ISSN: 2321-8169
4011 - 4014

Late Semantic Fusion Approach for the Retrieval of Multimedia Data

Shweta Satish Kadam
Student
Dept. of Computer Engg.
DGOI, COE, Daund, Pune
Email:ashwet.kdm@gmail.com

Usha Sunil Gound
Student
Dept. of Computer Engg.
DGOI, COE, Daund, Pune
Email: goundusha9@gmail.com

Prof. Tanaji A.Dhaigude
Assistant Professor
Dept. of Computer Engg.
DGOI, COE, Daund, Pune
Email:tanajidhaigude@gmail.com

Avinash Shivaji Gaikwad
Student
Dept. of E & TC Engg.
SCOE, Vadgaon (Bk),Pune
Email:avinash.gaikwad12@gmail.com

Abstract—In Multimedia information retrieval late semantic fusion is used to combine textual pre-filtering with an image re-ranking. Three steps are used for retrieval processes. Visual and textual techniques are combined to help the developed Multimedia Information Retrieval System to minimize the semantic gap for given query. In the paper, different late semantic fusion approaches i.e. Product, Enrich, MaxMerge and FilterN are used and for experiments publicly available ImageCLEF Wikipedia Collection is used.

Keywords-: *multimedia retrieval, textual-based information retrieval, content-based information retrieval, multimedia information fusion*

*****

I. INTRODUCTION

Multimedia data are widely used in all areas. All of us need this data in our day to day life as a personal work or professional work. But there is a big problem to retrieve multimedia information because this multimedia information is stored in a computational representation.

For multimedia retrieval number of techniques are existing out of which some system uses text based retrieval, some uses content based retrieval while some uses both text based and content based retrieval. Also there are large numbers of sources present to retrieve image. So the main task is to retrieve image which is most accurate. In proposed system both Text-based image retrieval (TBIR) and content-based image retrieval (CBIR) is used. TBIR system better captures the conceptual meaning while CBIR system is used to avoid false positives. For TBIR system metadata of Wikipedia and articles are considered as textual information and for CBIR system low level features of image i.e. color, texture etc. are considered as visual information. For TBIR system it uses whole dataset but for CBIR there is reduced database.

Database can be reduced by textual pre-filtering technique. It is observed that results obtained by using textual pre-filtering are much better than without using textual pre-filtering. After getting textual and visual result list, final list is obtained by using fusion technique. For fusion purpose late fusion is used than early fusion. Late fusion i.e. decision level fusion is more advantageous than early fusion in respect of simplicity and scalability. In late fusion each individual decision is obtained from individual feature of an image. In late fusion five different types of algorithms are used. These are Product, OWA operator, Enrich, MaxMerge and Filter N. TBIR subsystem consists of four different steps: Textual information extraction, textual preprocessing, indexation and search. CBIR subsystem consists of two different steps: Feature extraction and Similarity module.

Types of Fusion Techniques:

1. **Early Fusion**[9]: In early fusion approach feature representation of text and image are fused together using Joint features model [2]. Early fusion based on extracted features of information sources and combination of it. Advantage of early fusion approach is the correlation between multiple features and there is only one learning phase [1].
2. **Late Fusion**[1]: In late fusion algorithm the similarity scores are drawn from features of sources. Textual similarity is calculated from textual feature and visual similarity is calculated from visual features. The fusion carried out at decision level calculated from features is called late fusion. And after that some aggregation functions are used to combine these two similarities [2]. Aggregation function include mean average, product etc. Advantages of late fusion are Simplicity, scalability and flexibility [1].
3. **Transmedia Fusion**: The difference between late fusion and transmedia fusion lies in fusion function used. Instead of aggregation process diffusion process is used for fusion. This technique first uses one of the modalities and retrieve relevant documents and then to switch to the other modality and aggregate their results [5].

II. LITERATURE SURVEY

In this section we have reviewed the papers given in the references section.

1. In [6] author proposed metasearch model based on an optimal democratic voting procedure, the Borda Count and based on Bayesian inference and also investigated a model which obtains upper bounds on the performance of metasearch algorithms.
2. In [4] author presented experiments in ImageCLEF 2010 Campaign. Author assumes that textual module better captures the conceptual meaning of a topic. So that, the TBIR module works firstly and acts as a filter for CBIR, and the CBIR system starts working by reordering the textual result list. The CBIR system presents three different algorithms: the

 ISSN: 2321-8169
4011 - 4014

automatic, the query expansion and a logistic regression relevance feedback.

3. In [2] author proposed different techniques i.e. author semantically combines text and image retrieval results to get better fused result in the context of multimedia information retrieval. Using these techniques some observations are drawn that image and textual queries are expressed at different acceptable levels and that an only image query is often unclear. Overall, the semantic combination techniques overcome a conceptual barrier rather than a technical one: In these methods there is combination of late fusion and image reranking and also proposed techniques against late and cross-media fusion using 4 different ImageCLEF datasets.

4. In [11] author introduced a new task i.e. ImageCLEF 2009 Campaign used to retrieve photo. Author proposed an ad-hoc management of the topics delivered, and also generates different XML files for large number of caption of photos delivered. For this two different merging algorithms to merge textual and visual results were developed. Author's best run is at position 16th, in the 19th for MAP score of performance metrics, at position 11th, for a total of 84 submitted experiments of diversity metrics.

In [7] author gave an overview of different features used in content-based image retrieval and compares them quantitatively on four different tasks: stock photo retrieval, personal photo collection retrieval, building retrieval, and medical image retrieval. Five different available image databases are used for this experiments and the performance of image retrieval is investigated in detail. Due to this comparison of all features is possible and in future possibility of comparison of newly proposed features to these features.

III. PROPOSED METHOD

A. *Architecture Description*

Proposed system consists of three sub-systems i.e. TBIR (Textual based information retrieval), CBIR (Content based information retrieval) and fusion.

TBIR takes textual information from annotations and metadata and derive score S_t and CBIR takes information from low level features of an image and derive score S_i. CBIR used pre-filtered list so that database is minimized. CBIR works on that reduced database Fusion algorithms are used to merge these two results to obtain final result list. For fusion there are four different algorithms. These are product, MaxMerge, enrich and filter N.

B. *Text-Based Information Retrieval(TBIR) Sub-System*

By using textual information this subsystem retrieves relevant images for given query.There are four steps in TBIR including textual information extraction, textual preprocessing, indexation and retrieval.

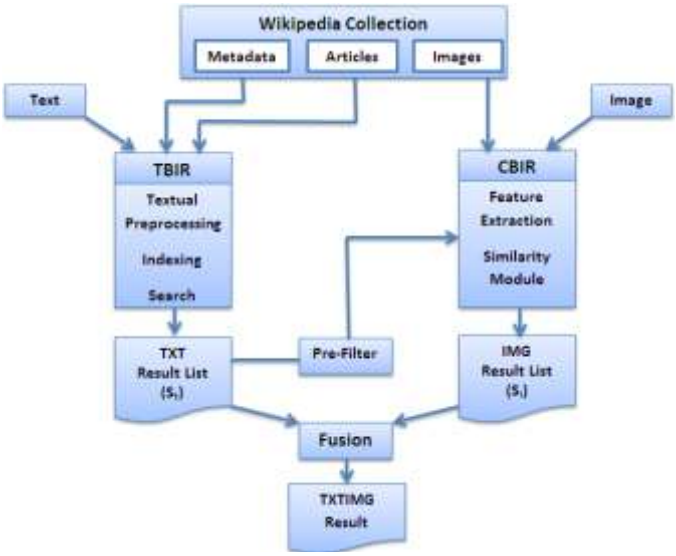

Fig. 1: Product Overview

1) Textual Information Extraction:

Metadata and articles are the sources to extract textual information. The XML tags like <name>, <caption> , <comment>, and <description> can be used as metadata for TBIR. From articles <title> and <category> fields are used to extract textual information

2) Textual Preprocessing:

It includes 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.

3) Indexation:

The extracted textual information is indexed using Lucene Library. Lucene library is high performance, full featured text search engine library written in Java.

4) Search:

In this step textual result list is obtained with retrieved images ranked by their similarity score (St).

C. *Content-Based Information Retrieval(CBIR) Sub-System*

CBIR sub-system uses image examples from topic and pre-filtered list to give more accurate result. From that it calculates relevance score (Si). Two steps are used for retrieval:

1) Feature Extraction:

Scale Invariant Feature Transformation (SIFT) is used to extract features from an image. SIFT is used to transforms an image into a "large collection
of local features vectors". SIFT is invariant to scale, rotation and translation i.e. feature vector doesnt change even image is scaled out or rotated or moved or translated from one position to another position.

2) Similarity Module:

To calculate similarity (Si) of each of images of the collection to the query, logistic regression relevance feedback algorithm is used.

 ISSN: 2321-8169
4011 - 4014

D. Fusion of TBIR and CBIR:

Four different algorithms are used to fuse to result list obtained from TBIR and CBIR.

1) Product:

The two result lists i.e. textual and visual are combined to obtain final result list. For that it takes product of relevance scores St and Si to calculate new relevance score.

If St has zero value i.e. image is not relevant then it does not appear in final list because St*Si=0.

2) Enrich:

There are two lists main list (textual result list) and support list (visual result list). If any image appears in both lists then relevance score of image in fused list can be calculated using formula

newrel = mainrel + suprel / (posrel+1)

where newrel is the relevance value in a fused list, mainrel is the relevance value in main list (textual), suprel is the relevance value in support list (visual) and posrel is position of image in support list.

3) MaxMerge:

This algorithm forms a final fused list by selecting highest relevance value for particular query from both lists independently.

4) FilterN:

This algorithm is used to remove images from textual list which are not appearing in first N results of visual mode. That means this algorithm eliminates images which are having low Si. This algorithm will clean textual results based on visual ones.

IV. RESULT

Database contain the set of records which are relevant to search topic. Records are assumed to be relevant or irrelevant. The actual retrieved set by using project must match with the set of relevant records. The performance of algorithm is carried out by using ”Precision” term. Precision means the probability hat a (randomly selected) retrieved document is relevant. It is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. It is usually expressed as a percentage.

After processing image using SIFT, extracted features are as follows:

Fig. 2: Image after SIFT feature extraction

SIFT features are represented in the form of XY Coordinates, scale and orientation as shown in following table, for above selected Test Image, 99 features are identified as shown below.

TABLE I: SIFT Features: XY Coordinates, Scale and Orientation

Sr.No.	X	Y	Scale	Orientation
1	108.446625	191.08086	1.2287662	6.5395913
2	18.3874	52.825363	6.441095	-2.882178
.	.	.	.	.
99	113.48065	76.63109	1.8838316	-0.618232

V. CONCLUSION

In this paper to retrieve more accurate image from database both Text based and content based information retrieval are used. TBIR works on whole database whereas CBIR uses reduced database. For this CBIR uses Textual pre-filtering method. Due to pre-filtering results are much more improved. Fusion algorithms are used to fuse these two relevant scores obtained from TBIR and CBIR. For fusion purpose different algorithms are used.

ACKNOWLEDGMENT

I would like to take this opportunity to express our profound gratitude and deep regard to my Project Guide Prof. Dhaigude T.A, for his guidance, support, advice, valuable feedback and constant encouragement throughout the duration of the project. His valuable suggestions were of immense help throughout my project work. Working under him was an extremely knowledgeable experience for me.

REFERENCES

[1] Xaro Benavent, Ana Garcia-Serrano, Ruben Granados, Joan Benavent, and Esther de Ves Multimedia Information Retrieval Based on Late Semantic Fusion Approaches: Experiments on a Wikipedia Image Collection, in Computer Science Department, Universidad de Valencia, Valencia 46022, Spain, 2013.

[2] 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.

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

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

[5] A. Garcia-Serrano, X. Benavent, R.Granados, and J. M.Goni-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: 9thWorkshop 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. 39

 ISSN: 2321-8169
4011 - 4014

[6] A. Popescu, T. Tsikrika, and J. Kludas Overview of the Wikipedia retrieval task at ImageCLEF 2010, in Proc. CLEF 2010 LabsWorkshop, Notebook Papers, Padua, Italy, 2010.
[7] T. Deselaers, D. Keysers, and H. Ney Features for image retrieval: An experimental comparison, Inf. Retrieval, vol. 11, pp. 77 107, Apr. 2008.
[8] Y. Rui, S. Huang, M. Ortega, and S. Mehrotra Relevance feedback: A power tool for interactive content-based image retrieval, IEEE Trans. Circuits Syst. Video Technol., vol. 8, no. 5, Sep. 1998.
[9] 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, Notebook Papers, Amsterdam, The Netherlands, 2011.
[10] D. G. Lowe Distinctive image features from scale-invariant keypoints, International J. Comput. Vision, vol. 60, no. 2, pp. 91 110, 2004.
[11] A. Garcia-Serrano, X. Benavent, R. Granados, E. de Ves, and J. Miguel Goni 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] E. A. Fox and J. A. Shaw Combination of multiple searches, in Proc. 2nd Text Retrieval Conf., 1993, pp. 243 252.
[13] M. Grubinger Analysis and Evaluation of Visual Information Systems Performance, Ph.D. thesis, School Comput. Sci. Math., Faculty Health, Engi., Sci., Victoria Univ., Melbourne, Australia, 2007.
[14] T. Leon, P. Zuccarello, G.Ayala, E. de Ves, and J. Domingo Applying logistic regression to relevance feedback in image retrieval systems, Pattern Recog., vol. 40, pp. 2621 2632, Jan. 2007.
[15] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain Content-based multimedia information retrieval: State of the art and challenges, ACM Trans. Multimedia Comp., Commun., Appl., vol. 2, no. 1, pp. 19, Feb. 2006.