ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 4, July 2014

MIRS: Text Based and Content Based Image Retrieval

Trupti S. Atre[1], K.V.Metre[2]

Abstract*— 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 solves the problem of the semantic gap of the given query. Five late semantic fusion approaches are discussed, which can be 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.

***Index Terms*— Content Based Image Retrieval, Late Fusion, Multimedia Information Retrieval, Text Based Image Retrieval.**

I. INTRODUCTION

The general objective of an Information Retrieval System is to minimize the overhead of a user locating needed information. The Multimedia Information Retrieval System (MIRS) is that which is able to store, retrieve, and maintain the information. Information in this context can be collected of text that can be numeric and date data, images, audio and video. In MIRS, there is difficulty in the communication between an information/image seeker/user and the image retrieval system. The user may have differing needs and knowledge about the image collection and an image retrieval system must support various forms for query formulation.

When an image is recorded there can be problem of sensory gap which defines the "gap between the object in the real world and the information in a description resulting from a recording of that different scene for an image". That is, when an image is provided to the dataset, from the image some information was present in the real world is automatically missing. So, this loss of information can be due to missed details, bad clarification or viewing angles or any imperfectness of the image capturing device like a camera. Another problem is the semantic gap which defines the "lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data for a user in a given situation". This reflects on the variation between the visual low-level features exhibited by an image and the semantic like objects, relationships, meanings and abstract of that image as perceived by a human. And other divides the semantic gap into the gap between the visual descriptors and the object levels and the gap between the labeled objects and the full semantics of an image [6]. The system uses the textual pre-filtering and image re-ranking for solving the problem of semantic gap in multimedia information retrieval. And also it uses the five different late fusion algorithms [1] that are Product, OWA operators, Enrich, MaxMerge and Filter N.

II. RELATED WORK

Multimedia information retrieval is retrieval of the text, image, audio, and video which may be the best developed technology or system. As early in 1986, image databases were being developed and deployed the system such as UC Berkeley's ImageQuery system, so "developers believe that this software was the first deployed multi-user networked digital image database system in the environment". Image data retrieval has time to grown-up. The system has allowed the area to address some difficult issues: image classification, query matching, image standards, attribute classification, and evaluation [8]. An image retrieval system has issues for audio and video retrieval systems as well. The Classification and querying also apply to the other forms of media which is to be used, although the media's unique properties necessitate different classification and query matching algorithms for each it used.

Multimedia Information Retrieval (MIR) is required for a vast multimedia data which is captured and stored, the special characteristics and requirements are significantly different from alphanumeric data. Text Based Information Retrieval (TBIR) and Content Based Information Retrieval (CBIR) have limited capacity to handle multimedia data effectively. The system also uses combination rules that are combMAX is the maximum

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 4, July 2014

combination, combMNZ is the product of maximum and non-zero numbers and combSUM is the sum combination [2]. It also used the strategy is to fuse the information at the feature level, which is early fusion and The other approach is decision level fusion or late fusion which fuses multiple modalities in the semantic space in the multimedia information retrieval. A combination of these approaches is also used as the hybrid fusion. So it uses the late fusion approach which is used for combining both textual and visual information of image search processes because its simplicity, scalability and flexibility.

Scale Invariant Feature Transform (SIFT) [5] is also used to transforms image data into scale-invariant coordinates relative to local features of an image. Main aspect of this approach is that it generates large numbers of features that closely cover the image over the full range of scales and locations. SIFT features are first extracted from a set of reference images and stored in a database for image matching and recognition. The new image is matched by individually comparing each feature from the new image to this previous database and finding candidate matching features based on Euclidean distance of their feature vectors. It can also use the fast nearest-neighbor algorithms that can perform this computation rapidly against large dataset [3].

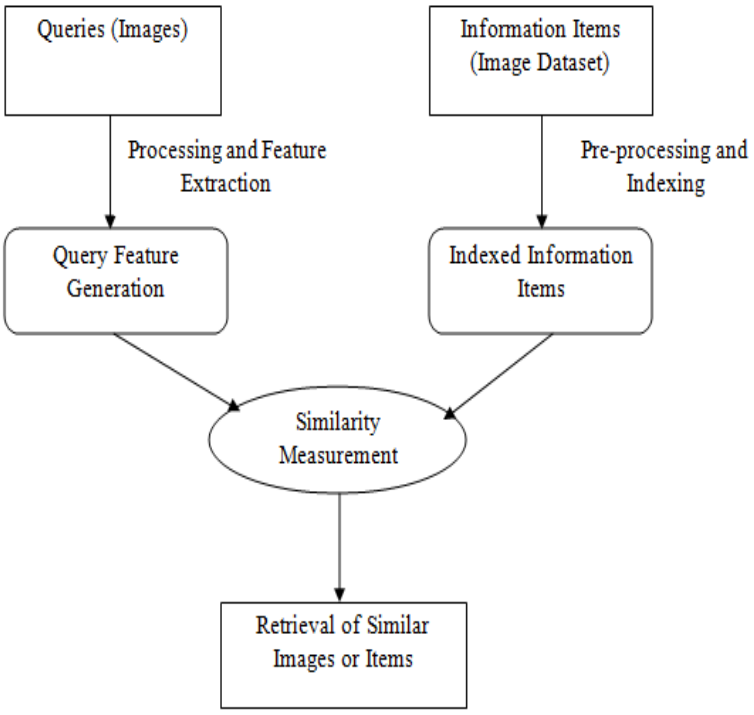

Fig.1. An overview of MIRS Operation

In the MIRS, query types can be in the form of Metadata-based quires, Annotation-based queries (event based queries), queries based on data patterns or features and Query by example for searching an image. It can be also used to retrieve video for its color, texture, segment, timestamp and regions. Image representation is done by visual content descriptor; it can be either global or local. The global descriptor uses the visual features of the whole image and a local descriptor uses the visual features of regions or objects to describe the image content, with the aid of region/object segmentation techniques.

Support vector machine (SVM) [4] is popular for data classification and related tasks to it. In the multimedia, SVMs are being used for different tasks including feature categorization, concept classification, face detection, text categorization, and modality fusion. SVM is used to solve a pattern classification problem, where the input to this classifier is the scores given by the individual classifier. The basic SVM method is extended to create a non-linear classifier by using the kernel concept, where every dot product in the basic SVM formalism is replaced using a non-linear kernel function. SVM-based fusion scheme also can be used. In the late fusion approach, to detect semantic concepts in videos using visual, audio and textual modalities. It uses a separate learning approach while fusing different modalities at the semantic level.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 4, July 2014

III. ALGORITHMIC STRATEGIES USED IN THE SYSTEM

A. LATE FUSION APPROACH

In the Image Retrieval, both visual and textual information are available, and late multimedia fusion approaches are based on combining the support from both the TBIR and CBIR subsystems. Provided decisions will be in the form of numerical similarities (probabilities or scores). Most basic fusion techniques use these scores (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 better than those of early fusion [6].A combination technique is uses called as *image re-ranking*, which consists of an initial step where the textual subsystem retrieves a set of ranked objects that followed by a reorder step of these objects according to the visual score (Si). In the system, the reorder step is carried out by the CBIR subsystem which computes the visual scores (Si) working only over the subset of selected objects by the TBIR subsystem. The five fusion algorithms used to implement the MIRS that can be organized in two categories, one for those based on the relevance score and other for the re-ranking. The algorithms are as follows:

Product/Join: 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/Join. The Product simulates the filtering when St is 0, which means no relevant image for the query and the image will never appear in the fused list ($S_t * S_i$ is 0).

OWA Operators: The ordered averaged weighted operator (OWA) transforms a finite number of inputs into a single output. By the OWA operator no weight is associated with any particular input and the relative magnitude of the input decides which weight corresponds to each input. In the system, the inputs are the textual and image scores (St and Si), which can be provide us the best information. As OWA operators are bounded by the max which is also known as an OR operator, and the min operators which is also known as the AND operator. OWA operator introduced a measure called *orness* to characterize the degree to which the aggregation is like an *or* (max) operation:

$$orness(w) = \frac{1}{n-1}\sum_{i=1}^{n}(n-i)\,w_i \qquad (1)$$

OWA operators with many of the weights close to their highest values will be *or-like* operators that is $orness(W) \leq 0.5$, while those operators with most of the weights close to their lowest values will be *and-like* operators that is $orness(W) \geq 0.5$.

Enrich: It uses two results lists that are a main list which is from the textual module and a support list which is from the visual one. If a positive result appears in both lists for the same query then the relevance of this result in the fused list can be increased in the following way:

$$newRel = mainRel + \frac{supRel}{posRel+1} \qquad (2)$$

Where *newRel* is the relevance value in the fused or merged list, *supRel* is the relevance value in the support list (Si), *mainRel* is the relevance value in the main list (St) and *posRel* is the position in the support list. So, Relevance values will be normalized from 0 to 1. All results appearing in the support list but not in the main list will be added at the end of the merged list. Then the relevance values will be normalized according to the lower value in the main list.

MaxMerge: It selects from the result lists to merge those retrieved images with a higher relevance or score for a specific query and independently of the subsystem (textual or visual) they belong to.

FilterN: It is used to remove from the textual results list those images not appearing in the first N results of the visual list, to eliminate the images that the visual module is not definite of; those with a low score Si. This technique will try to clean the textual results based on the visual result.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 4, July 2014

B. SYSTEM COMPONENTS AND WORKFLOW

These MIRS contains three modules to form architecture, as shown in Fig. 2. [1]. the figure illustrates the global system, which includes the TBIR (Text based image retrieval) module, the CBIR (Content based image retrieval) module, and the fusion module. The textual module works first as a filter for the visual one that works only with the sub-collection filtered by the textual section. Each section gets a ranked list based on a similarity score or probability. Score of the text (St) and score of an image (Si). The fusion module is used to merge these both Scores. For this fusion it uses the Product or Join algorithm. The TBIR subsystem uses the IDRA (Indexing and retrieving automatically) tool, that allows to preprocess the textual information associated with the images in the dataset and to index and retrieve using both its own implemented search engine. The CBIR subsystem uses its own low-level features or the CEDD (Color and edge directivity descriptor) [5] features and its own logistic regression relevance feedback algorithm. An automatic algorithm uses the Euclidean distance as the score for ranking images in the collection; it has been implemented in order to compare the performance of this distance with the logistic regression algorithm. Both the TBIR and CBIR subsystems, generates a ranked list with a certain probability and this information is merged at the fusion module which gives final result. Also merging algorithms are used inside the TBIR subsystem to fuse different textual result lists from monolingual preprocessing, and other fusing techniques are used inside the CBIR subsystem.

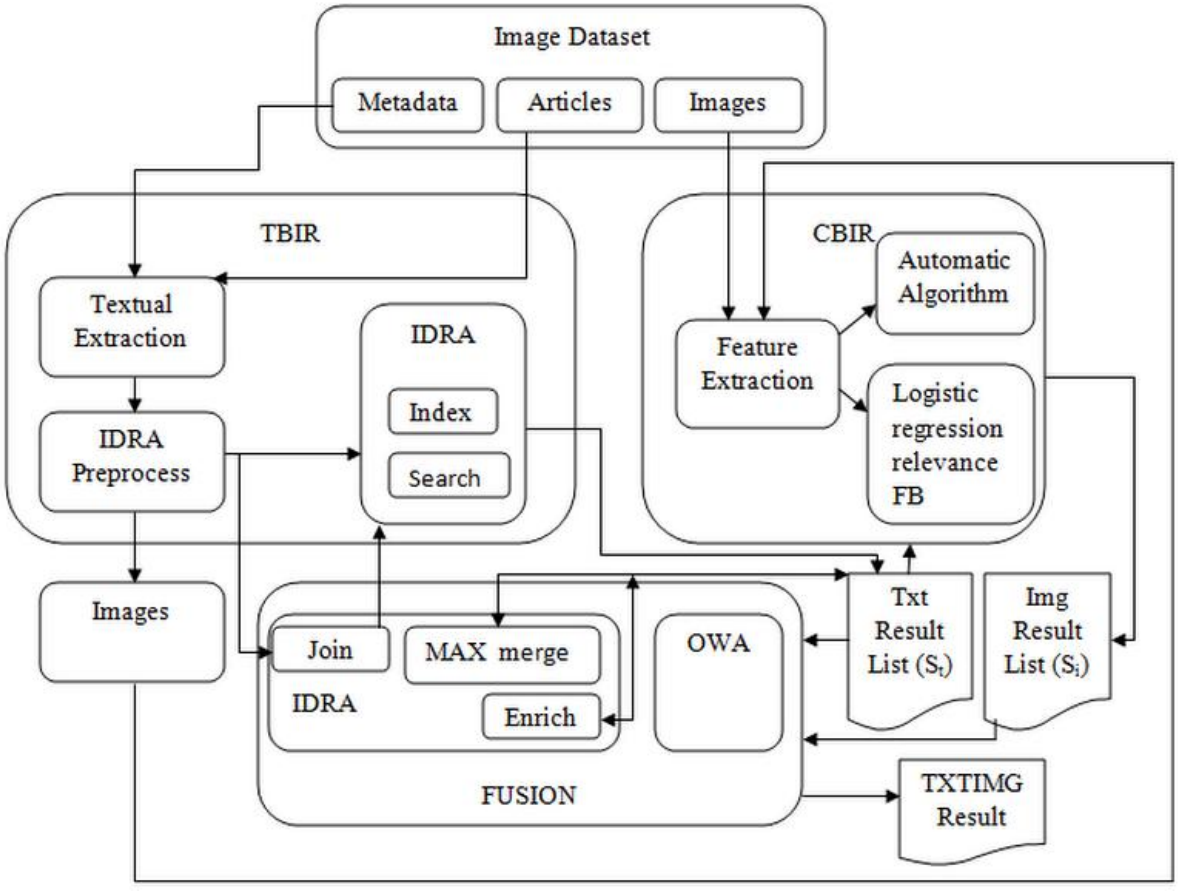

Fig.2. An overview of Fusion of TBIR & CBIR subsystems

Automatic algorithm: This is a standard algorithm in a CBIR system. An image in the dataset has an associated low level feature vector. For this algorithm the low level features are given by an organization, to calculate the similarity measurement between the feature vectors of each image on the dataset and N query images. A distance metric applied in the system can be the Euclidean distance. For N query images, get N visual result lists, individual for each query image in the topic. By using an OWA operator, N result lists are merged.

Logistic regression relevance feedback algorithm: The relevance is assessed relative to information require, not a query. A document is relevant if it addresses the stated information need, not because it just happens to

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 4, July 2014

Contain all the words in the query. Relevance feedback [7] describes the actions which are performed by a user to improve the results of a query by reformulating that query. An initial query formulated by a user may not gives the expected result of the user. Then user needs to change the query manually and re-execute the search until he/she gets expected result. The system learns a new query that better captures the users need for information. The user can enter his/her preferences at each iteration through the selection of relevant and non-relevant images from the dataset.

Let us consider the variable Y which is the random one that gives the user evaluation where Y=1 (image is positively evaluated) and Y=0 (negative evaluation of image). Each image can describe by using low-level features in such a way that the j-th image has the k-dimensional x_j (feature vector) associated. Our data will consist of (x_j, y_j), with j=1,…,n, where n is the total number of images, x_j is the feature vector and y_j the user evaluation (1=positive and 0=negative). The image feature vector x is known for any image and we intend to predict the associated value of Y. In this work, we have used a logistic regression where P(Y=1|x) i.e. the probability that Y=1 (the user evaluates the image positively) given the feature vector x, is related with the systematic part of the model (a linear combination of the feature vector) by means of the logit function. For a binary response variable Y and p explanatory variables $X_1,\ldots,Xp$, the model for $\pi(x)=P(Y=1|x)$ at values $x=(x_1,\ldots,x_p)$ of predictors is $\text{logit}[\pi(x)]=\alpha+\beta_1x_1+\ldots+\beta_px_p$, where $\text{logit}[\pi(x)]=\ln(\pi(x)/(1-\pi(x)))$[1]. The model parameters are obtained by maximizing the likelihood function given by:

$$l(\beta) = \prod^{n} \pi(x_i)^{y_i}[1-\pi(x_i)]^{1-y_i} \quad (3)$$

By using an iterative method, the maximum likelihood estimators (MLE) of the parameter vector β are calculated. The problem arises, when the number of positive plus negative images is typically smaller than the number of characteristics. In order to solve this problem, we can adjust different smaller regression models: each model considers only a subset of variables consisting of semantically related characteristics of the image which is selected. Each sub-model will get a different relevance probability to a given image x, for combine them in order to rank the database according to the users preferences. This problem has been solved by means of an ordered averaged weighted operator [3].

IV. CONCLUSION

In the MIR System detailed description and analysis of textual pre-filtering techniques are used. This textual pre-filtering technique can reduces the size of the multimedia database improving the final fused retrieval results of the system. The combination of textual pre-filtering and image re-ranking lists in a late fusion algorithm outperforms those without pre-filtering of the text. Textual information better captures the semantic meaning of a topic and that the image re-ranking (S_i) fused with the textual score (S_t) helps to overcome the semantic gap in between them. All this performance improvement will be significantly reducing the complexity of the CBIR process, in terms of both time and computation of it. The late fusion algorithms are analyzed and can use, so better results will obtained. The result can get by the value scores, which rely on the ranked positions. It will give best performance with the Join algorithm which means that both modality scores can take into account with the same significance. Duplicate detection of an image can also determined by MIRS.

ACKNOWLEDGMENT

The author wish to thank their guide, parents, god and MET's Institute of Engineering Bhujbal Knowledge City Nasik, for supporting and motivating for this work because without their blessing this was not possible.

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," IEEE Transactions On Multimedia, Vol. 15, No. 8, 2013.

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

[3] R. Granados, J. Benavent, X. Benavent, E. de Ves, and A. Garcia-Serrano, "Multimodal Information Approaches for the Wikipedia Collection at Image CLEF 2011," in Proc. CLEF 2011 Labs Workshop, Notebook Papers, Amsterdam, The Netherlands, 2011.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 3, Issue 4, July 2014

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

[5] S. A. Chatzichristo_s, K. Zagoris, Y. S. Boutalis, and N. Papamarkos, "Accurate image retrieval based on compact composite descriptors and relevance feedback in-formation," Int. J. Pattern Recog. Artif. Intell., vol. 24, no. 2, pp. 207 244, 2010.

[6] 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.

[7] 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. 119, 2006.

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

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

AUTHOR BIOGRAPHY

Ms. Trupti Subhash Atre is post graduate student of computer engineering at MET Bhujbal Knowledge City, Nasik under University of Pune.

K.V. Metre, is working at MET's IOE, Nashik, Maharashtra, India as an Assistant Professor. She completed BE from VNIT, Nagpur and post graduation in Computer Engineering from Dr. Babasaheb Ambedkar Technological University, Lonere. She is pursuing Ph.D. from RTM Nagpur University. She has presented papers at National and International conferences. Her areas of interest include Database, Operating System.