A COMMON REPRESENTATION FOR MULTIMEDIA DOCUMENTS

Ki Tai Jeong, B.A., M.S.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2002

APPROVED:

Mark E. Rorvig, Major Professor
Brian C. O'Connor, Committee Member and Coordinator of the program in Information Science
Min He Ji, Committee Member
Samantha Hastings, Committee Member
Philip M. Turner, Dean of the School of Library & Information Sciences
C. Neal Tate, Dean of the Robert B. Toulouse School of Graduate Studies

Jeong, Ki Tai, A Common Representation Format for Multimedia Documents. Doctor of Philosophy (Information Science), December 2002, 113 pp., 41 tables, 45 illustrations, references, 60 titles.

Multimedia documents are composed of multiple file format combinations, such as image and text, image and sound, or image, text and sound. The type of multimedia document determines the form of analysis for knowledge architecture design and retrieval methods. Over the last few decades, theories of text analysis have been proposed and applied effectively. In recent years, theories of image and sound analysis have been proposed to work with text retrieval systems and progressed quickly due in part to rapid progress in computer processing speed. Retrieval of multimedia documents formerly was divided into the categories of image and text, and image and sound. While standard retrieval process begins from text only, methods are developing that allow the retrieval process to be accomplished simultaneously using text and image.

Although image processing for feature extraction and text processing for term extractions are well understood, there are no prior methods that can combine these two features into a single data structure. This dissertation will introduce a common representation format for multimedia documents (CRFMD) composed of both images and text.

For image and text analysis, two techniques are used: the Lorenz Information Measurement and the Word Code. A new process named Jeong's Transform is demonstrated for extraction of text and image features, combining the two previous

measurements to form a single data structure. Finally, this single data structure is analyzed by using multi-dimensional scaling. This allows multimedia objects to be represented on a two-dimensional graph as vectors. The distance between vectors represents the magnitude of the difference between multimedia documents.

This study shows that image classification on a given test set is dramatically improved when text features are encoded together with image features. This effect appears to hold true even when the available text is diffused and is not uniform with the image features. This retrieval system works by representing a multimedia document as a single data structure. CRFMD is applicable to other areas of multimedia document retrieval and processing, such as medical image retrieval, World Wide Web searching, and museum collection retrieval.

ACKNOWLEDGMENTS

Without the encouragement of the late Dr. Mark E. Rorvig this dissertation could not be finished. I want to dedicate this dissertation to the late Dr. Mark E. Rorvig and thank Intel Co., who funded this research for two years. In addition I would like to acknowledge everyone who helped me continue my scholarly career, especially my wife Guja Jeong and my mother. Thanks God to make me achieve this accomplishment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Figure Page

CHAPTER 1. INTRODUCTION

Problem Description

The World Wide Web produces an abundant amount of multimedia documents on a daily basis. For example, images collected through satellite are used for forecasting weather, tracking ecological changes, and providing information on changes in space (Demers, 1999). Images are also used in the medical field to determine diagnosis (Liu, 1998). Image analyses are methods how images can be analyzed for retrieving and storing purpose. Image retrieval is a method how images can be retrieved from the storage for user's need. As massive repositories of images are created for such countless purposes, the need for better image analysis has grown and many theories of image analysis have been proposed (Goodrum et al., 2000).

However, use of images in information retrieval 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. In addition, two images can represent the same situation even if they are rotated, expanded, extracted, contracted, or colored. The text itself is easily understood and organized and has been used efficiently in information retrieval (Korfhage, 1997). The focus to date has been primarily on the use of features that can be computationally acquired from images, but little has been done to identify the visual attributes needed by users for various tasks and collections for searching (Goodrum et al., 2000).

Nevertheless, previously no method existed to combine these two features into a single data structure (Rorvig et al., 2000).

Objectives of the Study

Several multimedia document retrieval systems have been released by commercial vendors and proactive researchers, such as CONVERA™, VIRAGE™, IBM's QBIC™, AMORE™, ARTISAN™, BlobWorld™, and CANDID™ (Venters & Cooper, 1999). However, these systems have two data structures for multimedia documents: one for image, and the other for text (Eakins & Graham, 1999); therefore, for retrieval purposes, these systems need to establish a link between image and text (Venters & Cooper, 1999).

These two data structures used by commercial vendors must be maintained separately. Unfortunately anomalies could exist between the two data structures (Eakins & Graham, 1999). In that case the retrieval systems cannot use image data without linking text data structure to image data structure. However, the common representation format for multimedia documents explained in this paper has a single data structure (Jeong et al., 2001). The single combined data structure can solve the anomaly problem easily because the two data structures are combined. Not only can text terms be used to retrieve images, but also a text data structure can be used with an image data structure in retrieving images.

Research Questions

Content-based image retrieval uses several primitive image features. Features such as lines, edges, angles, grayscale, red, green, and blue color scale, pattern matching, and spatial proximity are used to extrapolate a meaning for limited image understanding and retrieval (Eakins & Graham, 1999). Although many such primitive measures are available, there is not yet a small set of optimal measures that

leads to perfect retrieval. Rather, it seems that more measures tend to work better than fewer (Jeong et al., 2001).

For this reason, twelve primitive features from images were extracted, namely red, green, blue, gray, distance A (distance from the origin of the image to a specific pixel), distance B (distance from side-A to a specific pixel), distance C (distance from side-B to a specific pixel), distance D (distance from side-C to a specific pixel), distance E (distance from side-D to a specific pixel), angle, Hough Transform value which is representing a kind of distance (Young, 1993), and density of image. All these features will be explained in chapter 4 in great detail.

Numerous methods to extract data from text documents have been known since 1960 (Salton et al., 1994). Typical of these are a term frequency method and a binary representation method. In this research, binary representation of textual data was chosen because of its simplicity and ease of transformation. Term frequency method uses the frequencies of each term in a document, but binary representation method uses "1" or "0" for each term depending on presence or absence of term in a document. A term extraction from the text document is the best-known way of making a binary representation through using a text representation method. Term extraction is a method of extracting words in a text document. This term set consists of terms from the entire document set. Extracted terms can then be represented in a binary representation for each document. A binary expression comes from a term list that is made from the entire document set using the term extraction method chosen for this research. The term list is used as a base, and a binary matrix is produced for representing each document and each term.

The binary matrix will be divided evenly into 11 groups called Word Code if that is possible and number of 1's is counted for each group. If the binary matrix is not divided evenly into 11 groups, then the 11^{th} group will have the rest of them. The 12^{th} group represents the total number of 1's in that document. After finding the largest frequency for each word group, the frequency of that group for each document is divided by the largest frequency for that group multiplied by 2. This is done so that the measurements for text document are on the same scale as measurements for image features. Using the above transform, twelve text measurements are made. Then, twelve measurements from image and twelve measurements from text are combined to construct a common representation format for multimedia documents. Constructing a common representation format for multimedia documents is easily accomplished through this method because it makes it possible for two totally different media formats to have a single data structure.

These are the hypotheses of the research:

1) A single data structure combining text measures and image measures is possible.

2) A combined representation format significantly improves the results of multimedia document retrieval.

CHAPTER 2. THEORIES AND RESEARCH ON THE NATURE OF CONTENT-BASED IMAGE RETRIEVAL

Introduction

A Brief History of Text Retrieval

Since the 1940s the problem of information storage and retrieval has attracted increasing attention for use with text documents. An example of an information storage and retrieval system is the MEDLINE™ system (McCarn & Leither, 1973) for on-line retrieval of medical information. This illustration shows by means of a black box what a typical information retrieval system looks like.

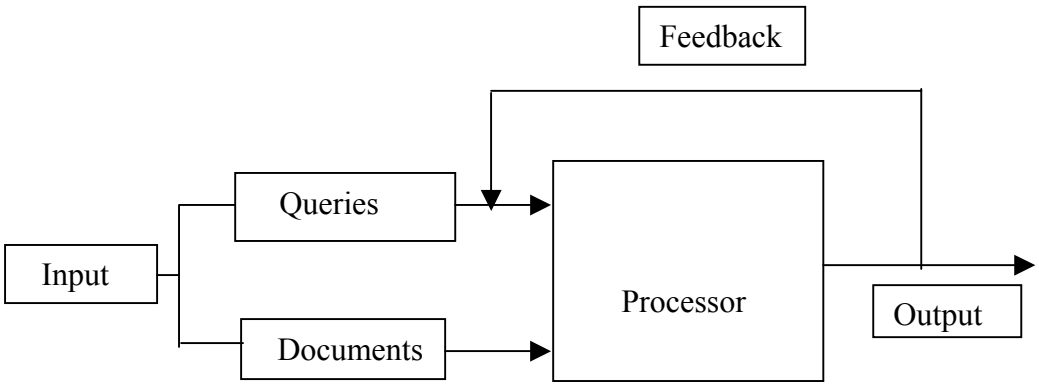

Figure 2.1 A Typical Information Retrieval System (Korfhage, 1997)

The diagram shows three components: input, processor and output. Feedback is a part of the processor. Although information retrieval can be divided in many ways, it seems that there are three main areas of research: content analysis, information structures, and evaluation. Briefly the first is concerned with describing the contents of text documents in a form suitable for computer processing; the second

with exploiting relationships between documents to improve the efficiency and effectiveness of retrieval strategies; the third with the measurement of the effectiveness of retrieval. Efficiency is usually measured in terms of the computer resources used such as central processing unit time, back-up time, and round-about time. Effectiveness of retrieval is also measured in terms of precision and recall measures.

For document representation Luhn (1957) used frequency counts of words in the document text to determine which words were sufficiently significant to represent or characterize the document in the computer. Thus a list of what might be called 'keyword' or 'term' was derived for each document. The use of statistical information about distribution of words in documents was further exploited by Maron and Kuhns (1960) who obtained statistical associations between keywords.

The term information structure covers specifically a logical organization of information, such as document representatives, for the purpose of information retrieval. The development in information structures has been fairly recent. The earlier experiments with text document retrieval systems usually adopted a serial file organization. More recently experiments have used clustered files for on-line retrieval. The organization of these files is produced by an automatic classification method. Good (1958) and Fairthorne (1961) were among the first to suggest that automatic classification might prove useful in document retrieval.

Evaluation of retrieval systems has proved extremely difficult. Senko (1969) in an excellent survey paper states: "Without a doubt, system evaluation is the most troublesome in information retrieval system ..." However, Lesk and Salton (1969) subsequently used a dichotomous scale on which a document is either relevant or non-

relevant, when subjected to a certain probability of error. They showed that this modification did not invalidate the results obtained for evaluation in terms of precision (the proportion of retrieved documents which are relevant) and recall (the proportion of relevant documents retrieved). Precision and Recall have a long history dating back to the Cranfield experiments (Korfhage, 1997) in the late 1950s.

A Brief History of Image Retrieval

With the rapid growth of digital technology in the last few years the potential use of digital images has increased enormously. Researchers in many professional fields are exploiting the opportunities offered by the ability to access and manipulate images in all kinds of new and exciting ways. Those explored in the early 1990s at National Aeronautics and Space Administration (NASA) (Rorvig, 1993; Rorvig et al., 1993), QBIC™ (Flickner et al., 1995) and VIRAGE™ (Gupta et al., 1996) are fairly indicative of the types of approaches available today.

Problems with past methods of image indexing (Enser, 1995) have led to the rise of interest in techniques for retrieving images on the basis of automatically-derived features such as color, texture and shape – a technology now generally referred to as content-based image retrieval (CBIR). To analyze and retrieve digital images, CBIR theories were created and are now being used in the market place worldwide (Eakins & Graham,1999; Ventors & Cooper, 1999) in the form of commercial products like QBIC (Flickner et al., 1995) and Virage (Gupta et al., 1996). CBIR has advantages and disadvantages with regard to the inherent nature of image analysis and retrieval.

Advantages:

1) Easy to extract features from image

2) Able to change extracted features to other forms such as histogram
3) Easy to build an automatic process

Disadvantages:

1) Hard to determine effectiveness
2) Unknown usability in handling real-life images
3) Difficult to choose features for extraction
4) Hard to get the semantic meaning of image from low level features
5) Difficult to process a specific region in the image
6) Limited markets of profit for CBIR (Goodrum et al., 2000; Eakins & Graham, 1999; Venters & Cooper, 1999)

Despite of the advantages and disadvantages CBIR is the most vigorously used image processing technique currently (Eakins & Graham, 1999; Venters & Cooper, 1999).

Pattern matching technique is another favorite technique in CBIR (Nadler and Smith, 1992). User's image needs may occur at a primitive level that taps directly into the visual attributes of an image, in which case the accompanying text would not be relevant. These attributes may best be presented by image exemplars and retrieved by systems performing pattern matches based on color, texture, shape, and other visual features (Hermes, 1995).

A Brief Explanation of Main Concepts Used in the Paper

To analyze image documents certain features have to be extracted. Eleven primitive features such as red, green, blue, gray, 5 distances, angle and Hough Transform, extracted from images could be transformed into histograms using frequencies in each feature. The density of image is calculated from the number of

edge-detected pixel. 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 like the Figure 2.2. The curve 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 to be the area under the Lorenz Curve. Clearly, $0 \leq LIM(p_1,...,p_n) \leq 0.5$. For any probability vector $(p_1,...,p_n)$, $LIM(p_1,...,p_n)$ can be computed by first ordering the p_i's in order from least to greatest, then calculating the area under the piecewise linear curve. Finally, the Lorenz Information Measure is the weighted sum of the Lorenz Curve (for example, 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.

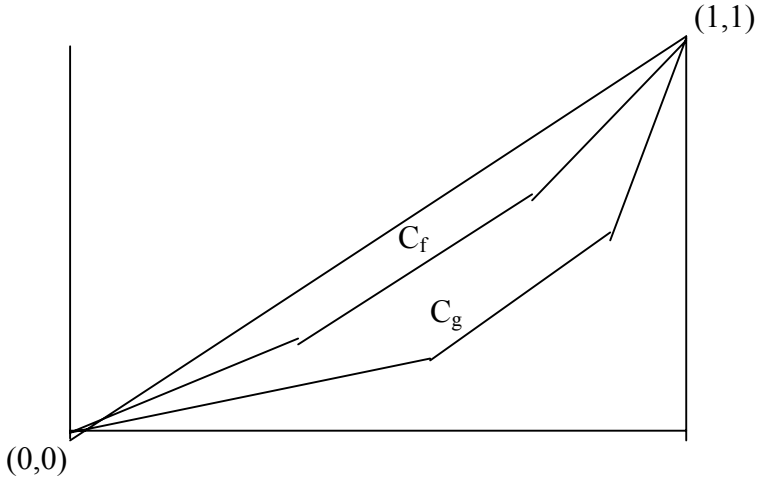

Figure 2.2 Lorenz Curve (Chang & Yang, 1982)

To analyze text documents the concept of codeword is adopted. Originally, the concept of codeword was described by Liu (1977). According to his explanation

the alphabet is the binary alphabet {0,1} and a sequence of letters from an alphabet is often referred to as a *word*. A *code* is a collection of words that are to be used to represent distinct messages. A word in a code is also called a *codeword*. For example, let x = 00101 and y = 10110, then x and y are codes and "0" and "1" are word or codeword. Let $\oplus$ be a binary operation, then x $\oplus$ y is a sequence of length n that has 1s in those positions x and y differ and has 0s in those positions x and y are the same. For the above example, x $\oplus$ y produces 10011. Originally, this idea was developed for the correction of error in information transmission. Even though the idea came from Liu (1977) it has been changed somewhat for this research.

The most important changes are a term list, a binary matrix, and the process to make them. Term list represents the appearance of words in the text documents and binary matrix is a table of term list for each document. For example, document A has "To be or not to be" and document B has "Life is to be happy". Then term list should be {to, be, or, not, life, is, happy} and binary matrix for document A should be {1,1,1,1,0,0,0} and binary matrix for document B should be {1,1,0,0,1,1,1}. Unlike Liu's approach, here the binary matrix is partitioned into sub-groups and each sub-group is called a word code.

Multi-dimensional scaling (MDS) as it is used today was invented by Shepard (1962a and 1962b). However, Torgerson (1958) proposed this technique in 1958. MDS is designed to analyze distance-like (dissimilarity) data in a way that displays the structure of the dissimilarity data as a geometrical picture. In other words MDS calculates Euclidean distances between given points and displays the result on the X-Y axis. Euclidean distance in 2-dimension means the shortest distance between two points. Figure 2.3 shows Euclidean distance between two fields.

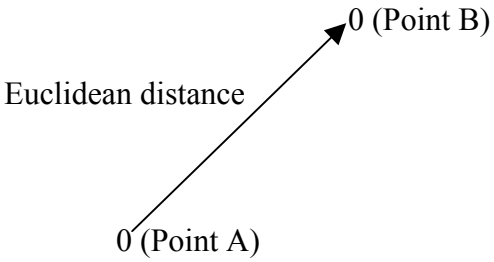

Figure 2.3 Euclidean Distance between Two Points

Here is the example that shows how Euclidean distance could be calculated. Lets assume there are three documents and each document has three terms and each term has frequencies in that document.

	Term A	Term B	Term C
Document 1	6	54	3
Document 2	4	42	5
Document 3	8	36	5

Table 2.1 Term Frequencies for Euclidean Distance Calculation

Formula for Euclidean distance:

Euclidean distance of Document 1 and Document 2 = square root of [((A of D-1 – A of D-2) / (Max (A) – Min (A)))^2 + ((B of D-1 – B of D-2) / (Max (B) – Min (B)))^2 + ((C of D-1 – C of D-2) / (Max (C) – Min (C)))^2]

Euclidean distance between Doc#1 and Doc#2 = square root of [$(2/4)^2$ + $(12/18)^2$ + $(2/2)^2$] = 1.9

Euclidean distance between Doc#1 and Doc#3 = square root of $[(2/4)^2 + (18/18)^2 + (2/2)^2] = 1.5$

Euclidean distance between Doc#2 and Doc#3 = square root of $[(4/4)^2 + (6/18)^2 + (0/2)^2] = 1.111$

Euclidean distances can be drawn on X-Y axis like the figure 2.4 using Multidimensional Scaling method.

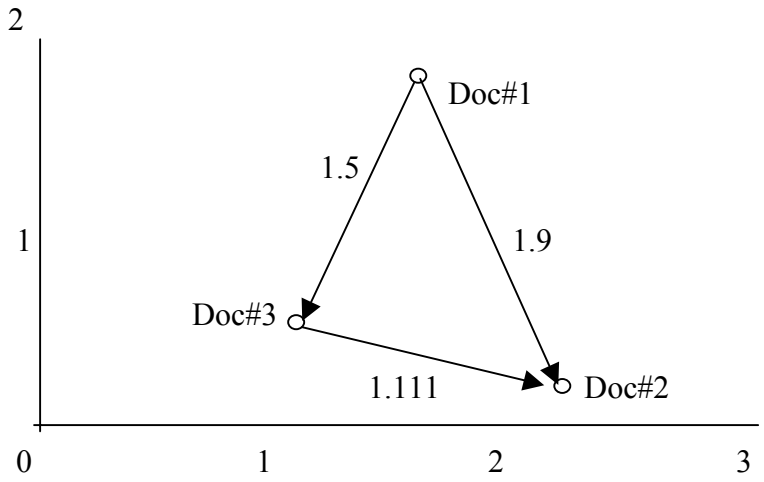

Figure 2.4 A Sample Illustration of Euclidean Distance using MDS

Current Techniques

Image Retrieval by Text

"Long before images could be retrieved by image-contents like color, texture, shape and semantic features, access to image collections was provided by librarians, curators, and archivists through the manual assignment of text descriptors and classification codes" (Goodrum et al., 2000). Text-based classification has a long enduring history including the ability to represent both general and specific instances where an object varies in levels of complexity.

Unfortunately, manual assignment of textual attributes introduces several

problems, such as time consumption, labor absorption, and high cost. Furthermore, manual indexing suffers from low term agreement between indexers (Markey, 1984) and between users while using queries for retrieval (Enser & McGregor, 1993).

More recently, automatic assignment of textual attributes to images has been conducted utilizing the text from captions, transcripts, close captioning, or verbal description for the blind that accompany some videos (Turner, 1998). Although these approaches greatly reduce the labor involved in manual assignment or keywords, they are only available with a small percentage of images. Furthermore, a user's image needs may occur at a primitive level that taps directly into the visual attributes of an image, in which case the accompanying text would not be relevant. These attributes may best be presented by image exemplars and retrieved by systems performing pattern matches based on color, texture, shape, and other visual features (Hermes, 1995).

Problems with text-based access to images have prompted increasing interest in the development of image-based solutions. But, in reality, almost all products on the market today are still using a text driven method for image retrieval.

Image Retrieval by Color

Color is a prominent attribute considered in image retrieval, and one of few that most researchers use. Researchers have added images to collections, extracted color features from the images, and transferred extracted color features to color histograms that show the proportion of pixels of each color within the image (Chang & Yang, 1982; Chang & Liu, 1984; Korfhage, 1997). The color histogram for each image is stored in the database for image retrieval. Users can provide color proportions or an example histogram to retrieve from within an image collection.

Based on the given data, retrieval systems use matching processes to decide which histogram is closest to the given data.

This matching technique, named as "histogram intersection", was developed by Swain and Ballard (1991). Variants of this technique have emerged, improving the original idea of Swain and Ballard by combining histogram intersection with other elements of spatial matching (Stricker & Dimai, 1996) and by using region based color histogram querying (Carson et al., 1997).

Image Retrieval by Texture

Because of its complexity, use of texture similarity does not currently seem to be useful in image retrieval. However, it may be useful in distinguishing areas of images with similar color, such as sky and sea, or grass and leaves (Eakins & Graham, 1999; Korfhage, 1997; Ma & Manjunath, 1998). Though there are several techniques to measure texture similarity, the most common technique involves second-order statistics.

This technique calculates the relative brightness of selected pixels from each image and measures texture similarities such as the degree of contrast, coarseness, directionality and regularity (Tamura et al., 1978), or periodicity, directionality and randomness (Liu & Picard, 1996). Ma and Manjunath (1998) disclosed a recent extension of texture similarity called texture thesaurus. Systems using texture similarity can retrieve from within an image collection by given texture queries or sample images similar in ways to color retrieval.

Image Retrieval by Shape

Shape retrieval is another aspect of image retrieval. Humans first begin matching images by shape. For machines, however, this is not so simple. All

possible features of shape have to be figured using appropriate methods; for example, edge detection method. The main types of shape features are global and local. Global features include ratio, circularity and moment invariants (Niblack et al., 1993). Local features are sets of consecutive boundary segments (Mehrotra & Gary, 1995).

Shape features are understandable in a two-dimensional space, but it is very difficult to grasp shape features from a three-dimensional space and to query three-dimensional images using two-dimensional input data (Niblack et al., 1993; Mehrotra & Gary, 1995). To form a query for a three-dimensional image using a three-dimensional image might be more difficult. Hence, there is no simple solution in shape retrieval.

Image Retrieval by Semantic Feature

Though the majority of image retrieval methods are developed at a primary stage, semantic feature retrieval focuses on more advanced measurements of scene recognition and object recognition. Scene recognition is often used when the image retrieval system is searching for images and identifying specific image over all. Hermes et al (1995) designed a system for scene recognition, which uses color, texture, region and spatial information to derive the most likely interpretation of the scene, generating keyword text descriptors that can be input into any text retrieval system.

Later, this idea was transformed to semantic visual templates and a visual thesaurus to retrieve the most likely relevant images (Chang, S.F., 1998). In conjunction with scene recognition, object recognition was suggested by Brooks (1981) and has been enhanced by several researchers for recognizing and classifying

objects using primitive features of the region, such as color, shape and texture, and not-textual information, such as its position and type of background in the image.

Image Retrieval by Other Types of Primitive Features

Another well-known image retrieval method is the wavelet transform, which allows measurements of an image to be taken at several different resolutions. Liang and Kuo (1998) reported a promising result of the wavelet transform method. There are two versions of retrieval by appearance (Ravela & Manmatha, 1998a). One is whole-image matching, the other is matching selected parts of an image.

Accessing images by spatial location is the essential aspect of geographical information systems. The basic concept is to retrieve maps that have been translated into numerical positions using longitude and latitude for a home address.

Commercial Products using Content-Based Image Retrieval

CONVERA™

CONVERA (formerly Excalibur™) is an image management system developed and distributed by the Excalibur Corp. This system supports image capture, image indexing and image retrieval. This system also supports three matching techniques: color, shape, and texture. Therefore this system enables images to be indexed, searched and retrieved through these three features characteristic.

The color function analyzes the global distribution of color within the entire image. The shape function measures the relative orientation, curvature, and contrast of lines in the image, and the texture function analyzes areas for periodicity, randomness, and roughness of fine-grained textures in images. This system has been designed to process two-dimensional grayscale or color image data which are supported by a common industry standard format: BMP, GIF, TIFF, PNG, PPM, etc.

(http://www.convera.com; Ventors & Cooper, 1999). Figure 2.5 is showing image management system, screening room, by CONVERA, explaining how it works.

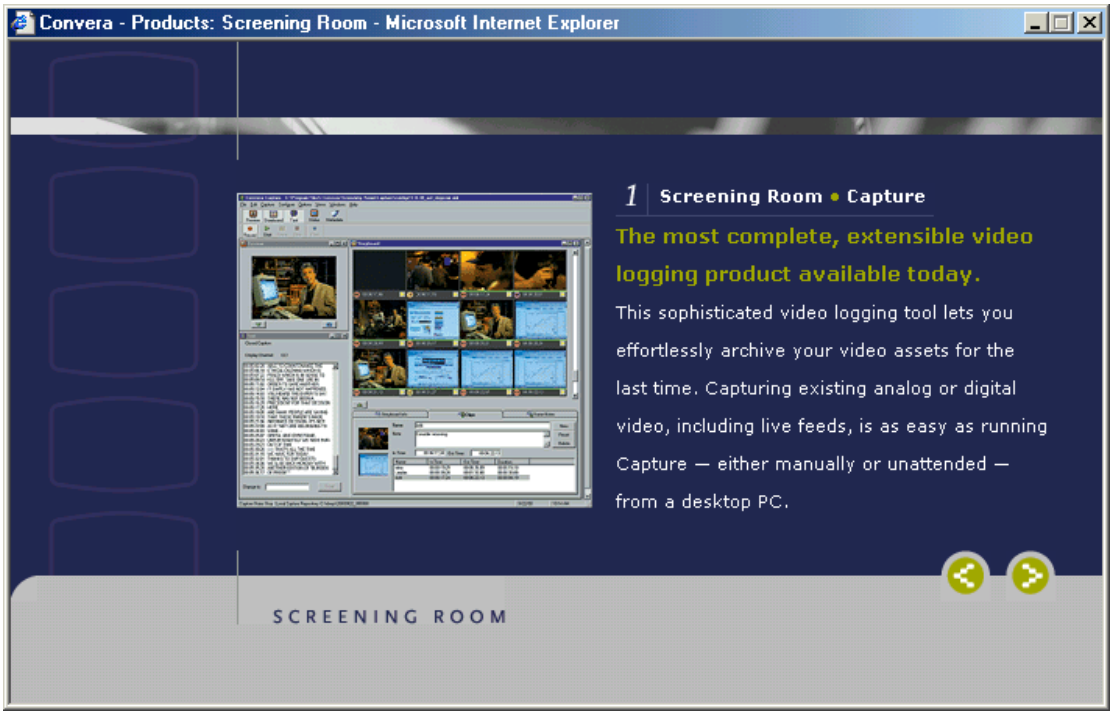

Figure 2.5 Image Management System, Screening Room, by CONVERA

QBIC

IBM Corp has developed the QBIC system that lets users make queries of large image databases based on visual image content -- properties such as color percentage, color layout, and textures occurring in the images. Such queries use the visual properties of images, so QBIC can match colors, textures and their positions without describing them in words. Content-based queries are often combined with text and keyword predicates to get powerful retrieval methods for image and multimedia databases.

This supports the ability of a query to retrieve a database object that is formed by data base management system (DBMS) or Oracle™. QBIC is unique among image retrieval systems in that there are three types of queries: simple query, multi-feature query, and multi-pass query. A simple query supports only one feature for retrieval, the multi-feature query uses more than one feature of color, shape, and texture, and multi-pass feature queries allow for the first results of a search to be used as the basis of the next search.

Like the CONVERA product, QBIC also supports the four matching features: global color, local color, shape, and texture. However, there is some difference between CONVERA and QBIC. For example, in QBIC, the global color function computes the average red, green and blue colors within the entire image, and the local color function computes the color distribution for both the dominant color and the variation for each image in a predetermined 256 color space.

For the shape function, QBIC computes a combination of area, circularity, eccentricity, and major axis orientation. For texture function, QBIC supports the area analysis for coarseness, contrast, and directionality.

Like CONVERA, QBIC also supports many kind of industrial standard image format (http://wwwqbic.almaden.ibm.com/, Ventors & Cooper, 1999).

VIRAGE™

VIRAGE executes the two primary functions of image analysis and image comparison like CONVERA and IBM's QBIC. However, VIRAGE differs from CONVERA and QBIC in several ways. First, it produces feature vector information from image analysis. VIRAGE also uses the feature vector information for image comparison, and then produces a score dataset from this process. A score set is used

to determine which image is close to the input data. VIRAGE also supports four basic functions: global color, local color, structure, and texture. The global color function calculates the distribution of color for the entire image set. Just as the local color function analyzes the distribution of color in the localized area of each image, the structure function is as same as the shape function, but the company uses different names for large-scale shapes. Finally, the texture function analyzes areas for periodicity, randomness, and roughness of fine-grained textures in images. Like the other products, VIRAGE also supports several industrial standard image formats (http://www.virage.com; Ventors & Cooper, 1999). Figure 2.7 is showing the main screen of VIRAGE.

Figure 2.7 VIRAGE Web Screen (http://www.virage.com)

Other Products

One of the windows-based content-based retrieval systems is ImageFinder™, developed by Attrasoft Corp (www.attrasoft.com). Unlike other image retrieval systems ImageFinder uses a neural network as its basic foundation. The feature matching technique of this system is based on complete or incomplete pattern matching. Performance problems can occur when this system searches huge amounts of images because of the inherent obstacle in speed in the pattern matching technique and accuracy in neural network. Figure 2.8 is showing the starting window of ImageFinder user interface for training of key image.

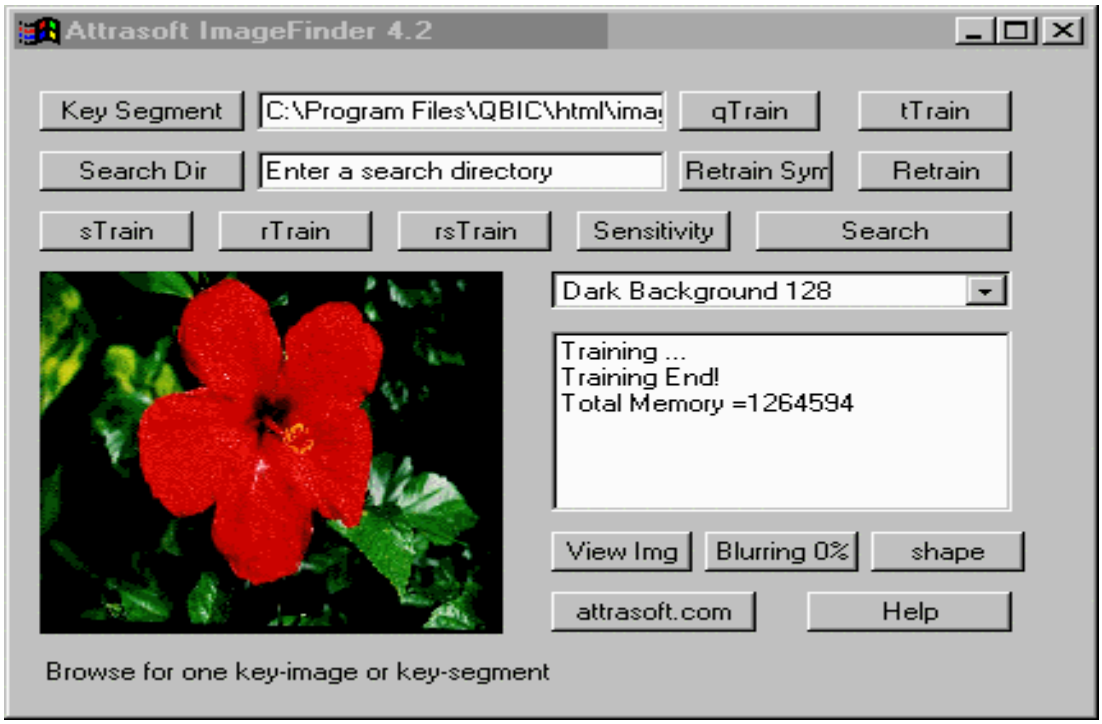

Figure 2.8 The ImageFinder User Interface

Imatch™, developed by Mario M. Westphal, is a shareware utility for the Windows operating system (www.imatch.com). Imatch allows users to perform a

variety of content-based image retrieval operations on their datasets, such as color similarity, color and shape (quick), color and shape (fuzzy), color percentage, and color distribution. Imatch also supports non-CBIR features to identify images with CRC checksum, duplicate scanner, and fuzzy filename.

Besides these systems, there are many prototypes for image retrieval based on CBIR, for example, the CANDID™ (Comparison Algorithm for Navigating Digital Image Database) system developed at Los Alamos National Laboratory (kelly@lanl.gov), the ARTISAN™ (Automatic Retrieval of Trademark Image by Shape ANalysis) developed at the University of Northumbria at Newcastle (http://www.artisan.demon.co.uk), the BlobWorld™ developed at the University of California, Berkeley, etc. (http://www.elib.cs.berkeley.edu/photos/blobworld)

Practical Applications of Content-Based Image Retrieval

Medical Diagnosis

Modern medical diagnostic technologies such as radiology, histopathology, and computerized tomography produce huge amounts of images everyday. Currently these images are being stored in hospitals around the globe whether they are stored in computer systems or not. Though the primary job of medical image retrieval systems is to retrieve the related patient's image with the patient's name, there is an increasing interest in the use of CBIR techniques to aid in diagnosis by identifying similar past cases.

The need for better image retrieval systems in the medical field is growing quickly, and developmental work on this kind of system is still in its beginning stages. Researchers in this field are still focusing on delivering basic functionality and usability ensuring that medical images can be successfully digitized, stored and

transmitted over local area networks without loss of quality. Ultimately, researchers hope to provide user-centered interfaces and integrate image storage and retrieval in order to better service for doctors and patients and management people.

In 1994, researchers at the University of Crete developed a 2-dimensional radiological image retrieval system, and Liu et al (1998) unveiled a 3-dimensional neurological image retrieval system at Carnegie-Mellon University. Both systems are targeted for assisting medical staff in diagnosing brain tumors.

Geographical Information Systems (GIS)

GIS combines numerical data with a map, a type of image, in several ways. The greatest benefit of GIS is that it provides explanations, which are statistically, geographically, and academically related. Commercial GIS systems, like ArcInfo™ and ArcView™ provided the capacity to search spatially referenced data by location or spatial attribute (Demers, 1999; ESRI, Inc. 1996; ESRI, Inc. 1997). This is a greatly useful function, but ArcInfo and ArcView still do not use CBIR. One team, Ma and Manjunath (1998), developed an experimental system which was aimed at identifying objects or regions within satellite images or digitized maps by shape, color, or texture similarity.

World Wide Web Searching

At an early age of Web search development, text-based image search engines have been adopted and progressed rapidly to meet the satisfaction of users. However, soon after the systems known to the users, developers knew that text-based image search engines would not be enough because of the speed of technological growth. A few years ago, several experimental content-based image searchers emerged for the World Wide Web. One of them is WebSEEK™ developed at the Department of

Electrical Engineering, Columbia University. This system supports color, spatial layout and textual matching features. The Web demonstration can be located at http://disney.ctr.columbia.edu/WebSEEK .

Another controversial and highly profitable topic on Web searching is identifying pornography and blocking access to pornographic content to children. One of the more basic problems involved in this process is to understand what constitutes pornography in the first place (Forsyth et al., 1997; Chan et al., 1999). This debate will continue, but in the meantime systems will continue to be developed to try and use CBIR and text/keyword extraction to achieve this end.

Other Areas of Interest in CBIR

Eventually CBIR will be a commonplace feature of nearly every field of interest. CBIR technology has already established itself in areas such as crime prevent, on war games for the military, architectural and engineering design, journalism, advertising, training and education, museums, libraries, toll gates for automatic fee collection, fashion, and interior design. Though all these fields currently use CBIR technology, they approach it from different perspectives.

For example, law enforcement agencies use CBIR technology in identifying the fingerprints and faces of suspects using similarity matching. A number of automatic fingerprint identification systems (AFIS) are now commercially available, including AFIX Tracker™ from Phoenix Group Inc, Pittsburg, Kansas (http://www.affix.com) and the Finger Search Engine™ from East Shore Technologies, Inc. of New York (http://www.east-shore.com). In the architectural and engineering design fields, CBIR can be used with computer aided design (CAD), and acts as a fundamental resource for architectural and engineering design

(http://vision.ucsd.edu/papers/manu/manu.html). The fields of journalism and advertising are also facing problems in being able to better utilize text and video archives. Archives are being produced every second, but it is difficult to get enough experts trained to do this job and then to pay them for the cost for this work, so automated systems significantly improve their performance. So there is a good reason to use CBIR technology for the automatic indexing of these archives, though such systems are not yet available widely.

Limitations of CBIR

Storage of Multimedia Documents

With the rapid growth of computer hardware peripheral devices, large amounts of data can be stored at relatively small costs. In general, to store every kind of multimedia document would necessitate unlimited storage, but in reality the amount of storage is limited. Multimedia documents need more storage space than do text only documents.

For example, one black and white image that is 400 by 400 pixels could require 160KB to be stored without the use of a compression technique. If it is a color image using RGB, then the size will grow to three times more than black and white approximately. That is the reason why many multimedia retrieval systems adopt various compression techniques to reduce the size of storage. To store one thousand images having 400 by 400 pixels in a RGB format without compression techniques would require 480MB (Eakins & Graham, 1999; Demers, 1999; Korfhage, 1997; Forsyth, 1997). However, if compression technique, WinZip™, offered by Microsoft Inc. is used, the size of storage will be saved more than 50 percent of 480MB.

Processing of Multimedia Documents

Methods of processing multimedia documents differ widely depend on the combination of multimedia documents. If it is combined with text and image multimedia retrieval system needs a method to retrieve image documents, but it is combined with image and sound multimedia retrieval system needs methods to retrieve images and sounds like CBIR technique for images and speech recognition technique for sound. Even though lots of multimedia processing methods proposed such as QBIC, CONVERA, VIRAGE, WebSEEk, etc., there is no compelling method and each one has various benefits and drawbacks. Multimedia retrieval systems based on the CBIR technique are relatively fast but do not currently give good results according to precision and recall (Venter & Cooper, 1999). On the other hand, multimedia retrieval systems based on a pattern matching technique are somewhat slow but give more reliable results. Depending on the purpose of the multimedia retrieval system, the method of processing and retrieval for multimedia document has to be chosen carefully.

Justification of This Research

Most image retrieval systems are text based such as Yahoo™, Google™ and Amazon.com™. But text-based image retrieval systems give some misclassified documents because there is misconnection between text keyword and image document.

Like QBIC computational image processing techniques provide a mechanism to retrieve images based on low-level image features such as color, shape, and texture. In this method there is still misclassification problem because images can not be

explained fully. The degree of misclassification of this method was shown in chapter 2.3.2.

Unlike the above two methods poly-representation methods generally provide better retrieval results like VIRAGE and CONVERA, but there are unique problems in combining textual features and visual features in a single representational space. Usually they maintain separate data structure for text and image documents. The main reason they give misclassified documents is that text documents are no longer used in retrieving image documents after the retrieval system is initiated using text keywords.

This research endeavors to explore the feasibility of combining text and image features into a single representational space using Jeong's Transform and tests the multimedia retrieval system of this approach to prove how much the above problems could be solved. To prove the improvement four experimental retrieval systems were built. For testing set of 26 multimedia documents http://archive4.lis.unt.edu/td26/www was built and this works exactly like poly-representational methods. Also, http://archive4.lis.unt.edu/tdt26/www was built in using a single representational space. Precision and recall of two systems are calculated and compared in chapter 5. For testing set of 994 multimedia documents http://archive4.lis.unt.edu/td/www was built and this works like poly-representational methods. Also, http://archive4.lis.unt.edu/tdt/www was built in using a single representational space. Precision of two systems are also calculated and compared in chapter 5.

CHAPTER 3. METHODOLOGY

Introduction

For this research, a new approach is introduced: a common representation format for text and image. To ensure the validity of this research, three experimental methodologies are being suggested; namely: a parallel comparison of retrieved results, a visual comparison of vector graphs, and a gradual expansion of grouping terms on test set of text retrieval conference (TREC) data. To illustrate this approach, two sets of multimedia documents are used: one from National Aeronautics and Space Administration (NASA) which has 26 multimedia documents and 105 multimedia documents and 994 multimedia documents and one from TREC text document which has 100 text documents.

The main idea for this common representation format is to obtain a single data structure for both text and image data. To accomplish this goal, the Word Code from a binary matrix for text data is modified and the Lorenz Information Measurement (LIM) is used for calculating the area of a histogram from the image data. The experiments are run to see the effect of grouping terms on text documents. The experiment is using a gradual expansion of grouping terms on a test set of TREC document to see the effect of different number of grouping terms. Grouping terms are expanding from 1 to 10, 20, 50, and 100 terms. Two more experiments, a parallel comparison of retrieved results and a visual comparison of vector graphs, were done. Parallel comparison provides precision and recall values in the case of retrieval results using image data only and retrieval results based on text and image features combined. Through the visual comparison of vector graphs the difference of shape on 2-dimension will be noticed.

Twelve measurements are generated from the Word Code and twelve measurements from the LIM. These measurements from both the text and image data are then combined to twenty-four measurements. These measurements are used for multi-dimensional scaling (MDS) analysis, which produces one vector for each multimedia document. All the vectors derived from MDS are then used to evaluate the closeness of the multimedia documents. The twenty-four measurements generated from the text and image documents are also used for multimedia document retrieval, and these retrieved images from the multimedia document are used as the heuristic judgment of this research.

A parallel Comparison of Retrieved Results

From many image representation methods, the content decomposition algorithm is chosen and combines three main components such as color, shape, and texture. Through this algorithm, pixel values of twelve components are extracted. These twelve components are pixel values of Red, Green, Blue, Gray, Distance-A, Distance-B, Distance-C, Distance-D, Distance-E, Hough Transform, Angle, and Density. All of the extracted pixel values except Density are transformed to histogram values. Density is already in the form of area because number of edge-detected pixels is used as Density here. The area of each histogram is then calculated. These twelve measurements (Table 3.1) are used for an image only retrieval system. From the text document, twelve measurements are captured using the procedure explained in 3.3 (A gradual expansion of grouping is done on test set of TREC data). For multimedia document retrieval, the twelve measurements from image data (Table 3.1) are combined with the twelve measurements from text data

(Table 3.2). These twenty-four measurements constitute a common representation format for multimedia documents.

Figure 3.1 Example of NASA image for S88E5001

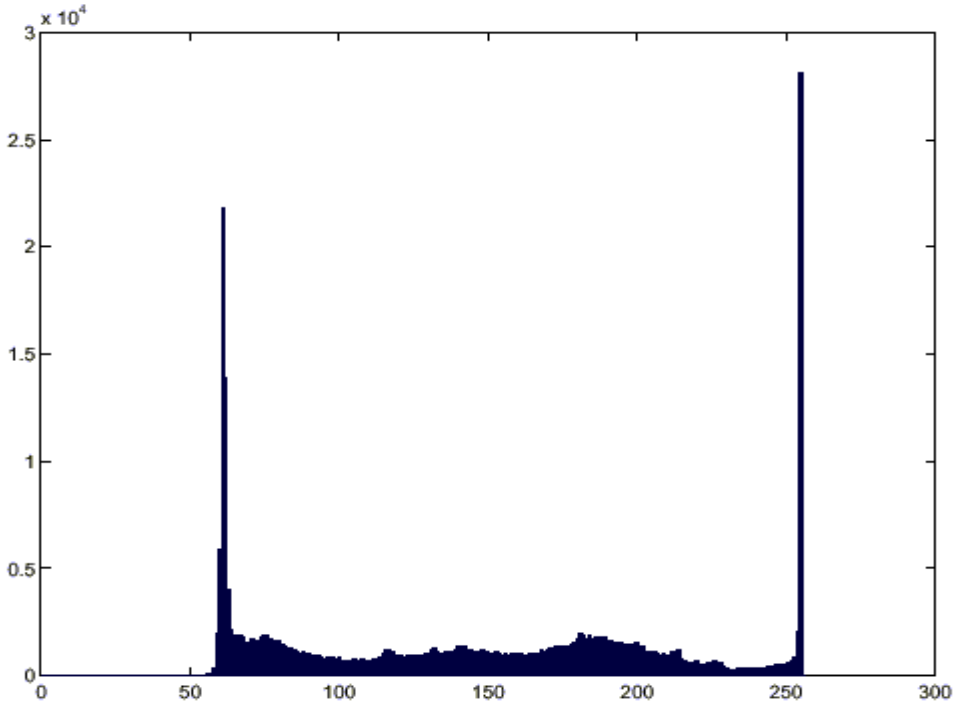

Figure 3.2 Example of the histogram for Red values of the ground instrumentation image shown above

Extractions / Images	Red	Green	..	..	..	..	..	..	..	Density
I_1	0.234	0.326	..	..	..	..	..	..	..	0.234
I_2	0.332	0.333	..	..	..	..	..	..	..	0.332
..	0.453	0.432	..	..	..	..	..	..	..	0.432
..	0.111	0.333	..	..	..	..	..	..	..	0.321
I_n	0.123	0.321	..	..	..	..	..	..	..	0.113

Table 3.1 Lorenz Information Measurements from Images

WCM / Text	G_1	G_2	..	..	..	..	..	..	..	G_{12}
T_1	0.1234	0.2326	..	..	..	..	..	..	..	0.1234
T_2	0.1332	0.3333	..	..	..	..	..	..	..	0.3332
..	0.2453	0.432	..	..	..	..	..	..	..	0.2432
..	0.3111	0.1333	..	..	..	..	..	..	..	0.1321
T_n	0.3123	0.1321	..	..	..	..	..	..	..	0.2113

Table 3.2 Word Code Measurements (WCM) from Text Documents

To retrieve and then evaluate multimedia documents, two retrieval systems are constructed; one for the image measurements only, the other for the text and image measurements combined. Text only retrieval system is not built because it is out of focus in this research.

Precision is defined as the proportion of retrieved documents that are relevant, P = w / n2. Recall is defined as the proportion of relevant documents that are retrieved, R = w / n1 (Korfhage, 1999).

From the table;

n1 = w + x

n2 = w + y

n = w + x + y + z

	Retrieved number of Docs	Not retrieved number of Docs
Relevant Docs	w	x
Not relevant Docs	y	z

Table 3.3 Contingency Table for Evaluating Retrieval

To test the result of this research, a person not related to the project prepared twenty-five questions with various possible answers for each question. Testers are graduate student volunteers from the School of Library and Information Sciences at the University of North Texas. 24 testers in total are divided into groups of three.

Two testing methods are used. The first method retrieves the images using the given text questions from the data, twelve measurements, constructed from image data only, and calculates the precision and recall. The second method retrieves the images using the given text questions from the common representation data of multimedia document, twenty-four measurements, to calculate the precision and recall. All conclusions are based on agreement of three testers in the group.

The Brighton Image Searcher shown below handles the process of retrieval for multimedia documents. The search starts with text terms, then the Retrieval System yields images related to the given text terms. From the retrieved images, tester can choose an image. The chosen image is then used to retrieve similar images from the multimedia document set. Finally, to calculate precision and recall the testers evaluated the retrieved images. Table 3.4 is showing the evaluation results over the

multimedia testing sets. Procedures and all the testing results will come in chapter 5 in great detail.

	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
994 multimedia documents	0.13				0.43			

Table 3.4 The Evaluation Results of Precision and Recall over Multimedia Testing Sets

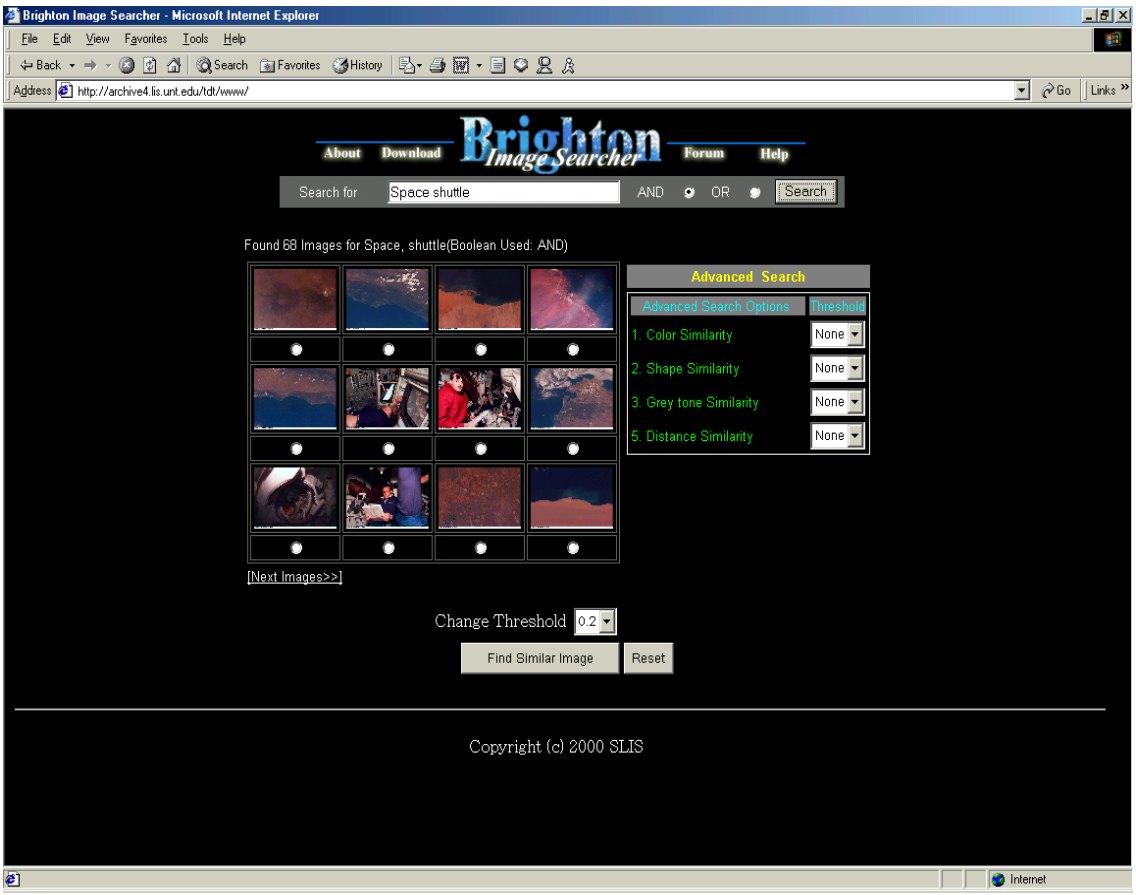

Figure 3.3 Image Retrieval System, the Brighton Image Searcher, for heuristic judgment

A Visual Comparison of Vector Graphs

Visualization in information retrieval is attractive in that it allows for users to see what the IR is doing in a way that's simpler to understand. In this research, there are four categories for image decomposition -- color, distance, angle, and texture, and it is possible to draw documents into a visualized vector graph.

For the visual comparison of vector graphs, all possible combinations are made without text measurements and all possible combinations are made with text measurements as shown in the Table 3.4 (The Combinations of LIM and Text

Measurement). The vector graphs of thirty combinations are then evaluated to find the combination that most reliably measures the distances between vectors that are representing documents to be representative of similarities between documents. Among those vector graphs, Figure 3.4 the vector graph for 26 images only (C15 in Table 3.5 without text measurement) is shown and it shows that there are some clusters.

Combination	Categories	Text Measurements
C1	Color	With Text, Without Text
C2	Color, Distance	..
C3	Color, Angle	..
C4	Color, Texture	..
..	..	..
..	..	..
..	..	..
C15	Color, Distance, Angle, Texture	With Text, Without Text

Table 3.5 The Combinations of LIM and Text Measurement

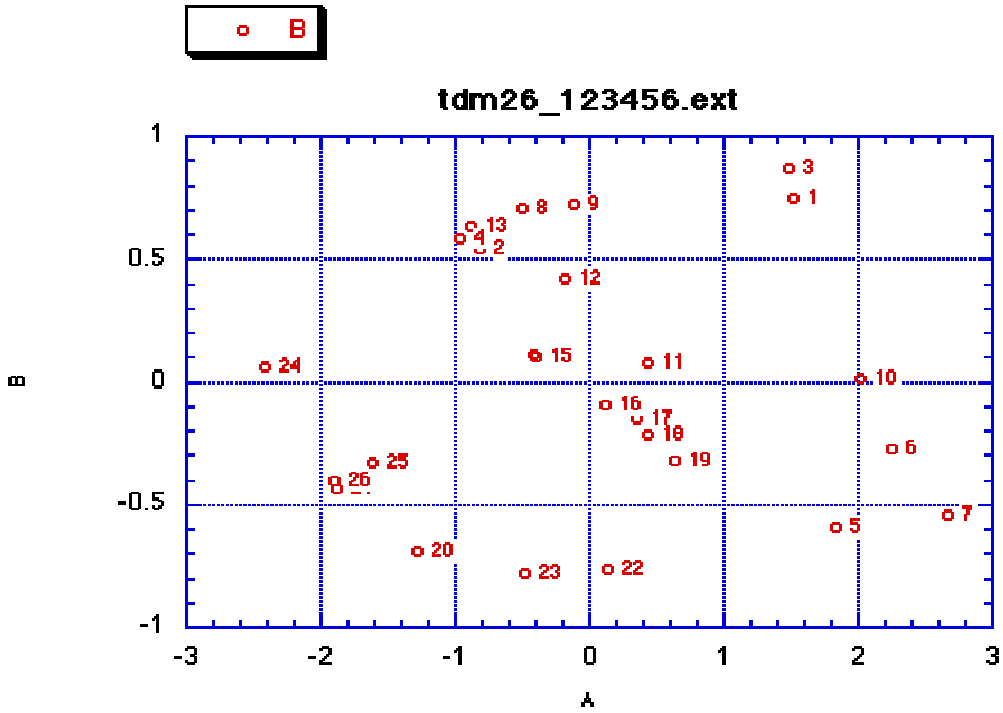

Figure 3.4 Vector Graph of 26 Images from NASA (Color, Distance, Angle and Density)

A Gradual Expansion of Grouping Terms on Test Set of TREC Data

Since 1990, TREC data has been widely used for the research purpose to develop and evaluate text retrieval systems. All terms extracted from the test set of TREC data are usually used in developing and evaluating the text retrieval system, but in this research, a grouping method of extracted terms is used, and visualized vectors on X-Y 2-dimensional graph are used to see if this method is more reliable.

The best-known text representation method is term extraction from a text document. A super term list is made from the whole documents set and a term list for each document is made based on the super term list. Extracted terms can be represented in several ways, including by term frequencies where the number of times

a word appears in a given document or by binary representation. The super term list is used as a base, and for each document, a binary matrix of documents and terms can be produced containing a combination of "0's" and "1's" (Table 3.7). In the binary matrix, a "0" means that the term does not appear in the document and a "1" means that the term does appear within the document.

Assume the terms are A_1 through A_n and the documents are D_1 through D_m. For the first experiment, a vector graph is drawn using MDS from the binary matrix of Table 3.7. For the second experiment, the n terms are divided into groups (also called Word Code) of 10 terms (Table 3.8), and each group having the frequencies of 1 of that group. For the third experiment, the n terms are divided into groups of 20 terms. Groups of 50 terms and groups of 100 terms are also tested. In the table 3.8, G_1 through G_p are the groups, and G_p represents total number of frequency of 1's in each document.

<DOC>
<DOCNO> AP890102-0137 </DOCNO>
<FILEID>AP-NR-01-02-89 1212EST</FILEID>
<FIRST>a a BC-EXP--AIDSSurvivors Adv05 01-02 1115</FIRST>
<SECOND>BC-EXP--AIDS Survivors, Adv 05,1144</SECOND>
<HEAD>$adv05</HEAD>
<HEAD>For release Thursday, Jan. 5, and thereafter</HEAD>
<HEAD>Long-Term AIDS Survivors Defy Odds</HEAD>
<HEAD>With LaserPhoto</HEAD>
<BYLINE>By BRENDA C. COLEMAN</BYLINE>
<BYLINE>Associated Press Writer</BYLINE>
<DATELINE>CHICAGO (AP) </DATELINE>
<TEXT>
Mike has lived twice as long as might have been expected when doctors diagnosed his AIDS. Dan Turner and Cristofer Shihar had one chance in five of seeing 1984. They don't know why they've survived what has been a death sentence for more than 45,000 Americans, but say it may be a matter of attitude. ``A lot of people don't die of the disease, they die because they give up," said Mike, a 34-year-old Chicagoan who was diagnosed with acquired immune deficienty syndrome in January 1984. He asked that his last name be withheld. According to Judith Wiker, a Chicago holistic therapist who says she has counseled hundreds of clients with AIDS or AIDS-related problems, Mike is one of many people with the disease who are well enough to feel and act normal. ``Is the virus somehow different?" asks Ann M. Hardy, a CDC epidemiologist now at the National Center for Health Statistics in Hyattsville, Md. ``Is it something in their immune system?" Does survival time hinge on the mildness or severity of the infections that attack people with AIDS? Or could the key really be a ``lifestyle-psychosocial type of thing" _ a positive attitude and emotional support? All of these possibilities are now being studied, either by the CDC or in studies funded by the National Institutes of Health. For the purposes of the CDC's 2-year-old study, long-term survivors were defined as people who lived at least three years after being diagnosed. Unlike the estimated hundreds of thousands of Americans who are infected with the AIDS virus but do not have symptoms, long-term survivors actually have battled one or more ailments that define acquired immune deficiency syndrome _ including Kaposi's sarcoma, pneumonia, damaged immune systems and severe weight loss. ``As of November 1988, we can assist somebody to stay alive and healthy for two years, with the current therapy," Piers said. ``And a great deal may occur in two years. We've seen an enormous change from 1986 to 1988. ``Many people being diagnosed now may benefit from breakthroughs that will totally change the surface of the disease."
</TEXT>
<NOTE>End Adv for Jan. 5</NOTE>
</DOC>

Table 3.6 An Example of TREC Data (AP890102-0137)

Term / Doc	A_1	A_2	..	..	..	..	..	..	..	A_n
D_1	1	0	..	..	..	..	..	..	..	0
D_2	0	0	..	..	..	..	..	..	..	0
..	0	0	..	..	..	..	..	..	..	0
..	0	0	..	..	..	..	..	..	..	0
D_m	0	0	..	..	..	..	..	..	..	0

Table 3.7 A Sample of Binary Matrix from TREC Text Documents

WCM / Doc	G_1	G_2	..	..	..	..	..	..	..	G_p
D_1	1	0	..	..	..	..	..	..	..	213
D_2	1	0	..	..	..	..	..	..	..	276
..	1	0	..	..	..	..	..	..	..	185
..	1	0	..	..	..	..	..	..	..	134
D_m	3	0	..	..	..	..	..	..	..	256

Table 3.8 A Sample of 10 Terms Word Code Matrix from the Binary Matrix

After finding the largest frequency from all the groups, the frequency of each group is divided by twice the largest frequency for the group. This is done because LIM for image has less than or equal to 0.5. As shown in the Table 3.9, Word Code

Matrix values of Table 3.8 are transformed using the rule explained above and all values are less than or equal to 0.5. Using value of each group per document, a vector graph is drawn using MDS. Using the process above, vector graphs are drawn using MDS for groups of 10 terms, groups of 20 terms, groups of 50 terms and groups of 100 terms. Figure 3.5 shows a vector graph of 10 terms of Word Code corresponding to Table 3.9. Finally, the vector graphs for all groups will be compared.

WCM / Doc	G_1	G_2	..	..	..	..	..	..	..	G_p
D_1	0.16667	0	..	..	..	..	..	..	..	0.198324
D_2	0.16667	0	..	..	..	..	..	..	..	0.256983
..	0.16667	0	..	..	..	..	..	..	..	0.172253
..	0.16667	0	..	..	..	..	..	..	..	0.124767
D_m	0.5	0	..	..	..	..	..	..	..	0.238361

Table 3.9 A Sample of 10 Terms Normalized Matrix from Table 3.8

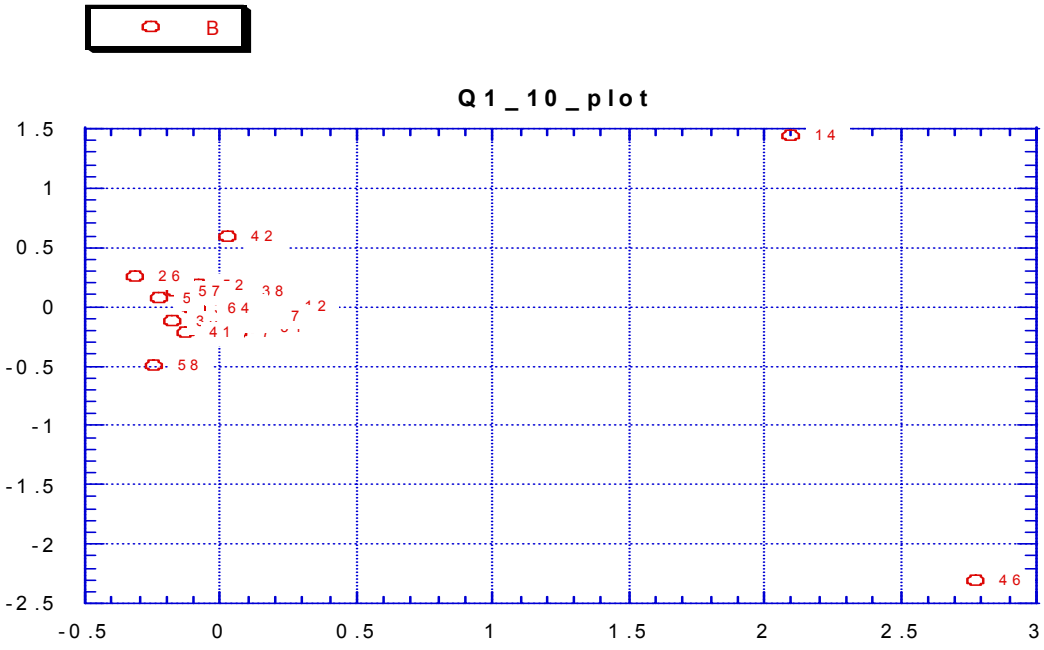

Figure 3.5 Vector Graph of 10 Terms of Word Code

CHAPTER 4. PROCESSING FOR TEST SETS

Introduction

For this research, three sets of multimedia documents were collected from the National Aeronautics and Space Administration (NASA) and one set of text document from the Text Retrieval Conference (TREC). The first set of NASA multimedia document was composed of 26 images and text documents, the second was composed of 105 images and text documents, and the third was composed of 994 images and text documents. For the TREC text documents, 100 TREC text documents were selected randomly from hundreds of text documents for this research.

Processing for the text documents from TREC is shown in Figure 4.31. According to the process, terms are extracted from text documents to make a super term list. Using the super term list a binary matrix is made. After that a Word Code matrix is made, and then it is transformed to the final normalized Word Code matrix using Jeong's Transform.

Multimedia documents from NASA were composed of two parts; image and text. For this test the first multimedia documents set from NASA were used and there were 26 multimedia documents. Using the process as explained in Figure 4.30 for image process and in Figure 4.31 for text process, image and text were analyzed separately. After that the two results were combined to obtain a single data structure. The main idea for this Common Representation Format is to obtain a single data structure to combine text and image. To accomplish this goal, the Word Code from a binary matrix for text data and the Lorenz Information Measurement (LIM) calculating the area from a histogram for the image data were used. Finally, this

common representation format from multimedia documents (CRFMD) was used to draw each vector graph using multi-dimensional scaling (MDS) and visualize the multimedia documents.

Processing for Text Documents

Term Extraction

Since 1990, TREC data has been widely used in a research to develop and evaluate text retrieval systems. Even though there are several ways to represent text documents, terms extracted from text documents do well represent text documents. Terms extracted from the test set of TREC data and terms consist of a super term list. As a test set of TREC data, 100 documents were randomly chosen as a test set and 6287 terms were extracted from the given test set. Table 4.1 is an example of a text document and Table 4.2 is the Perl program developed by Oyarce in 1999 used to extract terms from the test set. Table 4.3 is a part of super term list from among the 6287 terms extracted from the test set.

<DOCNO> AP890109-0313 </DOCNO>
<FILEID>AP-NR-01-09-89 1035EST</FILEID>
<FIRST>u f PM-Britain-GEC 01-09 0556</FIRST>
<SECOND>PM-Britain-GEC,0578</SECOND>
<HEAD>Government Looks at Possible Bid for British Electronics Giant</HEAD>
<DATELINE>LONDON (AP) </DATELINE>
<TEXT>
The government said today that it was looking at a possible bid for the electronics giant General Electric Co. PLC that an international consortium is expected to launch within days. The takeover, which analysts say could be worth between $11.5 billion and $14.2 billion, would be the largest in Britain. The consortium is expected to include Plessey PLC, another
electronics company which is the target of a $3 billion hostile takeover bid from GEC and Siemens AG of West Germany, another electronics company. Although no bid for GEC has been formally launched, the Office of Fair Trading has legal powers to look at a bid ``in contemplation.'' ``We really are looking at the situation to see who the participants are involved before we can take real active steps,'' said a spokesman for the office, who asked not to be identified. ``There hasn't actually been a statement of intention.'' The Office of Fair Trading usually reviews a bid and then makes a recommendation to the trade secretary on whether he should refer it to the monopolies commission for a full investigation. The bid speculation prompted heavy trading in GEC on London's Stock Exchange by midday Monday. A GEC spokesman, who wasn't identified in accordance with British practice, called the developments vague and inconclusive but said
that a takeover would be fought. ``This appears to be a self-interested attempt by the board of
Plessey and its advisers to form a consortium to break up GEC and therefore save Plessey,'' he said.
The possible bid for GEC began to be taken seriously after the investment firm Lazard Brothers and Co. said over the weekend that it had helped form a company called Metsun Ltd. to devise a proposal ``which may or may not'' lead to an offer for GEC. Metsun is headed by Sir John Cuckney, chairman of helicopter maker Westland PLC, which was at the center of a 1986 takeover
controversy that prompted the resignation of two British Cabinet ministers. Metsun was talking to possible partners both in Britain and abroad, Lazard Brothers said, without identifying them. French electronics company Thomson-CSF said it was considering joining the consortium. Meanwhile, Barclays Bank PLC confirmed it was putting together a $6.2 billion syndicated loan to help finance such a bid. GEC Managing Director Lord Weinstock said in a television interview that his company dropped Barclays Bank as one of its banks because the bank had ``behaved in a way that was not quite right.''
</TEXT>
</DOC>

Table 4.1 An Example of a TREC Document (AP890109-0313)

```
###################       MAIN      ###################
 getWORKfiles;          #  Reads filenames in input stream
&getSTOPw;              #  Reads Stop Words
foreach(@INfileList)    #  Starts proc. on input stream
  {
   $SEEin=$_;
   open (IN,"<".$SEEin);              # open file is input stream
   while (<IN>)                       # get in documents to id TEXT field
     {
       $t=$_;
            &UPDATEgd;
            &LOCALcount;
     }
close (IN);
  }
 UPDATEgd;
&printGD;
exit;

## Subroutines are excluded ##
```

Table 4.2 Term Extraction Program using Perl Language

```
threat:1
depression:1
regarding:1
wildlife:1
razed:1
evening:1
maintained:1
depositary:1
passenger:1
maximize:1
substantially:1
estimate:1
freeman:1
running:1
samuel:1
leave:1
independently:1
radio:1
paying:1
south:1
abroad:1
women:1
```

Table 4.3 A Part of Super Term List with Frequencies of Term

Binary Matrix

Extracted terms called a super term list can be represented in several ways, including by term frequencies and by binary representation. The binary representation method was chosen for this paper because of its simplicity. A super term list was made from the entire document set using the term extraction method. Then using the super term list, a term-document binary matrix was produced, which contained a combination of "0's" and "1's" as shown in Table 4.4 using a Perl program (Table 4.5) to identify whether or not a particular term existed in a particular document. In the binary matrix, a "0" means that the term does not appear in the document and a "1" means that the term does appear within the document. In Table 4.4, A_1, A_2 . . ., A_n represent terms extracted from the given TREC text document set and D_1, D_2 . . . D_m represent document from the given TREC text document set. For this experiment, 100 documents from TREC data set were chosen randomly and 6287 terms were extracted from the 100 documents. Table 4.5 shows a Perl program to generate a binary matrix like Table 4.4 using a super term list.

Term / Doc	A_1	A_2	..	..	..	..	..	..	..	A_n
D_1	1	0	..	..	..	..	..	..	..	0
D_2	0	0	..	..	..	..	..	..	..	0
..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..
D_m	0	0	..	..	..	..	..	..	..	0

Table 4.4 An Example of a Binary Matrix from TREC Text Documents

```
###################      MAIN     ###################
open(MatOUT,">$ofil");
&getWORKfiles;          #  Reads filenames in input stream
&getFeatureW;           #  Reads Feature Words
&clearMat;              #  Clear Matrix
foreach(@INfileList)    #  Starts proc. on input stream
  {
   $DocNameRead=$_;
   $SEEin=$_;
   open (IN,"<".$SEEin);              # open file is input stream
   &clearMat;
   while (<IN>)                       # get in documents to id TEXT field
     {
          &makeMat;
     }
   &printMat;
   close (IN);
  }
close(MatOUT);
exit;

## Subroutines are excluded ##
```

Table 4.5 A Binary Matrix Generating Program using Perl Language

Word Code

For the first experiment, the 6287 terms were divided into groups (also called Word Code) of 10 terms (Table 4.6) with each group having the frequencies of "1" in that group. For the second experiment, the 6287 terms were divided into groups of 20 terms (Table 4.7) with each group having the frequencies of "1" in that group. Groups of 50 terms (Table 4.8) and groups of 100 terms (Table 4.9) were also made using the matrix grouping program using Perl language shown in Table 4.10.

WCM / Doc	G_1	G_2	..	..	..	..	..	..	..	G_p
D_1	1	0	..	..	..	..	..	..	..	213
D_2	1	0	..	..	..	..	..	..	..	276
..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..
D_m	3	0	..	..	..	..	..	..	..	256

Table 4.6 An Example of 10 Terms Word Code Matrix from the Binary Matrix

WCM / Doc	G_1	G_2	..	..	..	..	..	..	..	G_p
D_1	1	1	..	..	..	..	..	..	..	213
D_2	1	0	..	..	..	..	..	..	..	276
..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..
D_m	3	0	..	..	..	..	..	..	..	256

Table 4.7 An Eample of 20 Terms Word Code Matrix from the Binary Matrix

WCM / Doc	G_1	G_2	..	..	..	..	..	..	..	G_p
D_1	2	2	..	..	..	..	..	..	..	213
D_2	2	3	..	..	..	..	..	..	..	276
..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..
D_m	3	5	..	..	..	..	..	..	..	256

Table 4.8 An Example of 50 Terms Word Code Matrix from the Binary Matrix

WCM / Doc	G_1	G_2	..	..	..	..	..	..	..	G_p
D_1	4	3	..	..	..	..	..	..	..	213
D_2	5	3	..	..	..	..	..	..	..	276
..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..
D_m	8	5	..	..	..	..	..	..	..	256

Table 4.9 An Example of 100 Terms Word Code Matrix from the Binary Matrix

```
##################  Definitions  ##################
print "input Matrix File = ";
$MatrixF=<STDIN>;
print "output Weight File = ";
$Ofil=<STDIN>;
print "Group Numbers = ";
$GroupN=<STDIN>;
$sw1=0;

##################      MAIN     ##################
open(WeightF,">$Ofil");
$SEEin=$MatrixF;
open (IN,"<".$SEEin);              # open file in input stream
while (<IN>)
     {
       $sw=1;
       &makeWeight;
     }
close (IN);
close(WeightF);
exit;

## Subroutines are excluded ##
```

Table 4.10 A Matrix Grouping Program using Perl Language

Each group matrix was normalized using the matrix normalizing program in order for each result to be less than or equal to 0.5.

The formula for normalization of group is this:

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

Table 4.11 shows the result of normalized values to Table 4.6 using the above formula and Table 4.15 shows the matrix normalizing program using Perl language. Table 4.12 represents the result of normalized values for Table 4.7, Table 4.13 to Table 4.8 and Table 4.14 to Table 4.9.

WCM / Doc	G_1	G_2	..	..	..	..	..	..	..	G_p
D_1	0.16667	0	..	..	..	..	..	..	..	0.198324
D_2	0.16667	0	..	..	..	..	..	..	..	0.256983
..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..
D_m	0.5	0	..	..	..	..	..	..	..	0.238361

Table 4.11 An Example of 10 Terms Normalized Matrix from Table 4.6

WCM / Doc	G_1	G_2	..	..	..	..	..	..	..	G_n
D_1	0.125	0.125	..	..	..	..	..	..	..	0.198324
D_2	0.125	0	..	..	..	..	..	..	..	0.256983
..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..
D_n	0.375	0	..	..	..	..	..	..	..	0.238361

Table 4.12 An Example of 20 Terms Normalized Matrix from Table 4.7

WCM / Doc	G_1	G_2	..	..	..	..	..	..	..	G_n
D_1	0.142857	0.125	..	..	..	..	..	..	..	0.198324
D_2	0.142857	0.1875	..	..	..	..	..	..	..	0.256983
..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..
D_n	0.214286	0.3125	..	..	..	..	..	..	..	0.238361

Table 4.13 An Example of 50 Terms Normalized Matrix from Table 4.8

WCM / Doc	G_1	G_2	..	..	..	..	..	..	..	G_n
D_1	0.181818	0.125000	..	..	..	..	..	..	..	0.198324
D_2	0.227273	0.125000	..	..	..	..	..	..	..	0.256983
..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..
D_n	0.363636	0.208333	..	..	..	..	..	..	..	0.238361

Table 4.14 An Example of 100 Terms Normalized Matrix from Table 4.9

```
##################  Definitions  ##################
print "input Weight File = ";
$MatrixF=<STDIN>;
print "output Normalized File = ";
$Ofil=<STDIN>;
$array = ();
###################      MAIN     ###################
open(WeightF,">$Ofil");
$SEEin=$MatrixF;
open (IN,"<".$SEEin);              # open file in input stream
$loop = 1;
while (<IN>)
     {
        &findHighest;
     }
close (IN);
open (IN,"<".$SEEin);              # open file in input stream
while (<IN>)
     {
    &makeNormal;
     }
close (IN);
close(WeightF);
exit;

## Subroutines are excluded ##
```

Table 4.15 A Matrix Normalizing Program using Perl Language

Multi-Dimensional Scale

The original binary matrix was used as it is without grouping to draw a vector graph using Multi-Dimensional Scaling (MDS). Next, the 10 term groups of Word Code (WC), 630 groups in total for this experiment, were used to draw a vector graph using MDS. The same was done for 20 term groups (316 groups in total), 50 term groups (127 groups in total) and 100 term groups of WC (64 groups in total) were also used to draw vector graphs using MDS. A sample SAS program is shown in Table 4.13.

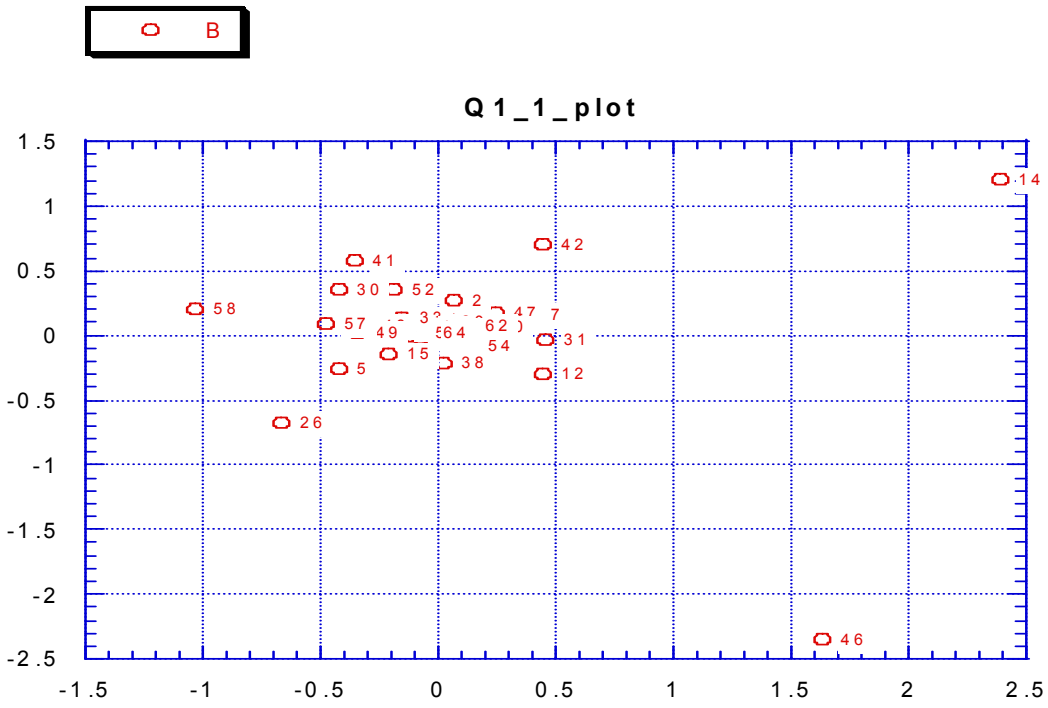

Figure 4.1 Vector Graph of Binary Matrix

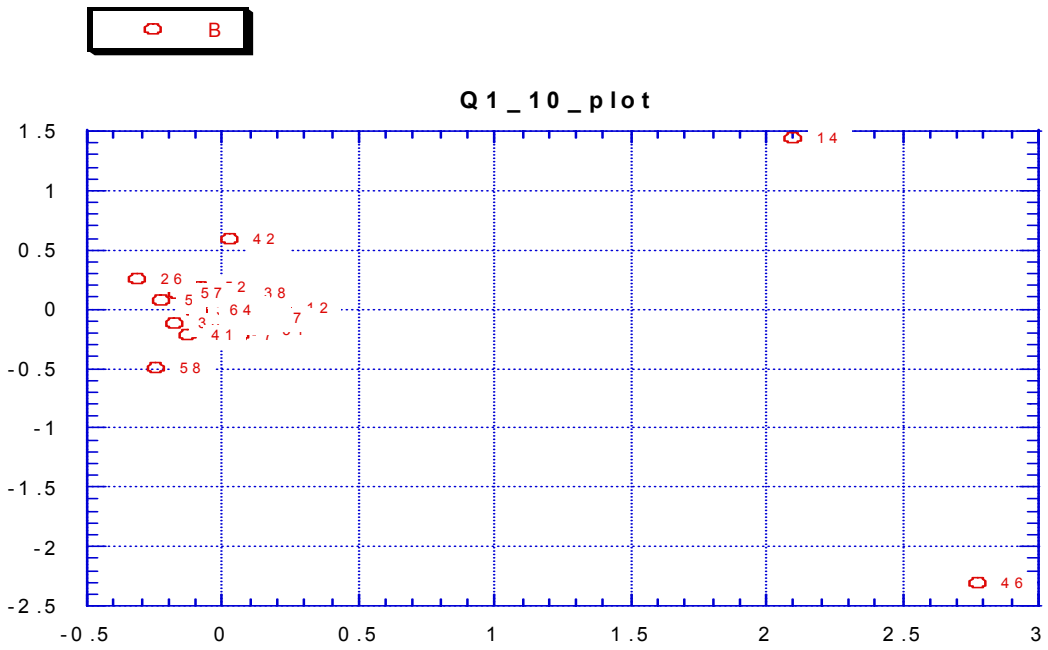

Figure 4.2 Vector Graph of 10 Term-Grouping of Word Code

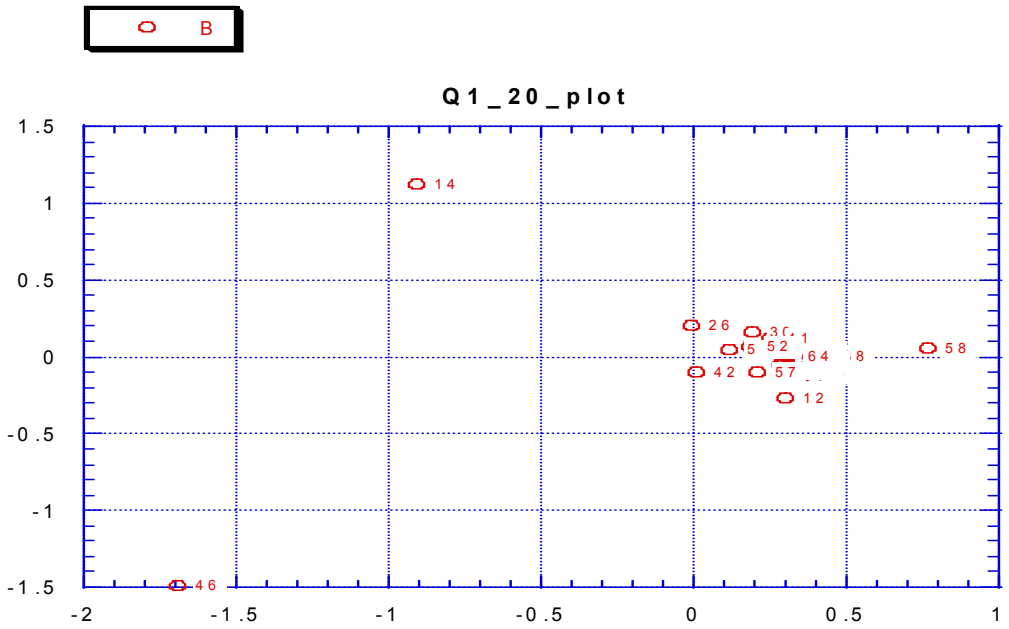

Figure 4.3 Vector Graph of 20 Term-Grouping of Word Code

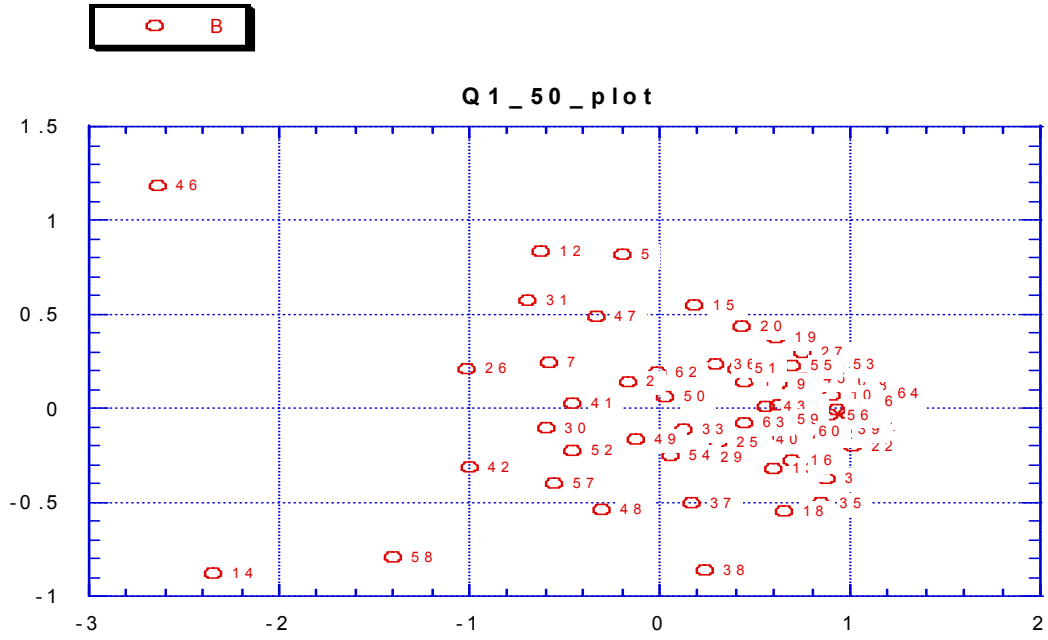

Figure 4.4 Vector Graph of 50 Term-Grouping of Word Code

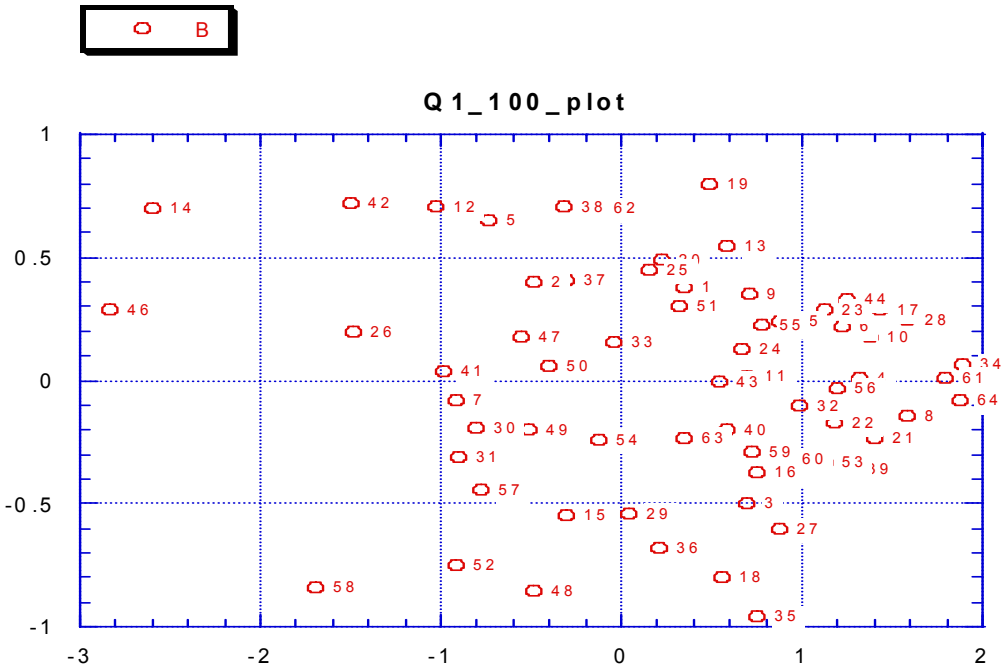

Figure 4.5 Vector Graph of 100 Term-Grouping of Word Code

Processing for Multimedia Documents of NASA

Feature Extraction

The content decomposition algorithm, which is composed of the three main components of color, shape, and texture, was chosen from among various image representation methods. All the algorithms developed for this research may be run on any industry standard image formats. The first component employed was color such as red, blue and green including grayscale extraction. Every color features have pixel values from “0” to “255” and each color feature shall have frequencies for pixel values from “0” to “255”. Those frequencies were transformed to histograms in intervals from 0-255 for red, green, blue, and gray. The red, green, and blue extractions were required before the grayscale histogram could be constructed.

The second component, shape, is composed of angle, five distances and Hough Transform value. Edge-detection method was used to get pixels from the image. Angle data were extracted from the edge-detected image ranged from zero degree to ninety degree for each pixel. Five distances such as Distance-A, Distance-B, Distance-C, Distance-D and Distance-E, and Hough Transform value were extracted from the edge-detected image. Distance-A represents the distances from the origin to the pixels. Distance-B represents the distances from the left side to the pixels. Distance-C represents the distances from the top side to the pixels. Distance-D represents the distances from the right side to the pixels. Distance-E represents the distances from the bottom side to the pixels. Hough Transform value is calculated using Hough Transform formula.

A third component, which is called "density", is also extracted from the edge-detected image. The number of edge-detected pixels is counted and then the number of pixels is translated into a single value using the following formula,

Density = Total number of edge-detected pixels / ((x-axis * y-axis) * 2).

This can also be called a texture component.

Histograms of Extracted Features

Shown below are histograms for 11 features; such as, Red, Green, Blue, Gray, Distance-A among 5 distances, Angle and Hough Transform value. Figure 4.6 is the base image used for histograms.

Figure 4.7 is the histogram for Red values.

Figure 4.8 is the histogram for Green values.

Figure 4.9 is the histogram for Blue values.

Figure 4.10 is the histogram for Gray values.

Figure 4.11 is the histogram for Distance-A values.

Figure 4.12 is the histogram for Angle values.

Figure 4.13 is the histogram for Hough values.

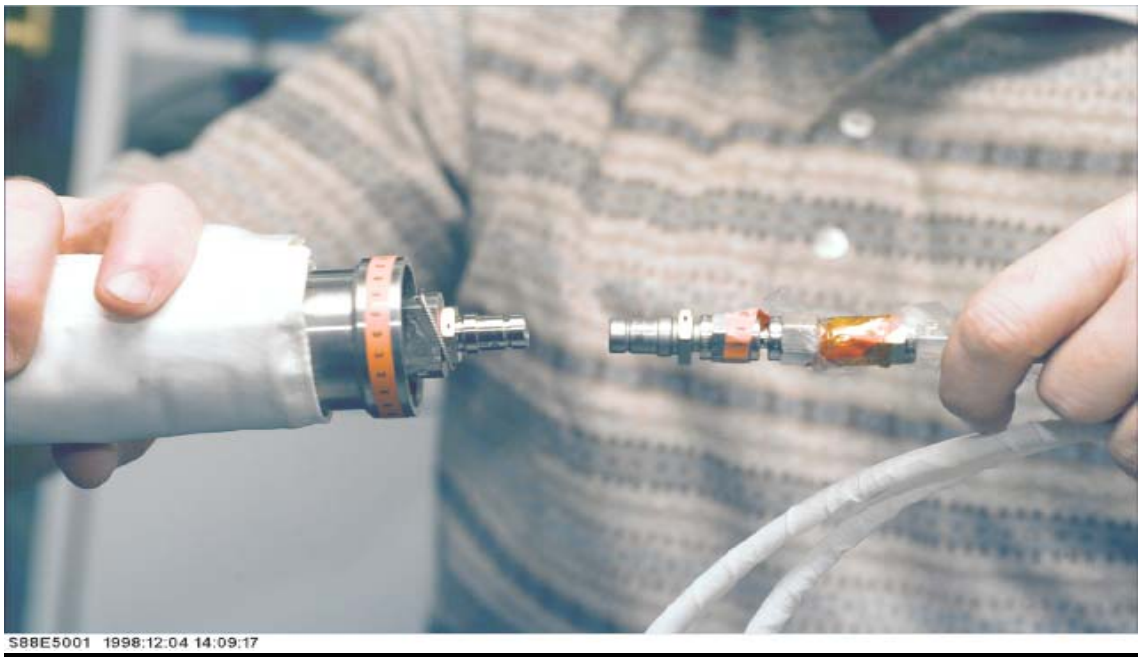

Figure 4.6 Example of NASA image for S88E5001

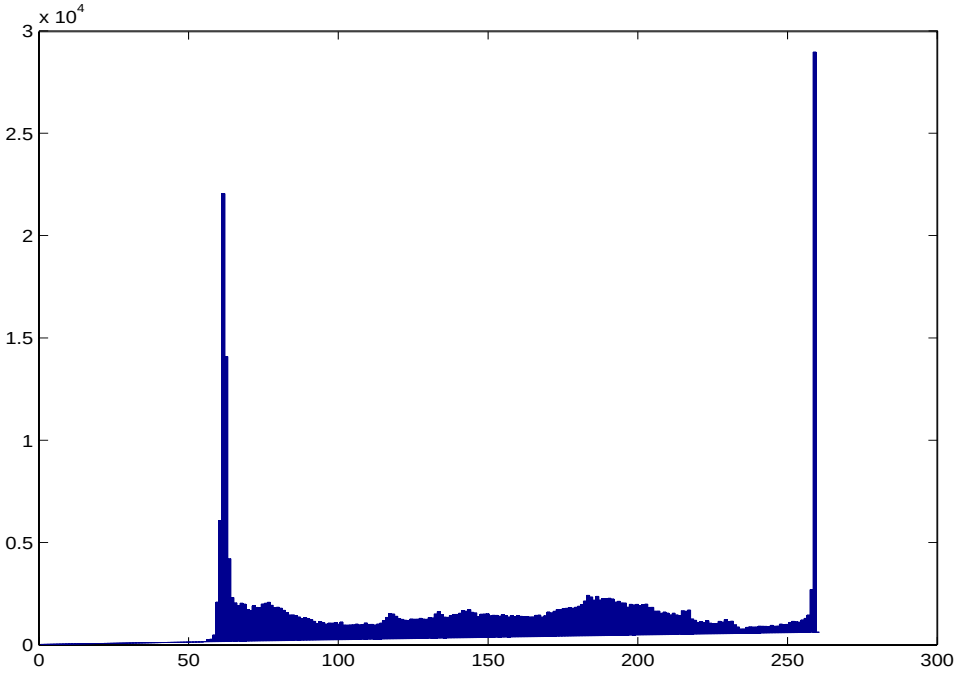

Figure 4.7 Example of the histogram for Red values

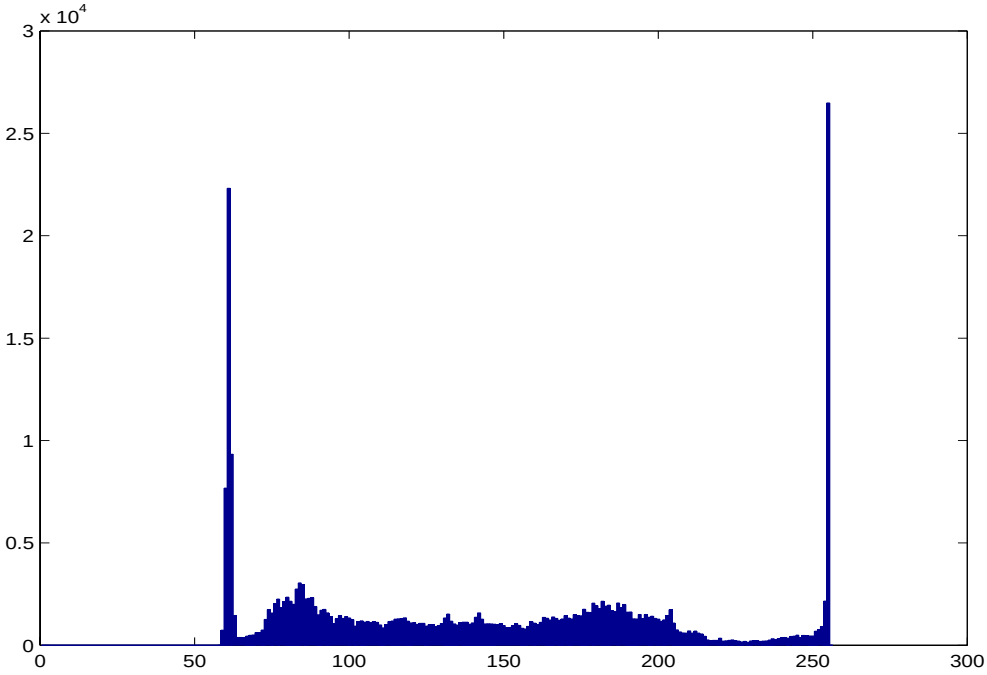

Figure 4.8 Example of the histogram for Green values

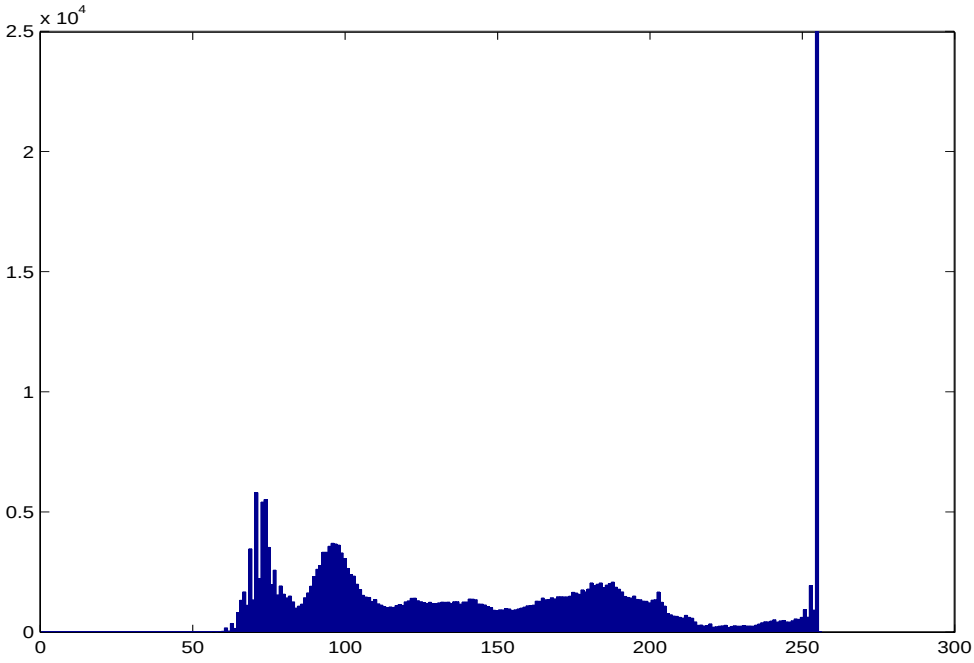

Figure 4.9 Example of the histogram for Blue values

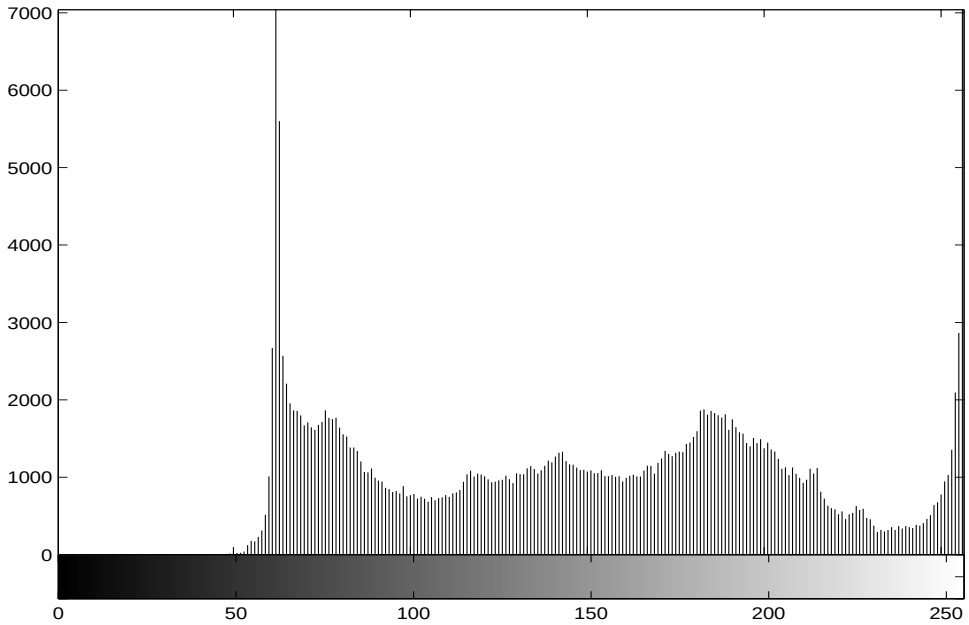

Figure 4.10 Example of the histogram for Gray values

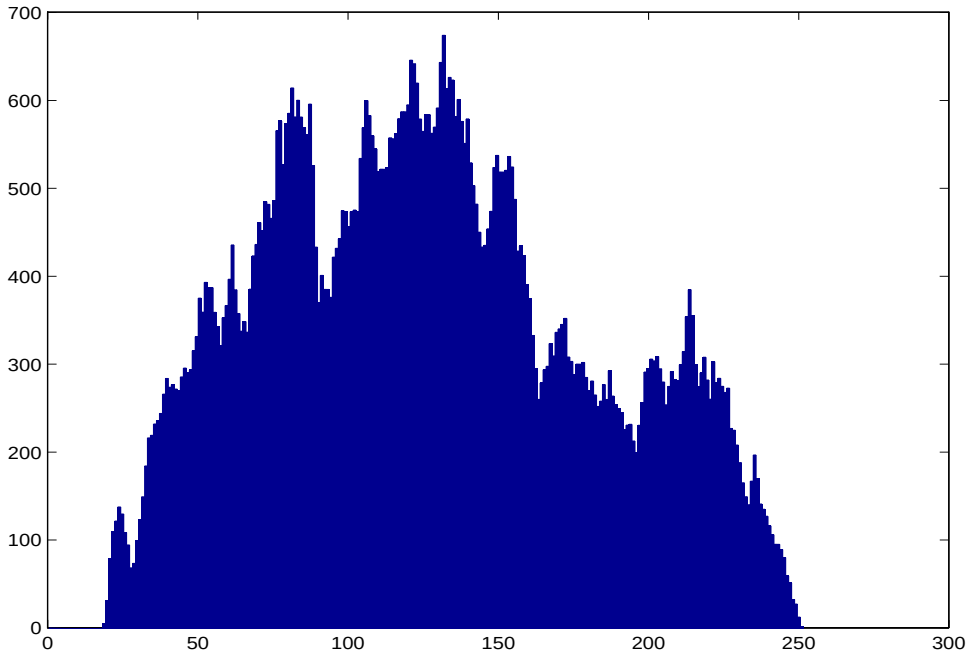

Figure 4.11 Example of the histogram for Distance-A values

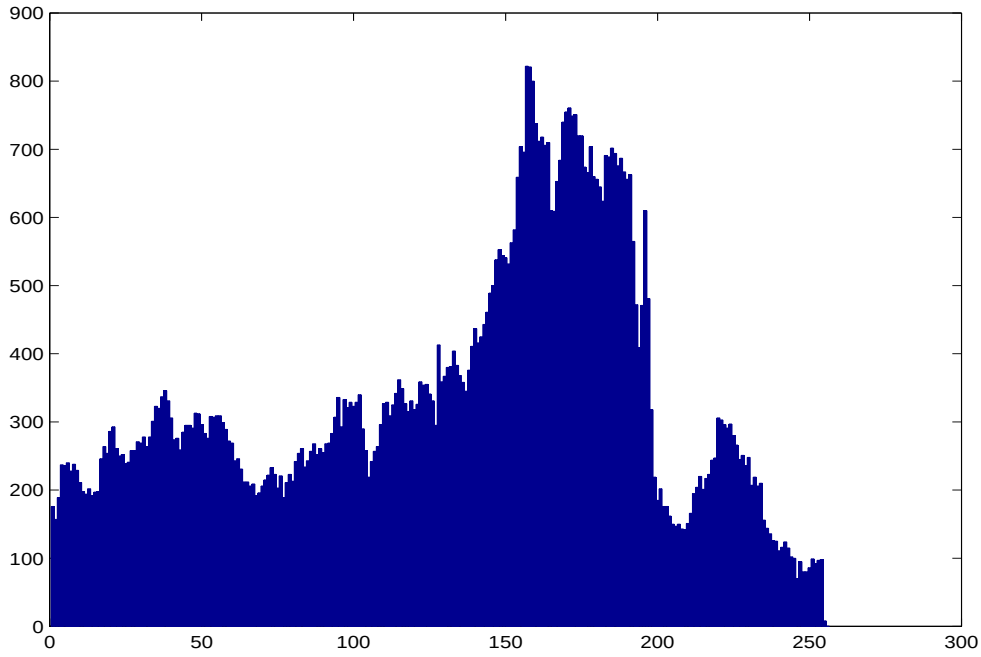

Figure 4.12 Example of the histogram for Angle values

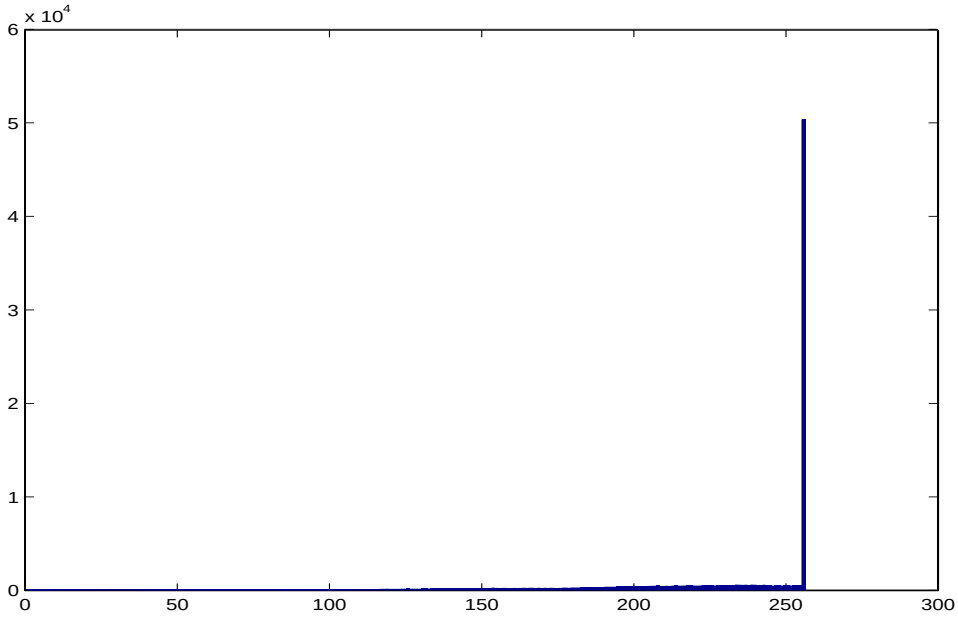

Figure 4.13 Example of the histogram for Hough Transform values

Lorenz Information Measurement

The final process of transforming the derived histograms into a single value utilized the technique known as a Lorenz Transformation (LT). This transformation allows for the unwieldy histograms of image values to be reduced to single real numbers, cutting storage space from several mega bytes per image to only 12 bytes each.

In a LT, histograms are sorted from the smallest to the largest value. The histogram is then scaled to fit into a square or rectangular depending on the height of histograms and width of intervals. This square or rectangular is divided into two triangles, and then the bottom triangle is only measured using the formula given under. The area measured was then used to represent the image property. In Figure 4.14, C_a and C_b showed as an example of Lorenz Information curve.

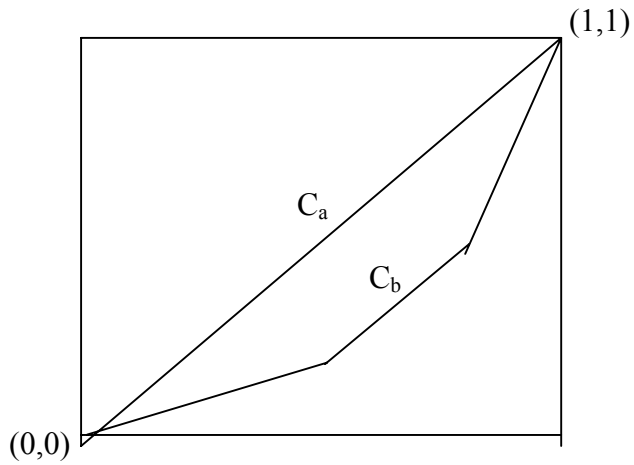

Figure 4.14 General shape of Lorenz Information curve (Chang & Yang, 1982)

Formula to compute LIM:

1) $P_0 = 0$
2) W = Width (Interval of histogram) has to be equal distribution
3) $P_i = P_{1,\ldots\ldots,}P_n$
4) $0 <= LIM (P_{1,\ldots\ldots,}P_n) <= 0.5$
5) LIM = {W*P1/2+(W*P1+W*(P2-P1)/2)+…..
   ….+(W*Pn-1+W*(Pn-Pn 1)/2)} / 2*# of pixels
   $= \{W*\sum^{n-1}P_i+W*(\sum^{n}(P_i-P_{i-1})/2)\}$ / 2*# of pixels
   $= W*(\sum^{n-1}P_i+\sum^{n}(P_i-P_{i-1})/2)$ / 2*# of pixels

An example to understand this formula by visualization is given below.

Let's assume

1) P0 = 0
2) P1 = 2, P2 = 3, P3 = 4, P4 = 5, P5 = 6
3) So, total number of pixels is 20.
4) W = 1

Then, according to the formula

LIM = 1 ((2+3+4+5) + (2+1+1+1+1)/2) / 2*20
= 17 / 40

= 0.425

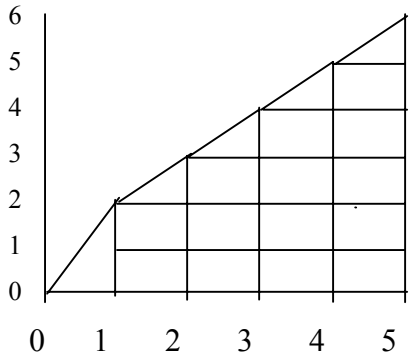

Figure 4.15 Visualized example of Lorenz Information curve

Here are the real LIM values computed by the above formula for the imageS88E5001 using a program (Table 4.17) developed based on Matlab software.

Red	Green	Blue	Gray	Hough	Dist-A	Angle
0.4915929	0.4900485	0.4908935	0.4925263	0.4077474	0.4972578	0.4982107
Dist-B	Dist-C	Dist-D	Dist-E	Density		
0.4959731	0.4979530	0.4976284	0.4978013	0.1265480		

Table 4.16 LIM values for the image S88E5001

```
% lim12.m
% This is for Lorenz Information Measurement.
% execution procedure
% matlab
% lim12
% quit
ofname=sprintf('/home/jove/fs/mer0007/sol/jktmat/data/sts88/lim12.dat');
fido = fopen(ofname,'w');
fnums = 26;
p = zeros(fnums,12);
for f = 5001:5026
if1=sprintf('/home/jove/fs/mer0007/sol/jktmat/data/sts88/hred%4.0f.dat',f);
if2=sprintf('/home/jove/fs/mer0007/sol/jktmat/data/sts88/hgreen%4.0f.dat',f);
if3=sprintf('/home/jove/fs/mer0007/sol/jktmat/data/sts88/hblue%4.0f.dat',f);
if4=sprintf('/home/jove/fs/mer0007/sol/jktmat/data/sts88/hgray%4.0f.dat',f);
if5=sprintf('/home/jove/fs/mer0007/sol/jktmat/data/sts88/hhough%4.0f.dat',f);
if6=sprintf('/home/jove/fs/mer0007/sol/jktmat/data/sts88/hdist%4.0f.dat',f);
if7=sprintf('/home/jove/fs/mer0007/sol/jktmat/data/sts88/hangle%4.0f.dat',f);
```

```
if8=sprintf('/home/jove/fs/mer0007/sol/jktmat/data/sts88/hdista%4.0f.dat',f);
if9=sprintf('/home/jove/fs/mer0007/sol/jktmat/data/sts88/hdistb%4.0f.dat',f);
if10=sprintf('/home/jove/fs/mer0007/sol/jktmat/data/sts88/hdistc%4.0f.dat',f);
if11=sprintf('/home/jove/fs/mer0007/sol/jktmat/data/sts88/hdistd%4.0f.dat',f);
if12=sprintf('/home/jove/fs/mer0007/sol/jktmat/data/sts88/mdensity%4.0f.dat',f
);
    fid1 = fopen(if1,'r');
    fid2 = fopen(if2,'r');
    fid3 = fopen(if3,'r');
    fid4 = fopen(if4,'r');
    fid5 = fopen(if5,'r');
    fid6 = fopen(if6,'r');
    fid7 = fopen(if7,'r');
    fid8 = fopen(if8,'r');
    fid9 = fopen(if9,'r');
    fid10 = fopen(if10,'r');
    fid11 = fopen(if11,'r');
    fid12 = fopen(if12,'r');
    x = 256;
    w = 1/x;
    d1 = fscanf(fid1,'%6d');
    d2 = fscanf(fid2,'%6d');
    d3 = fscanf(fid3,'%6d');
    d4 = fscanf(fid4,'%6d');
    d5 = fscanf(fid5,'%6d');
    d6 = fscanf(fid6,'%6d');
    d7 = fscanf(fid7,'%6d');
    d8 = fscanf(fid8,'%6d');
    d9 = fscanf(fid9,'%6d');
    d10 = fscanf(fid10,'%6d');
    d11 = fscanf(fid11,'%6d');
    d12 = fscanf(fid12,'%6d');
    sd1 = sort(d1);
    sd2 = sort(d2);
    sd3 = sort(d3);
    sd4 = sort(d4);
    sd5 = sort(d5);
    sd6 = sort(d6);
    sd7 = sort(d7);
    sd8 = sort(d8);
    sd9 = sort(d9);
    sd10 = sort(d10);
    sd11 = sort(d11);
    t = zeros(11);
    for i = 1:256
        t(1) = t(1) + sd1(i);
        t(2) = t(2) + sd2(i);
        t(3) = t(3) + sd3(i);
        t(4) = t(4) + sd4(i);
        t(5) = t(5) + sd5(i);
        t(6) = t(6) + sd6(i);
        t(7) = t(7) + sd7(i);
        t(8) = t(8) + sd8(i);
        t(9) = t(9) + sd9(i);
        t(10) = t(10) + sd10(i);
        t(11) = t(11) + sd11(i);
    end
    j = f - 5000;
    for i = 1:256
        p(j,1) = p(j,1) + (x - i) * sd1(i) / t(1);
        p(j,2) = p(j,2) + (x - i) * sd2(i) / t(2);
        p(j,3) = p(j,3) + (x - i) * sd3(i) / t(3);
        p(j,4) = p(j,4) + (x - i) * sd4(i) / t(4);
        p(j,5) = p(j,5) + (x - i) * sd5(i) / t(5);
```

```
        p(j,6) = p(j,6) + (x - i) * sd6(i) / t(6);
        p(j,7) = p(j,7) + (x - i) * sd7(i) / t(7);
        p(j,8) = p(j,8) + (x - i) * sd8(i) / t(8);
        p(j,9) = p(j,9) + (x - i) * sd9(i) / t(9);
        p(j,10) = p(j,10) + (x - i) * sd10(i) / t(10);
        p(j,11) = p(j,11) + (x - i) * sd11(i) / t(11);
    end
    for i = 1:11
        p(j,i) = w * (p(j,i) + 0.5);
    end
% for density
    p(j,12) = d12 / 320000;
    fclose(fid1);
    fclose(fid2);
    fclose(fid3);
    fclose(fid4);
    fclose(fid5);
    fclose(fid6);
    fclose(fid7);
    fclose(fid8);
    fclose(fid9);
    fclose(fid10);
    fclose(fid11);
    fclose(fid12);
end
for i = 1:fnums
    for j = 1:12
        fprintf(fido,'%10.7f  ',p(i,j));
    end
    fprintf(fido,'\n');
end
fclose(fido);
```

Table 4.17 Sample program to calculate LIM values for STS-88 26 images

Multi-Dimensional Scale

Those 12 content-based image features are grouped into 6 distinguished feature groups; such as, 1) RGB (red, green and blue) content-based image feature (CBIF), 2) Gray CBIF, 3) Hough CBIF, 4) Distance (distance-A, distance-B, distance-C, distance-D, distance-E) CBIF, 5) Angle CBIF, and 6) Density CBIF. From the six feature groups we can get six combination groups, and from the six combination groups 63 combinations should be derived according to these formulas: 1) ${}_6C_1 = 6$ 2) ${}_6C_2 = 15$ 3) ${}_6C_3 = 20$ 4) ${}_6C_4 = 15$ 5) ${}_6C_5 = 6$ 6) ${}_6C_6 = 1$. For combination group 1 (${}_6C_1 = 6$), six vector graphs were drawn. For combination group 2 (${}_6C_2 = 15$), fifteen

vector graphs were drawn. For combination group 3 (${}_6C_3 = 20$), twenty vector graphs were drawn. For combination group 4 (${}_6C_4 = 15$), fifteen vector graphs were drawn. For combination group 5 (${}_6C_5 = 6$), six vector graphs were drawn. For combination group 6 (${}_6C_6 = 1$), one vector graph was drawn. After that each vector graph was analyzed for choosing the best shape in clustering. It was assumed that similar images might be clustered closely and combination group 6 was comparatively well distributed and clustered vector graph among all vector graphs.

The following vector graphs were drawn by Statistical Analysis Software™ (SAS) using the multi-dimensional scaling method (MDS), using data provide by the LIM extracted from the NASA images. Figure 4.16 through figure 4.21 are representations for each combination group of the first test set STS-88 (26 images). In addition, Figure 4.22 is representing combination group six for 105 images and figure 4.23 is also representing combination group six for 994 images.

Figure 4.16 represents combination group one among 6 vector graphs. Figure 4.17 represents combination group two among 15 vector graphs. Figure 4.18 represents combination group three among 20 vector graphs. Figure 4.19 represents combination group four among 15 vector graphs. Figure 4.20 represents combination group four among 6 vector graphs. Figure 4.21 represents combination group six.

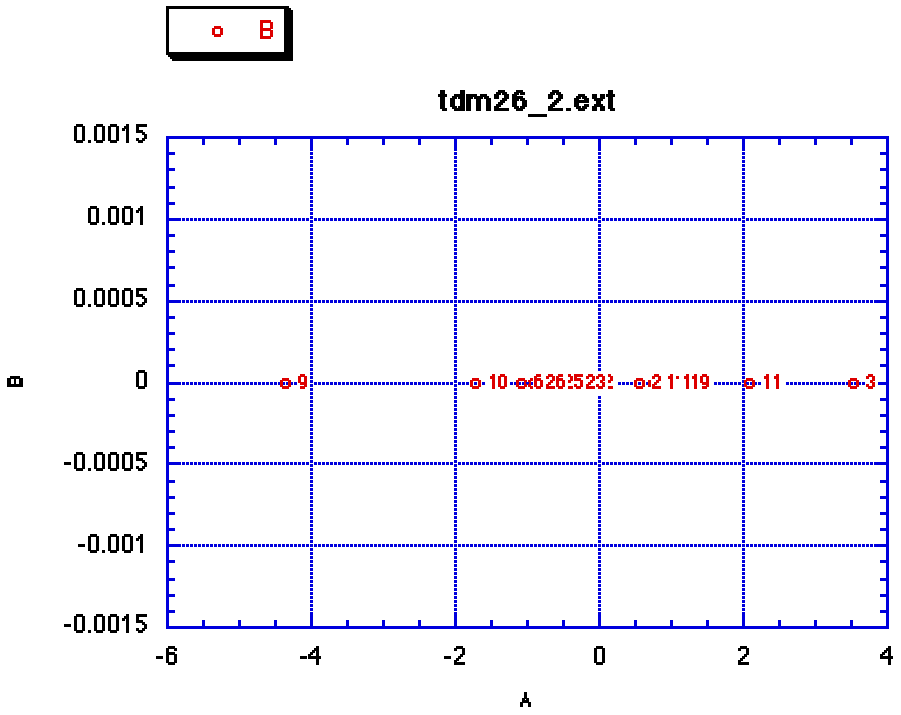

Figure 4.16 Vector graph for combination group one (STS-88)

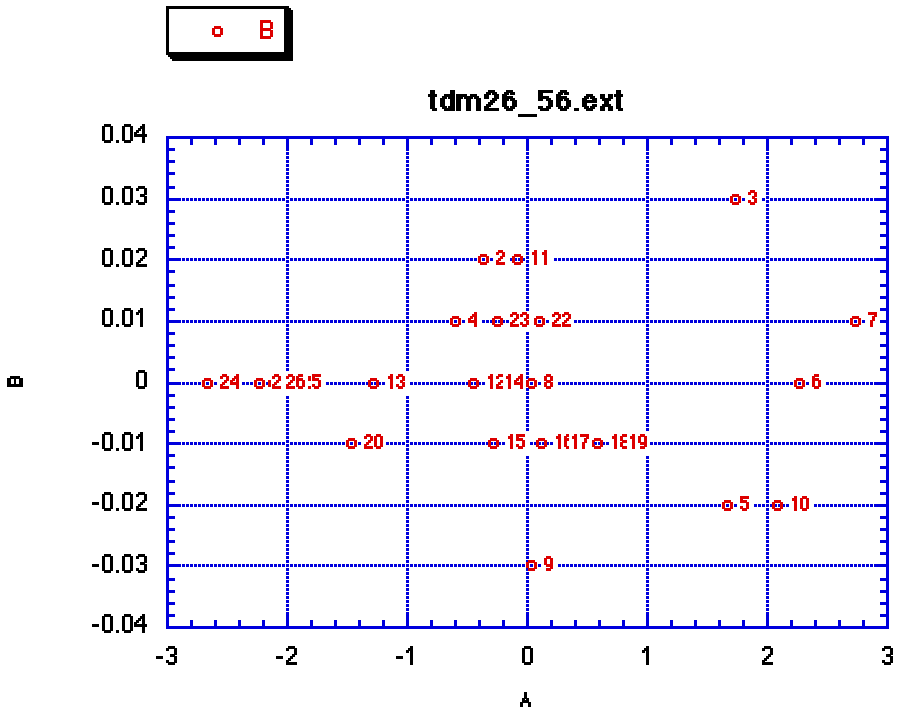

Figure 4.17 Vector graph for combination group two (STS-88)

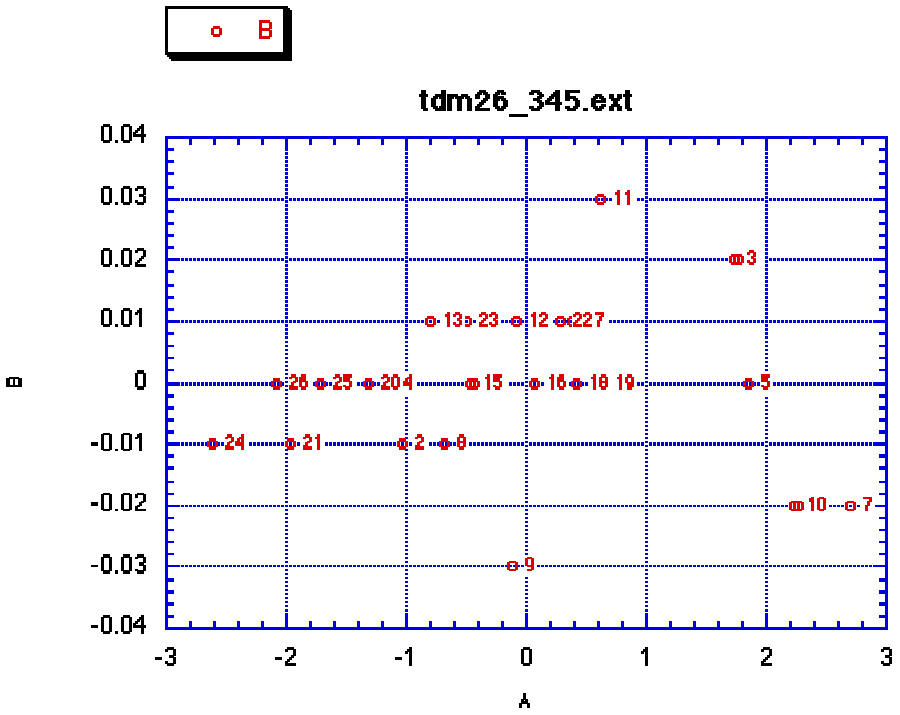

Figure 4.18 Vector graph for combination group three (STS-88)

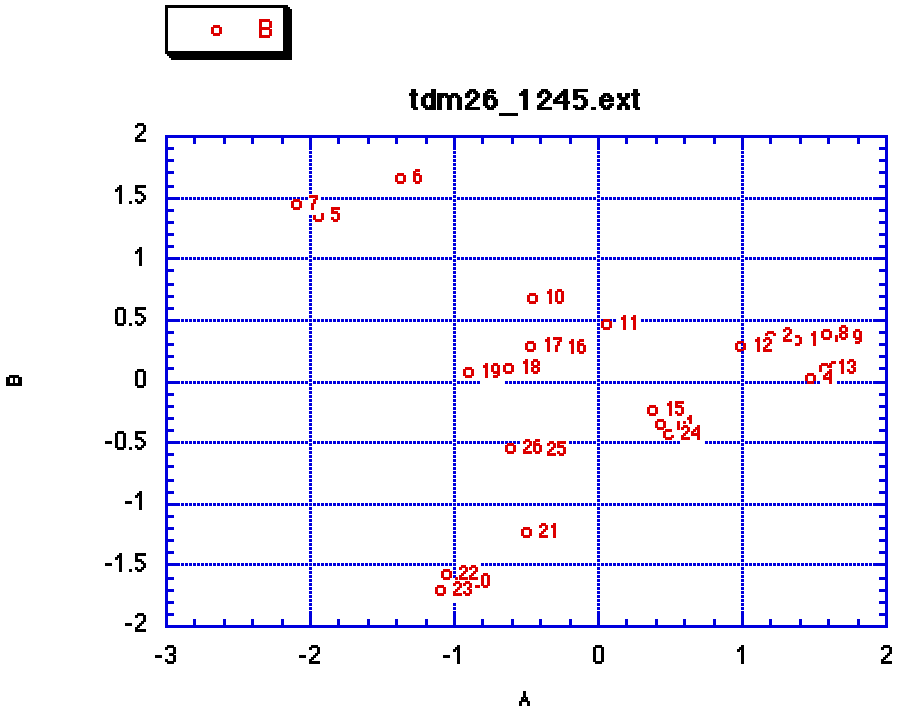

Figure 4.19 Vector graph for combination group four (STS-88)

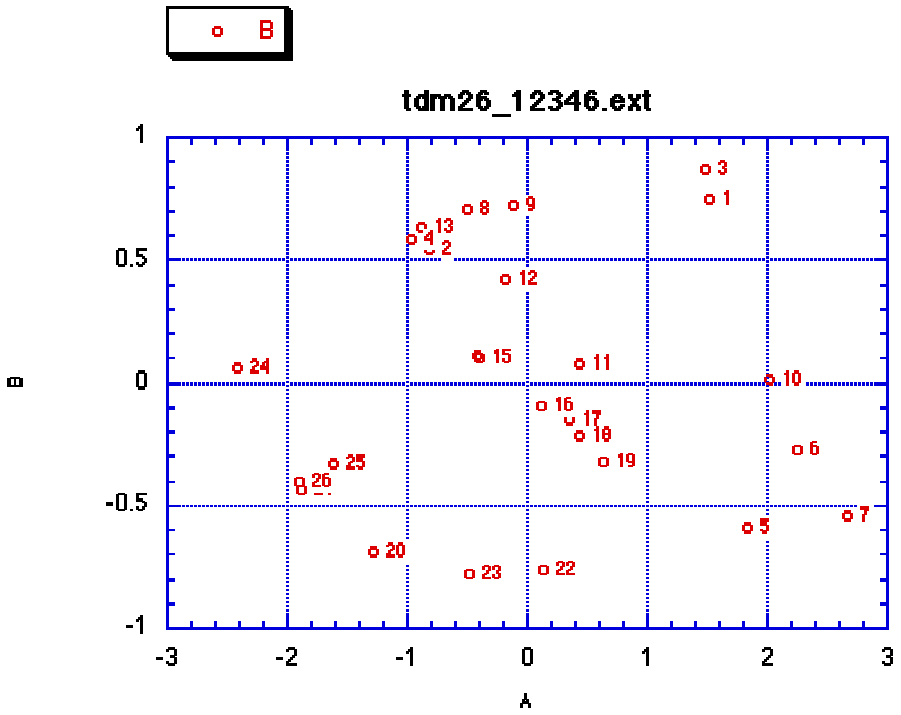

Figure 4.20 Vector graph for combination group five (STS-88)

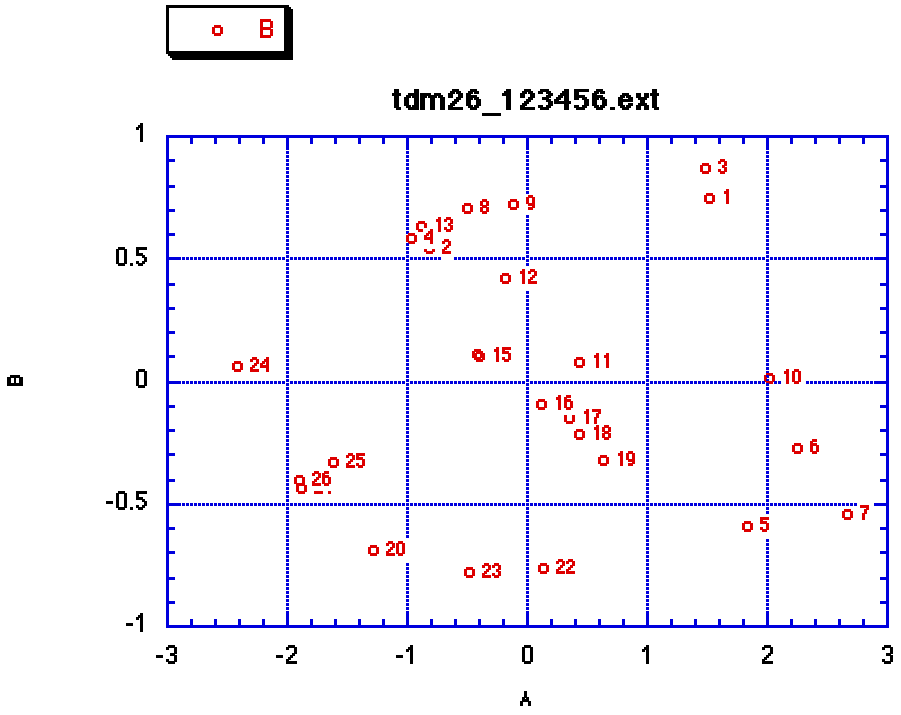

Figure 4.21 Vector graph for combination group six (STS-88)

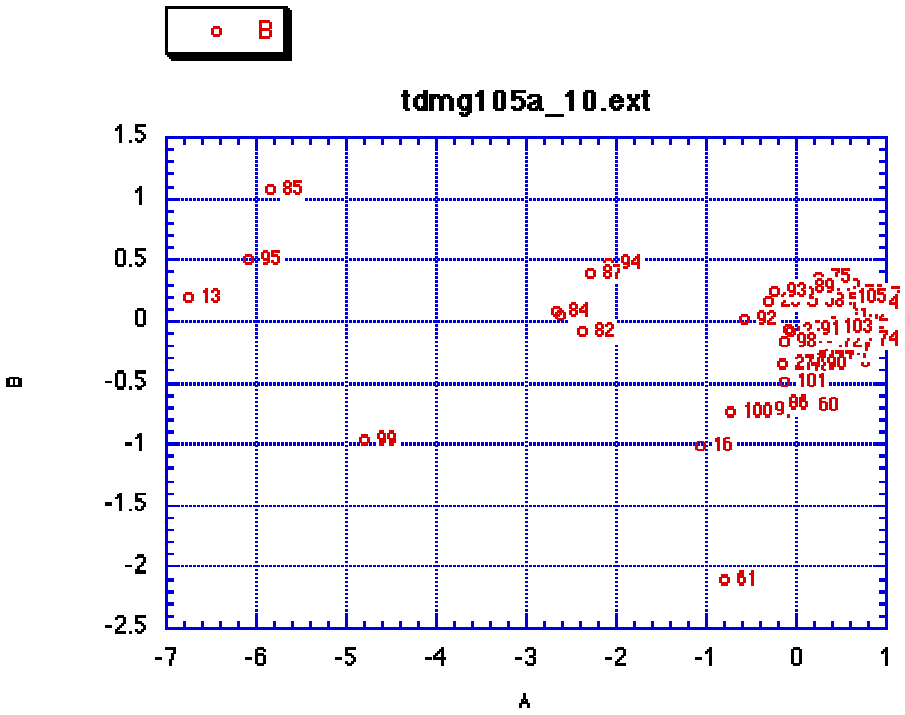

Figure 4.22 Vector graph for combination group six (STS-96)

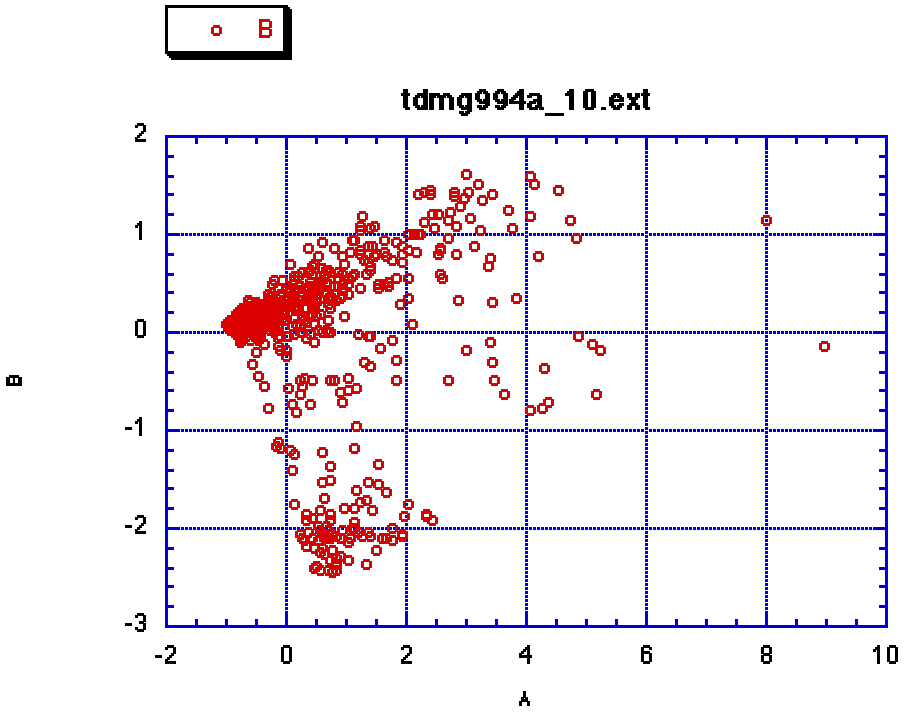

Figure 4.23 Vector graph for combination group six (STS-82)

Here are six more vector graphs. Figure 4.24 shows STS-88 26 text documents. There are only six vectors because many images in STS-88 use same text document to explain each image. Figure 4.25 shows STS-96 105 text documents and Figure 4.26 shows STS-82 994 text documents. There are not many vectors for STS-96 and STS-82 for the same reason like STS-88.

On the other hand, Figure 4.27 shows combined vector graph for STS-88 26 multimedia documents, Figure 4.28 shows combined vector graph for STS-96 105 multimedia documents and Figure 4.29 shows combined vector graph for STS-82 994 multimedia documents. These graphs are showing better distributed and clustered than not combined graphs, image only or text only graphs. Afterwards, a sample SAS program, which will draw the X-Y axis for a vector graph, is shown in Table 4.18.

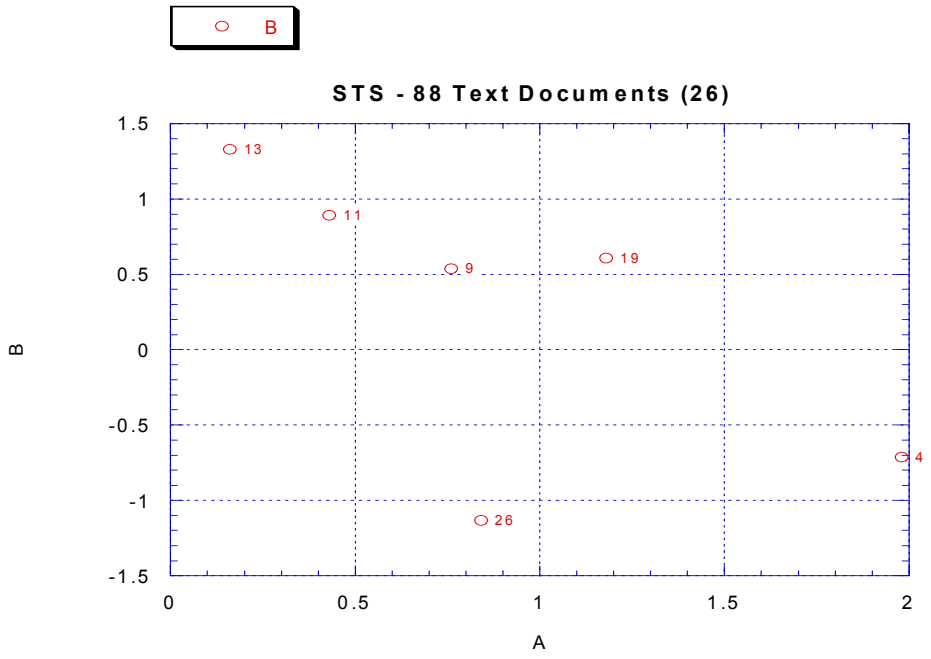

Figure 4.24 STS-88 Text Documents (26)

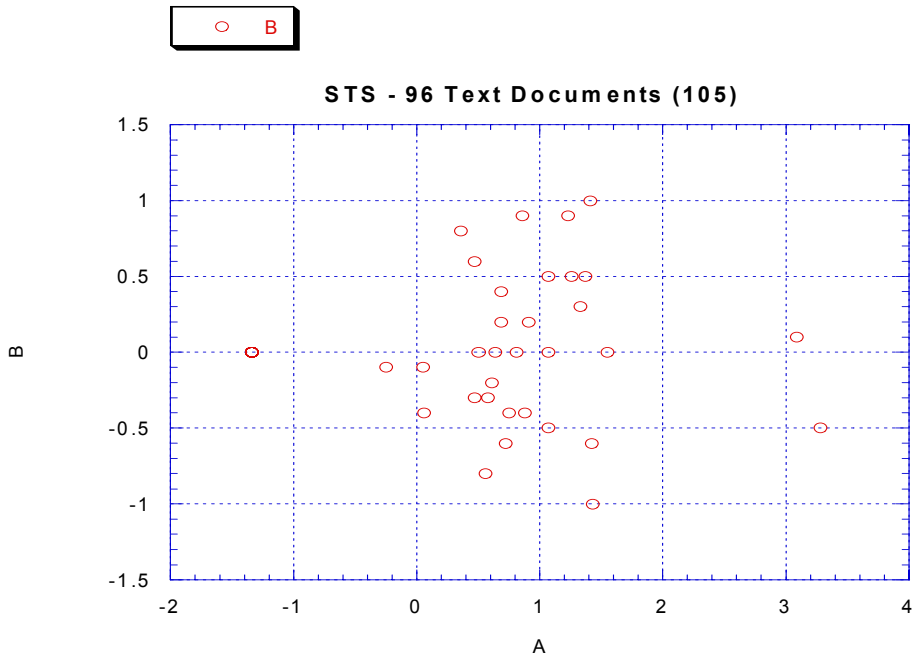

Figure 4.25 STS-96 Text Documents (105)

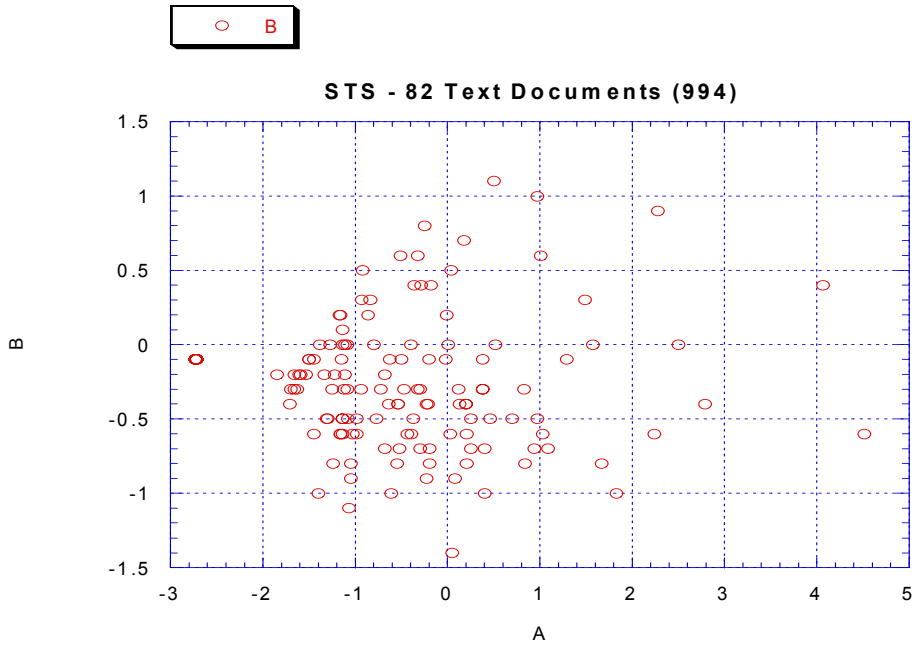

Figure 4.26 STS-82 Text Documents (994)

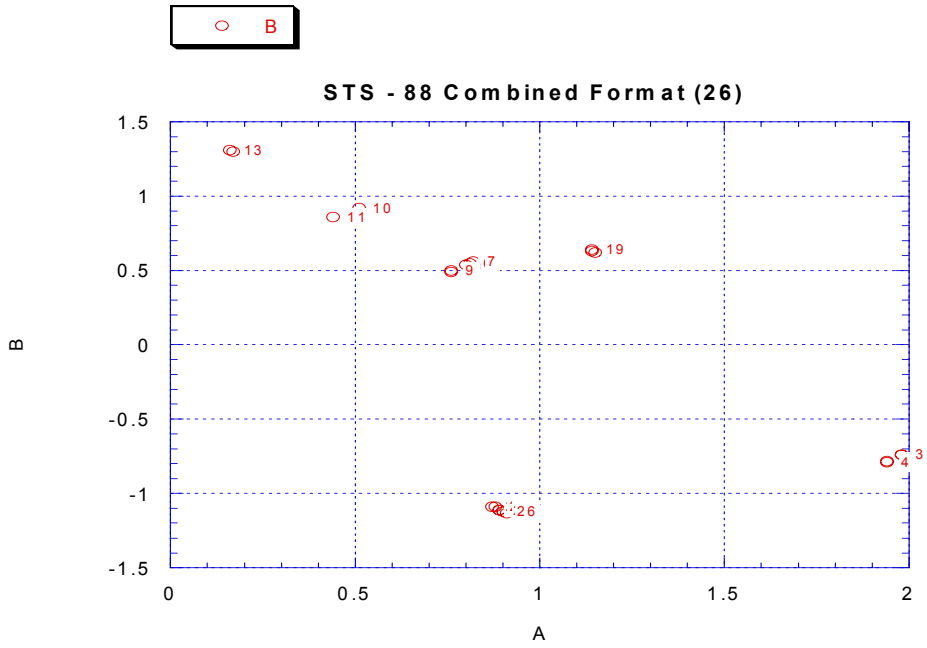

Figure 4.27 Combined Format for STS-88 Multimedia Documents (26)

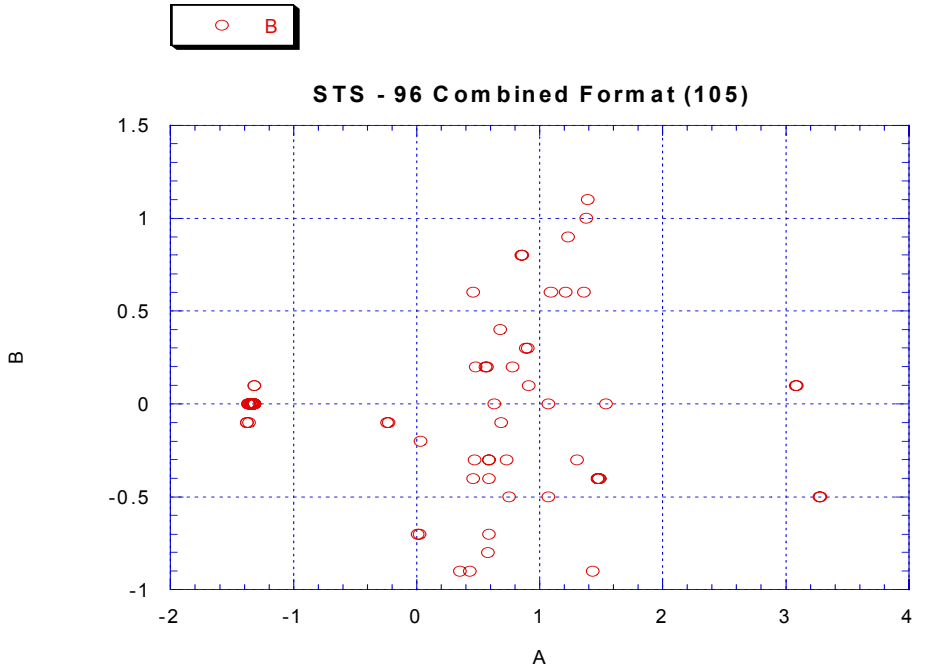

Figure 4.28 Combined Format for STS-96 Multimedia Documents (105)

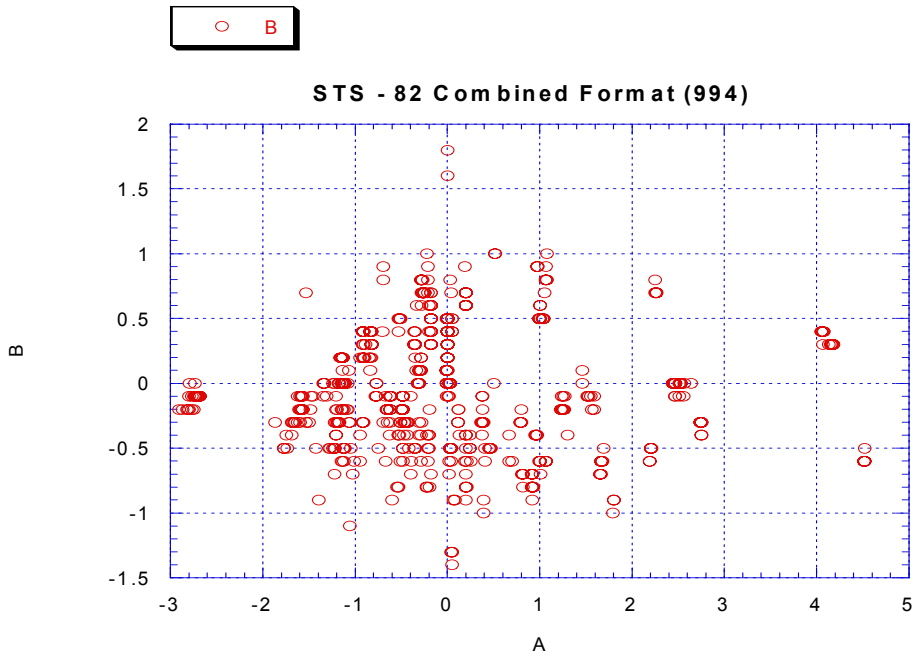

Figure 4.29 Combined Format for STS-82 Multimedia Documents (994)

```
filename io '/export/olddata/mrorvig/jktmat/data/sts88';
data d1;
   infile io(lim12.dat);
   input v1-v12;

proc print;


/***** MACRO TO COMPUTE SQUARED EUCLIDEAN DISTANCES
******************/

%MACRO E_DIST(RC=C, DATA=_LAST_, OUT=DISTOUT);
   PROC IML; XZX=0;
      USE &DATA; SETIN &DATA;
      READ ALL VAR _NUM_ INTO M(|COLNAME=V_NAMES|);
      CLOSE &DATA;
     %IF &RC=R %THEN %DO; M=M`; %END;
      NC=NCOL(M); D=J(NC,NC,0);
      START EDIST;
         DO I=1 TO NC-1; DO J=I+1 TO NC;
            D(|I,J|)=SSQ( M(|,I|)-M(|,J|) ); D(|J,I|)=D(|I,J|);
         END; END;
      FINISH;
      RUN EDIST;
     %IF &RC=C %THEN %DO;
         CREATE &OUT FROM D(|COLNAME=V_NAMES|);
     %END;
     %ELSE %DO; CREATE &OUT FROM D; %END;

      APPEND FROM D; CLOSE &OUT;
%MEND;

/*********************************************************************
**/


%e_dist(DATA=D1, RC=R, OUT=DIST);

 proc mds data=DIST
  level=loginterval
  dimension=2
  pfinal
  out=coord;
 run;

proc print data=coord;

 proc plot data=coord vtoh=1.7;
 PLOT DIM2*DIM1='*' $ _name_
    / box haxis=by .5 vaxis=by .5;
    where _type_='CONFIG';
run;
```

Table 4.18 A Sample SAS Program to get MDS Coordination for STS-88 LIM 12

Flowcharts for Multimedia Documents Processing

Two flowcharts, one for image processing and the other for text processing, are shown below in Figure 4.30 and Figure 4.31. A flow chart is one of the most efficient ways to explain the procedure of the job from beginning to end.

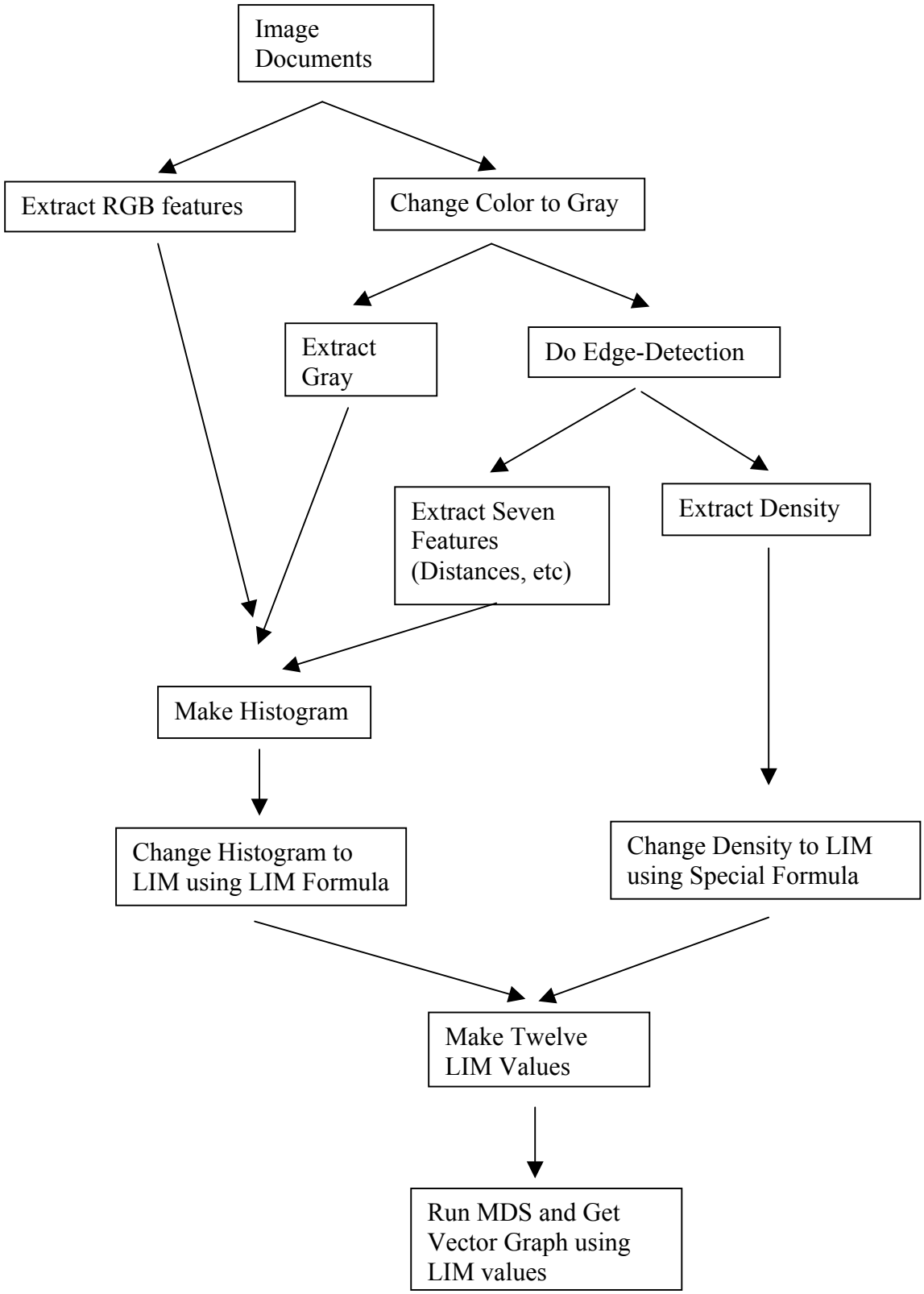

Figure 4.30 Flowchart for Image Document Process

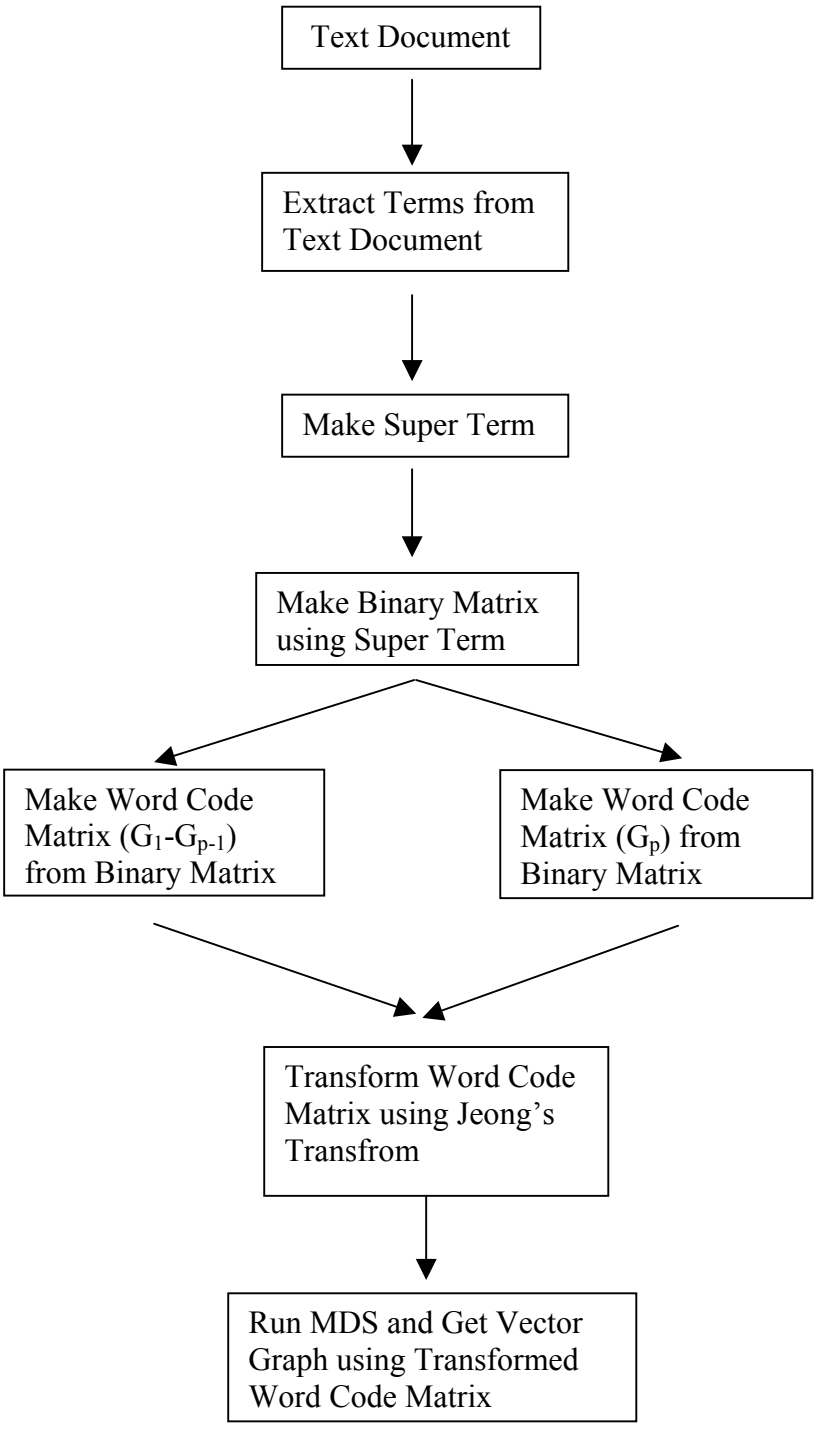

Figure 4.31 Flowchart for Text Document Process

CHAPTER 5. FINDINGS AND ANALYSIS

Introduction

A binary term-document matrix (Table 4.4) made from a set of text documents is more widely used in the processing of text documents than the process of directly using words from the text document. The binary term-document matrix does not combine well with image features as derived here, so a new format has to be used.

A grouping method was used so that there are as many text features as there are image features. The weight of each group shown in Table 4.6 through Table 4.9 represents the frequencies of 1's in that group. Then, the highest weight of the group must be identified for normalization. Finally, the weight of each group for each document is normalized so that all normalized values are less than or equal to 0.5. The reason for this procedure is so that the text features are to scale within image features utilizing the Lorenz Information Measurement (LIM). The entire processes for converting these text documents is named Jeong's Transform (JT).

Vector representation was used because the closeness of two vectors on the X-Y axis can be easily visualized. In vector representation, each vector stands for one of the images and its related text descriptions. Therefore, the distance between vectors could be described as how close the images and text descriptions were to each other. The process of vector representation was possible using the multi-dimensional scaling (MDS) method in statistical analysis software (SAS).

In analyzing the vectors a clustering method is used. If the vectors are clustered, it is said that those documents represented by the clustered vectors are very closely related. For this analysis, three sets of vector representations, namely a text

vector representation, an image vector representation, and a combined vector representation (image and text), were used, as shown on the graphs below. The combined vector representation from the National Aeronautics and Space Administration (NASA) image and text description set on the Hubble telescope repairing mission STS-88 strongly represents the closeness of the similar image and text documents. (Refer Figure 4.27)

Precision and recall measurements in Table 5.13 do also strongly support that using combined representation format in multimedia retrieval system is very valuable choice.

Findings

Development of Jeong's Transform

In processing text documents, a binary matrix shown in Table 4.4 is often used instead of directly using words from a text document. In this research, it was chosen particularly in order to combine totally different types of documents into one data structure that is called "A Common Representation Format". To create a balance between text documents and image documents and to combine these two features, the binary matrix has to be transformed to a new format that was not previously known.

To solve the problem of combining two different types of documents such as text documents and image documents, a grouping method is used so that there are 12 text features, as there are 12 image features currently. The weight of each group in the Word Code matrix represents the frequencies of 1's in that group for each document. Table 4.6 represents the weight of 10 term-groups, Table 4.7 represents the weight of 20 term-groups, Table 4.8 represents 50 term-groups and Table 4.9 represents 100 term-groups. After that the weight of each group has to be normalized

again using the highest weight of that group. Finally, the weight of each group is normalized and the normalized values are less than or equal to 0.5. The values of Lorenz Information Measurement (LIM) should be less than or equal to 0.5 according to the Lorenz theory. This is well explained in Chapter 4.3.3. The entire process of changing from a binary term-document matrix to a 12 Word Code matrix for text documents and combining 12 text measurements and 12 image measurements is named Jeong's Transform.

Test Set Analysis

Clustering of Vectors

Visualization in Information Retrieval is attractive, especially when it is possible to draw documents into a visualized vector graph. For this research, there were four categories for image decomposition -- color, distance, angle, and texture. For the visual comparison of vector graphs, fifteen combinations of these four categories were made without text measurements and fifteen combinations of these four categories were made with text measurements as shown in the Table 3.4 (The Combinations of LIM and Text Measurements). The vector graphs of thirty combinations are then drawn to find the combination that most reliably measures the distances between vectors that are representing documents.

Evaluation is done through the clustering method to support the precision and recall in Table 5.13. Human eyes can measure the distances between vectors and the distances represent the closeness between the vectors. In addition, if the vectors are clustered in any position on X-Y axis, the clustered vectors represent closely related documents. When the first test set, STS-88 26 images and text documents, are used, some clustering groups appeared in Figure 5.2, and are well clustered. Figure 5.1

shows the example images in that cluster. In Figure 5.2, the vectors representing multimedia documents, number 1 through 26, are shown. Figure 4.21 shows the image only vector graph and Figure 4.24 shows the text only vector graph. The combined vector graph of STS-96 105 multimedia documents is well clustered in comparison to the image only vector graph or the text only vector graph. Each vector graph is shown in Figure 4.22, Figure 4.25 and Figure 4.28. Also, the combined vector graph of STS-82 994 multimedia documents is well clustered in comparison to the image only vector graph or the text only vector graph. Each vector graph is shown in Figure 4.23, Figure 4.26 and Figure 4.29. For three test sets, STS-88 (26 multimedia documents), STS-96 (105 multimedia documents) and STS-82 (994 multimedia documents), combined vector graphs are all better distributed and clustered than image only vector graphs or text only vector graphs.

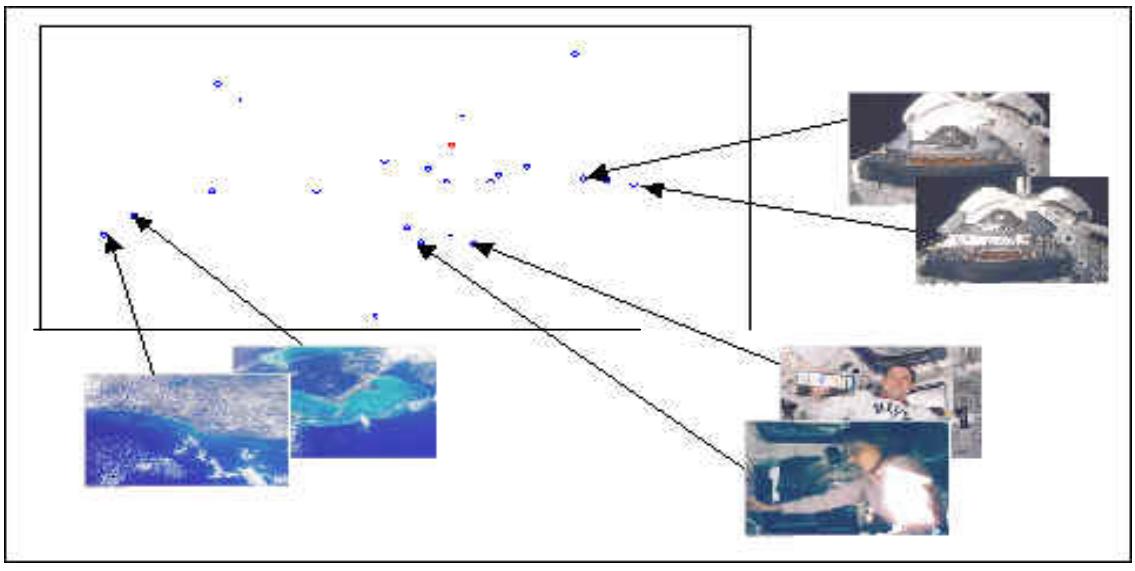

Figure 5.1 Display of the classification of images only mapped from STS88

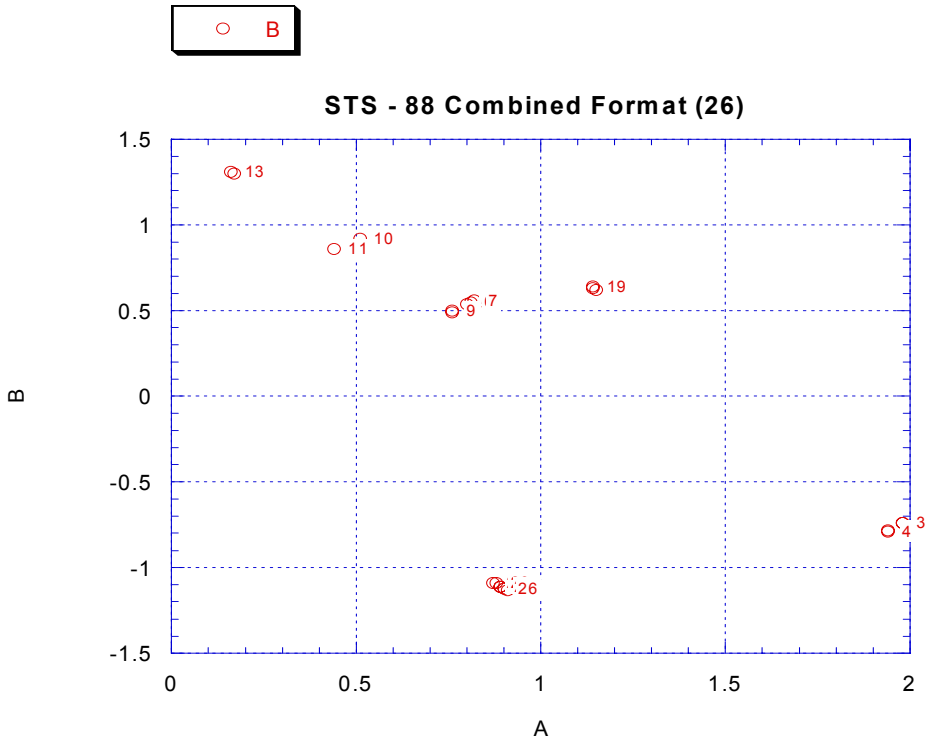

Figure 5.2 Display of the classification of images and text combined mapped from STS88

Precision and Recall based on Heuristic Judgment

Precision and recall measurements are very useful methods used to verify how the vector representation is correctly explaining the closeness between multimedia documents and how reliable the retrieval system is.

Precision is defined as the proportion of retrieved documents that are relevant,

$$P = w / n2.$$

Recall is defined as the proportion of relevant documents that are retrieved,

$$R = w / n1.$$

Where n1, n2, w, x, y and z are as follows.

From the table;

$n1 = w + x$

$n2 = w + y$

$n = w + x + y + z$

	Retrieved	Not retrieved
Relevant	w	x
Not relevant	y	z

Table 5.1 Contingency Table for Evaluating Retrieval

To compute precision and recall, 26 images are chosen. A table to show the similarity (Table 5.2) between the images is made by human. To test the result of this research, two testing methods are used. The first method retrieves the images using a term from the text documents, and records the retrieved results in the table (Table 5.3) under threshold 0.2, and calculates the precision and recall according to the formula, which is given the above. In here, threshold is used for the boundary of similarity between documents in the retrieval system. The first method is only using image document, which have 12 measurements. The second method is using image and text multimedia document, which have 24 measurements. Each one has 12 measurements. The third test uses the same method as the first one except that the threshold is changed to 0.1. The fourth one uses the same method as the second one except that the threshold is changed to 0.1.

The Brighton Image Searcher handles the process of retrieval for multimedia documents. For the first and the third method http://archive4.lis.unt.edu/td26/www is

used and for the second and the fourth http://archive4.lis.unt.edu/tdt26/www is used. The similarity of image number 1 through 26 is decided by tester's heuristics and recorded below the tables. For Table 5.2, Table 5.3, Table 5.5, Table 5.7 and Table 5.9, the numbers in the row represent document number and the numbers in the column also represent document number, which is similar to the row document number if it is marked as "0". Table 5.2 shows the similarities between images over 26 Images by human heuristics. For example, document number 1 in row 1 has five ovals, document number 1, 3, 6, 8 and 9. So these five documents might be retrieved if any one of five documents is given.

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

Table 5.2 A Similarity Table on 26 Images by Human Heuristics

Table 5.3 shows the retrieved results on 26 images only under threshold 0.2 using the Brighton Image Searcher (http://archive4.lis.unt.edu/td26/www) designed and implemented for this research. When image number 1 in row 1 was given, 25 images were retrieved except image number 10. This result means that image only measurements are not favorable for using alone.

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

Table 5.3 Retrieved Results on 26 Images only under Threshold 0.2

The contingency table (Table 5.4) is calculated from Table 5.2 and Table 5.3, and the precision and the recall is also calculated from the contingency table (Table 5.4). Precision and Recall for 26 images only under threshold 0.2 is calculated like the below.

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

P = 96 / 530

P = 0.181132

So, precision is approximately 18.11 %.

Recall = (Retrieved and Relevant) / ((Retrieved and Relevant) + (Not Retrieved but Relevant))

R = 96 / 96

R = 1

So, recall is 100 %.

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

Table 5.4 Contingency Table for 26 Images only under Threshold 0.2

Table 5.5 shows the retrieved results on 26 image and text combined measurements under threshold 0.2 using the Brighton Image Searcher (http://archive4.lis.unt.edu/tdt26/www) designed and implemented for this research. When image number 1 in row 1 was given, 6 images were retrieved. This result

means that the images retrieved are including the images driven by human heuristics on Table 5.2. These are very favorable results.

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

Table 5.5 Retrieved Results on 26 Image and Text Combined under Threshold 0.2

The contingency table (Table 5.6) is calculated from Table 5.2 and Table 5.5, and the precision and the recall is also calculated from the contingency table (Table 5.6). Precision and Recall for 26 images only under threshold 0.2 is calculated like the below.

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

P = 68 / 100

P = 0.68

So, precision is approximately 68 %.

Recall = (Retrieved and Relevant) / ((Retrieved and Relevant) + (Not Retrieved but Relevant))

R = 68 / 96

R = 0.7083333

So, recall is 70.83 %.

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

Table 5.6 Contingency Table for 26 Images and Text Combined under Threshold 0.2

Table 5.7 shows the retrieved results on 26 images only under threshold 0.1 using the Brighton Image Searcher (http://archive4.lis.unt.edu/td26/www) designed and implemented for this research. When image number 1 in row 1 was given, 14 images were retrieved. However, Table 5.7 shows that there are a lot more ovals

than the similarity table in Table 5.2. This result means that image only measurements are still not favorable for using alone.

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

Table 5.7 Retrieved Results on 26 Images only under Threshold 0.1

The contingency table (Table 5.8) is calculated from Table 5.2 and Table 5.7, and the precision and the recall is also calculated from the contingency table (Table 5.8). Precision and Recall for 26 images only under threshold 0.1 is calculated like the below.

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

P = 95 / 452

P = 0.2101769

So, precision is approximately 21.02 %.

Recall = (Retrieved and Relevant) / ((Retrieved and Relevant) + (Not Retrieved but Relevant))

R = 95 / 96

R = 0.9895833

So, recall is 98.96 %.

Precision is still very low, so this is not so favorable.

Image Number	Retrieved and Relevant	Retrieved but not Relevant	Not Retrieved but Relevant
1	5	9	1
2	1	21	0
3	6	15	0
4	2	19	0
5	2	19	0
6	6	16	0
7	1	21	0
8	6	15	0
9	6	15	0
10	4	0	0
11	4	1	0
12	4	0	0
13	1	5	0
14	4	0	0
15	2	18	0
16	2	18	0
17	4	16	0
18	4	17	0
19	4	17	0
20	4	17	0
21	6	14	0
22	4	16	0
23	4	16	0
24	4	16	0
25	1	20	0
26	4	16	0
Total	95	357	1

Table 5.8 Contingency Table for 26 Images only under Threshold 0.1

Table 5.9 shows the retrieved results on 26 image and text combined measurements under threshold 0.1 using the Brighton Image Searcher (http://archive4.lis.unt.edu/tdt26/www) designed and implemented for this research. When image number 1 in row 1 was given, 5 images were retrieved. This result

means that the images retrieved are including the images driven by human heuristics on Table 5.2. These are very favorable results.

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

Table 5.9 Retrieved Results on 26 Image and Text Combined under Threshold 0.1

The contingency table (Table 5.10) is calculated from Table 5.2 and Table 5.9, and the precision and the recall is also calculated from the contingency table (Table 5.10). Precision and Recall for 26 images only under threshold 0.1 is calculated like the below.

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

P = 66 / 92

P = 0.7173913

So, precision is approximately 71.74 %.

Recall = (Retrieved and Relevant) / ((Retrieved and Relevant) + (Not Retrieved but Relevant))

R = 66 / 96

R = 0.68.75

So, recall is 68.75 %.

These are very favorable.

Image Number	Retrieved and Relevant	Retrieved but not Relevant	Not Retrieved but Relevant
1	2	3	4
2	1	5	0
3	2	3	4
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	0	0
12	4	0	0
13	1	0	0
14	4	0	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	66	26	30

Table 5.10 Contingency Table for 26 Image and Text Combined under Threshold 0.1

Depending on the change of the threshold, the results of precision and recall changes. "Threshold" means, "the width of searching". If the threshold goes bigger, then it will give back more retrieved images, and if the threshold goes smaller, then it will give back less retrieved images. When the threshold changes from 0.2 to 0.1 for image only, the precision is changed from 18.11% to 21.02% and the recall is

changed from 100% to 98.96%. Then, when the threshold changes from 0.2 to 0.1 for image and text combined format, the precision is changed from 68% to 71.74% and the recall is changed from 70.83% to 68.75%. This trend demonstrates the inverse relationship between the precision and the recall, which is true for any set (Korfhage, 1997). According to these statistics the precision for a threshold of 0.2 is improved about 375%, when the combined representation format is used rather than image only format. For a threshold of 0.1, precision still improves 341%. The precision is still greatly improved when threshold 0.1 is used, about 341%. That is, the precision is improved from 18.11% to 70.83% for threshold 0.2 and 21.02% to 68.75% for threshold 0.1 when the combined representation format is used rather than image only format is used. Conclusion will be made from these statistics that precision might be improved when a combined format for multimedia documents is used.

Another test is designed for over 994 multimedia documents. Like the testing on 26 images, the first method uses 994 images for 12 measurements under the threshold 0.2. The second method uses 994 multimedia documents under the threshold 0.2 (Image and text combined, 24 measurements). It starts from the given text questions to retrieve the images. http://archive4.lis.unt.edu/td/www is used for the first method and http://archive4.lis.unt.edu/tdt/www is used for the second method. All conclusions are based on agreement of three testers in the group. On the other hand recall is not calculated because of time. It is too time consuming to decide a relative or non-relative relationship between 994 images. We already know that precision/recall have an inverse relationship, so we can limit our concern to

precision. The Table 5.11 shows 25 questions and Table 5.12 shows the retrieved results.

1. Discovery OB-103 launch and landing
2. Crew portrait in middeck wearing a various shirts
3. Orbiter name “Discovery” is visible on the orbiter
4. EVA tool preparation for upcoming Hubble Space Telescope servicing spacewalks
5. Commander Kenneth D.Bowersox looking out the aft flight desk window
6. Astronaut Tanner reads a checklist in the external airlock
7. External airlock from middeck of STS-82 Discovery
8. Orbiter name “Discovery” is visible on the orbiter
9. STIC & NICMOS EVA
10. Electronic image of Mark Lee and Steve Smith setting up for EVA
11. Hubble image over Earth limb
12. Hubble in orbiter payload bay
13. A bubble of contact lens solution enclosing an actual lens
14. EVA with astronaut and sunburst
15. Hubble EVA shots
16. Flat Stanley paper doll
17. Launch on February 11, 1997
18. Astronaut Mark Lee, wearing an extravehicular mobility unit (EMU), working on the Hubble Space Telescope with the assistance of the Remote Manipulator System (RMS) arm during mission
19. Crew portrait in the orbiter flight deck
20. Specialist Joe Tanner conducts Intravehicular Activity (IVA) during flight day 6
21. Astronaut Lee holding CCT camera and pointing it at the HST to survey the damage
22. Steve Smith waving to the camera near the end of the EVA
23. Ken Bowersox looing through the Crew Optical Alignment Sight (COAS)
24. Scott Horowitz fashioning Multilayer Insulation (MLI) patches
25. Hubble with golden solar arrays

Table 5.11 Twenty-five Questions made by NASA Employee

	Retrieved Result without Text Documents		Retrieved Result with Text Documents	
Questions	Retrieved Number of Images	Relative Number of Images	Retrieved Number of Images	Relative Number of Images
1	12	0	5	0
2	12	2	2	2
3	12	0	5	0
4	12	3	6	6
5	12	1	3	3
6	12	0	1	1
7	12	1	1	1
8	12	0	5	0
9	12	3	12	6
10	12	1	9	0
11	12	4	5	5
12	12	4	7	0
13	0	0	0	0
14	0	0	0	0
15	12	3	3	3
16	12	2	6	6
17	0	0	0	0
18	12	5	5	3
19	12	0	2	2
20	0	0	0	0
21	12	2	5	4
22	12	1	12	1
23	0	0	0	0
24	12	0	3	0
25	12	1	12	4
Total	**240**	**33**	**109**	**47**

Table 5.12 Retrieved Results by Three Testers

Using the retrieved results of Table 5.12, the precision is calculated like the below. The precision of image retrieval without using text document is,

0.1375 = 33 / 240

On the other hand, the precision of image retrieval using text document is,

0.4312 = 47 / 109

Note that the precision is increased almost 319% when image and text features are combined into a single data structure. Table 5.13 is showing the evaluation results of precision and recall over multimedia testing sets.

	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
994 multimedia documents	0.13				0.43			

Table 5.13 The Evaluation Results of Precision and Recall over Multimedia Testing Sets

CHAPTER 6. DISCUSSIONS AND CONCLUSION

Introduction

Multimedia documents exist in a variety of file formats, and various types of multimedia documents require different forms of analysis for knowledge architecture design and retrieval methods. Multiple multimedia retrieval systems by different commercial vendors have been researched, including CONVERA™, QBIC™, VIRAGE™, Image Finder™, Imatch™, CANDID™, ARTISAN™, etc. On the other hand, theories of text analysis have been proposed and applied effectively over the decades. In recent years, theories of image and sound analysis have been proposed in conjunction with text retrieval systems and implemented along with the rapid progress of computer hardware efficiency. Retrieval of multimedia documents formerly was divided into image and text, and image and sound. Also the existing process begins from text only, but now the retrieval process can be accomplished simultaneously using text and image features.

Although image processing for feature extraction and text processing for term extractions are well understood, there are no prior methods that can combine these two features into a single data structure. This dissertation introduces a common representation format for multimedia documents (CRFMD) composed of both images and text.

For image and text analysis, two techniques are used: the Lorenz Information Measure (LIM) and the Word Code (WC). A new process is demonstrated for extraction of text and image features, and then combination to produce a single data structure. Finally, this single data structure is analyzed by using multi-dimensional

scaling (MDS). Then multimedia objects are represented on a two-dimensional graph as vectors. The distance between vectors represents the magnitude of the similarity between multimedia documents.

Image classification on the given test set was dramatically improved when text features were encoded together with image features. This effect appears to hold true, even when the available text is diffused and is not uniform with the image features. In order to retrieve multimedia documents, a single data structure constructed through this process is used.

Discussions on Research Findings

In processing text documents, a binary matrix (Table 4.4) made from text document sets are often used. But the binary matrix itself was out of consideration for this research in combining two totally different types of documents into one data structure that is called "A Common Representation Format (CRF)". To combine these two documents: text and image document, the binary matrix has to be transformed to a new format.

To solve the problem of combining two different types of documents such as text document and image document, a grouping method is used to create the same number of text features as image features. The weight of each group (Table 4.6 through Table 4.9) represents the number of terms from that group that a particular document contained. Then the highest weight must be identified. Finally, the weight of each group is normalized (Table 4.11 through Table 4.14) and the normalized values are less than or equal to 0.5, due to the characteristics of the Lorenz Information Measurement. This is well explained in Chapter 4.3.3. Likewise, the binary matrix driven from National Aeronautics and Space Administration text

documents is transformed to twelve groups matrix. Then, twelve image measurements and twelve text measurements are combined for multi-dimensional scaling process. The entire process used to change from a binary matrix to twelve groups matrix for text document and to normalize the group matrix and to combine text measurements and image measurements is named Jeong's Transform.

The results of precision and recall supports that "A Common Representation Format" is a very promising theory for multimedia document retrieval system. The precision of the first test set (26 multimedia documents) is greatly improved from 18.11% to 70.83%, almost 375%, when combined measurements were used instead of using image measurements alone. It stands for the third test set (994 multimedia documents). The precision is improved almost 319%.

Discussions for Further Research

An interesting follow-up question is regarding the compatibility of other languages with this process. The Japanese language will be tested to determine the possibility of multilingual adaptation. Though Japanese is significantly different from English grammatically, there are similarities in the sentence components between the two languages. Though space is the primary delimiter between words, words are not conveniently separated by spaces in some cases. However, commas, periods, exclamation points, hyphens, *etc*. are also used just as in English. For this reason, the bigram approach, in which every two bytes are assumed as one word, will be tested, and then the results of the Japanese and English language may be compared.

For security reasons, the process of moving objects without text documents is very important. Currently, in most of security systems, moving objects are viewed, checked and analyzed by security personal. After that, the security on duty decides

how to handle the situation according to his or her personal heuristics.

Though the primary job of medical image retrieval systems is to retrieve the related patient's image with the patient's name, there is an increased interest in the use of content-based image retrieval (CBIR) techniques to aid in diagnosis by identifying similar past cases. In addition, this can be used in various ways, such as, surveillance of patients, virus tracking, heart movement, etc. For these reasons alone, processing of image itself without text documents has validity.

Finally, though all these fields such as crime prevention, war games in the military, architectural and engineering design, journalism, advertising, training in education, museums, libraries, toll gates for automatic fee collection, fashion, Geographic Information System, Web searching, and interior design, use their own image retrieval systems and approach it at different angles, but to combine this research in the above areas is greatly desired, and will define a large and active field of research.

Conclusion

Content-based image retrieval is the retrieval of imagery from a collection by means of internal measures of the information content of the images. Although CBIR has been available for more than a decade, most systems operate with only limited success. This paper describes a new method for incorporating text features into the CBIR content measures using Jeong's Transform, which was developed by the author. The results show a dramatic increase in the precision of image retrieval with the new text context measures on 26 multimedia documents from 18.11% to 70.83%, that is 375% improvement. The results indicate an increase in the precision

of image retrieval with the new text context measures on 994 multimedia documents from 13% to 43%, that is 319% improvement.

Though visualization exercises could not give digitized values, but visualization exercises on textual descriptions for image collections of greatly varying quality suggest that this technique will be successful in a variety of domains. Based on the visualized vector graphs, we could see the possible results figuratively.

The greatest achievement of this research was that Jeong's Transform is developed. Jeong's Transform represents the entire process changing from text documents to a certain number of text measurements and to combine text measurements and image measurements into a single data structure.

REFERENCES

Allen, B.L. (1996) Information Tasks :- Toward a User-Centered Approach to Information Systems. *Textbook, Academic Press*.

Brooks, R.A. (1981) Symbolic reasoning among 3-D models and 2-D images. Artificial Intelligence 17, pp. 285-348.

Carson, C.S. et al. (1997) Region-based Image Querying. *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, San Juan, Puerto Rico, pp. 42-49.

Chan, Y. et al. (1999) Building Systems to Block Pornography. *CIR-99: the Challenge of Image Retrieval*, Newcastle upon Tyne, February 25-26, 1999.

Chang, S.F. (1998) Semantic Visual Templates: Linking Visual Features to Semantics. *IEEE International Conference on Image Processing (ICIP '98)*, Chicago, Illinois, pp. 531-535.

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

Chang, S.K. & Liu, S.H. (1984) Picture Indexing and Abstraction Techniques for Pictorial Databases. *IEEE Transactions on Pattern Analysis and Machine Intelligence. 1984; 6(4)*, pp. 475-483, ISSN:0162-8828.

Demers, M.N. (1999) Fundamentals of Geographic Information Systems. *Second Edition, Textbook, John Wiley & Sons, Inc*.

ESRI, Inc. (1996) Getting to Know ArcView GIS: the Geographic Information System (GIS) for Everyone. *Textbook, Environmental Systems Research Institute, Inc. 1996.*

ESRI, Inc. (1997) Understanding GIS: The ARC/INFO Method. *Textbook, Environmental Systems Research Institute, Inc. 1997*.

Eakins, J.P. & 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, January, 1999.

Enser, P.G.B. (1995) Pictorial Information Retrieval. *Journal of Documentation*, 51(2): 126-170.

Enser, P. & McGregor, C. (1993) Analysis of Visual Information Retrieval Queries. *British Library Research and Development Report, 6104*.

Fairthorne, R.A. (1961) The Mathematics of Classification. *Towards Information Retrieval*, Butterworths, London, 1-10.

Flickner, M. et al (1995) Query by Image and Video Content: The QBIC System. *IEEE Computer, 28(9):* 23-32

Forsyth, D.A. et al (1997) Finding Pictures of Objects in Large Collections of Images. *Digital Image Access and Retrieval: 1996 Clinic on Library Applications of Data Processing*, pp. 118-139. Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.

Gastwirth, J.L. (1971) A General Definition of the Lorenz Curve. *Econometica 39*: 1037-1039

Good, I.J. (1958) Speculations Concerning Information Retrieval. *Research Report PC-78, IBM Research Centre*, Yorktown Heights, New York.

Goodrum, A., Rorvig, M., Jeong, K., Suresh, C. (2000) An Open Source Agenda for Research Linking Text and Image Content Features. *Journal of the American Society for Information Science*.

Gupta, A. et al (1996) The Virage Image Search Engine: An Open Framework for Image Management., Proc SPIE 2670, pp 76-87. *Storage and Retrieval for Image and Video Databases IV*

Hermes, T. et al. (1995) Image Retrieval for Information Systems. *Storage and Retrieval for Image and Video Databased III*, (Biblack, W.R. & Jain, R.C. eds), Proc SPIE 2420, pp. 394-405

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

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

Lesk, M.E. and Salton, G. (1969) Relevance Assessments and Retrieval System Evaluation. *Information Storage and Retrieval*, 4, 343-359.

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

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

Liu, F. & Picard, R.W. (1996) Periodicity, Directionality and Randomness: World Features for Image Modeling and Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7), pp.722-733.

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

Luhn, H.P. (1957) A Statistical Approach to Mechanised Encoding and Searching of Library Information. *IBM Journal of Research and Development*, 1, 309-317.

Ma, W.Y. & Manjunath, B.S. (1998) Texture Features for Browsing Large Aerial Photographs. *Journal of the American Society for Information Science, 49(7)*, pp. 633-648.

Markey, K. (1984) Indexer Consistency Tests: A Literature Review and Report of a Test of Consistency in Indexing Visual Materials. *Library and Information Science Research, 6*, 155-177.

Maron, M.E. & Kuhns, J.L. (1960) On Relevance, Probabilistic Indexing and Information Retrieval. *Journal of the ACM, 7*, 216-244.

McCarn, D.B. & Leiter, J. (1973) On-line Services in Medicine and Beyond. *Science*, 181, 318-324.

Mehrotra, R. & Gary, J.E. (1995) Similar-Shape Retrieval in Shape Data Management. *IEEE Computer, 28(9)*, pp. 57-62.

Nadler, M & Smith, E.P. (1992) Pattern Recognition Engineering. *Textbook, John Wiley & Sons Inc.*

Niblack, W. et al. (1993) The QBIC Project: Querying Images by Color, Texture and Shape. *IBM Research Report*, RJ-9203.

Ravela, S. & Manmatha, R. (1998a) Retrieving Images by Appearance. *Proceedings of IEEE International Conference on Computer Vision (IICV98),* Bombay, India, pp. 608-613.

Rorvig, M. (1993) A Method for Automatically Abstracting Visual Documents. *Journal of the American Society for Information Science*, 44(1): 040-056.

Rorvig, M. & Fitzpatrick, S. (1997) Visualization and Scaling of TREC Topic Document Sets. *International Journal of Information Processing and Management*, Sep. 19, 1997

Rorvig, M. & Fitzpatrick, S. (2000) "Shape Recovery: A Visual Method for Evaluation of Information Retrieval Experiments". *Journal of the American Society for Information Science*, 51(13): 1205-1210.

Rorvig, M., Fitzpatrick, S., Ladoulis, C.T., Vitthal, S. (1993) A New Machine Classification Method Applied to Human Peripheral Blood Leukocytes. *Information Processing and Management*, 29(6): 765-774.

Rorvig, M. & Jeong, K. (2000) A Common Representation Format for Multimedia Document. *Texas Center for Digital Knowledge*, School of Library and Information Sciences, University of North Texas: Denton, Texas.

Rorvig, M., Jeong, K., Suresh, C., Goodrum, A. (2000) Exploiting Image Primitives for Effective Retrieval. *CIR2000 – Third UK Conference on Image Retrieval*, 4-5 May 2000, Old Ship Hotel, Brighton, United Kingdom, University of Brighton: School of Information Management.

Rorvig, M., Smith, M. & Uemura, A. (1998) The N-Gram Hypothesis Applied to Matched Sets of Visualized Japanese-English Technical Documents. *Proceedings of the 1999 Annual Meeting of the America Society for Information Science, Knowledge: Creation, Organization, and Use*, October 31 – November 4, 1999, Washington D.C. (L. Woods. ED.) pp.359-364

Salton, Gerard, & James Allen. (1994) Text Retrieval using the Vector Processing Model. *Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval*, Las Vegas, pp. 9-22.

Senko M.E. (1969) Information Storage and Retrieval Systems. *In Advances in Information Systems Science*, (Edited by J. Tou) Plenum Press, New York.

Shepard, R.N. (1962a) The Analysis of Proximities: Multidimensional Scaling Method an Unknown Distance Function. I. *Psychometrika, 27(2):* 125-140.

Shepard, R.N. (1962b) The Analysis of Proximities: Multidimensional Scaling Method an Unknown Distance Function. II. *Psychometrika, 27(3):* 219-246.

Stricker, M. & Dimai, A. (1996) Color Indexing with Weak Spatial Constraints. *Storage and Retrieval for Image and Video Databases IV,* (Sethi, I K and Jain, R C, eds), Proc SSPIE 2670, pp. 29-40.

Swain, M.J. & Ballard, D.H. (1991) Color Indexing. *International Journal of Computer Vision*, 7(1), pp. 11-32.

Tamura, H. et al. (1978) Texture Features Corresponding to Visual Perception. *IEEE Transactions on Systems, Man and Cybernetics*, 8(6), pp.460-472.

Torgerson, W. (1958) Theory and Methods of Scaling. *New York: Wiley*.

Turner, J. (1998) Some Characteristics of Audio Description and the Corresponding Moving Image. *61st Annual Meeting of the American Society for*

Information Science, (Preston, C. Ed.), Pittsburgh, PA, October 25-29, 1998, pp. 108-117.

Venters, C.C. & Cooper, M. (1999) A Review of Content-Based Image Retrieval Systems. *University of Manchester*, 1999.

Wactlar, H.D. et al. (1996) Intelligent Access to Digital Video: the Informedia Project. *IEEE Computer 29(5),* pp. 46-52.

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