Image Retrieval using Late Fusion

Trupti S.Atre
Computer Engg. Department,
MET IOE, BKC, Adgaon, Nasik, Savitribai Phule Pune University, Maharashtra, India.

ABSTRACT
Multimedia data are used everywhere from huge digital study to the web, multimedia information is used in the professional or personal exercises. Enhancement of multimedia information retrieval can be used both the textual pre-filtering and image re-ranking. The textual and visual techniques are combined and then processes of retrieval are used to develop the multimedia information retrieval system to solve the problem of the semantic gap in the given query. For text based and content based image retrieval, late semantic fusion approaches can also used. The user can also use relevant items that have been found by the system to improve future searches, which is the basis behind logistic regression relevance feedback algorithm is used.

Keywords
Image Retrieval, Late Fusion, Multimedia Information Retrieval.

1. INTRODUCTION
The Multimedia Information Retrieval System is used to retrieve the information which can be in the form of text, images, audio and video or combination of them. The Multimedia Information Retrieval System (MIRS) is the system used to retrieve, store and also maintain the information. The problem arises for communication between information, user and the image retrieval system. It requires that the user have knowledge about the image dataset and the system also have query formulation for getting particular image or data.

The semantic gap is the problem which defines the "lack of coincidence between the information that can take out from the visual data and the analysis that the same data for a user in a given situation". So, semantic gap reflects on visual low-level features (For example, color and texture etc.) exhibited by an image and the semantic for example, objects, meanings, abstract of that image and relationship as perceived by a human. And also semantic gap is the gap between the visual descriptors and the object levels, the gap between the labeled objects and the full semantics of an image [7].Multimedia fusion takes an advantage of each mode and uses the different sources as corresponding information to get a particular image. In an image retrieval task, multimedia fusion can also solve the problem of semantic gap by using the textual pre-filtering and image re-ranking. Different late fusion algorithms that are Product, OWA operators, Enrich, MaxMerge and FilterN can also used. The SIFT algorithm (Scale Invariant Feature Transform) is used for extracting distinctive invariant features from images[1].

The combination rules are as combMAX(the maximum combination), combMNZ (the product of maximum and non-zero numbers) and combSUM(the sum combination)[10]. The approach to fuse the information at the feature level is early fusion and other approach is decision level is late fusion which fuses multiple modalities in the semantic space of the multimedia information retrieval system. By combining these approaches is known as the hybrid fusion of information. It can use the late fusion approach for combining both textual and visual information of image retrieval because of its scalability, flexibility and simplicity [2] [5].

The TBIR (text-based image retrieval) systems can enhance the conceptual meaning of the query than CBIR (content-based image retrieval) systems. For this we can use a textual pre-filtering approach. Then, the CBIR system enhances that the images visually similar from the low-level visual features but with different conceptual meaning of it. The CBIR process will be significantly reduced in terms of time and computation.

The fusion techniques in image retrieval are based on combining textual and visual results. The strategy then combine decisions coming from text and visual-based systems by mean of aggregation functions or classical combinations algorithms. Some of them can use weighted factors to assign different levels of confidence to each mode (textual or visual). The schema will be used in a textual pre-filtering step and semantically reduces the collection for the visual retrieval, that is followed by a textual and visual results fusion phase, and then results will show that how the retrieval performance is improved.

2. RELATED WORK
Multimedia information retrieval retrieves the information that can be in the form of the text, image, audio, and video which may be the best developed technology or system for better result of multimedia. Multimedia Information Retrieval system works with textual and visual information.

Two approaches can used to determine the robust regions in the image that are Scale Invariant Feature Transform (SIFT) [9] and Speeded Up Robust Features (SURF) [8]. The difference between these techniques is to find the salient regions in the image prior to the embedding process and to reveal the possible differences in their performance. SURF is faster and robust than SIFT. These both approaches do not only detect interest points or features but also propose a method for creating an invariant descriptor. This can be used to identify the found interest points and match them even under a variety of disturbing conditions like scale changes, rotation, changes in illumination or viewpoints or an image noise[8][9].

Support vector machine (SVM) is used for data classification and related tasks. In the multimedia information retrieval system, it uses SVMs for different tasks including text categorization, feature categorization, face detection, modality fusion, concept classification, etc. The SVM is a supervised learning method and is used as an optimal binary linear classifier in which a set of input data vectors are partitioned as belonging to either one of the two classes [2][13].

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.

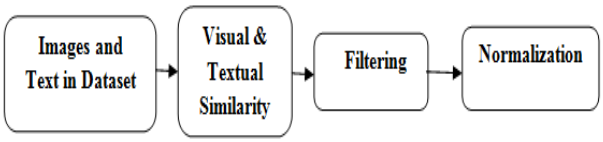

Fig 1: Filtering and Normalization

Filtering and normalization technique is illustrated in Fig 1.In multimedia retrieval model, it uses the text query based semantic filtering as a first level of information fusion. Filtering process categorized into semantic relationships such as multimedia similarities and relevance scores [4].

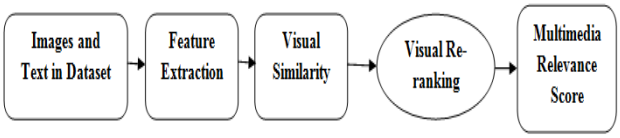

Fig 2: Visual Re-ranking

Visual Re-ranking is illustrated in Fig.2. It can be also called as Image Re-ranking. In Visual re-ranking or image re-ranking technique, visually similar images should have similar relevance scores. Different approaches are used to re-arrange the top retrieved items by the text similarities. There are two steps: by using the text query, firstly they use text based similarities in order to find the most relevant objects from a semantic viewpoint and then they employ the visual similarities between objects of the database in order to refine the textual similarities based ranking[4].

3. ALGORITHMIC STRATEGIES USED IN THE SYSTEM

3.1 Late Fusion Algorithms

The late multimedia fusion approach is based on combining the TBIR and CBIR subsystems which means that the fusion of the TBIR and CBIR subsystems. It provides decisions that will be in the form of numerical similarities (probabilities or scores). So, the probabilities (Pt from textual-based retrieval and Pi from the visual-based retrieval) merged or fused by means of aggregation functions. Late fusion algorithms are better than those of early fusion. A technique is called as image re-ranking, which retrieves a set of ranked objects from textual subsystem that followed by a reorder step of these objects according to the visual score (Pi). The CBIR subsystem which computes the visual scores (Pi) working only on the selected objects of the TBIR subsystem [6]. The five late fusion algorithms are as follows:

Product/Join: For relevance scores of both textual and visual retrieved images (Pt and Pi), two results lists are fused together. Both subsystems can have same importance for the final relevance of the images that can be calculated using the Product/Join. The Join simulates the filtering when Pt is 0, that means no relevant image for the query and the image will never appear in the fused list ($P_t * P_i$ is 0).

OWA Operators: The ordered averaged weighted operator (OWA) provides a finite number of inputs to perform a single output. None of the weight is associated with any particular input and the relative magnitude of the input by the OWA operator. That decides which weight corresponds to each input provided by an operator. The inputs are in the form of textual and image scores (Pt and Pi), that can provide us the best information [3]. The OR (max) and AND (min) operators can be used to find *orness* to characterize the degree to which the aggregation is like operation:

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

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

Enrich: Main list and support list are lists can be used here. A MR is the main list which is from the textual module and a SR is the support list which is from the visual module. If both lists will get positive result for the same query then it will increase the relevance of this result in the fused list as shown in the following way:

$$NR = MR + \frac{SR}{PR+1} \qquad (2)$$

Where, *NR is* the relevance value in the fused or merged list, *SR* is the relevance value in the support list (Pi), MR is the relevance value in the main list (Pt) and *PR* is the position in the support list. So, Relevance values will be normalized from 0 to 1. All results of the support list which are not in the main list will be added at the end of the fused list. And then the relevance values will be normalized according to the lower value in the main list which is the result of the textual module.

MaxMerge: It will get from the result lists to merge retrieved images with a higher relevance or score for a specific query, which are independent of the subsystem (textual or visual) they belongs to.

FilterN: It can be used to remove the textual results list from which images that are not appearing in the first N results of the visual list, the images are eliminated which are irrelevant; those with a low score Pi. This technique will try to clean the textual results that are based on the visual result.

3.2 System Architecture

There are three subsystems to form architecture of an Image Retrieval System, as shown in Fig. 3. The figure shows the overview of the system in which the TBI (Text based image) module and the CBI (Content based image) module, and the fusion module are illustrated. Pre-filtering will be done by textual module i.e. TBIR. In that, each section gets a ranked list based on a similarity score or probability. Score of the text (Pt) and score of an image (Pi). Score of the both CBIR and TBIR are merged by using the fusion module.

The TBIR subsystem can use the IDRA (Indexing and retrieving automatically) tool, which allows to preprocessing of the textual information related with the dataset images. The tool can also used to index and retrieve by using its own implemented search engine. The CBIR subsystem uses its own low-level features or the CEDD (Color and edge directivity descriptor) features [6]. This system also uses its logistic regression relevance feedback algorithm [6]. An automatic algorithm uses the Euclidean distance as the score for ranking images in the collection; it can be used to compare the performance of this distance with the logistic regression algorithm.

Both the TBIR and CBIR subsystems are used to generates a ranked list with a certain probability and this information is

merged at the fusion module which gives final result. And 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.

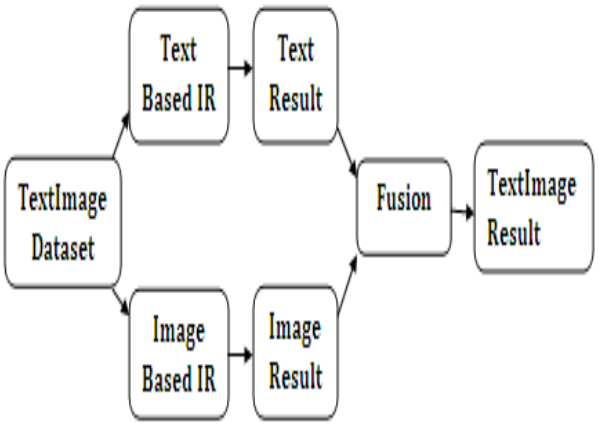

Fig 3: Image retrieval system

Automatic algorithm: In image retrieval system, this is the standard algorithm. For this algorithm the low level features are given, to determine the similarity measurement between the feature vectors of each image on the dataset and N query images. Euclidean distance can be used as a distance metric in the image retrieval system. N query images will get N visual result lists. By using an OWA operator, N result lists can merge.

Logistic regression relevance feedback algorithm: In the information retrieval system, there arises problem of finding relevant items to improve future searches from the dataset. To overcome this problem, the logistic regression relevance feedback algorithm can be used. In the relevance feedback concept, the new query should be based on the old query that modifies in the relevant items to increase the weight and in non-relevant items to decrease the weight. So, this technique not only modifies the terms in the original query but also allows expansion of new terms from the relevant items. This is also known as query reformulation [11][12].

Let us consider the variable B which is the random one that gives the user evaluation where B=1 (image is positively evaluated) and B=0 (negative evaluation of image). From the dataset each image can be described by using low-level features in such a way that the b^{th} image has the k-dimensional a_j (feature vector) associated. The data will consist of (a_j,b_j), with b=1,…,n, where n is the total number of images, a_j is the feature vector and b_j the user evaluation (1=positive and 0=negative). The image feature vector a is known for any image and we intend to predict the associated value of B. It can be used a logistic regression where P (B=1|a) i.e. the probability that B=1 (the user evaluates the image positively) given the feature vector a, 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 B and p explanatory variables A_1,…,A_p, the model for π(x)=P(Y=1|x) at values a=(a_1,…,a_p) of predictors is logit[π(a)]=α+$\beta_1 a_1$+…+$\beta_p a_p$, where logit[π(a)]=in(π(a)/(1-π(a)))[1]. The model parameters are obtained by maximizing the likelihood function given by:

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

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. Solution for this problem is that to 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. So, better solution for this problem is ordered averaged weighted operator [3].

4. CONCLUSION

The fusion techniques such as join algorithm using results of textual pre-filtering and image re-ranking methods can be used in the Image Retrieval System. This textual pre-filtering technique reduces the size of the multimedia database to improve the result of the final fused retrieval of the system. In the image re-ranking technique, the visual similar images can have similar relevance score. The image re-ranking (P_i) can fuse with the textual score (P_t) to overcome the problem of semantic gap. This performance improvement can reduce the complexity of the CBIR process. The late fusion algorithms are analyzed and can be used to give better results. Same approach can be used for video retrieval that can be text based or image based.

5. ACKNOWLEDGMENTS

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.

6. REFERENCES

[1] XaroBenavent, 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 ImageCLEF 2011," in Proc. CLEF 2011 Labs Workshop, Notebook Papers, Amsterdam, The Netherlands, 2011.

[4] Gabriela Csurka, Julien Ah-Pine, and Stéphane Clinchant, "Unsupervised Visual and Textual Information Fusion in Multimedia Retrieval - A Graph-based Point of View," arXiv:1401.6891v1 [cs.IR] 27 Jan 2014.

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

[6] 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.

[7] M. Grubinger, "Analysis and Evaluation of Visual Information Systems Perfor-mance," Ph.D. thesis, School Comput. Sci. Math., Faculty Health, Engi., Sci., Victoria Univ., Melbourne, Australia, 2007.

[8] Nagham Hamid, Abid Yahya, R. Badlishah Ahmad, and Osamah M. Al-Qershi, " A Comparison between Using SIFT and SURF for Characteristic Region Based Image Steganography" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3, May 2012.

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

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

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

[12] 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.

[13] Yogesh C. Pathak, S. A. Chhabria, "Performance Evolution of Eye and Hand Fusion for Diagonal Movement Gesture Recognition," International Journal of Advance Research in Computer Science and Management Studies, Volume 2, Issue 4, April 2014.