International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Image Retrieval Using Textual Pre-filtering and Visual Re-ranking

Aayushi Gupta[1], Rahul Soni[2], Nikhil Sabale[3], Aman Lad[4].

Abstract: This work shows the improvement inimage retrieval system by fusing the textual pre-filtering results combined with an image re-ranking for Multimedia Information Retrieval task. The goal of this paper is to fuse the text and image retrieval systems efficiently in the context of multimedia information retrieval system for a faster and more accurate result produced by this system.

Index Terms- Content-based information retrieval, multimedia information fusion, multimedia retrieval, textual-based information retrieval.

I. INTRODUCTION

As a result of the different information sources present in the multimedia resources (video, image, audio and text), multimediafusion has become a very interesting field of researchrecently for Information Retrieval (IR) and search in MultimediaDatabases or on theWeb. In the case of imageretrieval, textual and visual features both are usually provided:annotations or metadata (we use annotations only) as textual information, and low levelfeatures (color, texture, etc.) as visual information.[11]

In this paper we are interested in accessing a multimedia collection made of text/image objects or documentsin an efficient way. In order to better depict the context of this research work let us take the example of the Wikipedia collection. Traditional systems rely on text based searches.

The idea behind multimedia fusion is to exploit the individual advantages of each mode, and use different sources as integral information to accomplish a particular search task. In an image retrieval task, themultimedia fusion tries to help solving the semantic gap problem while obtaining accurate results.

There has been many researches in the past addressing text/image information fusion. The intuition behind the technique we are going to introduce is the following: since each media are semantically expressed at different levels, one should not combine them independently since most of information fusion techniques implemented so far do.

Our proposal is an asymmetric multimedia fusion strategy, which exploits the relationship of each mode. The schema consists in a pre-filtering textual step, which semantically reduces the database collection for the visual retrieval, followed by a text and image results fusion phase. Then the results will show how retrieval performance is improved, while the task is made scalable by a significant reduction of the image database collection.

II. PRIOR ART

Regarding text/image retrieval, we generally observe better performancesfortextbasedimagesearchsystemscomparedtocontent based image retrieval (CBIR) systems. However, most of research works in text/image information retrieval have shown that combining text and image information even with simple fusion strategies, allow one to increase multimedia retrieval results.

In many experiments reported in the literature [9], it has been shown that either late fusion or transmedia fusion approaches have been performing better than early fusion techniques.

III. MULTIMEDIA INFORMATION RETRIEVAL

Usually, Multimedia Information Retrieval is addressed from a textual sentiment in most of the existing commercial tools, using annotations or metadata information related with imagesor videos. In this paper we deal with both textual and visual information, carrying out both monomdal and multimodal experiments.

Multimedia fusion tries to use the different media sources as complementary information thus increasing the accuracy of the retrieved results to helpsolving the problem of semantic gap, referred to the difficulty in understanding of information that the user perceives from the low level features of the multimedia data. Particularly, in Image Retrieval case, the semantic gap is the lack of correlation between the information from visual characteristics (e.g., histograms) and the perception of these data by a user in a specific situation (visually similar images to the query can be very different in terms of meaningin terms of low level features).

The early fusion approach is based on the extracted features (visual, text, audio, motion,metadata, etc.) from the different

information sources, and their combination at this level [11]. The main advantages of this level of fusion are the possibilities of using the correlation between the multiple features, and that only one learning phase in the combined feature vector is required. Difficulties are related to the synchronization between the features, and with the need to represent all features in the same format before fusion.

IV. LATE FUSION IN IMAGE RETRIEVAL

In any Image Retrieval task it is widely known that text-based search is usually more efficient than visual-based one. However, it is also known that when it is possible to combine textual and visual information in the correct way, taking advantage of each of the modalities, the combination will be beneficial to multimedia retrieval. Because of the semantic gap problem, obtaining of good results is very difficult for CBIR systems, but "content-based methods can likely improve retrieval results accuracy."

In the Image Retrieval task, where both visual and textual information are available, late multimedia fusion strategies are based on combining the evidence from both the TBIR and CBIR subsystems. These results will be in the form of numerical similarities (scores). Most basic fusion techniques use these scores and combine them by means of aggregation functions. The Late fusion algorithms between text and visual modalities are known to perform better than those of early fusion

A. Text-Based Information Retrieval (TBIR) Sub-System

This sub-system is in charge of retrieving relevant images for a query given by usertaking into account the textual information available in the collection. Various steps are required in order to accomplish this task: information extraction, textual preprocessing, indexation and retrieval. Text-based ranked resultslist of images will be obtained, containing the relevance or score (St) of the retrieved result of images for the concrete query.

B. Textual Information Extraction:

The XML tags extracted in the experiments presented are: the <name> and <caption> for English language.

C. Textual Preprocessing

This component processes the selected text in following three steps:

1) characters with no statistical meaning, like punctuation marks or accents, are eliminated
2) semantic empty words (*stopwords*) from specifics lists are excluded
3) stemmingor derived words to their stem.

Search: Preprocessed topic texts are launched, obtaining the textual (TXT) results list with the retrieved images ranked by their similarity score (St). Depending on this, we will obtain a monolingual result list of images.

D. Content-Based Information Retrieval (CBIR) Sub-System

The CBIR sub-system is in charge of retrieving a list of relevant images taking into consideration the image examples given by the topic. The two main steps of the CBIR sub-system are: the feature extraction and the similarity module. The CBIR sub-system ranks an image result list based on the image score (Si) for each query given by user.

1) *Feature Extraction:* The visual low-level features for all the images in the database for each topic are extracted using the CEDD given by the ImageCLEF2011 organization. The CEDD descriptors, includes more than one feature in a compact histogram (color and texture information) belong to the family of Compact Composite Descriptors. The CEDD structure consists of 6 texture areas. In particular, each texture area is divided into 24 sub-regions, with each sub-region describing a color. They require an enormous amount of training data and lead to huge amount of computing times to create these models. These reasons make them almost impracticable for general CBIR systems.
2) *Similarity module:* The similarity module calculatethe similarity (Si) of each image of the collection to the given query. The algorithm calculates the probability of an image belonging to a set of those images sought by the query, and models the probability as the output of a generalized linear model whose inputs are the visual low-level image features.
3) *Multimedia Information Retrieval (MBIR) Sub-System*

Late Fusion Based on Relevance Scores Product (Si*St):two results lists are fused together to combine the relevance scores of both

Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

textual and visual retrieved images (StandSi). Both these subsystems will have the same importance for the resulting list: the final relevance of the images will be calculated using the Product

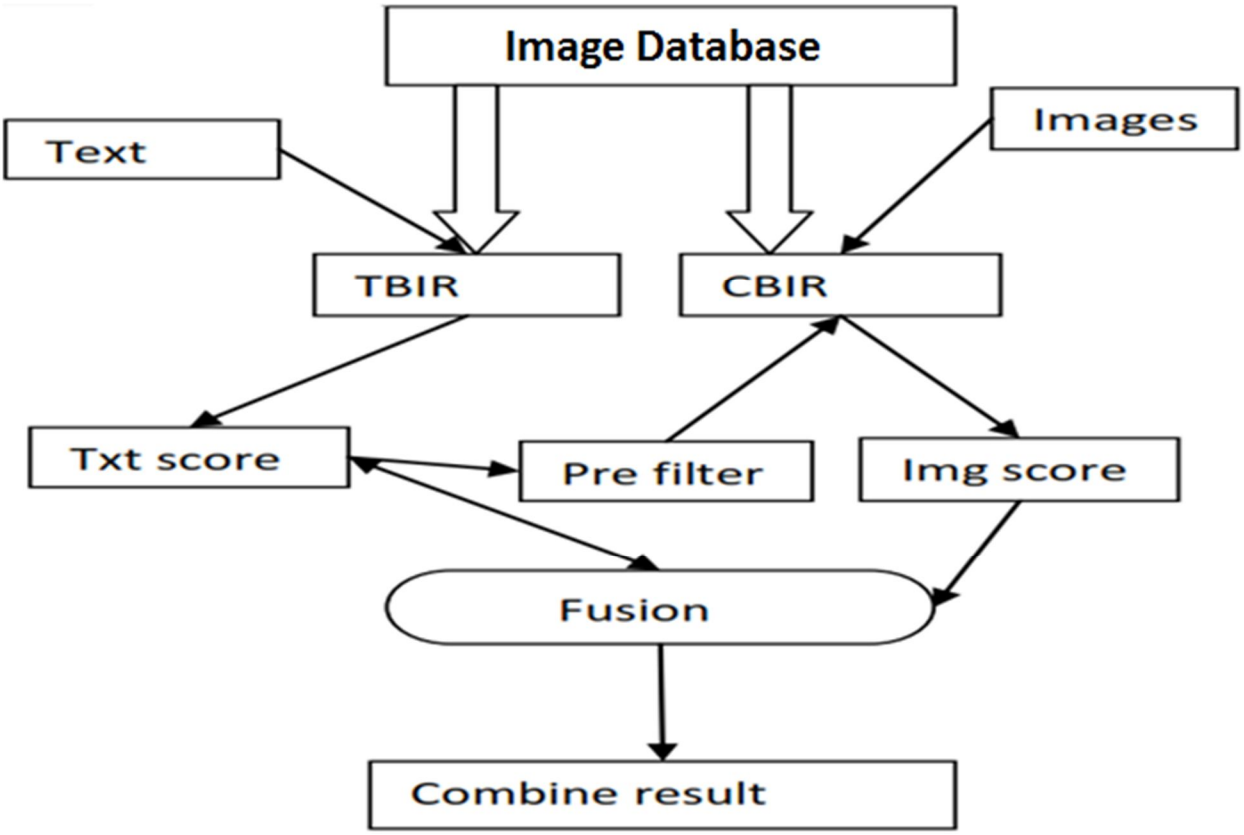

Fig: System Architecture.

V. CONCLUSION

We discussed a detailed description of textual pre-filtering techniques. These textual pre-filtering techniques reduce in a suitable way the size of the multimedia database thus improving the final fused image retrieval results. The combination of textual pre-filtering and image re-ranking in a late fusion algorithm outperforms those without pre-filtering. It seems that textual information better captures the semantic meaning of a topic and hence the image re-ranking fused with the textual scorehelps to overcome the semantic gap. All the performance improvement can be carried out while significantly reducing the complexity of the CBIR process, in regard of both time and computation.

Different media, such as texts and images, are expressed at different semantic levels. And therefore, one modality usually outperforms the other one when accessing a multimedia collection with the help of monomedia search systems. Despite this observation, media are in fact complementary and their aggregation can improve the retrieval performance.

The best performance can been obtained with the Product algorithm in which both modality scores are taken into account with the same importance and then combining to achieve multimedia information retrieval results.

REFERENCES

[1] J. A. Aslamand, M. Montague, "Models for metasearch,"Proc. 24th Annu. Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, LA, USA, 2001, pp. 276–284.

[2] P. K. Atrey,M. A. Hossain, A. El Saddik, andM. S. Kankanballi, "Multimedia Fusion forMultimedia Analysis: A Survey," Multimedia Syst., vol. 16, pp. 345–379, 2010.

[3] S. Clinchant, G. Csurka, and J. Ah-Pine, "Semantic combination of textual and visual information in multimedia retrieval," in Proc. 1st ACM Int. Conference Multimedia Retrieval, New York, NY, USA, 2011.

[4] S. A. Chatzichristofis, K. Zagoris, Y. S. Boutalis, and N. Papamarkos, "Accurate image retrieval based on compact composite descriptors andrelevance feedback information," Int. J. Pattern Recognition. Artif.Intell., vol. 24, no. 2, pp. 207–244, Feb. 2010, World Scientific.

[5] G. Csurka,S.Clinchant, andA. Popescu, "XRCEandCEALIST's Participation at Wikipedia Retrieval of ImageCLEF 2011," in CLEF 2011Working Notes, V.

Petras, P. Forner, and P. Clough, Eds., Amsterdam, The Netherlands, Sep. 2011.
[6] M. Grubinger, "Analysis and Evaluation of Visual Information Systems Performance," Ph.D. thesis, School Comput. Sci. Math., Faculty Health, Engineering, Sci., Victoria University, Melbourne, Australia, 2007.
[7] R. Granados, J. Benavent, X. Benavent, E. de Ves, and A. Garcia-Serrano, "Multimodal Information Approaches for the Wikipedia Collection at ImageCLEF 2011," in Proc. CLEF 2011 Labs Workshop, Amsterdam, The Netherlands, 2011.
[8] J. Kludas, E. Bruno, and S. Marchand-Maillet, "Information fusion in multimedia information retrieval," in AMR Int. Workshop Retrieval,User Semantics, 2007.
[9] "Semantic Combination of Textual and Visual Information in Multimedia Retrieval", StéphaneClinchant, Julien Ah-Pine, Gabriela Csurka, ICMR '11 Trento, Italy.
[10] T. Leon, P. Zuccarello,G.Ayala, E. de Ves, and J.Domingo, "Applying logistic regression to relevance feedback in image retrieval systems," Pattern Recognition, vol. 40, pp. 2621–2632, Jan. 2007.
[11] J. Benavent, X. Benavent, E. de Ves, A. García-Serrano and R. Granados, "Experiences at ImageCLEF 2010 using CBIR and TBIR mixing information approaches," in Proc. CLEF 2010, Padua, Italy.
[12]]"Multimedia Information Retrieval Based on Late Semantic Fusion Approaches: Experiments on a Wikipedia Image Collection",XaroBenavent, Ana Garcia-Serrano, RubenGranados, Joan Benavent,and Esther deVes,IEEEtransactionsonmultimedia,vol.15,no.8,december2013