

Text- and Content-based Approaches to Image Retrieval for the ImageCLEF 2009 Medical Retrieval Track

Matthew Simpson, Md Mahmudur Rahman,
Dina Demner-Fushman, Sameer Antani, George R. Thoma
Lister Hill National Center for Biomedical Communications
National Library of Medicine, NIH, Bethesda, MD, USA
{simpsonmatt, rahmanmm, ddemner, santani}@mail.nih.gov

Abstract

This article describes the participation of the Image and Text Integration (ITI) group from the United States National Library of Medicine (NLM) in the ImageCLEF 2009 medical retrieval track. Our methods encompass a variety of techniques relating to document summarization and text- and content-based image retrieval. Our text-based approach utilizes the Unified Medical Language System (UMLS) synonymy of concepts identified in information requests and image-related text to retrieve semantically relevant images. Our content-based approaches utilize similarity metrics based on computed “visual concepts” to identify visually similar images. In this article we present an overview of these approaches, discuss our experiences combining them into multimodal retrieval strategies, and describe our submitted runs and results.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.7 Digital Libraries; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—*Object Recognition*

General Terms

Measurement, Performance, Experimentation

Keywords

Image Retrieval, CBIR, Medical Imaging, Ontologies, UMLS

1 Introduction

This article describes the participation of the Image and Text Integration (ITI) group from the United States National Library of Medicine (NLM) in the ImageCLEF 2009 medical retrieval track. This is our second year participating in ImageCLEFmed.

ImgeCLEFmed’09 [17] consists of two medical retrieval tasks. In the first task, a set of ad-hoc information requests are given, and the goal is to retrieve the most relevant images pertaining to each topic. In the second task, a set of case-based information requests are given, and the goal is to retrieve the most relevant articles describing case studies pertaining to the topic case.

In the following sections, we describe our text-based approach (Section 2), which is suitable for both retrieval tasks, and several content-based approaches (Section 3) to the ad-hoc retrieval task. Our text-based approach relies on mapping information requests and image-related text to concepts in the Unified Medical Language System (UMLS) [13] Metathesaurus, and our content-based approaches analogously rely on mapping medical images to “visual concepts” using machine learning techniques.

In Section 4, we suggest strategies for combining our text- and content-based approaches, describe our submitted runs, and present their results. For the ad-hoc retrieval task, our best run, a multimodal feedback approach, achieved a Mean Average Precision (MAP) of 0.38, and our best automatic run, a text-based approach, achieved a MAP of 0.35. For the case-based retrieval task our automatic text-based approach achieved a MAP of 0.34 and was ranked 1st among all case-based run submissions.

2 Text-based Image Retrieval

In this section we describe our text-based approach to image retrieval. Effective text-based medical image retrieval requires (1) a document representation that contains the most pertinent information describing the content of the image and potential information needs and (2) a retrieval strategy that is appropriate for the biomedical domain.

Our document representation consists of several automatically extracted search areas in addition to the image captions provided in the ImageCLEFmed’09 [17] collection. These fields include the title of the article in which the image appears, the article’s abstract, a brief mention (one sentence) of the image from the article’s full text, and the Medical Subject Headings (MeSH terms) assigned to the article. MeSH is a controlled vocabulary created by NLM to index biomedical articles. We provide a summary of each caption according to a structured representation of information needs that are relevant to the principles of evidence-based practise [7]. This search area includes automatically extracted fields relating to anatomy, diagnosis, population group, etc.

We use the Essie [12] search engine to index this collection of image documents and retrieve relevant images. Essie was originally developed by NLM to support the online registry of clinical research studies at ClinicalTrials.gov [15], and now it serves several other information retrieval systems at NLM. Key features of Essie that make it particularly well-suited to the medical retrieval track include its automatic expansion of query terms along synonymy relationships in the UMLS Metathesaurus and its ability to weight term occurrences according the location of the document in which they occur. For example, term occurrences in an image caption can be given a higher weight than occurrences in the abstract of the article in which the image appears. Essie also expands query terms to include morphological variants derived from the UMLS SPECIALIST Lexicon instead of stemming.

To construct queries for each information request, we map topics to the UMLS using the MetaMap [1] tool and represent terms relating to image modality, clinical findings, and anatomy with their preferred UMLS names. Thus, each query consists of the conjunction of a set of UMLS concepts that are expanded by Essie during the retrieval process. For extracted modality terms that cannot be mapped to the UMLS, we perform an automatic term expansion based on a list of image modalities (originally created by the authors using RadLex¹ as a starting point[6]) which we expanded using the UMLS synonymy and manually augmented with missing terms (mostly abbreviations) based on the authors’ experience creating the ITI modalities hierarchy.

2.1 Case-based Retrieval Task

Our retrieval strategy for the case-based retrieval task is identical to that of the ad-hoc task. However, since the retrieval unit of the case-based task is an entire article, to construct an appropriate document representation we perform a simple union of all the search areas for each image in the

¹<http://radlex.org/viewer>.

article. That is, a case-based document consists of a title, abstract, MeSH terms, and the caption, mention, and structured caption summary of each image contained in the article.

3 Content-based Image Retrieval

In content-based image retrieval (CBIR), access to information is performed at a perceptual level based on automatically extracted low-level features (e.g., color, texture, shape, etc.) [19]. The performance of a CBIR system depends on the underlying image representation, usually in the form of a feature vector. Due to the limitations of the low-level features in CBIR and motivated by a learning paradigm, we explore classification at both the global collection level and the local individual image level in our submitted runs for ImageCLEFmed’09 [17]. In addition to the off-line supervised learning approach, we incorporate users’ semantic perceptions interactively in the retrieval loop based on relevance feedback (RF) information. The following sections describe our feature representation schemes and the retrieval methods applied to the various visual and multimodal submitted runs.

3.1 Image Feature Representation

To generate the feature vectors at different levels of abstraction, we extract both visual concept-based feature based on a “bag of concepts” model comprising color and texture patches from local image regions [21] and various low-level global features including color, edge, and texture.

3.1.1 Visual Concept-based Image Representation

In the ImageCLEFmed’09 collection [18], it is possible to identify specific local patches in images that are perceptually and/or semantically distinguishable, such as homogeneous texture patterns in grey level radiological images and varying color and texture structures in microscopic pathology and dermoscopic images. The content of these local patches can be effectively modeled as “visual concepts” [21] by using supervised learning based classification techniques such as the Support Vector Machine (SVM).

For concept model generation, we utilize a voting-based multi-class SVM known as *one-against-one* or pairwise coupling (PWC) [11]. In developing training samples for this SVM, only local image patches that map to visual concept models are used. To accurately automatically segment and unambiguously and consistently label image segments, a fixed-partition based approach is used to divide the entire image space into an $(r \times r)$ grid of non-overlapping regions. Manual selection is applied to limit such patches in the training set to those that have a majority of their area (80%) covered by a single concept. In order to train the SVMs based on the local concept categories, a set of L labels are assigned as $C = \{c_1, \dots, c_i, \dots, c_L\}$, where each $c_i \in C$ characterizes a local concept category. The training set of the local patches that comprise color and texture moment-based features, is annotated manually with the concept labels in a mutually exclusive way. Images in the data set are annotated with local concept labels by partitioning each image I_j into an equivalent $r \times r$ grid of l region vectors as $\{\mathbf{x}_{1_j}, \dots, \mathbf{x}_{k_j}, \dots, \mathbf{x}_{l_j}\}$, where each $\mathbf{x}_{k_j} \in \mathbb{R}^d$ is a d -dimensional combined color and texture feature vector. For each \mathbf{x}_{k_j} , the local concept category probabilities are determined by the prediction of the multi-class SVMs:

$$p_{ik_j} = P(y = i \mid \mathbf{x}_{k_j}), \quad 1 \leq i \leq L. \quad (1)$$

The category label of x_{k_j} is determined by the maximum probability score. Thus, the entire image is represented as a two-dimensional index linked to the concept labels assigned for each region. Based on this encoding scheme, an image I_j can be represented as a vector in a local concept space as

$$\mathbf{f}_j^{\text{Concept}} = [f_{1_j}, \dots, f_{i_j}, \dots, f_{L_j}]^T \quad (2)$$

where each f_{i_j} corresponds to the normalized frequency of a concept $c_i, 1 \leq i \leq L$ in image I_j . The feature vector $\mathbf{f}^{\text{Concept}}$ is viewed as a local concept distribution from a probabilistic viewpoint.

According to the notion of total probability [9], an element f_{i_j} can be defined as

$$f_{i_j} = \sum_{k_j=1}^l P_{i|k_j} P_k = \frac{1}{l} \sum_{k_j=1}^l P_{i|k_j} \quad (3)$$

where P_k is the probability of a region selected from image I_j being the k_j th region, which is $1/l$, and $P_{i|k_j}$ is the conditional probability that the selected k_j th region in I_j maps to the concept c_i . In the context of the concept vector $\mathbf{f}_j^{\text{concept}}$, the value of $P_{i|k_j}$ is 1 if the region k_j is mapped to the c_i concept, or 0 otherwise. Due to the crisp membership value, this feature representation is sensitive to quantization errors. However, based on the probabilistic values of each region, an image I_j is represented as $\mathbf{f}_j^{\text{PVCV}} = [\hat{f}_{1_j} \cdots \hat{f}_{i_j} \cdots \hat{f}_{L_j}]^T$, where

$$\hat{f}_{i_j} = \sum_{k=1}^l p_{ik_j} P_k = \frac{1}{l} \sum_{k=1}^l p_{ik_j}; \quad \text{for } i = 1, 2, \dots, L \quad (4)$$

where p_{ik_j} is determined based on (1). In contrast to the simple concept vector $\mathbf{f}^{\text{concept}}$, this vector representation considers not only the similarity of different region vectors from different concepts but also the dissimilarity of those region vectors mapped to the same concepts.

3.1.2 Low Level Global Feature Representation

In addition to the visual concepts of local image patches, we extract the following global features:

- **Color Feature:** To represent the spatial structure of images, we utilize the Color Layout Descriptor (CLD) of MPEG-7 [3]. The CLD represents the spatial layout of the images in a compact form. It is obtained by applying the discrete cosine transformation (DCT) on the 2D array of local representative colors in the $YCbCr$ color space, where Y is the luma component and Cb and Cr are the blue and red chroma components. Each channel is represented by 8 bits and each of the 3 channels is averaged separately for the 8×8 image blocks. We extract a CLD with 10 Y , 3 Cb , and 3 Cr to form a 16-dimensional feature vector.

Images may also be represented as Color Coherence Vector (CCV) [20], where a particular color’s coherence is defined as the degree to which pixels of that color are members of large similarly-colored regions. A CCV stores the number of coherent versus incoherent pixels with each color. By separating coherent pixels from incoherent pixels, CCV’s provide finer distinctions than color histograms.

- **Edge Feature:** To represent the global edge feature, the spatial distribution of edges are utilized by the Edge Histogram Descriptor (EHD) [3]. The EHD represents local edge distribution in an image by dividing the image into 4×4 sub-images and generating a histogram from the edges present in each of these sub-images. Edges in the image are categorized into five types—namely vertical, horizontal, 45° diagonal, 135° diagonal and non-directional edges. Finally, a histogram with $16 \times 5 = 80$ bins is obtained, corresponding to a 80-dimensional feature vector.

In addition, a histogram of edge direction is constructed, where the edge information contained in the images is processed and generated by using the Canny edge detection algorithm (with $\sigma = 1$, Gaussian masks of size = 9, low threshold = 1, and high threshold = 255). The corresponding edge directions are quantized into 72 bins of 5° each. Scale invariance is achieved by normalizing the histograms with respect to the number of edge points in the image.

- **Texture Feature:** We extract texture features from the grey level co-occurrence matrix (GLCM) [10] of each image. In order to obtain efficient descriptors, the information contained in GLCM is traditionally condensed into a few statistical features. Four GLCM’s for four different orientations (horizontal 0° , vertical 90° , and two diagonals— 45° and 135°)

are obtained and normalized to the entries [0,1] by dividing each entry by total number of pixels. Higher order features, such as energy, entropy, contrast, homogeneity and maximum probability are measured based on averaging features in GLCMs to form a 20-dimensional feature vector for an entire image.

- **Average Grey Level Feature:**

For different categories or within the same category, images in a collection may vary in size and undergo translations. Resizing them into a thumbnail of a fixed size can reduce the translational error and some of the noise due to the artifacts present in the images, especially for images in medical domain. Hence, a feature extraction is performed from the low-resolution scaled images where each image is converted to a gray-level image (one channel only) and scaled down to the size 64×64 regardless of the original aspect ratio. Next, the down-scaled image is partitioned further with a 16×16 grid to form small blocks of (4×4) pixels. The average gray value of each block is measured and concatenated to form a 256-dimensional feature vector.

- **Other Features:** We extract two additional features using the Lucene image retrieval (LIRE) library [14] including the Color Edge Direction Descriptor (CEDD) and the Fuzzy Color Texture Histogram (FCTH). CEDD incorporates color and texture information into one single histogram and requires low computational power compared to MPEG-7 descriptors. To extract texture information, CEDD uses a fuzzy version of the five digital filters proposed by the MPEG-7 EHD, which forms 6 texture areas [4]. This descriptor is appropriate for retrieving images even in cases with deformation, noise and smoothing. In contrast, FCTH uses the high frequency bands of the Haar Wavelet Transform in a fuzzy system to form 8 texture areas [5].

3.2 Fusion-based Image Similarity Matching

It is difficult to find a unique representation to compare images accurately for all types of queries. Feature descriptors at different levels of image representation are in diverse forms and are often complementary in nature. Data fusion, or multiple-evidence combination, describes a range of techniques where multiple pieces of information are combined to achieve improvements in retrieval effectiveness [8]. CBIR also adopts some of the ideas from data fusion, where the most commonly used approach is a linear combination of similarity matching scores of different features with pre-determined weights. In this framework, the similarity between a query image I_q and target image I_j is described as

$$\text{Sim}(I_q, I_j) = \sum_F \omega^F \text{Sim}^F(I_q, I_j) \quad (5)$$

where $F \in \{\text{Concept, EHD, CLD, CCV, CEDD, FCTH, etc.}\}$ and ω^F are the weights within the different image representations. We now present several linear combination schemes including ones based on the online category prediction of a query image and on relevance feedback.

3.2.1 Category-Specific Similarity Fusion

In this approach, the category of a query image at a global level is determined based on the SVM learning on a training set of 5000 images of 32 manually assigned and mutually exclusive categories from the ImageCLEFmed'05-07 collections [16]. Images are classified into three levels of detail as shown in Figure 1. For the SVM training, the radial basis function (RBF) is used and a 10-fold cross-validation is conducted to find the best tunable parameters C and γ of the RBF kernel. Only the best performing features are used in SVM classification. Our SVM implementation is based on the *LIBSVM* package [2].

Based on the online categorization of a query image, precomputed category-specific feature weights (e.g., ω^F) are subsequently utilized in the linear similarity matching function. Based on this scheme, for example, a color feature will have more weight for microscopic pathology

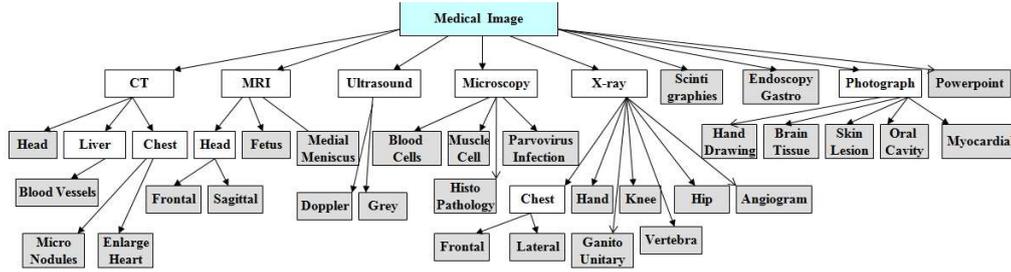


Figure 1: Classification structure of the training set.

and dermatology images, whereas edge- and texture-related features will have more weight for radiographs.

In addition, to find the optimal weights we consider the 10-fold cross validation accuracy of each feature. The accuracies are based on SVM classification of the images in the training set of 5000 images. The weights are normalized based on the accuracies of the features subject to $0 \leq \omega^F \leq 1$ and $\sum \omega^F = 1$ for $F \in \{\text{Concept, EHD, CLD, CCV, CEDD, FCTH, etc.}\}$.

3.2.2 Image Similarity Fusion Based on Relevance Feedback (RF)

We used a feedback-based similarity fusion technique, where feature weights are updated at each iteration by considering both the precision and the rank order of relevant images in individual result lists. As a result, the final rank-based retrieval is obtained through an adaptive and linear weighted combination of overall similarity, fusing individual level similarities.

In this approach, to update the feature weights (e.g., ω^F), we first perform similarity matching based on equal weighting of each feature. After this initial retrieval, a relevance judgement is manually provided for the top K returned images. We then measure the effectiveness of the top K images as

$$E = \frac{\sum_{i=1}^K \text{Rank}(i)}{K/2} \times P(K) \quad (6)$$

where $\text{Rank}(i) = 0$ if the image in rank position i is not relevant and $\text{Rank}(i) = (K - i)/(K - 1)$ for relevant images. Hence, the function $\text{Rank}(i)$ is monotonically decreasing from one (if the image at rank position one is relevant) down to zero (e.g., for a relevant image at rank position K). On the other hand, $P(K) = R_K/K$ is the precision at the top K , where R_k is the number of relevant images in the top K retrieved results. Hence, equation (6) is basically the product of two factors: rank order and precision. The rank order factor takes into account the position in the retrieval set of the relevant images, whereas the precision is a measure of the retrieval accuracy, regardless of the position. Generally, the rank order factor is heavily biased for the position in the ranked list over the total number of relevant images, and the precision value ignores the rank order of the images. To balance both criteria, we use a performance measure that is the product of the rank order factor and precision. If there is more overlap between the relevant images of a particular retrieval set and the one provided through feedback, the performance score will be higher. Both terms on the right side of equation (6) will be one if all the top K returned images are considered relevant. The raw performance scores obtained by the above procedure are then normalized by the total score as $\hat{E} = \omega^F$ to generate the updated feature weights respectively. For the next iteration of retrieval with the same query, these modified weights are utilized for the similarity matching function by

$$\text{Sim}(I_q, I_j) = \sum_F \hat{\omega}^F \text{Sim}^F(I_q, I_j) \quad (7)$$

This weight updating process might be continued as long as relevance judgements are available or until no changes are noticed due to the system convergence.

4 Submitted Runs and Results

This section provides descriptions and retrieval results of our submitted textual and visual runs as well as our attempts at integrating the text-based and CBIR-based approaches.

4.1 Ad-hoc Retrieval Task

We submitted the following 9 runs for the 25 ad-hoc topics [17]:

1. *ceb_essie2_automatic*: This is a textual run utilizing the approach described in Section 2. Based on our previous experience with the ImageCLEFmed'08 [18] collection, we weighted the caption and title search areas more heavily than the other areas.
2. *cbir_fusion_category*: This is a visual run based on the category-specific similarity fusion approach described in Section 3.2.1. For this run, we selected only one query image for each topic and considered all features for similarity fusion as described in Section 3.1. For each query, the category was determined based on SVM trained on 5000 images from ImageCLEFmed'05-07 collections [16]. The individual preassigned feature weights were selected based on the category-specific rules and utilized in the linear combination of similarity matching functions.
3. *cbir_fusion_merge*: This is a visual run similar to the above (*cbir_fusion_category*), but instead of only considering one image for each topic, we considered every query image for each topic and generated separate ranked lists for each retrieval result. For each topic, we took the top 500 retrieved images corresponding to each query image and merged them into a single ranked list for the topic.
4. *cbir_fusion_cv_merge*: This is a visual run similar to the above (*cbir_fusion_merge*), but instead of utilizing category-specific rules for feature weights, we found the optimal weights by considering the normalized cross validation accuracies of each feature as described in Section 3.2.1. We merged the top 500 retrieved images for each query image into a single ranked list as before.
5. *multimodal_text_qe_cbir*: This is a mixed run that combines the approaches described in *ceb_essie2_interactive* and *cbir_fusion_category*. For each topic, we first performed the textual search. We then manually selected 3-5 of the highest ranked retrieved images as relevant. Finally, we computed the mean vector of these retrieved images and used it as the query for the visual search.
6. *multimodal_text_rerank*: This is a mixed run that combines the approaches described in *ceb_essie2_interactive* and *cbir_fusion_category*. For each topic, we first performed the textual search and then re-ranked the retrieved images based on the scores of the visual search.
7. *ceb_interactive_with_pad*: This is a mixed run that interactively combines all of the above approaches (1-6) in a text-based relevance feedback approach. For each topic, we manually selected relevant images from the top ten retrieved images of each of the above approaches. We then selected additional query terms from the document representation of the relevant images (described in Section 2), and used this expanded query as the input to the textual search described in *ceb_essie2_automatic*. We ranked these additional retrieved images below the ones manually selected as relevant.
8. *text_manual_cbir_rf*: This is a mixed run similar to the approach described in *multimodal_text_qe_cbir*. However, instead of manually choosing 3-5 images from the textual retrieval results, we automatically selected the top 5 images from *ceb_interactive_with_pad*. We computed the mean vector of these images and used it as the input query to the approach described in *cbir_fusion_category*.

File Name	ID	Mode	Type	Recall	MAP	P@5
ITI_26_8_1244841659565.txt	ceb_interactive_with_pad	Mixed	Feedback	0.65	0.38	0.74
ITI_26_8_1243447590820.txt	ceb_essie2_automatic	Textual	Automatic	0.66	0.35	0.65
ITI_26_8_1244811028909.txt	multimodal_text_rerank	Mixed	Automatic	0.66	0.27	0.49
ITI_26_8_1244842970604.txt	text_manual_cbir_rf	Mixed	Feedback	0.21	0.04	0.28
ITI_26_8_1244811851777.txt	multimodal_text_qe_cbir	Mixed	Manual	0.19	0.04	0.27
ITI_26_8_1244813032166.txt	cbir_fusion_cv_merge	Visual	Automatic	0.12	0.01	0.09
ITI_26_8_1244813305029.txt	cbir_fusion_merge	Visual	Automatic	0.12	0.01	0.08
ITI_26_8_1244846828228.txt	cbir_rf	Visual	Feedback	0.13	0.01	0.06
ITI_26_8_1244812535094.txt	cbir_fusion_category	Visual	Automatic	0.13	0.01	0.06

Table 1: Results of the 9 Submitted Runs for the Ad-hoc Task

File Name	ID	Mode	Type	Recall	MAP	P@5
ITI_26_8_1243520633864.txt	ceb_cases_essie2_automatic	Textual	Automatic	0.78	0.34	0.32

Table 2: Results of the Submitted Run for the Case-based Task

- cbir_rf*: This is a visual feedback approach based on *cbir_fusion_category*. We manually selected 5 highly ranked images from the visual retrieval results as relevant. We then computed the mean vector of these retrieved images and used it as the query for another iteration of the visual search.

Table 1 presents the results of our submitted runs for the ad-hoc topics. *ceb_interactive_with_pad*, a multimodal relevance feedback approach, achieved the highest precision (MAP = 0.38) of our submitted runs. This run was ranked 1st among all submitted multimodal approaches and 1st among all feedback approaches, although its MAP is lower than some automatic runs submitted by other participating groups. The noticeable increase in Precision at 5 retrieved images (P@5) of *ceb_interactive_with_pad* is inherent in its retrieval strategy—the highest ranked images were manually selected from the top 10 retrieved images from 6 other approaches. *ceb_essie2_automatic* (MAP = 0.35) ranked 14th among automatic textual runs (ITI the 5th ranked group). *cbir_fusion_merge* (MAP = 0.01) ranked 1st among submitted visual runs although this result is likely not statistically significant. Finally, among multimodal automatic approaches, *multimodal_text_rerank* (MAP = 0.27) ranked 8th (ITI the 4th ranked group).

For our three best runs, we evaluated the statistical significance of the increase in precision using the two-sided Wilcoxon signed rank test. At the 0.05 significance level, the differences in precision between *ceb_interactive_with_pad* and *ceb_cases_essie2_automatic* and between *ceb_cases_essie2_automatic* and *multimodal_text_rerank* are not significant ($p = 0.059$ and $p = 0.057$, respectively), which is consistent with the null hypothesis of having the same mean. However, *ceb_interactive_with_pad* significantly improves the precision of *multimodal_text_rerank* ($p < 0.001$).

4.2 Case-based Retrieval Task

We submitted the following run for the 5 case-based topics [17]:

- ceb_cases_essie2_automatic*: This is a textual run based on the approach described in Section 2. We weighted the caption, title and anatomy search areas heavier than the other areas and favored articles indexed with MeSH terms indicative of case studies or clinical trials. Examples of such terms include “Case Reports,” “Case-Control Studies” and “Cross-Sectional Studies” among several others.

Table 2 presents the retrieval results of our submitted case-based run. *ceb_cases_essie2_automatic* achieved a MAP of 0.34 and was ranked 1st among all case-based submissions. *sinai_TA_cbt* (MAP = 0.26) was ranked 2nd.

5 Conclusion

This article describes the retrieval strategies of the ITI group for the ImageCLEF 2009 medical retrieval track. For the ad-hoc task, we submitted 9 runs that include various combinations of our text- and content-based approaches in different retrieval scenarios. We submitted one automatic textual run for the case-based task. Many of our submitted runs were successful—most notably our case-based run, which was ranked 1st among all case-based run submissions.

Our results indicate that content-based approaches to image retrieval are not yet advanced enough to achieve the precision of text-based approaches, and in many cases can reduce the precision of text-based approaches when combined in a multimodal automatic scheme. However, precision can be improved by combining text- and content-based approaches in relevance feedback retrieval scenarios.

Acknowledgments

The authors wish to thank Haiying Guan, Rodney L. Long and Zhiyun Xue for their valuable input during group discussions and time spent preparing relevance judgements for our interactive experiments.

References

- [1] A. R. Aronson. Effective mapping of biomedical text to the UMLS metathesaurus: The MetaMap program. In *Proc. of the Annual Symp. of the American Medical Informatics Association (AMIA)*, pages 17–21, 2001.
- [2] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [3] S.-F. Chang, T. Sikora, and A. Puri. Overview of the MPEG-7 standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):688–695, 2001.
- [4] S. A. Chatzichristofis and Y. S. Boutalis. CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. In A. Gasteratos, M. Vincze, and J. K. Tsotsos, editors, *Proceedings of the 6th International Conference on Computer Vision Systems*, volume 5008 of *Lecture Notes in Computer Science*, pages 312–322. Springer-Verlag Berlin Heidelberg, 2008.
- [5] S. A. Chatzichristofis and Y. S. Boutalis. FCTH: Fuzzy color and texture histogram: A low level feature for accurate image retrieval. In *Proceedings of the 9th International Workshop on Image Analysis for Multimedia Interactive Services*, pages 191–196, 2008.
- [6] D. Demner-Fushman, S. Antani, M. Simpson, and G. Thoma. Combining medical domain ontological knowledge and low-level image features for multimedia indexing. In *Proc. of the Language Resources for Content-Based Image Retrieval Workshop (OntoImage)*, pages 18–23, 2008.
- [7] D. Demner-Fushman and J. Lin. Answering clinical questions with knowledge-based and statistical techniques. *Computational Linguistics*, 33(1):63–103, 2007.
- [8] E. A. Fox and J. A. Shaw. Combination of multiple searches. In *Overview of the Second Text Retrieval Conference (TREC-2)*, pages 243–252, 1994.
- [9] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 2nd edition, 1990.

- [10] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621, 1973.
- [11] T. Hastie and R. Tibshirani. Classification by pairwise coupling. *The Annals of Statistics*, 26(2):451–471, 1998.
- [12] N. C. Ide, R. F. Loane, and D. Demner-Fushman. Essie: A concept-based search engine for structured biomedical text. *Journal of the American Medical Informatics Association*, 1(3):253–263, 2007.
- [13] D. Lindberg, B. Humphreys, and A. McCray. The unified medical language system. *Methods of Information in Medicine*, 32(4):281–291, 1993.
- [14] M. Lux and S. A. Chatzichristofis. LIRe: Lucene image retrieval—an extensible java CBIR library. In *Proceedings of the 16th ACM International Conference on Multimedia*, pages 1085–1088, 2008.
- [15] A. T. McCray and N. C. Ide. Design and implementation of a national clinical trials registry. *Journal of the American Medical Informatics Association*, 7(3):313–323, 2000.
- [16] H. Müller, T. Deselaers, E. Dim, J. Kalpathy-Cramer, T. M. Deserno, and W. Hersh. Overview of the imageCLEFmed 2007 medical retrieval and annotation tasks. In *Working Notes for the CLEF 2007 Workshop*, 2007.
- [17] H. Müller, J. Kalpathy-Cramer, I. Eggel, S. Bedrick, S. Radhouani, B. Bakke, C. E. Kahn, Jr., and W. Hersh. Overview of the CLEF 2009 medical image retrieval track. In *CLEF Working Notes 2009*, 2009.
- [18] H. Müller, J. Kalpathy-Cramer, C. E. Kahn, Jr., W. Hatt, S. Bedrick, and W. Hersh. Overview of the imageCLEFmed 2008 medical image retrieval task. In *Working Notes for the CLEF 2008 Workshop*, 2008.
- [19] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *International Journal of Medical Informatics*, 73(1):1–23, 2004.
- [20] G. Pass, R. Zabih, and J. Miller. Comparing images using color coherence vectors. In *Proceedings of the 4th ACM International Conference on Multimedia*, pages 65–73, 1996.
- [21] M. M. Rahman, S. Antani, and G. Thoma. A medical image retrieval framework in correlation enhanced visual concept feature space. In *Proceedings of the 22nd IEEE International Symposium on Computer-Based Medical Systems*, 2009.