CLASSIFICATION AND RETRIEVAL OF ENDOSCOPIC IMAGES FROM THE CLINICAL OUTCOMES RESEARCH INITIATIVE (CORI) COLLECTION

By

Jayashree Kalpathy-Cramer, PhD

A THESIS

Presented to the Department of Medical Informatics and Clinical Epidemiology

and the Oregon Health & Science University

School of Medicine

in partial fulfillment of

the requirements for the degree of

Master of Science

June, 2009

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Master's Thesis of

Jayashree Kalpathy-Cramer

"Classification and Retrieval of Endoscopic Images from the Clinical Outcomes Research Initiative (CORI) Collection"

Has been approved

Thesis Advisor – William Hersh, MD

Committee Member – Judith Logan, MD, MS

Committee Member – Xubo Song, PhD

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would to begin by acknowledging my thesis advisor Dr. William Hersh for all his guidance, encouragement and help with this thesis as well as throughout my informatics training program. The course that I took in Information Retrieval set me on my current path for which I am grateful!

I would like to thank my thesis committee members Dr. Judith Logan and Dr. Xubo Song for their support with this project and others as well as guidance during the last three years. I look forward to continuing these collaborations.

I would like to express my sincere appreciation to the National Library of Medicine for providing me with the funding for this fellowship which enabled me to make a significant and opportune mid-career course correction into academia.

The DMICE staff, including Andrea Ilg, Diane Doctor and Lynne Schwabe, have been immensely helpful during my time at OHSU. Their help has been instrumental in completing this program.

I have been lucky to have had great co-workers and friends –thanks! Finally, I would like to acknowledge the support and encouragement and love that my parents, husband Alan and son Avilash have shown me. Thank you all so much for the joy you bring me every day.

ABSTRACT

There has been a substantial growth in the number of images being created every day in healthcare settings. Effective image annotation and retrieval can be useful in the clinical care of patients, education and research.

Traditionally, image retrieval systems have been text-based, relying on the annotations or captions associated with the images. Although text-based information retrieval methods are mature and well-researched, they are limited by the quality and availability of the annotations associated with the images. Advances in techniques in computer vision have led to methods for using the image itself as the search entity.

The goal of our project was to create an image retrieval system a set of 1500 upper endoscopic images from the Clinical Outcomes Research Initiative Collection. We have created a web-based multimodal image retrieval system written using the Ruby on Rails framework. Ferret, a ruby port of Lucene was used for the text indexing of the annotations for the text-based retrieval. Our database also contains a number of visual features created using image processing algorithms that allows users to perform content-based retrieval. When operating in a “query-by-example” mode, our system retrieves an ordered set of images from the test collection that are “similar” in visual content to the image being queried. We also evaluated the performance of a variety of image

features and machine learning classifiers that can be used to automatically annotate the image with an image class consisting of one of eight findings. We developed a hybrid algorithm for image classification that showed improved performance compared to commonly-used classification algorithms. This enabled us to provided text-based querying capability where search words from a controlled vocabulary retrieve a set of pre-classified and annotated images matching the search criteria. Our intention was to enable users to query using either a sample image, keywords or desired image class to retrieve "similar images" from the system, along with a display of the associated information from these images.

Although CBIR has great potential in patient care, research and education, purely content-based image retrieval can be quite challenging for clinical purposes due to the semantic gap. Low level global features like color and texture may not be sufficient for classification of findings. However, combining visual and textual information can greatly improve retrieval performance. Additionally, the use of distance metric learning and relevance feedback can help the system produce results that are more relevant to the user.

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

Introduction

Image retrieval is a burgeoning area of research in medical informatics [1, 2]. With the increasing utilization of digital imaging in all aspects of health care and medical research, there has been a substantial growth in the number of images being created every day in healthcare settings. Consequently, there is a critical need to manage the storage and retrieval of these image collections, whether in stored in Picture Archival and Communication Systems (PACS) or in patient health records or on the web. Effective image annotation and retrieval can be useful in the clinical care of patients, education and research [2, 3]. Image retrieval can be used by clinicians to generate differential diagnosis, monitor response to therapy and for quality control. Medical students and residents have indicated that effective image retrieval can be useful for self-education [4]. A good image retrieval system can also be beneficial to healthcare practitioners for patient education. Data-mining of large image collections can provide useful information for researchers. Examples include prevalence of certain findings including polyps during routine screening [5], visual characteristics associated with malignancy in mammography [6, 7], and prediction of response to radiation therapy based on FDG-PET [8].

Traditionally, image retrieval systems have been text-based, relying on the annotations or captions associated with the images [9]. Although text-based information retrieval methods are mature and well-researched, they are limited by the quality of the annotations applied to the images. These techniques have limitations as 1) the annotations are subjective and context sensitive, can be quite limited in scope or completely absent; 2) manually annotating images is labor and time intensive, and error prone; 3) annotation are very "noisy" if they are automatically extracted from the surrounding text; and 4) there is far more information in an image than can be abstracted using a limited number of words. In addition, these annotations can experience a "semantic gap" as described by Smeulders et al. [10], where the annotation may not fully describe the semantic context of the image.

Advances in techniques in computer vision have led to methods for using the image itself as the search entity. In content-based image retrieval (CBIR), the visual information from the image is mathematically abstracted and compared to similar abstractions of all images in the database [10]. These features could include the color, shape or texture of images. An ordered-list of images that are visually most similar to the sample image is presented to the user.

However, the success of these methods when applied to a collection of diverse content can be limited. It is generally felt that the application of CBIR techniques to a more homogenous collection, like medical image databases, can prove to be useful [2]. Also, as shown in recent years, a combination of text-based and content-based image retrieval techniques can be very successful [11, 12].

The Clinical Outcomes Research Initiative (CORI) [13], is a research initiative that "acquires information that will improve the quality of clinical practice in gastroenterology." The data collected under this initiative "has been analyzed to examine endoscopic practice patterns, to develop research hypotheses, as a resource for prospective research on topics such as colon polyp surveillance, and to support quality measure reporting." In addition to case reports and pathological findings, images are also collected and archived in the CORI. Currently, over 400 endoscopists participate in this project at almost 90 sites, providing reports on procedures. Many of the participating sites also collect and send images attained during endoscopic procedures to the data warehouse. More than 150,000 images are currently available in the system. However, the annotations associated with these images are highly variable and have not been validated. OHSU has created a manually curated digital library collection consisting of 1500 images from eight endoscopic findings for the upper GI tract. Associated textual information includes the finding, size, anatomical location and comments about visual appearance.

The goal of our project was to create an image retrieval system for this subset of the CORI collection. Our system retrieves an ordered set of images from the test collection that are "similar" in visual content to the image being queried when operating in a "query-by-example" mode. It also has limited text-based querying capability where search words from a controlled vocabulary retrieve a set of pre-classified and annotated images matching the search criteria. Our intention is that either a query image or keywords can be used to retrieve “similar images” from the system, along with a display of the associated information from these images.

CHAPTER 2

Background

Advances in computing capabilities and the pervasiveness of digital images in the last few decades have resulted in significant strides in the areas of image processing, computer vision and pattern recognition. Algorithms have been developed to analyze images, classify images and automatically annotate images as well as retrieve similar images. Computer aided detection (CAD) is routinely used in medicine in areas as diverse as mammography, pathology and endoscopy [7, 14].

Image retrieval

Traditionally, researchers have taken three broad approaches in the task of images retrieval: text-based retrieval, content-based retrieval and automatic annotation based retrieval.

Text-based image retrieval

Image retrieval most commonly has been accomplished using the textual annotations associated with image. Most popular web-based image search engines including Google images[1], Yahoo! image search[2], and Bing[3] as well as

---

[1] http://images.google.com

[2] http://images.search.yahoo.com

[3] http://www.bing.com/images

medical image retrieval engines including Goldminer[4] [15] and Yottalook![5] perform text-based image retrieval. The search entities are keywords and the results are images. Although this is still the most commonly used approach in non-academic situations, there can be numerous issues with this approach. Manually annotating images is a very labor-intensive task. Not all images have well-curated annotations. The annotations can depend on the annotator and the context in which they were labeled. Other issues include the problem with synonymy as shown in Figure 1 below where the word "mum" could refer to mother or chrysanthemums.

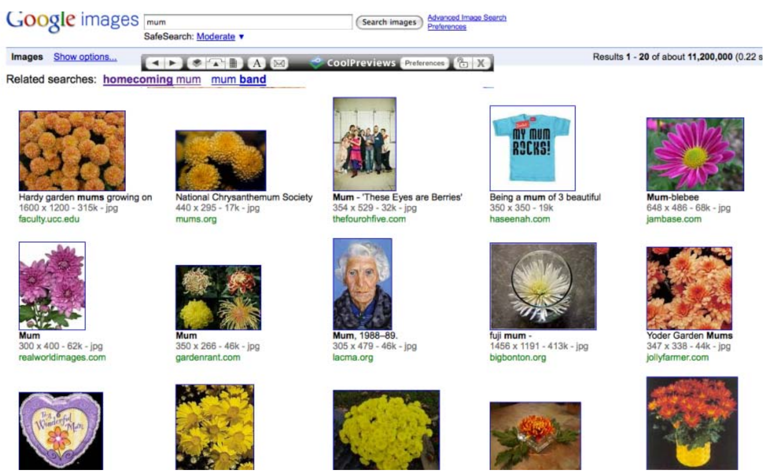

Figure 1 Searching for "mum" in Google shows images for different synonyms of the word

---

[4] http://goldminer.arrs.org/

[5] http://www.yottalook.com

Content-based image retrieval

Advances in image processing and the prevalence of digital images led to the notion of using the image itself as the search entity. The early nineties saw the advent of some of the early CBIR systems. The idea behind these systems was that the user would upload an image and the system would return an ordered list of images that are visually most similar to the search image. These systems were typically used for color-images of real-life scenes (animals, cars, airplanes, faces etc). Reviews of these systems can be found at [10,16-18] There were many different techniques based on image properties like color, texture, shape.

One of more successful of the early systems was the QBIC (Query by Image Content) system from IBM [19]. This system uses color histograms, texture [20, 21] and shape descriptors (moments) for image retrieval.

Other pioneering systems include the BlobWorld system [22]and the NETRA system. Blobworld was developed at the University of California, Berkeley and uses color, texture, location and shape of the regions (blobs) and background as features. NETRA [23], developed at UC, Santa Barbara, also uses color, texture using Gabor filters [24], shape and spatial location as features.

However, content-based image retrieval systems often suffer from the "semantic and sensory gaps" [10]. Smeulders et al. identified the 'semantic gap' as "the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a given user in a given situation." The goal of many researchers in image retrieval continues to be to find ways to close that gap [25-27].

One manifestation of the semantic gap in medical images is the difference between the image itself and the contents of the annotations associated with the image. The same image may be annotated quite differently depending on the user, his training, and/or his medical expertise. The 'sensory gap' between "the object in the world and the information in a computational description derived from a recording of the scene" is also pertinent in medicine. This sensory gap can be demonstrated by the differences between the actual tumor in the physical world and how it is imaged under various modalities (e.g., CT or MRI) and views (prone or supine).

In comparing image retrieval to text retrieval, Smeulders et al. [10] note the lack of a sensory gap in text retrieval. They also note that the difference between the semantic gap in text retrieval (between keywords to full text) to that in image retrieval. The differences in semantic and sensory gaps between textual and visual retrieval may shed some light on why image retrieval systems currently do not perform as well as their textual counterparts.

Many reviews, including a recent one performed by Datta et al. [18], have compared the image retrieval performance of various color, texture, shape and saliency features that were extracted from images in a database. However, establishing the correspondence between the semantics of an image and its mathematical properties such as color, shape and texture remains a very active area for research.

Automatic annotation and classification

More recently, automatic annotation-based techniques have been developed to label images with image tags [28-30]. Images can be organized and retrieved easily based on the annotations. Typically, these methods require a training set of labeled images consisting of tags from a constrained vocabulary. Using these training images, image tags can be propagated to new images. Retrieval is then accomplished using the image tags. Behold[6] and ALIPR[7] are examples of image search engines that use automatically generated image tags.

Image Retrieval in Medicine

Mueller et al. have performed an extensive review of the use of image retrieval in medicine [2]. Image retrieval in medicine is most commonly performed within the purview of PACS systems, where the images are retrieved using either the patient or study ID. However, Mueller et al. argue that content-based methods can be a useful functionality to be integrated with PACS systems. Mueller et al. identify teaching, research and diagnostics as three primary domains for applying image retrieval. They identify an important use of purely

[6] http://www.behold.cc

[7] http://www.alipr.com

visual (or content-based) image retrieval, namely, “Visual features do not only allow the retrieval of cases with patients having similar diagnoses but also cases with visual similarity but different diagnoses.”

They note various image retrieval engines used in many clinical disciplines including for high resolution CT images of the lung (ASSERT) [31], spine x-rays [32,33], pathological images, CT’s, mammograms, dermatology and varied collection (IRMA) [21,34].

MedGIFT [35,36] is an adaptation of the GIFT, an open source CBIR system developed at the University Hospitals of Geneva. The MedGIFT system uses color histogram in HSV space and Gabor coefficients for texture characterization. This system uses an inverted file structure (similar to text based method) for improve retrieval efficiency. Other notable image retrieval systems used in medicine include SPIRS [37,38] for spine x-rays and BRISC for images from the Lung Image Database Consortium (LIDC) [39]. SPIRS allows users to query using either text or images.

Image processing and machine learning in endoscopy

Image processing and machine learning techniques have been applied to numerous medical applications in both the grey scale and color image domains. Kodogiannis et al. created a system for computer aided diagnosis in endoscopy using image processing and machine learning [40-42]. In [41], the authors first create histogram in the RGB and HSV spaces. Statistical measures of standard deviation, variance, skew, kurtosis, energy, entropy, inverse difference moment, contrast and covariance are used to create a 54-dimension feature vector. Different techniques including Extended Normalized radial basis function networks, adaptive fuzzy logic systems and fuzzy integrals are then used to create a fusion of multiple classifiers. A fairly high accuracy of classification was achieved.

Karkanis et al. [43] used color wavelet covariance based feature vectors, followed by Linear Discriminant Analysis (LDA) for the detection of abnormal colonic regions.

Artificial neural networks were used by Tjoa et al. [44,45] to detect abnormal colonoscopic area. They compared the use of Backpropagation and Adaptive resonance theory networks. They used a variety of segmentation methods and extracted features based on color and shape of the segmented regions.

User Search Behavior and Interaction

Researchers [3,4,10,18] have developed models to describe the search behavior of users. Users can be categorized in three broad classes: a) 'browser' as someone looking for pictures with no clear end goal, b) 'surfer' as a user searching with a moderate clarity of an end goal, and c) a 'searcher' who is very clear about for what she is searching. Tasks can similarly be classified as "lookup", "learn", and "investigate" [46]. Many leading scientists [47] in information science have stressed the importance of taking into account user behavior, whether someone is searching for a particular document/image or browsing with no particular end in mind or something in between, when designing user interface and information retrieval systems. User interaction, in the form of relevance feedback, can be useful for setting the context and training the system for the particular user and search.

Information Seeking Behavior of Users of Medical Image Systems

Image retrieval systems are expected to be useful tools for a variety of consumers of health information, from clinicians performing diagnosis to the general public trying to understand health conditions. However, only a few studies [3, 4, 48] have looked at the use-behavior of image retrieval system users. Mueller et al. noted that many clinicians store reference images from past cases, often on their personal computers. Most of the clinicians interviewed do not believe that the CBIR systems in medicine are ready to be used in a clinical setting. They identified "recommendations for search techniques that do not exist but are regarded as very useful: Search by pathology; Search by anatomic region; Search by visual similarity; Search by multi-modality combined to find similar cases; Indexation of the entire PACS by keywords regarding the pathology." We believe that it is very important to understand the needs and use patterns of real users in designing user-centric, clinically useful image retrieval systems.

Relevance Feedback and Similarity Learning

Relevance feedback has been used quite successfully in text retrieval [49-51]. In relevance feedback, the user reviews the results that are initially returned from a given search and provides feedback to the system about whether he thinks that the document or image retrieved was pertinent to the intended search. This can be explicit, where the user states which documents are relevant, or implicit, where the relevance in inferred from user behavior. Relevance feedback can be useful when the query is hard to formulate, but the user may recognize the document as being pertinent when he sees it. Such feedback can help the user identify new search terms or discover concepts. Many researchers believe that relevance feedback can be critical in image retrieval as a method to bridge the semantic gap [52]. This feedback can be used in the short term where the results of the subsequent search are enhanced using the information provided by the user. In addition, it can be used for long term learning to provide training data for machine learning. This can be used to adapt the system for a particular user [53].

Similarity learning is another attractive approach to tuning the system to meet the user's needs [54, 55]. Early research in CBIR systems evaluated the use of similarity measures such as Euclidean distance, Earth mover's distance, Histogram intersection, and Mahalanobis distance for retrieval performance [56, 57]. However, Squire et al. [58] noted that "current systems face great difficulties, due to the fact that perceived image similarity is both subjective and task dependent." They proposed a distance-learning network that allows the system to learn distances that better correlate to human perception. Ma et al. [59] have previously shown improved image retrieval performance by learning similarity. More recently, in computer vision, Chopra et al. [60] and others [55,61] have demonstrated that learning similarity measures can significantly improve performance by transforming the feature space such that objects that are semantically similar will have reduced distance in the new space while objects that are dissimilar will have large distances.

Multimodal Fusion for Improved Retrieval

Multimodal fusion is an extremely promising approach to image retrieval [21, 62, 63]. This approach take advantage of both the visual and textual information that can be associated with an image. The textual information can include annotations, captions or text contained on a web page near the image, while the visual information can include the color, shape or texture contained within the image. The query itself could be multimodal, with the user supplied key words as well as sample images. By leveraging both types of available information, this approach seeks to support multimodal queries in which the user could supply key words (e.g. text) as well as sample images to improve the likelihood of more precise results. However, as noted by Datta et al. [18], this is still an emerging area of research, with only a small number of actual systems that employ this technique. Some participants of ImageCLEF 2006 and 2007 have had successful submissions using a variety of algorithms to combine textual and visual features

[63, 64]. The text-based annotations associated with the image can also be used as supervised training data for visual classification. Class labels can be then learned from training examples to aid in image retrieval [65].

Evaluation of Classification and Image Retrieval

The Cranfield methodology is typically used to evaluate the effectiveness of information retrieval systems using test collections and relevance judgements [66-68]. Commonly used metrics for information retrieval include precision, recall, and mean average precision (MAP). In order to evaluate the performance of an information retrieval system, we require a set of topics, a set of documents (or images) that are retrieved by the search system and a set of judgements to indicate the relevance of the returned document or image to the user's search. Relevance is typically binary, with the returned documents being deemed either relevant if they satisfy the user's information need or not relevant if they do not. Judgements are performed by domain experts. Precision is the fraction of the documents retrieved that are relevant to the user's information need. It is often specified given a number of documents retrieved. For example, P(10) is the number of relevant documents in the top 10 documents retrieved. It is analogous to positive predictive value. Recall is the fraction of the successfully retrieved documents that are relevant to the query. It is analogous to sensitivity [66, 69]. Hence, a system with high precision will return mostly relevant documents, while a system with high recall will return all the relevant documents in the database.

However, there is limited research to indicate the preference of real users of medical image retrieval systems as to which measures mostly closely correlate to user satisfaction with the retrieval results. Search engines usually utilize precision as it has been shown that most users of search engines typically do not view more than the first or second page of results [70, 71]. However, there are many clinical and research applications where an exhaustive search may be necessary and recall would be preferred, such as a genomics researcher interpreting the vast output of a microarray [72] or the clinical epidemiologist performing a systematic review [73].

In the case of a multi-class image classification evaluation, the typical metric used to compare performance of systems is the accuracy or hit rate. This is the ratio of number of images that are correctly classified to the total number of images.

Recently, Cohen's kappa [74] has been suggested as a better metric for comparing classification accuracies since the hit rate does not account for classification by chance.

In the case of a multi-class classifier, Cohen's Kappa is defined at

$$K = \frac{N * \sum_{i=1}^{I} x_{ii} - \sum_{i=1}^{I} x_i x_j}{N^2 - \sum_{i=1}^{I} x_i x_j}$$

where x_{ii} is the count of cases in the main diagonal ,N is the number of examples, I is the number of classes, and x_j and xi are the column and row total counts, respectively. Cohen's kappa can be calculated in a straightforward manner based on the confusion matrix obtained by the classifier.

CHAPTER 3

Methodology

The basic goal of image retrieval is to find images that are most 'similar' to a given query image. We framed this problem as both an unsupervised problem of finding the most similar images to a query image as well as a supervised machine learning problem of image categorization, given labels on a training set or pair-wise constraints between images as during a relevance feedback phase. Image retrieval systems typically consist of two phases. During the indexing phase, image features are extracted and stored from all images in the database. During the retrieval phase, image features are extracted from the query image and compared to all similar features that were indexed. An ordered list of most similar images is presented to the user.

Image Collection

The CORI test collection that was used in this project included about 1500 endoscopic images that have been categorized into eight pathological findings: Barrett's, esophagitis, nodules, polyps, stenosis, tumors, ulcers, and varices. In addition, the images in this collection have corresponding textual information about the location at which these images were acquired, associated findings and optional comments. The images are typically about 450x450 pixels in size and 8bits/channel RGB images. An example image and associated annotation is given below.

Title: Upper Gastrointestinal Polyps	
MeSH Subject: Endoscopy, Gastrointestinal; Intestinal Polyps	
Description: Maximum size: 2 mm, sessile, located in distal esophagus	

Figure 2 Example of image and associated annotation in the CORI collection

Framework of image retrieval system

A framework of generic image retrieval systems is show below in Figure 3. In the pre-processing phase, the images are first cropped, resized and potentially transformed to a different color space. In the indexing phase, feature vectors are then extracted for all images using one or more algorithms from the feature extraction algorithm library. Feature vectors and references to the images are stored in the relational database. The annotations associated with the images are also stored in the database and indexed.

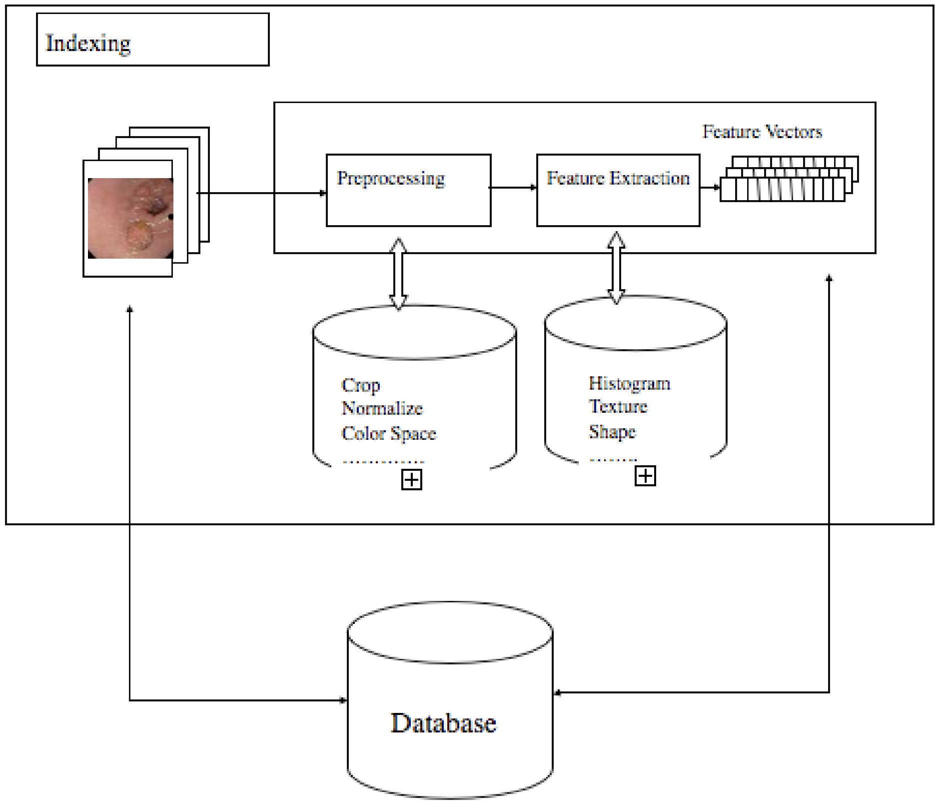

Figure 3: Framework for the indexing phase of an image retrieval system

If labels are known for a subset of images, this information can be used to improve the retrieval process using machine learning. If sufficient images are available for training and the labels are of a small number of categories, the unlabeled images can be categorized using supervised classification techniques. The system allows such an automatic annotation phase to be performed. First, a set of training images and associated annotations are identified. We then extract tags from a constrained vocabulary for these images. These include location and finding for the endoscopic images. Using machine learning algorithms, we create a classifier for the tags which are then propagated to the test images. We then evaluate the performance of the classifier. If the classifiers perform well, the entire database of images is labeled. These labels can be used in the retrieval phase by allowing users to search using keywords.

In the retrieval phase, the query image is preprocessed in a similar manner to the database images. Feature vectors are extracted similarly. Distances between the feature vector of the query image and all images in the database are computed using one of the distance measures from the distance measure library. The most similar images are then presented to the user.

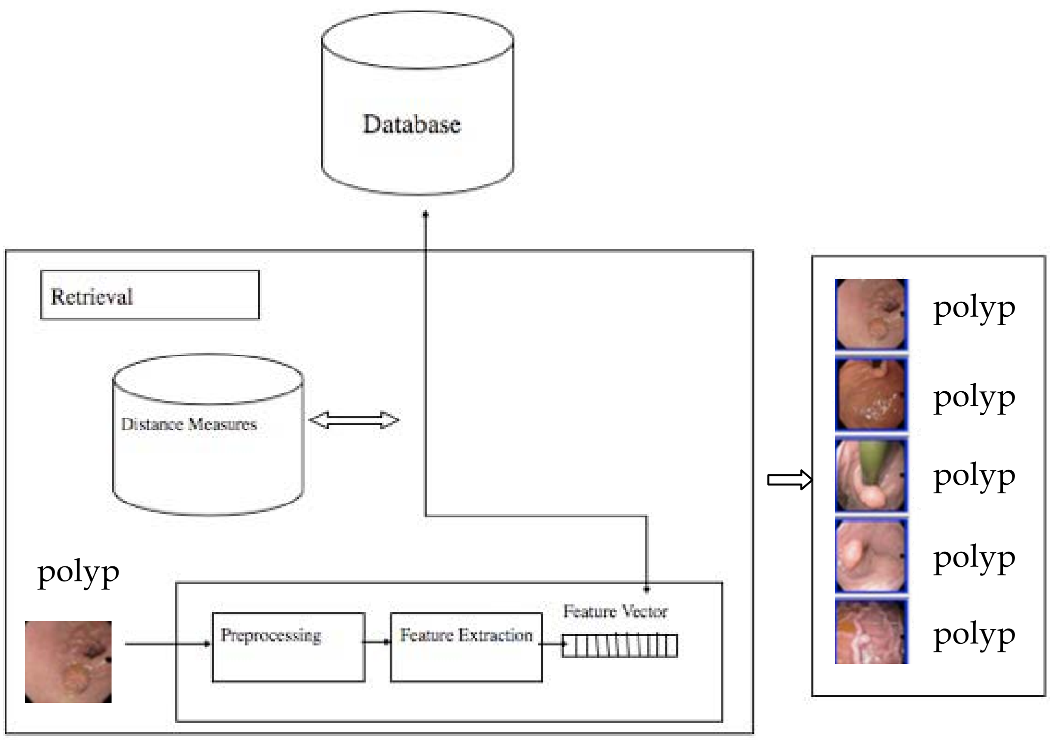

Figure 4 Framework for the retrieval phase of an image retrieval system

Alternatively, using the class labels, the distance space between images can be transformed in such a way that distances between images in the same class is minimized while distances between images in different classes is increased. Finally, users can indicate their preference for the images retrieved by using relevance feedback. This allows the distances between images to be updated to more closely reflect the users' notion of similarity. Relevance feedback can also be used for distance metric learning by establishing must-link and cannot-link constraints between images.

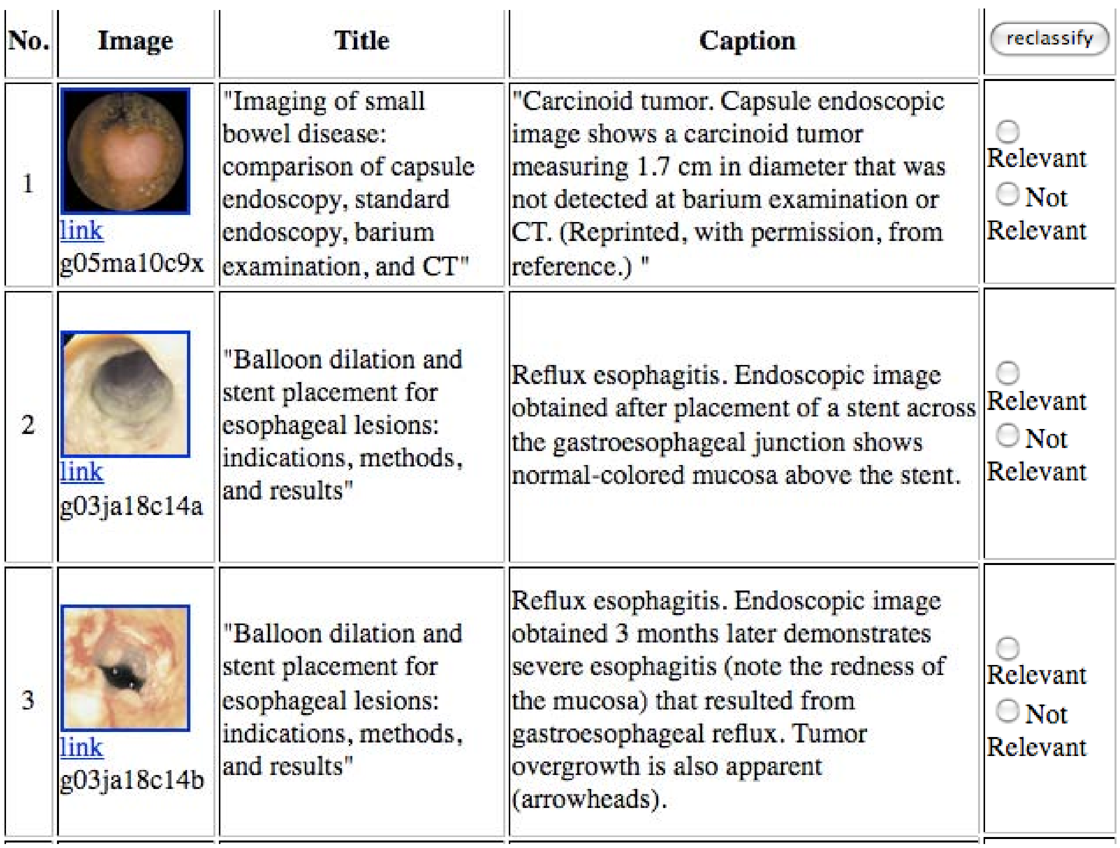

Figure 5 Graphical User Interface of OHSU's image retrieval system that enables users to provide feedback about the relevance of images.

System architecture

As described above, the basic components of an image retrieval system include a database to store images and annotations, associated tags and/or features, image processing algorithms, an indexing system, a search engine, and a user interface.

Database and Web Application

We used the Ruby programming language[8], with the open source Ruby On Rails web application framework[9] to build our web-based image retrieval system. The image locations, annotations and other user-created features were stored in a PostgreSQL relational database[10]. There are also fields for storing visual features that were created using image processing techniques. The relational database allows us to maintain the mappings between the images associated annotations, as well as other features that are derived using image processing and supervised machine learning methods. This format facilitates retrieval using both textual and visual techniques. The text annotations in the collection as well as some visual features are currently indexed and we continue to add indexable fields for incorporating new visual and textual features.

---

[8] http://www.ruby-lang.org/en/

[9] http://www.rubyonrails.org

[10] http://www.postgresql.org/

Text-based Search Engine

Ferret[11], a Ruby port of the popular Lucene [12] search engine, was used in our system for indexing and searching the text-based annotations. This system uses the well known TF-IDF (term frequency/inverse document frequency) to rank documents containing the search terms. Ferret allows us to perform Boolean and fuzzy searches as well as many of the classical Natural Language Processing (NLP) techniques, including the use of stop words and stemming.

Image processing

Most of the computationally intensive portions of the CBIR aspect of our system were implemented in MATLAB[13] using open-source toolboxes and in-house software. This included the pre-processing, feature extraction and distance metric learning. The distance matrices were then exported and stored as text-files or in the database in order to easily visualize the efficacy of the various algorithms.

---

[11] http://ferret.davebalmain.com

[12] http://lucene.apache.org/java/docs/

[13] http://www.mathworks.com

Preprocessing

We created a library consisting of a variety of routines to preprocess images. Some of the images in our collection had wide, dark, frames while others had no border at all, while a majority had triangular dark patches at the corners.

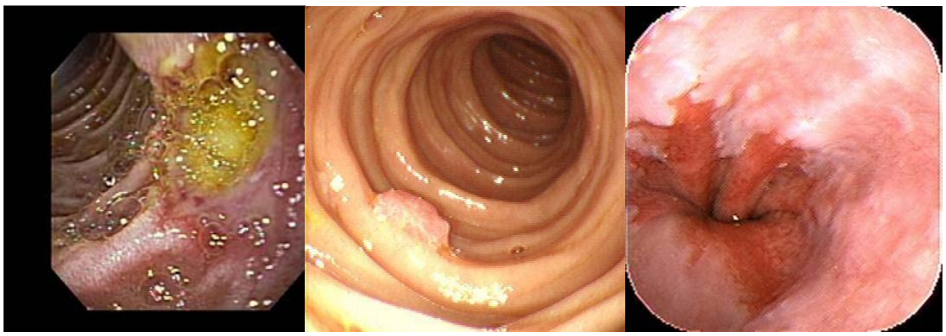

Figure 6 Variation in dark frames around images needs to be addressed before feature extraction

This variation caused problems with most feature-extraction algorithms as features from the edges would be included along with features from the main region of interest. In comparing these images, the dark edge features were a prominent source of variation between images. We created an algorithm to create a mask to separate the frame from the main image. We developed two techniques for this. A Canny edge filter was first used with appropriate thresholds. A non-parametric snake was initialized at the outer boundary of the image and converged within two iterations on the desired boundary, if it existed. We also created a simple heuristic method where, starting at all 4 outer edges, we considered all pixels to be the mask until a certain pixel intensity threshold was reached. The mask was saved for all images and used prior to feature extraction. In the case of some feature, we cropped the images and resized them to 256x256. We also normalized the images to have a mean of 0 and a variance of 1.

The images in our collection were 8 bit images RGB images. However, as the RGB color space is not believed to correspond well to human perception, the images are transformed to HSV or CIE L*a*b* color spaces.

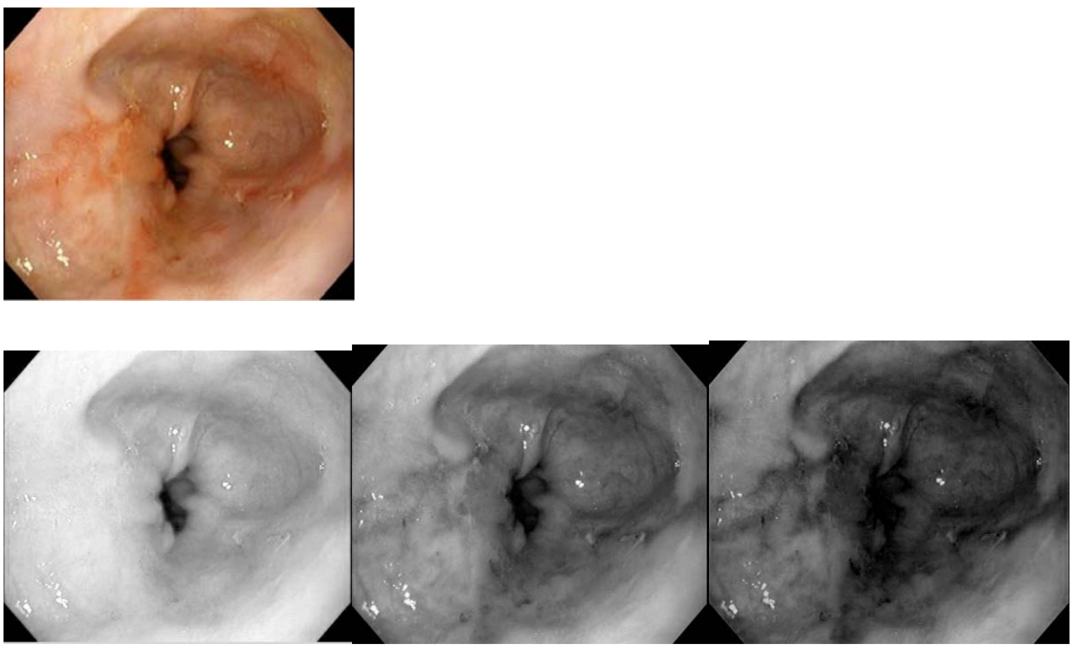

Figure 7 a) Original RGB image b) Red channel c) Green channel d) Blue channel

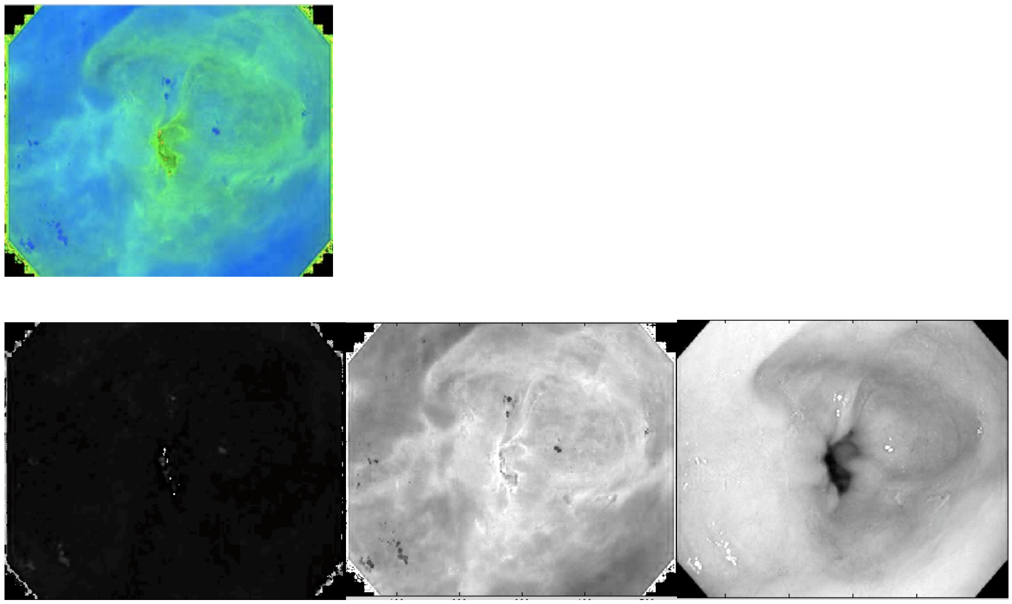

Figure 8 a) Original HSV image b) Hue c) Saturation d) Value (false colors)

Feature extraction

Image processing techniques are employed to create an n-dimensional feature vector for each image that can be considered a "signature" of the image. Feature extraction is a critical component of any image retrieval systems. Ideal features are those that have the ability to be semantically discriminatory i.e distances using such features from images that are close to the query image should small while distances to images from semantically different images should be large. Below are a set of routines that were used to create feature vectors used in our visual retrieval system.

Icon: this is a simple, but surprisingly effective feature vector in many cases [76]. The images is resized to 16x16 using bilinear interpolation and converted to a 256-dimensional vector that is compared with other similar features in the database. In the case of color images, a 16x16x3 icon is used, as shown below.

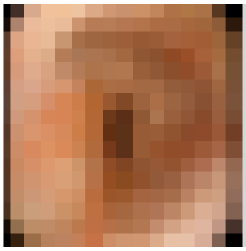

Figure 9 16x16x3 icon of an endoscopic image

GLCM: Four gray level co-occurrence matrices (GLCM) [77, 78] matrices with offsets of 1 pixel, 0, 45, 90 and 135 degrees were created for the image after rescaling the image to 16 levels. GLCM statistics of contrast, correlation, energy, homogeneity and entropy were calculated for each matrix. A 20 dimensional vector was created for each image by concatenating the 5 dimensional vector obtained by each of the four matrices.

GLCM2: In order to capture the spatial variation of the images in a coarse manner, the resized image (256x256) was partitioned into 5 squares of size 128x128 pixels (top left, top right, bottom left, bottom right, centre). A gray level correlation matrix was created for each partition. A 20 dimensional vector was created for each partition. Subsequently, the 5 vectors from each of the partitions were concatenated to created feature vector of dimension 100.

Histogram: We created a 32-bin gray-scale intensity histogram as well as a 96-bin color histogram for each image.

DCT: A global discrete cosine transform was created for each image. The upper left (10x10) vectors were concatenated and used as inputs.

Gist: Images were convolved with a set of 32 multiscale-oriented Gabor filters [79-81]. We created a 512- dimensional vector using statistics from these filters. Principal component analysis was then used to reduce the dimensionality of the vector to 80. The 512-dimensinal vector as well as the 80-dimensional vector were evaluated. Feature vectors were also created by concatenating one or more of the above features.

Texture: Texture-based features are typically derived using filter response of a set of filters with the image. We have evaluated a variety of filters including Gabor, wavelet, Maximum response (MR8), Leung-Malik (LM) filter bank, and the Schmid Filter bank [82].

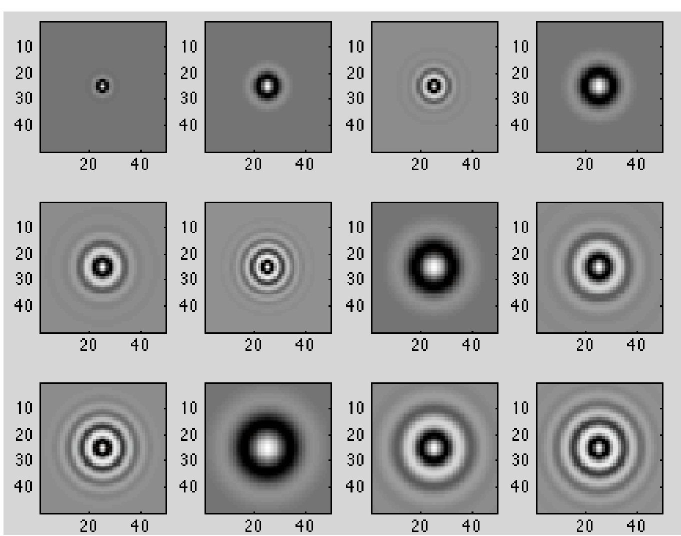

Figure 10 a set of 12 radial Schmid (S) filters

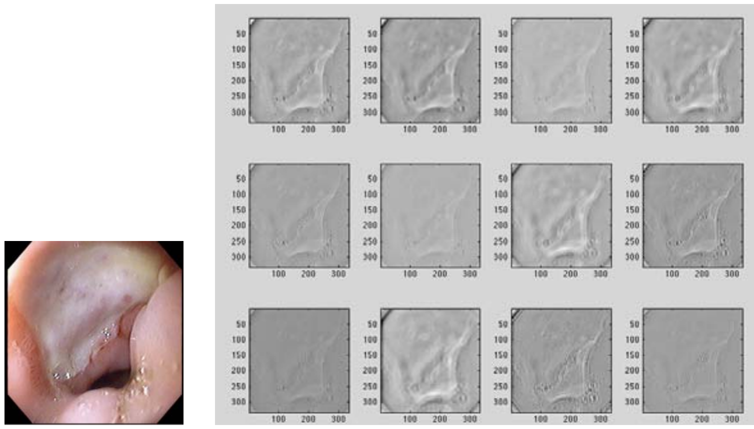

Figure 11 An endoscopic image with the response of the S filter bank

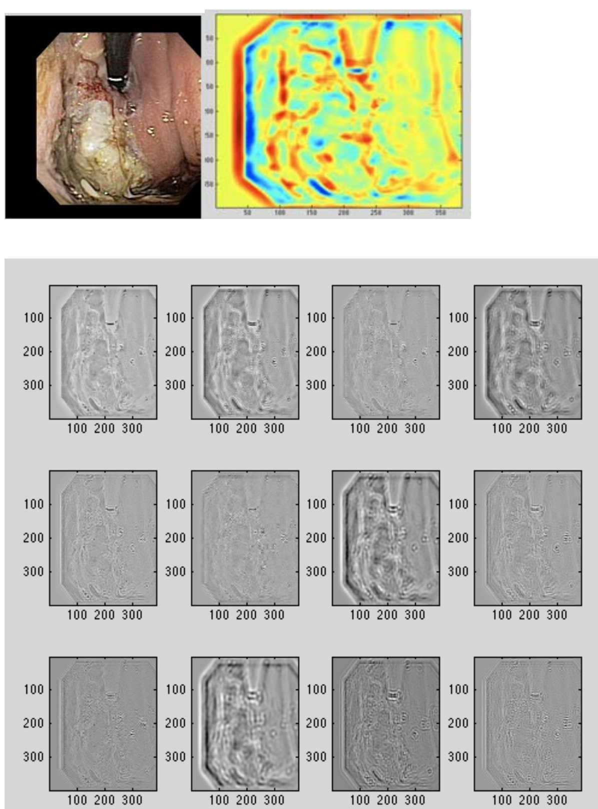

Figure 12 an endoscopic image with the sum of the responses of the filter bank as well as individual filters

Textons: Textons [83-85] are a popular and effective class of algorithm for texture categorization. They are basic repeated units of texture learnt by clustering responses to a set of filters, where the cluster centers are labeled 'textons'. As shown in the schematic below, filter responses from a set of images from each category are aggregated and quantized into textons using the k-means clustering algorithm. These are combined into a texton dictionary in the training stage as shown in Figure 13.

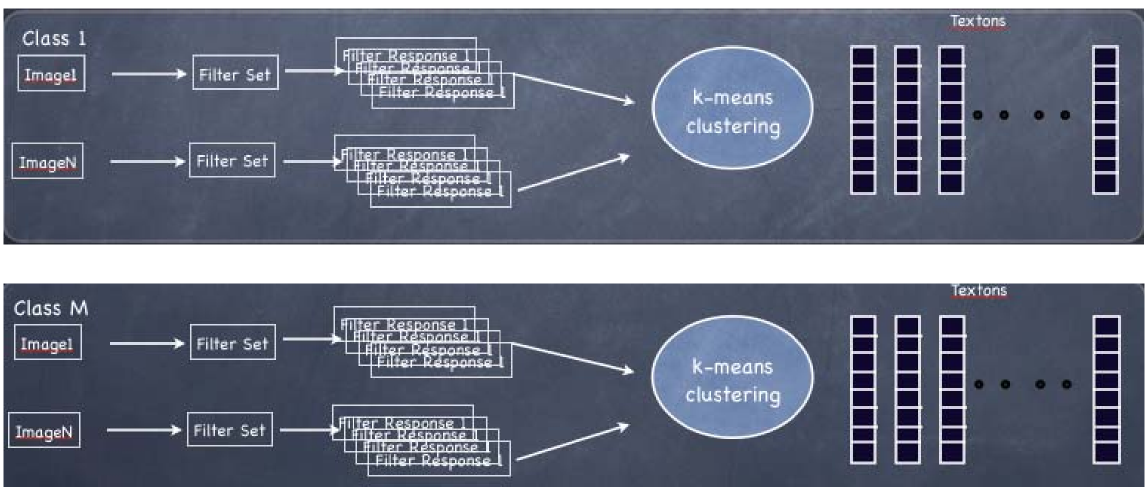

Figure 13 Creation of texton dictionary by quantizing filter responses

In the class-model creation stage, images in the training set are convolved with a filter bank and the response of each pixel is labeled with the nearest texton. A histogram of texton frequency is created for each image in the training set Multiple models for each class are created by quantizing the filter responses of the images as shown below in Figure 14.

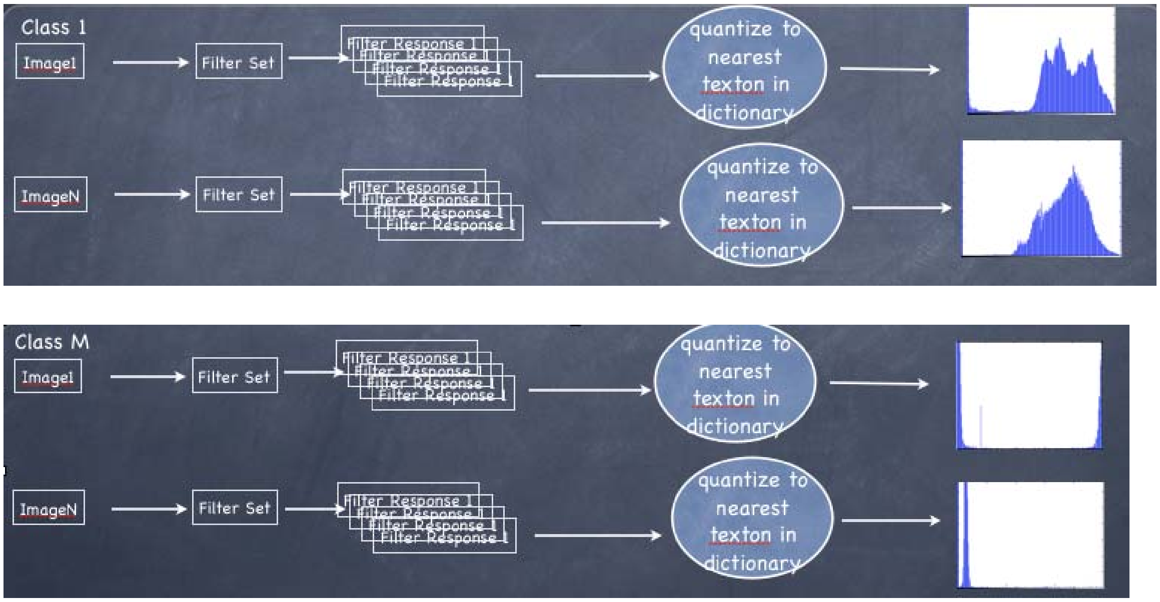

Figure 14 Class-model creation stage where images are convolved with filter bank to create histogram of textons

Finally, each new image is classified based on the chi-squared distance of its histogram of textons to all other histrograms from the training set as shown below. Additionally, more sophisticated classifiers can be used.

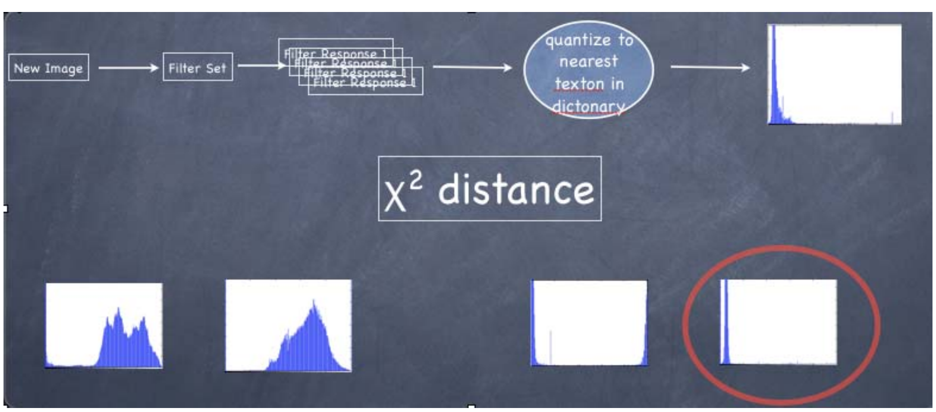

Figure 15 Classification of image by minimizing chi-squared distance of image-to-image histogram

Modified textons: We also created modified textons using image patches instead of filter responses as described in Varma and Zisserman [82]. In this case, instead of aggregating filter responses, random pixels in the image were sampled and an feature vector consisting of the pixel values in the neighborhood were used to create textons.

Naive Bayes NearestNeighbor (NBNN):

This surprisingly effective algorithm [86] uses a nearest neighbor approach with image descriptors like those used for textons. However, it differs from the texton -based approaches in two major aspects : there is no quantization of images descriptors, which the authors argue causes degeneration in performance; and image-to-class distances are calculated instead of image-to-image. A set of training images are identified and image descriptors for all these images are calculated. Similarly, descriptors for the test image are calculated. The class that minimizes

$$\sum_{i-1}^{n} \| d_i - NN_c(d_i) \|^2$$

where $d_i$are the image descriptors of the test image, $NN_C(d_i)$ is the nearest neighbor descriptor of d_i in class C is chosen as the class for the test image

Classifiers

Three types of classifiers were evaluated using a variety of the above features in categorizing images into one of 8 classes. All features were evaluated using the nearest neighbor classifier as a baseline. In addition to the Euclidean distance, the Mahalanobis distance was also used. The classifier is the simplest and does not require training. On feature sets that performed well, Support Vector Machines (SVM)[14] and a multi-layer perceptron (MLP) were used as additional classifier. All three classifiers were open-source implementations. The MLP, using the Netlab toolbox [87] had a two layer structure, with a hidden layer of approximately 50-400 nodes. A variety of combinations of the above image features were used as inputs. All inputs to the neural network (the image feature vectors) were normalized using the training set to have a mean of zero and variance of 1. The architecture was optimized using the training and development sets provided.

Affinity propagation: Frey et al. [88] recently formulated an algorithm for clustering they found to be significantly faster than many other clustering algorithms. It identifies exemplars are cluster centers using message passing. It is robust and does not require the number of clusters to be know a-priori but can be forced to converge to a required number of clusters. It also utilizes the incorporation of user-mandated exemplars.

Hybrid Algorithm: We formulated a hybrid algorithm for the classification task. We first identified a small set of exemplar training images from each class using

[14] http://www.igi.tugraz.at/aschwaig/software

the affinity propagation algorithm. Patch-based image descriptors were used for the feature vectors. Finally, we used the NBNN classifier and minimized the image-class distances, not the image-image distances.

*Distance metric learning:*In addition to the image features, the distances used to compute similarity between images is also critically important in image retrieval. Historically, distances were calculated using Euclidean norm or other static measures like the Minkowski distance, earth-mover's distance, histogram intersection [56, 89]. However, the user's notion of similarity is subjective and context-dependent. Distance learning algorithms use the distribution of the data to inform the distance spaces, either using the class information or pair-wised constraints. Relevant Component Analysis [91] is a simple and efficient algorithm for learning the Mahalanobis metric.

We have evaluated the performance of a variety of distance metric-learning algorithms and enhanced constrained affinity propagation as a method to incorporate relevance feedback. We established pair-wise constraints in the form of must-link within class and cannot-link between class images on 20-100 images per class.

Constrained affinity propagation [90]: Pair-wise constraints are specified on a small number of initial pairs. The affinity measures based on features is available for all image pairs. Each time a new constraint is provided in the form of relevance feedback, a row and column of the affinity matrix are updated, reflecting a new order for similar images.

Evaluation

The system was evaluated both for classification as well as retrieval. In the case of the classification task, we evaluated the efficacy of the various feature vectors and classifiers for both the findings and the location task. 10-fold cross-validation was used when the time for training/testing was less than 15 minutes per fold. When more computationally-intensive algorithms were used, a 50-50 training-test split was used for evaluation.

Classification rate (or error rate) is the most commonly-used measure for classification accuracy. However, as described in the previous chapter, it has limitations as it does not account for classification by chance. We also measured Cohen's kappa using the confusion matrix.

Precision at 20 was measured for each category given using the original distance matrices as well as those obtained using the pair-wise constraints.

CHAPTER 4

Results and Discussions

Our three-pronged approach to building the image retrieval system consisted of a text-based retrieval engine that indexed any available annotations, an automatic annotation system based on visual features that could be used to attach class labels to the images and the use of distance-metric learning for relevance feedback. For the evaluation shown in this chapter, we used the subset of CORI images for which the class labels as well as annotations were available. Here, we present the results obtained using the various image processing and machine learning algorithms that we evaluated for classification and retrieval of endoscopic image.

Pre-processing results

We found that it was critical to pre-process the images. Sample results without pre-processing to remove the uneven black border can be seen below. In both cases, the top image in the list was the search image. As can be seen, it is the shape of the border similarity that predominates the ordering of the results rather than the image itself.

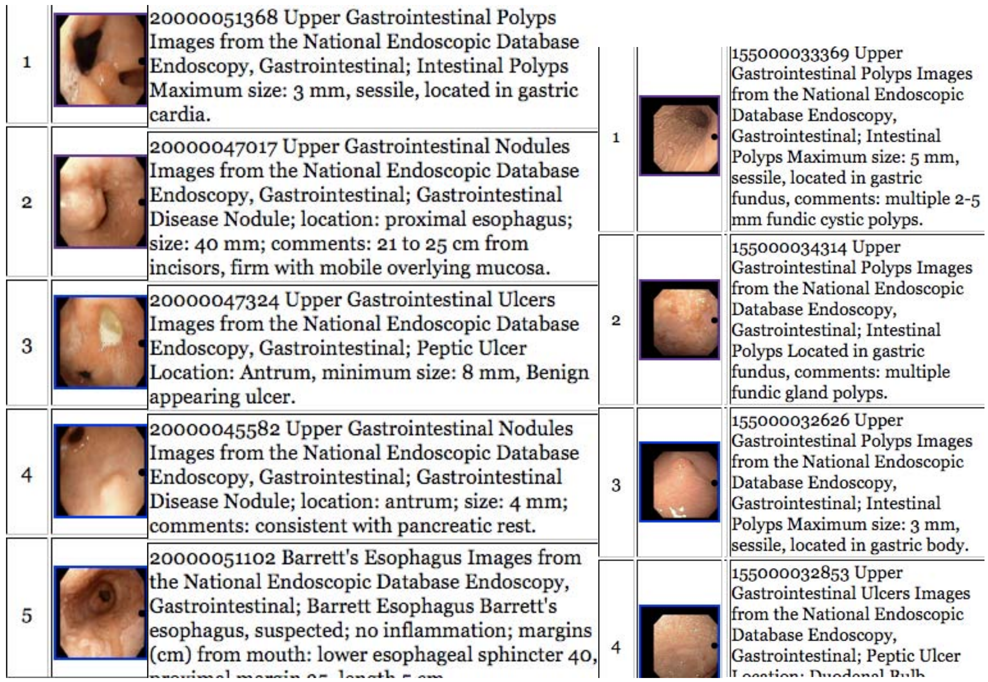

Figure 16 shows the importance of pre-processing prior to feature extraction. Similarity of images retrieved without pre-processing is primarily determined by the shape of the border rather than the image contents.

Our frame detection worked relatively well to generate masks. We reviewed all 1500 images and the cropping was within a few pixels of the ideal border. Only pixels that were not masked out (white pixels) were used for the feature extraction. As seen in figure 17, we have three images with vastly different black borders. Below each figure is the corresponding mask resulting from our simple frame detection algorithm.

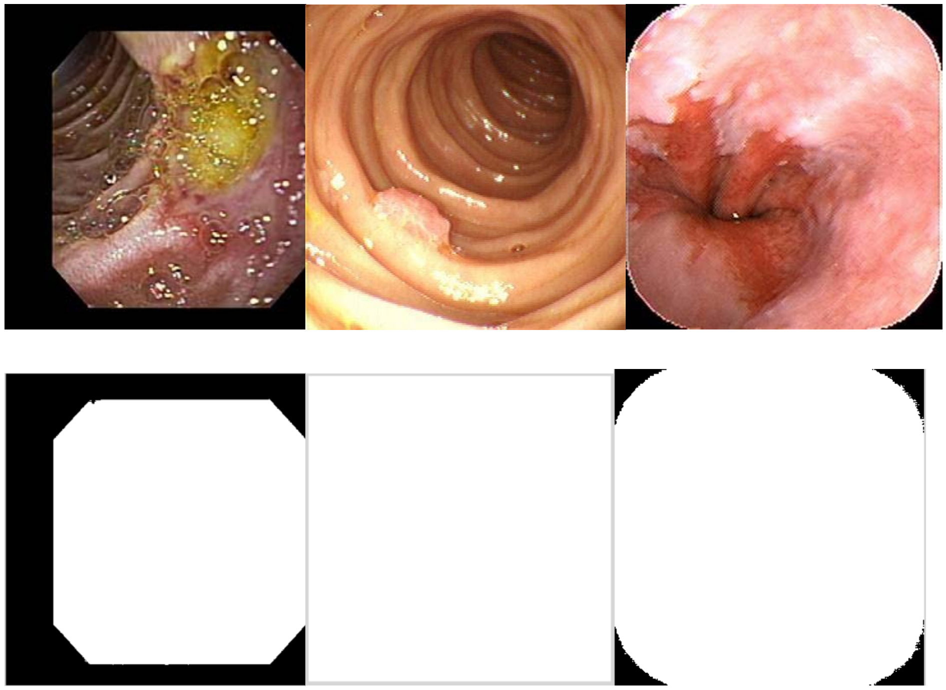

Figure 17 Effectiveness of frame detection algorithm- lower row contains mask for the dark frames. Only the white areas are used for feature extraction

Classification Results

In this section, we present the results for the various eight-class classifiers that were described previously. The goal was to annotate images with class labels consisting of the finding and the location. Results provided are for a 10-fold cross validation where the data set was divided into ten random sets, nine of which are used as the training set and the last used as the ten set. This is repeated ten times. A nearest-neighbor classifier was used for the different sets of features as shown in table 1 below.

Table 1 Results for nearest neighbor classifier

Feature	Feature Size	N	Classification Rate	Kappa
Grey scale histogram	256	5	24.8%	0.12
DCT	100	5	25.3%	0.14
Gist	512	5	36.1%	0.17
GLCM	20	5	33.4%	0.16
GLCM2	100	5	36.2%	0.18

Results for a neural network classifier using some of the same feature vectors as shown above is given below in table 2. There is a substantial improvement in classification accuracy with the more sophisticated classifier. However, kappa does not always reflect the same level of improvement. This is primarily due to the uneven distribution of the class membership sizes. By more frequently classifying larger number of images as belonging to the most populous classes, the hit rate improves. But since kappa corrects for likelihood due to chance, it does not see a similar improvement.

Table 2 Classification results for a multi-layer perceptron classifier

Feature	Feature Size	Classification Rate	Kappa
Grey scale histogram	256	34.2%	0.13
DCT	100	29.6%	0.14
Gist	512	44.3%	0.18
GLCM	20	42.6%	0.17
GLCM2	100	43.8%	0.19

Table 3 Classification results for textons, NBNN and our hybrid algorithm

Feature	Classification Rate	Kappa
Textons (Schmidt filters)	38.1%	0.20
Textons (image patch)	42.4%	0.25
Textons +RCA	44.3%	0.27
NBNN	47.6%	0.3
Hybrid	51.8%	0.35

These results demonstrate that local features including textons are more effective that global features like histograms. We found, similar to the work performed by Varma and Zisserman [82] that patch-based textons are more effective than the filter-based textons. Use of distance metric learning algorithms like RCA improves the classification performance by conveying the user's notion of similarity to the system. The Naive-Bayes Nearest Neighbor classifier performed better than the filter-based or patch-based textons. Our hybrid algorithm that used Affinity propagation for identifying the class exemplars, patch-based descriptors and the NBNN classifier showed a lot of promise, producing the best classification results.

Thus, we have demonstrated some success with the ability of our system to provide class labels to endoscopic images. This can be quite useful in either classifying unknown images or validating the labels of images for which the labels are extracted from noisy textual data.

Retrieval results

We created a set of simple queries for each finding similar to those used in ImageCLEF [9,63]. An example query was: "Show me endoscopic images of polyps". In order to evaluate the performance of the system in the absence of real users, we created "qrels" using the class information. Since this was a well-curated collection, we did not perform evaluate the text-based retrieval using the provided annotation as it is expected to be close to perfect. However, we evaluated the system using annotation-based retrieval where images were tagged with the purported finding based on our hybrid algorithm. This was to simulate situations where the class labels are not known on a larger test set but are available on a small training set. Our overall mean average precision (MAP) was 0.14 and our P30 was 0.31. Although these numbers seem low compared to text-based methods, they are quite comparable to the visual runs submitted by most participants at ImageCLEF in 2007 and 2008. Additionally, by combining textual and visual methods, the precision of the system can be improved substantially as seen in [11]. We also evaluated the use of our modified constrained affinity propagation-based relevance feedback. This algorithm uses the user-provided relevances as pair-wise constraints to update the distance matrices. Not surpsingly, our precision improved by the use of relevance feedback. Our MAP improved to 0.16 and our P30 improved to 0.42.

CHAPTER 5

Conclusions and Future Plans

We have created a web-based multimodal image retrieval system written using the Ruby on Rails framework. Ferret, a ruby port of Lucene was used for the text indexing. Our database also contains a number of image-based features as described in previous chapters. This system indexes the text-based annotations associated with the images, uses image processing to perform supervised automatic annotation based on available class labels, and facilitates interactive searching by allowing users to provide relevance feedback. We have also implemented an open source content-based image search engine (FIRE) created by Deselaers et al. [21].

Although CBIR has great potential in patient care, research and education, purely content-based image retrieval can be quite challenging for clinical purposes due to the semantic gap. Low level global features like color and texture may not be sufficient for classification of findings. In order to achieve better performance, segmentation and detection of abnormalities might be necessary. Retrieval, unlike classification, can be somewhat ambiguous and user and specific need dependent. However, combining visual and textual information can greatly improve retrieval performance. The use of distance metric learning and relevance feedback can help the system produce results that are more relevant to the user.

Future plans

Although our system has some basic image processing and machine learning tools written in ruby, most of our development has been in MATLAB and C++. However, we would like to port the algorithms written in MATLAB to C++ and create Ruby wrappers that will enable these functions to be called from within our web-based environment. We plan on release these as well as the entire system architecture as an open-source system.

User Evaluation with Larger Collection

We now have access to 250,000 images from the CORI collection. We plan to create a robust web-based system to retrieve these images using text and image-based techniques. We hope to recruit users from OHSU's Gastroenterology Department to evaluate the system. We are particularly interested in the area of learning to rank based on pairwise-constraints. Affinity propagation using spectral clustering will be used to evaluate if the retrieval performance can be significantly improved using a small amount of relevance feedback.

Metrics Evaluation

Information retrieval has given us popular metrics to evaluate search engines. These include precision, recall and mean average precision (MAP). However, there has been little work in medical image retrieval that demonstrates which, if any, of these measures is most meaningful to clinicians and students. It is also likely that these measures should be evaluated within the context of the search task. Identifying suitable measures that take into account the role of the searcher and their task, and that correlate to user satisfaction with their search can be immensely useful to builders of clinical image retrieval systems. We would like to assess the correlation between user satisfaction with the system will be measured and standard information retrieval measures, taking into account the role of the searcher and their task.

References

1. Tagare HD, Jaffe CC, Duncan J. Medical image databases: A content-based retrieval approach. J Am Med Inform Assoc 1997;4(3):184-98.

2. Müller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. International Journal of Medical Informatics 2004;73(1):1-23.

3. Muller H, Despont-Gros C, Hersh W, Jensen J, Lovis C, Geissbuhler A. Health care professionals' image use and search behaviour. Proceedings of Medical Informatics Europe (MIE 2006), Maastricht, Netherlands 2006:24-32.

4. Hersh, W. A qualitative task analysis of biomedical image use and retrieval; Imageclef/MUSCLE workshop on image retrieval evaluation, Vienna, Austria. 2005.

5. Lieberman DA, Holub J, Eisen G, Kraemer D, Morris CD. Prevalence of polyps greater than 9 mm in a consortium of diverse clinical practice settings in the united states. Clinical Gastroenterology and Hepatology 2005;3(8):798-805.

6. Orel SG, Kay N, Reynolds C, Sullivan DC. Radiology 1999;211(3):845-50.

7. Freer TW, Ulissey MJ. Radiology 2001;220(3):781-6.

8. Kapoor V, McCook BM, Torok FS. Radiographics 2004;24(2):523-43.

9. Hersh WR, Müller H, Jensen JR, Yang J, Gorman PN, Ruch P. Advancing biomedical image retrieval: Development and analysis of a test collection. J Am Med Inform Assoc 2006;13(5):488-96.

10. Smeulders AWM, Worring M, Santini S, Gupta A, Jain R. Content-Based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence 2000;22, 1349-1380.

11. Kalpathy-Cramer J, Hersh W. Automatic image modality based classification and annotation to improve medical image retrieval. Studies in Health Technology and Informatics 2007;129(2):1334.

12. Müller H, Deselaers T, Kim E, Kalpathy–Cramer J, Deserno TM, Hersh W, Overview of the ImageCLEFmed 2007 medical retrieval and medical annotation tasks; Advances in multilingual and multimodal information retrieval: 8Th workshop of the cross-language evaluation forum, clef 2007, Budapest, Hungary, September 19-21, 2007, Springer; 2008.

13. Logan JR, McCashland TM, Lieberman DA. Evaluation of reliability in the use of endoscopic terminology. Stud Health Technol Inform 2004;107(Pt 1):396-400.

14. Ameling S, Wirth S, Paulus D. Methods for polyp detection in colonoscopy videos: A review. University of Koblenz-Landau 2008;14:2008.

15. Kahn Jr CE, Thao C. Goldminer: A radiology image search engine. American Journal of Roentgenology 2007;188(6):1475.

16. Veltkamp RC, Tanase M, Sent D. Features in content-based image retrieval systems: A survey. State-Of-The-Art in Content-Based Image and Video Retrieval 1999:97-124.

17. Rui Y, Huang TS, Chang SF. Image retrieval: Current techniques, promising

directions, and open issues. Journal of Visual Communication and Image Representation 1999;10(1):39-62.

18. Datta R, Joshi D, Li J, Wang JZ. Image retrieval: Ideas, influences, and trends of the new age. 2008.

19. Flickner M, Sawhney H, Niblack W, Ashley J, Huang Q, Dom B, et al. Query by image and video content: The QBIC system. Computer 1995;28(9):23-32.

20. Tamura H, Mori S, Yamawaki T. Textural features corresponding to visual perception. IEEE Transactions on Systems, Man and Cybernetics 1978;8(6):460-73.

21. Deselaers T, Weyand T, Keysers D, Macherey W, Ney H. FIRE in Imageclef 2005: Combining content-based image retrieval with textual information retrieval. LNCS 2006;4022:652.

22. Carson C, Thomas M, Belongie S, Hellerstein JM, Malik J. Blobworld: A system for region-based image indexing and retrieval. LNCS 1999:509-16.

23. Ma WY, Manjunath BS. Netra: A toolbox for navigating large image databases. Multimedia Systems 1999;7(3):184-98.

24. Zhang, D.S., Wong, A., Indrawan, M., Lu, G , Content-Based image retrieval using gabor texture features; Proc. IEEE Pacific-Rim Conference on Multimedia. 2000.

25. Zhao R, Grosky WI. Narrowing the semantic gap-improved text-based web document retrieval using visual features. IEEE Transactions on Multimedia 2002;4(2):189-200.

26. Zhao R, Grosky WI. Negotiating the semantic gap: From feature maps to semantic landscapes. Pattern Recognition 2002;35(3):593-600.

27. Hare JS, Lewis PH, Enser PGB, Sandom CJ. Mind the gap: Another look at the problem of the semantic gap in image retrieval. 2006.

28. Yavlinksy A, Heesch D, An online system for gathering image similarity judgements; Proceedings of the 15th international Conference on Multimedia. 2007;565-568.

29. Jin R, Chai JY, Si L, Effective automatic image annotation via a coherent language model and active learning; Proceedings of the 12th annual ACM international conference on multimedia. 2004; 892—899, New York, NY, USA

30. Jeon J , Lavrenko V , Manmatha R, Automatic image annotation and retrieval using cross-media relevance models; Proceedings of the 26th annual international ACM SIGIR conference on research and development in Information Retrieval. 2003.

31. Aisen AM, Broderick LS, Winer-Muram H, Brodley CE, Kak AC, Pavlopoulou C, et al. Automated storage and retrieval of thin-section CT images to assist diagnosis: System description and preliminary assessment. Radiology 2003, Jul;228(1):265-70.

32. Antani S, Long LR, Thoma GR, Content-Based image retrieval for large biomedical image archives; Proceedings of 11th world congress on medical informatics (MEDINFO). 2004.

33. Long LR, Antani SK, Thoma GR. Image informatics at a national research center. Computerized Medical Imaging and Graphics 2005;29(2-3):171-93.

34. Lehmann TM, Gold MO, Thies C, Fischer B, Spitzer K, Keysers D, et al. Content-Based image retrieval in medical applications. Methods of Information in Medicine 2004;43(4):354-61.

35. Muller H, Rosset A, Vallee JP, Geissbuhler A. Integrating content-based visual access methods into a medical case database. Studies in Health Technology and Informatics 2003:480-5.

36. Muller H, Geissbuhler A, Marty J, Lovis C, Ruch P. The use of MedGIFT and EasyIR for ImageCLEF 2005. Lecture Notes in Computer Science 2006;4022:724.

37. Hsu W, Long LR, Antani S, SPIRS: A framework for content-based image retrieval from large biomedical databases; Medinfo 2007: Proceedings of the 12th world congress on health (medical) informatics: Building sustainable health systems. 2007.

38. Hsu W, Antani S, Long LR, Neve L, Thoma GR. SPIRS: A web-based image retrieval system for large biomedical databases. International Journal of Medical Informatics 2009;78:13-24.

39. Lam MO, Disney T, Raicu DS, Furst J, Channin DS. BRISC—an open source pulmonary nodule image retrieval framework. Journal of Digital Imaging 2007;20:63-71.

40. Kodogiannis V, Chowdrey HS. Multi network classification scheme for

57. Puzicha J, Hofmann T, Buhmann JM, Non-Parametric similarity measures for unsupervised texture segmentation and image retrieval; IEEE CVPR. 1997.

58. Squire DM, Learning a similarity-based distance measure for image database organization from human partitionings of an image set; Proc. Of the 4th IEEE workshop on applications of computer vision (WACV'98). Princeton, NJ, USA

59. Ma WY , Manjunath BS, T exture features and learning similarity; 1996 IEEE computer society conference on computer vision and pattern recognition, 1996. Proceedings CVPR'96. 1996.

60. Learning a similarity metric discriminatively, with application to face verification; IEEE computer society conference on computer vision and pattern recognition, 2005. CVPR 2005. 2005.

61. Lim JH, Wu JK, Singh S, Narasimhalu D. Learning similarity matching in multimedia content-based retrieval. IEEE Transactions on Knowledge and Data Engineering 2001:846-50.

62. Westerveld T, de Vries AP, Experimental result analysis for a generative probabilistic image retrieval model; Proceedings of the 26th annual international ACM SIGIR conference on research and development in information retrieval. 2003.

63. Muller H, Deselaers T, Kim E, Kalpathy--Cramer J, Deserno TM, Hersh W. Overview of the imageclef 2007 medical retrieval and annotation tasks. Cross-Language Retrieval in Image Collections (Imageclef) 2007.

computer-aided diagnosis in clinical endoscopy. 2004.

41. Kodogiannis VS , Computer-Aided diagnosis in clinical endoscopy using neuro-fuzzy systems; 2004 IEEE international conference on Fuzzy Systems, 2004. Proceedings. 2004.

42. VS Kodogiannis, M Boulougoura, Neural network-based approach for the classification of wireless-capsule endoscopic images; 2005 IEEE international joint conference on neural networks, 2005. IJCNN'05. Proceedings. 2005.

43. Karkanis SA, Iakovidis DK, Maroulis DE, Karras DA, Tzivras M. Computer-Aided tumor detection in endoscopic video using color wavelet features. IEEE Transactions on Information Technology in Biomedicine 2003;7(3):141-52.

44. Tjoa MP, Kirshnan SM, Kugean C, Wang P, Doraiswami R. Segmentation of Clinical Endoscopic Image Based on Homogeneity and Hue 2001.

45. Zheng MM, Krishnan SM, Tjoa MP. A fusion-based clinical decision support for disease diagnosis from endoscopic images. Computers in Biology and Medicine 2005;35(3):259-74.

46. Marchionini G. Exploratory search: From finding to understanding. Communications of the ACM 2006;49(4):41-6.

47. Marchionini G. Interfaces for end-user information seeking. Journal of the American Society for Information Science 1992;43(2):156-63.

48. Hersh W, Pentecost J, Hickam D. A task-oriented approach to information retrieval evaluation. Journal of the American Society for Information Science

1996;47(1):50-6.

49. Buckley C, Salton G, Optimization of relevance feedback weights; Proceedings of the 18th annual international ACM SIGIR conference on research and development in information retrieval. 1995.

50. Salton G, Buckley C. Improving retrieval performance by relevance feedback. Readings in Information Retrieval 1997:355-64.

51. Wu H, Salton G. The estimation of term relevance weights using relevance feedback. Journal of Documentation 1981;37(4):194-214.

52. Zhou XS, Huang TS. Relevance feedback in image retrieval: A comprehensive review. Multimedia Systems 2003;8(6):536-44.

53. Müller H, Pun T, Squire D. Learning from user behavior in image retrieval: Application of market basket analysis. International Journal of Computer Vision 2004;56(1):65-77.

54. Gudivada VN, Raghavan VV. Design and evaluation of algorithms for image retrieval by spatial similarity. ACM Transactions on Information Systems (TOIS) 1995;13(2):115-44.

55. Guo GD, Jain AK, Ma WY, Zhang HJ. Learning similarity measure for natural image retrieval with relevance feedback. IEEE Transactions on Neural Networks 2002;13(4):811-20.

56. Rubner Y, Tomasi C, Guibas LJ. The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision 2000;40(2):99-121.

64. Kalpathy-CramerJ, Hersh W, Medical image retrieval and automatic annotation: OHSU at ImageCLEF 2007; Advances in Multilingual and Multimodal Information Retrieval, LNCS 5152, Springer; 2008.

65. Deselaers T, Müller H, Clough P, Ney H, Lehmann TM. The CLEF 2005 automatic medical image annotation task. International Journal of Computer Vision 2007;74(1):51-8.

66. Voorhees EM, Harman DK. TREC: Experiment and evaluation in information retrieval. MIT Press; 2005.

67. Buckley C, Voorhees E, Retrieval evaluation with incomplete information; Proceedings of the 27th annual international ACM SIGIR conference on research and development in information retrieval. 2004; 25-32, Sheffield, UK.

68. Singhal A. Modern information retrieval: A brief overview. IEEE Data Engineering Bulletin 2001;24(4):35-43.

69. Hersh WR. Information retrieval: A health and biomedical perspective. Springer Verlag; 2008.

70. BJ Jansen, A Spink, J Bateman, T Saracevic , Real life information retrieval: A study of user queries on the web; ACM SIGIR forum. 1998; 5--17, New York, NY, USA

71. Wolfram D, Spink A, Jansen BJ, Saracevic T. Vox populi: The public searching of the web. Journal of the American Society for Information Science and Technology, 2001, 52(12), 1073-1074.

72. Mobasheri A, Airley R, Foster CS, Schulze-Tanzil G, Shakibaei M. Post-Genomic applications of tissue microarrays: Basic research, prognostic oncology, clinical genomics and drug discovery. Histology and Histopathology 2004;19(1):325-35.

73. Cohen AM, Hersh WR, Peterson K, Yen PY. Journal of the American Medical Informatics Association 2006;13(2):206-19.

74. Ben-David A. What's wrong with hit ratio? IEEE Intelligent Systems 2006;21(6):68-70.

75. Ozertem U, Erdogmus D, OGI B. Nonparametric snakes. IEEE Transactions on Image Processing 2007;16(9):2361-8.

76. Deselaers T, Keysers D, Ney H. Features for image retrieval: A quantitative comparison. Lecture Notes in Computer Science 2004:228-36.

77. Haralick RM. Statistical and structural approaches to texture. Proceedings of the IEEE 1979;67(5):786-804.

78. Aksoy S , Haralick RM, Textural features for image database retrieval; IEEE workshop on content-based access of image and video libraries, 1998. Proceedings. 1998.

79. Oliva A. Gist of the scene. Neurobiology of Attention 2005:251-6.

80. Oliva A, Torralba A. Building the gist of a scene: The role of global image features in recognition. Progress in Brain Research 2006;155:23.

81. Torralba A, Fergus R, Weiss Y,Small codes and large image databases for

recognition; IEEE conference on computer vision and pattern recognition, 2008. CVPR 2008. 2008.

82. Varma M , Zisserman A,Texture classification: Are filter banks necessary? 2003 IEEE computer society conference on computer vision and pattern recognition, 2003. Proceedings. 2003.

83. Leung T, Malik J. Representing and recognizing the visual appearance of materials using three-dimensional textons. International Journal of Computer Vision 2001;43(1):29-44.

84. Chantler M, Schmidt M, Petrou M, McGunnigle G. The effect of illuminant rotation on texture filters: Lissajous's ellipses. Lecture Notes in Computer Science 2002:289-303.

85. Chantler M, Petrou M, Penirsche A, Schmidt M, McGunnigle G. Classifying surface texture while simultaneously estimating illumination direction. International Journal of Computer Vision 2005;62(1):83-96.

86. Boiman O, Shechtman E, Irani M, In defense of nearest-neighbor based image classification; IEEE conference on computer vision and pattern recognition, 2008. CVPR 2008. 2008.

87. Nabney I. NETLAB: Algorithms for pattern recognition. Springer Verlag; 2002.

88. Frey BJ, Dueck D. Clustering by passing messages between data points. Science 2007;315(5814):972.

89. Cha SH, Srihari SN. On measuring the distance between histograms. Pattern Recognition 2002;35(6):1355-70.

90. Lu Z, Constrained spectral clustering through affinity propagation; IEEE conference on computer vision and pattern recognition, 2008. CVPR 2008. 2008.

91. Bar-Hillel A, Hertz T, Shental N, Weinshall D, Learning distance functions using equivalence relations; Proc. ICML 2003, Washington DC