A Study on the Improvement of Retrieval Efficiency Based on the CRFMD*

공통기술표현포맷에 기반한 다매체자료의 검색효율 향상에 관한 연구

박일종(Il-Jong Park)**, 정기태(Ki-Tai Jeong)***

ABSTRACT

In recent years, theories of image and sound analysis have been proposed to work with text retrieval systems and have progressed quickly with the rapid progress in data processing speeds. This study proposes a common representation format for multimedia documents (CRFMD) composed of both images and text to form a single data structure. It also shows that image classification of a given test set is dramatically improved when text features are encoded together with image features. CRFMD might be applicable to other areas of multimedia document retrieval and processing, such as medical image retrieval, World Wide Web searching, and museum collection retrieval.

초 록

최근 수년 동안 영상자료와 음성자료 분석에 대한 이론들이 텍스트자료 검색 시스템과 함께 사용되기 위해서 제안되어 왔으며 데이터 처리 속도의 급격한 향상과 함께 발전되어 왔다. 일반적 검색 방법들은 단지 텍스트만을 사용하지만 텍스트와 그림을 동시에 사용하는 검색 방법 또한 최근에 제안되어 왔다. 본 연구는 다매체자료의 공통기술표현포맷(CRFMD)이라는 이름으로 화상자료와 텍스트자료를 하나의 자료 구조로 통합하는 방법을 제안하고 있으며, 주어진 테스트자료에 대한 화상자료의 유사성 분석에서 텍스트와 그림의 형태소를 함께 사용하였을 때 현격히 개선되어 짐을 보여주고 있다. CRFMD는 의료문서 검색, WWW 검색, 박물관 소장품 검색과 같은 다양한 분야의 다매체자료 검색 및 처리에 응용될 수가 있을 것이다.

Keywords : Multimedia Document Retrieval, Common Representation Format for Multimedia Document (CRFMD), Precision, Recall, Text Retrieval System, Sound Analysis, Image Analysis
다매체자료 검색, 다매체자료 공통표현, 정확률, 재현율, 텍스트검색시스템, 음성자료분석, 화상자료분석

---

* 본 논문은 2003-2004학년도 저자의 계명대 연구년 논문의 일부로서 제출되기 위하여 작성되었음.
** 계명대학교 문헌정보학과 부교수 (ipark@kmu.ac.kr)
*** Assistant Professor, Computer Science Department, Lane College
■ Received : 7 August 2006
■ Accepted : 20 September 2006

1. Introduction

The World Wide Web produces an abundant amount of multimedia documents on a daily basis. Images collected from satellites are used for forecasting weather, tracking ecological changes, and providing information on changes in space (Demers 2000). Images are also used in medical diagnosis (Liu Y et al. 1998). Image analysis is a method for determining how images are analyzed and stored for future use, and image retrieval is used to describe any method used to retrieve images from storage. As massive repositories of images are created, the need for better image analysis and retrieval has grown and many theories of image analysis and retrieval have been proposed (Goodrum et al. 2000).

Text documents are easily understood and organized and have been used efficiently in information retrieval for many years (Korfhage 1997). The use of images in information retrieval is still in a nascent stage of development, and has been less successful than text retrieval. Although shape, color, and texture are undoubtedly important for image representation, there is little understanding of how best to analyze these attributes for actual image retrieval. Moreover, multiple images can represent the same event even if they are rotated, expanded, extracted, contracted, or colored. The focus of image retrieval research to date has been primarily on the use of features that can be computationally acquired from images ; little has been done to identify the visual attributes needed by users for various tasks or to characterize collections for searching (Goodrum et al. 2000 ; Jorgensen et al. 2001). No previous research has successfully combined text and image description and retrieval features into a single data structure (Rorvig et al. 2000).

Several multimedia document retrieval systems have been released by commercial vendors and proactive researchers, including CONVERA™, VIRAGE™, IBM's QBIC™, AMORE™, ARTISAN™, BlobWorld™, and CANDID™ (Venters & Cooper 2000). Each of these systems uses two separate data structures for multimedia documents : one for images and the other for text (Eakins & Graham 1999).

These two data structures must be maintained separately. Unfortunately, anomalies, which are defined as broken links, may exist between the two data structures (Eakins & Graham 1999), in which case these retrieval systems cannot use image data without linking first to the text data structure. In contrast, the common representation format for multimedia documents described in this paper employs a single data structure (Jeong et al. 2001). That combines both text and image data structures in order to retrieve images.

Also, the research project described here was designed to test two hypotheses :

1) A single data structure combining text measures and image measures is possible.
2) A combined representation format

improves the results of multimedia document retrieval.

2. FEATURES AND TESTING

2.1. Image Features

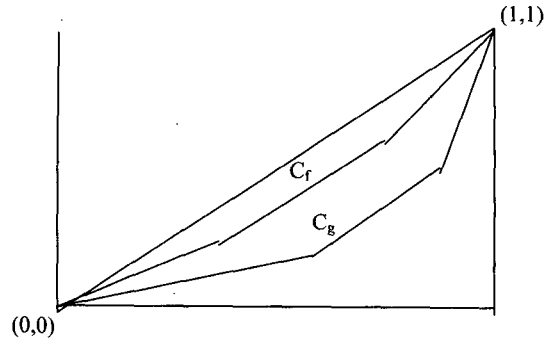
Content-based image retrieval is typically based on primitive image features that are used to extrapolate a meaning for limited image understanding and retrieval (Eakins & Graham 1999). Although many such primitive measures are available, there currently is not a set of optimal measures that leads to perfect retrieval. Previous research suggests that more measures tend to work better than fewer (Jeong et al. 2001).

For this study, twelve primitive features from images were extracted :

1. red
2. green
3. blue
4. gray
5. distance A (distance from the origin of the image to a specific pixel)
6. distance B (distance from side-A to a specific pixel)
7. distance C (distance from side-B to a specific pixel)
8. distance D (distance from side-C to a specific pixel)
9. distance E (distance from side-D to a specific pixel)
10. angle (angle from the origin of the image to a specific pixel)
11. Hough Transform value (representing a kind of distance) (Young 1993)
12. density of image

Analysis of image documents requires extraction of features that can be defined as essential document characteristics. This research project examined twelve primitive features-red, green, blue, gray, five distances, angle, the Hough Transform value, and density-extracted from images. These twelve features were chosen because they were easily extracted from images, and can be expressed as areas under the Lorenz Information Curve. Eleven features except density were transformed into histograms using the frequencies of each feature. For example, pixels for red can be valued as one of 0 thru 255, and extracted pixel values can be transformed into red histogram. Image density, which explains the textile quality of the image, is calculated from the number of edge-detected pixels.

The curve, derived from the histograms, is called the Lorenz Curve or the Lorenz Information Curve (Gastwirth 1971 ; Chang & Yang 1982). It can be seen that once the histogram h is given, the Lorenz Curve is completely specified as in 〈Figure 1〉. The curves C_f and C_g represent the Lorenz Curves. The Lorenz Information Measure (LIM) (Lorenz 1893 ; Chang & Yang 1982) $LIM(p_1,...,p_n)$ is defined as the area under the Lorenz Curve where the area is $0 \leq LIM(p_1,...,p_n) \leq 0.5$. For any probability vector $(p_1,...,p_n)$), $LIM(p_1,...,p_n)$) can be computed by first ordering the p_is from least to greatest, then calculating the area



〈Figure 1〉 Lorenz Curve (Chang & Yang 1982)

under the piecewise linear curve. Finally, the Lorenz Information Measure is the weighted sum of the Lorenz Curve (C_f or C_g), so that LIM can be regarded as a global measure of information content because each distinct height of the Lorenz Curve represents the amount of information content in the image.

2.2. Text Features

Numerous methods to extract data from text documents have been suggested since 1960 (Salton et al. 1994). Typical of these are a term frequency method and a binary representation method. The term frequency method uses the frequencies of each term in a document, while the binary representation method uses "1" or "0" for each term depending on the presence or absence of the term in a document. In this study, a binary representation of textual data was chosen because of its simplicity and ease of transformation. Text analysis for term extraction in this study was based on the concept of the codeword. This idea was originally developed by Liu (1977) for the correction of errors in information transmission. The codeword concept has been adapted for this study to include a term list, a binary matrix, a Word Code, and a process for generating the term list, binary matrices, and Word Codes. The term list represents the appearance of words in the text documents that comprise a dataset. For example, imagine a dataset containing two documents - document A and document B. Document A contains the expression "To be or not to be" and document B contains the expression "Life is to be happy." The term list for the two-document dataset is {to, be, or, not, life, is, happy}. The binary matrix for document A is {1,1,1,1,0,0,0}. The binary matrix for document B is {1,1,0,0,1,1,1}. This study differs from Liu's approach in that the binary matrix is partitioned into sub-groups and each sub-group is called a Word Code.

1750 terms were extracted from twenty-six text documents excluding 298 stop words and it was called as super term list.

Number-S88E5001 to S88E5002 2 Images
-Format- DIG. FILE
-Date- Photo Date : 12/04/98 Logbook Entry : 12/06/98
-Title-
Inflight maintenance of GIRA cable
-Keywords-
CABLES MAINTENANCE CREW PROCEDURES (INFLIGHT) MIDDECK ASTRONAUTS ONBOARD ACTIVITIES STS-88 ENDEAVOUR (ORBITER)
-Text-
Close-up view of the Galley Iodine Removal System (GIRA) being held by STS-88 mission specialist Jerry Ross on the Endeavour's middeck (5001). The cable was found to have a bent pin and an in-flight maintenace was performed on Flight Day 1. View of Ross holding the cable with Mission specialist James Newman in the background (5002).
-END-

〈Figure 2〉 Example of NASA image for S88E5001

Using this super term list, a document-term binary matrix was drawn as explained above. Among the twenty-six text documents, the smallest one had seven lines and the longest one had thirty lines ; the mode was fifteen lines. NASA employees wrote the images' descriptions, which were related to space work. A sample image and accompanying text document of the image is presented in 〈Figure 2〉.

Twelve text features were derived to create an equal number of text and image features. The decision to identify twelve text features was a conscious effort to achieve matching numbers of image and text features and does not necessary represent an optimal number of features. To get twelve text features the binary matrix will be divided evenly into eleven groups called the Word Code and the number of 1s will be counted for each group. If the binary matrix is not divided evenly into eleven groups, then the 11^{th} group will have the remainder. The 12^{th} group represents the total number of 1s in that document, and explains the textile quality of the text document like density of image explains the textile quality of the image. Each group matrix was normalized in order for each result to be less than or equal to 0.5 because LIM measurements are less than or equal to 0.5. The procedures to get twelve measurements for text documents are named as Jeong' s transform.

The formula for normalization of a group is this :

Normalized value = Group weight / (The highest value of that group * 2)

Using the above transform, twelve text measurements are made.

The twelve measurements from the image and the twelve measurements from the text were combined in linear to construct a common representation format for multimedia documents. The first hypothesis - that it is possible to derive a single data structure to represent both text and image features - was supported by

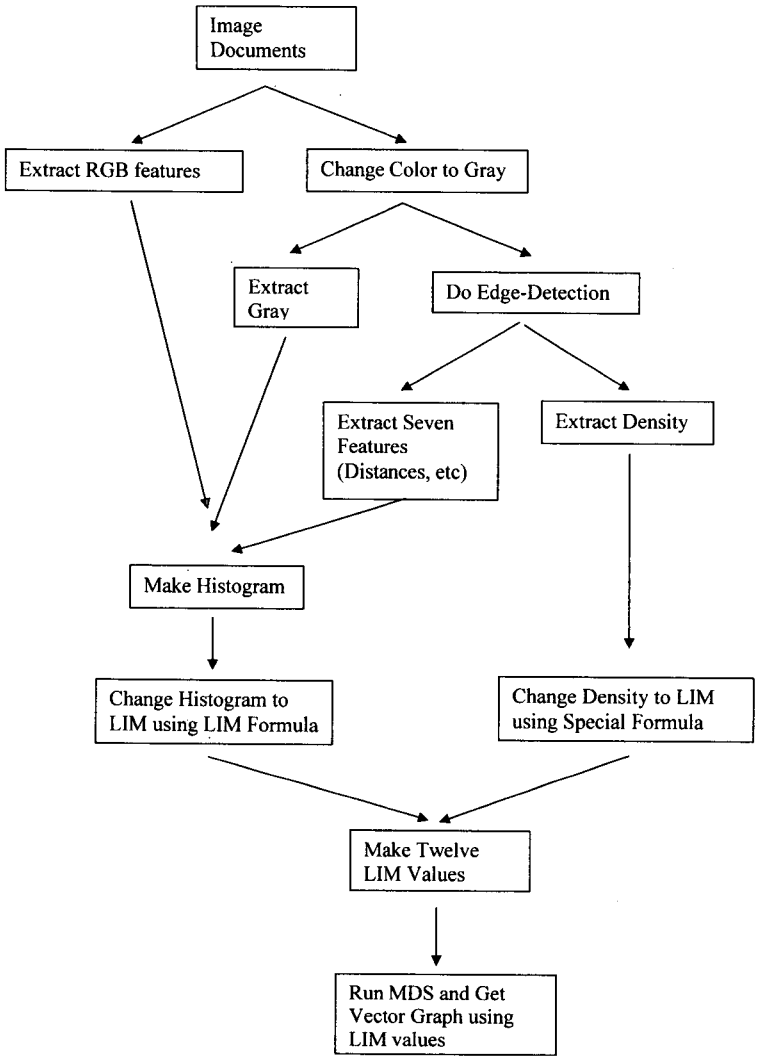
〈Table 1〉 A Similarity Table on 26 Images by Human Heuristics

	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6
1	x		x			x		x	x												x					
2		x																								
3	x		x			x		x	x												x					
4				x	x																					
5				x	x																					
6	x		x			x		x	x												x					
7							x																			
8	x		x			x		x	x												x					
9	x		x			x		x	x												x					
10										x	x	x		x												
11										x	x	x		x												
12										x	x	x		x												
13													x													
14										x	x	x		x												
15															x	x										
16															x	x										
17																	x	x	x	x						
18																	x	x	x	x						
19																	x	x	x	x						
20																	x	x	x	x						
21	x		x			x		x	x												x					
22																						x	x	x		x
23																						x	x	x		x
24																						x	x	x		x
25																									x	
26																						x	x	x		x

combining twelve text measurements and twelve image measurements in linear.

2.3. Testing Precision and Recall

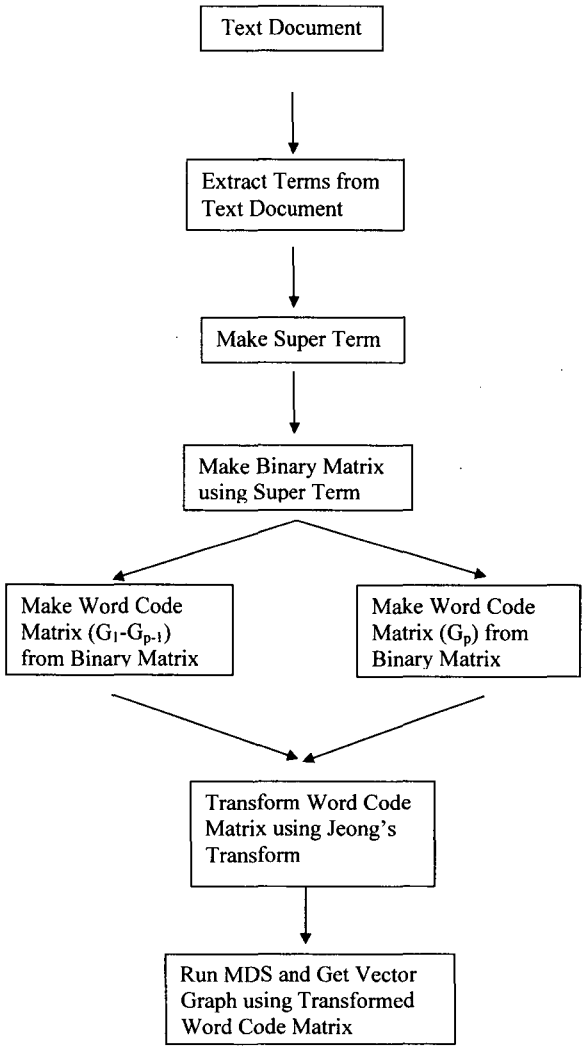
To compute precision and recall, the twenty-six images were first presented to two human testers who judged the images for their similarity. This was done by comparisons each of the images to all other images in the dataset and assigning a binary value of similar. No criteria or content for similarity were given to the testers. 〈Table 1〉 presents the results of



〈Figure 4〉 Flowchart for Image Document Process

their judgments.

In case of disagreement between the two testers, the author acted as an arbitrator. Fortunately, there was only one disagreement, and the author resolved the disagreement. The twenty-six images were grouped in six homogeneous groups, but the images were quite different between the groups. The images are numbered from one to twenty-six and the numbers in rows and columns of 〈table 1〉 represent image numbers. 〈table 1〉 shows the similarities between images. For example, image number 1 in row has marked in image number 1, 3, 6, 8, 9 and 21. These six images were decided to be very similar



〈Figure 5〉 Flowchart for Text Document Process

by two human testers. If any one of these six images is used as a query for image retrieval, a human would expect that these six images should be retrieved. They examined twenty-six images, judged the similarities, and recorded their judgments in 〈table 1〉

The second hypothesis - that a combined representation format will improve the results of multimedia retrieval - was tested using a parallel comparison of precision and recall over retrieved results. Testing was carried out on a dataset of twenty-six multimedia documents drawn from a NASA database of annotated images.

Histograms for image features were generated for each image in the dataset. These twelve image measurements were used for an image-only retrieval system ; the procedure for generating these measurements is presented graphically in 〈Figure 4〉.

From the textual descriptions accompanying each image, twelve text measurements were captured using the procedure presented in 〈Figure 5〉. The twelve image measurements from the twenty-six image documents were then combined with the twelve text measurements from the accompanying text documents in linear. These twenty-four measurements constitute a common representation format for multimedia documents that is used to define a combined retrieval system.

Two retrieval systems were constructed. The image-only retrieval system (http://slis-linux.labnet.ou.edu/tdlimage/) uses the twelve image measurements to retrieve images ; the combined retrieval system (http://slis-linux.labnet.ou.edu/tdlcomb/) uses the twenty-four combined measurements derived from image and text features to retrieve images. Text-only retrieval system was not tested because it was not the focus of this research. The Brighton Image Searcher (http://slis-linux.labnet.ou.edu/tdlimage/, http://slis-linux.labnet.ou.edu/tdlcomb/) was used to construct searches and it starts with text.

The term, "Space", was chosen arbitrarily from the 1750 terms in the dataset, then both Retrieval Systems yield images linked to the given text term. From the retrieved images, the tester chose an image. The chosen image is then used as a query to retrieve similar images from the multimedia document set. Finally, to calculate precision and recall, the two testers who were graduate students evaluated the retrieved images based on 〈Table 1〉.

3. Findings and Analysis (Findings through test set analysis)

Precision and Recall based on Heuristic Judgment

For this study, four test cases based on two testing methods were conducted. In case 1, the image-only retrieval system, a search was initiated using a keyword, "Space". Images associated with that keyword were displayed and the searcher selected an image one by one among retrieved images, then executed the "Find Similar Image" search option with a threshold of 0.2 and the selected image was used as a query. The threshold value controls the sensitivity of the search ; a larger threshold value can be expected to produce a greater number of retrieved documents than a smaller threshold value. The system then used the twelve image features, which were coded in HTML using 〈META〉 tag with feature name and value, to generate a set of image documents. Precision and recall were calculated manually based on searcher assessment of

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html ; charset=window-1252">
<meta name=="GENERATOR" content="Microsoft FrontPage 4.0">
<meta name="ProgID" content="FrontPage.Editor.Document">
<meta name="Red" content="0.4503067">
<meta name="Green" content="0.4424329">
<meta name="Blue" content="0.4863451">
<meta name="Grey" content="0.4685448">
<meta name="Hough Transform" content="0.3573107">
<meta name="Line length" content="0.2545683">
<meta name="Angles" content="0.4978859">
<meta name="Right Side Distance from Origin" content="0.496774">
<meta name="Top Side Distance from Origin" content="0.4981645">
<meta name="Left Side Distance from Origin" content="0.4979944">
<meta name="Bottom Side Distance from Origin" content="0.4981703">
<meta name="Density" content="0.1523178">
<meta name="Group1" content="0.4615385">
<meta name="Group2" content="0.4583333">
<meta name="Group3" content="0.1363636">
<meta name="Group4" content="0.3846154">
<meta name="Group5" content="0.3000000">
<meta name="Group6" content="0.3181818">
<meta name="Group7" content="0.2000000">
<meta name="Group8" content="0.1666667">
<meta name="Group9" content="0.5000000">
<meta name="Group10" content="0.2222222">
<meta name="Group11" content="0.3461538">
<meta name="Group12" content="0.3515625">
<title>NASA IMAGE AND TEXT MEASUREMENTS</title>
</head>
```

〈Figure 7〉 HTML format using 〈META〉 tag for twenty-four measurements

the retrieved set of image documents.

In case 2, the combined retrieval system, images were retrieved using the same process as used in the image-only system. The combined retrieval system generated a retrieved document set based on combined features of the twelve image features and the twelve text features in linear.

Case 3 and case 4 repeated the process using a threshold value of 0.1 to determine variations in precision and recall based on the sensitivity of the test. In another words, testing of the system relies on a single query posted to two systems with two thresholds set.

Image-Only Retrieval, Threshold = 0.2

〈Table 2〉 shows the retrieved results for the twenty-six images with a threshold of 0.2 (case 1). All of the test searches produced very large retrieved sets. When image number 1 was selected as a query image, for instance, twenty-five images were retrieved (all of the images except image number 10). These results imply that image-only measurements are not effective when used alone.

〈Table 2〉 Retrived Results on 26 Images-only under Threshold 0.2

	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6
1	x	x	x	x	x	X	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
2	x	x	x	x	x	X	x	x	x		x		x	x	x	x	x	x	x	x	x	x	x	x	x	x
3	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
4	x	x	x	x	x	X	x	x	x		x		x		x	x	x	x	x	x	x	x	x	x	x	x
5	x	x	x	x	x	X	x	x	x		x		x		x	x	x	x	x	x	x	x	x	x	x	x
6	x	x	x	x	x	X	x	x	x		x		x	x	x	x	x	x	x	x	x	x	x	x	x	x
7	x	x	x	x	x	X	x	x	x		x		x	x	x	x	x	x	x	x	x	x	x	x	x	x
8	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
9	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
10										x	x	x	x	x												
11	x	x		x	x	X	x			x	x	x	x	x												
12	x									x	x	x	x	x												
13	x	x	x	x	x	X	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
14	x	x				X	x			x	x	x	x	x												
15	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
16	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
17	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
18	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
19	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
20	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
21	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
22	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
23	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
24	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
25	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x
26	x	x	x	x	x	X	x	x	x				x		x	x	x	x	x	x	x	x	x	x	x	x

A contingency table 〈Table 3〉 was calculated from 〈Table 1〉 and 〈Table 2〉 and used to calculate precision and recall. Precision and recall under the 0.2 threshold were calculated using the formula :

Precision = (Retrieved and Relevant) / ((Retrieved and Relevant) + (Retrieved but not Relevant))

P = 96 / 530
= 0.181132
= 18.11 %.

Recall = (Retrieved and Relevant) /

〈Table 3〉 Contingency Table for 26 Images-only under Threshold 0.2

Image Number	Retrieved and Relevant	Retrieved but not Relevant	Not Retrieved but Relevant
1	6	19	0
2	1	23	0
3	6	16	0
4	2	21	0
5	2	21	0
6	6	18	0
7	1	23	0
8	6	16	0
9	6	16	0
10	4	1	0
11	4	7	0
12	4	2	0
13	1	25	0
14	4	5	0
15	2	20	0
16	2	20	0
17	4	18	0
18	4	18	0
19	4	18	0
20	4	18	0
21	6	16	0
22	4	18	0
23	4	18	0
24	4	18	0
25	1	21	0
26	4	18	0
Total	96	434	0

((Retrieved and Relevant) + (Not Retrieved but Relevant))

R = 96 / 96
= 1
= 100 %.

Image-and-Text Combined Retrieval, Threshold = 0.2

〈Table 8〉 shows the retrieved results for the twenty-six image-and-text combined measurements with a threshold of 0.2 (case 2). Retrieved sets were much smaller

〈Table 4〉 Retrieved Results on 26 Images-and-Text Combined under Threshold 0.2

	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6
1	x	x	x	x	x	x																				
2	x	x	x	x	x	x																				
3	x	x	x	x	x	x																				
4	x	x	x	x	x	x																				
5	x	x	x	x	x	x																				
6	x	x	x	x	x	x																				
7							x	x	x																	
8							x	x	x																	
9							x	x	x																	
10										x	x	x		x												
11										x	x	x	x	x												
12										x	x	x	x	x												
13											x	x	x	x												
14										x	x	x	x	x												
15															x	x										
16															x	x										
17																	x	x	x	x						
18																	x	x	x	x						
19																	x	x	x	x						
20																	x	x	x	x						
21																					x					
22																						x	x	x		
23																						x	x	x		
24																						x	x	x		
25																									x	
26																										x

than those generated by the first case. When image number 1 was used as a query image, for instance, six images (1,2,3,4,5,6) were retrieved. Among six images, three images (1,3,6) were judged as similar images according to Table 1. In other words, 50% of the retrieved images were similar. These are very positive results.

The contingency table for case 2 is shown in 〈Table 5〉. Precision and Recall for the twenty-six image-and-text combined documents under the threshold 0.2 were calculated using the formula :

〈Table 5〉 Contingency Table for 26 Images-and-Text Combined under Threshold 0.2

Image Number	Retrieved and Relevant	Retrieved but not Relevant	Not Retrieved but Relevant
1	3	3	3
2	1	5	0
3	3	3	3
4	2	4	0
5	2	4	0
6	3	3	3
7	1	2	0
8	2	1	4
9	2	1	4
10	4	0	0
11	4	1	0
12	4	1	0
13	1	3	0
14	4	1	0
15	2	0	0
16	2	0	0
17	4	0	0
18	4	0	0
19	4	0	0
20	4	0	0
21	1	0	5
22	3	0	1
23	3	0	1
24	3	0	1
25	1	0	0
26	1	0	3
Total	68	32	28

Precision = 68 / 100
= 0.68
= 68 %.

Recall = 68 / 96
= 0.7083333
= 70.83 %.

According to this calculation, it seems like to return high precision and high recall. The reason will be that this research is testing a small set of data which is a

〈Table 6〉 The Evaluation Results of Precision and Recall over Multimedia Testing Set

	Image Only				Image and Text Combined			
	Threshold 0.2		Threshold 0.1		Threshold 0.2		Threshold 0.1	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
26 multimedia documents	0.18	1.00	0.21	0.98	0.68	0.70	0.71	0.68

homogeneous test set.

Image-Only Retrieval, Threshold = 0.1

The table which could show the retrieved results for the twenty-six with a threshold of 0.1 (case 3) was omitted owing to the generally limited amount of journal articles. Retrieved sets are still relatively large. When image number 1 was selected as a query image, for instance, fourteen images were retrieved. Selecting a more sensitive threshold, then, did not improve search results.

The contingency table for case 3 was also omitted here because of the same reason. Precision and Recall for the twenty-six images-only with threshold of 0.1 were calculated using the formula :

Precision = 95 / 452
= 0.2101769
=21.02 %.

Recall = 95 / 96
= 0.9895833
= 98.96 %.

Image-and-Text Combined Retrieval, Threshold = 0,1

The table which could show the retrieved results for the twenty-six images-and-text combined measurements with a threshold of 0.1 (case 4) was omitted here owing to the generally limited amount of journal articles. Also, the contingency table is omitted because of the same reason. Precision and Recall for the twenty-six images-and-text combined under the threshold 0.1 were calculated using the formula :

Precision = 66 / 92
= 0.7173913
= 71.74 %.

Recall = 66 / 96
= 0.68.75
= 68.75 %.

Variations in Precision and Recall

Precision and recall are to some extent depending on the sensitivity of the threshold value. Setting a more sensitive threshold value reduced recall and increased precision for both the image-only retrieval system and the image-and-text retrieval system. This trend demonstrates the inverse

relationship between precision and recall, which is true for any set (Korfhage 1997). According to these statistics, precision for a threshold of 0.2 is improved by about 275% when the combined representation format is used rather than the image?only format. Precision is improved by about 241% when a threshold of 0.1 is used. The conclusion can be made from these statistics that precision may be improved when a combined format for multimedia documents is used. 〈Table 6〉 shows the evaluation results of precision and recall over multimedia testing sets.

4. Discussion and Conclusion

In processing text documents, a binary matrix made from text document dataset is often used. In this study the binary matrix itself was transformed into twelve measurements using Jeong' s Transform and two different types of documents (text documents and image documents) were combined into one data structure to form A Common Representation Format for Multimedia Documents (CRFMD).

The results for precision and recall support the hypothesis that "the CRFMD would improve the results of multimedia document retrieval comparing to the image-only representation format for multimedia documents." The test results show 275% improvement when a threshold 0f 0.2 is used and 241% improvement when a threshold of 0.1 is used.

Even though two hypotheses were supported, the optimal number of features to include in a CRFMD is subject to exploration. It seems axiomatic that the greater the number of image and text features, the greater the precision of a search of a CRFMD system.

For this research, a small set of multimedia documents which was homogeneous (twenty-six image and text documents) used, but the test results were very promising. As a continuing project a medium set of multimedia documents (994) is undergoing testing to verify the high precision and high recall. After that a large set of multimedia documents (probably more than 100,000) will be tested. In addition, heterogeneous test cases will be tested.

Some fields, such as museums, libraries, Geographic Information System, and Web searching, use an image retrieval system with different data structure and approach it at different angles. However, it will be a very challenging research to combine this research with the above areas.

References

Chang, S.K. and Yang, C.C. 1982. "Picture Information Measures for Similarity Retrieval." *Computer Vision, Graphics, and Image Processing* 23, 366-375.

Demers, M.N. 2000. *Fundamentals of*

Geographic Information Systems. 2nd ed. New York : John Wiley & Sons.

Eakins, J.P. and Graham, M.E. 1999. "Content-based Image Retrieval : a report to the JISC technology applications programme." *Institute for Image Data Research*, University of Northumbria at Newcastle.

Gastwirth, J.L. 2001. "A General Definition of the Lorenz Curve." *Econometrica 39* : 1037-1039, November 6.

Goodrum, Abby A., Mark Rorvig, Ki Tai Jeong and C. Suresh. 2001. "An Open Source Agenda for Research Linking Text and Image Content Features." *Journal of the American Society for Information & Technology*, 52(11), 948 - 953.

Jeong, K., Rorvig, M., Jeon, J. and Weng, N. 2001. "Image Retrieval by Content Measure Metadata Coding." *CIR 2001, Tenth International World Wide Web Conference*, Hong Kong.

Jorgensen, C., Jaimes, A, Benitez, A, and Chang, S. 2001. "A Conceptual Framework and Empirical Research for Classifying Visual Descriptors." *Journal of the American Society for Information Science & Technology*, 52 : 938-947.

Korfhage, R.R. 1997. *Information Storage and Retrieval.* Wiley Computer Publishing.

Liu, C.L. 1977. *Elements of Discrete Mathematics.* McGraw-Hill, Inc. 225-228.

Liu, Y. et al. 1998. "Content-based 3-D Neuroradiologic Image Retrieval : Preliminary Results." *IEEE Int'l Workshop on Content-based Access of Image and Video Databases (CAIVD'98)*, Bombay, India, 91-100.

Lorenz, M.O. 1893. "Methods of measuring the Concentration of Wealth." *J. of Amer. Statist. Assoc.* 9 : 209-219

Park, Il-Jong K. 1997. "Comparing Major US OPAC Systems for Developing Countries." Libri : Int'l J. of Lib. and Info. Services. 47(4) ; 234-242.

Rorvig, M. and Jeong, K. 2000. "A Common Representation Format for Multimedia Documents." *Texas Center for Digital Knowledge*, SLIS, U. of N. Texas : Denton, TX.

Salton, Gerard, and James Allen. 2004. "Text Retrieval using the Vector Processing Model." *Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval*, Las Vegas, Nevada, 9-22.

Venters, C.C. and Cooper, M.D. 2000. "A Review of Content-Based Image Retrieval Systems." *The Joint Information Systems Committee.* March.

Young, David. 1993. "Hough Transform." *Sussex Computer Vision*, http://www.cogs.susx.ac.uk/users/davidy/teachvision/vision4.html