DEMIR at ImageCLEFwiki 2011: Evaluating Different Weighting Schemes in Information Retrieval

Tolga Berber, Ali Hosseinzadeh Vahid, Okan Ozturkmenoglu, Roghaiyeh Gachpaz Hamed and Adil Alpkocak

Dokuz Eylul University
Dept. of Computer Engineering, DEMIR Research Group
Tinaztepe, 35160 Izmir, Turkey
tberber@cs.deu.edu.tr, ali_h_vahid@yahoo.com, okan.ozturkmenoglu@deu.edu.tr, ramisa_84@yahoo.com, alpkocak@cs.deu.edu.tr

Abstract. This paper present the participation details of DEMIR (Dokuz Eylul University Multimedia Information Retrieval) research team at ImageCLEFwiki2011. This year we investigate on evaluating of different weighting models on text retrieval performance. In the case of low-level feature selection, we extracted different features and examined their performance to choose the proper feature for our experiments. Thereupon to apply late fusion for best gained result of image and textual features. In these experiments we found that choice of proper weighting model may crucially affect the performance of any information retrieval system and also we found that although linear weighted fusion is simplest and frequently used method. The results clearly show that combining text-based and content-based image retrieval with a proper fusion technique improves the performance.

Keywords: Information Retrieval, Low-level Features, Linear Weighted Fusion, Combination Algorithms & Methods, Feature Extraction and Selection, Late fusion.

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 2011Wikipedia Retrieval task. The main focus of this work is to improve results by examine of different weighting models for retrieved text data and then choose the best low level feature of figures for fusion with text data result. During the combination of text and low level features we check the variation of methods to gain the best result.

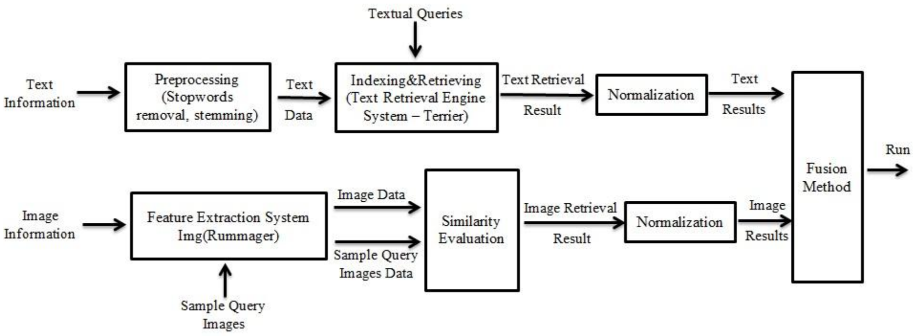

Fig. 1. Basic block diagram of our retrieval system.

The rest of the paper is organized as follows: In Section 2 we describe our text retrieval and weighting models examination. In Section 3 we describe the image features extraction & selection phase. In section 4 we present and compare late fusion combination methods and we conclude and propose the future work in section 5.

2 Different Weighting Models in Text Retrieval

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

Terrier provides implementation of the following weighting models: [1],[2]

- **BB2:** Bose-Einstein model for randomness, the ratio of two Bernoulli's processes for first normalization, and Normalization 2 for term frequency normalization.
- **BM25:** The BM25 probabilistic model.
- **DFR_BM25:** The DFR version of BM25.
- **IFB2:** Inverse Term Frequency model for randomness, the ratio of two Bernoulli's processes for first normalization, and Normalization 2 for term frequency normalization.
- **In_expB2:** Inverse expected document frequency model for randomness, the ratio of two Bernoulli's processes for first normalization, and Normalization 2 for term frequency normalization.
- **In_expC2:** Inverse expected document frequency model for randomness, the ratio of two Bernoulli's processes for first normalization, and Normalization 2 for term frequency normalization with natural logarithm.
- **InL2:** Inverse document frequency model for randomness, Laplace succession for first normalization, and Normalization 2 for term frequency normalization.

- **PL2:** Poisson estimation for randomness, Laplace succession for first normalization, and Normalization 2 for term frequency normalization.
- **TF_IDF:** The *tf*×*idf* weighting function, where *tf* is given by Robertson's *tf* and *idf* is given by the standard Sparck Jones' *idf*.

We performed our experiments using three ImageCLEF2010 Wikipedia track monolingual test collection and using all them together. We start from a traditional bag-of-words representation of pre- processed texts that preprocessing includes stemming (Porter stemmer [3] for English, Snowball for German and French) and stop word removal. As illustrated in Figure2, we acquired that best result in almost all experiment using IFB2 model, so we use it to submit our base-line run on Image CLEF2011 Wikipedia track textual metadata. (RunID 5 in Table1)

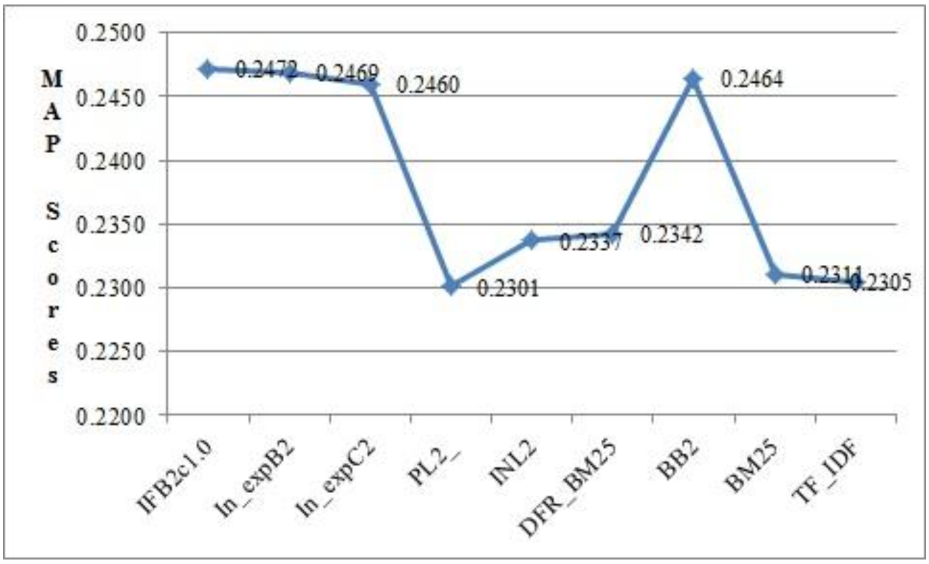

Fig. 2. Comparison of MAP in different weighting schemes retrieved results on ImageCLEF2010 Wikipedia.

3 Low-level feature Extraction

Feature extraction is one of the major aspects of a typical content-based information retrieval (CBIR) system. We call these low-level features because most of them are extracted directly from digital representations of objects in the database and have little or nothing to do with human perception. We utilized the Img(Rummager) application[4], is developed in the Automatic Control Systems & Robotics Laboratory at the Democritus University of Thrace-Greece, and extract features explained below for all images in ImageCLEF2011 test collection and query examples:

- **CEDD:** This feature is called "Color and Edge Directivity Descriptor" and incorporates in histogram color and texture information. CEDD size is limited to

54 bytes per image, rendering this descriptor suitable for use in 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[4]

- **FCTH:** This feature is named "Fuzzy Color and Texture Histogram" and results from the combination of 3 fuzzy systems include histogram, color and texture information. FCTH size is limited to 72 bytes per image, also suitable for use in large image databases.[5]
- **BTDH:** The Scalable Fuzzy Brightness and Texture Directionality Histogram, was specially conceived for representing radiology images. It combines brightness and texture characteristics and their spatial distribution in one compact vector by using a two-unit fuzzy system. [6]
- **EHD:** This Edge Histogram Descriptor proposed for MPEG-7 expresses only the local edge distribution in the image and is designed to contain only 80 bins for this purpose. The EHD basically represents the distribution of 5 types of edges in each local area called a sub-image that is defined by dividing the image space into 4x4 non-overlapping blocks. Thus, the image partition always yields 16 equal-sized sub-images regardless of the size of the original image. Edges in the sub-images are categorized into 5 types: vertical, horizontal, 45-degree diagonal, 135-degree diagonal and non-directional edges. Thus, the histogram for each sub-image represents the relative frequency of occurrence of the 5 types of edges in the corresponding sub-image and contains 5 bins.[7]
- **SCD:** "Scalable Color Descriptor" is one of the most basic descriptions of color a feature is a color histogram encoded by a Haar transform. It uses the HSV colors space uniformly quantized to 255 bins.
- **CLD:** "Color Layout Descriptor" represents the spatial layout of color images in a very compact form. It is based on generating a tiny (8x8) thumbnail of an image, which is encoded via Discrete Cosine Transform (DCT) and quantized. As well as efficient visual matching, this also offers a quick way to visualize the appearance of an image, by reconstructing an approximation of the thumbnail, by inverting the DCT.

After extracting features, we gain an *n*-dimensional feature space per feature. For query processing, we had to mapping all the objects in the database and the query onto this space and then evaluating the similarity/ difference between the vector corresponding to the query and the vectors representing the data. We selected the Euclidean distance, one of commonly used similarity and distance functions for measuring distances between points in the 3D space, as distance/similarity function and based on obtained similarity scores, we found that CEDD and FCTH are the best descriptors for image retrieval based on low level features only. Therefore we submitted our visual only base point run for CEDD feature. (RunID 6 in table 1) Moreover we use these features for multimodal fusion in next experiments, as explain below:

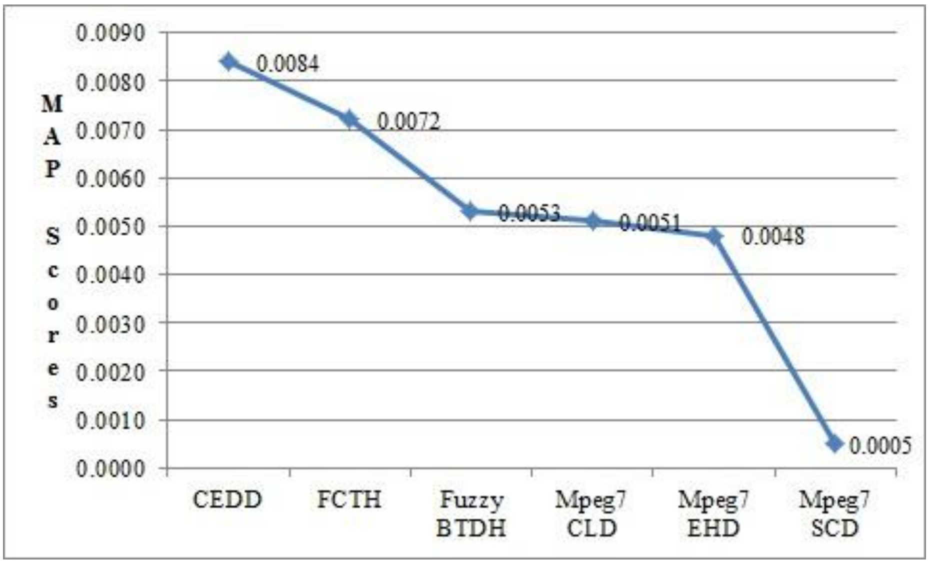

Fig. 3. Comparison of low level feature performance on WikiImageCLEF2010

4 Fusion of multimodality

Multimedia fusion is referred to as integration of multiple media, their associated features, or the intermediate decisions in order to perform an analysis task, has gained much attention of many researchers in recent times. The fusion of multiple modalities can provide complementary information and increase the accuracy of the overall decision-making process [8].

The fusion of different modalities is generally performed at two levels: feature level or early fusion and decision level or late fusion. Some researchers have also followed a hybrid approach by performing fusion at the feature as well as the decision level. In the feature level or early fusion approach, the features, some distinguishable properties of a media stream, extracted from input data are first combined and then sent as input to a single analysis unit that performs the analysis task. In the decision level or late fusion approach, the analysis units first provide the local decisions D_1 to D_n that are obtained based on individual features F_1 to F_n. Then a decision fusion unit combines local decisions to make a fused decision vector that is analyzed further to obtain a final decision D about the task or the hypothesis. To achievement the advantages of both the feature level and the decision level fusion strategies, several researchers have opted to use a hybrid fusion strategy, which is a combination of both feature and decision level strategies.

The decision level fusion strategy has many advantages over feature fusion. For instance, the decisions (at the semantic level) usually have the same representation. Therefore, the fusion of decisions becomes easier. Moreover, the decision level fusion strategy offers scalability (i.e. graceful upgrading or degradation) in terms of the modalities used in the fusion process, which is difficult to achieve in the feature level fusion. Another advantage of late fusion strategy is that it allows us to use the most suitable methods for analyzing each single modality and this provides much more flexibility than the early fusion. Because of these profits, we exerted Linear Weighted Fusion, one of the simplest and most widely used methods on our extracted CEDD and FCTH similarity scores and similarity scores that gained from text retrieval as explained in previous chapters. Since different retrieval results can generate quite different ranges of similarity values, a normalization method should be applied to get accurate and correct results. Hence we apply Max-Min normalization on similarity values to ensure that the range of these features is between 0 and 1. Then we applied Fagin's Combination Algorithms [9] for Ranked Input Sets putting on different combination method and based on our previous study [10], we found that the method called Comb-SUM for summing the similarity values provide the best results. So we combined the CEDD and FCTH features normalized similarity scores with textual similarity score in this manner and submitted two other runs (RunID 3, 4 in Table 1). On the other hand, the Weighted Sum function applied in the same manner but differing on multiplying each individual similarity with a weight value and submitted two the best runs of our experiments. As shown in Table 1, we applied Weighted Comb-SUM combination method with multiplying CEDD feature by 2 and result of retrieved textual feature by 3 and submitted result as RunID 2 in Table1. Also in RunID 1 in Table 1 as our best ranked run, first we use Comb-SUM combination for fused CEDD and FCTH features similarity scores then combined them using Weighted Comb-SUM with two folds of retrieved textual features.

Table 1. Results of our runs at WikiImageCLEF2011.

RunID	Rank	Type	MAP	P@10	P@20	Bpref
1	53	Mixed	0.2432	0.4520	0.3420	0.2564
2	54	Mixed	0.2426	0.4640	0.3460	0.2573
3	56	Mixed	0.2401	0.4340	0.3370	0.2554
4	57	Mixed	0.2394	0.4500	0.3390	0.2529
5	58	Text	0.2369	0.4180	0.3320	0.2476
6	109	Visual	0.0044	0.0340	0.0280	0.0115

So it is apparent when we combine the results of different modalities, all of the performance evaluation factors in retrieval system improved.

5 Conclusion

In this year, we examined effects of different weighting models on text retrieval and found that the role of proper weighting model selection is to improve the performance of text retrieval systems. Also, we compare MAP of different extracted low-level features normalized similarity scores and due to this comparison we select CEDD and FCTH descriptors as suitable features to utilize for fusion to textual results. Also due to analogy of combination methods in our previous studies, we acquire choosing a suitable combination method for fusion improved the results. The results clearly show that combining text-based and content-based image retrieval with a proper fusion technique improves the performance.

References

1. The Terrier IR Platform, http://terrier.org/docs/v2.2.1/
2. Amati, Giambattista: Probability models for information retrieval based on divergence from randomness, PhD Thesis, University of Glasgow Faculty of Information and Mathematical Sciences Department of Computing Science (2003)
3. Porter, M.F.: An algorithm for suffix stripping, Program: electronic library and information systems, vol. 14, iss. 3, pp. 130--137 (1980)
4. Chatzichristofis, S.A., Boutalis, Y.S., Lux, M.: Img(Rummager): An Interactive Content Based Image Retrieval System. In: 2nd International Workshop on Similarity Search and Applications, pp. 151-153. IEEE Computer Society, Washington (2009)
5. Chatzichristofis, S.A., Boutalis, Y.S.: FCTH: Fuzzy Color and Texture Histogram - A Low Level Feature for Accurate Image Retrieval. In: 9th International Workshop on Image Analysis for Multimedia Interactive Services, vol., no., pp.191-196. Klagenfurt, Austria (2008)
6. Chatzichristofis, S.A., Boutalis Y.S.: Content based radiology image retrieval using a fuzzy rule based scalable composite descriptor. Multimedia Tools and Applications, vol. 46, iss. 2, pp. 493--519 (2009)
7. Won C. S., Park D. K., Park S.J.: Efficient Use of MPEG-7 Edge Histogram Descriptor, ETRI Journal, vol. 24, no. 1 (2002)
8. Pradeep K. Atrey, Anwar Hossain M.: Multimodal fusion for multimedia analysis, Multimedia Systems, vol 16, pp. 345--379 (2010)
9. Fagin R, Lotem A, Naor M.: Optimal aggregation algorithms for middleware, In: Journal of Computer and System Sciences, vol. 66, pp. 614--656 (2003)
10. Ulker T.: Analysis and comparison of combination algorithms for joining ranked inputs, MSc Thesis, Dokuz Eylül University Department of Computer Engineering, Izmir, Turkey (2003)