

# DEMIR at ImageCLEFMed 2012: Inter-modality and Intra-Modality Integrated Combination Retrieval

Ali Hosseinzadeh Vahid, Adil Alpkocak, Roghaiyeh Gachpaz Hamed, Nefise  
Meltem Ceylan, and Okan Ozturkmenoglu

Dokuz Eylul University  
Department of Computer Engineering  
DEMIR Dokuz Eylul Multimedia Information Retrieval Research Group  
Tinaztepe, 35160 Izmir, Turkey

ali.vahid@st.cs.deu.edu.tr, alpkocak@cs.deu.edu.tr, ramisa\_84@yahoo.com  
meltem.ceylan@gmail.com, okan.ozturkmenoglu@deu.edu.tr

**Abstract.** This paper present the details of participation of DEMIR (Dokuz Eylul University Multimedia Information Retrieval) research team to the ImageCLEF 2012 Medical Retrieval task. This year, we evaluated impact of our proposed Integrated Combination method and Explicit Graded relevance feedback on the most descriptive low level features of images and best text retrieval result. We improved results by examination of different level of integrated approach for retrieved text data and low level features. We tested multi-modality image retrieval in ImageCLEF 2012 medical retrieval task and obtained the best rank in visual retrieval due to our experiments. The results clearly show that proper combination of different modalities improve the overall retrieval performance.

**Keywords:** Integrated Combination, fusion methods, Relevance Feedback, Multimodal Information Retrieval, Content-based Image Retrieval, Medical Image Retrieval, Information Retrieval

## 1 Introduction

In this paper we present the experiments performed by Dokuz Eylul University Multimedia Information Retrieval (DEMIR) Group, Turkey, in the context of our participation to the ImageCLEF 2012 Medical Image retrieval task [1]. The main focus of this work is evaluation of result improvement using our proposed Integrated Combination Multimodal Retrieval (ICMR) system. To acquire the optimum result, not only we applied inter-modality integrated combination of text and low-level image features; also we utilized this method to combine result of different low level features of images. On the other hand, we performed the experiments for narrowing down the data

collection by defining and filtering out of irrelevant documents. Also we checked the impact of relevance feedback on performance of medical image retrieval system.

After analysis of integrated combination method (Section 2), we describe our techniques for filtering the data collection out and for image query expansion using relevance feedback (Section 3 , 4). After we present experiments on ImageCLEF 2012 Medical track data (Section 5), then Section 6 concludes the paper by pointing out the open issues and possible avenues of further research in the area of multimodal fusion techniques for content-based image retrieval.

## **2 INTEGRATED COMBINATION multimodal RETRIEVAL**

It seems that the best performance of multi-modality fusion appears when we can put all relevant retrieved documents of each fused modalities in response to a query into relevant retrieved document set of fused modalities. More formally, Union of modalities' relevant retrieved document sets can achieve the best result in fusion. However, there are some documents that are relevant but not appear in any individual modalities' result set. This is the limit of multimodality fusion. Because of that, in literature of multimodal data fusion, some of the authors claimed that data fusion algorithms are competitive in performance and is not devoid of risks and sometimes can degrade the retrieval performance.

Our previous studies illustrated that main reason for this downside originated from threshold on number of fused document (1). In most fusion methods, combination applied partially, often on top 1000 relevant retrieved documents of each modality. While in our proposed integrated approach, all documents of data collections in each modality participate in combination, based on their similarity score or rank position.

In order to evaluate the impact of our proposed method on improvement of overall system performance, we designed and implemented an experimental integrated combination multimodal retrieval system. In this system, each modality was preprocessed and indexed separately. Then the results obtained from different modalities in response to each query combined based on selected combination methods and produced the final result set of retrieved documents. This combination can be inter-modality

(between results of different modalities) or intra-modality (between different features of a modality).

## **2.1 Textual Features**

In order to simplify the work, we split the XML file for textual metadata and represented each image in the collection as a structured document of xml file. We also expanded the XML file using related article full text, abstract and title as new tags. We used Terrier IR Platform API, open source search engine written in Java and is developed at the School of Computing Science, University of Glasgow (2), for our Text Based Information Retrieval subsystem. Terrier provides both efficient and effective search methods for large-scale document collections. To introduce flexibility to the processing and transformation of textual information, it requires a preprocessing in different ways. The order in which transformations were applied is as follows: 1) special characters deletion: characters with no meaning, like punctuation marks or blanks, are all eliminated; 2) stop words removal: discarding of semantically empty words, very high frequency words, 3) token normalization: converting all words to lower case 4) stemming: we used the Porter stemmer (3) as a process for removing the commoner morphological endings from words in English.

Since the choice of the weighting model may crucially affect the performance of any information retrieval system, first of all we decided to work on evaluating the relative merits and drawbacks of different weighting models using Terrier IR Platform (2), open source search engine written in Java and is developed at the School of Computing Science, University of Glasgow.

Since choice of the weighting model may crucially affect the performance of any information retrieval system, we performed our experiments on textual features using TF-IDF, the most common and famous weighting model in IR systems.

## **2.2 Visual Features**

The visual content of the images is mapped into a new space named the feature space. A feature is a set of characteristics of the image. In addition, a feature can further be enriched with information about the spatial distribution of the characteristic that it describes. Selection of right features, typically represented as vectors in this space, is the major aspect to attain discriminative and sufficient retrieval systems. We extracted features for all

images in test collection and query examples using Rummager tool (4), which is developed in the Automatic Control Systems & Robotics Laboratory at the Democritus University of Thrace-Greece. Although considerable design and experimental work, and rigorous testing, have been performed in MPEG-7 to arrive at efficient image descriptors for similarity matching, but no single generic color descriptor exists that can be used for all foreseen applications. As a result, a range of descriptors has been standardized, each suitable for achieving specific similarity-matching functionalities. On the other hand, there is some new and well known set of composite descriptors that combine two or more feature types. Here we introduce some of such new and well known set of composite descriptors then present the experimental results of comparison on performance of extracted features to select the most effective features as a visual modality:

### **2.2.1 Color and edge directivity descriptor (CEDD)**

The CEDD includes texture information produced by the six-bin histogram of the fuzzy system that uses the five digital filters proposed by the MPEG-7 EHD. Additionally, for color information the CEDD uses the 24-bin color histogram produced by the 24-bin fuzzy-linking system. Overall, the final histogram has 144 regions. Each Image Block interacts successively with all the fuzzy systems. In the Texture Unit, the Image Block is separated into four regions called Sub Blocks. The value of each Sub Block is the mean value of the luminosity of the pixels it contains. The luminosity values are derived from a YIQ color space transformation. Each Image Block interacts with the five digital filters proposed by MPEG-7 EHD, and with the use of the pentagonal diagram it is classified in one or more texture categories. Then, in the Color Unit, every Image Block is converted to the HSV color space. The mean values of H, S and V are calculated and become inputs to the fuzzy system that produces the fuzzy ten-bin histogram. Then, the second fuzzy system uses the mean values of S and V as well as the position number of the bin (or bins) resulting from the previous fuzzy ten-bin unit, calculates the hue of the color and produces the fuzzy 24-bin histogram. The combination of the three fuzzy systems will finally classify the Image Block. The process is

repeated for all the image blocks. At the completion of the process, the histogram is normalized and quantized (5).

### **2.2.2 Fuzzy color and texture histogram (FCTH)**

The FCTH descriptor includes the texture information produced in the eight-bin histogram of the fuzzy system that uses the high frequency bands of the Haar wavelet transform. For color information, the descriptor uses the 24-bin color histogram produced by the 24-bin fuzzy-linking system. Overall, the final histogram includes 192 regions. Each Image Block interacts successively with all the fuzzy systems in the exact manner demonstrated in CEDD production. Each Image Block is transformed into the YIQ color space and transformed with the Haar Wavelet transform. The fLH, fHL and fHH values are calculated and with the use of the fuzzy system that classifies the f coefficients, this Image Block is classified in one of the eight output bins. Next, the same Image Block is transformed into the HSV color space and the mean H, S and V block values are calculated. These values become inputs to the fuzzy system that forms the ten-bin fuzzy color histogram. Then, the next fuzzy system uses the mean values of S and V as well as the position number of the bin (or bins) resulting from the previous fuzzy ten-bin unit, to calculate the hue of the color and create the fuzzy 24-bin histogram. The combined three fuzzy systems therefore classify the Image Block. The process is repeated for all the blocks of the image. At the completion of the process, the histogram is normalized and quantized (6).

### **2.2.3 Brightness and Texture Directionality Histogram (BTDH)**

This feature is very similar to FCTH feature. The main difference from FCTH feature is using brightness instead of color histogram. This descriptor uses brightness and texture characteristics as well as the spatial distribution of these characteristics in one compact 1D vector. The most important characteristic of the proposed descriptor is that its size adapts according to the storage capabilities of the application that is using it. This characteristic renders the descriptor appropriate for use in large medical (or gray scale) image databases. To extract the proposed descriptor, a two unit fuzzy system is used. To extract the brightness information, a fuzzy unit classifies the brightness values of the image's pixels into  $L_{\text{Bright}}$  clusters. The cluster centers are calculated using the Gustafson Kessel Fuzzy Classifier.

The texture information embodied in the proposed descriptor comes from the Directionality histogram. This feature is part of the well-known Tamura texture features. Fractal Scanning method through the Hilbert Curve or the Z-Grid method is used to capture the spatial distribution of brightness and texture information (7)

After extracting features, we examined the performance of all extracted feature and perceived that compact composite features like CEDD and FCTH have satisfactorily retrieval result on our image collection and require noticeably lower computational power and storage space.(as Figure 1)

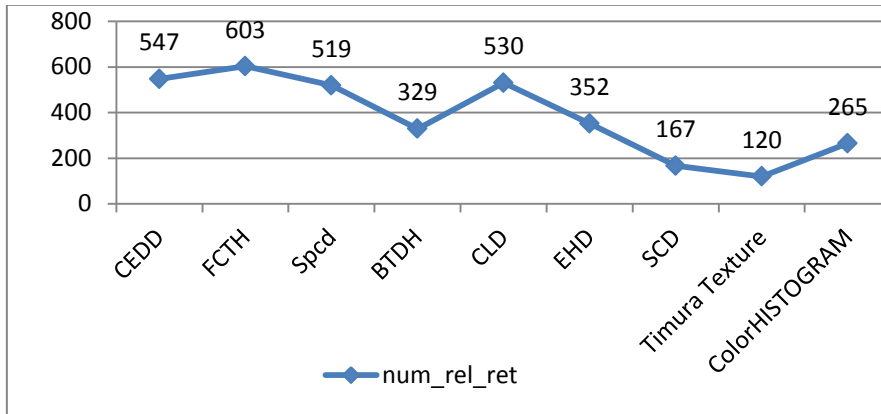


Figure 1. Comparison on performance of different low level features

Also, we assessed performance of different similarity function on Compact Composite features and comprehended Euclidean distance on CEDD and FCTH features and Cosine distance function on SPCD and BTDH produce the best performance(as illustrated in Figure2). Then we calculate the similarity between query and dataset objects using Euclidean distance in matching phase. Then we sorted all of dataset images in a descending list based on the value of similarity score in corresponding to each query example image.

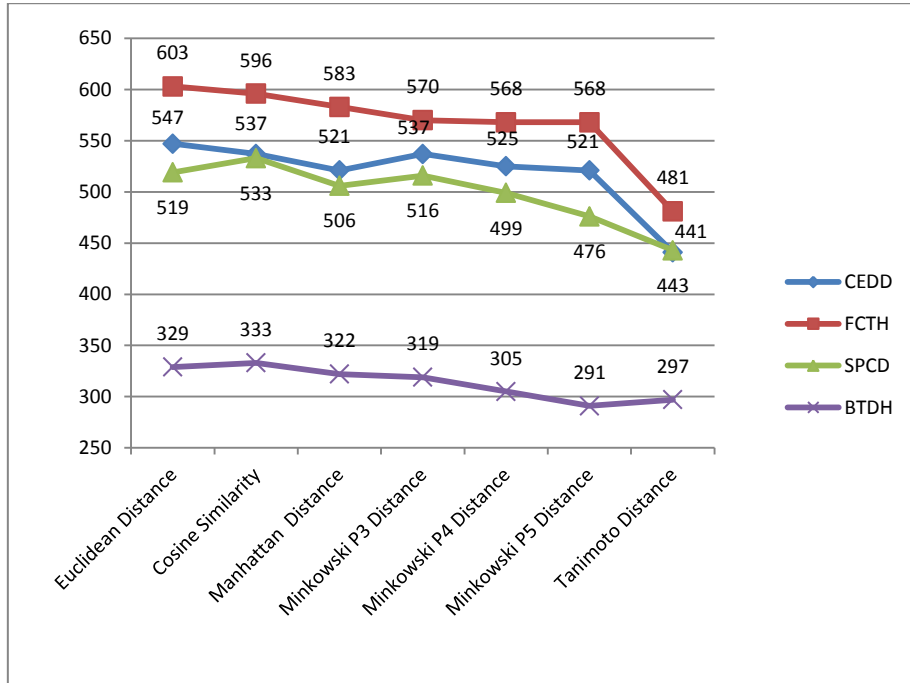


Figure 2. Distance function performance evaluation of different features based on number of relevant retrieved documents.

### 3 Query Expansion with Relevance feedback

The idea behind relevance feedback is to take the results that are initially returned from a given query and to use information about whether or not those results are relevant to perform a new query. Firstly we examined blind or Pseudo relevance feedback on ImageCLEFmed 2011 data collection and assumed that the top 3 ranked documents were most relevant then considered them as query image for each topic but found that this approach could not increase the performance of system. Therefore we performed Explicit Graded relevance feedback to indicate the most relevant images retrieved from a typical query based on assessment of non-professional person. Then we employ those images as query image of the topics.

### 4 Narrowing down

The database includes over 300,000 images of the biomedical open access literature from journals of BioMed Central at the PubMed Central database associated with their original articles in the journals. Therefore, there are many irrelative generic biomedical illustrations that make the collection so

noisy. This noisy documents influence on reduction of system precision. To solve this problem, we filter out such documents using classification decision tree methods and narrowing data collection down to about 35000 images.

## **5 Experimentations**

In order to assess the above mentioned methods, we set up a set of experiments on the data collection of ImageCLEF 2012 Medical retrieval. We submitted 24 runs to ImageCLEF Ad-hoc Medical Retrieval in five categories as following:

### **5.1 Baseline text-only runs**

This category includes the runs for baseline retrieval in textual modality on xml metadata, called as DEMIR\_R1 to DEMIR\_R4. Difference between runs of this category is on tags that are indexed. DEMIR\_R1 index just caption of images in original article while DEMIR\_R2 indexed full articles correspond to image. Caption of the images and abstract of the corresponded articles were indexed at DEMIR\_R3. Title of the article joined to indexed contents in DEMIR\_R4. As illustrated in Table1 results shows that caption of images is comprehensive and effective content of metadata to index. On the other hand, although full article has more irrelative content but is more effective than abstract.

### **5.2 Baseline and combined visual-only runs**

As we mentioned in previous sections, CEDD, FCTH and CLD descriptors outperform other low level features. We submitted retrieval result of them as DEMIR\_R5, DEMIR\_R6 and DEMIR\_R7 respectively. Beside these runs, we also investigate on impact of intra-modality combination between these features. We applied our proposed integrated approach on these features in DEMIR\_R8 and DEMIR\_R9. Also we used feature concatenation method on the synchronous feature vectors of all images in data collection and topics and concatenated them as the joint feature vector in DEMIR\_R10 and DEMIR\_R11. Results show that the best performance of this category appears in DEMIR\_R9, where we combined all of these descriptors based on our method, using CombSUM function.



### **5.3 ICMR based runs**

In this category, we examine our claims in ImageCLEF and confirm them again. We performed integrated weighted CombSUM combination such that coefficient of text modality was 1.7 folds of visual modality. Consequently, system performance arise to the best performance in contrast with other fusion methods in same condition. It is obvious that system performance in DEMIR\_R14 and DEMIR\_R15 that combined baseline textual and visual runs in integrated approach using weighted sum, are more batter than performance of DEMIR\_R12 and DEMIR\_R13 respectively. This viewpoint is true also in subject of DEMIR\_R17 compared by DEMIR\_16. In these later runs, we apply two-level integrated combination; in intra-modality level between low level image features and in inter-modality level between result of pervious level as a visual modality (DEMIR\_R9) and baseline run of textual modality (DEMIR\_R1).

### **5.4 Runs on narrowed down dataset**

In this category, we repeated some of submitted runs on new narrowed down and filtered out dataset but any of these new versions of runs did not improve the performance of total system. We submitted this category as DEMIR\_R18 till DEMIR\_R21.

### **5.5 Query expansion with relevance feedback**

To submit runs of this class, firstly we chose three most relevance document from retrieved document set of each base run , if it was existence in top 50 retrieved document set. Then these chosen images were joint to image query set of each topic then run retrieval system again to obtained new result. In our experiments, selection of most relevant document is based on only textual, only visual and both of visual and textual information respectively in DEMIR\_R22, DEMIR\_R23, DEMIR\_R24. Although this manner decrease the performance of system when we consider at textual properties for selection of relevant documents, but it improve the effectiveness of system in visual modality so that DEMIR\_R23 could obtain the best rank of visual runs.

**Table 1.** Runs of DEMIR group in ImageCLEFMed 2012.

Category	Run ID	Rank	Type	modalities or Features	MAP	GMAP	BPREF	P@10	P@30
Textual	DEMIR_R1	6	Textual	Caption	0.1905	0.0531	0.1822	0.3318	0.2152
Textual	DEMIR_R2	10	Textual	Caption , Full Text	0.1865	0.0502	0.1981	0.25	0.1515
Textual	DEMIR_R3	15	Textual	Caption , Abstract	0.1819	0.0679	0.1921	0.2864	0.1909
Textual	DEMIR_R4	16	Textual	Caption , Abstract , Title	0.1814	0.0693	0.1829	0.2864	0.1894
Visual	DEMIR_R5	10	Visual	CEDD	0,004	0,0003	0,0091	0,0364	0,0409
Visual	DEMIR_R6	12	Visual	FCTH	0,0034	0,0003	0,01	0,0318	0,0318
Visual	DEMIR_R7	18	Visual	CLD	0,0026	0,0002	0,0058	0,0318	0,0242
Visual	DEMIR_R8	17	Visual	FCTH + CEDD	0,0027	0,0003	0,0099	0,0045	0,0212
Visual	DEMIR_R9	2	Visual	FCTH + CEDD+CLD	0,0092	0,0005	0,019	0,05	0,0424
Visual	DEMIR_R10	15	Visual	FCTH , CEDD	0,0032	0,0003	0,0111	0,0227	0,0303
Visual	DEMIR_R11	5	Visual	FCTH , CEDD,CLD	0,0064	0,0003	0,0154	0,0409	0,0318
ICMR	DEMIR_R12	17	Mixed	R1, R5	0.1682	0.0478	0.1825	0.3136	0.2061
ICMR	DEMIR_R13	13	Mixed	R1, R6	0.1816	0.0527	0.1912	0.3409	0.2076
ICMR	DEMIR_R14	5	Mixed	R1, R5	0.1954	0.0566	0.2096	0.3455	0.2182
ICMR	DEMIR_R15	4	Mixed	R1, R6	0.2085	0.0621	0.2204	0.3545	0.2152
ICMR	DEMIR_R16	12	Mixed	R9,R1	0.1817	0.0574	0.1997	0.3409	0.2121
ICMR	DEMIR_R17	3	Mixed	R9,R1	0.2111	0.0645	0.2241	0.3636	0.2242

Filter out	DEMIR_R18	34	Textual	R1	0.1447	0.0313	0.1445	0.2864	0.1742
Filter out	DEMIR_R19	8	Visual	R6	0,0042	0,0004	0,0109	0,0409	0,0364
Filter out	DEMIR_R20	4	Visual	R9	0,0066	0,0003	0,0141	0,0318	0,0288
Filter out	DEMIR_R21	18	Mixed	R17	0.1637	0.0349	0.1705	0.2773	0.1758
RF	DEMIR_22	41	Textual	R1	0.1372	0.0466	0.1683	0.3	0.1818
RF	DEMIR_R23	1	Visual	R9	0,0101	0,0004	0,0193	0,0591	0,0439
RF	DEMIR_R24	19	Mixed	R17	0.1589	0.0424	0.1773	0.3136	0.1985

## 6 Conclusion

In this year, we examined the impact of our integrated combination method on different level of multimodality retrieval system and found that proper combination model can improve the performance of multimodal retrieval systems. In order to investigate on possible combinations methods, we first do some evaluations to determine the appropriate low-level features and distance functions. Then we presented a two level integrated combination method for multimodal CBIR systems. We also perform a visual explicit relevance feedback method and examine its effects on visual, textual and mixed runs. Meanwhile we narrowed down data collection by filtering irrelative generic biomedical illustrations out of data collection using visual low level features.

Our finding illustrated that:

- Explicit relevance feedback based on visual similarity of image improves the performance of medical image retrieval system.
- Elimination of noisy and irrelevant images from dataset can progress effectiveness and efficiency of medical image retrieval system.
- It is clear that integrated combination multimodal retrieval improves the performance of result system in all combination level, regardless of inter or intra-modality level.

- In the best combination of textual and visual modalities, weight for textual modality is about 1.7 folds of visual modality weight.

Our study can be extended in several ways, in future. First, it would be good to apply this experimentation results to other medical image collection and verify that our findings produces the similar results on similar CBIR systems. Second, the impact of combination functions on system performance can also be further investigated. Lastly, our study can be extended into other domain of CBIR Systems rather than medical domain.

## 7 References

1. *DEMIR at ImageCLEFMed 2011: Evaluation of Fusion Techniques for Multimodal Content-based Medical Image Retrieval*. **Alpkocak, A. , Ozturkmenoglu, O., Berber, T., Hosseinzadeh Vahid, A., Gachpaz Hamed, R.** s.l. : CLEF (Notebook Papers/Labs/Workshop) 2011, 2010.
2. *Terrier: A High Performance and Scalable Information Retrieval Platform*. **Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C., Lioma, Ch.** 2006. ACM SIGIR'06 Workshop on Open Source Information Retrieval.
3. *An algorithm for suffix stripping*. **Porter, M. F.** 3, 1980, Electronic library and information systems, Vol. 14, pp. 130-137.
4. *IMG(RUMMAGER): AN INTERACTIVE CONTENT BASED IMAGE RETRIEVAL SYSTEM*. **Chatzichristofis, S. A., Boutalis, Y. S., Lux, M.** Prague, Czech Republic. : IEEE Computer Society, 2009. 2nd International Conference on Similarity Search and Applications (SISAP). pp. 151-153.
5. *CEDD: COLOR AND EDGE DIRECTIVITY DESCRIPTOR – A COMPACT DESCRIPTOR FOR IMAGE INDEXING AND RETRIEVAL*. **Chatzichristofis, S. A., Boutalis, Y. S.** Santorini, Greece : s.n., 2008. 6th International Conference in advanced research on Computer Vision Systems (ICVS),. pp. 312-322.
6. *FCTH: FUZZY COLOR AND TEXTURE HISTOGRAM- A LOW LEVEL FEATURE FOR ACCURATE IMAGE RETRIEVAL*. **Chatzichristofis, S. A., Boutalis, Y. S.** Klagenfurt, Austria : s.n., 2008. 9th International Workshop on Image

Analysis for Multimedia Interactive Services (WIAMIS), IEEE Computer Society,. pp. 191-196.

*7. CONTENT BASED RADIOLOGY IMAGE RETRIEVAL USING A FUZZY RULE BASED SCALABLE COMPOSITE DESCRIPTOR. Chatzichristofis, S. A. , Boutalis, Y. S.* 2-3, 2010, Multimedia Tools and Applications, Vol. 46, pp. 493-519.

*8. ACCURATE IMAGE RETRIEVAL BASED ON COMPACT COMPOSITE DESCRIPTORS AND RELEVANCE FEEDBACK INFORMATION. Chatzichristofis, S. A., Zagoris, K., Boutalis, Y. S., Papamarkos, N.* 2, 2010, International Journal of Pattern Recognition and Artificial Intelligence (IJPAI), Vol. 24, pp. 207-244.