國 立 交 通 大 學

資訊工程學系

博 士 論 文

醫學影像資料庫之研究

A Study on Medical Image Database

研 究 生：鄭培成

指導教授：楊維邦 教授

錢炳全 教授

中 華 民 國 九 十 五 年 十 二 月

醫學影像資料庫之研究

A Study on Medical Image Database

研 究 生：鄭培成　　Student：Pei-Cheng Cheng

指導教授：楊維邦 博士　　Advisor：Dr. Wei-Pang Yang

錢炳全 博士　　Dr. Been-Chian Chien

國 立 交 通 大 學

資 訊 工 程 學 系

博 士 論 文

A Dissertation Submitted to

Department of Computer Science

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Computer Science

December 2006

Hsinchu, Taiwan, Republic of China

中華民國九十五年十二月

醫學影像資料庫之研究

學生：鄭培成　　　　　　　　　　　　　　　　指導教授：楊維邦 博士

錢炳全 博士

國立交通大學資訊科學與工程研究所 博士班

摘　　要

隨著數位化科技以及網路通訊的快速發展，影像資料在醫療院所、數位典藏以及網際網路上被大量的產生，日漸提高了影像檢索技術的需求。另一方面由於網路的快速連結，跨越了國界，各種語言的資料都可以在網路上檢索。本論文主要研究一個跨語言的醫學影像查詢系統，讓使用者可以使用熟悉的語言或是使用影像內容來找尋相關的醫學影像。

本論文探討的兩個研究主題，為醫學影像內容查詢以及跨語言的資訊檢索。在醫學影像內容查詢方面，我們提出了一些針對學影像內容特徵的表示法，相關的相似度計算方法，以及新的接受使用者回饋的模式，在實驗結果中顯示所提出的特徵表示法可以有效的提高準確率及查全率。在跨語言方面我們針對在翻譯時所常遇到的歧義性的問題，提出了利用本體知識來改善的方法，實驗結果顯示我們所提出解決歧義性問題的方法，可以有效提高查詢結果的準確率。

關鍵字： *醫學影像查詢; 跨語言查詢; 相關回饋查詢;*

A Study on Medical Image Database

Student: Pei–Cheng Cheng

Advisors: Dr. Wei-Pang Yang
Dr. Been-Chian Chien

Institute of Computer Science and Engineering
National Chiao Tung University

ABSTRACT

The importance of digital image retrieval techniques increases in the emerging fields of medical imaging, picture archiving and communication systems. In this dissertation, a bilingual medical image database system is proposed for users to retrieve medical images. The principal objective is to provide users a medical image retrieval system to find similar diagnosis and to obtain useful information for treatment.

This dissertation relates to two areas – medical image retrieval and cross-language information retrieval. We proposed an effective representation for content-based medical image retrieval and an approach to address the translation ambiguity problem for cross-language image retrieval. Furthermore, a novel relevance feedback mechanism is proposed to improve the retrieval effectiveness by interacting with users.

Keywords: *Medical image retrieval; Cross-language retrieval; Relevance feedback*

誌　　謝

論文能夠順利完成，首先要感謝我的指導教授楊維邦教授及錢炳全教授，多年來悉心的教導以及鼓勵,使我不論在學業上以及待人處事方面,都能夠更進一層的學習。兩位老師平時的循循善誘以及諄諄教誨都給予我適時的鼓舞作用，讓我可以持續堅持下去，完成這本論文。

我也要特別感謝袁賢銘教授、孫春在教授、陳榮傑教授、唐傳義教授、項潔教授、黃明祥教授以及曾定章教授對於本論文的不吝指教及給予本論文諸多寶貴的意見。

還有感謝資料庫實驗室的伙伴們，讓我在實驗室裡獲得許多的協助以及關心，使我能夠渡過這個漫長的等待，我畢業囉! 學弟 加油!

最後要感謝我的家人對於我百分之百的支持,長久以來對於我的關懷以及不求回報的付出，是我努力不懈的原動力，非常感謝對我非常關心的父母，爸爸、媽媽以及哥哥，謝謝你們!我想你們最想聽到的話就是”我畢業了”吧。

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Image capture capabilities are evolving so rapidly, extreme amount of images are produced daily. The importance of digital image retrieval techniques increases in the emerging fields of publication on the internet, variety trademarks, and the medical imaging (Geneva radiology department generate 20,000 images per day), etc. An image is worth thousand of words especially in medical application. It is a hard work to retrieve an image from thousand of images by browsing one bye one. Retrieving large amount of images rely on annotation is expensive and time consume. On the other hand, human has subjective viewpoint, image that annotate by different people may be described into different annotation.

Content Based Image Retrieval (CBIR) is a new technology to assist image finding. CBIR retrieve image by itself without annotation. CBIR allow user to query image database by image example, partial region of image, sketch contour, or dominate color, etc. IBM in 1995 have developed the QBIC[20] system which lets user make queries of large image databases based on visual image content properties such as color percentages, color layout, and textures occurring in the images. User can match colors, textures and their positions without describing them in words. Content based image retrieval offer an alternate method for user to retrieve desired images. CBIR is a very convenient and economic approach for image retrieval system because it is an automatic method to analyze images.

Imaging systems and image archives have often been described as an important economic and clinical factor in the hospital environment [22]. Several methods from computer vision and image processing have already been proposed for the use in medicine [51]. Medical images have often been used for retrieval systems, and the medical domain is often cited as one of the principal application domains for content-based access technologies [6] [32] [48] [64] in terms of potential impact.

Images by their nature are language independent, but often they are accompanied by texts semantically related to the image (e.g. textual captions or metadata). Images can then be retrieved using primitive features based on pixels with form the contents of an image (e.g. using a visual exemplar), abstracted features expressed through text or a combination of both. The language used to express the associated texts or textual queries should not affect retrieval, i.e. an image with a caption written in English should be searchable in languages other than English.

In this dissertation, we study on a cross medical image retrieval system to assist clinical student learning or patient to understand his condition. In this chapter the rationale of cross medical image retrieval system is introduced. The status of medical image retrieval is also reviewed. The main results of this dissertation are stated in the last of each chapter.

1.1 The Rationale of Cross Medical Image Retrieval System

The use of CBIR systems is becoming an important factor in medical imaging research. In this dissertation we study CBIR systems and determine how associated cross-language text can be used in combination with CBIR to improve retrieval and ranking in particular medical domain. The goal is to find images that are similar with respect to modality (CT, radiograph, MRI...), with respect to the anatomic region shown (lung, liver, head ...) and sometimes with respect to the radiological protocol (such as a contrast agent.).

The medical image retrieval systems allow user can use image example to retrieve related documents or use native language to query another language documents or images. Figure 1-1 is the data model of cross medical image retrieval system. When user issues query by image example first, visual feature was used to find visual similar images. The images are associated with case notes; a written description of a previous diagnosis for an illness the image identifies (e.g. Figure 1-2). Case notes consist of several fields including: a diagnosis, a free textual description, clinical presentation, keywords and title. Case notes are mixed language written in either English or French. We can use the associated notes to improve the query result. Extracting the possible keyword from the visual similar result as query words used to execute text query. In the last phase, we can combine the text and visual result to refine the query result.

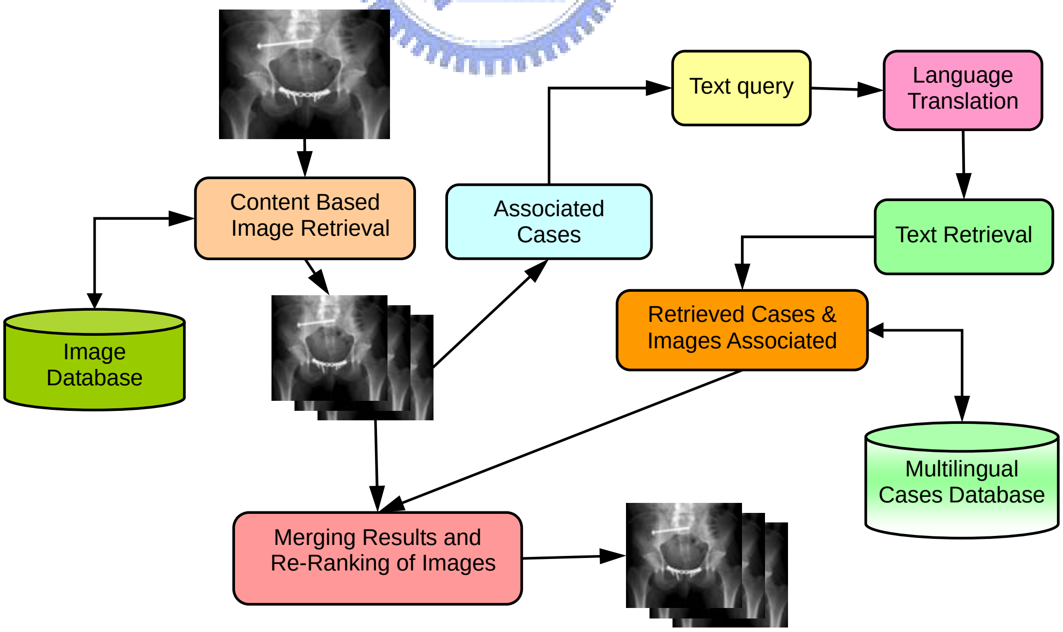


Figure 1-1: The data model of our system

An example case and images

<Description>

X ray: Mass effect within the soft tissues of the proximal part of the left calf, difficult to outline, seen only as it displaces the fat planes. There are no calcifications. The adjacent bone is normal.

MRI: Oval mass within the medial gastrocnemius muscle, very well delineated, slightly lobulated (axial cuts). Its structure is heterogeneous. On T1, it is slightly hyperintense compared to the adjacent muscle (red arrow). It is isointense to fat on proton density images (DP), very hyperintense on T2 and IR with some hypointense areas in its centre. After injection of contrast medium, there is marked enhancement except for the central area, which remains hypointense.

Arteriography: there is hypervascularity of the soft tissues outside the medial tibial plateau by vessels arising mostly from the genicular arteries.

…..

</Description>

Figure 1-2: A case note from the medical case

1.2 Content Based Medical Image Retrieval

In the past years, content-based image retrieval has been one of the most hot research areas in the field of computer vision. The commercial QBIC [20] system is definitely the most well known system. Another commercial system for image and video retrieval is Virage [2][25] that have well known commercial customers such as CNN. In the academia, some systems including Candid [32], Photobook [50], and Netra [37] use simple color and texture characteristics to describe image content. The Blobword system [5][9] exploits higher-level information, such as segmented objects of images, for queries. A system that is available free of charge is the GNU Image Finding Tool (GIFT) [67]. Some systems are available as demonstration version on the Web such as Viper, WIPE or Compass.

Content based image retrieval system contains two main phases: first phase is feature representation, the system extract the features from image automatic. The feature selection dependent on the application, good representation will be the first condition to obtain better performance. The second phase is similar metric designing, the similar metric must correspondent to the representation of images.

The general application of image retrieval to broad image databases has experienced limited success, principally due to the difficulty of quantifying image similarity for unconstrained image classes (e.g., all images on the Internet). We expect that medical imaging will be an ideal application of CBIR, because of the more-limited definition of image classes, and because the meaning and interpretation of medical images is better understood and characterized. In this thesis, we design several image feature representation methods and similar metric for medical images content based retrieval. In Chapter 2, the experiment result shows that proposed method is out performance than previous work.

1.3 Combine Text and Visual Feature for Medical Image Retrieval

Although content based approach can find similar images automatically, content based method use image features to retrieve image sometimes the query result is unexpected by user because the result rely on image features automatic produced by system. Image feature is unlike keyword which is easy reasonable by user.

Content based image retrieval still has some problem need to be solved. Two of the common problems are sensory gap and semantic gap. The loss between the actual structure and the representation in a digital image is called sensory gap. The another problem is semantic gap, Even systems using segments and local features such as Blobworld [9] are still far away from identifying objects reliably. How can we use limited feature domain to represent real world infinite possible objects. No system offers interpretation of images or even medium level concepts as they can easily be captured with text. This loss of information from an image to a representation by features is called the semantic gap. The more a retrieval application is specialized for a certain, limited domain, the smaller the gap can be shortened. Content based queries are often combined with text and keyword predicate to get powerful retrieval methods for image and multimedia databases. In the experiment result show that combines the text and visual feature will get better performance.

The rapid growth in size of the World Wide Web (WWW) and the increasing speed of internet cause the world become globalization. People can get information from various countries on the internet. Thus, multilingual information retrieval problem becomes a hot research field.

Cross-language retrieval is the retrieval of any type of object (texts, images, products, etc.) composed or indexed in one language (the target language) with a query formulated in another language (the source language). There may be any number of source languages and any number of target languages.

In many collections (e.g. historic or stock photographic archives, medical databases and art/history collections), images are often accompanied by some kind of text (e.g. metadata or captions) semantically related to the image. Retrieval can then be performed using primitive features based on pixels which form an image's content (CBIR); using abstracted textual features assigned to the image, or a combination of both. The language used to express the associated texts or metadata should not affect the success of retrieval, i.e. an image with English captions should be searchable in languages other than English.

In Chapter 3, retrieving medical image with multilingual annotation is studied. While translating query language to another language will cause the ambiguous problem, we proposed an ontology-based approach to overcoming the problem of translation ambiguity. The experimental results show that the proposed approach can effectively remove the irrelevant results and improve the precision.

1.4 Medical Image Retrieval with Relevance Feedback

Many researches show that relevance feedback method really improves the performance of content based approach results. Users often have subjective viewpoint. Some user may want to find similar color images and others may want to find similar shape images. Furthermore, the color definition of similar red for each person is different. The system initially hard to know the user's biased like and the system can't designed for specify users without profile of users. Relevance feedback learn user's preference reflect to the system. System according to the user offered example (positive examples or negative examples) to reformulate the query. The relevance feedback mechanism improves the precision of result respect for user.

In Chapter 4, we propose a new relevance feedback mechanism to fast trace the user's interesting. Previous works use positive and negative example to reformulate the query example. In this thesis, an ordering query examples model was proposed, user can further express the prefer images in an order. Based on the ordering information, we can refine the result more fast and precise.

1.5 Motivation & Objective

Content based image retrieval method is a good and only choice to retrieval images while images do not have explanatory notes. The increasing reliance of modern medicine on diagnostic techniques such as radiology, histopathology, and computerized tomography has led to the explosion in the number and importance of medical images now stored by most hospitals. While the prime requirement for medical imaging systems is to be able to display images relating to a named patient, there is increasing interest in the use of CBIR techniques to aid diagnosis by identifying similar past cases.

In the world, people confront with variety disease and accident. In different country, the similar cases of illness may occur. The rare diseases especially need reference similar past cases to aid diagnosis. There is a problem need to be conquered that the diagnosis cases was annotated in different language. In this thesis, we also study the cross-language information retrieval method to add the robust of medical image systems and offer user can query in native language. We try to combine the keyword and image to improve the precision of the medical image systems. The images illustrate the fact without keyword, thus the cross language problem is not exist. The similar images (found by CBIR) may offer a connection to link the similar diagnosis cases which were annotated in different language. The similar content of images may be annotated by different languages, but they all describe the same things. In the world there exist many usefully diagnosis cases, but the diagnosis are described in different languages. Based on the medical image, we may offer a cross language environment for user to query similar past cases which are written bye different languages.

In previous related work, user only can mark the result image which is relevant or not. The relevance feedback mechanism relies on the positive examples or negative examples to refine the query formulation. In this dissertation we design a novel relevance feedback model to receive more information from user. User can set a prefer order of similar images of result. According to the rank images set by user, the systems is more easy to tune the weight of each image features. We expect this reformulation mechanism can fast close to the user's interesting than previous feedback mechanism.

In this study we research what technologies can be used in the application of medical image retrieval. In the limited domain some technologies may need modify to suit the specify application. We proposed a wavelet based signature for images combine and modify previous methods to analysis the coefficient of wavelet. The representation we designed consider the human's perceptual cooperate with low level image features have better performance than previous works.

In this dissertation contain two main research topics. One is medical image retrieval problem and the other is about bilingual language information retrieval. In the medical image retrieval issue, we modify exist content based techniques used in general propose images to suit medical images retrieval. We also develop several new representations of medical images from human's conceptual and the correspondent metrics of new representation was designed. In the cross language information retrieval problem, we propose an ontology method to solve the translation ambiguity problem. The occurrence probability of transformed words in the same class will help to solve the translation ambiguity problem.

In the previous related works, user can refine the results by relevant feedback mechanism. User can mark images as relevant or not from result for retrieval system. The system according to user's interesting that is positive examples tunes the weight of related features to improve the result. In the experiment result of previous related work show that positive example will improve the results. Previous works only allow user to mark relevant or not, the system need several times to refine the results. In is dissertation we design a new mechanism allow user to offer more information for system to fast close to his interesting. Proposed feedback mechanism allows user to rank the results by similarity. According to the ranking of selected images we can get the information for weighting the image features.

1.6 Organization of This Dissertation

The remainder of this thesis is organized as follows. In Chapter 2, we review previous related visual medical image researches and proposed an efficient representation correspond to user's viewpoint and similar matching metric. In Chapter 3, we combine text and visual feature to improve the accuracy of query result and propose an ontology method to conquer the ambiguity problem of cross language retrieval. In Chapter 4, relevance feedback methods were surveyed and a novel feedback mechanism is proposed. Last, the conclusion and further works is described in Chapter 5.

Chapter 2
Content Based Medical Image Retrieval

The number of digitally produced medical images is rising strongly. In the radiology department of the University Hospital of Geneva alone, the number of images produced per day in 2002 was 12,000, and it is still rising. The management and the access to these large image repositories are increasing the need for tools that effectively filter and efficiently search through large amounts of visual data. Most access to these systems is based on the patient identification or study characteristics (modality, study description) [35] as it is also defined in the DICOM standard [52].

Imaging systems and image archives have often been described as an important economic and clinical factor in the hospital environment [22]. Several methods from computer vision and image processing have already been proposed for the use in medicine [51]. Medical images have often been used for retrieval systems, and the medical domain is often cited as one of the principal application domains for content-based access technologies [6] [32] [48] [64] in terms of potential impact.

Two exceptions seem to be the Assert system on the classification of high resolution CTs of the lung [1] [63] and the Image Retrieval in Medical Applications (IRMA) system for the classification of images into anatomical areas, modalities and view points [33]. Assert system allow the physician to circles one or more Pathology Bearing Regions (PBR) in the query image. The system then retrieves the n most visually similar images from the database using an index comprised of a combination of localized features of the PBRs and of the global image. The database consists of High Resolution Computed Tomography of the lung. CBIR is particularly needed for this domain because the current state of the art in diagnosis, for those cases not immediately recognizable, is to consult a published atlas of lung pathologies. Assert system saves the radiologist from the laborious task of paging through the atlas looking for an image that matches the pathology of their current patient.

Image Retrieval in Medical Applications (IRMA) is a cooperative project of the Department of Diagnostic Radiology, the Department of Medical Informatics, Division of Medical Image Processing and the Chair of Computer Science VI at the Aachen University of Technology. Aim of the project is the development and implementation of high-level methods for content-based image retrieval with prototypical application to medico-diagnostic tasks on a radiological image archive.

They want to perform semantic and formalized queries on the medical image database which includes intra- and inter-individual variance and diseases. Example tasks are the staging of a patient's therapy or the retrieval of images with similar diagnostic findings in large electronic archives. Formal content-based queries also take into account the technical conditions of the examination and the image acquisition modalities.

The system ought to classify and register radiological images in a general way without restriction to a certain diagnostic problem or question. Methods of pattern recognition and structural analysis are used to describe the image content in a feature based, formal and generalized way. The formalized and normalized description of the images is then used as a mean to compare images in the archive which allows a fast and reliable retrieval.

In addition to the queries on an existing electronic archive, the automatic classification and indexing allows a simple insertion of conventional radiographs into the system. Automated classification of radiographs based on global features with respect to imaging modality, direction, body region examined and biological system under investigation.

Identification of image features that are relevant for medical diagnosis these features are derived from a-priori classified and registered images. The resulting system must retrieve images similar to a query image with respect to a selected set of features. These features can, for example, be based on the visual similarity of certain image structures. Figure 2-1 is a concept view of content based image retrieval system.

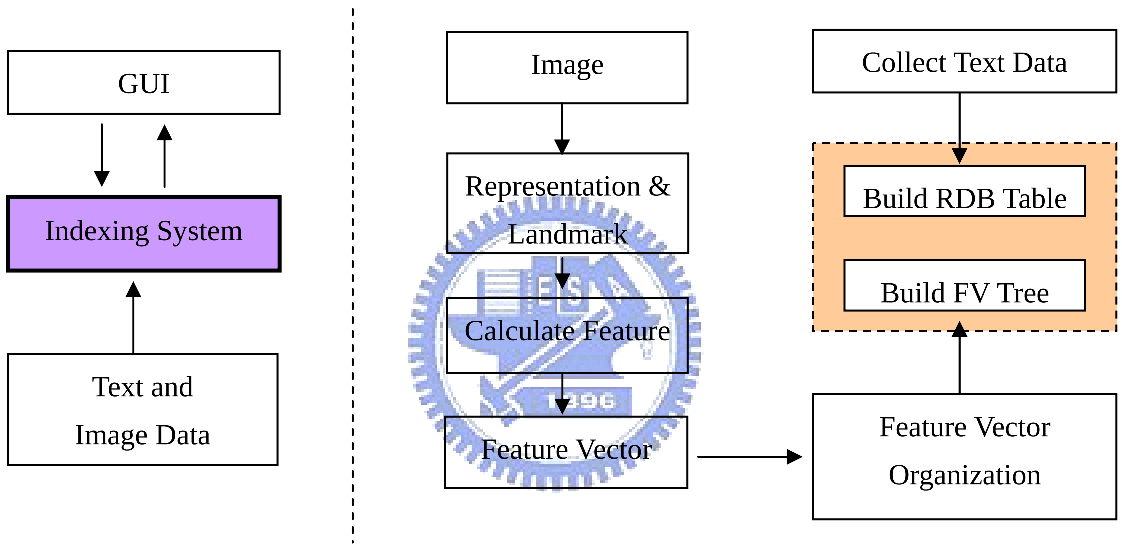


Figure 2-1: A concept view of content based image retrieval system

Content-based retrieval has also been proposed several times from the medical community for the inclusion into various applications [7] [69], often without any implementation. Still for a real medical application of content-based retrieval methods and the integration of these tools into medical practice a very close cooperation between the two fields is necessary for a longer period of time and not simply an exchange of data or a list of the necessary functionality.

Most existing content-based image retrieval (CBIR) systems rely on computing a global signature for each image based on color, texture and shape information [65].

However, medical image usually consists of gray level variations in highly localized regions in the image. Thus, the content-based image retrieval method in color based can't apply to medical images directly. In the gray level images, the contrast of gray level is more important than illuminate in human's viewpoint. In this dissertation, we consider the human's conceptual and we design a wavelet-based image representation method. Wavelets have proven their efficiency in image retrieval, for their capability in capturing texture and shape information [1] [42] [49]. Making use of wavelet sub-band decomposition, relevant information about the structure of data can be computed, which can serve as a low dimensional feature vector. A wavelet-based medical image retrieval system, which is mainly based on textural information, has shown satisfactory results in retrieving HRCT lung images [31].

One of the most significant problems in content-based image retrieval results from the lack of a common test-bed for researchers. Although many published articles report on content-based retrieval results using color photographs, there has been little effort in establishing a benchmark set of images and queries. It is very important that image databases are made available free of charge for the comparison and verification of algorithms. Only such reference databases allow comparing systems and to have a reference for the evaluation that is done based on the same images. ImageCLEF[1] [12] offers numerous medical images for evaluation and has many benefits in advancing the technology and utilization of content-based image retrieval systems.

Compared to text retrieval little is known about how to search for images, although it has been an extremely active domain in the fields of computer vision and information retrieval [41][56][64][69]. Benchmarks such as ImageCLEF [12] allow evaluating algorithms compared to other systems and deliver insights into the

[1] http://clef.isti.cnr.it/

techniques that perform well. They offered a default CBIR system that is GIFT[2]. This software tool is open source and can be used by other participants of ImageCLEF. The feature sets that are used by GIFT are:

–Local color features at different scales by partitioning the images successively into four equally sized regions (four times) and taking the mode color of each region as a descriptor;

–Global color features in the form of a color histogram, compared by a simple histogram intersection;

–Local texture features by partitioning the image and applying Gabor filters in various scales and directions, quantized into 10 strengths;

–Global texture features represented as a simple histogram of responses of the local Gabor filters in various directions and scales.

In this dissertation we want to study a medical image retrieval system that contains variety of type images for clinical student to learning or patient to understand his condition. The image contains variety of type image, thus we consider global and local image features expect to describe variety of type images. We based on the wavelet coefficient to extract dominate gray value and texture information. For efficient retrieving we design several new representations of features. The following is our proposed feature representation methods.

[2] http://www.gnu.org/software/gift/

2.1 Proposed Methods for Medical Image Comparison

In color images, users are usually attracted by the change of colors more than the positions of objects. Thus, we use *color histogram* as the feature of color images to retrieve similar color images. Color histogram is suitable to compare images in many applications. Color histogram is computationally efficient, and generally insensitive to small changes in the camera position.

Color histogram has some drawbacks. Color histogram provides less spatial information; it merely describes which colors are present in an image, and in what quantities. Because gray images encompass few colors (usually 256 gray levels), directly using color histogram in gray images will get bad retrieval results. For gray images, we must emphasize spatial relationship analysis; furthermore, object and contrast analysis is important for medical images; therefore, three kinds of features that can indicate the spatial, coherence, and shape characteristics, *gray-spatial histogram*, *coherence moment*, and *gray correlogram*, are employed as the features of gray images.

In the following we describe the four kinds of features, one for color images and three for gray images, used in this paper.

2.1.1 Color Image Features

Color histogram [68] is a basic method and has good performance for presenting image content. The color histogram method gathers statistics about the proportion of each color as the signature of an image. Let C be a set of colors, $(c_1, c_2 \ldots c_m) \in C$ that can occur in an image. Let I be an image that consists of pixels p(x, y)[3]. The

[3] p(x, y) indicates the color of the corresponding pixel as well.

color histogram $H(I)$ of image I is a vector ($h_1, h_2, \ldots, h_i, \ldots, h_m$), in which each bucket h_i counts the ratio of pixels of color c_i in I. Suppose that p is the color level of a pixel. Then the histogram of I for color c_i is defined as Eq. (1):

$$h_{c_i}(I) = \Pr_{p \in I}\{p \in c_i\} \quad (1)$$

In other words, $h_{c_i}(I)$ corresponds to the probability of any pixel in I being of the color c_i. For comparing the similarity of two images I and I', the distance between the histograms of I and I' can be calculated using a standard method (such as the L_1 distance or L_2 distance). Then, the image in the image database most similar to a query image I is the one having the smallest histogram distance with I.

Any two colors have a degree of similarity. Color histogram is hard to express the similar characteristic between neighbor bins. Each pixel does not only assign a single color. We set an interval range δ to extend the color of each pixel. Then the histogram of image I is redefined as the Eq. (2):

$$h_{c_i}(I) = \frac{\sum_{j=1}^{m} \frac{[p_j - \frac{\delta}{2}, p_j + \frac{\delta}{2}] \cap c_i}{\delta}}{m} \quad (2)$$

Where p_j is a pixel of image, and m is the total number of pixels.

The colors of an image are represented in HSV (Hue, Saturation, Value) space, which is closer to human perception than spaces such as RGB (Red, Green, Blue) or CMY (Cyan, Magenta, Yellow). In implementation, we quantize HSV space into 18 hues, 2 saturations and, 4 values, with additional 4 levels of gray values; as a result, there are a total of 148 bins.

Using the modified color histogram, the similarity of two color images q and d is defined as Eq. (3):

$$\text{SIMcolor}(\text{H(q)}, \text{H(d)}) = \frac{\text{H(q)} \cap \text{H(d)}}{|\text{H(q)}|} = \frac{\sum_{i=1}^{n} \min(h_i(q), h_i(d))}{\sum_{i=1}^{n} h_i(q)} \tag{3}$$

2.1.2 Gray Image Features

Gray images are different from color images in human's perception. Gray images have fewer colors than color images, only 256 gray levels in each gray image. Human's visual perception is influenced by the contrast of an image. The contrast of an image from the viewpoint of human is relative rather than absolute. To emphasize the contrast of an image and handle images with less illuminative influence, we normalize the value of pixels before quantization. In this paper we propose a relative normalization method. First, we cluster the whole image into four clusters by the K-means cluster method [26]. We sort the four clusters ascendant according to their mean values. We shift the mean of the first cluster to value 50 and the fourth cluster to value 200; then each pixel in a cluster is multiplied by a relative weight to normalize. Let m_{c1} is the mean value of cluster 1 and m_{c4} is the mean value of cluster 4. The normalization formula of pixel p(x, y) is defined as Eq. (4).

$$p(x, y)_{normal} = (p(x, y) - (m_{c1} - 50)) \times \frac{200}{(m_{c4} - m_{c1})} \tag{4}$$

After normalization, we resize each image into 128*128 pixels, and use one level wavelet with Haar wavelet function [66] to generate the low frequency and high frequency sub-images. Wavelets are mathematical functions that decompose signals into different frequency components, and then analyze each component with a resolution matching its scale. The main advantage of wavelets over Fourier analysis is that they allow better resolution in space and frequency [23]. Consequently, the wavelet transform (WT) is more efficient in analyzing periodic signals, such as images, especially if they contain impulsive sharp changes. The continuous WT of a function f(x) using a wavelet functions basis is defined as

$$F(i,j)=\int f(x)\psi_{i,j}(x)dx \tag{5}$$

The basis of wavelet function is obtained by shifting and scaling a single mother wavelet function $\psi(x)$:

$$\psi_{i,j}(x)=\frac{1}{\sqrt{i}}\psi(\frac{x-j}{i})dx; \qquad i>0 \tag{6}$$

Where i is the scale factor and j is the shift value. The mother wavelet should satisfy the zero average condition. The Discrete WT is obtained by taking $i=2^n$ and $b\in Z$.

The analyzed signal can be a 2-D image. The 2-D analysis is performed as a product of two 1-D basis functions

$$\psi_{i1,j1;i2,j2}(x1,x2)=\psi_{i1,j1}(x1)\cdot\psi_{i2,j2}(x2) \tag{7}$$

This yields a decomposition of the image into 4 subbands called the approximation (the low frequency components in 1 subband) and the details (the high frequency components in 3 subbands). From the whole set of approximation and details coefficients, it is possible to reconstruct the original image again. The decomposition process can be iterated with successive approximations being decomposed in turn.

Process an image using the low pass filter will obtain an image that is more consistent than the original one; on the contrary, processing an image using the high pass filter will obtain an image that has high variation. The high-frequency part keeps the contour of the image. After wavelet translation, there are four sub-bands denoted by Low_Low (LL), Low_Height (LH), Height_Low (HL), Height_Height (HH). High-frequency pixels may be important in medical images for doctor diagnoses. By performing the OR operation for LH, HL, and HH bands, we get the contour of a medical image.

2.1.3 Gray-Spatial Histogram

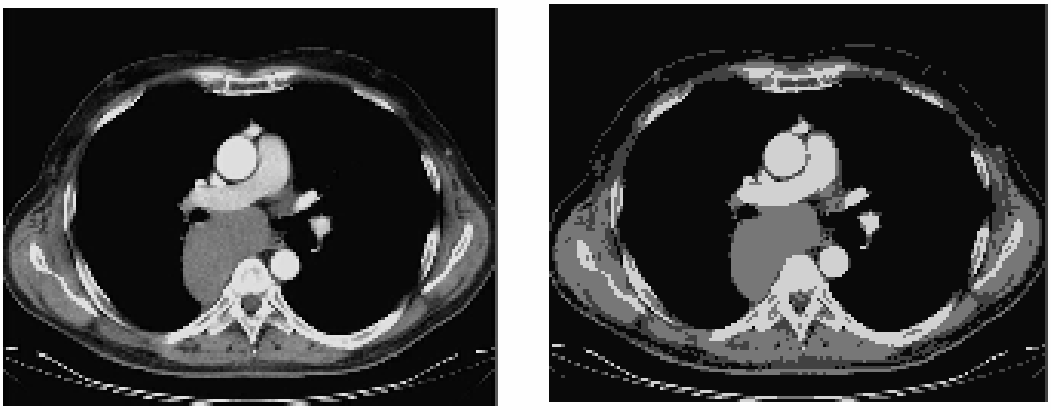
In a gray image the spatial relationship is very important especially in medical images. Medical images always contain particular anatomic regions (lung, liver, head, and so on); therefore, similar images have similar spatial structures. We add spatial information into the histogram so we call this representation as gray-spatial histogram in order to distinguish from color histogram. We use the LL band for gray-spatial histogram and coherence analysis. To get the gray-spatial histogram, we divide the LL band image into nine areas. The gray values are quantized into 16 levels for computational efficiency.

The gray-spatial feature estimates the probability of each gray level that appears in a particular area. The probability equation is defined in Eq. (2), where δ is set to 10. The gray-spatial histogram of an image has a total of 144 bins.

2.1.4. Coherence Moment

One of the problems to design an image representation is the semantic gap. The state-of-the-art technology still cannot reliably identify objects. The coherence moment feature attempts to describe the features from the human's viewpoint in order to reduce the semantic gap.

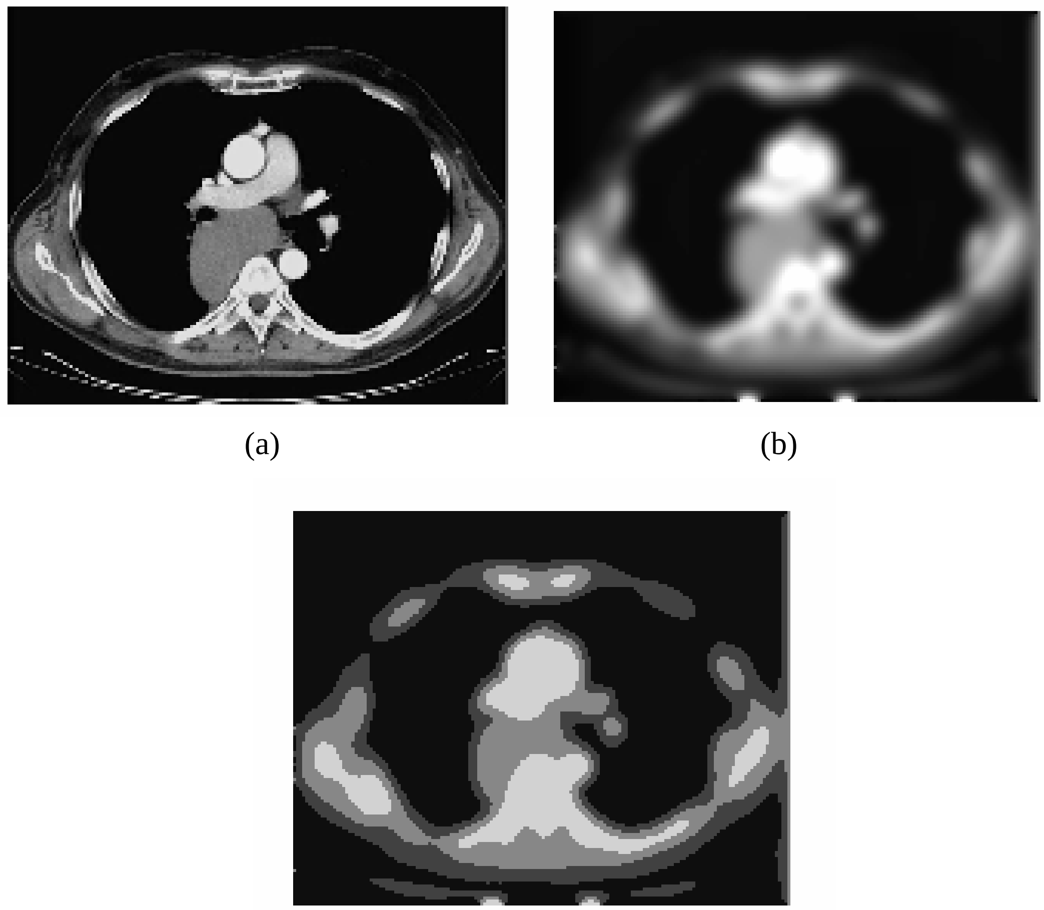
We cluster an image into four classes by the K-means algorithm. Figure 2-2 is an example. Figure 2-2 (a) is the original image and Figure 2-2 (b) is four-level gray image. We almost can not visually find the difference between the two images. After clustering an image into four classes, we calculate the number of pixels (COH_κ), mean value of gray value (COH_μ) and standard variance of gray value (COH_ρ) in each class. For each class, we group connected pixels in eight directions as an object. If an object is bigger than 5% of the whole image, we denote it as a big object; otherwise it is a small object. We count how many big objects (COH_o) and small objects (COH_ν) in each class, and use COH_o and COH_ν as parts of image features.



(a) (b)

Figure 2-2: (a) original image with 256 levels; (b) new image after clustering with only 4 levels

We intend to detect the reciprocal among neighbor pixels and eliminate the effect of noise; we apply smoothing method to the original image. If two images are similar, they will also be similar after smoothing. If their spatial distributions are quite different, they may have different result after smoothing. After smoothing, we cluster an image into four classes and calculate the number of big objects (COH_τ) and small objects (COH_ω). Figure 2-3 is an example. Each pixel will be influenced by its neighboring pixels. Two close objects of the same class may be merged into one object. Then, we can analyze the variation between the two images before and after smoothing. The coherence moment of each class is a seven-feature vector, (COH_κ, COH_μ, COH_ρ, COH_o, COH_ν, COH_τ, COH_ω). The coherence moment of an image is a 28-feature vector that combines the coherence moments of the four classes.



(a) (b)

(c)

Figure 2-3 (a) original image; (b) image after smoothing; (c) image after clustering into four classes.

2.1.5 Gray Correlogram

The contour of a medical image contains rich information. Diseases can be easily detected in the high frequency domain. But, in this task we are going to find similar medical images, not to detect the affected part. A broken bone in the contour may be different from the healthy one. Thus we choose a representation that can estimate the partial similarity of two images and can be easy to calculate their global similarity.

We analyze the high frequency part by our modified correlogram algorithm. The definition of the correlogram [28][47] is as Eq. (8). Let D denote a set of fixed distances {d_1, d_2, d_3,…, d_n}. The correlogram of an image *I* is defined as the probability of a color pair (ci, cj) at a distance d.

$$\gamma^{d}_{c_i,c_j}(I) = \Pr_{p_1 \in c_i, p_2 \in I} \{ p_2 \in c_j \| \, p_1 - p_2 \, |= d \} \tag{8}$$

For computational efficiency, the autocorrelogram is defined as Eq. (9)

$$\lambda^{d}_{c_i}(I) = \Pr_{p_1 \in c_i, p_2 \in I} \{ p_2 \in c_i \| \, p_1 - p_2 \, |= d \} \tag{9}$$

The contrast of a gray image dominates human's perception. If two images have different gray levels they still may be visually similar. Thus the correlogram method cannot be used directly.

Our modified correlogram algorithm works as follows. First we sort the pixels of the high frequency part decadently. Then we order the results of the preceding sorting by the ascendant distances of pixels to the center of the image. The distance of a pixel to the image center is measured by the L2 distance. After sorting by gray value and distance to the image center, we select the top 20 percent of pixels and the gray values higher than a threshold to estimate the autocorrelogram histogram. We set the threshold zero in this task. Any two pixels have a distance, and we estimate the probability that the distance falls within an interval. The distance intervals we set are {[0,2], (2,4], (4,6], (6,8], (8,12], (12,16], (16,26], (26,36], (36,46], (46,56], (56,66],(66,76] (76,100]}. The high frequent part comprises 64*64 pixels, thus the maximum distance will be smaller than 100. The first n pixels will have n*(n+1)/2 numbers of distances. We calculate the probability of each interval to form the correlogram vector similarity metric

2.1.6 Gabor Texture Features

Gabor filter is widely adopted to extract texture features from images for image retrieval [37], and has been shown to be very efficient. Gabor filters are a group of wavelets, with each wavelet capturing energy at a specific frequency and a specific direction. Expanding a signal using this basis provides a localized frequency description, therefore capturing local features/energy of the signal. Texture features can then be extracted from this group of energy distributions. The scale (frequency) and orientation tunable property of Gabor filter makes it especially useful for texture analysis.

The Gabor wavelet transformation W_{mn} of Image I(x, y) derived from Gabor filters according to [37] is defined in Eq. (10)

$$W_{mn}(x,y) = \int I(x,y) g^{*}_{mn}(x - x_1, y - y_1) dx_1 dy_1. \tag{10}$$

The mean μ_{mn} and standard deviation σ_{mn} of the magnitude $|W_{mn}|$ are used for the feature vector, as shown in Eq. (11).

$$\mu_{mn}(x,y) = \iint |W_{mn}(x,y)|\, dxdy, \text{ and } \delta_{mn} = \sqrt{\iint (|W_{mn}(x,y)| - u_{mn})^2 dxdy}. \tag{11}$$

This image feature is constructed by μ_{mn} and σ_{mn} of different scales and orientations. Our experiment uses four (S=4) as the scale and six (K=6) as the orientation to construct a 48 feature vectors $\bar{f}$, as shown in Eq. (12).

$$\bar{f} = [u_{00}, \delta_{00}, u_{01}, \delta_{01}, \ldots\ldots, u_{35}, \delta_{35}]. \tag{12}$$

2.1.7 Similarity Metric

While an image has features to represent it, we need a metric to measure the similarity between two feature vectors (and consequently, the similarity between two images). The similarity metric of color histogram is defined as Eq. (3) and that of gray-spatial histogram is defined as Eq. (13):

$$\mathrm{SIM}_{\mathrm{gray\_spatial}}(\mathrm{H(q)},\ \mathrm{H(d)}) = \frac{\mathrm{H(q)} \cap \mathrm{H(d)}}{|\ \mathrm{H(q)}\ |} = \frac{\sum_{i=1}^{n} \min(h_i(q), h_i(d))}{\sum_{i=1}^{n} h_i(q)} \tag{13}$$

The similarity metric of the coherence moment is defined as Eq. (14)

$$\begin{aligned} DIS_{coh}(COH(q),COH(d)) = \sum_{i=1}^{4classes}(&|COH_{\kappa}^{q_i} - COH_{\kappa}^{d_i}| + |COH_{\mu}^{q_i} - COH_{\mu}^{d_i}| \times |COH_{\rho}^{q_i} - COH_{\rho}^{d_i}| + \\ &|COH_{o}^{q_i} - COH_{o}^{d_i}| + |(COH_{\nu}^{q_i})^{1/2} - (COH_{\nu}^{d_i})^{1/2}| + \\ &|COH_{\tau}^{q_i} - COH_{\tau}^{d_i}| \times 2 + |(COH_{\omega}^{q_i})^{1/2} - (COH_{\omega}^{d_i})^{1/2}|) \end{aligned} \tag{14}$$

The correlogram metric is defined as Eq. (15):

$$\mathrm{DIS}_{\mathrm{hf}}(\mathrm{H(q)},\ \mathrm{H(d)}) = \frac{\sum_{i=1}^{n} |h_i(q) - h_i(d)|}{\sum_{i=1}^{n} |h_i(q) + h_i(d)|} \tag{15}$$

The similarity of two images Q and D is measured by Eq. (16):

$$\begin{aligned} \mathrm{SIM}_{\mathrm{image}}(\mathrm{Q, D}) = \mathrm{W}_1 \times \mathrm{SIM}_{\mathrm{color}}&(\mathrm{H(Q),H(D)} + \\ &\mathrm{W}_2 \times \mathrm{SIM}_{\mathrm{gray\text{-}spatial}}(\mathrm{H(Q),H(D)} + \\ &\mathrm{W}_3 \times 1/(1+\mathrm{DIS}_{\mathrm{coh}}(\mathrm{COH(Q),COH(D)})) + \mathrm{W}_4 \times \\ &1/(1+\mathrm{DIS}_{\mathrm{hf}}(\mathrm{COH(Q),COH(D)})), \end{aligned} \tag{16}$$

where W_i is the weight of feature *i*. In this task the database contains color and gray images. When the user queries an image by example, we first determine whether the example is color or gray. We calculate the color histogram, if the four bins of gray values occupy more than 80% of the whole image, we decide the query image is gray; otherwise it is color. If the input is a color image, then we set W_1=10, W_2=0.1, W_1=10, and W_1=10; otherwise we set W_1=0.1, W_2=1, W_1=100, and W_1=100.

2.2 The Experiment and Result

We use the ImageCLEF 2004 evaluation to evaluate the performance of our system. The dataset for the medical retrieval task is called CasImage and consists of 8,725 anonymised medical images, e.g. scans, and X–rays from the University Hospitals of Geneva. The majority of images are associated with *case notes*, a written description of a previous diagnosis for an illness the image identifies. Case notes are written in XML and consist of several fields including: a diagnosis, free-text description, clinical presentation, keywords and title. The task is multilingual because case notes are mixed language written in either English or French (approx. 80%).

An example case notes field for description and corresponding images is shown in Figure 2-4. Not all case notes have entries for each field and the text itself reflects real clinical data in that it contains mixed–case text, spelling errors, erroneous French accents and un–grammatical sentences as well as some entirely empty case notes. In the dataset there are 2,078 cases to be exploited during retrieval (e.g. for query expansion). Around 1,500 of the 8,725 images in the collection are not attached to case notes and 207 case notes are empty. The case notes may be used to refine images which are visually similar to ensure they match modality and anatomic region.

The process of evaluation and the format of results employ the trec_eval tool. There are 26 queries. The corresponding answer images of every query were judged as either relevant or partially relevant by at least 2 assessors.

```
<?xml version='1.0' encoding='iso-8859-1' ?>
<CASIMAGE_CASE>
<ID>
2526
</ID>
<Description>
Bassin du 28.02.1985 :

Status avant et aprËs rÉduction. Avant rÉduction, luxation
compl`Ete du fÉmur, avec fracture avec fragments du cotyle.
Apr`Es rÉduction, interposition de l'un de ces fragments entre
la tte fÉmorale et le toit du cotyle.

</Description>

<Diagnosis>
Luxation postÉrieure du fÉmur gauche associÉe ? une fracture
multifragmentaire d
</Diagnosis>
........
```

Figure 2-4: An example of medical case note and associated images

In this task, we have two experiments. The first run, VIS, uses the visual features of the query image to query the database. The second run, VWF, is the result where the user manually selects the relevant images as positive examples. In the results summary, the mean average precision of the first run VIS of our system is 0.3788. The mean average precision of run2 (VWF) is 0.4474 (refer to appendix). Figure 2-5 shows the precision and recall graphs.

The results show that the image features we propose can represent the medical image content well. The medical image's background is very similar. Relevance feedback can extract the dominant features; thus it can improve the performance strongly. This result is better than that of the GIFT system (MAP=0.3791), used as the baseline for medical imageCLEF 2004 [12]. It is also even better than the best work in visual method (MAP=0.4214) [12]. For convenience, we have tabulated them in table 2-1 where '1','2','3' indicate the top 3 result of the medical imageCLEF 2004 published in [12], respectively.

In this paper we consider that the contrast of a grayscale image dominates human perception. We use a relative normalization method to reduce the impact of illumination. Figure 2-6 is the result of an example query returned by our system. It can be observed from Figure 2-6: that F_10/9719.jpg and F_17/16870.jpg are darker than the query image (def_queries/1.jpg), but our system still can find them out. Figure 2-7 and Figure 2-8 are some of the query results of our system.

Rank	1	2	3	Ours
MAP	0.4214	0.4189	0.3824	0.4474

Table 2-1 Results reported at imageCLEF2004

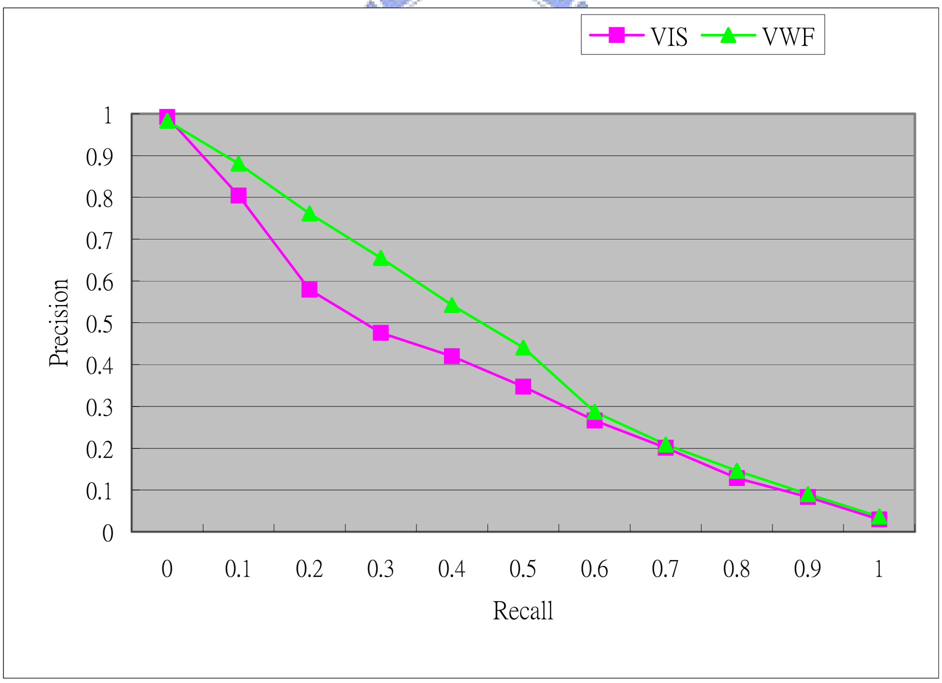


Figure 2-5: Precision vs. Recall graphs without and with feedback. VIS is the result without relevance feedback. VWF is the result with manual feedback

Figure 2-6: Result of an example query 'Pelvic'

Figure 2-7: Result of example query 'lung'

Figure 2-8: Result of example query ‘hand’

The first run has accuracy above 50% in the first 20 images. The really similar images may have similar features in some aspect. The misjudged images are always less consistent. So we try to refine the initial result by the automatic feed back mechanism. We cluster the first 20 images into six classes. If the class contains diverse images, the center of the class will become farther, and consequently more different, from the query image. Thus we can improve the result by our feedback method.

In this section we propose several image features to represent medical images. Although the color histogram of content-based image retrieval methods has good performance in general-purpose color images, unlike general-purpose color images the X-ray images only contain gray level pixels. Thus, we concentrate on the contrast representation of images.

The image representations we propose have obtained good results in this task. Our representation is immune to defective illumination. A total of 322 features are used. It is very efficient in computation.

An image represents thousands of words. An image can be viewed from various aspects; furthermore, different people may have different interpretations of the same image. This means that many parameters need to be tuned. In the future, we will try to learn the user behavior and tune those parameters by learning methods.

Chapter 3

Combine Text and Visual Feature for Medical Image Retrieval

Images by their very nature are language independent, but often they are accompanied by texts semantically related to the image (e.g. textual captions or metadata). Images can then be retrieved using primitive features based on pixels with form the contents of an image (e.g. using a visual exemplar), abstracted features expressed through text or a combination of both. The language used to express the associated texts or textual queries should not affect retrieval, i.e. an image with a caption written in English should be searchable in languages other than English.

There is an increasing amount of full text material in various languages available through the Internet and other information suppliers. The world of globalization is increasing, many countries have been unified. The European project to unify European countries is a very important example in order to eliminate broader for the cooperation, global and large market, real international and free business. The high-developed technologies in network infrastructure and Internet set the platform of the cooperation and globalization. Indeed, the business should be global and worldwide oriented. Thus, the issues of the multilinguality arise in order to overcome the remaining technical barriers that still separate countries and cultures.

There are 6,700 languages spoken in 228 countries and English is used as the native language only 6 % of the World population, but English is the dominant language of the collections, resources and services in the Internet. Actually, the

English language is widely used on the Internet. About 60 % of the world online population is represented by English and 30% by European languages [44]. Approximately 147 M people are connected to the Internet (US and Canada about 87 M, Europe 33.25 M, Asia/Pacific 22 M, and Africa 800,000). However, the size of Web sites and Internet users from other countries (non-English countries) is increasing progressively, so that the multilingual problem becomes more and more important.

Recently, a lot of digital libraries will be set up containing large collections of information in a large number of languages. However, it is impractical to submit a query in each language in order to retrieve these multilingual documents. Therefore, a multilingual retrieval environment is essential for benefiting from worldwide information resources. Most research and development activities are focused on only one language. But, this is not the objective of the digital libraries and Internet philosophy. Both technologies aim at establishing a global digital library containing all information resources from different areas, different countries and in different languages.

The access to those materials should be possible for the worldwide community and never be restricted because of non-understanding languages. Thus, the EU funded some projects addressing the multilingual issues. For instance, in the ESPRIT project EMIR (European Multilingual Information Retrieval) a commercial information retrieval system SPIRIT has been developed which supports French, English, German, Dutch and Russian. UNESCO launched some projects in order to democratize and globalize the access to the world cultural patrimony, such as Memory of the World. Recently, UNESCO has started the MEDLIB[4] project which aims at creating a virtual library for the Mediterranean region.

[4] http://www.unesco.org

Cross-language Information retrieval allows user query document in another language. In 1996, ACM SIGIR Workshop for Multilingual Information Retrieval names it as Cross-Language Information Retrieval. Many international conferences focus on Cross-Language information retrieval issues, like as ACL Annual Meeting，ACM SIGIR99 (SIGIR00), etc. The organizations TREC, CLEF, and NTCIR are three major Cross-Language Evaluations to evaluate Cross-language technologies.

3.1 Previous Work for Cross-Language Document Retrieval

The approach of cross-language information retrieval allows a user to formulate a query in one language and to retrieve documents in others. The controlled vocabulary is the first and the traditional technique widely used in libraries and documentation centers. Documents are indexed manually using fixed terms which are also used for queries. These terms can be also indexed in multiple languages and maintained in a so-called thesaurus.

Using dictionary-based technique queries will be translated into a language in which a document may be found. The corpus-based technique analyzes large collections of existing texts and automatically extracts the information needed on which the translation will be based. However, this technique tends to require the integration of linguistic constraints, because the use of only statistical techniques by extracting information can introduce errors and thus achieve bad performance [24]. Latent Semantic Indexing (LSI) is a new approach and a new experiment in multilingual information retrieval which allows a user to retrieve documents by concept and meaning and not only by pattern matching. In following, we will review those three approaches.

3.1.1 Dictionary-Based Approach

The size of public domain and commercial dictionaries in multiple languages on the Internet is increasing steadily. As an example we cite a few of them: Collins COBUILD English Language Dictionary and its series in major European languages, Leo Online Dictionary, Oxford Advanced Learner's Dictionary of Current English, and Webster’s New Collegiate Dictionary.

Electronic monolingual and bilingual dictionaries build a solid platform for developing multilingual applications. Using dictionary-based technique queries will be translated into a language in which a document may be found. However, this technique sometimes achieves unsatisfactory results because of ambiguities. Many words do not have only one translation and the alternate translations have very different meanings. Moreover, the scope of a dictionary is limited. It lacks in particular a technical and topical terminology which is very crucial for a correct translation. Nevertheless, this technique can be used for implementing simple dictionary-based application or can be combined with other approaches to overcome the above mentioned drawbacks.

Using electronic dictionary-based approach for query translation has achieved an effectiveness of 40-60% in comparison with monolingual retrieval [4] [29]. In [27] a multilingual search engine, called TITAN, has been developed. Based on a bilingual dictionary it allows to translate queries from Japanese to English and English to Japanese. TITAN helps Japanese users to search in the Web using their own language. However, this system suffers again from polysemy.

3.1.2 Corpus-based Technique

The corpus-based technique seems to be promising. It analyzes large collections of existing texts (corpora) and automatically extracts the information needed on which the translation will be based. Corpora are collections of information in electronic form to support e.g. spelling and grammar checkers, and hyphenation routines (lexicographer, word extractor or parser, glossary tools).

These corpora are used by researchers to evaluate the performance of their solutions, such TREC collections for cross-language retrieval. A few examples of mono, bi- and multilingual corpora are Brown Corpus, Hansard and United Nation documents respectively. The Hansard Corpus[5] contains parallel texts in English and Canadian French collected during six years by the Canadian Parliament. The Brown Corpus consists of more than one million words of American English. It was published 1961 and it is now available at the ICAME[6] (International Computer Archive of Modern English). Interested readers are referred to the survey about electronic corpora and related resources in [17].

In the United States, the WordNet Project at Princeton has created a large network of word senses in English related by semantic relations such as synonymy, part-whole, and is-a relation [19] and [39]. Similar work has been launched in Europe, called EuroWordNet [21]. These semantic taxonomies in EuroWordNet, have been developed for Dutch, Italian and Spanish and are planned to be extended to other European languages. Related activities have been launched in Europe, such as ACQUILEX[7] (Acquisition of Lexical Knowledge for Natural Language Processing Systems), ESPRIT MULTILEX[8] (Multi-Functional Standardized Lexicon for

[5] http://morph.ldc.upenn.edu/ldc/news/release/hansard.html
[6] http://www.hd.uib.no/icame.html
[7] http://www.cl.cam.ac.uk/Research/NL/acquilex
[8] http://www.twente.research.ec.org/esp-syn/text/5304.html

European Community Languages).

The main problems associated with dictionary-based CLIR are (1) untranslatable search keys due to the limitations of general dictionaries, (2) the processing of inflected words, (3) phrase identification and translation, and (4) lexical ambiguity in source and target languages. The category of untranslatable keys involves new compound words, special terms, and cross-lingual spelling variants, i.e., equivalent words in different languages which differ slightly in spelling, particularly proper names and loanwords. In this dissertation *translation ambiguity* refers to the increase of irrelevant word senses in translation due to lexical ambiguity in the source and target languages.

The collection may contain parallel and/or comparable corpora. A parallel corpus is a collection which may contain documents and their translations. A comparable corpus is a document collection in which documents are aligned based on the similarity between the topics which they address. Document alignment deals with documents that cover similar stories, events, etc. For instance, the newspapers are often describing political, social, economical events and other stories in different languages. Some news agencies spend a long time in translating such international articles, for example, from English to their local languages (e.g. Spanish, Arabic). These high-quality parallel corpora can be used as efficient input for evaluating cross-language techniques.

Sheridan and Ballerini developed an automatic thesaurus construction based on a collection of comparable multilingual documents [62]. Using the information retrieval system Spider this approach has been tested on comparable news articles in German and Italian (SDA News collection) addressing same topics at the same time. Sheridan and Ballerini reported that queries in German against Italian documents achieve about 32% of the best Spider performance on Italian retrieval, using relevance feedback. Other experiments on English, French and German have been presented in [72]. The document alignment in this work was based on indicators, such as proper nouns, numbers, dates, etc. There is also alignment based on term similarity as in latent semantic analysis. This allows mapping text between those documents in different languages.

3.1.3 Indexing by Latent Semantic Analysis

In previous approaches the ambiguity of terms and their dependency leads to poor results. Latent Semantic Indexing (LSI) is a new approach and a new experiment in multilingual information retrieval which allows a user to retrieve documents by concept and meaning and not only by pattern matching. If the query words have not been matched, this does not mean that no document is relevant. In contrast, there are many relevant documents which, however, do not contain the query term word by word. This is the problem of synonymy. The linguist will express for example his request differently as computer scientist. The documents do not contain all possible terms that all users will submit. Using thesauri to overcome this issue remains ineffective, since expanding query to unsuitable terms decreases the precision drastically.

The latent semantic indexing analysis is based on singular-value decomposition [16]. Considering the term-document matrix terms and documents that are very close will be ordered according to their degree of "semantic" neighborhood. The result of LSI analysis is a reduced model that describes the similarity between term-term, document-document, and term-document relationship. The similarity between objects can be computed as cosine between their representing vectors.

The results of the LSI approach have been compared with those of a term matching method (SMART). Two standard document collections MED and CISI (1033 medical documents and 30 queries, 1460 information science abstracts and 35 queries) have been used. It has been showed that LSI yields better results than term matching.

Davis and Dunning [14] have applied LSI to cross-language text retrieval. Their experiments on the TREC collection achieved approximately 75 % of the average precision in comparison to the monolingual system on the same material [15]. The collection contains about 173,000 Spanish newswire articles. 25 queries have been translated from Spanish to English manually. Their results reported in TREC-5 showed that the use of only dictionary-based query expansion yields approx. 50 % of the average precision in comparison to results of the multilingual system. This degradation can be explained by the ambiguity of term translation using dictionaries.

This technique has been used by Oard [45] as the basis for multilingual filtering experiments, and encouraging results have been achieved. The representation of documents by LSI is “economical” through eliminating redundancy. It reduces the dimensionality of document representation (or modeling), and the polysemy as well. However, updating (adding new terms and documents) in representation matrices is time-consuming.

3.2 Combine Text and Visual Feature for Medical Image Database

In this proposed we want to design a cross-language medical image retrieval system. The data for the experiment were taken from a medical case database called casImage [55]. This database contains almost 9000 images from more than 2000 medical cases. Images contain annotation in XML format but these annotations are very rudimentary and are not at all controlled with respect to quality or fields that have to be filled in. About 10% of the records do not contain any annotation. A majority of the documents are in French (70%), with around 20% being in English.

Figure 3-1 shows a few example images from the database. These images are among others the query topic for the performance evaluation. Figure 3-2 is the data model of the medical image retrieval system. The database contains multilingual patient diagnosis documents. User can use native language to retrieve other language patient cases. Our proposed system allows user can use keyword or medical image example to query medical image cases.

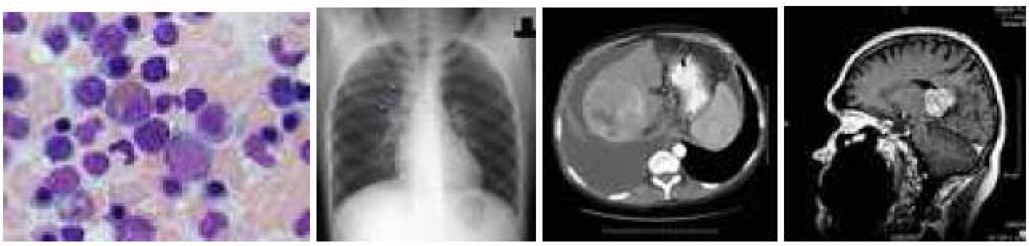


Figure 3-1: Example from the CasImage collection

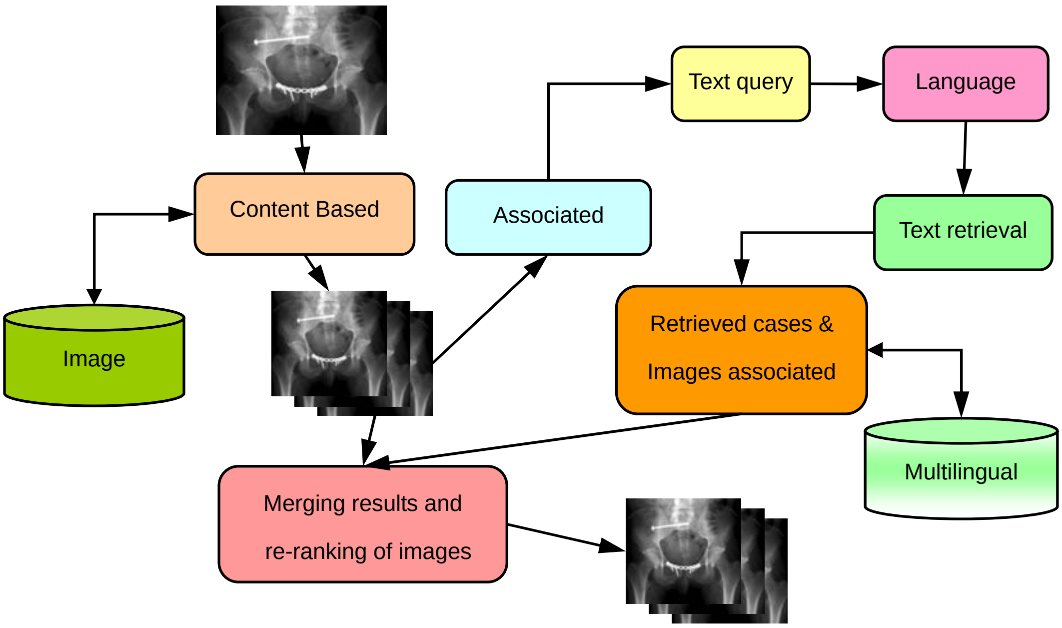


Figure 3-2: The data model of medical image retrieval system

In the ImageCLEFmed collections of annotations are in English, French and German. The overall multilingual search process is shown in Fig. 3-3. Given an initial query Q, the system performs the cross-language retrieval, and returns a set of relevant documents to user. We use the representation expressing a query as a vector in the vector space model [59]. The Textual Vector Representation is defined as following. Let W ($|W| = n$) be the set of significant keywords in the corpus. For a document D, its textual vector representation (i.e., D_T) is defined as Eq. (17),

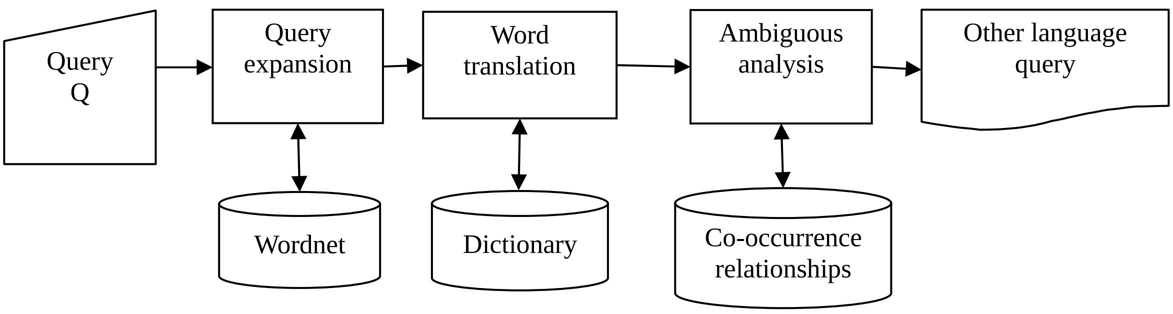


Figure 3-3: Text based multilingual query translation flowchart

$$D_T = < w_{t_1}(D_T), \ldots, w_{t_n}(D_T) > \qquad (17)$$

Where the n dimensions indicate the weighting of a keyword t_i in D_T, which is measured by TF-IDF [59], as computed in Eq. (18);

$$w_{t_i}(D_T) = \frac{tf_{t_i,D_T}}{\max tf} \times \log \frac{N}{n_{t_i}} \tag{18}$$

In Eq.(18), $\frac{tf_{t_i,D_T}}{\max tf}$ stands for the normalized frequency of t_i in D_T, max tf is the maximum number of occurrences of any keyword in D_T, N indicates the number of documents in the corpus, and n_{t_i} denotes the number of documents in whose caption t_i appears.

In the above, we introduce the textual vector representation for documents. As for a query Q, one problem is that since Q_T is given in English, it is necessary to translate Q_T into French and German, which is the languages used in the document collection.

A short query usually cannot cover as many useful search terms as possible because of the lack of sufficient words. We perform the query expansion process to add new terms to the original query. The additional search terms is taken from a thesaurus – WordNet [39]. For each expansion English term, it is then translated into one or several corresponding French and German words by looking it up in a dictionary[9].

Assume $AfterExpansion(Q_T) = \{e_1, ..., e_h\}$ is the set of all English words obtained after query expansion and query translation, it is obvious that $AfterExpansion(Q_T)$ may contain a lot of words which are not correct translations or useful search terms. To resolve the translation ambiguity problem, we define *word co-occurrence relationships* to determine final query terms. If the co-occurrence frequency of e_i and e_j in the corpus is greater than a predefined threshold, both e_i and e_j are regarded as useful search terms.

So far, we have a set of search terms, $AfterDisambiguity(Q_T)$ which is presented

[9] http://www.freelang.net/

as Eq.(19),

$$AfterDisambiguity(Q_T) = \{e_i, e_j \mid e_i, e_j \in AfterTranslation(Q_T) \\ \& \; e_i, e_j \text{ have a significant co-occurrence}\} \quad (19)$$

After giving the definition of $AfterDisambiguity(Q_T)$, for a query Q, its textual vector representation (i.e., Q_T) is defined in Eq. (20),

$$Q_T = < w_{t_1}(Q_T), \ldots, w_{t_n}(Q_T) > \quad (20)$$

Where $w_{t_i}(Q_T)$ is the weighting of a keyword t_i in Q_T, which is measured as Eq. (10), $w_{c_i}(Q_T)$ indicates whether there exists an $e_j \in AfterDisambiguity(Q_T)$.

In Eq. (21), W is the set of significant keywords as defined before, $\frac{tf_{t_i,Q_T}}{\max tf}$ stands for the normalized frequency of t_i in $AfterDisambiguity(Q_T)$, maxtf is the maximum number of occurrences of any keyword in $AfterDisambiguity(Q_T)$, N indicates the number of images in the corpus, and n_{t_i} denotes the number of images in whose caption t_i appears.

$$w_{t_i}(Q_T) = \left\{ \frac{tf_{t_i,Q_T}}{\max tf} \times \log \frac{N}{n_{t_i}} \right. \quad (21)$$

3.3 Reduce the Cross-Language Translation Ambiguity

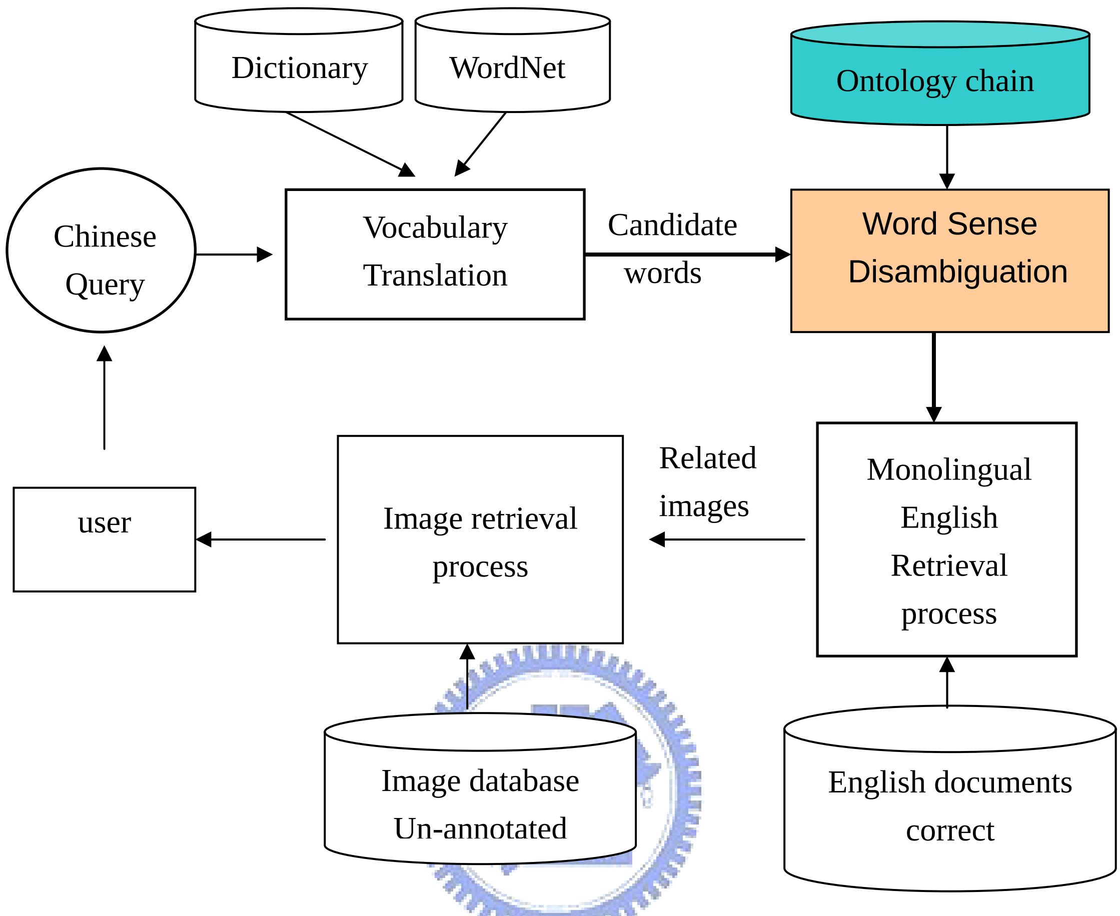


Figure 3-4: The cross medical image system architecture

Figure 3-4 is the overview of our cross medical image system architecture. User can use Chinese terms to retrieve related medical document. The associated medical images of document can be used as image query example. The image process retrieves the similar image by image features.

The Dictionaries and WordNet approach are designed for general domain resource and the creation time is very expansive. The relationship definitions are still not enough to gain good results. The ambiguous translation is the main problem that gets bad result. Davis (1997) had discussed that use category will reduce the ambiguity problem. In this proposed we want to solve the ambiguity problem by using proposed ontology chain method.

In the corpus approach need large amount parallel bilingual document for reference. In this proposed we try to find similar concept bilingual document by similar medical images. When user uses Chinese language to retrieve English language document case, we can get relevance medical images. In the multilingual database, we can create a relationship between English document and French document by similar medical image. Those document have similar medical images, it must have similar concept in some aspect.

3.3.1 Cross-Language Retrieval System Query by Keyword

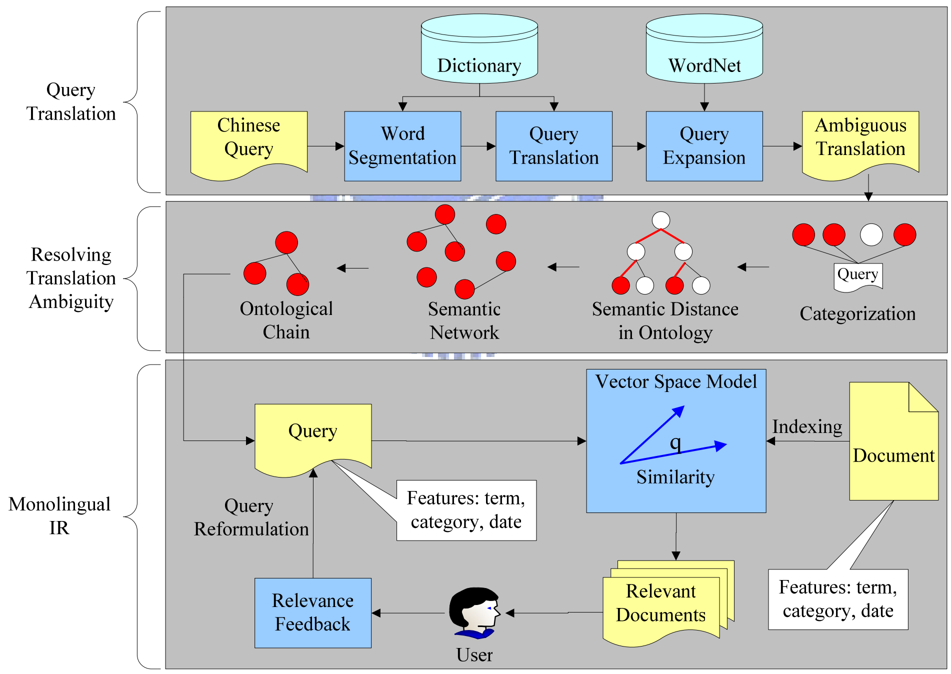


Figure 3-5: Proposed system flowchart

Figure 3-5 is the proposed cross-language system flowchart for keyword based retrieval. It contains three modules that are: Query Translation, Resolving Translation Ambiguity, and Monolingual Information Retrieval. Describe detail as following:

- Query Translation: Translate the Chinese query keyword into possible English words by dictionary and its synonym, Hypernym and hyponym expansion by WordNet.

- Resolving Translation Ambiguity: After query Translation process, the produced possible translation English words may cause ambiguous. We proposed an ontology chain approach to conquer the ambiguity problem. The ontology was created by expert. We use the category to construct an ontology map. We use ontology chain to refine the candidate translated keywords, puck up the meaningful candidate keyword as the query terms for monolingual retrieval.
- Monolingual Retrieval Process: Find the document that related to the query terms.

Ontology is a formal explicit specification. It constructs a common conceptualization for users. Ontology contains principal concept and relationship among concepts, especially usefully in specific domain. Figure 3-6 is an example of ontology. In our previous works, we use the St Andrews data set from ImageClef2004 to evaluate our proposed methods. The St Andrews dataset consists of 28,133 photographs from St Andrews University Library photographic collection which holds one of the largest and most important collections of historic photography in Scotland.

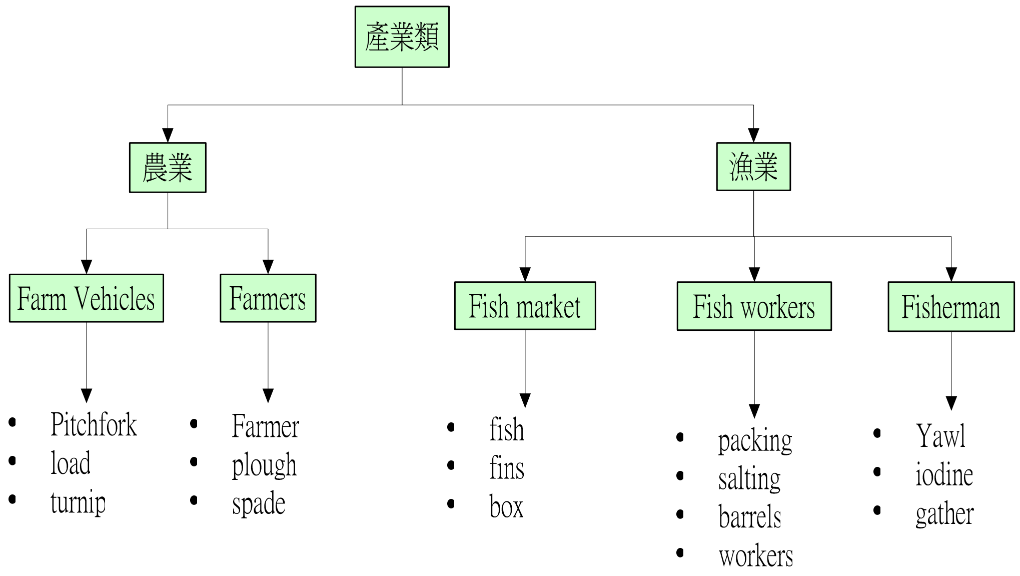


Figure 3-6: An example of ontology

The St Andrews data set from CLEF2004, each document belong to more than one category. It contains 946 categories. We arrange the categories to construct an ontology map that define the relationship between each category. Creating the ontology by category is easier than construct ontology by terms in time complexity.

When the translations of query terms contain multiple meaning will cause ambiguity result. We use the ontological chain to evaluate the most appropriate meaning.

The related similarity between ontology node L_i and query express Q is defined as Eq. (22).

$$Sim(L_i, Q) = \sqrt{\frac{\sum_{j=1}^{N} t_{ij}^{2}}{N}} \tag{22}$$

Where t_{ij} is the frequencies of translated English terms, N is the number of Chinese query terms.

We let arbitrary two node of ontology have a distance. For example as show in Figure 3-6 the “herring” and “fish processing” have a path 3. Based on the similarity of query terms and the distance we can define a relationship between two nodes.

$$\mathrm{Rel}(L_i, L_j) = \frac{Sim(L_i, Q) \times Sim(L_j, Q)}{distance(L_i, L_j)} \tag{23}$$

Figure 3-7 is a relationship map between nodes. We set a threshold 1.5 to eliminate less important relationships. The connected nodes will construct many semantic concepts. We can eliminate the translated English term that do no fall in the connected nodes. The sum of relationships of each island connected nodes can be used as important degree for ranking candidate English terms.

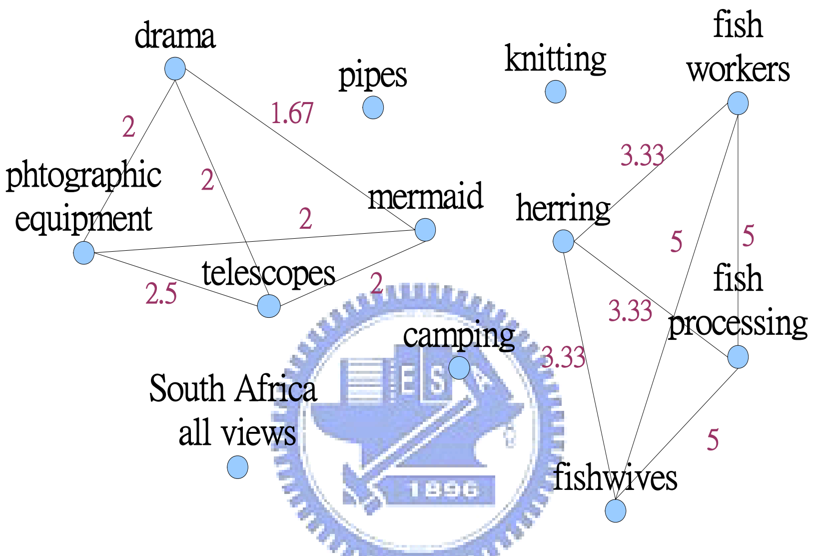


Figure 3-7: Semantic relationship map

3.4.1 Experiment Result for Medical Image Collections

The entire ImageCLEFmed *library* consists of multiple collections. Each *collection* is organized as cases that represent a group of related images and annotations. Each *case* consists of a group of images and an optional annotation. Each *image* is part of a case and has optional associated annotations consisting of metadata (e.g., HEAL tagging) and/or a textual annotation.

In practical medical images retrieval systems, it is natural for doctors to retrieve the related images by the descriptions of requirement. After textual query, the system retrieves related images by annotations for users to browse. The user can further rank the images in visual. Combining the textual and visually features will help the user to find the desire image more precise. In ImageCLEF 2005, medical image retrieval task contains 25 queries for evaluation, the queries mixed visual images and semantic textual to retrieve desire images. The visual queries use image examples to find similar images, each topic contain at least one image example. The semantic textual queries allow user query by a sentence, which high-level semantic concept are hard to be derived from images directly. The goal of this task is to examine how the visual feature can improve the query result.

All submissions of participants in this task were classified into automatic runs and manual runs. The automatic runs mean that the system at the query process without human manual intervened. In the automatic category, the methods can be classified into three sub-categories: Text only, Visual only and Mixed retrieval (visual and textual) according to the feature used. The category "Text only" means that systems use textual feature only to retrieve relevant images. "Visual only" category means that systems only use visual image feature without combine textual annotation to retrieve similar images. Mixed retrieval means the systems combine the visual and textual feature to retrieve images.

In this task, we have submitted ten runs for the mixed retrieval of automatic runs and six runs for the visual only of automatic runs. In the content-based approach, we combine four proposed image features by weighted adjusting to retrieve related images. The weight of features we set at the system initial and do not have any further user intervention while query is processing. Table 3-1 lists the query result of visual only runs and the setting weight of four image features. Table 3-2 lists the result of mixed retrieval runs and the setting weight of image features and textual features. The difference of each runs is the weighted setting of features.

The query topics contain color and gray images. We first examine the queries image is color or gray image by **color/gray feature**. According to the image is color or gray set different weight for image features. In the Table 3-1, “C” denotes that query image is color image and “G” denotes that query image is gray image. We submit six runs for visual only category. The run, “nctu_visual_auto_a8”, has the better result in our experiment. The weight of each feature are set equal, it means that four image features have the same importance. The result also shows that visual only approach has a bottleneck because the query topics contain semantic queries.

Submission runs	The weight of Image features								Result	
Image features	Coherence		Gray HIS		Color HIS		Facade		MAP	Rank of runs
detected color or gray	C	G	C	G	C	G	C	G		
visual_auto_a1	0.3	0.2	0.3	0.5	1	0.2	1	1	0.0628	14
visual_auto_a2	0.3	0.2	0.5	0.3	0.3	0.5	1	1	0.0649	10
visual_auto_a3	0.5	1	0.5	1	1	0.5	1	1	0.0661	8
visual_auto_a5	0.1	0.2	0.1	0.5	1	0.5	0.5	1	0.0631	13
visual_auto_a7	0.3	0.2	0.3	0.5	1	0.2	1	0.5	0.0644	11
visual_auto_a8	1	1	1	1	1	1	1	1	0.0672	7

Table 3-1: The query result of visual only runs and the weight of visual image features

Submission runs	The weight of Image features													
Image features	Coherence		Gray HIS		Color HIS		Facade		visual		textual		MAP	Rank
Detected color or gray	C	G	C	G	C	G	C	G	C	G	C	G		
visual+Text_auto_1	0.3	0.2	0.3	0.5	1	0.2	1	1	1	1	0.8	0.1	0.2276	10
visual+Text_auto_2	0.3	0.2	0.5	0.3	0.3	0.5	1	1	1	1	0.8	0.1	0.2127	14
visual+Text_auto_3	0.5	1	0.5	1	1	0.5	1	1	1	1	0.8	0.1	0.2286	9
visual+Text_auto_4	0.3	0.2	0.3	0.5	1	0.2	1	1	1	1	1	0.2	0.2389	3
visual+Text_auto_5	0.1	0.2	0.1	0.5	1	0.5	0.5	1	1	1	0.8	0.1	0.2246	12
visual+Text_auto_6	0.3	0.2	0.3	0.5	1	0.2	1	1	1	1	1	1	0.2318	7
visual+Text_auto_7	0.3	0.2	0.3	0.5	1	0.2	1	0.5	1	1	0.8	0.1	0.2265	11
visual+Text_auto_8	1	1	1	1	1	1	1	1	1	1	1	1	0.2324	6
visual+Text_auto_9	1	1	1	1	1	1	1	1	1	1	0.1	0.1	0.0906	22
visual+Text_auto_10	1	1	1	1	1	1	1	1	0.1	0.1	1	1	0.1941	15

Table 3-2: The result of mixed retrieval runs and the weight of visual image features and textual features

The setting weights of mixed runs and results are listed in the Table 3-2. The result of run8, run9 and run10 illustrate that combine the visual and textual feature will get better results than single features. Run8 assume that the significant of visual and textual feature are equal. Run9 emphasizes the weight of visual features and Run10 emphasizes the weight of textual features. The result shows that text-based approach is better than content-based approach, but the content-based approach can improve the textual result.

3.4.2 Experiment Result for St Andrews collections

We test our method using the CLEF English-Chinese ImageCLEF [11] test collection. The collection includes about 30,000 historic photographs with British English semi-structured captions. ImageCLEF is the cross-language image retrieval track that is run as part of the Cross Language Evaluation Forum (CLEF) campaign. The campaign is run in a similar manner as the TREC and NTCIR information retrieval evaluations. This is a bilingual query translation task from a source language to English. The goal of this task lies in finding as many relevant images as possible from the St. Andrews image collection.

Each image has an accompanying textual description consisting of 8 distinct fields. These fields can be used individually or collectively to facilitate image retrieval. The 28,133 captions consist of 44,085 terms and 1,348,474 word occurrences; the maximum caption length is 316 words, but on average 48 words in length. All captions are written in British English, although the language also contains colloquial expressions. Approximately 81% of captions contain text in all fields, the rest generally without the description field. In most cases the image description is a grammatical sentence of around 15 words.

Each image in the St Andrews collection has been assigned to one or more descriptive categories, e.g. “airports”, “airships”, “flowers”, “beach scenes” and “breweries”. On average, each image is assigned to 4 categories. We use the categories to manually construct the ontology. Related categories will be gathered into a class from bottom to up by experts to construct a hierarchic structure, which is the ontology used in this paper. Totally the ontology has 946 categories.

The language used to express the associated texts or textual queries should not affect retrieval, i.e. an image with a caption written in English should be searchable in languages other than English. In the experiment we will show that ontology approach will improve the precision by 13%, which is better than the result without dis-ambiguity process.

The cross language evaluation forum includes data set, 25 query topics and query answer to evaluate related systems. We evaluate three models; they are Mono-lingual IR (Mono-lingual Information Retrieval), CLIR (Cross Language Information Retrieval), and Ontology-based CLIR. The Mono-lingual IR uses the original English queries with expansion by WordNet to retrieve related documents; it is the baseline in evaluation.

The CLIR model uses Chinese queries to retrieve related documents; it first translates the Chinese terms to all possible English terms by a dictionary. After translation, the CLIR model also expands synonym by WordNet and uses the possible English terms as query without dis-ambiguous analysis. The CLIR model does not consider the ambiguity problem that is used to compare with Ontology-based CLIR model. The Ontology-based CLIR analysis the candidate terms based on Ontology chain trying to reduce the ambiguity translation problem.

The performance is normally evaluated based on precision and recall as defined in the following equations:

$$\text{precision} = \frac{\textit{number} \text{ of correct answer retrieved}}{\text{total number of retrieved}},$$

$$recall = \frac{\textit{number} \text{ of correct answer retrieved}}{\textit{total number of} \text{ actual answer}}.$$

We used the UMASS and Lemur versions of the standard trec_eval tool to compute the precision, recall and mean average precision scores. This provides the "standard" information retrieval evaluation measures, e.g. precision at a given rank cut-off, average precision across 11 recall points, and single-valued summaries for each measure.

Figure 3-8 is the precision/recall of experiment, the result show that ontology-based CLIR approach is better than normal CLIR and close to the Mono-lingual IR.

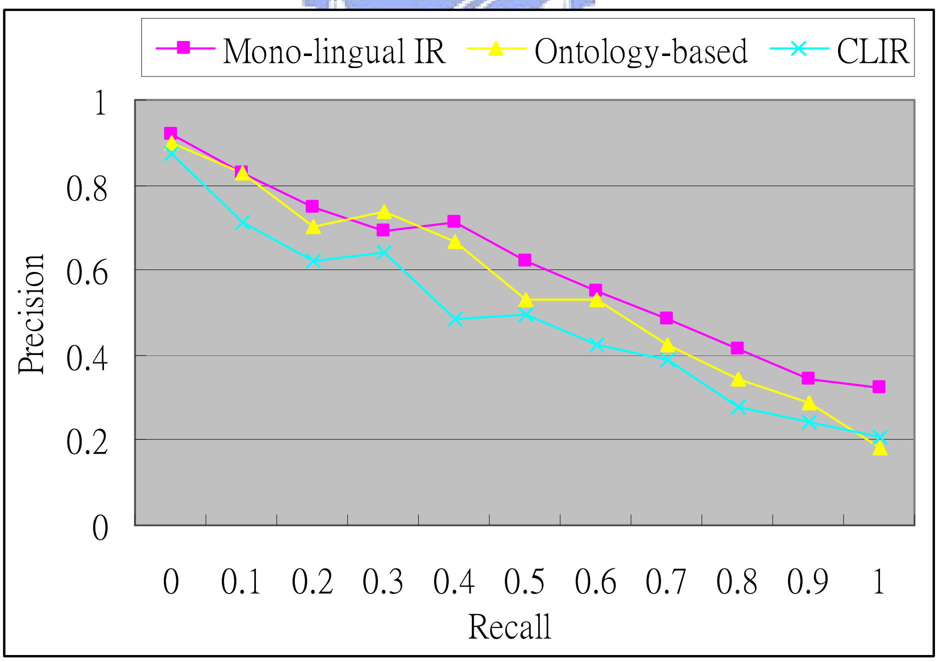


Figure 3-8: The precision vs. recall

Table 3-3 is the experiment result. The mean average precision (MAP) of normal CLIR is 49.18% and ontology based CLIR is 55.18. The Mono-lingual IR directly uses native language for query without translation ambiguity for comparison. It shows that Ontology-based CLIR improve the performance from 49% to 55 %, and reach to the 92% precision of mono lingual IR. The mono-lingual IR using the original English query avoids the translator ambiguity, so that it has the best performance in the experiment result.

The normal CLIR translated Chinese to English may have miss translation or ambiguity translated will cause find more irrelevant documents lowering the precision. And the ontology based CLIR furthermore filter out irrelevant translated terms. The experiment result shows that the Ontology chain is useful to recognize the acceptation of query. Translated terms must have the same concept to consist a query. Based on the specific domain, compare to the query context can reducing the ambiguity. The proposed Ontology based CLIR effective to locate the correct translation increasing a 10% rate comparing to monolingual IR.

	Mono IR	CLIR	Ontology
MAP	60.63%	49.18%	55.18%

Table 3-3: The mean average precision

A novel ontology-based approach for reducing the ambiguity of bilingual translation has been presented in this paper. Using a dictionary to translate a Chinese term into English may cause the problem of translation ambiguity. We successfully find the best translation by analyzing the relevance of all possible translated terms. A metric based on the ontology for co-occurrence concept evaluation has been proposed. The experimental results show that the proposed approach can effectively remove the irrelevant results and improve the precision.

The Ontology is very useful in the synonym and polysemy analysis. A word in different domain may have different lexical meaning. The ontology is domain dependent. In this paper we define the ontology by manual for St. Andrews corpus; we now try to cluster the words into concept capable of automatic constructing the ontology.

In the WWW application, each page can be viewed as an ontology node. In the digital library the similar pages will be placed together. We can reference the link to construct related concept. Our proposed mechanism will easy apply to WWW environment.

Chapter 4

Medical Image Retrieval with Relevance Feedback

In this chapter we will describe the relevance feedback mechanism help to improve the retrieval result for image query. In this dissertation we design a new model for user to do relevance feedback in image retrieval. The proposed feedback method not only allow user to choice positive examples and negative examples but also can observe the features of user's interest. We expect that our new relevance feedback mechanism will get more information from user and fast tune the system to suit user's interesting.

Content-based image retrieval (CBIR) is a process to find images similar in visual content to a given query from an image database. It is usually using low level features, such as color, texture or shape features, extracted from the images themselves to compare similarity. There is much research addressing the performance of content-based image retrieval method are still limited especially in the two aspects of retrieval accuracy and response time.

The limited retrieval accuracy is because of the gay between semantic concepts and low-level image features, which is a problem in content-based image retrieval. An image can be explained in many aspects by different people. For example, different types of features have different significance; user sometime want to find similar color images and sometime want to find similar shape images that is very subjective. An important issue is how to derive the weighting of different features. For the flexibility, the system in the initial weighting of features always equal, but there is no universal formula for all queries. The relevance feedback technique can be used to reduce the gap [13] [36] [38] [57].

Relevance feedback, developed from information retrieval [54], is a supervised learning technique used to improve the effectiveness of information retrieval systems. User marks the image of query result as positive or negative examples to improve the system's performance. For a given query, the system first retrieves a list of ranked images according to predefined similarity metrics. Then, the user select a set of positive and negative example form the result images, and the system reformulate the query and retrieves a new list of images. The main problem is how to incorporate positive and negative examples to refine the query and how to adjust the similarity measure according to the feedback.

4.1 Previous Relevance Feedback Works

The original relevance feedback method, in which the vector model [8] [59] [61] is used for document retrieval, can be illustrated by the Rocchio's formula [54] as following equations

$$Q^{'} = \alpha Q + \beta(\frac{1}{N_{R^{'}}}\sum_{i \in D_{R^{'}}} D_i) - \gamma(\frac{1}{N_{N^{'}}}\sum_{i \in D_{N^{'}}} D_i)$$

Where α, β, and γ are suitable constants and $N_{R'}$ and $N_{N'}$ are the number of documents in D'_R and D'_N, respectively. That is, for a given initial query Q, and a set of relevant documents D'_R and nonrelevant documents D'_N given by the user, the optimal new query, $Q^{'}$, is the one that is moved toward positive example points and away from negative example points. This technique is also implemented in many content-based image retrieval systems [30] [36]. Experiments show that the retrieval performance can be improved considerably by using this approach.

The weighting method (e.g., [30] [57] [58]) adjusts larger weight with more important dimensions and smaller weight with less important ones. For example, [58] generalizes a relevance feedback framework of the low-level feature based relevance feedback methods. And ideal query vector for each feature i is described by the weighted sum of all positive feedback images as

$$q_i = \frac{\pi^T Y_i}{\sum_{j=1}^{n} \pi_j},$$

Where Y_i is the $n \times K_i$ (K_i is the length of feature *i*) training sample matrix for feature I obtained by stacking the n positive feedback training vectors X_i^+ into a matrix. The n element vector $\pi = [\pi_1, \pi_2, \ldots \pi_n]$ represents the degree of relevance of each of the n positive feedback images, which can be determined by the user at each feedback interaction. The system then uses q_i as the optimal query to evaluate the relevance of the images in database. This strategy is widely used by many other image retrieval and relevance feedback systems [30][57][58].

Quicklook [11], the innovative part of the system is the relevance feedback method. After the relevant images are selected, each of their features contributes to the new query feature vector if its distance to the average over the relevant images' feature is sufficiently large (three times the standard deviation). The new query feature vector is the average of the contributing features.

ImageRETRO [71] let I_s be the image set after s reductions (filtering) and let F denote the set of 10 color features described. The image set is clustered based on an automatically selected feature subset F_s of F. The images from I_s are ranked independently for each feature in F_s, and each such ranking is divided into 4 clusters (corresponding to a reduction of 25%) and each cluster centroid is chosen as the cluster representative. The union of these representatives for all rankings forms the representative set of I_s, which will be shown to the user for the next reduction. The choice of feature subset F_s at stage s in the retrieval process is based on statistical analysis. For each feature, the variance of the feature values of all images is computed and F_s are made of the features with highest variances that do not highly correlate with each other.

The Webseek [65] let user has the possibility of selecting positive and negative examples from the result of a query in order to reformulate the query. If the set of relevant images is denoted by I_r and the set of nonrelevant images is I_n then the new query histogram at the (k+1)th feedback iteration is computed by

$$h_q^{k+1} = \| \alpha h_q^k + \beta \sum_{i \in I_r} h_i - \gamma \sum_{j \in I_n} h_j \| .$$

Bayesian estimation methods have been used in the probabilistic approaches to relevance feedback. Cox et al. [40], Vasconcelos and Lippman [70], Meilhac and Nastar [38] all used Bayesian learning to incorporate user feedbacks to update the probability distribution of all the images in the database. They consider the feedback examples as a sequence of independent queries and try to minimize the retrieval error by Bayesian rules. That is, given a sequence of queries, they try to minimize the probability of retrieval error as

$$\begin{aligned} g(x) &= \arg\max_i P(y = i \mid x_1, ..., x_t) \\ &= \arg\max_{\mathrm{i}} \{ P(x_t \mid y = i) P(y = i \mid x_1, ..., x_{t-1}) \}. \end{aligned}$$

Where $\{x_1, \ldots, xt\}$ is a sequence of queries (feedback examples) and $\mathrm{P}(y=i|x_1,\ldots,x_t)$ is a prior belief about the ability of the ith image class to explain the queries.

PicHunter [13] implements a probabilistic relevance feedback mechanism, which tries to predict the target image the user wants based on his actions (the images he selects as similar to the target in each iteration of a query session). A vector is used for retaining each image's probability of being the target. The vector is updated during each iteration of the relevance feedback, based on the history of the session (images displayed by the system and user's actions in previous iterations).

The updating formula is based on Bayes' rule. If the n database images are noted T_j, j=1,…,n, and the history of the session through iteration t is denoted $H_t=\{D_1,A_1,D_2,A_2,\ldots\ldots,D_t,A_t\}$, with D_j and A_j being the images displayed by the system and, respectively, the action taken by the user at the iteration j, then the iterative update of the probability estimate of an image T_i being the target, given the history H_t, is:

$$P(T=T_i \mid H_t) = P(T=T_i \mid D_t, A_t, H_{t-1}) = \frac{P(A_t \mid T=T_i, D_t, H_{t-1})P(T=T_i \mid H_{t-1})}{\sum_{j=1}^{n} P(A_t \mid T=T_j, D_t, H_{t-1})P(T=T_j \mid H_{t-1})}$$

in computing the probability of a user to take a certain action A_t given the history so far and the fact that the target is indeed T_i, namely $P(A_t|T=T_i,D_t,H_{t-1})$, a few models were tested. One approach is to estimate the probability of the user to pick an image X_a from $X_1,\ldots,X_{nt}$ by

$$P_{soft\,\min}(A=a \mid X_1,\ldots,X_{n_t},T) = \frac{\exp(-d(X_a,T)/\sigma)}{\sum_{i=1}^{n_t}\exp(-d(X_i,T)/\sigma)}$$

and in the case of choosing any number of images, to assume that each image is selected independently acording to a $p_{softmin}$.

Efforts have also been made to address the problem of slow response time in content-based image retrieval, the problem being caused mainly by the high dimensionality of the feature space, typically hundreds to thousand. Ng and Sedighian [43] made direct use of eigenimages, a method from face recognition [34], to carry out the dimension reduction, Faloutsos and Lin [18], Chandrasekaren et al. [10] and Brunelli and Mich [53] used principal Component analysis (PCA) to perform the dimension reduction in feature spaces. Experimental results in these works show that most real image feature sets can be considerably reduced in dimension without significant degradation in retrieval quality.

Previous researches only allow user marked positive examples and/or negative examples of results. It is successfully used in the document vector based, because they rely on the keyword vectors. In the image features, they contain variety different features, like as color, shape, and textures etc. User may concentrate on the similar color than we may give color more significant weighting. In this paper we proposed a more robust relevance feedback mechanism to adjust the weighting of different features.

4.2 Proposed Relevance Feedback Mechanism

Image retrieval always use low level feature to retrieve similar images. It is more far from human's semantic concept than document retrieval system. On the other hand, the query of image retrieval system usually more confuse than keyword based retrieval systems. Thus, image retrieval system need to design image query interface to allow user to describe an image or query by similar example. While in the process, the system also try to learn user's interesting until find the objective.

Relevance feedback mechanism expects to extract the interesting of user from user interaction. One of the main distinct between image retrieval system and document retrieval system is the result of document need more time to realize than image result. User can fast judge which images of results are more similar to query image in image retrieval systems. User can decide the image is relevant or not in a glance. Thus, image retrieval system is more suitable to interact with user in query process.

Previous researches allow user feedback with positive examples and/or negative examples to reformulate the query example. In this propose, we design a new feedback mechanism to refine the weighting of variety features to follow user's interesting.

We allow user to offer a sequence of relevant image in similar order. Based on the ranking sequence we can estimate each designed features how close to user's conceptual. If one feature is more closes to user's conceptual then the sequence of ranking result must more close to user's query ranking sequence.

Let user input the sequence is p1 is more similar to target image than P2 than we denote p1<p2. If the similar degree of P1 and P2 are the same than we denote p1=p2. (p1<p2<p3<p4<p5<p6) is a ranking sequence from left to right that similar to target.

In the system, each feature will affect the result ranking sequence. We can analysis each feature how close to user's defined sequence to adjust the weight individually. For example if feature F1 output sequence is (p1<p2<p3<p4<p6<p5) and feature F2 output sequence is (p6<p5<p4<p3<p2<p1). We can find that feature F1 is more close to user's conceptual and feature F2 is not. We may low down the weight of feature F2 and rise up the weight of feature F1 will get better result. The user behavior problem will become sequence comparison.

We use the R_{norm} method to evaluate how two sequences are closing. The R_{nor} comparison is defined as following:

Definition: Let I be a finite set of images with a user-defined preference relation ≥ that is complete and transitive (weak order). Let Δ^{user} be the rank ordering of I induced by the user preference relation. Also, let Δ^{system} be some rank ordering of I induced by the similarity values computed by an image retrieval system. Then R_{norm} is defined as

$$R_{norm}(\Delta^{system}, \Delta^{expert}) = \frac{1}{2}\left(1 + \frac{S^{+} - S^{-}}{S^{+}_{\max}}\right),$$

where S^{+} is the number of image pairs where a better image is ranked ahead of a worse on by Δ^{system}; S^- is the number of pairs where a worse image is ranked ahead of a better one by Δ^{system}; S^- is the number of pairs where a worse image is ranked ahead of a better on by Δ^{system}; and $S^+_{\max}$ is the maximum possible number of S^+ from Δ^{user}. It should be noted that the calculation of S^+, S^-, $S^+_{\max}$ is based on the ranking of image pairs in Δ^{system} relative to the ranking of corresponding image pairs in Δ^{user}.

Example: consider the following two rank orderings:

$$\Delta^{user}=(p1,p4<p2,p3<p5) \text{ and } \Delta^{system}=(p5<p2,p4<p1,p3).$$

According to the user, p1 and p4 have the highest preference, followed by p2 and p3 at the next level of preference, followed by p5 at the lowest level of preference. The user considers p1 equivalent to p4 and p2 equivalent to p3. Δ^{system} is interpreted in a similar way. Here we have,

$S^+_{\max}=\{(p1,p2), (p1,p3), (p1,p5), (p4,p2), (p4,p3), (p4,p5), (p2,p5), (p3,p5)\}=8$,

$S^+=\{(p4,p3)\}=1$,

$S^-=\{(p5,p2),(p5,p4),(p5,p1),(p5,p3),(p2,p1)\}=5$.

Therefore,

$R_{norm}=1/2(1+(1-5)/8)=0.25$.

R_{norm} values range from 0 to 1.0 and a value of 1.0 indicates that the system provided rank ordering of the database images close to the rank ordering of the database images provided by the user.

Assume there are n features $(f_1,f_2,\ldots\ldots,f_n)$ for image retrieval system and each weight of features is $(w_1,w_2,\ldots.,w_n)$. After user feedback the ranked result to the system. We can estimate the R_{norm} for each feature $(r_1,r_2,\ldots\ldots r_n)$. Then we define the new weight of each feature as:

$$w_i = \frac{r_i}{\sum_{j=1}^{n} r_j}$$

The system then uses the new weight for each feature to generate query result. This mechanism allows the features present in any types (image features or text vectors) will more flexible and robust than previous researches.

We design a graphic user interface to show how the new feedback model can be integrated into a content-based image retrieval system. Previous relevance feedback mechanisms only offer the user to choose positive or negative examples. Giving too few positive examples distorts the result; on the other hand, giving too many negative examples will confuse the system. The reason is that all positive examples are alike in a way; but each negative example is negative in its own way. Our proposed model allows the user to provide more information in the feedback phase. With the same number of judged examples we can get more information in our graphic user interface. In this manner, the iterations of feedback processes can be reduced.

We define a new mechanism for the user to weight various features based on his interests. According to the result of the system, user can re-rank his preferred priority. It is inconvenient for the user to give each image a value of similarity degree. The user is usually difficult in defining a value about similarity, but the user can distinguish which images are more similar than the others. We develop a friendly user interface for the user to easily express his intention. Figure 4-1 is the graphic user interface of our system. The user can click the resultant images and put it into the ranking box. The priority is reduced gradually by following “>” symbols.

As shown in Figure 4-1, the top-left image is the query image. We can specify the query image from a file or the right window. The user first queries a medical image by example and obtain the list of resultant similar medical images. From the similar resultant images, the user picks up the most similar images into the ranking box. The user can use the “>” and “=” buttons to adjust the priority. The symbol “>” means that the preceding image is more important than the following image. The symbol “=” means that the importance of the preceding image is equal to the following image. This graphic user interface allows the user to easily list preferred ranking result. The system then exploits the list in the rank box to evaluate the weight of features and refine the query by the method proposed in Section 3.

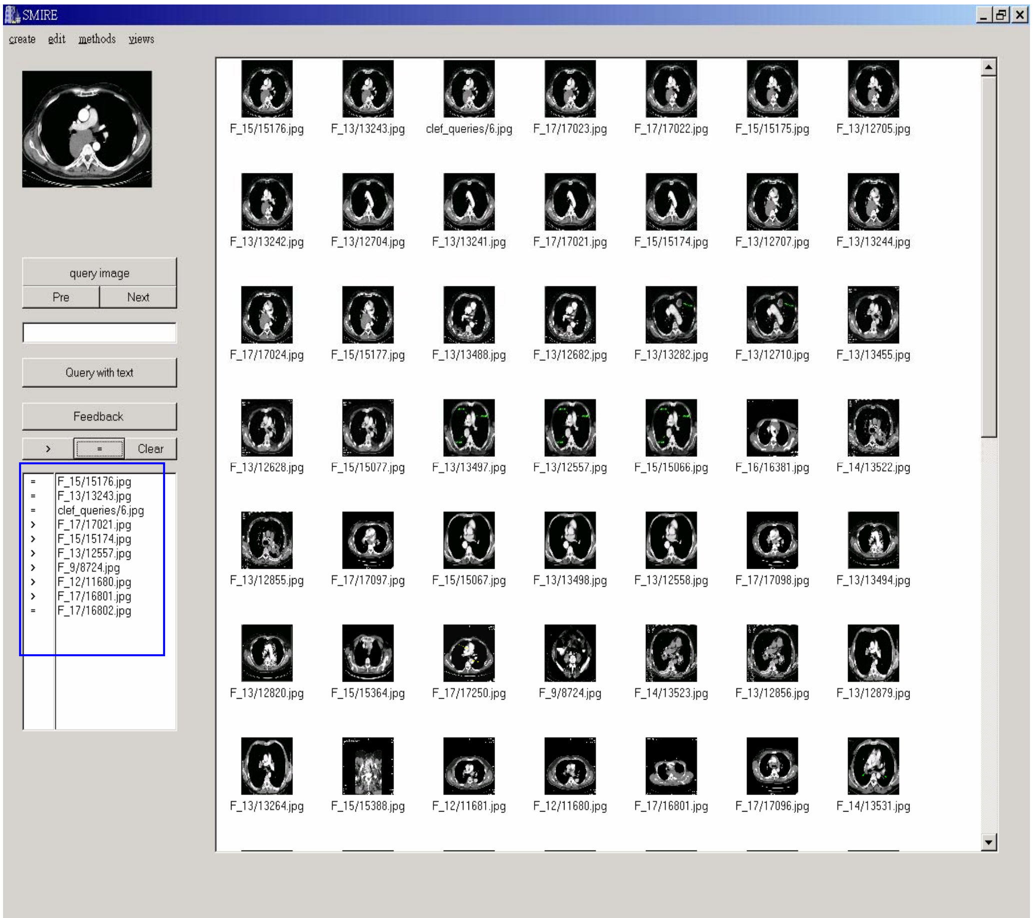


Figure 4-1: Graphic user interface for the proposed CBIR system.

4.3. Experiment Result

Although many content-based image retrieval methods have been proposed, there are few benchmarks for evaluation. In the ImageCLEF 2004 forum, a forum for comparing the performance of content-based image retrieval methods is first proposed. The ImageCLEF 2004 forum contains 9916 medical images for evaluation. In this paper we follow the ImageCLEF 2004 evaluation to evaluate the performance of the feedback mechanism. The process of evaluation and the format of results employ the trec_eval tool. There are 26 queries for test. The corresponding answer images of each query were judged as either relevant or partially relevant by at least 2 assessors.

We conduct three experiments. Color Histogram, Gray Level Histogram, Semantic Moment, and Shape Correlogram are the four features for retrieving similar medical images. To show that the proposed relevance feedback mechanism is very flexible, the types of image features we use are quite different. The first experiment, called BASE, uses the visual features of the query example to query the database without relevance feedback. The second experiment, called RUI, is the weighting method that associates larger weights with more important dimensions and smaller weights with less important ones. This method generalizes a relevance feedback framework of the physical features based on positive feedback examples. We normalize different concept features by Gaussian normalization, and the weights of concept features are equal. The weight of individual feature vectors applies the feedback method proposed in. The experiment, denoted as ARF (Adaptive Relevance Feedback), is the result that uses the proposed feedback mechanism. The system integrates the four features by Gaussian normalization in the first run. While the second run, we adjust the weight of concept features by the R_{norm} method. The weights of color histogram and gray-spatial histogram are adjusted. The weight of Coherence moment and Gray correlogram are tuned by the R_{norm} method. The test result shows that the feedback mechanisms (RUI and ARF) have better result than the mechanism without relevance feedback (BASE). While the user feedbacks its interests to the system, the proposed method (ARF) is more precise and quicker to reach user's requirement. Figure 4-2 shows the precision and recall graphs. RUI and ARF curves are the result after conducting relevance feedback three times

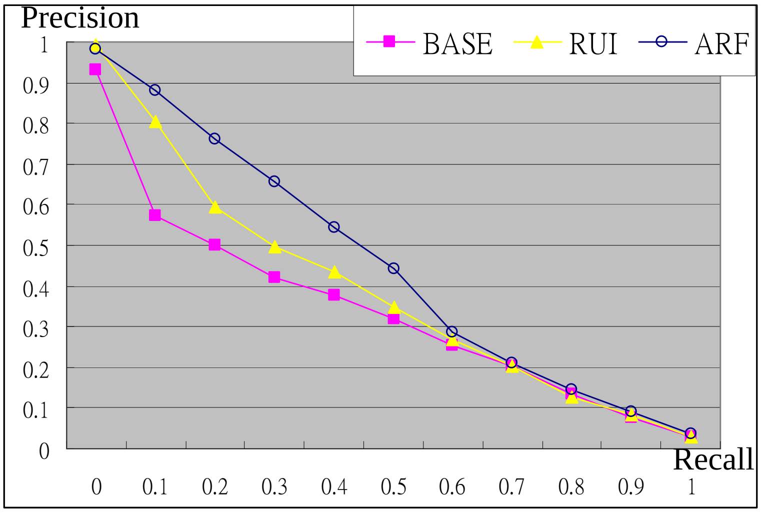


Figure 4-2: The precision vs. recall graphs of average 26 queries

The mean average precision of BASE is 0.3273. The mean average precision of RUI is 0.3884. The mean average precision of ARF is 0.447. Table 4-1 is the mean average precision and relevance feedback iterations. As shown in Table 4-1, the ARF method reaches the user's interests faster than the RUI method.

The experimental result shows that the proposed feedback method can be used for integrating arbitrary concept features. We can estimate which features are more important although the scales of features are different.

Iterations / methods	0	1	2	3
RUI	0.327	0.367	0.374	0.388
ARF	0.327	0.401	0.432	0.447

Table 4-1: The mean average precision at n-iteration relevance feedback.

In this paper we develop a new relevance feedback mechanism to improve content-based image retrieval. The two-level feature modulation mechanism according to user's interests enhances the result significantly. Uniform and equal calibration of features is easy to adjust the feature's weight, but some features are not so trivial. The proposed method can treat various types of features in the concept level and is more robust than previous works.

It is easy to integrate our feedback mechanism into existent content-based image retrieval methods. Furthermore, the feedback mechanism can be applied to CBIR applications other than medical images. In the future, we will use the feedback mechanism to combine visual feature and textual features.

Chapter 5

Conclusions and Future Research

This dissertation proposes the investigation of design and implementation of a cross language medical image retrieval system. Three new approaches are considered to provide an environment for users to retrieval medical diagnosis cases. They are medical image content retrieval, user oriented relevance feedback mechanism, and ontology chain for disambiguate cross language retrieval.

Several features are designed for similar medical image searching. Proposed method considers for humans perceptual, concentrating on the contrast representation of images are more suitable for user. Our representation is immune in defective illumination. It is very efficient in searching computation times. The image representation we proposed has obtained good results in ImageCLEF 2004.

We also proposed a new feedback mechanism help user to find interesting aspect, thus will improve the query result. The relevance feedback model can be applied to diversity similar methods. The mechanism is more flexible and robust than previous relevance feedback methods.

At the least, we integrate cross-language approach and medical image retrieval method to allow user more convenient and effective to searching data. The ontology chain will reduce the ambiguity of translated words. Based on the similar image, a multilingual document crosses images retrieval become possible without large parallel corpus.

This thesis makes the following contributions:

- A new integrated multilingual medical image retrieval system supports the capability of medical image content retrieval and cross language retrieval.
- A new representation for medical image indexing to improve the query result.
- A new feedback mechanism for user to easy regulate diverse methods for user oriented searching.
- Cross text and image query, the document always contains text and images in medical image database. We proposed a system allow user to use text or image query. Cross the text and image searching will get more related documents.

An image represents thousands of words. An image can be viewed from various aspects; furthermore, different people may have different interpretations of the same image. This means that many parameters need to be tuned. In the future, we will try to learn the user behavior and tune those parameters by learning methods.

Bibliography

[1] A. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C.-R. Shyu, and A. Marchiori, “Automated storage and retrieval of thin–section CT images to assist diagnosis: System description and preliminary assessment,” *Radiology*, vol. 228, pp. 265–270, 2003.

[2] S. Ardizzoni, I. Bartolini, and M. Patella, “Windsurf: Region-based image Retrieva1 Using Wave1ets,” *Proc. Int. Workshop on Similarity Search* (IW0SS), pp.167-173, 1999.

[3] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jain, and C.-F. Shu, “The Virage image search engine: An open framework for image management,” in: I. K. Sethi, R. C. Jain (Eds.), *Storage & Retrieval for Image and Video Databases IV*, Vol. 2670 of IS&T/SPIE Proceedings, San Jose, CA, USA, pp. 76-87, 1996.

[4] L. Ballesteros, and W.B. Croft, “Dictionary-based methods for cross-lingual information retrieval,” *Proc. of the 7th Int. DEXA Conference on Database and Expert Systems Applications*, 1996.

[5] S. Belongie, C. Carson, H. Greenspan, and J. Malik, “Color and texture based image segmentation using EM and its application to content-based image retrieval,” in: *Proc. of the International Conference on Computer Vision (ICCV'98)*, Bombay, India, pp. 675-682, 1998.

[6] S. Beretti, A. Del Bimbo, and P. Pala, “Content-based retrieval of 3D cellular structures,” in: *Proceedings of the second International Conference on Multimedia and Exposition* (ICME'2001), IEEE Computer Society, Tokyo, Japan, pp. 1096-1099, 2001.

[7] W. D. Bidgood, B. Bray, N. Brown, A. R. Mori, K. A. Spackman, A. Golichowsky, R. H. Jones, L. Korman, B. Dove, L. Hildebrand, and M. Berg, “Image acquisition context: Procedure description attributes for clinically relevant indexing and selective retrieval of biomedical images,” *Journal of the American Medical Informatics Association*, vol.6(1), pp.61-75, 1999.

[8] C. Buckley and G. Salton, “Optimization of relevance feedback weights,” in *Proc. SIGIR’95*, 1995.

[9] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik, “Blobworld: A system for region-based image indexing and retrieval,” in: D. P. Huijsmans, A. W. M. Smeulders (Eds.), *Third International Conference On Visual Information Systems* (VISUAL' 99), no. 1614 in Lecture Notes in Computer Science, Springer Verlag, Amsterdam, the Netherlands, pp.509-516, 1999.

[10] S. Chandrasekaran, “An eigenspace update algorithm for image analysis,” *CVGIP: Graph. Models Image Process*, 1997.

[11] G. Ciocca, I. Gagliardi, and R. Schettini, “Quicklook2: An integrated multimedia system,” *International Journal of Visual Languages and Computing,* Vol. 12 (1), pp. 81-103, 2001.

[12] P. Clough, M. Sanderson, and H. Muller, “The CLEF Cross Language Image Retrieval Track (ImageCLEF) 2004,” *the 8th in the series of European Digigtal Library Conferences*, ECDL 2004, September, Bath, UK, 2004.

[13] I. J. Cox, T. P. Minka, T. V. Papathomas, and P. N. Yianilos, “The Bayesian image retrieval system, pichunter: Theory, implementation, and psychophysical experiments,” *IEEE Trans. Image Processing-Special Issue on Digital Libraries*, 2000.

[14] M. W. Davis and Ted E. Dunning, “A TREC evaluation of query translation methods for multi-lingual text retrieval,” *In D. K. Harman, TREC-4, NIST*, Nov. 1995.

[15] M. W. Davis, and W. C. Ogden, “Implementing cross-language text retrieval system for large-scale text collection on the World Wide Web,” *In AAAI Symposium on Cross-language Text and Speech Retrieval*, American for Artificial Intelligence, 1997.

[16] S. Deerwester, S. T. Dumais, and R. Harshman, “Indexing by Latent Semantic analysis,” *Journal of the American Society for Information Science*, vol. 41 (6), Sept. 1990.

[17] J. A. Edwards, “Survey of Electronic Corpora and Related Resources for Language Researchers,” *In Edwards & Lampert (eds) Talking data: Transcription and Coding in Discourse Research*, NJ, Erlbaum, London and Hillsdale, 1993.

[18] C. Faloutsos and K. Lin, “Fastmap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia,” in *Proc.SIGMOD*, pp. 163-174, 1995.

[19] C. Fellbaum, WordNet (eds.): An Electronic Lexical Database. MIT Press, 1998.

[20] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, “Query by Image and Video Content: The QBIC system,” *IEEE Computer*, vol. 28 (9), pp. 23-32, 1995.

[21] J. Gilarranz, J. Gonzalo, and F. Verdejo. “An Approach to Conceptual Text Retrieval Using the EuroWordNet Multilingual Semantic Database,” *in Working Notes of AAAI Spring Symposium on Cross-Language Text and*

Speech Retrieval, Stanford, CA, 1997

[22] R. A. Greenes, and J. F. Brinkley, "Imaging systems," in: Medical Informatics: *Computer Applications in Healthcare (2nd edition), Springer,* New York, pp. 485- 538, 2000.

[23] J.C. Goswami and A. K. Chan, "Fundamentals of Wave1ets, Theory, Algorithms, and Applications," A Wiley Inter-science Publication, 1999.

[24] H. Haddouti, "Multilinguality Issues in Digital Libraries," *Proceedings of the EuroMed Net'98 Conference Nicosia*, March 3-7, 1998

[25] A. Hampapur, A. Gupta, B. Horowitz, C.F. Shu, C. Fuller, J. Bach, M. Gorkani, R. Jain, "Virage video engine," in: I. K. Sethi, R. C. Jain (Eds.), *Storage and Retrieval for Image and Video Databases V*, Vol. 3022 of SPIE Proceedings, pp. 352-360, 1997.

[26] J. Han, and M. Kamber, "Data mining: concepts and techniques," *Academic press*, San Diego, CA, USA, 2001.

[27] Y. Hayashi, « TITAN: A Cross-linguistic Search Engine for the WWW," *in Working Notes of AAAI Spring Symposium on Cross-Language Text and Speech Retrieval*, Stanford, CA, 1997.

[28] J. Huang, S. R. Kumar, M. Mitra, W. J. Zhu, and R. Zabih, "Image indexing Using Color Correlograms," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico 1997.

[29] D. A. Hull, and G. Grefenstette, "Querying across languages. A dictionary-based approach to multilingual information retrieval," *In Proceedings of the 19th ACM SIGIR Conference*, 1996

[30] Y. Ishikawa, R. Subramanya, and C. Faloutsos, "Mindreader: Query databases through multiple examples," in *Proc. 24th VLDB Conf.*, New York, 1998.

[31] O. H. Karam, A. H. Hamad, S. Ghoneimy, G. Abdel Rahman, and S. Rady, "Wavelet-Based Retrieval for Medical Image Databases," *Proceedings of the international Conference on intelligent Computing and Information systems* (ICICIS), pp. 309-314, 2002.

[32] M. Cannon, and D. R. Hush, "Query by image example: the CANDID approach," in: W. Niblack, R. C. Jain (Eds.), *Storage and Retrieval for Image and Video Databases III*, Vol. 2420 of SPIE Proceedings, pp. 238-248, 1995.

[33] D. Keysers, J. Dahmen, H. Ney, B. B. Wein, and T. M. Lehmann, "A statistical framework for model–based image retrieval in medical applications," *Journal of Electronic Imaging,* vol. 12 (1), pp.59–68, 2003.

[34] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 103-108, Jan. 1990.

[35] T. M. Lehmann, M. O. Guld, C. Thies, B. Fischer, M. Keysers, D. Kohnen, H. Schubert, and B. B. Wein, "Content–based image retrieval in medical applications for picture archiving and communication systems," in: *Medical Imaging of SPIE Proceedings*, Vol. 5033, San Diego, California, USA, 2003.
[36] Y. Lu, C. Hu, X. Zhu, H. Zhang, and Q. Yang, "A unified framework for semantics and feature based relevance feedback in image retrieval systems," in *Proc. 8th ACM Multimedia Int. Conf.*, Los Angeles, CA, Nov. 2000.
[37] W. Y. Ma, Y. Deng, and B. S. Manjunath, "Tools for texture- and color-based search of images," in: B. E. Rogowitz, T. N. Pappas (Eds.), *Human Vision and Electronic Imaging II of SPIE Proceedings*, Vol. 3016, San Jose, CA, pp. 496-507, 1997.
[38] C. Meilhac and C. Nastar, "Relevance feedback and category search in image databases," in *IEEE Int. Conf. Multimedia Computing and Systems*, 1999.
[39] G. Miller et al. Five Papers on WordNet. CSL Report 43. Cognitive Science Laboratory, Princeton University, 1990 (http://www.cogsci.princeton.edu/~wn/)
[40] I. J. Cox, T. P. Minka, T. V. Papathomas, and P. N. Yianilos, "The Bayesian image retrieval system, pichunter: Theory, implementation, and psychophysical experiments," *IEEE Trans. Image Processing-Special Issue on Digital Libraries*, 2000.
[41] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medicine clinical benefits and future directions," *International Journal of Medical Informatics*, vol. 73, pp. 1-23, 2004.
[42] A. Natscv, R. Rastogi and K. shim, "WALRUS: A simiiarity Retrieval Algorithm for image Databases," *SIGMOD Conference*, 1999.
[43] R. Ng and A. Sedighian, "Evaluating multi-dimensional indexing structures for images transformed by principal component analysis," in *Proc.SPIE Storage and Retrieval for Image and Video Databases*, 1996.
[44] NUA, Internet Consultancy and Developer, October 1998 (http://www.nua.net).
[45] D. W. Oard, "Adaptive Vector Space Text Filtering for Monolingual and Cross-language Applications," Ph.D. Thesis, University of Maryland, College Park, 1996.
[46] M. R. Ogiela, and R. Tadeusiewicz, "Semantic oriented syntactic algorithms for content recognition and understanding of images in medical databases," in: *Proceedings of the second International Conference on Multimedia and Exposition* (ICME'2001), IEEE Computer Society, Tokyo, Japan, pp. 621–624, 2001.

[47] T. Ojala, M. Rautiainen, E. Matinmikko and M. Aittola, “Semantic Image Retrieval with HSV Correlograms,” *Proceedings of 12th Scandinavian Conference on Image Analysis*, Bergen, Norway 2001.

[48] S. C. Orphanoudakis, C. E. Chronaki, and S. Kostomanolakis, “I2Cnet: A system for the indexing,” *Storage and retrieval of medical images by content, Medical Informatics*, vol. 19(2), pp. 109-122, 1994.

[49] A. H. Paquet, S. Zahir and R. K. Ward, “Wavelets Packets-based Image Retrieva1,” *Proceedings of the IEEE Internationa1 Conference on Acoustics Speech and Signa1 Processing* (ICASSP), 2002.

[50] A. Pentland, R. W. Picard, and S. Sclaro, “Photobook: Tools for content-based manipulation of image databases,” *International Journal of Computer Vision*, vol.18 (3), pp. 233-254, 1996.

[51] T. Pun, G. Gerig, O. Ratib, “Image analysis and computer vision in medicine,” *Computerized Medical Imaging and Graphics*, vol. 18(2), pp. 85-96, 1994.

[52] B. Revet, “DICOM Cook Book for Implementations in Madalities,” *Philips Medical Systems*, Eindhoven, Netherlands, 1997.

[53] B. Roberto and M. Ornella, “Image retrieval by examples,” *IEEE Trans.Multimedia*, vol. 2, Sept. 2000.

[54] J. J. Rocchio, “Relevance feedback in information retrieval,” in *The SMART Retrieval System: Experiments in Automatic Document Processing*, G. Salton, Ed. Englewood Cliffs, NJ: Prentice-Hall, pp.313-323, 1971.

[55] A. Rosset, H. Muller, M. Martins, N. Dfouni, J. P. Vallee, and O. Ratib, “Casimage Project a digital teaching files authoring environment,” *Journal of Thoracic Imaging*, vol. 19(2), pp.1-6, 2004.

[56] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, “Relevance feedback: A power tool for interactive content based image retrieval,” *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 644-655, 1998.

[57] Y. Rui and T. S. Huang, “Relevance feedback: A power tool for interactive content-based image retrieval,” *IEEE Circuits Systems for Video Technology*, vol. 8, no. 5, 1999.

[58] Y. Rui and T. S. Huang, “A novel relevance feedback technique in image retrieval,” *ACM Multimedia*, 1999.

[59] G. Salton and M. J. McGill, “Introduction to Modern Information Retrieval,” New York: McGraw-Hill, 1983.

[60] G. Salton, “Automatic Text Processing,” Addison-Wesley Publishing Company, 1988.

[61] W. M. Shaw, “Term-relevance computation and perfect retrieval performance,” *Information Processing Manage*, vol. 31, pp. 491–498, 1995.

[62] P. Sheridan, and J. P. Ballerini, "Experiments in Multilingual Retrieval Using the Spider System," *In Proceeding of the 19th Annual International ACM SIGIR*, 1996

[63] C. R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick, "ASSERT: A physician in the loop content based retrieval system for HRCT image databases," *Computer Vision and Image Understanding (special issue on content–based access for image and video libraries)*, vol. 75, pp.111–132, 1999.

[64] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, No. 12, pp. 1349-1380, 2000.

[65] J. R. Smith, "Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis," PhD thesis, Graduate School of Arts and Sciences, Columbia University, February 1997.

[66] E. J. Stollnitz, T. D. DeRose, and D. H. Salesin, "Wavelets for computer Graphics–Theory and Applications," San Francisco, CA: Morgan Kaufmann Publilshers, Inc, 1996.

[67] D. M. Squire, W. Muller, H. Muller, and T. Pun, "Content-based query of image databases," in aspirations from text retrieval, *Pattern Recognition Letters (Selected Papers from The 11th Scandinavian Conference on Image Analysis SCIA '99)*, vol. 21, pp. 1193-1198, 2000.

[68] M. J. Swain and D. H. Ballard, "Color Indexing," *International Journal of Computer Vision,* Vol. 7, pp.11-32, 1991.

[69] H. D. Tagare, C. Jaffe, and J. Duncan, "Medical image databases: A content–based retrieval approach," *Journal of the American Medical Informatics Association*, vol. 4 (3), pp. 184-198, 1997.

[70] N. Vasconcelos and A. Lippman, "Learning from user feedback in image retrieval systems," in *Proc. NIPS'99*, Denver, CO, 1999.

[71] J. Vendrig, M. Worring, and A. W. M. Smeulders, "Filter image browsing: Exploiting interaction in image retrieval," *Visual Information and Information Systems, Proceedings of the Third International Conference VISUAL '99, Amsterdam, The Netherlands,* Lecture Notes in Computer Science 1614. Springer, pp. 147-154, 1999.

[72] M. Wechsler, and P. Schäuble, "Multilingual Information Retrieval Based on Document Alignment Techniques," *In Lecture Notes in Computer Science. Ed. Ch. Nikolau, C. Stephanidis. Second European Conference on Research and Advanced Technology for Digital Libraries ECDL´98*, Crete, 1998.

Appendix

Topic: 1
Total number of relevant docs = 235
Total number of retrieved relevant docs = 171
Average (non-interpolated) precision = 0.551375630938718
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 1
prec at 0.2 = 0.972972972972973
prec at 0.3 = 0.972972972972973
prec at 0.4 = 0.803418803418803
prec at 0.5 = 0.6
prec at 0.6 = 0.456310679611651
prec at 0.7 = 0.193960511033682
prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 1
prec at 10 = 1
prec at 15 = 1
prec at 20 = 1
prec at 30 = 1
prec at 100 = 0.84
prec at 200 = 0.6
prec at 500 = 0.312
prec at 1000 = 0.171
Breakeven Precision: 0.540425531914894

Topic: 2
Total number of relevant docs = 320
Total number of retrieved relevant docs = 202
Average (non-interpolated) precision = 0.463059255430413
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 1
prec at 0.2 = 0.970588235294118
prec at 0.3 = 0.825396825396825
prec at 0.4 = 0.731428571428571
prec at 0.5 = 0.450980392156863
prec at 0.6 = 0.25668449197861
prec at 0.7 = 0
prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:

prec at 5 = 1
prec at 10 = 1
prec at 15 = 1
prec at 20 = 1
prec at 30 = 1
prec at 100 = 0.84
prec at 200 = 0.65
prec at 500 = 0.356
prec at 1000 = 0.202
Breakeven Precision: 0.49375

Topic: 3
Total number of relevant docs = 72
Total number of retrieved relevant docs = 46
Average (non-interpolated) precision = 0.13061167135495
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 0.225
prec at 0.2 = 0.192771084337349
prec at 0.3 = 0.120192307692308
prec at 0.4 = 0.119834710743802
prec at 0.5 = 0.080338266384778
prec at 0.6 = 0.0581241743725231
prec at 0.7 = 0
prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 0.6
prec at 10 = 0.4
prec at 15 = 0.4
prec at 20 = 0.35
prec at 30 = 0.233333333333333
prec at 100 = 0.17
prec at 200 = 0.115
prec at 500 = 0.076
prec at 1000 = 0.046
Breakeven Precision: 0.180555555555556

Topic: 4
Total number of relevant docs = 43
Total number of retrieved relevant docs = 27
Average (non-interpolated) precision = 0.154008972803451
Interpolated precsion at recalls:
prec at 0 = 0.555555555555556
prec at 0.1 = 0.555555555555556
prec at 0.2 = 0.555555555555556
prec at 0.3 = 0.220588235294118
prec at 0.4 = 0.114649681528662
prec at 0.5 = 0.0751633986928105

prec at 0.6 = 0.027807486631016
prec at 0.7 = 0
prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 0.4
prec at 10 = 0.4
prec at 15 = 0.466666666666667
prec at 20 = 0.5
prec at 30 = 0.333333333333333
prec at 100 = 0.15
prec at 200 = 0.1
prec at 500 = 0.048
prec at 1000 = 0.027
Breakeven Precision: 0.255813953488372

Topic: 5
Total number of relevant docs = 84
Total number of retrieved relevant docs = 67
Average (non-interpolated) precision = 0.477904545235569
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 1
prec at 0.2 = 0.869565217391304
prec at 0.3 = 0.771428571428571
prec at 0.4 = 0.692307692307692
prec at 0.5 = 0.448979591836735
prec at 0.6 = 0.264248704663212
prec at 0.7 = 0.214532871972318
prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 1
prec at 10 = 1
prec at 15 = 0.933333333333333
prec at 20 = 0.85
prec at 30 = 0.766666666666667
prec at 100 = 0.44
prec at 200 = 0.26
prec at 500 = 0.134
prec at 1000 = 0.067
Breakeven Precision: 0.488095238095238

Topic: 6
Total number of relevant docs = 252
Total number of retrieved relevant docs = 212
Average (non-interpolated) precision = 0.579034878225824
Interpolated precsion at recalls:

prec at 0 = 1
prec at 0.1 = 1
prec at 0.2 = 0.964285714285714
prec at 0.3 = 0.894117647058824
prec at 0.4 = 0.751824817518248
prec at 0.5 = 0.661458333333333
prec at 0.6 = 0.484177215189873
prec at 0.7 = 0.374207188160677
prec at 0.8 = 0.226666666666667
prec at 0.9 = 0
prec at 1 = 0

Non-interpolated precsion at docs:

prec at 5 = 1
prec at 10 = 1
prec at 15 = 1
prec at 20 = 1
prec at 30 = 1
prec at 100 = 0.85
prec at 200 = 0.64
prec at 500 = 0.354
prec at 1000 = 0.0212

Breakeven Precision: 0.547619047619048

Topic: 7
Total number of relevant docs = 48
Total number of retrieved relevant docs = 48
Average (non-interpolated) precision = 0.874008594922796
Interpolated precsion at recalls:

prec at 0 = 1
prec at 0.1 = 1
prec at 0.2 = 1
prec at 0.3 = 0.96
prec at 0.4 = 0.96
prec at 0.5 = 0.96
prec at 0.6 = 0.852941176470588
prec at 0.7 = 0.822222222222222
prec at 0.8 = 0.816326530612245
prec at 0.9 = 0.8
prec at 1 = 0.440366972477064

Non-interpolated precsion at docs:

prec at 5 = 1
prec at 10 = 1
prec at 15 = 0.933333333333333
prec at 20 = 0.95
prec at 30 = 0.833333333333333
prec at 100 = 0.47
prec at 200 = 0.24
prec at 500 = 0.096
prec at 1000 = 0.048

Breakeven Precision: 0.8125

Topic: 8
Total number of relevant docs = 117
Total number of retrieved relevant docs = 75
Average (non-interpolated) precision = 0.422060236256825
Interpolated precsion at recalls:
  prec at 0 = 1
  prec at 0.1 = 1
  prec at 0.2 = 0.961538461538462
  prec at 0.3 = 0.8
  prec at 0.4 = 0.7
  prec at 0.5 = 0.215328467153285
  prec at 0.6 = 0.109567901234568
  prec at 0.7 = 0
  prec at 0.8 = 0
  prec at 0.9 = 0
  prec at 1 = 0
Non-interpolated precsion at docs:
  prec at 5 = 1
  prec at 10 = 1
  prec at 15 = 1
  prec at 20 = 0.95
  prec at 30 = 0.933333333333333
  prec at 100 = 0.5
  prec at 200 = 0.28
  prec at 500 = 0.14
  prec at 1000 = 0.075
Breakeven Precision: 0.427350427350427

Topic: 9
Total number of relevant docs = 43
Total number of retrieved relevant docs = 32
Average (non-interpolated) precision = 0.297845201627166
Interpolated precsion at recalls:
  prec at 0 = 1
  prec at 0.1 = 0.777777777777778
  prec at 0.2 = 0.6
  prec at 0.3 = 0.590909090909091
  prec at 0.4 = 0.292307692307692
  prec at 0.5 = 0.239130434782609
  prec at 0.6 = 0.12621359223301
  prec at 0.7 = 0.0355097365406644
  prec at 0.8 = 0
  prec at 0.9 = 0
  prec at 1 = 0
Non-interpolated precsion at docs:
  prec at 5 = 0.6
  prec at 10 = 0.7
  prec at 15 = 0.6
  prec at 20 = 0.55

prec at 30 = 0.466666666666667
prec at 100 = 0.22
prec at 200 = 0.125
prec at 500 = 0.056
prec at 1000 = 0.032
Breakeven Precision: 0.325581395348837

Topic: 10
Total number of relevant docs = 79
Total number of retrieved relevant docs = 62
Average (non-interpolated) precision = 0.427997764235216
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 1
prec at 0.2 = 0.707317073170732
prec at 0.3 = 0.707317073170732
prec at 0.4 = 0.627118644067797
prec at 0.5 = 0.493975903614458
prec at 0.6 = 0.261538461538462
prec at 0.7 = 0.100538599640934
prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 1
prec at 10 = 0.8
prec at 15 = 0.8
prec at 20 = 0.6
prec at 30 = 0.633333333333333
prec at 100 = 0.43
prec at 200 = 0.255
prec at 500 = 0.11
prec at 1000 = 0.062
Breakeven Precision: 0.468354430379747

Topic: 11
Total number of relevant docs = 9
Total number of retrieved relevant docs = 8
Average (non-interpolated) precision = 0.559645427727475
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 1
prec at 0.2 = 1
prec at 0.3 = 1
prec at 0.4 = 1
prec at 0.5 = 1
prec at 0.6 = 0.0126582278481013
prec at 0.7 = 0.0126582278481013
prec at 0.8 = 0.0126582278481013
prec at 0.9 = 0

prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 1
prec at 10 = 0.5
prec at 15 = 0.333333333333333
prec at 20 = 0.25
prec at 30 = 0.166666666666667
prec at 100 = 0.05
prec at 200 = 0.025
prec at 500 = 0.01
prec at 1000 = 0.008
Breakeven Precision: 0.555555555555556

Topic: 12
Total number of relevant docs = 179
Total number of retrieved relevant docs = 149
Average (non-interpolated) precision = 0.408620265605011
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 0.95
prec at 0.2 = 0.692307692307692
prec at 0.3 = 0.551020408163265
prec at 0.4 = 0.436363636363636
prec at 0.5 = 0.365853658536585
prec at 0.6 = 0.281984334203655
prec at 0.7 = 0.235514018691589
prec at 0.8 = 0.181704260651629
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 1
prec at 10 = 1
prec at 15 = 1
prec at 20 = 0.95
prec at 30 = 0.733333333333333
prec at 100 = 0.54
prec at 200 = 0.415
prec at 500 = 0.246
prec at 1000 = 0.149
Breakeven Precision: 0.418994413407821

Topic: 13
Total number of relevant docs = 95
Total number of retrieved relevant docs = 45
Average (non-interpolated) precision = 0.23803780818125
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 0.916666666666667
prec at 0.2 = 0.648648648648649
prec at 0.3 = 0.211678832116788

prec at 0.4 = 0.0861678004535147
prec at 0.5 = 0
prec at 0.6 = 0
prec at 0.7 = 0
prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 0.8
prec at 10 = 0.9
prec at 15 = 0.866666666666667
prec at 20 = 0.7
prec at 30 = 0.6
prec at 100 = 0.27
prec at 200 = 0.16
prec at 500 = 0.08
prec at 1000 = 0.045
Breakeven Precision: 0.284210526315789

Topic: 14
Total number of relevant docs = 11
Total number of retrieved relevant docs = 6
Average (non-interpolated) precision = 0.421717171717172
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 1
prec at 0.2 = 1
prec at 0.3 = 1
prec at 0.4 = 0.416666666666667
prec at 0.5 = 0.222222222222222
prec at 0.6 = 0
prec at 0.7 = 0
prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 0.8
prec at 10 = 0.4
prec at 15 = 0.333333333333333
prec at 20 = 0.25
prec at 30 = 0.2
prec at 100 = 0.06
prec at 200 = 0.03
prec at 500 = 0.012
prec at 1000 = 0.006
Breakeven Precision: 0.363636363636364

Topic: 15
Total number of relevant docs = 252
Total number of retrieved relevant docs = 228

Average (non-interpolated) precision = 0.744893033620958
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 1
prec at 0.2 = 1
prec at 0.3 = 1
prec at 0.4 = 0.962616822429907
prec at 0.5 = 0.921428571428571
prec at 0.6 = 0.777777777777778
prec at 0.7 = 0.639285714285714
prec at 0.8 = 0.426778242677824
prec at 0.9 = 0.258503401360544
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 1
prec at 10 = 1
prec at 15 = 1
prec at 20 = 1
prec at 30 = 1
prec at 100 = 0.99
prec at 200 = 0.77
prec at 500 = 0.41
prec at 1000 = 0.228
Breakeven Precision: 0.698412698412698

Topic: 16
Total number of relevant docs = 141
Total number of retrieved relevant docs = 118
Average (non-interpolated) precision = 0.56028344470182
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 0.962962962962963
prec at 0.2 = 0.957446808510638
prec at 0.3 = 0.957446808510638
prec at 0.4 = 0.77027027027027
prec at 0.5 = 0.660714285714286
prec at 0.6 = 0.382882882882883
prec at 0.7 = 0.267567567567568
prec at 0.8 = 0.167159763313609
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 1
prec at 10 = 0.9
prec at 15 = 0.933333333333333
prec at 20 = 0.95
prec at 30 = 0.933333333333333
prec at 100 = 0.65
prec at 200 = 0.415
prec at 500 = 0.21

prec at 1000 = 0.118
Breakeven Precision: 0.553191489361702

Topic: 17
Total number of relevant docs = 31
Total number of retrieved relevant docs = 20
Average (non-interpolated) precision = 0.273684242448477
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 1
prec at 0.2 = 0.538461538461538
prec at 0.3 = 0.478260869565217
prec at 0.4 = 0.26
prec at 0.5 = 0.0780487804878049
prec at 0.6 = 0.0261707988980716
prec at 0.7 = 0
prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 0.8
prec at 10 = 0.4
prec at 15 = 0.466666666666667
prec at 20 = 0.45
prec at 30 = 0.366666666666667
prec at 100 = 0.13
prec at 200 = 0.075
prec at 500 = 0.034
prec at 1000 = 0.020
Breakeven Precision: 0.354838709677419

Topic: 18
Total number of relevant docs = 78
Total number of retrieved relevant docs = 62
Average (non-interpolated) precision = 0.264569720858244
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 0.666666666666667
prec at 0.2 = 0.432432432432432
prec at 0.3 = 0.301204819277108
prec at 0.4 = 0.293103448275862
prec at 0.5 = 0.258064516129032
prec at 0.6 = 0.225961538461538
prec at 0.7 = 0.149863760217984
prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 0.6
prec at 10 = 0.6

prec at 15 = 0.533333333333333
prec at 20 = 0.45
prec at 30 = 0.366666666666667
prec at 100 = 0.28
prec at 200 = 0.225
prec at 500 = 0.116
prec at 1000 = 0.062
Breakeven Precision: 0.294871794871795

Topic: 19
Total number of relevant docs = 114
Total number of retrieved relevant docs = 114
Average (non-interpolated) precision = 0.824666946796749
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 1
prec at 0.2 = 1
prec at 0.3 = 0.978260869565217
prec at 0.4 = 0.95
prec at 0.5 = 0.95
prec at 0.6 = 0.8625
prec at 0.7 = 0.661157024793388
prec at 0.8 = 0.603896103896104
prec at 0.9 = 0.527638190954774
prec at 1 = 0.339285714285714
Non-interpolated precsion at docs:
prec at 5 = 1
prec at 10 = 1
prec at 15 = 1
prec at 20 = 1
prec at 30 = 0.966666666666667
prec at 100 = 0.77
prec at 200 = 0.525
prec at 500 = 0.228
prec at 1000 = 0.114
Breakeven Precision: 0.684210526315789

Topic: 20
Total number of relevant docs = 27
Total number of retrieved relevant docs = 14
Average (non-interpolated) precision = 0.188472478585917
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 1
prec at 0.2 = 0.176470588235294
prec at 0.3 = 0.025
prec at 0.4 = 0.0192616372391653
prec at 0.5 = 0.0159453302961276
prec at 0.6 = 0
prec at 0.7 = 0

prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 0.8
prec at 10 = 0.5
prec at 15 = 0.333333333333333
prec at 20 = 0.25
prec at 30 = 0.166666666666667
prec at 100 = 0.06
prec at 200 = 0.035
prec at 500 = 0.018
prec at 1000 = 0.014
Breakeven Precision: 0.185185185185185

Topic: 21
Total number of relevant docs = 90
Total number of retrieved relevant docs = 52
Average (non-interpolated) precision = 0.137801726998012
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 0.391304347826087
prec at 0.2 = 0.333333333333333
prec at 0.3 = 0.149171270718232
prec at 0.4 = 0.101694915254237
prec at 0.5 = 0.0583016476552598
prec at 0.6 = 0
prec at 0.7 = 0
prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 0.4
prec at 10 = 0.5
prec at 15 = 0.4
prec at 20 = 0.4
prec at 30 = 0.366666666666667
prec at 100 = 0.21
prec at 200 = 0.14
prec at 500 = 0.082
prec at 1000 = 0.052
Breakeven Precision: 0.222222222222222

Topic: 22
Total number of relevant docs = 171
Total number of retrieved relevant docs = 128
Average (non-interpolated) precision = 0.339972316781161
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 0.642857142857143

prec at 0.2 = 0.6
prec at 0.3 = 0.536082474226804
prec at 0.4 = 0.376344086021505
prec at 0.5 = 0.296666666666667
prec at 0.6 = 0.247596153846154
prec at 0.7 = 0.189952904238619
prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 1
prec at 10 = 0.9
prec at 15 = 0.733333333333333
prec at 20 = 0.75
prec at 30 = 0.6
prec at 100 = 0.52
prec at 200 = 0.355
prec at 500 = 0.222
prec at 1000 = 0.128
Breakeven Precision: 0.380116959064327

Topic: 23
Total number of relevant docs = 74
Total number of retrieved relevant docs = 38
Average (non-interpolated) precision = 0.246097034531665
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 0.823529411764706
prec at 0.2 = 0.714285714285714
prec at 0.3 = 0.27710843373494
prec at 0.4 = 0.08
prec at 0.5 = 0.0480519480519481
prec at 0.6 = 0
prec at 0.7 = 0
prec at 0.8 = 0
prec at 0.9 = 0
prec at 1 = 0
Non-interpolated precsion at docs:
prec at 5 = 1
prec at 10 = 0.7
prec at 15 = 0.8
prec at 20 = 0.7
prec at 30 = 0.566666666666667
prec at 100 = 0.23
prec at 200 = 0.125
prec at 500 = 0.064
prec at 1000 = 0.038
Breakeven Precision: 0.297297297297297

Topic: 24

Total number of relevant docs = 409
Total number of retrieved relevant docs = 357
Average (non-interpolated) precision = 0.707249273799482
Interpolated precsion at recalls:
    prec at 0 = 1
    prec at 0.1 = 0.98
    prec at 0.2 = 0.964705882352941
    prec at 0.3 = 0.953846153846154
    prec at 0.4 = 0.923076923076923
    prec at 0.5 = 0.826612903225806
    prec at 0.6 = 0.719298245614035
    prec at 0.7 = 0.642384105960265
    prec at 0.8 = 0.515723270440252
    prec at 0.9 = 0
    prec at 1 = 0
Non-interpolated precsion at docs:
    prec at 5 = 1
    prec at 10 = 1
    prec at 15 = 1
    prec at 20 = 1
    prec at 30 = 1
    prec at 100 = 0.96
    prec at 200 = 0.9
    prec at 500 = 0.596
    prec at 1000 = 0.357
Breakeven Precision: 0.67481662591687

Topic: 25
Total number of relevant docs = 64
Total number of retrieved relevant docs = 59
Average (non-interpolated) precision = 0.662344255945024
Interpolated precsion at recalls:
    prec at 0 = 1
    prec at 0.1 = 1
    prec at 0.2 = 0.933333333333333
    prec at 0.3 = 0.740740740740741
    prec at 0.4 = 0.684210526315789
    prec at 0.5 = 0.666666666666667
    prec at 0.6 = 0.655737704918033
    prec at 0.7 = 0.584415584415584
    prec at 0.8 = 0.557894736842105
    prec at 0.9 = 0.453125
    prec at 1 = 0
Non-interpolated precsion at docs:
    prec at 5 = 1
    prec at 10 = 1
    prec at 15 = 0.933333333333333
    prec at 20 = 0.8
    prec at 30 = 0.7
    prec at 100 = 0.54

prec at 200 = 0.29
prec at 500 = 0.116
prec at 1000 = 0.059
Breakeven Precision: 0.640625

Topic: 26
Total number of relevant docs = 53
Total number of retrieved relevant docs = 53
Average (non-interpolated) precision = 0.678035220077291
Interpolated precsion at recalls:
prec at 0 = 1
prec at 0.1 = 1
prec at 0.2 = 1
prec at 0.3 = 1
prec at 0.4 = 0.958333333333333
prec at 0.5 = 0.870967741935484
prec at 0.6 = 0.363636363636364
prec at 0.7 = 0.306451612903226
prec at 0.8 = 0.3
prec at 0.9 = 0.3
prec at 1 = 0.174917491749175
Non-interpolated precsion at docs:
prec at 5 = 1
prec at 10 = 1
prec at 15 = 1
prec at 20 = 1
prec at 30 = 0.866666666666667
prec at 100 = 0.33
prec at 200 = 0.255
prec at 500 = 0.106
prec at 1000 = 0.053
Breakeven Precision: 0.566037735849057

Set average over 26 topics
Set average (non-interpolated) precision = 0.447461427669486
Set total number of relevant docs = 3091
Set total number of retrieved relevant docs = 2393
Set average interpolated precision at recalls:
avg prec at 0 = 0.982905982905983
avg prec at 0.1 = 0.880473866618368
avg prec at 0.2 = 0.761000780247991
avg prec at 0.3 = 0.654720938630329
avg prec at 0.4 = 0.542730795347003
avg prec at 0.5 = 0.44095768180659
avg prec at 0.6 = 0.286685304308082
avg prec at 0.7 = 0.208854678865097
avg prec at 0.8 = 0.146492607805713
avg prec at 0.9 = 0.0899717920121276
avg prec at 1 = 0.0367142376350751
Set non-interpolated precsion at docs:

avg prec at 5 = 0.876923076923077
avg prec at 10 = 0.792307692307692
avg prec at 15 = 0.761538461538462
avg prec at 20 = 0.717307692307692
avg prec at 30 = 0.646153846153846
avg prec at 100 = 0.442307692307692
avg prec at 200 = 0.307884615384615
avg prec at 500 = 0.162769230769231
avg prec at 1000 = 0.0111538461538462
Set breakeven precision = 0.450548795493923