Late Semantic Fusion Approaches for Multimedia Information Retrieval with Automatic Tag Generation

Pooja Shinde
P.G.Student
Department of Computer Engineering
Late G.N.Sapkal College of Engineering,
Anjaneri, Nasik

J.V. Shinde
Assistant Professor
Department of Computer Engineering
Late G.N.Sapkal College of Engineering, Anjaneri,
Nasik

ABSTRACT

Image retrieval in general and content based image retrieval in particular are well known research fields in information retrieval management. An image contains several types of visual information which are difficult to extract and combine manually by humans. The main goal of this paper is to show multimedia information retrieval task using the combination of textual pre-filtering and image re-ranking. The combination of textual and visual techniques and retrieval processes used to develop the multimedia information retrieval system by which we solves the problem of the semantic gap of the given query. Five late semantic fusion approaches are used for text based and content based image retrieval of any dataset. The logistic regression relevance feedback algorithm is used to determine the similarity between the images from the dataset to the query.

General Terms

Algorithm, Performance, Experimentation, Design.

Keywords

Multimedia information fusion, multimedia retrieval, tag recommendation, late fusion,content based image retrieval.

1. INTRODUCTION

With the advances in the different computer technologies and the development in the world wide web, there has been an explosion in the amount and the complexity of the digital data being generated, stored, transmitted, analyzed, accessed, searching is performed based on different parameters [1].The various Multimedia based Information Retrieval Systems (MIRS) is the system which is able to store, retrieve, maintain, update and manage the information in the database. In Multimedia based information retrieval system, there is somewhat difficulty in the communication between an information or images, users and the image retrieval system.

The users may have different needs and knowledge about the images collection in database and an image retrieval system, it must support different forms of query formulation in the system. When an image is observed there can be problem of semantic gap which defines gap between the image in the real world and the information consist in a description which is resulting from a recording of that different scene for different image. That is, when an image is provided to different dataset, from the image some information was present in the real world is automatically missing. So, the loss of information can be due to the missed details, bad clarification or different viewing angles or any type of imperfectness of an image capturing device like a camera.

In this paper we discussed, Semantic filtering method that seeks to enhance the similarities between multimedia items when they are composed of both a visual and a textual part. We discuss how such a filter based approach works on the text query can better cope with the semantic gap in the case of MBIR system. We propose to use this approach as a first level of the fusion process of visual and textual information in our multimedia relevance model and automatic tag are also provided to user for uploaded images.

The global system includes three main subsystems: the TBIR, the CBIR and the merging module i.e. MBIR. The TBIR subsystem uses the own implemented tool IDRA, in charge of indexing and retrieving textual annotations from images. We perform automatically tag recommendation to the system. Social tagging becomes increasingly important in large scale community contributed photos on social websites. We will try to improve the result to facilitate generating high quality social tags, tag recommendation by automatically assigning relevant tag to a photo draws a particular research interest.

2. LITERATURE REVIEW

Multimedia based information retrieval of the different text, images, audio, and video are may be the best developed technology. As early in the 1986, image databases were being developed and deployed the system as UC Berkeley's Query Image on system, so Developers and researchers believe that , this software was the first deployed multiuser based networked digital image database system in the environment.[3] In the early times, multimedia based fusion techniques had gained much more attention of many researchers due to the benefit it has provides for various multimedia based information analysis tasks. One of the earliest considerations is to decide what strategy to follow when fusing multiple modalities. The most widely used strategy is to fuse the information at the different feature level, which is also known as early fusion strategy.[4][5] The other approach is decision level fusion or late fusion which fuses multiple modalities in the semantic space. [6]

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, 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).When multimedia approaches are used, several aspects have to be taken into account in order to select the most appropriate.

The Image CLEF 2011 based Photo Annotation and Conceptbased Retrieval Tasks The experiments have been carried out using the Wikipedia collection at Image CLEF 2011 that it contains almost 240 thousand socially annotated images in their different Wikipedia articles (also provided), 50 multimedia based topics and the related relevance judgments.[7][8] Both task differentiate among approaches that consider solely visual information, approaches that rely only on textual information in form of image metadata and user tags, and multi-modal approaches that combine both information sources. Results indicate that while the annotation task shows promising results, the conceptbased retrieval task is much harder to solve, especially for specific information needs.

Semantic Combination of Textual and Visual Information in Multimedia Retrieval In this paper they discuss about the combination of textual and visual information in retrieval system. In this paper we are interested in accessing in an efficient way a multimedia collection made of text/image objects or documents. In order to better depict the context of this research work let us take the example of the Wikipedia collection. They proposed semantic combination models for textual and visual information fusion can significantly improve multimedia retrieval performances while being more robust with respect to the choice of the mixing parameter. The problem of the optimal weight assignment to the different modalities under a varying context is an open problem. Since they usually have different confidence levels in the different modalities for accomplishing various analysis tasks, the problem of dynamic computation of the confidence information for the different streams for various tasks, becomes challenging and worth researching in future.

3. PROPOSED SYSTEM

The TBIR subsystem acts firstly over the whole images of the database, acting as a filter to the CBIR system selecting the relevant images for a certain query.

In a second step, the CBIR system works over the set of filtered images reordering this list taking into account the visual information of the image.

The CBIR system generates different visual result lists depending on the number of query Images.

3.1 Text-based Index and Retrieval

This module is in charge of the textual image retrieval using the metadata supplied for the images in the collection. IDRA tool extracts, selects, preprocesses and indexes the metadata information, for later search and retrieve the most relevant images for the queries. After this process, a ranked results list is obtained for each textual experiment.

3.2 Content Based Information Retrieval

Content-based image retrieval, a technique which uses visual contents to search images from large scale image databases according to users' interests, has been an active and fast advancing research area In this case of retrieval system we need to focus on the metadata of the given objects.[9] Here it requires having an index prepared like structure which will hold the metadata related with all the objects or images in
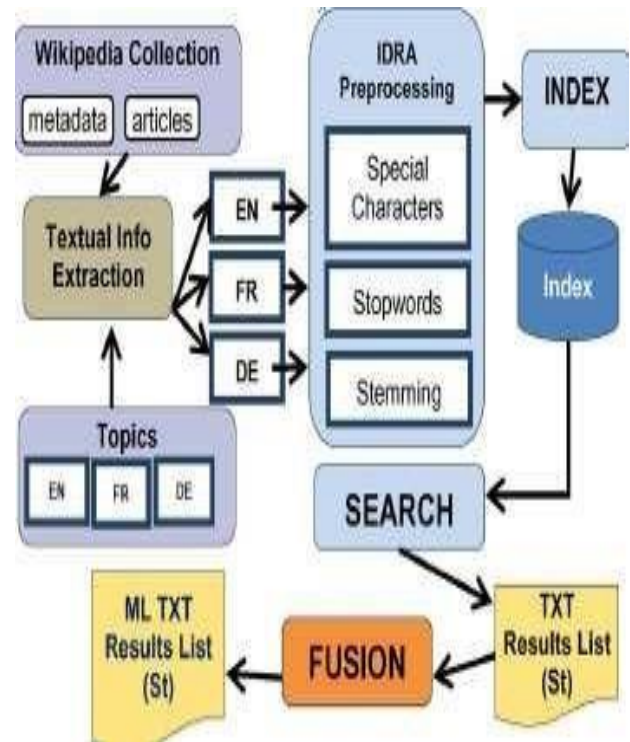
case of our proposed system.

Fig1: Text-based Index and Retrieval

3.2 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 given by the Image CLEF 2011 organization. The CEDD descriptors, which include more than one feature in a compact histogram (color and texture information), belong to the family of Compact Composite Descriptors. [10][11]The structure of CEDD consists of 6 texture areas. In particular, each texture area is separated into 24 sub-regions, with each sub-region describing a color.

3.3 Similarity Module

The similarity module uses our own logistic regression relevance feedback algorithm to calculate the Similarity (S_i) of each of the images of the collection to the query. [13]The algorithm calculates the probability of an image belonging to a set of those images sought by the query, and models the log it of this probability as the output of a generalized linear model whose inputs are the visual low-level image features.
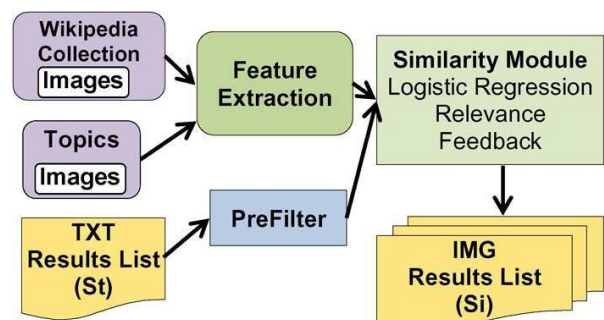

Fig 2: Content based image retrievals

3.4 Flow of our proposed system will be as follows:

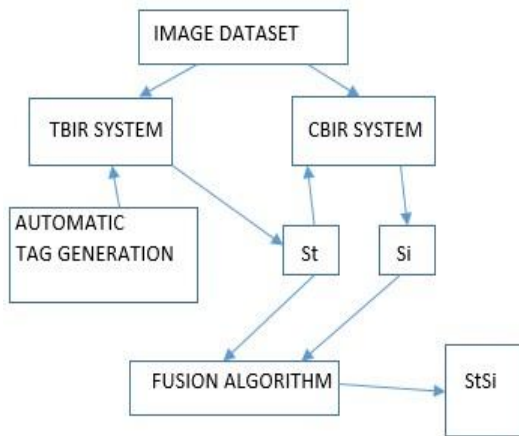1. Read the query and extract the keywords.

Fig 3: Proposed Architecture

3.5 Add Image To Data Set

1. Browse the specify image from file system.

2. Specify a title and description for selected image.

3. Provide search keyword for selected image.

4. Read basic low level information from selected image using CEDD7 algorithm. (The CBIR subsystem uses the CEDD7 and Si from Content based search.

5. While calculating the result of tag based system automatically tags are generated based on Geolocation for better result. Then calculate the Product of St and Si then calculate the orness (w), and then find the new Relevance score.

The Late fusion ranking algorithm filter N algorithm is applied to the fused list.

4. ALGORITHMIC STRATEGY

Data Set Manager

In this section, we are going manage (Add New Image, Delete image , Update , Auto generate new tag for image, Modify images , Searching , Search image based on different parameters) Relevance feedback algorithm to get the score (Si) for each Image).

1. It is used calculate the keyword position, no of keyword, and Keyword density.

2. Store all image information to db.

3. For more image, repeat step 1.

4.1 View All Image From Data Set

1. Load all image info from db.

2. Display the data into table.

3. User can delete an image from table.

4.2 Search

In this section, user can search any images which is related to their search type. There are 2 types of search 1. Filename based search and 2. Annotation or Tag based search.

4.4 Filename based search

1. User enter a search word.

2. System will send the keyword and verify the keyword in Filename field.

3. If any image contains the keyword on their filename, not in Image content.

4. Display the verified images on result.

4.5 Annotation or Tag based search

1. User enter a search word.

2. System will send the keyword and verify the keyword in Search tag field.

3. If any image contains the keyword on their tags, not in Filename and content.

4. Display the verified images on result.

4.6 Content based image search

1. User upload the image to search.

2. System will send the image and verify the content of image.

3. CBIR use the result of TBIR result list as input.

4. Display the verified images on result.

4.7 Multimedia based image search

1. Take both the image list (TBIR CBIR) as input.

2. System will work on result list obtained from TBIR and CBIR system and apply semantic fusion algorithm to obtain result.

3. Display the fused image result list.

5. RESULT AND ANALAYSIS

The image should be retrieve as soon as user will enter a keyword for searching any image. Therefore need arise to calculate the ranking of given keyword which is then use with respect to image searching technique. Along with TBIR system user also upload image for searching image all features of image are compared along with query image i.e. CBIR system and finally both the result list are compared to generate final multimedia image list is generated.

The proposed method is compared with the semantic fusion algorithms. Comparison with TBIR and CBIR in the application area of Annotation-Based Image Retrieval with Precision versus Recall diagrams on Wikipedia collection at Image CLEF Database tell that the proposed approach achieves better retrieval scores. In this paper use Wikipedia collection at ImageCLEF Database and calculate the precision and recall for all database images. The comparison between existing and proposed system is performed on the basis of ranking of the retrieved result.

Precision: The precision is defined as the ratio of the number of relevant images retrieved and total number of images in the collection.

Recall: Recall rate is defined as the ratio of number of relevant images retrieved and to the total number of relevant images in the collection.

The figure (4,5) shows precision vs. recall comparison

between MBIR system for ranking of images for the classes of the Wikipedia Collection at ImageCLEF2011 databases.

6. CONCLUSION AND FUTURE SCOPE

The proposed application is to develop an image retrieval application which can perform identity check of an image. The objective is to progress towards the user fulfillment by returning images that have a higher probability to be accepted, meaning is that those images will be highly relevant to the users query. Each feature defines a multidimensional space where images are points, and the similarity between images is computed as the distance between points. Therefore the main goal of this work is to provide detailed description and analysis of textual prefiltering techniques. These textual prefiltering techniques reduce in a suitable way the size of the multimedia database improving the final fused retrieval results. Experiments show that the combination of textual pre-filtering and image re-ranking lists in a late fusion algorithm outperforms those without pre-filtering. 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. Notice that all this performance improvement is carried out while significantly reducing the complexity of the CBIR process, in terms of both time and computation. With respect to the late fusion algorithms analyzed, better results are obtained with those that work only with the value scores than others, which rely on the ranked positions.
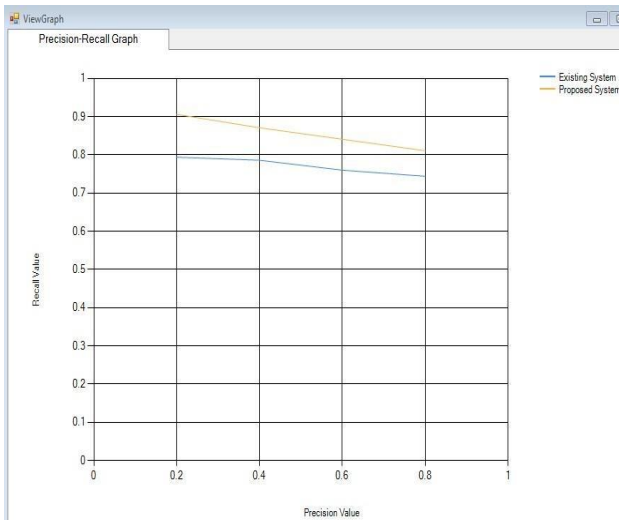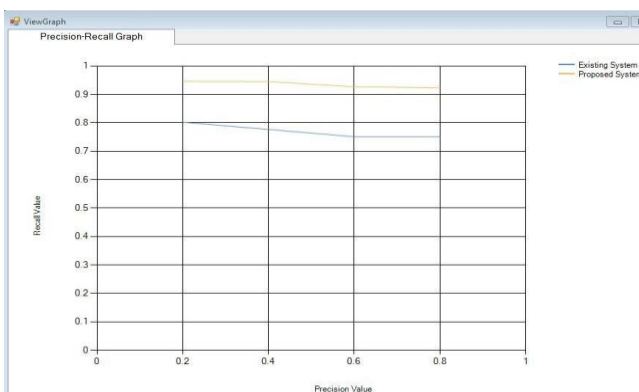

Fig4: Result for MBIR System: African

Fig 5: Result for MBIR System: Animal.

7. REFERENCES

[1] Xaro Benavent, Ana Garcia-Serrano, Ruben Granados, Joan Benavent, and Es- ther de Ves,*"Multimedia Information Retrieval Based on Late Semantic Fusion Ap-proaches: Experiments on a Wikipedia Image Collection"*, *IEEE* transactions on multimedia, vol.15.

[2] S. Santini and R. Jain, *"Semantic combination of textual and visual information in multimedia retrieval"*,*IEEE* Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 9, pp. 871-883, Sept. 1999.

[3] J. A. AslamandM. Montague *"Models for metasearch "*, Proc. *IEEE* Intl Conf. Multimedia and Expo, July 2005.

[4] B.J. Jansen, A. Spink, and T. Saracevic, *"Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web"*, Information Processing and Management, vol. 36, no. 2, pp. 207-227, 2000.

[5] R. Datta, D. Joshi, J. Li, and J.Z. Wang, and E.L. Miller, *"Image Retrieval: Ideas, Influences, and Trends of the New Age"*ACM Computing Surveys, vol. 40, no. 2, pp. 1-60, 2010.

[6] D. Joshi, J.Z. Wang, and J. Li ,*"Semantic combination of textual and visual information in multimedia retrieval "*, Proc. ACM 14th Ann. Intl Conf. Multimedia, 2006.

[7] J. Li and J. Wang ,*"Real-Time Computerized Annotation of Pictures"*, Proc. ACM 14th Ann. Intl Conf. Multimedia, 2006.

[8] T. Hofmann, *"Probabilistic Latent Semantic Indexing"*, Proc. 22nd Intl Conf. Research and Development in Information Retrieval (SIGIR 99), 1999.

[9] T. Hofmann ,*'Unsupervised Learning by Probabilistic Latent Semantic Analysis"*, Machine Learning, vol. 42, no. 1/2, pp. 177- 196, 2001.

[10] D.M. Blei and A.Y. Ng, and M.I. Jordan,*"Latent Dirichlet Allocation"*,J. Machine Learning Research, vol. 3, pp. 993-1022, 2003.

[11] A. Popescu, T. Tsikrika, and J. Kludas, *"Overview of the wikipedia retrieval task at ImageCLEF 2010"*,Proc. Natl Academy of Sciences USA, vol. 101, no. suppl. 1, pp. 5228-5235, 2009.

[12] 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.

[13] P. K. Atrey,M. A. Hossain, A. El Saddik, andM. S. Kankanballi,*Multimodal Fusion for Multimedia Analysis: A Survey,* in Multimedia Syst., vol. 16, pp. 345 379, 2010.

[14] R. Granados, J. Benavent, X. Benavent, E. de Ves, and A. GarciaSerrano, *Multimodal Information Approaches for the Wikipedia Collection at ImageCLEF 2011,* in Proc. CLEF 2011 Labs Workshop, Notebook Papers, Amsterdam, The Netherlands, 2011.

[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. Tsikrika, A. Popescu, and J. Kludas, *Overview of the Wikipedia image retrieval task at ImageCLEF 2011,* in

Proc. CLEF 2011 Labs Workshop, Notebook Papers, Amsterdam, The Netherlands, 2011

[17] CHATZICHRISTOFIS, S. A., BOUTALIS, Y. S., LUX,*"An interactive content based image retrieval system. In Second International Workshop on Similarity Search and Applications SISAP 2009"*. in Prague (Czech Republic), 2009, p. 151 153.

[18] 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.

[19] M. Steyvers, P. Smyth, M. Rosen-Zvi, and T. Griffiths,*"Probabilistic Author-Topic Models for Information Discovery"*,Proc. 10th ACM SIGKDD Conf. Knowledge Discovery and Data Mining, 2004.

[20] Z. Guo, S. Zhu, Y. Chi, Z. Zhang, and Y. Gong, *"A Latent Topic Model for Linked Documents"*, Proc. 32nd Intl ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2009.

[21] T.-T. Pham, N.E. Maillot, J.-H. Lim, and J.-P. Chevallet,*"Latent Semantic Fusion Model for Image Retrieval and Annotation"*, Proc. 16th ACM Conf. Information and Knowledge Management (CIKM), 2007.

[22] J. Fan and Y. Gao, and H. Luo, *"Integrating Concept Ontology and Multitask Learning to Achieve More Effective Classier Training for Multilevel Image Annotation"*,*IEEE* Trans. Image Processing, vol. 17, no. 3, pp. 407-426, Mar. 2008.