FCSE at ImageCLEF 2012: Evaluating techniques for medical image retrieval

Ivan Kitanovski, Ivica Dimitrovski and Suzana Loskovska

Faculty of Computer Science and Engineering, Skopje, Macedonia
{ivan.kitanovski, ivica.dimitrovski, suzana.loshkovska}@finki.ukim.mk

Abstract. This paper presents the details of the participation of FCSE (Faculty of Computer Science and Engineering) research team in ImageCLEF 2012 medical retrieval task. We investigated by evaluating different weighting models for text retrieval. In the case of the visual retrieval, we focused on extracting low-level features and examining their performance. For, the multimodal retrieval we used late fusion to combine the best text and visual results. We found that the choice of weighting model for text retrieval dramatically influences the outcome of the multimodal retrieval. We tested the multimodal retrieval on data from ImageCLEF 2011 medical task and based on that we submitted new experiments for ImageCLEF 2012. The results show that fusing different modalities in the retrieval can improve the overall retrieval performance.

Keywords: Information Retrieval, Medical Imaging, Content-based Image Retrieval, Medical Image Retrieval

1 Introduction

In this paper we present the experiments performed by the Faculty of Computer Science and Engineering (FSCE) team for the medical retrieval task at ImageCLEF [1] 2012.

The task of medical image retrieval consists of retrieving the most relevant images to a given query from a database of images. Medical image retrieval from medical image databases does not aim to replace the physician by predicting the disease of a particular case but to assist him/her in diagnosis. By consulting the output of a medical image retrieval system, the physician can gain more confidence in his/her decision or even consider other possibilities.

There are two forms of medical image retrieval: text-based (textual) and content-based (visual) [1]. In text-based image retrieval images are usually manually annotated with keywords or a short caption, which describe their content, or in the case of medical images the keywords are related to modality of the image, the present body part, the disease or anomaly depicted. In the latter stage, the user provides textual queries and the retrieval is performed using traditional text retrieval techniques. In visual retrieval the images are represented using descriptors (automatically generated) which describe the visual content of the images. Descriptors

are usually numerical by nature and are represented as vectors of numbers [2]. In the retrieval phase, the user provides visual queries (query images) and the retrieval is performed by comparing descriptors of the query images to those of all images in the database [3].

Recently, multimodal image retrieval arises as an active research topic [4] and is also part of medical task of ImageCLEF. Multimodal image retrieval is the process of using both text-based and visual-based techniques for retrieval. In multimodal retrieval the user provides textual queries and query images and retrieval should provide an ordered set of images related to that complex query. The authors of [5] use late fusion to combine the results from text-based and visual-based retrieval. For the text-based retrieval, they use a bag-of-words representation on the image captions and DFR-BM25 model for the retrieval. In the visual-based retrieval, they describe the images using a low-level feature called CEDD, and the retrieval is performed using Img(Rummager). Then, using late fusion they combine the results. The most efficient strategy was a linear combination scheme. In [6], the authors use Late Semantic Combination for multimodal retrieval. They represent each image caption with a bag-of-words representation and for the retrieval they compare several models: Dirichlet Smoothed Language (DIR), Log-logistic Information-Based Model (LGD), Smoothed Power Law Information-Based Model (SPL) and Lexical Entailment based IR Model (AX). In the visual retrieval, the images are described with ORH and COL features and they use dot product as similarity measure for the visual retrieval. In [7] the authors use linear late fusion for multimodal retrieval. The text-based retrieval is performed using Lucene and the visual-based using Lira. The multimodal retrieval is performed by linear combination of scores from the text-based and visual-based retrieval and re-ranking. The authors of [8] first perform text-based retrieval and use those results as a filter for the visual-based retrieval. They use low-level texture and color features (CEDD) for the visual-based retrieval. Text-based retrieval is performed using a Lucene.

The paper is organized as follows: Section 2 presents the feature set which is used for the text-based and visual-based retrieval. Multimodal retrieval techniques are presented in section 3. The evaluation of our used features is described in section 4. Section 5 provides the submitted runs. The concluding remarks are given in section 6.

2 Feature Set

Feature selection is a very important part in every information retrieval system, since it directly influences the performance. In this paper we analyze text features for the text-based retrieval and visual features for visual-based retrieval.

2.1 Textual Features

Text-based retrieval is needed when we have text describing the image content i.e. image caption. From the related work we can conclude that regarding the text-based retrieval a traditional bag-of-words representation can be used for the image caption.

The image captions are first pre-processed. Pre-processing includes stemming and stop words removal [5], which is needed so we can extract only the vital information.

The choice of a weighting model may crucially affect the retrieval hence we evaluated the positive and negative sides of different weighting models. We evaluated the following models: PL2 [9], BM25 [9], DFR-BM2 [9], BB2 [9] and one of the most popular TF-IDF [10]. We choose these models as one the most commonly used in real practice.

2.2 Visual Features

Related work shows that low-level features are typically used in content-based image retrieval systems, since they typically deal with large image databases. These features are called low-level because they have little or nothing with human perception. We decided to use the following features:

- Color and Edge Directivity Descriptor (CEDD) combines EHD feature [11] with color histogram information. This descriptor is limited in size to 54 bytes per image, which makes it appropriate for large image databases. Important attribute of the CEDD is the low computational power needed for its extraction, in comparison to the needs of the most MPEG-7 descriptors.
- Fuzzy Color and Texture Histogram (FCTH) is a fuzzy version of CEDD feature which contains fuzzy set of color and texture histogram [12]. FCTH contains results from the combination of three fuzzy systems including histogram, color and texture information. This feature is limited in size to 72 bytes per image and that makes it suitable for use in large image databases.
- The Scalable Fuzzy Brightness and Texture Directionality Histogram (BTDH), was originally created for representing radiology images [13]. BTDH is very similar to FCTH feature. The main difference from FCTH feature is using brightness instead of color histogram. It combines brightness and texture characteristics and their spatial distribution in one compact vector by using a two-unit fuzzy system. This feature does not contain color data, since it was meant for grayscale images.

3 Fusion Techniques for Multimodal Retrieval

Multimodal information retrieval refers to the task of using multiple media to perform a retrieval task. Multimodal retrieval is usually done by fusing multiple modalities. Fusing multiple modalities can improve the overall accuracy in the decision making process [14].

The fusion of multiple modalities can be performed at feature level or decision level. In fusion at feature level, also known as early fusion, various features extracted from the input data are combined in some fashion and then that newly created feature is sent as input to the module that performs the analysis task. In fusion at decision level, also known as late fusion, the analysis units first provide the local decisions that are obtained based on individual features. Afterwards a decision fusion unit combines local decisions to create a new fused decision vector which is analyzed to provide a

final decision about the task. To utilize the merits of both approaches, researchers have attempted to create hybrid techniques which are a combination of both feature and decision level techniques.

Related work shows that the late fusion strategy is frequently used. Late fusion has many advantages over early fusion. The decisions usually have the same representation. For instance, the result of both text-based and visual-based retrieval is an ordered list of images. Hence, the implementation of fusion techniques becomes much easier. Furthermore, late fusion allows for modularity and scalability in terms of modalities used in the fusion process, which is quite difficult to achieve with early fusion techniques. Additionally, late fusion allows us to use the optimal analyzing methods for each modality separately which cannot be done with early fusion techniques.

Because of these merits we used late fusion strategy in our experiments. We turned to Linear Weighed Fusion strategy, one of the simplest and most widely used methods. We applied this strategy to results obtained from the separate text-based and visual-based retrievals. Each, retrieval contains an ordered list of images with computed similarity scores. The weighted average function is applied by multiplying each individual similarity with a weight value. The weight assignment to individual scores defines the importance of each modality in the decision making process. If a modality has a high weight it will have significant impact on the final results and vice versa. In this way we can control the influence of individual modalities.

4 Evaluating Features on ImageCLEF2011

The evaluation consists of three kinds of retrieval: text-based, visual-based and multimodal retrieval. We use the text-based and visual-based retrieval to find the best weighting model and descriptor, which we latter use in the multimodal retrieval.

The data for the evaluation is provided from the collection of the ImageCLEF 2011 medical task [1]. The collection contains textual and visual information. It consists of 230088 images, each described with a short text (image caption). The queries which we used for testing are the same which were provided for the medical retrieval task. Participants were given a set of 30 textual queries with 2-3 sample images for each query. The queries are classified into textual, visual or mixed (multimodal) based on the data and techniques used. For the text-based retrieval, we only use the text queries. On the other hand, for the visual-based retrieval we use the images provided for each query. Finally, for the multimodal retrieval we use both text and image data provided for every query.

The text-based retrieval was performed using Terrier IR Platform [15], open source search engine written in Java which is developed at School of Computer Science, University of Glasgow. For stemming we used Porter stemmer [16] for English, since the image captions and text queries are in English. Terrier also has a predefined stop words list, which we use in the preprocessing stage. All weighting models which we analyze are integrated in Terrier. The results from these evaluations in presented on Table 1.

Table 1. Comparison of weighting models in text-based retrieval

Model	*MAP*	*P10*	*P20*	*Rprec*	*# of rel. docs*
BB2	0.2059	0.3700	0.3100	0.2425	1472
BM25	0.2144	0.3633	0.3200	0.2449	1504
DFR-BM25	0.2054	0.3533	0.2967	0.2426	1494
PL2	0.1970	0.3533	0.2967	0.2413	1474
TF-IDF	0.2048	0.3533	0.3033	0.2398	1482

The visual-based retrieval was performed with the aid of the Img(Rummager) application [17], developed in the Automatic Control Systems & Robotics Laboratory at the Democritus University of Thrace-Greece. CEDD, FCTH and BDTH features are implemented in the application. The retrieval stage greatly relies on the distance/similarity function used to quantitatively compare the images. We compute the similarity score based on Tanimoto distance [11], since it is one most frequently used methods for the visual features which we use to describe the images. Since, there were multiple images per query we used an averaging technique in this stage [5]. Here we calculated a mean descriptor from all images in a query, thus creating one new feature vector which will be passed as a query. Visual results are presented at Table 2.

Table 2. Comparison of features in visual-based retrieval

Feature	*MAP*	*P10*	*P20*	*Rprec*	*# of rel. docs*
CEDD	0.0142	0.0867	0.0733	0.0401	552
FCTH	0.0134	0.0633	0.0483	0.0342	621
BDTH	0.0053	0.0419	0.0372	0.0216	217

The multimodal retrieval is performed using late fusion strategy. In this stage we pick the best weighting model for text-based retrieval and the best performing descriptor for visual-based retrieval. Then, we combine the results from the separate retrievals using linear combination. The formula by which we calculated the score for each image in the retrieval is the following:

$$(text\_score * w_1 + visual\_score * w_2) / 100 = score \quad (1)$$

After comparing different studies [18] and experimenting with various parameters we determined to multiply the text score with 85 and the visual score with 15, thus giving a greater influence to the text-based component.

Before we combine the score we need to normalize them to get more valid and accurate results since different modalities calculate different ranges of values in the similarity score. Here we apply the most common used method for normalization i.e.

Min-Max normalization [19]. This normalization ensures that the values of the scores are in the range from 0 to 1. The lowest value is set to 0 and the highest value is set to 1. This allows us to compare values that are measured using different scales. After normalization takes place, we turn to linear combination of the modified retrieval scores.

In this case, we make three types of experiment to assess the change in retrieval performance. First, we make linear combination of the text-based and visual-based retrieval. The second approach slightly modifies the text-based retrieval with query expansion, since the text-based retrieval has the crucial impact on final result. The third approach uses query expansion and word weighting. This approach assigns weights to special words, in our case image modalities (i.e. MRI, CT, X-RAY etc.). We added a weight of 2.5 to these words using query language of Terrier. The results (Table 3) show that there is an improvement of the retrieval compared to text-based retrieval in every multimodal experiment.

Table 3. Results of the multimodal retrieval experiments

Mode	*MAP*	*P10*	*P20*	*Rprec*	*# of rel. docs*
mixed	0.2148	0.3600	0.3233	0.2579	1531
mixed + qe	0.2179	0.3833	0.3433	0.2577	1483
mixed + ww	0.2232	0.3933	0.3467	0.2568	1458

5 Submitted Results on ImageCLEF 2012

After evaluating the performance of different visual features, weighting models and linear combination parameters we could evaluate which performs best under which conditions. We made another experiment, only now using ImageCLEF 2012 data and submitted the results only from the best performing techniques. For text-based retrieval we submitted the run using BM25 weighting model using query expansion and word weights and for the visual-based retrieval we submitted the run using CEDD descriptor. Finally, for the multimodal retrieval we submitted the linear combination of the two previous modalities. The results from our runs on ImgeCLEF 2012 are presented on Table 4.

Table 4. Runs of FCSE group in ImageCLEFMed 2012

Type	*MAP*	*GM-MAP*	*bpref*	*P10*	*P30*
text	0.1763	0.0498	0.1773	0.2909	0.1864
visual	0.0041	0.0003	0.0105	0.0318	0.0364
mixed	0.1794	0.049	0.1851	0.3	0.1894

6 Conclusion

In this paper we explained in detail our participation in ImageCLEF 2012. We examined the effects of different weighting models for text-based retrieval and concluded that the choice of a weighting model dramatically influences retrieval. In the case of visual-based retrieval, we compared several low level visual features and found that CEDD descriptor to be the best suited for this type of task.

Additionally, we investigated in late fusion for multimodal retrieval. We used linear combination for late fusion of the text-based and visual-based retrieval results. The obtained results show that by combining the two modalities the overall retrieval performance can be improved.

Medical image retrieval is a crucial task which can aid the work of medical practitioners. It is a very complex which can be improved in many aspects from improving current weighting models to developing or modifying features to describe the image content and creating different techniques to combine these two modalities.

References

1. Kalpathy--Cramer, J., Muller, H., Bedrick, S., Eggel, I. and de Herrera, A.G.S. and Tsikrika, T.: Overview of the CLEF 2011 medical image classification and retrieval tasks (2011)
2. Sonka, M., Hlavac, V., Boyle, R. et al.: Image processing, analysis, and machine vision. PWS publishing Pacific Grove, CA. 2 (1999)
3. Deb, S., Zhang, Y.: An overview of content-based image retrieval techniques. 18th International Conference on Advanced Information Networking and Applications. 1. pp. 59—64 (2004)
4. Muller, H., Kalpathy--Cramer, J. et al.: Overview of the CLEF 2009 medical image retrieval track. Multilingual Information Access Evaluation II. Multimedia Experiments. pp. 72—84 (2010)
5. Alpkocak, A., Ozturkmenoglu, O., Berber, T., Vahid, A.H., Hamed, R.G.: DEMIR at ImageCLEFMed 2011: Evaluation of Fusion Techniques for Multimodal Content-based Medical Image Retrieval. 12th Workshop of the Cross-Language Evaluation Forum (CLEF), Amsterdam, Netherlands (2011)
6. Csurka, G., Clinchant, S., Jacquet, G.: XRCE's Participation at Medical Image Modality Classification and Ad-hoc Retrieval Tasks of ImageCLEF 2011 (2011)
7. Gkoufas, Y., Morou, A., Kalamboukis, T.: IPL at ImageCLEF 2011 Medical Retrieval Task. Working Notes of CLEF (2011)
8. Castellanos, A., Benavent, X., Benavent, J., Garcia-Serrano, A.: UNED-UV at Medical Retrieval Task of ImageCLEF (2011)
9. Amati, G., Van Rijsbergen, C.J.: Probabilistic models of information retrieval based on measuring the divergence from randomness. ACM Transactions on Information Systems (TOIS). 20. pp. 357--389 (2002)
10. Hiemstra, D.: A probabilistic justification for using tf-idf term weighting in information retrieval. International Journal on Digital Libraries. 3. pp. 131--139 (2000)
11. Chatzichristofis, S., Boutalis, Y.: Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. Computer Vision Systems. pp. 313--322 (2008)

12. Chatzichristofis, S.A., Boutalis, Y.S.: Fcth: Fuzzy color and texture histogram-a low level feature for accurate image retrieval. Ninth International Workshop on Image Analysis for Multimedia Interactive Services. pp. 191--196 (2008)
13. Chatzichristofis, S.A., Boutalis, Y.S.: Content based radiology image retrieval using a fuzzy rule based scalable composite descriptor. Multimedia Tools and Applications. 46. pp. 493--519 (2010)
14. Atrey, P.K., Hossain, M.A., El Saddik, A., Kankanhalli, M.S.: Multimodal fusion for multimedia analysis: a survey. Multimedia Systems. 16. pp. 345--379 (2010)
15. Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C., Johnson, D.: Terrier information retrieval platform. Advances in Information Retrieval. pp. 517--519 (2005)
16. Porter, M.F.: An algorithm for suffix stripping (1980)
17. Chatzichristofis, S.A., Boutalis, Y.S., Lux, M.: Img (rummager): An interactive content based image retrieval system. Second International Workshop on Similarity Search and Applications. pp.151--153 (2009)
18. Croft, W.B.: Combining approaches to information retrieval. Advances in information retrieval. pp. 1--36 (2002)
19. Jain, A., Nandakumar, K., Ross, A.: Score normalization in multimodal biometric systems. Pattern recognition. 38. pp. 2270--2285 (2005)