A Review Paper on Multimedia Information Retrieval based on Late Semantic Fusion Approaches

Harsha J. Kolhe
University of Pune, PG Student,
Department of Computer Engineering,
SITRC College of Engineering,
Nashik-422213

Amitkumar Manekar
Assistant Professor,
Department of Computer Engineering,
SITRC College of Engineering,
Nashik-422213

ABSTRACT
Multimedia information retrieval combines the images and data. Multimedia information retrieval task show the improvement of using textual pre-filtering combined with an image re-ranking. Multimedia fusion has very interesting field for research in recent times for Information Retrieval (IR) and search in Multimedia Databases or on the Web. There are three developed environment to overcome the semantic gap in a given query. There are several different late fusion algorithms to overcome the semantic gap.

Keywords
Content-based information retrieval, multimedia information fusion, multimedia retrieval, textual-based information retrieval.

1. INTRODUCTION
As a result of the different information sources present in a multimedia resource (video, image, audio and text), multimedia fusion has become in a very interesting field of research in recent times for Information Retrieval (IR) and search in Multimedia Databases or on the Web. In the particular case of image retrieval, both textual and visual features are usually provided: annotations or metadata as textual information, and low level features (color, texture, etc.) as visual information. The idea behind multimedia fusion is to exploit the individual advantages of each mode, and use the different sources as complementary information to accomplish a particular search task. In an image retrieval task, multimedia fusion tries to help in solving the semantic gap problem while obtaining accurate results.

Main proposal of this is to present several late semantic fusion algorithms that combine textual pre-filtering with visual re-ranking in order to solve the semantic gap in a Multimedia formation Retrieval (MIR) setting.

2. RELATED WORK
Recently, multimedia fusion has become very interesting field for research. We briefly review related work below.

Multimedia Information Retrieval is usually addressed from a textual point of view in most of the existing commercial tools, using annotations or metadata information associated with images or videos.

In this work we deal with both textual and visual information, carrying out both monomodal and multimodal experiments using different multimedia fusion techniques and algorithms.

Multimedia fusion tries to use the different media sources as complementary information to increase the accuracy of the retrieved results [15], in order to help in solving the semantic gap problem, referred to the difficulty in understanding the information that the user perceives from the low level characteristics of the multimedia data. Specifically, in the case of Image Retrieval, the semantic gap is the lack of correspondence between the information from visual features (e.g., histograms) and the interpretation of these data by a user in a certain situation (visually similar images to the query in terms of low level features can be very different in terms of meaning).

2.1 Problem Analysis
In any Image Retrieval task it is well known that text-based search is usually more efficient than visual-based one [5]. However, it is also known that when it is possible to combine textual and visual information in the correct way, taking advantage of each one of the modalities, the combination will be beneficial to multimedia retrieval [4].

Because of the problem of the semantic gap, the obtaining of good results is very difficult for CBIR systems [5][13], but "content-based methods can potentially improve retrieval accuracy even when text annotations are present by giving additional insight into the media collections".

Within the task of Image Retrieval, where both visual and textual information are available, late multimedia fusion approaches are based on combining the evidence from both the TBIR and CBIR subsystems. These decisions will be in the form of numerical similarities (scores). Most basic fusion techniques use these scores (denoted here in after as St from textual based retrieval and Si from the visual based) and merge them by means of aggregation functions. Late fusion algorithms between text and visual modalities are known to perform better than those of early fusion.

3. PROPOSED ARCHITECTURE
3.1 Architecture Environment
To carry out the experiments, a three-subsystem architecture was developed (Fig 1.): Text Based Image Retrieval), CBIR(Content Based Image Retrieval).

Both the textual (TBIR) and the visual subsystem (CBIR) obtain a ranked list of images based on a similarity scores (St and Si) for a given query.

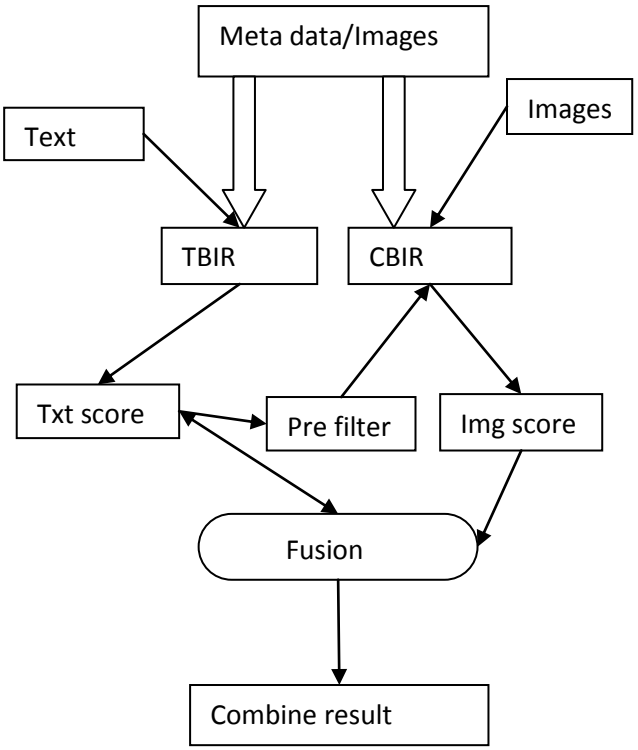

Figure 1 Environment Overview

Firstly, TBIR uses the textual information from the annotations (metadata and articles) to obtain these scores (St). This textual pre-filtered list is then used by the CBIR sub-system.

It extracts the visual information from the given example images of the topic and generates a similarity score (Si). The fusion sub-system is in charge of merging these two lists of results, taking into account the scores and rankings, in order to obtain the final result list.

3.2 Text-based Information Retrieval (TBIR) Sub-System

This sub-system (Fig.2) is in charge of retrieving relevant images for a given query taking into account the textual information available in the collection. Different steps are required in order to accomplish this task: information extraction, textual preprocessing, indexation and retrieval. A text-based ranked results list of images will be obtained, containing the relevance or score (St) of the retrieved images for the concrete query.

Textual information extraction: This component selects the textual information that describes the images coming from both metadata and articles and this information will be separated by language: English, French or Dutch.

Textual Preprocessing: This component processes the selected text by using IDRA tool[11] in three steps,

1. characters with no statistical meaning like punctuation marks or accents are eliminated,
2. exclusion of semantic empty words (stopwords) from specifics lists for each language and
3. stemming or derived words to their stem

Indexation: After preprocessing the textual information the data is indexed using the white space analyzer which just separates the tokens.

Search: Preprocessed topic texts are against launched the index to obtaining the textual (TXT) results list with the retrieved images ranked by their similarity score (st).

3.3 Content-based Information Retrieval (CBIR) Sub-System

The CBIR sub-system (Fig.3) is in charge of retrieving a list of relevant images taking into account the image examples given by the topic.

The CBIR sub-system ranks an image result list based on the image score (Si) for each given query.

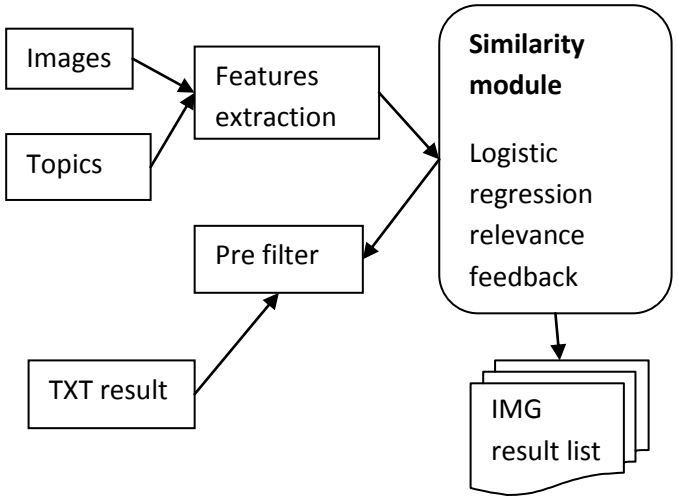

Figure 2 CBIR Subsystem

Feature Extraction: The visual low-level features for all the images in the database for the example images for each topic are extracted using the CEDD[4].

Similarity module: The similarity module uses our own logistic regression relevance feedback algorithm[16] to calculate the Similarity (Si) of each of the images of the collection to the query.

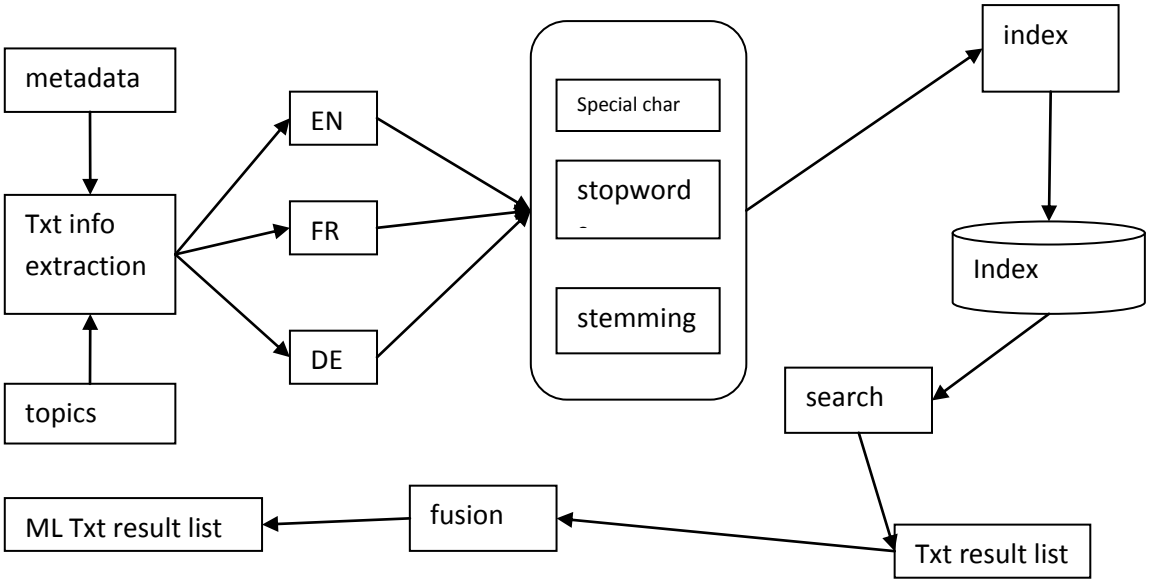

Figure 3 TBIR Subsystem

4. METHODOLOGY

There are several algorithms for Late Fusion approaches like,

Product (st,si): two results lists are fused together to combine the relevance scores of both textual and visual retrieved images (St and Si). Both subsystems will have the same importance for the resulting list: the final relevance of the images will be calculated using the Product. Notice that the Product simulates the filtering when St is 0 (no relevant image for the query), so the image will never appear in the fused list.

Filter N: this algorithm is used to remove from the textual results list those images not appearing in the first N results of the visual list. The idea is to eliminate the images that the visual module is not very sure of; those with a low score Si. This technique will try to clean the textual results based on the visual ones.

5. PERFORMANCE ANALYSIS

The best performance has been obtained with the Product algorithm that means that both modality scores [23] are taken into account with the same importance.

6. CONCLUSION

The detailed description of textual pre-filtering techniques. The textual pre-filtering techniques reduce the size of database and improving the fused list result. It seems that textual information better captures the semantic meaning of a topic and that the image re-ranking fused with the textual score helps to overcome the semantic gap. The developed environment for retrieving the images also studied.

Late fusion algorithms are used and with respect to this algorithm better results are obtained.

7. REFERENCES

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

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

[3] J. Benavent, X. Benavent, E. de Ves, R. Granados, and A. García-Ser-rano, "Experiences at Image CLEF 2010 using CBIR and TBIR mixing information approaches," in Proc. CLEF 2010, Padua, Italy, 978-88-904810-2-4, Notebook papers.

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

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

[6] G. Csurka, S. Clinchant, and A. Popescu, "XRCE and CEA LIST's Participation at Wikipedia Retrieval of Image CLEF 2011," in CLEF 2011 Working Notes, V. Petras, P. Forner, and P. Clough, Eds., Amsterdam, The Netherlands, Sep. 2011.

[7] A. Depeursinge and H. Müller, "Fusion Techniques for Combining Textual and Visual Information Retrieval," in Image CLEF: Experi-mental Evaluation in Visual Information Retrieval. Berlin, Germany: Springer-Verlag, 2010, ch. 6, pp. 95–114.

[8] T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: An experimental comparison," Inf. Retrieval, vol. 11, pp. 77–107, Apr. 2008.

[9] H. Escalante, C. Hernadez, L. Sucar, and M. Montes, "Late fusion of heterogeneous methods for multimedia image retrieval," in Proc. 1st ACM Int. Conf. Multimedia Inf. Retrieval, 2008, pp. 172–179.

[10] E. A. Fox and J. A. Shaw, "Combination of multiple searches," in Proc. 2nd Text Retrieval Conf., 1993, pp. 243–252.

[11] A. García-Serrano, X. Benavent, R. Granados, E. de Ves, and J. Miguel Goñi, "Multimedia Retrieval by Means of

Merge of Results from Textual and Content Based Retrieval Subsystems," in Multilingual Information Access Evaluation II. Multimedia Experiments: 10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009, Corfu, Greece, September 30 - October 2, 2009, Revised Selected Papers. Berlin, Germany: Springer-Verlag, 2010, pp. 142–149.

[12] A. García-Serrano, X. Benavent, R. Granados, and J. M. Goñi-Menoyo, "Some results using different approaches to merge visual and text-based features in CLEF'08 photo collection," in Evaluating Systems for Multilingual and Multimodal Information Access: 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers. Berlin, Germany: Springer-Verlag, 2009, pp. 568–571.

[13] R. Granados, J. Benavent, X. Benavent, E. de Ves, and A. Garcia-Ser- rano, "Multimodal Information Approaches for the Wikipedia Collec-tion at ImageCLEF 2011," in Proc. CLEF 2011 Labs Workshop, Note-book Papers, Amsterdam, The Netherlands, 2011.

[14] M. Grubinger, "Analysis and Evaluation of Visual Information Sys-tems Performance," Ph.D. thesis, School Comput. Sci. Math., Faculty Health, Engi., Sci., Victoria Univ., Melbourne, Australia, 2007.

[15] J. Kludas, E. Bruno, and S. Marchand-Maillet, "Information fusion in multimedia information retrieval," in AMR Int. Workshop Retrieval, User Semantics, 2007.

[16] T.Leon, P. Zuccarello, G. Ayala, E. de Ves, and J. Do mingo, "Applying logistic regression to relevance feedback in image retrieval systems,"Information. Pattern Recog., vol. 40, pp. 2621–2632, Jan. 2007.

[17] M. S. Lew, N. Sebe, C. Djeraba and R. Jain, "Content-based mul-timedia information retrieval: State of the art and challenges," ACM Trans. Multimedia Comp., Commun., Appl., vol. 2, no. 1, pp. 1–19, Feb. 2006.

[18] D. G. Lowe, "Distinctive image features from scale-invariant key-points," International J. Comput. Vision, vol. 60, no. 2, pp. 91–110, 2004.

[19] M. Montague and J. A. Aslam, "Condorcet fusion for improved re-trieval," in Proc. 11th Int. Conf. Inf. Knowledge Manage., McLean, VA, USA, 2002, pp. 538–548.

[20] "ImageCLEF: Experimental Evaluation in Visual Information Retrieval," in The Information Retrieval Series, H. Müller, P. Clough, T. Deselaers, and B. Caputo, Eds. New York, NY, USA: Springer-Verlag, 2010, vol. 32.