Multimodal Medical Image Retrieval Improving Precision at ImageCLEF 2009

Saïd Radhouani, Jayashree Kalpathy-Cramer, Steven Bedrick, Brian Bakke, William Hersh

Department of Medical Informatics & Clinical Epidemiology
Oregon Health and Science University (OHSU)
Portland, OR, USA
radhouan@ohsu.edu

Abstract. We present results from Oregon Health & Science University's participation in the medical retrieval task of ImageCLEF 2009. This year, we focused on improving retrieval performance, especially early precision, in the task of solving medical multimodal queries. These queries contain visual data, given as a set of image-examples, and textual data, provided as a set of words belonging to three dimensions: Anatomy, Pathology, and Modality. To solve these queries, we use both textual and visual data in order to better interpret the semantic content of the queries. Indeed, using the textual data associated with the image, it is relatively easy to extract anatomy and pathology, but it is challenging to extract the modality, since this is not always explicitly described in the text. To overcome this problem, we utilized the visual data. We combined both text-based and visual-based search techniques to provide a unique ranked list of relevant documents for each query. The obtained results showed that our approach outperforms our baseline by 43% in MAP and 71% in precision at top 5 documents. This is due to the use of domain dimensions and the combination of both visual-based and text-based search techniques.

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.3.4 Systems and Software; H.3.7 Digital Libraries; H.2.3 [**Database Management**]: Languages—*Query Languages*

General Terms

Measurement, Performance, Experimentation

Keywords

Medical Image Retrieval, Performance Evaluation, Image Classification, Image Modality Extraction, Domain Dimensions

1   Introduction

Advances in digital imaging technologies and the increasing prevalence of Picture Archival and Communication Systems (PACS) have led to a substantial growth in the number of digital images stored in hospitals and medical systems in recent years. Medical images can form an essential component of a patient's health records, and the ability to retrieve them can be useful in several tasks, including diagnosis, education, research, and so on.

Image retrieval systems (IRS) do not currently perform as well as their text counterparts [1]. Medical and other IRS have historically relied on indexing annotations or captions associated with the images. The last few decades, however, have seen advancements in the area of content-based image retrieval (CBIR) [2, 3]. Although CBIR systems have demonstrated success in fairly constrained medical domains including pathology, dermatology, chest radiology, and mammography, they have demonstrated poor performance when applied to databases with a wide spectrum of imaging modalities, anatomies, and pathologies [1, 4–6].

In this paper, we address the problem of solving medical multimodal queries, by focusing especially on improving early precision (i.e., precision at top 5 documents). The queries we deal with in ImageCLEF 2009 contain visual data, given as a set of image-examples, and textual data, provided by a set of words belonging to the categories Anatomy, Pathology, and Modality. Using only a visual-based search technique, it is relatively feasible to identify the modality of a medical image, but it is very challenging to extract from an image the anatomy or the pathology (e.g., slight fracture of a bone). On the other hand, using a text-based search technique, it is relatively easy to extract the anatomy and the pathology from a text, but it is not obvious to identify the modality. Indeed, this latter is not always explicitly described in the text, since the writer might use general words, such as "this image ...," to write their medical report. To overcome these problems, retrieval performance, especially early precision, can be improved demonstrably by merging the results of textual and visual search techniques [7–9].

In the rest of this paper, we first present a brief description of our system (Section 2). Sections 3 and 4 are dedicated respectively to our visual-based search technique and text-based search technique. We describe the official runs in Section 5, and the corresponding results in Section 6. Finally, we conclude this paper and provide some perspectives (Section 7).

2   System Description of Our Adaptive Medical Image Retrieval System

Starting in 2007, we created and have continued to evolve a multimodal image retrieval system based on an open-source framework that allows the incorporation of user search preferences. We designed a flexible database schema that allows us to easily incorporate new collections while facilitating retrieval using both text and visual techniques. The 2009 ImageCLEF collection consists of 74,902

medical images and annotations associated with them [10]. This collection contains images and captions from *Radiology* and *Radiographics*, two Radiological Society of North America (RSNA) journals.

2.1 Database and Web Application

We used the Ruby programming language [1], with the open source Ruby On Rails[2] web application framework. The PostgreSQL[3] relational database was used to store mapping between the images and the various fields associated with the image. The title, full caption and precise caption, as provided in the data distribution, were indexed. The captions and titles in the collection are currently indexed and we continue to add indexable fields for incorporating visual information. The data distribution included an xml file with the image id, the captions of the images, the titles of the journal articles in which the image had appeared and the PubMed ID of the journal article. In addition, a compressed file containing the approximately 74,900 images was provided.

2.2 Query Parser and Search Engine

Our system presents a variety of search options to the user including Boolean OR, AND, and "exact match." There are also options to perform fuzzy searches, as well as a custom query parser. A critical aspect of our system is the query parser, written in Ruby. Ferret, a Ruby port of the popular Lucene system, was used in our system as the underlying search engine. The custom query parser performs stop-word removal using a specially-constructed list of stop-words. The custom query parser is highly customizable, and the user has several configuration options from which to choose. The first such option is modality limitation. If the user selects this option, the query is parsed to extract the desired modality, if available. Using the modality fields described in the previous section, only those images that are of the desired modality are returned. This is expected to improve the precision, as only images of the desired modality would be included within the result set. However, there could be a loss in recall if the process of modality extraction and classification is inaccurate.

The system is linked to the UMLS Metathesaurus; the user may choose to perform manual or automatic query expansion using synonyms from the Metathesarus. In the manual mode, a list of synonyms is presented to the user, which the user can choose to add to the query. In the automatic mode, all synonyms of the UMLS preferred term are added to the query. Another configuration option is the "stem and star" option, in which all the terms in the query are first stemmed. A wildcard (*) is then appended to the word to allow the search of words containing the desired root. The last option allows the user to only send unique terms to the search engine. This can be useful when using the UMLS option, as many of the synonyms have a lot of overlap in the preferred terms.

---

[1] http://www.ruby-lang.org/

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

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

3   Modality Classification and Annotation

Medical images often begin life with rich metadata in the form of DICOM headers describing their imaging modality or anatomy. However, since most teaching or on-line image collections are made up of compressed standalone JPEG files, it is very common for medical images to exist *sans* metadata. In previous work [8], we described a modality classifier that could identify the imaging modality for medical images using supervised machine learning. We extended that work to the new dataset used for ImageCLEF 2009. We created additional tables in the database to store image information that was created using a variety of image processing techniques in MATLAB[4]. These include color and intensity histograms as well as measures of texture using gray-level co-occurrence matrices and discrete cosine transforms. These features can be used to find images that are visually similar to the query image.

One of the biggest challenges in creating such a modality classifier is creating a labeled training dataset of sufficient size and quality. Our system, as previously described [8], relied on a external training set of modality-labeled images for its supervised learning. In 2009, as in 2008, we did not use any external databases for training the modality classifier. Instead, a Ruby text parser was written to extract the modality from the image captions for all images in the collection using regular expressions as well as a simple Bayesian classifier.

Note that the quality and accuracy of these labels are not as good as in case of the external training set used in previous experiments. Images where a unique modality could be identified based on the caption were used for training the modality classifier. Grey scale images were classified into a set of modalities including x-rays, CT, MRI, ultrasound and nuclear medicine. Color image classes include gross pathology, microscopy, and endoscopy. The rest of the dataset (i.e., images for which zero or more than one modalities were parsed) was classified using the above classifier. We created two fields in the database for the modality that were indexed by our search engine. The first field contained the modality as extracted by the text parser, and the second contained the modality resulting from the classification process using visual features.

4   Text Processing and Analysis

Our text processing module is very simple; after applying a stop-word list, we index documents and queries using the Vector Space Model (VSM). During the retrieval process, a list of relevant documents can be ranked with regard to their relevance to the corresponding query.

While no specific treatment was applied to documents, we theorized that it would be useful to use external knowledge to interpret the queries' semantic content. Indeed, each query contains a precise description of a user need materialized by a set of words belonging to three semantic categories: modality of

---

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

the image (e.g., MRI, x-ray, etc.), anatomy (e.g., leg, head, etc.), and pathology (e.g., cancer, fracture, etc.). We call these categories "domain dimensions," and define them as follow: "A dimension of a domain is a concept used to express the themes in this domain" [11]. The idea behind our approach is that, in a given domain, a theme can be developed with reference to a set of dimensions of this domain. For instance, a Physician wishing to write a report about a medical image, first they focus on a domain (Medicine), next they refer to specific dimensions of this domain (e.g., Anatomy), then they choose words from this dimension (e.g., femur), and finally they write their report.

In order to resolve CLEF queries, we proposed using the domain dimensions to interpret their semantic content. To do so, we first need to define the dimensions. For this purpose, we use external resources, such as ontologies or thesaurus, to define each dimension by a hierarchy of concepts. Every concept is denoted by a set of words. Thereafter, to identify dimensions from a query, we extract query's words depending on the dimension hierarchy they belong to. Once dimensions are extracted from each query, we use them to search for relevant documents. In particular, we use Boolean operators on query's dimensions in order to reformulate the initial text of the query and better represent its semantic content. For instance, if we assume that a relevant document must contain all the dimensions belonging to the query, we should use the operator AND between the query's words that represent these dimensions in order to query the document collection.

Mainly, our querying process contains two steps. The first step consists in using the initial query's text to search for documents based on the VSM. The result of this step is a list of ranked documents called D. The second step consists of selecting, from D, those documents that satisfy the Boolean expression formulated based on the domain dimensions.

5   Runs Submitted

We submitted a total of 9 offical runs. The search options for the different runs are provided in Table 1. All our runs are automatic; 2 of them are based on textual data and 8 are based on both textual and visual data (mixed).

The "ohsu_j_no_mod" run is based on the VSM, where each document/query is represented by a vector of words. The result of this run, considered as the baseline, will be compared to those obtained by the other runs, which are based domain dimensions and/or both textual and visual data.

5.1   Modality Extraction-Based Runs

We submitted two mixed runs that used the automatically extracted modality to filter results. The custom query parser first extracted the desired modality from the query, if it existed. The "ohsu_j_mod1" run used the custom parser to remove stop-words from the query and limit the results to the desired modality. This run was expected to have high precision but potentially lower recall as it

Table 1. Description of OHSU runs.

Run Name	Retrieval Method	Run Type	Data Used
ohsu_j_umls	Mixed	Automatic	Full caption
ohsu_j_mod1	Mixed	Automatic	Full caption
OHSU_SR1	Textual	Automatic	Full caption
OHSU_SR2	Mixed	Automatic	Full caption
OHSU_SR3	Mixed	Automatic	Full caption
OHSU_SR4	Mixed	Automatic	Full caption
OHSU_SR5	Mixed	Automatic	Full caption
OHSU_SR6	Mixed	Automatic	Full caption
ohsu_j_no_mod	Textual	Automatic	Full caption

did not use any term expansion. Also, if the modality classier was not accurate or the modality extraction from the textual query was too strict, the results could be limited. In order to try to increase the recall, we also submitted a run labeled "ohsu_j_umls" where term expansion based on the UMLS metathesaurus was used.

5.2 Dimensions-Based Runs

With a view to defining the domain dimensions, we utilized the UMLS metathesaurus. The domain dimensions were defined using the UMLS semantic types as follows:

Anatomy: "Body Part," "Organ, or Organ Component," "Body Space or Junction," "Body Location or Region," and "Cell"

Pathology: "Sign or Symptom," "Finding," "Pathologic Function," "Injury or Poisoning," "Disease or Syndrome," "Neoplastic Process," "Neoplasms," "Anatomical Abnormality," "Congenital Abnormality," adn "Acquired Abnormality"

Modality: "Manufactured Object," and "Diagnostic Procedure"

After automatically extracting the dimensions from each query, we used them to perform the following runs.

OHSU_SR1: We use dimensions to rank the retrieved documents for a given query. Documents that contain the three dimensions belonging to the query are considered as most relevant and are therefore highly ranked. They are followed by those documents that are only missing the modality. Indeed, the latter dimension is not always explicitly described in the text, since the writer might use general words, such as "this image ...," to write their medical report. Finally, in the third rank, we find documents that contain at least one of the query dimensions.

Since the modality is not always explicitly described in the text, we use the visual data to extract it from images (Section 3). Thereafter, we use it to re-rank the document list obtained using the textual data. In particular, documents of

which the modality has been extracted from image are ranked on the top. In all the following runs, we apply this technique using the result of the "ohsu_j_mod1" run. For simplicity of writing, we call this process "checking modality technique."

OHSU_SR2: Re-rank the document list obtained during run OHSU_SR1 by applying the "checking modality technique."

OHSU_SR3: Re-rank the document list obtained during run ohsu_j_no_mod by applying the "checking modality technique."

OHSU_SR4: We select, for each query, those documents that contain the Anatomy and the Pathology belonging to this query. The obtained result is re-ranked by applying the "checking modality technique."

OHSU_SR5: From the result of the "OHSU_SR2" run, we randomly select one image from each article. Indeed, in CLEF documents, a textual article might contain more than one image. In this run, if an article is retrieved by the textual approach, we randomly select one of its images (instead of keeping all images as we have been doing for the other runs).

OHSU_SR6: We select from the result of the "OHSU_SR1" run, only those documents of which the modality has been extracted during the "ohsu_j_mod1" run.

6 Results and Discussion

For each query, the results are measured by mean average precision of top 1000 documents (MAP), precision at 10 documents (p@10), and precision at 5 documents (p@5). The results given by the baseline run are 0.1223 (MAP), 0.416 (p@10), and 0.38 (p@5). Obtained results are presented in Table 2 where rows correspond to the runs, and values correspond to their corresponding results. We also include data from the best runs (based on MAP and p@5) in ImageCLEF 2009 campaign.

Table 2. Results of OHSU runs.

Run Name	MAP	P@5	P@10
ohsu_j_umls	0.1753	**0.712**	**0.664**
ohsu_j_mod1	0.1698	0.592	0.552
OHSU_SR1	**0.1756**	0.592	0.536
OHSU_SR2	0.1582	0.624	0.54
OHSU_SR3	0.1511	0.608	0.524
OHSU_SR4	0.1147	0.6	0.484
OHSU_SR5	0.1133	0.584	0.516
OHSU_SR6	0.1646	0.68	0.612
ohsu_j_no_mod	0.1223	0.416	0.38
Best in p@5	0.3775	0.744	0.716
Best in MAP	0.4293	0.696	0.664

As described above, our system has been designed to improve precision, perhaps at the expense of recall. Since we do not use any advanced natural language processing and we filter images based on purported modality, we were expecting a relatively low recall. Consequently, as the MAP is highly dependent and limited by recall, we believe that it makes more sense to compare our results to those obtained by the other ImageCLEF 2009's participants at an early precision level (p@5 or p@10). We have a high early precision (p@5 = 0.712) obtained by performing the "ohsu_j_umls" run. This is the second best result obtained in the ImageCLEF 2009 campaign; the first being equal to 0.744 obtained by an interactive run. In fact, it had the highest p@5 for automatic runs in Image-CLEF2009.

Independent of the other ImageCLEF 2009 participants, most of our runs outperform our baseline, reaching an improvement of 43% in MAP (OHSU_SR1) and 71% in p@5 (ohsu_j_umls). From the result of the "OHSU_SR1" run, we notice that the use of domain dimensions is of great interest in solving medical multimodal queries. Indeed, by using domain dimensions, we highlight the "relevant words" that describe the queries' semantic content. Using these words, the system can retrieve only documents that contain the anatomy, the modality, and the pathology described in the query text.

We notice that at p@5 and p@10, all our mixed runs outperform our baseline. This is not surprising, because the first ranked documents are those that have been retrieved both by the text-based search technique and the visual-based search technique. This supports our previous conclusions, which confirm that the retrieval performance can be improved demonstrably by merging the results of textual and visual search techniques [7–9].

Compared to the baseline, our results decreased in two runs in terms of MAP. First, in the "OHSU_SR4" run, where the modality dimension was ignored during the querying process. This decrease might be explained by the fact that the modality is described in some documents, and its use is thought to beneficial. This is notably obvious in the "OHSU_SR1" run when we used the three dimensions and obtained the highest performance. Secondly, in the "OHSU_SR5" run, only one image was randomly selected from each article. An increase in the performance would have been surprising, since this technique is not accurate at all, and there is a high risk that the selected image is irrelevant. Indeed, it is better to keep all images of each article, thus, if one of them is relevant to the corresponding query, it will be retrieved.

7 Conclusions and Future Work

In order to improve early precision in the task of solving medical multimodal images, we combined a text-based search technique with a visual-based one. The first technique consists in using domain dimensions to highlight relevant words that describe the queries' semantic content. While anatomy and pathology are relatively easy to identify from textual documents, it is quite challenging to identify the modality dimension. To overcome this problem, we used a visual-

based search technique that allows automatically extracting the modality from images. The obtained results in terms of precision at p@5 and p@10 are very encouraging and outperform our baseline.

Among the ImageCLEF 2009's particians, we obtained the second best overall and best automatic result in terms of precision at p@5. However, in terms of MAP, even though our results outperform our baseline, they are significantly below the best performance obtained in ImageCLEF 2009. This was expected, since we were only focused on improving early precision and did not use any advanced natural language processing to improve the recall. For our future work, we will focus on this issue. We believe that our current text-based search technique has room for improvement. We plan to use further textual processing, such as term expansion and pseudo-relevance feedback, in order to improve our recall, and hope to be able to compare our results to the best ImageCLEF 2009' performance.

Acknowledgements

References

1. Hersh, W.R., Müller, H., Jensen, J.R., Yang, J., Gorman, P.N., Ruch, P.: Advancing biomedical image retrieval: Development and analysis of a test collection. J Am Med Inform Assoc (June 2006) M2082
2. Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. Pattern Analysis and Machine Intelligence, IEEE Transactions on **22**(12) (2000) 1349–1380
3. Tagare, H.D., Jaffe, C.C., Duncan, J.: Medical image databases: A content-based retrieval approach. J Am Med Inform Assoc **4**(3) (May 1997) 184–198
4. Aisen, A.M., Broderick, L.S., Winer-Muram, H., Brodley, C.E., Kak, A.C., Pavlopoulou, C., Dy, J., Shyu, C.R., Marchiori, A.: Automated storage and retrieval of thin-section ct images to assist diagnosis: System description and preliminary assessment. Radiology **228**(1) (July 2003) 265–270
5. Schmid-Saugeona, P., Guillodb, J., Thirana, J.P.: Towards a computer-aided diagnosis system for pigmented skin lesions. Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society **27**(1) (2003) 65–78 PMID: 12573891.
6. 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 **73**(1) (February 2004) 1–23
7. Hersh, W., Kalpathy-Cramer, J., Jensen, J.: Medical image retrieval and automated annotation: Ohsu at imageclef 2006. In Peters, C., Clough, P., Gey, F.C., Karlgren, J., Magnini, B., Oard, D.W., de Rijke, M., Stempfhuber, M., eds.: CLEF. Volume 4730 of Lecture Notes in Computer Science., Springer (2006) 660–669

8. Kalpathy-Cramer, J., Hersh, W.: Automatic image modality based classification and annotation to improve medical image retrieval. Studies in Health Technology and Informatics **129**(Pt 2) (2007) 1334–8 PMID: 17911931.
9. Radhouani, S., HweeLim, J., pierre Chevallet, J., Falquet, G.: Combining textual and visual ontologies to solve medical multimodal queries. Multimedia and Expo, IEEE International Conference on **0** (2006) 1853–1856
10. MÃ¼ller, H., Kalpathy-Cramer, J., Eggel, I., Bedrick, S., Radhouani, S., Bakke, B., Jr., C.K., Hersh, W.: Overview of the medical retrieval task at imageclef 2009. Working Notes of the CLEF 2009 workshop, Corfu, Greece (2009)
11. Radhouani, S.: Un modèle de recherche d'information orienté précision fondé sur les dimensions de domaine. PhD thesis, University of Geneva, Switzerland, and University of Grenoble, France (2008)