านค้นหา (browsing) และดัชนี (Index) ของหนังสือจะช่วยผู้อ่านเรียกกลับคืน (retrieval) ซึ่งทั้งสารบัญและดัชนีจะมีความสำคัญเท่า ๆ กันที่จะช่วยให้ผู้อ่านเข้าถึงเนื้อหาหนังสือ แต่ในกรณีของข้อมูลวิดีโอ เป็นเรื่อง โชคร้ายที่ขาดแคลนทั้งสารบัญและดัชนี เทคนิคต่าง ๆ จึงถูกต้องการอย่างเร่งด่วนในการที่จะนำมาสร้างสารบัญและดัชนีของวิดีโอทั้งที่เป็นแบบอัด โนมัติหรือกึ่งอัด โนมัติ ทั้งนี้เพื่อความสะดวกสบายในการค้นหาและการเรียกกลับคืน ซึ่งจะเป็นพลังและความคล่องตัวอย่างสูง ถ้าสามารถออกแบบการเข้าถึงองค์ประกอบของวิดีโอ ด้วยการกระทำให้เป็นโครงการเดียวกัน ได้เป็นผลสำเร็จ

2.2 ภาพการมองเห็น (Visual)

ภาพวิดีโอแบบดิจิทัล (Digital Video) เกิดจากการนำภาพวิดีโอที่ถ่ายด้วยกล้องถ่ายวิดีโอ ม้วนเทปวิดีโอ หรือกล้องถ่ายภาพยนตร์ นำมาบันทึกให้อยู่ในรูปแบบไฟล์ในคอมพิวเตอร์ โดยใช้ฮาร์ดแวร์ (Hardware) พิเศษที่เรียกว่าบอร์คการจับภาพวิดีโอ (Group Based Representation) ในการจับภาพวิดีโอมาเป็นไฟล์ภาพในรูปแบบดิจิทัล

วิดีโอ (Video) มีอยู่หลายชนิดไม่ว่าจะเป็น เอวีไอ (AVI), เอ็มโอวี (MOV), เอ็มเพก (MPEG) หรือจะเป็นวิดีโอแบบดิจิทัล (Digital Video) ที่ใช้ดูบนอินเทอร์เน็ต (Internet) เช่น เรียดวิดีโอ (Real Video) เป็นต้น ซึ่งก็มีลักษณะคล้าย ๆ กัน ต่างกันที่คุณภาพของภาพ ความต่อเนื่องของภาพ (Playback-Rate) และขนาดของไฟล์ (Compression) ที่จะมีขนาดเล็กใหญ่แตกต่างกันไป



รูปที่ 2.5 แสดงการอ้างอิงของวิดีโอ

เฟรม (Frame) ภาพยนตร์ที่บีบอัดโดยเอ็มเพก-1 (MPEG1) จะมีอยู่ 3 ชนิดคือ เฟรมอินทรา (Intra-Picture : I), เฟรมพรีดิกทีฟ (Predictive : P) และเฟรมไบ-ไดเรกชันนัล (Bi-directional : B) โดยที่ไอเฟรม (I frame) จะเป็นเฟรม (frame) สำหรับให้เฟรม (frame) อื่น ๆ อ้างอิงและเก็บข้อมูลภาพของ

ตัวเองไว้ครบถ้วน เฟรม (P frame) จะอ้างอิงข้อมูลของเวกเตอร์เคลื่อนที่ (motion vector) และที่เป็นเศษเหลือ (residual) มาจากเฟรม (frame) ก่อนหน้าเท่านั้น ส่วนบีเฟรม (B frame) จะอ้างอิงมาจากทั้งพี (P) และไอ (I) ที่อยู่ก่อนหน้าและข้างหลังด้วยทำให้ลดข้อมูลที่ต้องส่งเพิ่มในบางเฟรม (frame) ลงไปได้

2.2.1 ทฤษฎีภาพ

2.2.1.1 แบบจำลองภาพการมองเห็นโดยใช้ดัชนีของวิดีโอปรับตัวได้

(Visual Modeling by Adaptive Video Indexing : AVI)

ข้อมูลของวิดีโอจะประกอบไปด้วยภาพที่ต่อเนื่องเป็นลำดับ ซึ่งวิดีโอที่มีเนื้อหาคล้ายคลึงกัน โดยปกติจะบรรจุภาพที่คล้ายคลึงกัน สมการต่อไปนี้จะถูกใช้บ่อยเพื่อใช้ในการอธิบายว่าวิดีโอ อ้างถึงเขตความคล้ายคลึงกันของภาพได้อย่างไร

$$D_{I_x} = \{(X_i, f_i) | X_i \in IR^p, i = 1, 2, \dots, N\} \quad (2.1)$$

D_{I_x} = ตัวอธิบายองค์ประกอบเบื้องต้น (Primary content descriptor) สำหรับวิดีโอช่วงที่

I_x

X_i = เวกเตอร์ที่บ่งบอกคุณลักษณะของวิดีโอ

f_i = เฟรม (frame) ของวิดีโอที่ i อยู่นภายใน I_x

N = จำนวนทั้งหมดของเฟรม (frame)

I_x = ช่วงของวิดีโอ ที่อาจจะเป็นช็อต (Shot), ฉาก (Scene) และคลิปเนื้อเรื่อง (Story Clips)

แบบจำลองเอวีไอ (AVI Model) ถูกนิยามโดยความน่าจะเป็นของการค้นหาแบบจำลองของเฟรม (Frame model) หรือแม่แบบ (Template) M , ในวิดีโอที่ถูกนำเข้ามา ซึ่งถูกกำหนดให้เป็นสมการอย่างง่ายดังนี้

$$P(M_i) = (N \times N_i)^{-1} \sum_{j=1}^{N \times N_i} I(j = I_{M_i}) \quad (2.2)$$

I_{M_i} = ป้ายชื่อ (Label) ของแบบจำลองเวกเตอร์ (Model vector) M_i

$N \times N_i$ = จำนวนรวมของป้ายชื่อ (Label) ที่ใช้สับคั่นวิดีโอที่ถูกนำเข้ามา

ซึ่งฟังก์ชัน $I(\bullet) = 1$ ถ้าอาร์กิวเมนต์ (Argument) เป็นจริง และจะเท่ากับ 0 ถ้าเป็นกรณีอื่น

ให้ M เป็นเซตของแบบจำลอง (Model) โดย $M = [M_1, \dots, M_i, \dots, M_T]$, $M_i \in IR^p$ ซึ่ง T เป็นจำนวนรวมของแบบจำลอง (Model) ซึ่งแบบจำลอง (Model) เหล่านี้จะถูกสร้างขึ้นและปรับปรุงให้ดีขึ้นโดยเวกเตอร์การฝึก (Training vector) X_j , $j = 1, \dots, J$ ซึ่งได้มาจากตัวอธิบายเบื้องต้น (Primary descriptors) ของวิดีโอที่นำไปในฐานะข้อมูล ในที่นี้สมมุติว่า จำนวนของแบบจำลองเวกเตอร์ (Model Vectors) มีนัยสำคัญน้อยกว่าจำนวนของเวกเตอร์การฝึก (Training vector) $T \ll J$

งานนี้ กระบวนการออปติไมเซชัน (Optimization) จะถูกทำให้บรรลุผลสำเร็จโดยแอปพลิเคชัน (Application) ของอัลกอริทึมการเรียนรู้แบบแข่งขัน (Competitive Learning Algorithm) ไปยังเซตการฝึก (Training Set) X_j , $j = 1, \dots, J$ และพื้นที่ ของ X จะถูกแสดงคุณลักษณะ โดยลักษณะเด่นของสีฮิสโตแกรม (Color Histogram Feature) โดยใช้พื้นที่ของสีเอชเอสวี (HSV Color Space)

เพื่อให้บรรลุผลของการสับคั่นวิดีโอ-ตัวอธิบายที่-2 (Secondary-descriptor) จะถูกสร้างขึ้นเพื่อบ่งบอกคุณลักษณะพิเศษ (Characterize) ของข้อมูลข่าวสารที่เป็นที่ว่างชั่วคราว (Spatio-temporal Information) ทำการสร้างเซต (Set) ของป้ายชื่อ (Labels) ผ่านทางฟังก์ชันการจับป้ายชื่อแบบทวีคูณ (Multiple-label-mapping Function) $F(X): IR^{48} \rightarrow IR^{N_i}$, ซึ่งในแต่ละ $X_i, i \in \{1, \dots, N\}$ จะถูกจับคู่ไปบนวอโรนอย สเปซ (Voronoi Space)

$$X_i \Rightarrow \langle t^*, R_i^{N_i} \rangle \Rightarrow \rho^{(X_i)}, t^* = \arg \min_{t \in \{1, \dots, T\}} (\|X_i - M_t\|), R_i^{N_i} = \bigcup_{j=1}^{N_i} M_{t^*, j} \quad (2.3)$$

ซึ่ง $R_i^{N_i}$ เป็นบริเวณที่ประกอบด้วยวินิ่ง โหนด (Wining node) M_{t^*} และบริเวณใกล้เคียงกับ $N_i - 1$ วอโรนอย เซล (Voronoi cells) ไปยัง โหนด (node) M_{t^*}

เซตของป้ายชื่อ (Labels) , $P^{(X_i)} = I_{i,1}^{X_i}, I_{i,2}^{X_i}, \dots, I_{i,N_1}^{X_i}$, จะประกอบด้วยป้ายชื่อที่ถูกจัดเข้าพวกกัน (Associated labels) ของวอโรนอย เซล (Voronoi cells) ใน $R_i^{N_1}$ อาจกล่าวได้ว่าป้ายชื่อ (Labels) ทั้งหมด $I_{i,j}^{X_i}, j=1, \dots, N_1$ เป็นตัวแทนแบบจำลอง (Model) ที่จับที่จับคู่กันอย่างดีที่สุดในพื้นที่ (Top) N_1 ซึ่งมีความคล้ายคลึงกันเป็นส่วนมากไปยังเวกเตอร์นำเข้า (Input Vector) X การจับคู่ (Mapping) กันของเฟรมวิดีโอ (Video frame) ทั้งหมด ให้ผลลัพธ์ อยู่ในรูปเซตของป้ายชื่อ (Labels), $p^{X_i}, i=1, 2, \dots, N$ ซึ่งจะถูกเรียงต่อกันเป็นอนุกรมไปในเวกเตอร์ลักษณะเด่นแบบเดี่ยว (Single feature vector) $W_{I_i} = [W_1, \dots, W_i, \dots, W_T]$ สอดคล้องกับสมการที่ 2.2 ซึ่งแต่ละ W_i ได้มาโดย

$$W_i = P(M_i) = (N \times N_1)^{-1} \sum_{i=1}^N \sum_{j=1}^{N_1} I(I_j^{X_i} = l_{M_i}) \quad (2.4)$$

ซึ่งถูกกำหนดไว้ว่า มีเพียงจำนวนน้อยของ W_i ที่มีค่าไม่เป็น 0 เนื่องจาก $T \gg N$, ดังนั้นผลลัพธ์ของเวกเตอร์ W_{I_i} จะมีจำนวนน้อยมาก (Very Sparse) ซึ่งจะคำนึงถึงช่องว่างของการเก็บรักษาที่มีประสิทธิภาพ (Efficient Storage Space) และความเร็วในการจับคู่ของเวกเตอร์ (Fast Vector Matching)

เวกเตอร์ที่มีจำนวนน้อยนี้สามารถกำหนดแบบฟอร์มได้ดังนี้

$$\hat{W}_{I_i} = [(t, w_t)], t \in X, X \subset \{1, 2, \dots, T\}, \quad (2.5)$$

ซึ่ง X เป็นเซตของดัชนี t ที่เชื่อมโยงกับส่วนประกอบที่ไม่เป็น 0 (Nonzero component) w_t โดยวิธีนี้ จะทำให้จับคู่ความคล้ายคลึงกันระหว่าง 2 ช่วงวิดีโอใด ๆ I_x และ I_y ที่สามารถคำนวณได้โดย

$$S_{xy}^{(v)} = 1 - \frac{\hat{W}_{I_x} \cdot \hat{W}_{I_y}}{\left(\|\hat{W}_{I_x}\| \times \|\hat{W}_{I_y}\| \right)} \quad (2.6)$$

ซึ่ง $S_{xy}^{(v)}$ เป็นค่าความคล้ายคลึงกันระหว่าง I_x และ I_y และเทอม (Term) ที่ 2 ในสมการ (2.6) จะถูกกำหนดโดยการวัดโคไซน์ที่นิยมใช้กัน (Popular Cosine Measure)

2.3 เสียง (Audio)

เสียงเป็นส่วนประกอบสำคัญที่จะช่วยเกื้อหนุนให้โปรแกรมมัลติมีเดีย (Multimedia) นั้นมีความน่าสนใจและดึงดูดมากขึ้น ชนิดของไฟล์เสียงมีอยู่ด้วยกัน 3 รูปแบบได้แก่ เวฟฟอร์มออডিโอ (Waveform Audio), มิดิ (MIDI) และซีดี ออดิโอ (CD Audio) ซึ่งเวฟฟอร์มออডিโอ (Waveform Audio) เป็นชนิดที่นิยมใช้กันมากที่สุดในโปรแกรมมัลติมีเดีย (Multimedia) เกิดจากการอัดเสียงหรือเพลง โดยผ่านซาวด์การ์ด (Sound Card) ให้เข้ามาอยู่ในรูปแบบของไฟล์ในคอมพิวเตอร์ ซึ่งสามารถบันทึกไฟล์ได้หลายแบบฟอร์แมต (Format) ด้วยกัน ขึ้นอยู่กับคุณสมบัติของซาวด์การ์ด (Sound Card) และซอฟต์แวร์ (Software) ที่ใช้ นอกจากนี้คุณภาพของเสียง (Sample Rate) ก็เป็นสิ่งที่ควรคำนึงในการบันทึก เนื่องจากคุณภาพของเสียงนั้นจะมีผลกับขนาดของไฟล์ ยิ่งใช้คุณภาพเสียงดีเท่าไร ไฟล์ ก็จะมีขนาดใหญ่ขึ้นเท่านั้น เช่น 11.025 Khz เหมาะกับเสียงพูด บรรยาย หรือเสียงเอฟเฟกต์ (Effect) ประกอบเสียงโทรศัพท์ (Telephone Audio), 22.05 Khz เหมาะกับเสียงเพลงแบบเป็นเสียงพื้นหลังในวิทยุ (Background Radio Audio), 44.1 Khz เหมาะกับงานคุณภาพสูง ซีดี ออดิโอ (CD Audio) ส่วนมิดิ (MIDI) คือ รูปแบบไฟล์ที่เก็บบันทึกตัวโน้ต (Notes) เอาไว้ เมื่อต้องการเล่นเสียงเพลงคอมพิวเตอร์ จะทำการอ่านรูปแบบของเครื่องดนตรี พร้อมทั้งตัวโน้ต (Notes) และเล่นออกมาเป็นเสียงดนตรีตามที่กำหนด รูปแบบของไฟล์แบบมิดิ (MIDI) นี้จะมีขนาดไฟล์ที่เล็กกว่าชนิดอื่น ๆ แต่คุณภาพของเสียงนั้นขึ้นอยู่กับการ์ดเสียง (Hardware : sound card) ว่ามีเวฟเทเบิล (Wavetable) หรือไม่ ถ้ามีคอมพิวเตอร์จะเล่นเสียงผ่านเวฟเทเบิล (Wavetable) ของ การ์ดเสียง (Sound card) ซึ่งจะเก็บเสียงดนตรีนั้น ๆ ไว้ ทำให้คุณภาพเสียงเหมือนเล่นจากเครื่องดนตรีจริงๆ แต่หาก ไม่มีคอมพิวเตอร์จะทำการสังเคราะห์เสียงเหล่านั้นขึ้นมา ซึ่งคุณภาพจะไม่ได้เท่าที่ควร ส่วนซีดี ออดิโอ (CD Audio) เป็นเสียงที่ผ่านการบันทึกในรูปแบบเวฟฟอร์ม (Waveform) ให้อยู่ในรูปแบบของแผ่นซีดีปัจจุบัน

2.3.1 ทฤษฎีเสียง

2.3.1.1 อัลกอริทึมสำหรับการขยายคุณลักษณะเด่นของเสียง

(Feature Extraction Algorithm of Audio)

วิธีนี้อาจถูกแบ่งเป็นคลิปเพื่อใช้กับอัลกอริทึม (Algorithm) ที่มีอยู่ ส่วน Audio ก็อาจถูกแบ่งแยกจากวิธี โอคลิป และสัญญาณ (Signal) ก็จะถูกทำให้ดีขึ้น โดยอยู่ในรูปของระดับการสุ่ม (Sampling Rate)

ในแต่ละส่วนของเสียงจะถูกแยกออกโดยการใส่การแปลงเวฟเลทแบบไม่ต่อเนื่อง 1 มิติ (One-dimensional Discrete Wavelet Transformation : DWT) ซึ่งการแปลงเวฟเลทแบบนี้ จะแยก สัญญาณเสียง (Audio signal) ออกเป็น 2 ชั้น-แบนด์ (Sub-band) ในแต่ละเวฟเลท สเกล (Wavelet scale) คือ ชั้น-แบนด์ความถี่ต่ำ (Low frequency sub-band) กับชั้น-แบนด์ความถี่สูง (High frequency sub-band) ซึ่งการแปลงเวฟเลท (Wavelet transformation) จะเสนอผลลัพธ์ สเกล

(Multi-scale) ที่เป็นตัวแทนข้อมูลของเสียง คำนึงสามารถสร้างโครงสร้างของการสืบค้นบนพื้นฐานคุณสมบัติของสเกล (Scale) เหล่านี้ได้ ยิ่งไปกว่านั้นสัญญาณเสียงยังเป็นสัญญาณที่ไม่คงที่ ซึ่งเนื้อหาของความถี่จะเกี่ยวข้องกับเวลา ซึ่งการแปลงเวฟเลต (Wavelet transformation) จะนำเสนอข้อมูลทั้งความถี่และเวลาไปพร้อม ๆ กัน

ด้วยคุณสมบัติเหล่านี้ของการแปลงเวฟเลต (Wavelet transformation) ที่ใช้เพื่อแยกสัญญาณเสียง ซึ่งจะเป็นพื้นฐานของการนำเสนอเสียง (Audio) ในการจำลองรูปแบบการกระจายของสัมประสิทธิ์เวฟเลต (Wavelet coefficients) ใน ชั้น-แบนด์ความถี่สูง (High frequency sub-band) โดยการรวม 2 ลาปลาซ (Laplacians) ที่เข้าสู่ 0 ซึ่งพารามิเตอร์ (Parameters) ของการรวมนี้ จะถูกบ่งบอกถึงคุณลักษณะสำหรับสืบค้นองค์ประกอบเสียง (Audio content) โดยมีรูปแบบจำลองดังนี้

$$p(w_i) = \alpha_1 p_1(w_i|b_1) + \alpha_2 p_2(w_i|b_2), \alpha_1 + \alpha_2 = 1 \quad (2.7)$$

α_1 และ α_2 เป็นความน่าจะเป็นที่ผสมกันของ 2 ส่วนประกอบ (Components) p_1 และ p_2
 w_i เป็นสัมประสิทธิ์เวฟเลต (Wavelet coefficients)

b_1 และ b_2 เป็นพารามิเตอร์ (Parameters) ของการกระจายลาปลาซ (Laplacian distribution) p_1 และ p_2 .

ส่วนประกอบของลาปลาซ (Laplacian component) ในสมการ (2.7) ถูกกำหนดโดย

$$p_1(w_i|b_1) = \frac{1}{2b_1} \exp(-|w_i|/b_1) \quad (2.8)$$

รูปร่าง (Shape) ของการกระจายลาปลาซ (Laplacian distribution) ถูกกำหนดค่า โดยสัญญาณพารามิเตอร์ (Signal parameter) b

อี-สเต็ป (E-Step) : สำหรับวงรอบการทำซ้ำที่ n รอบ (n -th Iterative cycle), อี-สเต็ป (E-Step) จะคำนวณ 2 ความน่าจะเป็น สำหรับแต่ละสัมประสิทธิ์เวฟเลต (Wavelet coefficients)

$$p_{ii}(n) = \frac{\alpha_1(n)p(w_i|b_1(n))}{\alpha_1(n)p(w_i|b_1(n)) + \alpha_2(n)p(w_i|b_2(n))} \quad (2.9)$$

$$p_{2i}(n) = \frac{\alpha_2(n)p(w_i|b_2(n))}{\alpha_1(n)p(w_i|b_1(n)) + \alpha_2(n)p(w_i|b_2(n))} \quad (2.10)$$

เอ็ม-สเต็ป (M-Step) : ในเอ็ม-สเต็ป (M-Step), พารามิเตอร์ (Parameter) $[b_1, b_2]$ และความน่าจะเป็นของอันก่อน (Prior Probabilities) $[\alpha_1, \alpha_2]$ ถูกปรับปรุง (Update)

$$\alpha_1(n+1) = \frac{1}{K} \sum_{i=1}^K p_{1i}(n), \quad \alpha_2(n+1) = \frac{1}{K} \sum_{i=1}^K p_{2i}(n), \quad (2.11)$$

$$b_1(n+1) = \frac{\sum_{i=1}^K |w_i| p_{1i}(n)}{K \alpha_1(n+1)}, \quad b_2(n+1) = \frac{\sum_{i=1}^K |w_i| p_{2i}(n)}{K \alpha_2(n+1)} \quad (2.12)$$

ซึ่ง K เป็นจำนวนของสัมประสิทธิ์เวฟเลต (Wavelet coefficients) ในซับ-แบนด์ (Sub-band) ปัจจุบัน ซึ่งรูปแบบของส่วนประกอบ (Component) ต่อไปนี้ ลักษณะเด่นเวกเตอร์ (Feature vector) จะใช้สำหรับการสืบค้นคลิปของเสียง (Audio clips)

(1) ค่ากลาง และส่วนเบี่ยงเบนมาตรฐานของสัมประสิทธิ์เวฟเลต (Wavelet coefficients) ในซับ-แบนด์ความถี่ต่ำ (Low frequency sub-band)

(2) แบบจำลองพารามิเตอร์ (Model parameters) $[\alpha_1, b_1, b_2]$ คำนวณสำหรับแต่ละซับ-แบนด์ความถี่สูง (High frequency sub-band)

ฟังก์ชันความคล้ายคลึง (Similarity Function) ที่ใช้สำหรับจับคู่ของเวกเตอร์ลักษณะเด่นของเสียง (Audio feature vectors) ถูกอธิบายโดยระยะทางแบบยูคลิดีเนียน (Euclidean distance) ดังนี้

$$S_{xy}^{(n)} = \sqrt{\sum_{i=1}^P (F_x[i] - F_y[i])^2}, \quad (2.13)$$

ซึ่ง F_x และ F_y เป็นเวกเตอร์ลักษณะเด่นของเสียง (Audio feature vectors) ของวีดีโอคลิป I_x และ I_y

2.4 หลักการการจัดการของโปรแกรม

ไฟล์วีดิโอที่ได้จะทำการแบ่งออกเป็น 2 ส่วน เพื่อความง่ายในการเปรียบเทียบกันข้อมูลที่มีอยู่ในฐานข้อมูล-คือส่วนที่เป็นตัวชี้ของภาพและ-ตัวชี้ของเสียง-เพื่อที่จะนำไปวิเคราะห์กับไฟล์

วีดิโอที่ต้องการนำไปเปรียบเทียบกัน

โปรแกรมที่ได้พัฒนาขึ้นมาจะมีวิธีการจัดเก็บไฟล์ เพื่อจัดการกับไฟล์ที่มีขนาดใหญ่ ก็จะจำแนกไฟล์วีดิโอออกเป็น 2- ส่วน คือส่วนตัวชี้ของภาพและตัวชี้ของเสียง เนื่องจากในการเก็บข้อมูลตัวชี้ของภาพลงในฐานข้อมูลมีขนาดใหญ่มากถึงตัวชี้ 2000 ตัว เพราะฉะนั้นจึงจำเป็นต้องทำ

การแปลงให้อยู่ในรูปแบบ เลขฐาน 2 (Binary) แล้วนำไปเก็บในฐานข้อมูลในรูปแบบของชุดข้อมูลให้เป็นเพียงตัวชี้ตัวเดียว

จะใช้ Vb.net เพื่อเชื่อมต่อกับฐานข้อมูล โดยจะดึงข้อมูลในฐานข้อมูล ขึ้นมาเพื่อเปรียบเทียบไฟล์วีดิโอที่เหมือนกันในฐานข้อมูลที่ต้องการแล้วแสดงไฟล์วีดิโอออกมาให้ผู้ใช้งาน

2.4.1 การจัดการกับฐานข้อมูล

การจัดเก็บตัวชี้ของภาพจะเก็บเหมือนเป็นชุดข้อมูลของ เมื่ออยากจะเก็บอะไรลงไปก็เก็บได้ โดยที่ Database จะถือว่านี่คือก้อนของข้อมูลอันหนึ่ง แต่เก็บเป็นเหมือนข้อความ แล้วเวลาจะใช้ข้อมูลก็จะเข้าไปจัดการกับข้างในข้อความแต่ละตัว เหมือนการเก็บข้อมูลใน คอลัมน์ ให้น้อยลง ก็จะทำให้การค้นหาในดาต้าเบสเร็วขึ้น

ตารางที่ 2.1 การจัดเก็บฐานข้อมูลใน Database เดิม

ชื่อเรื่อง	component 1	component 2	component 3	component 4	...	component n
Name1.avi	A1	A2	A3	A4	...	An
Name2.avi	B1	B2	B3	B4	...	Bn
Name3.avi	C1	C2	C3	C4	...	Cn
Name4.avi	D1	D2	D3	D4	...	Dn
.
.
NameM.avi	Z1	Z2	Z3	Z4	...	Zn

ตารางที่ 2.2 แสดงการเก็บของตัวชี้ของภาพให้เป็นชุดข้อมูล

ชื่อเรื่อง	Data frame
Name1.avi	A1A2A3. An
Name2.avi	B1B2B3. Bn
Name3.avi	C1C2C3 Cn
NameM.avi	Z1Z2Z3 Zn

จากตารางที่ 2.1 แสดงถึงการจัดการ กับฐานข้อมูลแบบเดิมโดยที่ค่าของ A B C ถึง Z เป็นค่าของเลขฐานสอง (Binary) โดยค่าที่เก็บใน Component ถ้ามีค่ามากกว่า 0 จะเก็บค่าในเลขฐานสอง (Binary) เป็น 1 แต่ถ้าค่าที่เก็บใน Component มีค่าน้อยกว่าหรือเท่ากับ 0 จะเก็บค่าในเลขฐานสอง (Binary) เป็น 0 และเก็บค่าไว้แต่ละคอลัมน์ และเมื่อดูจากตารางที่ 2.2 จะเห็นว่าเมื่อจัดการกับฐานข้อมูลใหม่จะได้ข้อมูลที่กระชับมากขึ้น และจะทำให้การค้นหาที่ซับซ้อนและเร็วขึ้นด้วย ซึ่งจะทำให้ค้นหาไฟล์ได้ดีกว่าเดิม โดยข้อมูลที่เก็บไว้ในฐานข้อมูลนี้ จะเป็นตัวชี้ของไฟล์วิดีโอนามสกุล .avi ซึ่งมีอยู่แล้วในฐานข้อมูลเดิม

A	B	X
0	0	0
0	1	1
1	0	1
1	1	0

$$X = A \oplus B$$

รูปที่ 2.6 แสดงวงจร XOR เกท

จากรูป 2.6 จะใช้เกท XOR ในการเปรียบเทียบ ตัวชี้ของภาพของ Query กับ ตัวชี้ของภาพ
 ในฐานข้อมูล

2.4.2 การเปรียบเทียบตัวชี้ของภาพ (Compare Picture)

ลองพิจารณาวงจรของ XOR ซึ่ง $X = A \oplus B$ จากคณิตศาสตร์บูลีน จะเห็นว่า
 ซึ่งสามารถเขียนเป็นลอจิกได้ดังรูป 2.6 จะสังเกตเห็นว่า ค่าของ A กับ B ที่เหมือนกันจะทำงาน
 เพื่อให้ลอจิกเอาต์พุต "0" ในขณะที่ค่า A กับ B ที่ต่างกันจะทำงานเพื่อให้ลอจิกเอาต์พุต "1" จำนวน

2.4.3 การเปรียบเทียบตัวชี้ของเสียง (Compare Audio)

เนื่องจากเก็บตัวชี้ของภาพในรูปแบบ เลข จะเปรียบเทียบ ความใกล้เคียงกันของเสียงจึงใช้
 หลักของระยะทางแบบยูคลิเดียน (Euclidean distance) มาใช้ในการเปรียบเทียบ

การหาระยะทาง (Euclidean distance) ระหว่างจุด $P = (p_1, p_2, \dots, p_n)$ และ $Q = (q_1, q_2, \dots, q_n)$
 จะได้ว่า

$$d(P,Q) = \|P-Q\| = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

$d(P,Q)$ = ระยะทาง(Euclidean Distance)

p_i = เป็นค่าของ Component เสียงที่เลือก

q_i = เป็นค่าของ Component เสียงที่นำมาเปรียบเทียบ

2.4.4 การจัดเรียงค่าลำดับโดยใช้ Ranking

ในการจัดลำดับจะจัดลำดับของความเหมือนจากมาก ไปหาน้อย จะใช้ Bubble Sort ในการจัดการ Bubble Sort เป็นการเรียงลำดับข้อมูลในสัญกรณ์ Big-O(N^2) เช่นเดียวกับการเรียงลำดับข้อมูลแบบ Selection Sort และยังมีลักษณะการเรียงที่คล้ายกัน กล่าวคือจะหาค่าที่ต้องการเรียง มาที่ละค่า โดยจะเริ่มที่ตำแหน่งแรก หรือตำแหน่งสุดท้ายของแถวลำดับก็ได้ สิ่งที่แตกต่างกันสำหรับการเรียงแบบ Bubble Sort คือจะไม่นำค่าที่ต้องการเรียงมาเทียบกับทุกค่าเหมือน Selection Sort แต่จะเปรียบเทียบกับค่าถัดไปทีละคู่ไปเรื่อยๆ เพื่อดันค่าที่ต้องการมาด้านหน้าหรือด้านหลัง

ขั้นตอนการเรียงลำดับข้อมูลมีดังนี้

1. เริ่มจากการนำค่าแรก มาเปรียบเทียบกับค่าถัดไป ซึ่งก็คือค่าที่สองในแถวลำดับ ถ้าไม่ตรงตามจุดประสงค์จะทำการสลับค่าทั้งสอง จากนั้นขยับมาเปรียบเทียบ ระหว่างค่า ที่ 2 และที่ 3 เช่นใน บรรทัดที่ 1-2 ของตัวอย่าง ทำเช่นนี้ไปเรื่อยๆ จนกว่าจะถึงค่าสุดท้าย ก็จะได้ค่าที่เรียงแล้วค่าแรกอยู่ ณ ตำแหน่งสุดท้ายของ แถวลำดับ
2. ดำเนินการเช่นเดียวกับข้อ 1. โดยเริ่มต้น จากการเปรียบเทียบ ค่าที่ 2 กับ ค่าที่ 3 และเปรียบเทียบ ทีละคู่ไปเรื่อยๆ จนถึงค่า รองสุดท้าย ก็จะได้ค่าที่เรียงแล้ว ค่าที่ 2 อยู่ ณ ตำแหน่งรองสุดท้าย ของแถวลำดับ เช่น บรรทัดที่ 6-9 ของตัวอย่าง
3. ดำเนินการเช่นเดียวกับข้อ 1. จนครบทุกค่า ก็จะได้แถวลำดับที่เรียงตามความประสงค์ นำ Bubble Sort มาทำการจัดเรียงลำดับความคล้ายคลึงกันที่เปรียบเทียบมาแล้วของทั้งภาพ และเสียงดังรูป 2.7

1)	รวม 1	8	20	30	3	4
2)		8	20	30	3	4
3)		8	20	30	3	4
4)		8	20	3	30	4
5)	รวม 2	8	20	3	4	30
6)		8	20	3	4	30
7)		8	20	3	4	30
8)		8	3	20	4	30
9)	รวม 3	8	3	4	20	30
10)		3	8	4	20	30
11)		3	8	4	20	30
12)		3	4	8	20	30
13)	รวม 4	3	4	8	20	30
14)	ผลลัพธ์	3	4	8	20	30

รูปที่ 2.7 การเรียงค่าโดยใช้ Bubble Sort

2.5 การรวมภาพและเสียง

การใช้ภาพเสียงเพื่อการสืบค้น จะใช้การรวมคุณลักษณะเด่นของภาพและเสียง ซึ่งคุณลักษณะของภาพจะใช้เทคนิคเอวีไอ (Adaptive Video Indexing : AVI) ซึ่งเป็นการวางตำแหน่งจุดเน้นที่แข็งแกร่งบนลักษณะเด่นที่ถูกต้องของข้อมูลข่าวสารที่เป็นที่ว่างชั่วคราว (Spatio-temporal Information) ภายในวิดีโอคลิปและคุณลักษณะของเสียง โดยการใช้วิธีความถี่ของช่วงเวลาเชิงสถิติ (Statistical Time-frequency) ซึ่งเป็นการประยุกต์ใช้แบบจำลองลาปลาซผสม (Laplacian Mixture Model : LMM) เข้าไปสู่ค่าสัมประสิทธิ์เวฟเลต (Wavelet coefficients) ดังที่ได้กล่าวไปแล้วข้างต้น

การค้นหาและเรียกคืนกลับวิดีโอโดยใช้การรวมภาพและเสียงเข้าด้วยกันจะมีความยืดหยุ่นและมีมาตราส่วนในการตัด (Scalable) เป็นอย่างสูง สามารถประยุกต์ใช้กับฐานข้อมูลวิดีโอได้หลากหลายชนิดอย่างมีประสิทธิภาพ

2.5.1 ทฤษฎีการรวมภาพและเสียง

2.5.1.1 ลำดับค่าของการรวมภาพและเสียง (Audio and Visual Score Ranking)

คุณลักษณะของภาพและเสียง อาจไม่ถูกรวมกันอย่างตรง ๆ เพราะความแตกต่างกันในด้านโครงสร้างทางฟิสิกส์หรือมิติของภาพและเสียง ดังนั้นการได้มาซึ่งการจัดอันดับของวิดีโอ จะต้องแยกกันระหว่างฐานข้อมูลของคุณลักษณะของเสียงและของภาพ โดยผลลัพธ์ในการจัดอันดับเหล่านี้ จะถูกรวบรวมเพื่อนำไปใช้ในการตัดสินใจการจัดอันดับที่คล้ายคลึงกันในขั้นสุดท้าย การใช้ฐานข้อมูลคุณลักษณะของภาพ สามารถเชื่อมโยงเข้ากับตัวชี้อันดับ ($Rank_i^{(v)}$) ของมัน ซึ่งในที่นี้ $Rank_i^{(v)} < Rank_j^{(v)}$ ถ้า $S_{qi}^{(v)} < S_{qj}^{(v)}, \forall j \neq i$ ในทำนองเดียวกัน การจัดอันดับของฐานข้อมูลคุณลักษณะของเสียงจะได้ค่าความคล้ายคลึง (Similarity scores) , $S_{qi}^{(a)}, i=1, \dots, M$ ซึ่งจะถูกนำไปใช้เพื่อให้ได้ตัวชี้อันดับ (Ranking indexs) , $Rank_i^{(a)}, i=1, \dots, M$ สำหรับลำดับที่ i ของช่วงวิดีโอ I_i ใด ๆ ผลลัพธ์ของ $Rank_i^{(v)}$ และ $Rank_i^{(a)}$ จะถูกรวมกันเพื่อให้ได้ ค่าความคล้ายคลึง (Similarity scores) ใหม่ ดังนี้

$$S_{qi}^{(v,a)} = Rank_i^{(v)} + \xi \times Rank_i^{(a)} \quad (2.14)$$

ซึ่ง ξ เป็นตัวประกอบของมาตราวัด (Scaling factor) ซึ่ง $0 \leq \xi \leq 1$ เพื่อควบคุมผลกระทบที่ได้จากผลลัพธ์ของการจัดอันดับคุณลักษณะของเสียง การรวมกันนี้จะได้เซตใหม่ของค่าความคล้ายคลึง (Similarity scores) , $S_{qi}^{(v,a)}, i=1, \dots, M$ ซึ่งจะถูกจัดเรียงเพื่อให้ได้เซตของการเรียกคืนกลับ (Retrieval set) และกำหนดว่าองค์ประกอบของเสียง (Audio content) อาจจะมีจำนวนของชั้น (Number of class) น้อยกว่าองค์ประกอบของภาพ (Visual content) ดังนั้น พารามิเตอร์ (Parameter)

)

๕ สามารถนำมาใช้เพื่อถ่วงน้ำหนักองค์ประกอบของเสียง (Audio content) เพื่อปรับปรุงเรียกคืนกลับ (retrieval) ให้ดียิ่งขึ้น แต่โปรแกรมนี้จะทำการใช้ ๕ เท่ากับ 1

)



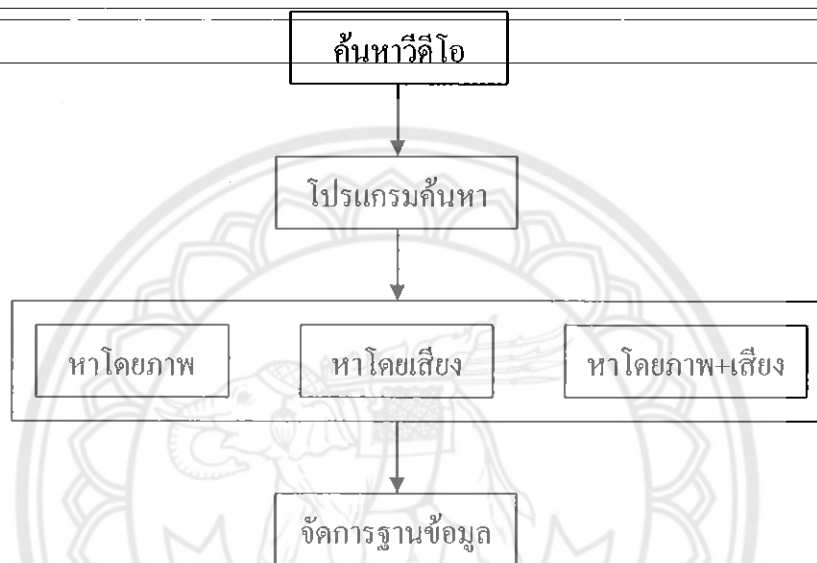
)

)

บทที่ 3

วิธีการดำเนินการ

ในการทดลองสืบค้นวิดีโอคลิป จะสามารถแยกประเภทของการสืบค้นและแบ่งขั้นตอนการดำเนินงานออกเป็นส่วนต่าง ๆ ดังนี้



รูปที่ 3.1 ขั้นตอนการค้นหาวีดิโอคลิป

3.1 การจัดการกับฐานข้อมูล

การจัดการกับฐานข้อมูล ขั้นแรกต้องหาวิธีการดึงข้อมูลจาก MATLAB มาเก็บข้อมูลไว้ใน Microsoft Access ก่อน เพราะว่าการเก็บข้อมูลใน MATLAB ในส่วนของ VB.NET ไม่สามารถทำการเชื่อมต่อได้โดยตรง จึงได้นำส่วนของข้อมูลที่มีอยู่ใน MATLAB แปลง Database ที่มีอยู่ใน MATLAB มาเก็บไว้ใน Microsoft Excel ก่อน เนื่องจากไม่สามารถนำข้อมูลจาก MATLAB มาเก็บไว้ใน Microsoft Access ได้โดยตรง

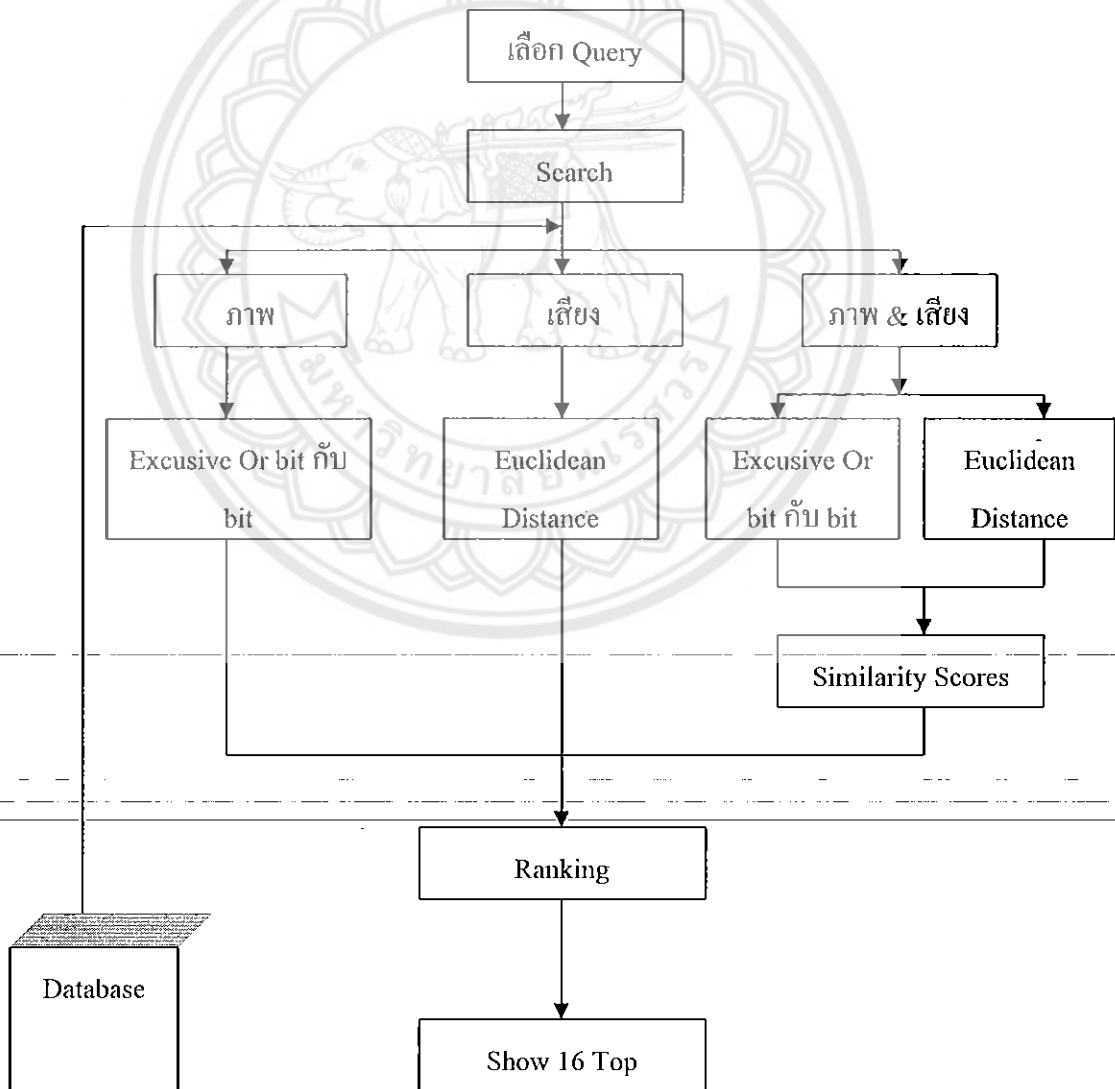
จากนั้นนำข้อมูลจาก Microsoft Excel มาเก็บไว้ใน Microsoft Access โดยชื่อวิดีโอ, ตัวชี้ (Index) ของภาพ เก็บไว้ในตารางชื่อ Keyframes โดยค่าของภาพจะถูกเก็บค่าไว้ถนัดในคอลัมน์ เดียว 6000 แถว และเก็บค่าไว้เป็น type MEMO ส่วนฐานข้อมูลของเสียงจะถูกเก็บไว้ค่าในตารางชื่อ Audio โดยเก็บไว้ทั้งหมด 29 คอลัมน์ 6000 แถว

3.2 โปรแกรมสืบค้นวีดีโอคลิป

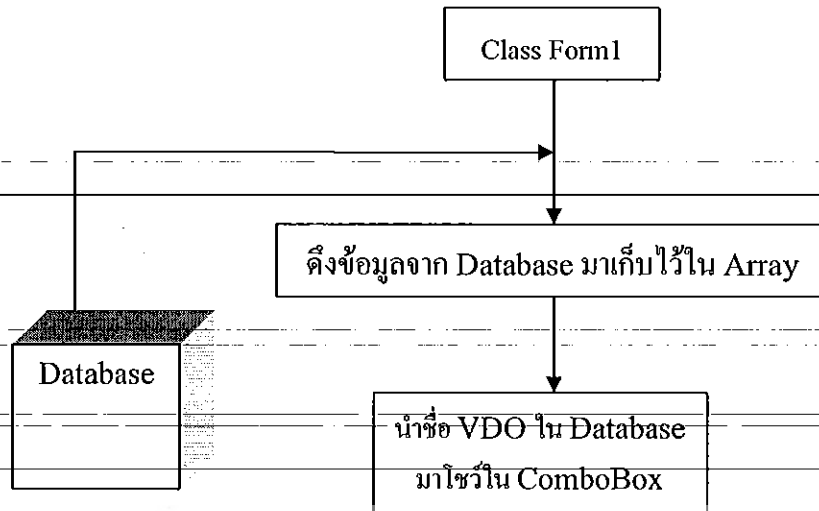
การค้นหาวีดีโอคลิป โดยแยกค้นหา ภาพ, เสียง, หรือ ภาพและเสียง ในการค้นหาจะต้องทำการสอบถาม (Query) ที่ต้องการค้นหา แล้วจึงผ่านกระบวนการของการค้นหา

การสืบค้นวีดีโอคลิป

การค้นหาวีดีโอคลิปสามารถค้นหาได้ 3 แบบ คือ การค้นหาโดยใช้เสียง การค้นหาโดยใช้ภาพ-และการค้นหาโดยใช้ภาพและเสียง-การค้นหาโดยใช้ภาพจะนำค่าที่เก็บเป็นสตริงมาเปรียบเทียบกับจนครบทั้งภาพ (6000x1) การค้นหาโดยเสียงจะนำเสียง (6000x29) มาเปรียบเทียบกับโดยใช้ยูคลิดีเนียน (Euclidean Distance) ส่วนการค้นหาภาพและเสียง ต้องนำทั้งข้อมูลภาพมาผ่านกระบวนการของภาพและข้อมูลเสียงมาผ่านกระบวนการของเสียงและนำค่าที่ได้มารวมกัน ดังรูปที่ 3.2



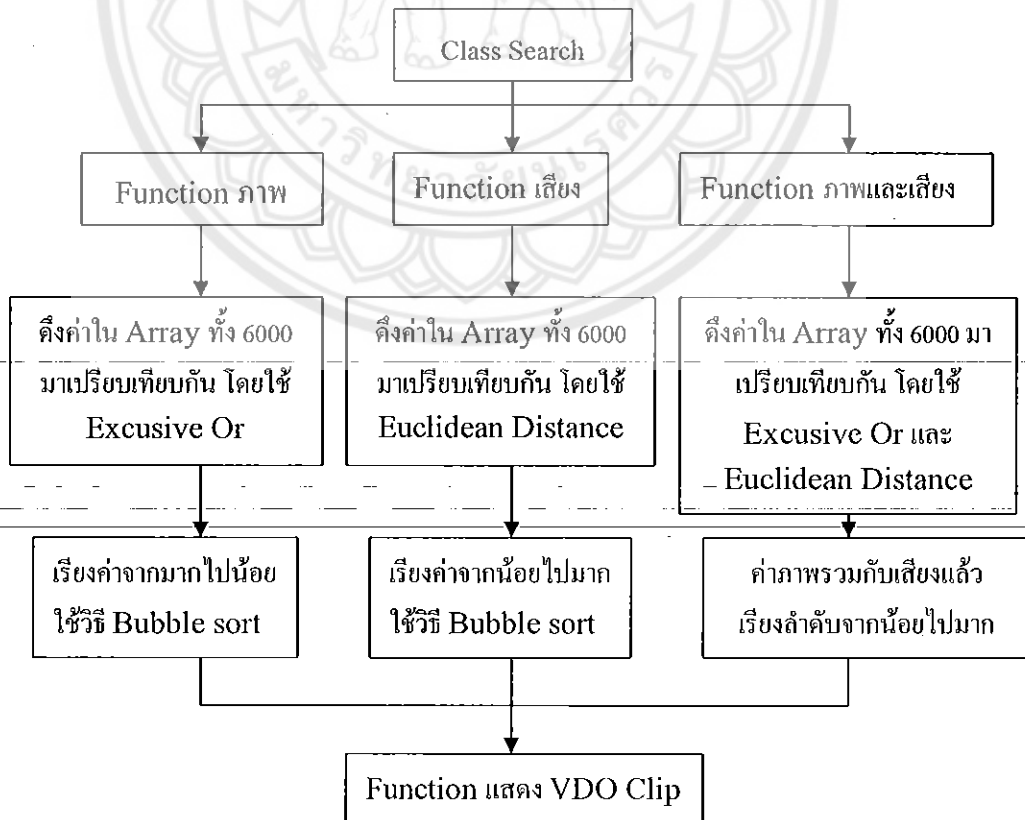
รูปที่ 3.2 แสดงภาพรวม การค้นหาวีดีโอคลิป



รูปที่ 3.3 แสดงภาพรวม การค้นหาวิดีโอคลิปโดยโปรแกรมของ Class Form1

การสืบค้นวิดีโอคลิป โดยโปรแกรม

Class Form1 ดึงตัวชี้ (Index) ของคลิปวิดีโอ จากฐานข้อมูล (Database) ดังรูปที่ 3.3 มาเก็บไว้ในอาร์เรย์ จนครบ 6000 คลิปวิดีโอ แล้วนำชื่อของ คลิปวิดีโอ ในฐานข้อมูล (Database) ทั้ง 6000 ชื่อมาโชว์ใน ComboBox



รูปที่ 3.4 แสดงภาพรวมการค้นหาวิดีโอคลิปโดยโปรแกรมของ Class Search

จากรูปที่ 3.4 Class Search ทำการค้นหาคลิปวิดีโอ (VDO Clip) ทั้ง 16 คลิปที่เหมือนกัน ขึ้นมาโชว์ใน PictureBox โดยการค้นหา (Search) จะแบ่งออกเป็นฟังก์ชัน (Function) ย่อยได้เป็น 4 ฟังก์ชัน คือ

1. ภาพ จะทำการค้นหาแต่ภาพโดยใช้อันกอร์ทิมการ Exclusive OR จากนั้นจะนำค่าที่ได้มาจัดเรียงค่าจากมากไปน้อย จากนั้นจะนำ 16 ค่าแรกมาโชว์ใน PictureBox
2. เสียง จะทำการค้นหาแต่เสียงโดยใช้ อันกอร์ทิม การหาระยะแบบยูคลิเดียน (Euclidean Distance) จากนั้นจะนำค่าที่ได้มาจัดเรียงจากน้อยไปมาก จากนั้นจะนำ 16 ค่าแรกมาโชว์ใน PictureBox
3. ภาพและเสียง จะทำการค้นหาทั้งภาพและเสียงโดยใช้อันกอร์ทิม การ Exclusive OR และอันกอร์ทิม การหาระยะแบบยูคลิเดียน (Euclidean Distance) โดยนำ ลำดับ (Rank) ที่ได้ทั้งภาพและเสียงมารวมกันแล้วเรียงจากน้อยไปมาก จากนั้นจะนำ 16 ค่าแรกมาโชว์ใน PictureBox
4. แสดงภาพ โดยจะทำการแสดง คลิปวิดีโอ (VDO Clip) ทั้ง 16 คลิป ที่มีความใกล้เคียงกับ คลิปที่เลือกนำมาแสดงใน PictureBox

การเปรียบเทียบภาพและเสียง

ในการเปรียบเทียบภาพและเสียงจะใช้การเปรียบเทียบแบบภาพและการเปรียบเทียบแบบเสียง โดยจะนำลำดับ (Rank) ของภาพและเสียงมารวมกัน ซึ่งสามารถแยกวิธีการเปรียบเทียบได้ดังนี้

เปรียบเทียบภาพโดยใช้ Exclusive OR

$$X = A \oplus B$$

X = Exclusive OR

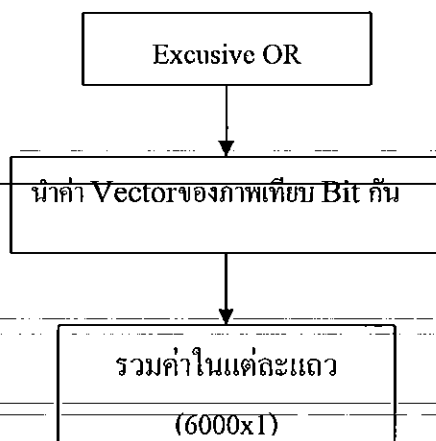
A = เป็นค่า Index ของภาพที่เลือก

B = เป็นค่า Index ของภาพที่นำมาเปรียบเทียบ

ร/ร.

93315

2550



รูปที่ 3.5 แสดงการทำงานใน Exclusive OR

จากรูปที่ 3.5 การทำงานของ Exclusive OR ของภาพมีกระบวนการ ซึ่งเมื่อทำกระบวนการนี้เสร็จก็จะได้ออกมาเป็น Exclusive OR ของภาพ ซึ่งต่างก็มีขนาด 6000x1

เปรียบเทียบเสียงแบบยูคลิเดียน (Euclidean Distance)

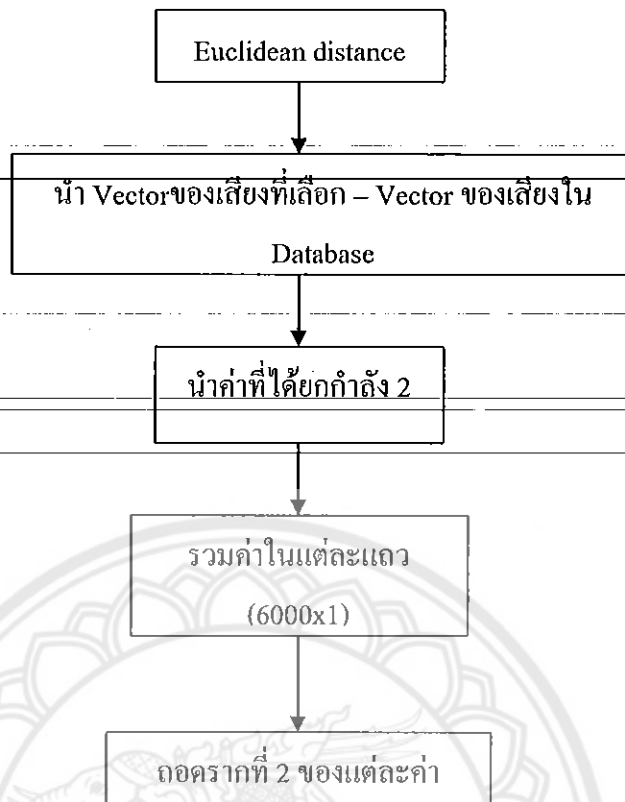
การหาระยะทาง (Euclidean distance) ระหว่างจุด $P = (p_1, p_2, \dots, p_n)$ และ $Q = (q_1, q_2, \dots, q_n)$ จะได้ว่า

$$d(P, Q) = \|P - Q\| = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

$d(P, Q)$ = ระยะทาง (Euclidean Distance)

p_i = เป็นค่าของ Component เสียงที่เลือก

q_i = เป็นค่าของ Component เสียงที่นำมาเปรียบเทียบ

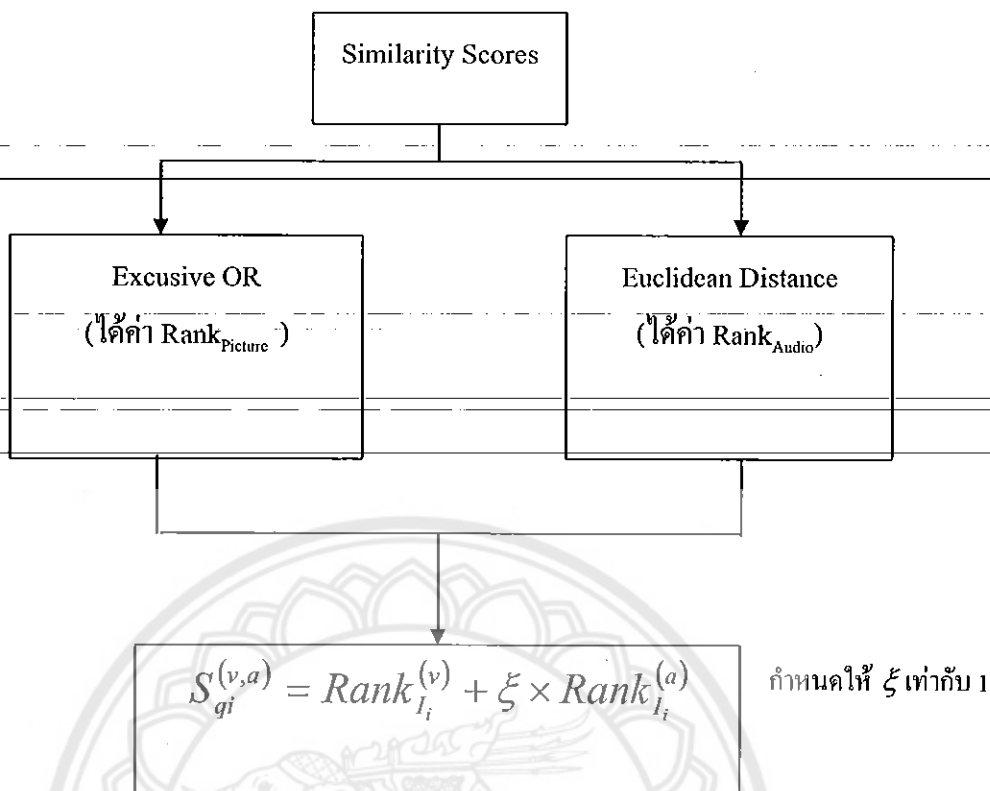


รูปที่ 3.6 แสดงการทำงาน Euclidean distance ของเสียง

จากรูปที่ 3.6 การทำงานของการหาระยะทางแบบยูคลิเดียน (Euclidean distance) ของเสียง มีกระบวนการทำงาน ซึ่งเมื่อทำกระบวนการนี้เสร็จก็จะได้ออกมาเป็นค่าระยะทางของเสียง ซึ่งต่างก็มีขนาด 6000×1

หาค่าความคล้ายคลึง

การหาค่าความคล้ายคลึงของภาพและเสียงมีกระบวนการทำงาน ดังรูปที่ 3.7 ซึ่งเมื่อทำกระบวนการนี้จะคำนวณหาลำดับ (Rank) ของภาพ และลำดับ (Rank) ของเสียง เสร็จก็จะได้อา ค่าความคล้ายคลึงของภาพและเสียง ซึ่งต่างก็มีขนาด 6000×1 โดยการคำนวณหาลำดับ (Rank) ของภาพและเสียงสามารถคำนวณได้จาก $S_{ij}^{(v,s)} = Rank_{I_i}^{(v)} + \xi \times Rank_{I_i}^{(s)}$ โดยกำหนดให้ ξ เท่ากับ 1



รูปที่ 3.7 แสดงการหาค่าความคล้ายคลึงของภาพและเสียง

การเรียงลำดับ (Ranking)

การเรียงลำดับจะแบ่งเป็น ภาพ, เสียง และภาพกับเสียง มีกระบวนการทำงาน ดังรูปที่ 3.8 โดยจะนำข้อมูลที่ผ่านมากระบวนการแล้วมาเรียงลำดับข้อมูล และจะนำ 16 Clip แรกมาแสดง

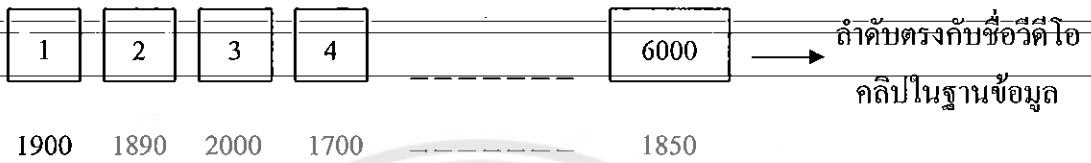


รูปที่ 3.8 แสดงการเรียงลำดับของภาพและเสียง

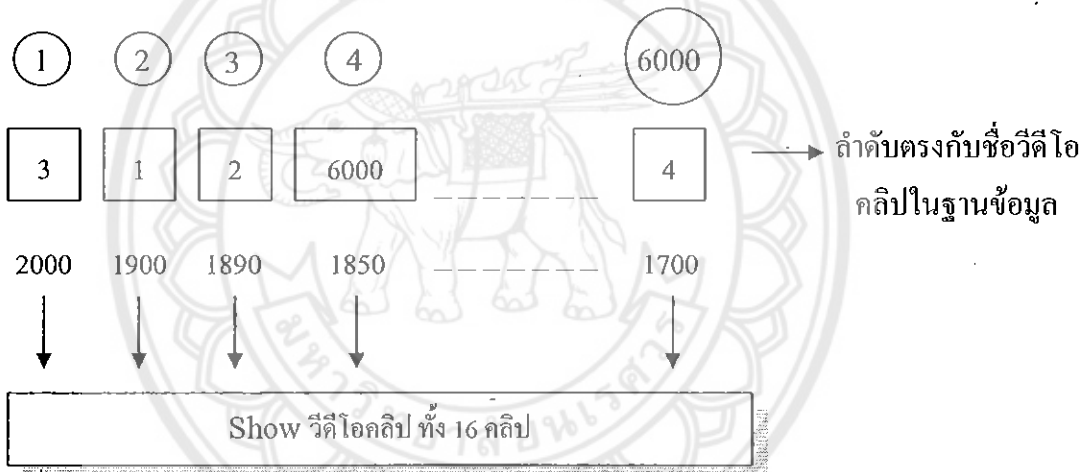
การหาระยะของภาพ (Distance picture)

การเรียงลำดับ (Rank) ภาพ จะเรียงค่า ระยะ (Distance) จากมากไปหาน้อย เพราะว่าค่าที่มากที่สุดก็คือค่าที่มีจำนวน บิต เท่ากันมากที่สุด ซึ่งจะมีค่าใกล้เคียงกัน เมื่อกระบวนการสิ้นสุดจะได้ลำดับค่าระยะ (Distance) ของภาพ ซึ่งเรียงจากค่ามากไปหาน้อย ดังรูปที่ 3.9

ใช้การเปรียบเทียบแบบ Exclusive OR



เรียงลำดับ (Rank) จากมากไปน้อย

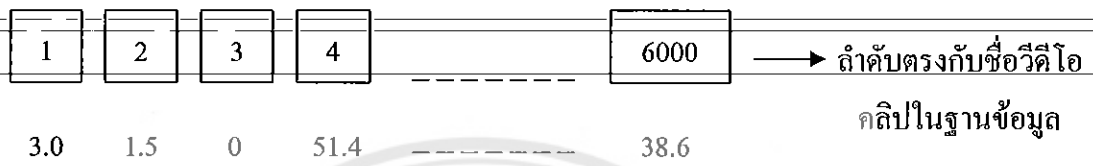


รูปที่ 3.9 แสดงการทำงานการ เรียงลำดับ (Rank) ภาพ

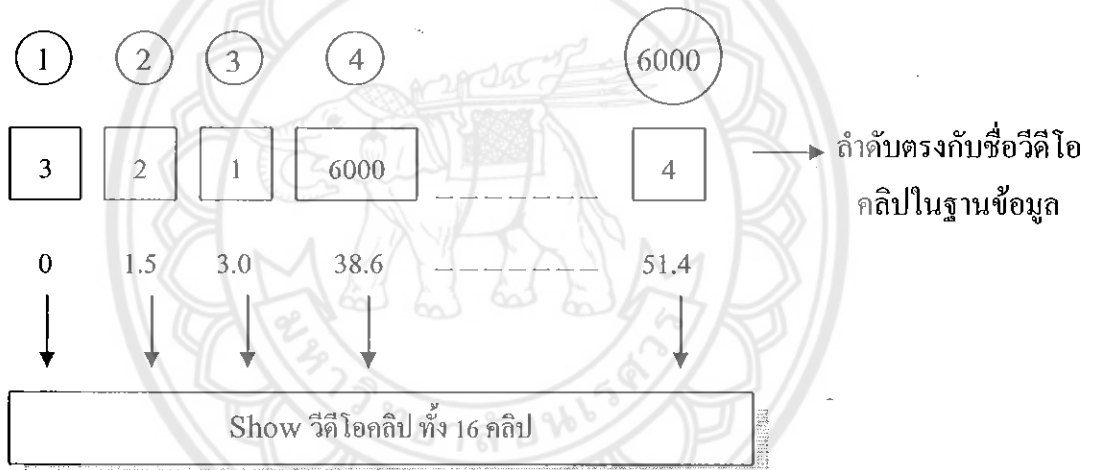
การหาระยะของเสียง (Distance audio)

การเรียงลำดับ (Rank) เสียง จะเรียงค่าระยะ (Distance) จากน้อยไปหามาก เพราะว่าค่าที่น้อยที่สุดก็คือค่าที่มี ระยะ (Distance) ค่าใกล้เคียงกัน เมื่อกระบวนการสิ้นสุดจะได้ลำดับค่าระยะ (Distance) ของเสียง ซึ่งเรียงจากค่าน้อยไปหามาก ดังรูปที่ 3.10

ใช้การเปรียบเทียบแบบยูคลิดีเนียน (Euclidean distance)



การเรียงลำดับ (Rank) จากน้อยไปมาก



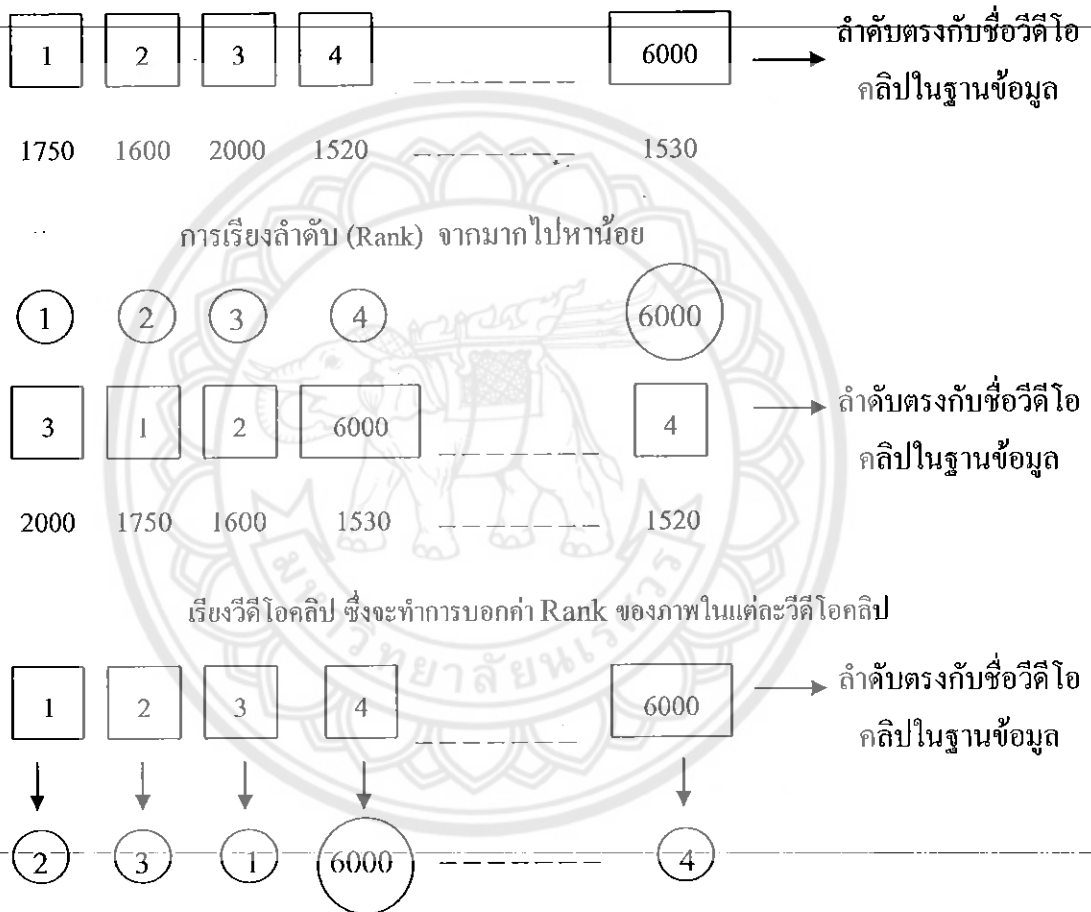
รูปที่ 3.10 แสดงการทำงานการการเรียงลำดับ (Rank) เสียง

ระยะของภาพและเสียง (Distance picture and audio)

การเรียงระยะของภาพ (Distance picture)

เมื่อได้ค่าระยะ (Distance) ของภาพแล้วก็จะทำการเรียงลำดับ (Rank) จากมากไปน้อย จากนั้นจะนำค่าวีดิโอคลิปมาเรียงจากวีดิโอคลิปที่ 1 ถึงวีดิโอคลิป 6000 ซึ่งวีดิโอแต่ละคลิปจะทำการบอกค่าลำดับ (Rank) ไว้เพื่อนำไปคำนวณต่อไป ดังรูปที่ 3.11

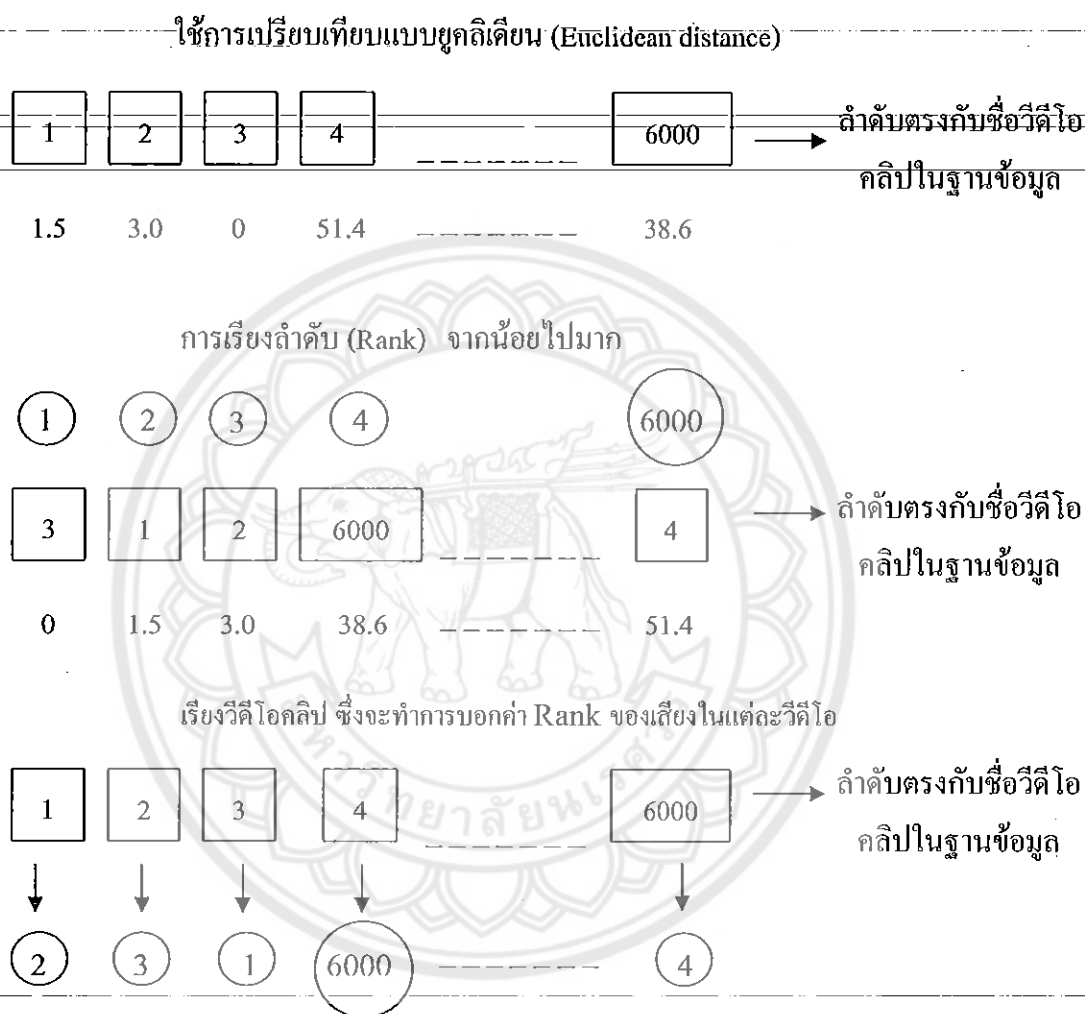
ใช้การเปรียบเทียบแบบ Exclusive OR



รูปที่ 3.11 แสดงการการเรียงระยะของภาพ (Distance picture)

การเรียงระยะของเสียง (Distance audio)

เมื่อได้ค่าระยะ (Distance) ของเสียงแล้ว ก็จะทำการเรียงลำดับ (Rank) จากน้อยไปมาก จากนั้นจะนำวิดีโอคลิปมาเรียงจากวิดีโอคลิปที่ 1 ถึงวิดีโอคลิป 6000 ซึ่งวิดีโอแต่ละคลิปจะทำการบอกค่าลำดับ (Rank) ไว้เพื่อนำไปคำนวณต่อไป ดังรูปที่ 3.12

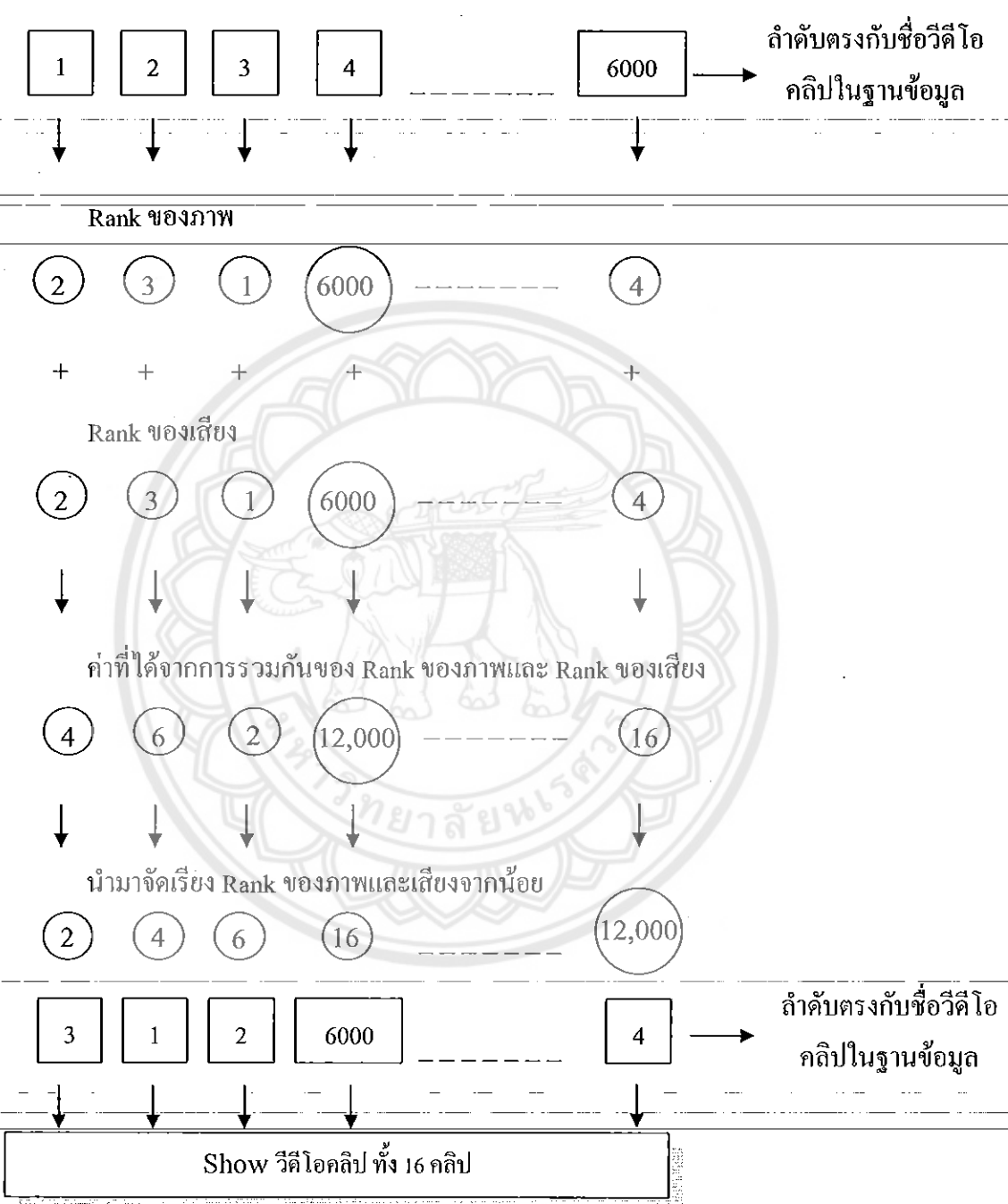


รูปที่ 3.12 แสดงการเรียงระยะของเสียง (Distance audio)

การรวมภาพและเสียง

นำลำดับ (Rank) ของภาพและเสียงมารวมกัน โดยการรวมกันสามารถคำนวณจากสมการ

$$S_{qi}^{(v,a)} = Rank_{I_i}^{(v)} + \xi \times Rank_{I_i}^{(a)} \text{ โดยกำหนดให้ } \xi \text{ เท่ากับ } 1$$



รูปที่ 3.13 แสดงการทำงานการรวมภาพและเสียง

การรวมกันของภาพและเสียง หาได้จากสมการ $S_{gi}^{(v,a)} = Rank_{i_i}^{(v)} + \xi \times Rank_{i_i}^{(a)}$ โดยกำหนดให้ ξ เท่ากับ 1 จากนั้นจะนำค่าลำดับ (Rank) ของภาพมารวมกับค่าลำดับ (Rank) ของเสียง เมื่อได้ค่าลำดับ (Rank) ของภาพและเสียงแล้ว จึงนำลำดับ (Rank) ของภาพและเสียงมาเรียงลำดับจากน้อยไปมาก จากนั้นจะนำ 16 ลำดับแรกมาแสดง



บทที่ 4

ผลการทดลอง

การทดลองค้นหาไฟล์วีดีโอคลิปของ โปรแกรม โดยการค้นหานั้นจะใช้ทั้งภาพและเสียงในการค้นหา ผลที่ได้เป็นผลมาจากกรณีศึกษา และใช้หลักการเปรียบเทียบผลของค่า Precision ซึ่งเป็นค่าที่ได้จากการคำนวณดังต่อไปนี้

$$\% \text{ ความถูกต้อง} = (\text{จำนวนวิดีโอที่ถูกเลือก} / \text{จำนวนวิดีโอทั้งหมด}) \times 100 \quad 4.1$$

$$\text{เช่น } \% \text{ ความถูกต้อง} = \left(\frac{10}{16} \right) \times 100 = 62.5 \%$$

แล้วนำค่าความถูกต้อง (Precision) ของตัวสอบถาม (Query) ตัวอย่างทั้งหมด มารวมกันแล้วหาค่าเฉลี่ย ก็จะได้เป็นค่าความถูกต้อง (Precision) ของตัวสอบถาม (Query) ออกมา

4.1 สิ่งที่ต้องเตรียมก่อนทำการทดลอง

- 1) โปรแกรมค้นหาวีดีโอคลิปที่ทำการพัฒนา มาจาก โปรแกรม Microsoft Visual Basic .NET (VB.NET)
- 2) วิดีโอคลิปในฐานะข้อมูลทั้งหมด 6000 ไฟล์ ซึ่งเป็นไฟล์ .AVI ทั้งหมด
- 3) ไฟล์ .JPG ที่เป็นฐานข้อมูลของ Query ทั้งหมด 6000 ไฟล์ เพื่อนำมาใช้เป็นตัว Query ในการค้นหาไฟล์วีดีโอคลิป

4.2 ผลการเปรียบเทียบความเร็วในการสืบค้นวีดีโอคลิปด้วยโปรแกรม

ตารางที่ 4.1 บันทึกผลการทดลองเปรียบเทียบเวลาที่ใช้ในการค้นไฟล์วีดีโอ กับ โปรแกรม นาย ฉัตรชัย สุขสอน[3]

โปรเจกต์	โปรเจกต์นาย ฉัตรชัย สุขสอน			โปรเจกต์ทำ		
	ภาพ(6000x2000) distance	เสียง(6000x29) distance	รวมภาพและเสียง	ภาพ(6000x2000) XOR	เสียง(6000x29) Euclidin	รวมภาพและเสียง
1. patriot_clip1000	23.46	7.85	28.76	3.07	2.65	3.5
2. postman_clip857	22.08	7.96	28.87	3.07	2.65	3.82
3. junaid7_clip14	23.14	8.17	28.90	3.18	2.33	3.39
4. titanic_clip1174	23.46	8.28	28.87	3.18	2.54	3.5
5. titanic_clip999	23.35	8.30	28.76	2.97	2.44	3.5

จากตารางจะแสดงให้เห็นค่าผลเฉลี่ยของการจับเวลา ของ โครงานนาย ฉัตรชัย สุขสอน[3] กับ โครงานที่ได้ทำซึ่งเห็นได้ว่า การค้นหาแต่ละแบบของ โครงานที่ได้ทำขึ้นมาจะใช้เวลาในการค้นหาวีดีโอคลิปได้เร็วกว่า

4.3 การทดลองค้นหาวีดีโอคลิปโดยใช้ภาพ, เสียง และการรวมภาพและเสียง

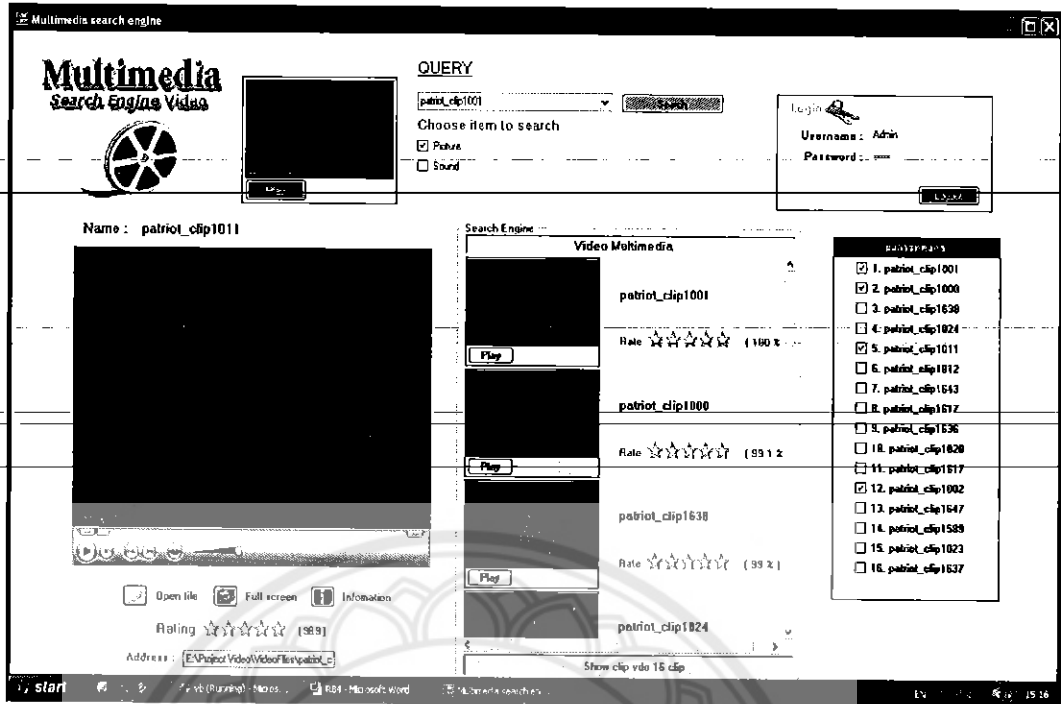
ในการทดลองจะให้ผู้ทดลองทำการเลือกวีดีโอคลิปที่มีความใกล้เคียงกับตัวสอบถาม (Query) จากวีดีโอคลิปทั้ง-16 คลิป จากนั้นจะนำค่าที่ได้มาหาค่าเฉลี่ย ซึ่งจะมีขั้นตอน ดังนี้-

- 1) ทำการเลือก ตัวสอบถาม (Query) ที่ต้องการค้นหา
- 2) โปรแกรมจะทำการเรียกเก็บค่าของภาพและเสียง เพื่อไปผ่านกระบวนการต่าง ๆ ตามที่ได้กล่าวในบทที่-3 หัวข้อที่ 3.2
- 3) แสดงผลของการค้นหาวีดีโอคลิป โดยจะทำการแสดง 16 คลิป ที่ได้จากการเรียง
- 4) ทำการเลือกวีดีโอคลิป แล้วทำการเช็คเครื่องหมายถูกที่ CheckBox เมื่อคลิปมีความ

ใกล้เคียงกับตัวสอบถาม (Query)

รายชื่อผู้ที่ทำการทดลองโดยใช้ผู้ทดลอง 10 คน ประกอบด้วย

1. นาย ฤทธิจักร เครือสาร
2. นายชัชชา ชาวเหนือ
3. นายชยศ กองจันทร์
4. นางสาวนันทา บุญประคอง
5. นายพิริยการ วงศ์ประสิทธิ์
6. นางสาวพัชรา ภักดี
7. ด.ญ.ธันวารัตน์ ชาวเหนือ
8. นางสาวพรรณภทรา สุดแสง
9. นายนพดล โดบันลือภพ
10. นายธีระพงษ์ บุญเกี่ยวงศ์



รูปที่ 4.1 แสดงผลการค้นหาวีดิโอคลิปด้วยภาพ

ตารางที่ 4.2 บันทึกผลการทดลองการสืบค้นด้วยภาพในวีดิโอคลิปเฉลี่ย โดยผู้ใช้

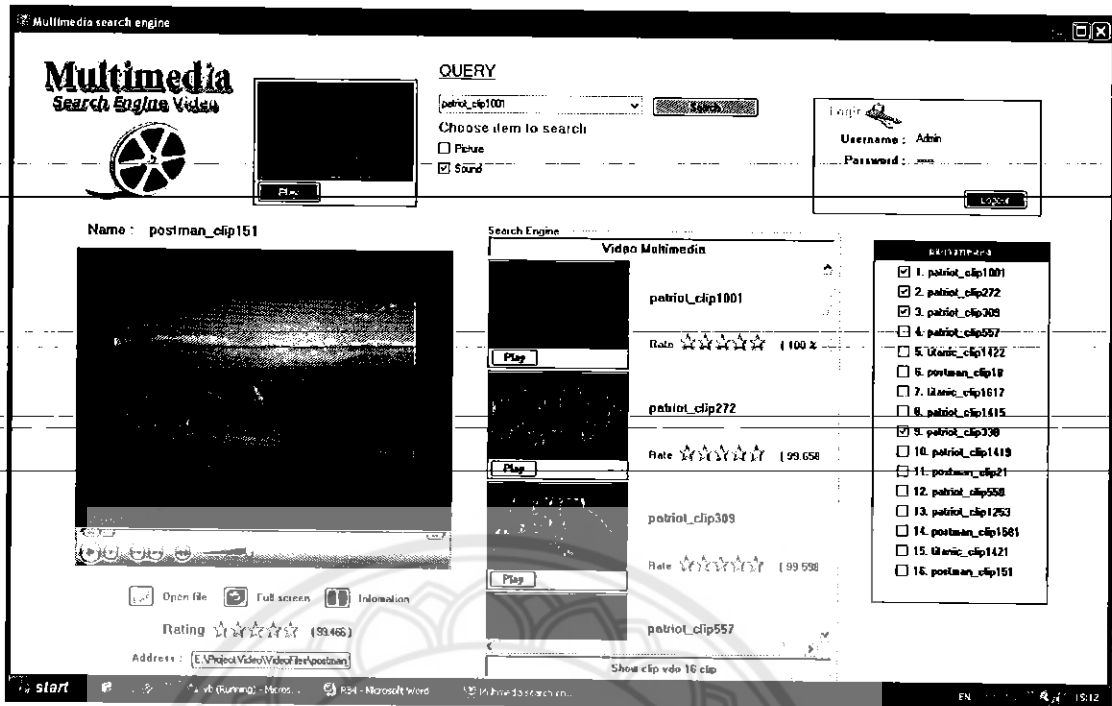
ชื่อคลิป(ภาพเหตุการณ์)	ผู้ใช้ (User)	ภาพ										ค่าเฉลี่ยจากผู้ใช้ 10 ครั้ง (AVG)
		ค่าความถูกต้อง (%Precision) ของผู้ใช้ 16 วีดิโอคลิป										
		1	2	3	4	5	6	7	8	9	10	
1.patrio_clip1002(ฉากสีฟ้า)		62.50	50.00	50.00	56.25	50.00	43.75	43.75	50.00	56.25	50.00	51.25
2.patrio_clip1441(ถนนรวม)		50.00	43.75	37.50	37.50	56.25	50.00	50.00	43.75	50.00	56.25	47.50
3.fitaide_clip1553(บ้าน: เถ)		43.75	43.75	43.75	50.00	56.25	50.00	43.75	37.50	31.25	43.75	44.38
4.postman_clip10(บ่อน้ำ)		31.25	31.25	25.00	25.00	37.50	31.25	31.25	25.00	37.50	31.25	30.63
5.postman_clip514(ถนนเอเซีย)		43.75	37.50	31.25	31.25	50.00	43.75	43.75	37.50	43.75	43.75	40.63
6.jamald14_clip15(Music Video)		43.75	43.75	25.00	31.25	62.50	68.75	43.75	50.00	68.75	68.75	50.63
7.jamald2_clip30(Music Video)		43.75	37.50	43.75	43.75	31.25	25.00	37.50	37.50	37.50	43.75	38.13
8.jamald2_clip40(Music Video)		18.75	25.00	18.75	25.00	18.75	25.00	31.25	12.50	25.00	25.00	22.50
9.jamald10_clip14(Music Video)		6.25	6.25	37.50	37.50	37.50	50.00	31.25	37.50	31.25	37.50	31.25
10.jamald10_clip2(Music Video)		56.25	56.25	100.00	100.00	100.00	100.00	56.25	62.50	93.75	100.00	82.50
AVG %Precision(10 Clip)		40.00	37.50	41.25	43.75	50.00	48.75	41.25	39.38	47.50	50.00	41.91

ช่องสี่ฟ้าเข้ม คือผลเฉลี่ยของค่าความถูกต้องทั้งหมด ของค้นหาคลิปแบบภาพ

จากการค้นหาวีดิโอคลิป จากตัวสอบถาม (Query) ต้นแบบ เช่น patriot_clip1001 ผลจากการค้น จะได้วีดิโอคลิปที่ใกล้เคียงตัวอย่างทั้งหมด 4 วีดิโอคลิป ดังรูปที่ 4.1 แล้วนำมาทำการคำนวณหา %ค่าถูกต้อง (Precision) จะได้ว่า

$$\% \text{ Precision} = \left(\frac{4}{16} \right) \times 100 = 25\%$$

แล้วนำค่าที่ได้บันทึกผลลงใน ตารางที่ 4.2



รูปที่ 4.2 แสดงผลการค้นหาวีดิโอคลิปด้วยเสียง

ตารางที่ 4.3 บันทึกรายการทดลองการสืบค้นด้วยเสียงในวีดิโอคลิปเฉลี่ยโดยผู้ใช้

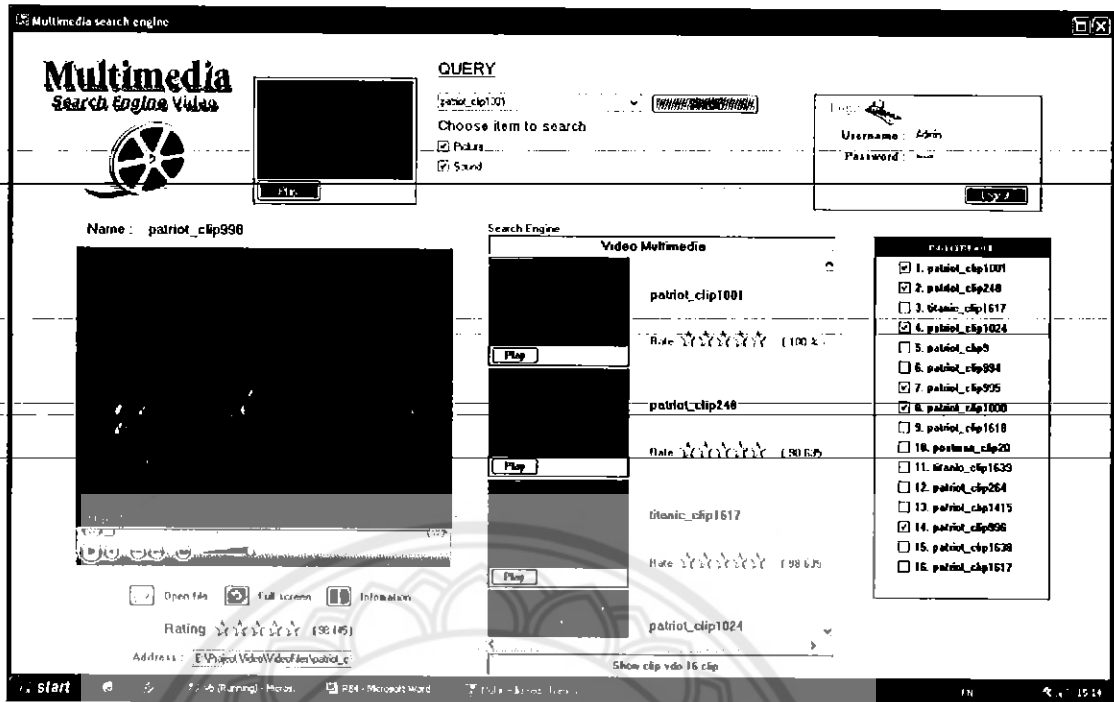
ชื่อคลิป(จากเหตุการณ์)	ผู้	สืบค้น										ค่าเฉลี่ยจากผู้ใช้ 10 ครั้ง (AVG)
		ค่าความถูกต้อง(%Precision) ของผู้ใช้ 16 วีดิโอคลิป										
		1	2	3	4	5	6	7	8	9	10	
1.patrio_clip1002(รถถังสีฟ้า)		25	25	25	25	37.5	43.75	31.25	37.5	31.25	37.5	31.875
2.patrio_clip1441(ทหารอเมริกัน)		50	43.75	68.75	62.5	100	93.75	68.75	75	93.75	93.75	75
3.titanic_clip1553(ไททานิค)		43.75	43.75	31.25	37.5	43.75	43.75	37.5	43.75	43.75	37.5	40.625
4.postman_clip10(โปสการ์ด)		37.5	37.5	43.75	43.75	43.75	37.5	43.75	37.5	37.5	43.75	40.625
5.postman_clip514(โปสการ์ด)		43.75	43.75	25	25	37.5	43.75	25	37.5	25	25	33.125
6.jamaid14_clip15(Music Video)		62.5	56.25	62.5	62.5	62.5	68.75	56.25	62.5	62.5	62.5	61.875
7.jamaid2_clip30(Music Video)		56.25	56.25	62.5	68.75	68.75	62.5	62.5	68.75	56.25	56.25	61.875
8.jamaid2_clip40(Music Video)		56.25	50	43.75	50	18.75	18.75	50	56.25	56.25	62.5	46.25
9.jamaid10_clip14(Music Video)		56.25	50	62.5	62.5	56.25	50	62.5	56.25	31.25	37.5	52.5
10.jamaid10_clip2(Music Video)		31.25	31.25	31.25	25	31.25	37.5	25	31.25	37.5	37.5	31.875
AVG %Precision(10 Clip)		46.25	43.75	45.625	46.25	50	50	46.25	50.625	47.5	49.375	46.25

ช่องสี่ฟ้าเข้ม คือผลเฉลี่ยของค่าความถูกต้องทั้งหมด ของค้นหาคลิปแบบเสียง

จากการค้นหาวีดิโอคลิป จากการสอบถาม (Query) ต้นแบบ เช่น patriot_clip1001 ผลจากการค้น จะได้อีวีโอคลิปที่ใกล้เคียงตัวอย่างทั้งหมด 4 วีดิโอคลิป ดังรูปที่ 4.2 แล้วนำมาทำการคำนวณหา %ค่าถูกต้อง (Precision) จะได้ว่า

$$\% \text{ Precision} = \left(\frac{4}{16} \right) \times 100 = 25\%$$

แล้วนำค่าที่ได้บันทึกผลลงใน ตารางที่ 4.3



รูปที่ 4.3 แสดงผลการค้นหาวิดีโอคลิปด้วยภาพและเสียง

ตารางที่ 4.4 บันทึกผลการทดลองการสืบค้นด้วยภาพและเสียงในวิดีโอคลิปเฉลี่ยโดยผู้วิจัย

ชื่อคลิป(เลขที่ผลการค้น)	ผู้วิจัย	ภาพและเสียง										ค่าเฉลี่ยของผู้วิจัย 10 ครั้ง (AVG)
		ค่าความถูกต้อง (%Precision) ของผู้วิจัย 16 วิดีโอคลิป										
		1	2	3	4	5	6	7	8	9	10	
1 patriot_clip1002(นกที่สี่)		37.5	50	56.25	50	75	75	62.5	62.5	75	75	61.875
2 patriot_clip1441(ถนนบรม)		75	68.75	75	68.75	100	100	75	75	93.75	100	83.125
3 titanix_clip1553(โถง:10)		68.75	68.75	56.25	56.25	68.75	62.5	68.75	75	56.25	50	63.125
4 postman_clip10(บ๊อง)		37.5	37.5	31.25	31.25	50	50	37.5	43.75	50	50	41.875
5 postman_clip514(กานแดงจ)		50	50	25	25	43.75	50	50	50	43.75	43.75	43.125
6 jannaid14_clip15(Music Video)		50	56.25	62.5	68.75	68.75	68.75	56.25	68.75	68.75	62.5	63.125
7 jannaid2_clip30(Music Video)		68.75	62.5	43.75	50	50	50	62.5	68.75	50	50	55.625
8 jannaid2_clip40(Music Video)		56.25	50	50	50	25	25	56.25	50	56.25	50	46.875
9 jannaid10_clip14(Music Video)		50	50	68.75	68.75	50	56.25	56.25	56.25	68.75	56.25	58.125
10 jannaid10_clip2(Music Video)		37.5	37.5	31.25	37.5	12.5	12.5	31.25	25	37.5	31.25	29.375
AVG %Precision(10 Clip)		53.125	53.125	50	50.625	54.375	55	55.625	57.5	60	56.875	

ช่องสี่ฟ้าเข้ม คือผลเฉลี่ยของค่าความถูกต้องของค้นหาคlipแบบ รวมภาพและเสียง

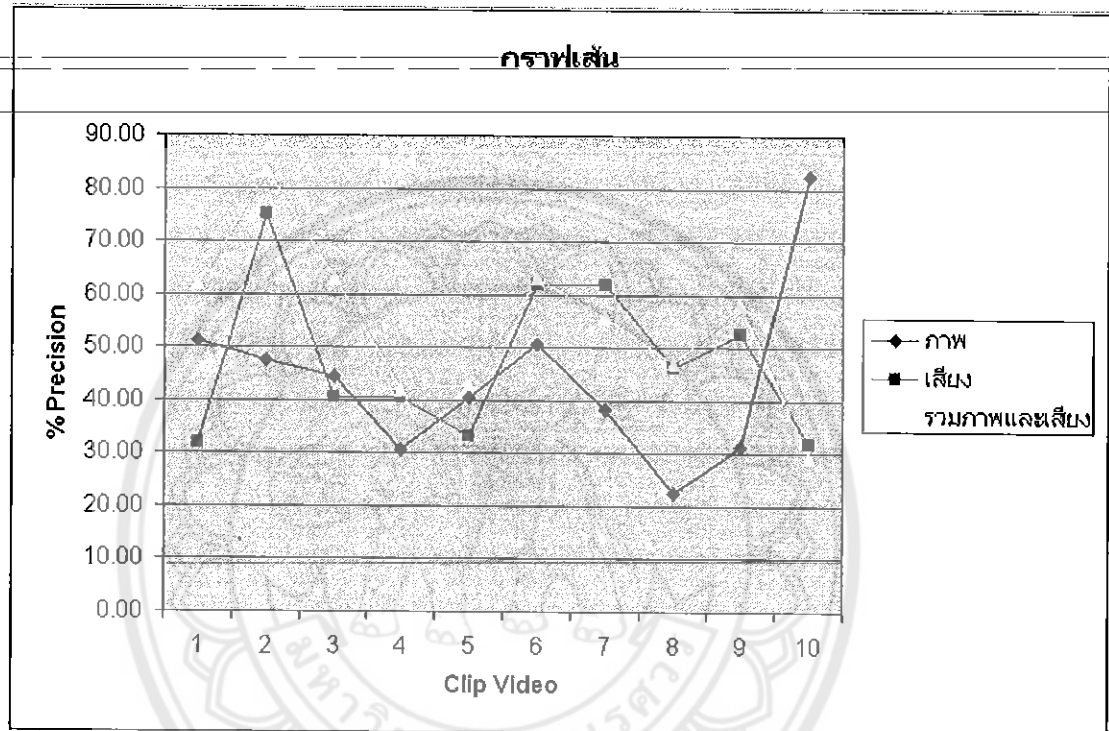
จากการค้นหาวิดีโอคลิป จากการสอบถาม (Query) ต้นแบบ เช่น patriot_clip1001 ผลจากการค้น จะได้วิดีโอคลิปที่ใกล้เคียงตัวอย่างทั้งหมด 6 วิดีโอคลิป ดังรูปที่ 4.3 แล้วนำมาทำการคำนวณหา %ค่าถูกต้อง (Precision) จะได้ว่า

$$\% \text{ Precision} = \left(\frac{6}{16} \right) \times 100 = 37.5 \%$$

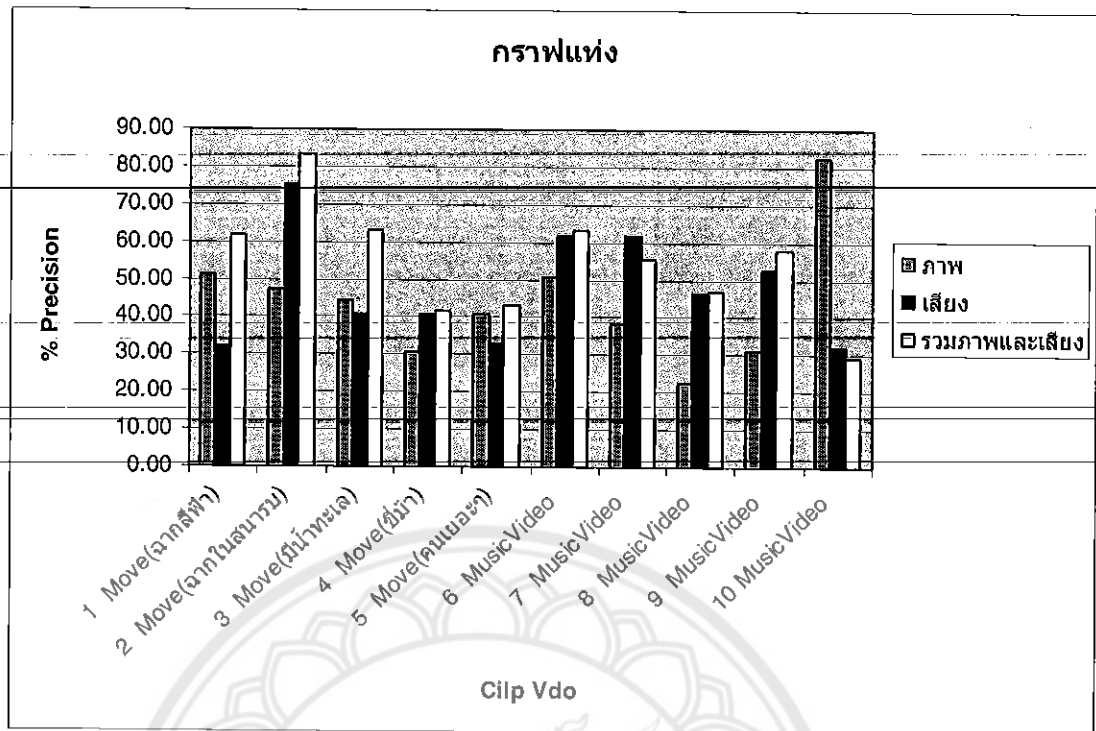
แล้วนำค่าที่ได้บันทึกผลลงใน ตารางที่ 4.4

4.4 เปรียบเทียบผลการทดลอง

นำผลการทดลองที่หาค่าถูกต้อง (Precision) ของวีดิโอคลิปตัวอย่างแต่ละคลิปมาทำการพล็อตกราฟ เปรียบเทียบการสืบค้นไฟล์วีดิโอคลิป จะเห็นได้ว่าค่าความถูกต้องของการสืบค้นแบบภาพและเสียงรวมกัน โดยส่วนใหญ่จะมีค่าสูงกว่าการสืบค้นแบบภาพ หรือแบบเสียง ดังรูปที่ 4.4



รูปที่ 4.4 แสดงกราฟเปรียบเทียบการสืบค้นวีดิโอแบบภาพ, เสียง และรวมภาพกับเสียง



รูปที่ 4.5 แสดงกราฟแท่งเปรียบเทียบการสืบค้นวิดีโอจากกราฟแท่ง

จากรูป จะเห็นได้ว่า จากบางฉากหรือคลิปบางคลิปอาจค้นหาแบบเสียงอย่างเดียวได้ดีกว่า หรือภาพเพียงอย่างเดียวได้ดีกว่า เช่น ดูจากประเภทของคลิป เมื่อดูจากกราฟแล้ว เป็นคลิปหนัง (Movie clip) จะเห็นได้ว่าโดยส่วนใหญ่ การค้นหา แบบภาพอย่างเดียวจะมีความถูกต้องมากกว่า การค้นหาแบบเสียงเพียงอย่างเดียว แต่พอเป็นคลิปประเภท มิวสิควิดีโอ เห็นได้ว่าการค้นหาแบบเสียงอย่างเดียวจะมีค่าความถูกต้องมากกว่าการค้นหาแบบภาพอย่างเดียว

4.5 การเปรียบเทียบอัลกอริทึมของระบบที่พัฒนาจากนายจักรชัย สุขสอน[3] และระบบที่พัฒนาขึ้น

การเปรียบเทียบของ 2 ระบบ จะทำการคำนวณจากอัลกอริทึมที่ได้นำไปใช้ในการสืบค้นวิธีโอคลิปทั้ง 2 ระบบ ซึ่งการคำนวณจากอัลกอริทึมสามารถคำนวณได้ ดังนี้

4.5.1 ระบบที่พัฒนาจากนายจักรชัย สุขสอน[3] ในการสืบค้นภาพและเสียง

ใช้สูตร Distance ในการเปรียบเทียบภาพ ดังนี้

$$d = \frac{\vec{V}_q \cdot \vec{V}}{\|\vec{V}_q\| \times \|\vec{V}\|}$$

d = Distance

\vec{V}_q = เวกเตอร์ของตัวสอบถาม (Query)

\vec{V} = เวกเตอร์ทั้งหมดของฐานข้อมูล

$\|\vec{V}_q\|$ = ขนาดของตัวสอบถาม (Query)

$\|\vec{V}\|$ = ขนาดของฐานข้อมูล

การหาค่า $\|\vec{V}_q\| = \sqrt{\sum_{i=1}^n v_q^2}$

ขั้นที่ 1 v_q^2 ยกกำลังสอง n Component

ขั้นที่ 2 $\sum_{i=1}^n v_q^2$ บวกค่าของขั้นตอนที่ 1 n-1 Component

ขั้นที่ 3 $\sqrt{\sum_{i=1}^n v_q^2}$ ถอดรากที่สอง 1 ครั้ง

เพราะฉะนั้น การเปรียบเทียบทั้งหมดเท่ากับ $2n$ ครั้ง

การหาค่า $\|\vec{V}\| = \sqrt{\sum_{i=1}^n v^2}$

ขั้นที่ 1 v^2 ยกกำลังสองทั้งหมด n Component

ขั้นที่ 2 $\sum_{i=1}^n v^2$ บวกค่าของขั้นตอนที่ 1 ทั้งหมด n-1 Component

ขั้นที่ 3 $\sqrt{\sum_{i=1}^n v^2}$ ถอดรากที่สอง 1 ครั้ง

เพราะฉะนั้น การเปรียบเทียบทั้งหมดเท่ากับ $2n$ ครั้ง

กรณีในการสืบค้น \vec{V}_q คู่กับ \vec{V}

ขั้นที่ 1 นำค่ามาคูณทั้งหมด n ครั้ง

ขั้นที่ 2 นำค่า \vec{V}_q บวกกับค่า \vec{V} ทั้งหมด $n-1$ ครั้ง

กรณีในการสืบค้น $\|\vec{V}_q\|$ คู่กับ $\|\vec{V}\|$

ขั้นที่ 3 นำค่ามาคูณทั้งหมด n ครั้ง

กรณีในการสืบค้น $d = \frac{\vec{V}_q \cdot \vec{V}}{\|\vec{V}_q\| \times \|\vec{V}\|}$

ขั้นที่ 4 นำค่ามาหารกัน ใช้การเปรียบเทียบ 1 ครั้ง

การเปรียบเทียบเท่ากับ $3n$ ครั้ง

เปรียบเทียบจนครบทั้งหมดในฐานข้อมูล ใช้การเปรียบเทียบ m ครั้ง

เพราะฉะนั้น การเปรียบเทียบทั้งหมดเท่ากับ $(2n + 2n + 3n) \times m = 6nm$ ครั้ง

ใช้สูตร Distance ในการเปรียบเทียบเสียง ดังนี้

การเปรียบเทียบของเสียง ใช้สูตร Distance เช่นเดียวกับภาพ ดังนั้นการเปรียบเทียบของเสียงจะใช้การเปรียบเทียบเท่ากับ $6nm$ ครั้ง เช่นเดียวกับภาพ

การรวมกันของภาพและเสียงในการสืบค้นวีดีโอคลิป

เพราะฉะนั้นการเปรียบเทียบทั้งหมดเท่ากับ $6nm + 6nm = 12nm$ ครั้ง

4.5.2 ระบบที่ได้พัฒนาการสืบค้นภาพและเสียง

การสืบค้นภาพจะใช้สูตร Exclusive OR ดังนี้

$$X = A \oplus B$$

กรณีในการสืบค้น Exclusive OR

ขั้นที่ 1 เปรียบเทียบกับวีดีโอคลิปที่ 1 ใช้การเปรียบเทียบ n ครั้ง

ขั้นที่ 2 นำค่ามาบวกกัน n ครั้ง

การเปรียบเทียบเท่ากับ $2n$ ครั้ง

การสืบค้นเสียงจะใช้สูตรยูคลิดีเนียน (Euclidean Distance) ดังนี้

$$d(P, Q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

กรณีในการสืบค้น $(p_i - q_i)^2$

ขั้นที่ 1 นำค่ามาลบกัน ใช้การเปรียบเทียบ n ครั้ง

ขั้นที่ 2 ยกกำลัง 2 ใช้การเปรียบเทียบ n ครั้ง

กรณีในการสืบค้น $\sqrt{\sum_{i=1}^n (p_i - q_i)^2}$

ขั้นที่ 3 หาผลรวม ใช้การเปรียบเทียบ $n-1$ ครั้ง

ขั้นที่ 4 ถอดรากที่ 2 เปรียบเทียบกัน 1 ครั้ง

การเปรียบเทียบเท่ากับ $3n$ ครั้ง

เปรียบเทียบจนครบทั้งหมดในฐานข้อมูล ใช้การเปรียบเทียบ m ครั้ง

เพราะฉะนั้น การเปรียบเทียบทั้งหมดเท่ากับ $(2n+3n) \times m = 5nm$ ครั้ง

4.5.3 ผลการเปรียบเทียบในการดึงข้อมูลจากฐานข้อมูล

การดึงข้อมูลจากฐานข้อมูลมาเก็บค่าไว้ใน Array ซึ่งมี 2 วิธี

1. จากโปรเจกต์เคมของนายฉัตรชัย สุขสอน[3] ในการค้นหาไฟล์ภาพและเสียง

ในฐานข้อมูลมีขนาด $m \times n$ ซึ่งเก็บข้อมูลของภาพและเสียง เพราะฉะนั้นในการดึงเวกเตอร์

จากฐานข้อมูลต้องใช้การดึง $(m \times n) \times 2 = 2nm$ ครั้ง

2. จากโครงการที่ได้ทำ

ในฐานข้อมูลภาพมีขนาด $m \times 1$ และฐานข้อมูลเสียงมีขนาด $m \times n$ เพราะฉะนั้นในการดึงเวกเตอร์จากฐานข้อมูลต้องใช้การดึง $(m \times n) + m = nm + m$ ครั้ง

4.5.4 ความเร็วในการสืบค้นวีดิโอคลิป

การสืบค้นด้วยภาพและเสียงเปรียบเทียบของนายฉัตรชัย สุขสอน[3] เทียบกับการพัฒนาโครงการใหม่ ในการค้นหาไฟล์ภาพและเสียง

จำนวนครั้งที่ดีกว่าเท่ากับ $\frac{14nm}{(5n + n + 1)m}$ ครั้ง

โดยกำหนด n คือความยาวของเวกเตอร์

m คือข้อมูลทั้งหมด 6000 วิดีโอคลิป

บทที่ 5

บทสรุป

5.1 สรุปผลการทดลอง

โครงการที่ได้พัฒนาขึ้น เห็นได้ชัดว่าเวลาในการสืบค้น ไฟล์วีดีโอของโปรแกรม สามารถสืบค้นไฟล์ได้เร็วกว่า และขนาดของไฟล์ในฐานข้อมูลก็มีขนาดเล็กกว่าโครงการเดิมของนาย จัทรชัย สุขสอน[3] และการสืบค้นไฟล์ยังสามารถสืบค้นได้ 3 แบบ ซึ่งผลของการสืบค้นแต่ละวิธีมีค่าที่แตกต่างกันตาม %ความถูกต้อง

จากผลการทดลองบทที่ 4 การสืบค้นไฟล์วีดีโอคลิปโดยใช้วิธีการภาพและเสียงรวมกัน จะสามารถสืบค้นไฟล์วีดีโอคลิปที่มีความถูกต้องมากที่สุด จากการสืบค้นไฟล์วีดีโอทั้งสามแบบคือการสืบค้นด้วยภาพอย่างเดียว การสืบค้นด้วยเสียงอย่างเดียว และการสืบค้นไฟล์วีดีโอคลิปโดยใช้ภาพและเสียงรวมกัน แต่ในบางกรณีการสืบค้นด้วยเสียงเพียงอย่างเดียวอาจมีความถูกต้องในการสืบค้นได้มากกว่าเช่น การสืบค้นไฟล์วีดีโอคลิปประเภท มิวสิควีดีโอบางไฟล์ อาจสืบค้นได้ดีกว่าการสืบค้นไฟล์วีดีโอคลิปโดยใช้ภาพและเสียงรวมกัน ดูได้จากเส้นสีม่วงเทียบกับเส้นสีเหลืองในกราฟในรูป 4.4 ในบทที่ 4 ผลการทดลอง

เมื่อดูจากค่าเฉลี่ยของความถูกต้อง (Average Precision) จากช่องฟ้าเข้มของทั้งสามตารางในบทที่ 4 ในผลการทดลอง คือ ตาราง 4.1, 4.2, 4.3 แล้ว ก็จะสรุปได้ว่าการสืบค้นไฟล์วีดีโอแบบภาพและเสียงรวมกัน จะสามารถสืบค้นไฟล์วีดีโอได้มีความถูกต้องมากที่สุดเพราะมีค่าเฉลี่ยของความถูกต้อง (Average Precision) เท่ากับ 54.625% และรองลงมาคือการสืบค้นไฟล์วีดีโอด้วยเสียง มีค่าเฉลี่ยของความถูกต้อง (Average Precision) เท่ากับ 47.5625% และการสืบค้นไฟล์วีดีโอด้วยภาพจะมีค่าน้อยสุดมีค่าเฉลี่ยของความถูกต้อง (Average Precision) เท่ากับ 43.94%

จากการทดลองทั้งหมดจะเห็นได้ว่า การใช้การรวมกันของภาพและเสียงมีความถูกต้องในการสืบค้นวีดีโอคลิปมากกว่าการใช้ ภาพหรือเสียงเพียงอย่างเดียว แต่ถึงอย่างไรก็ตาม โครงการนี้ก็สามารถเลือกใช้การสืบค้นไฟล์วีดีโอแบบใดก็ได้ตามประเภทของคำสอบถาม(Query) เพื่อที่จะให้ผลออกมาได้ถูกต้องที่สุด

5.2 ปัญหาที่พบ

1. วีดีโอคลิปที่นำมาเปรียบเทียบบางวีดีโอคลิปอาจจะมีเพียงคลิปเดียว ที่เป็นคลิปโดดๆ ซึ่งไม่เหมือนคลิปอื่นที่อยู่ในฐานข้อมูล ดังนั้น โปรแกรมที่พัฒนาขึ้นมาอาจสืบค้นไฟล์วีดีโอคลิปไม่ถูกต้องเท่าที่ควร
2. เนื่องจากโครงนี้ได้เก็บข้อมูลของภาพในรูปแบบ เลขฐานสองเพื่อเพิ่มความเร็วในการสืบค้น จึงอาจทำให้ประสิทธิภาพความถูกต้องลดลงบ้าง

5.3 ข้อเสนอแนะ

1. เมื่อมีการพัฒนาโปรแกรมต่อ ควรที่เพิ่มส่วนที่รับข้อมูลจากภายนอกได้ด้วย ก็สามารถรับคลิปวีดีโอจากภายนอกมาเปรียบเทียบได้
2. เมื่อมีการพัฒนาโปรแกรมต่อ โปรแกรมการสืบค้นควรมีความแม่นยำมากกว่านี้
3. เมื่อมีการพัฒนาโปรแกรมต่อ ควรที่จะนำบางส่วนของคลิปมาทำการเปรียบเทียบกับวีดีโอคลิปทั้งหมดได้ คือการตัดช่วงหนึ่งของคลิปแล้วนำช่วงนั้นไปเปรียบเทียบกับวีดีโอคลิปทั้งหมด



เอกสารอ้างอิง

- [1] สุรสิทธิ์ คิวประสพศักดิ์. นันทน์ แวงโสภณ. อินไซท์ Visual Basic .NET ฉบับสมบูรณ์. บริษัท โปรวิชั่น จำกัด. 2546
- [2] ศุภชัย สมพานิช. **Advanced Database Programming ด้วย VB 6 + VB.NET. DEV BOOK.** 2548
- [3] ภัทรชัย สุขสอน. การสืบค้นไฟล์วิดีโอด้วยองค์ประกอบของภาพและเสียง (Video retrieval using visual and audio content). มหาวิทยาลัยนเรศวร. 2548
- [4] "The MathWorks." [Online]. Available: <http://www.mathworks.com/products/matlabx1>
- [5] "Microsoft Access." [Online]. Available: <http://www.ict.pyo.nu.ac.th/uthais>



ประวัติผู้เขียนโครงการ



ชื่อ นายชัชชา ชาวเหนือ

ภูมิลำเนา 86/26 หมู่ 2 ต.ท่าทอง อ.เมือง จ.พิษณุโลก 65000

ประวัติการศึกษา

- จบการศึกษาระดับมัธยมจาก โรงเรียนพะเยาพิทยาคม
- ปัจจุบันกำลังศึกษาในระดับปริญญาตรีชั้นปีที่ 4

สาขาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

มหาวิทยาลัยนเรศวร

E-mail: ohmchat@hotmail.com



ชื่อ นายสุทธิจักร เครือสาร

ภูมิลำเนา 169/2 หมู่ 4 ต.ต๋น อ.เมืองพะเยา จ.พะเยา

ประวัติการศึกษา

- จบการศึกษาระดับมัธยมจาก โรงเรียนพะเยาพิทยาคม
- ปัจจุบันกำลังศึกษาในระดับปริญญาตรีชั้นปีที่ 4

สาขาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

มหาวิทยาลัยนเรศวร

E-mail: fujinkn_zero@hotmail.com