VOL. 10, NO. 9, MAY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

WEIGHTED RELEVANCE FEEDBACK AND KEYWORD BASED MEDICAL IMAGE ANNOTATION

Vijay Jeyakumar and Bommanna Raja Kanagaraj
Department of Biomedical Engineering, Centre for System Design and Information Processing, PSNA College of Engineering and Technology, Dindigul, India
E-Mail: ece.vijay@gmail.com

ABSTRACT

Information retrieval plays a major role in all the fields such as medicine, industry, information technology and research. In order to retrieve the medical images from healthcare section, feature based retrieval methods are used to enable automation. In this paper, we proposed a new framework to make the retrieval system to outperform well. In this case, we have taken the advantage of feature based retrieval system along with relevance feedback mechanism in the first stage and text based query refinement at the second stage on heterogeneous datasets. In CBIR, rank order and precision values calculated in each iteration make appropriate weight adjustments among different features. As the image retrieval system is highly subject to user semantic concepts, the text based query is also adapted to enable the retrieval system to perform better.

Keywords: CBIR, feature extraction, feature selection, fusion, image retrieval, relevance feedback, TBIR.

1. INTRODUCTION

Image retrieval is a poor stepchild to other forms of information retrieval (IR). Now a day's millions and millions of users are searching for data based on text from the internet [1], fewer search and retrieve images every day. Generally Image retrieval systems take two approaches for indexing and retrieval of data. First approach is, indexing and retrieval of the textual annotations associated with images termed as Text based image retrieval system (TBIR) [2]. A number of commercial systems employ this approach, such as Google Images (www.images.google.com) and Flickr (www.flickr.com). A second approach, called visual or content-based is adopting various image processing techniques to find features of the images, such as color, shape and texture [3]. Content based image retrieval (CBIR) is the most prevalent mechanism to describe image contents [4]. The need for content-based techniques become obvious when considering the enormous amounts of digital images produced and stored on every day e.g. by digital cameras or digital imaging methods in medicine. The content description method is used to extract low level features automatically. Unfortunately, it is often difficult to translate an image retrieval requirement into a statistical distribution of low level features [5]. In addition to this, the descriptions from these features may not necessarily provide a meaningful representation of an object. There are other ongoing researches studying on higher level or associating low level and high level to facilitate effective retrieval. So when considering the large amount of medical images, an essential mechanism is required which can classify and search medical images at the different semantic level. As the medical images are related to so many criteria's such as treatment history, diagnosis suggested, symptoms, directionality, anatomy, patient information and timeline of the image captured, simple feature based retrieval method is meaningless. In order to overcome this problem, text supported like metafeature along with CBIR will improve the overall system performance.

2. FUSION TECHNIQUES

Another challenge involved in content based image retrieval method is finding suitable feature extraction methods for all kind of medical images. Therefore, to increase the efficiency of the retrieval system in medical domain is to assist the users by combining the two approaches (CBIR and TBIR) with simple fusion techniques. The fusion technique can be categorized as: 1) Early fusion and 2) Late fusion. The early fusion is also referred to as fusion in feature space, unimodal features are extracted from different data are described into single vector. Choice of choosing classifiers for unimodal features is difficult. Various early fusion methods are adopted for image classification [6], scene recognition [7], face detections and traffic monitoring [8]. The late fusion is referred to as decision level fusion or semantic level fusion. Based on the nature of the features, classifiers are learned to take decisions like yes or no by providing a score. Some of the experiments show that feature based retrieval system performs well for mono modality images. Whereas the late fusion methods works for similarity findings than the feature based method. This method can also be used to combine two different approaches such as TBIR and CBIR by aggregating functions. Gui *et al* [9] proposed a system which works based on the text associated terms like URL, title or keyword using tf-idf strategy. The MAP provided by the system is 0.27. Clinchant *et al* [10] introduced a fusion technique using image re ranking on four datasets and the MAP of the system is 0.396. Some of the retrieval systems adopted a) classification based late fusion and b) Rule based late fusion method. The classification based late fusion method utilized various classifiers such as SVM and Bayesian classifiers to maximize the expected semantic results of the users for multimedia, multimodality and text features

VOL. 10, NO. 9, MAY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

[11-13]. The rule based fusion method [14-15] provides the scores to the classified objects and texts by the classifiers and the weights of the scores are fused and ordered using relevance feedback mechanism. This weight based late fusion approach is tested in multimedia retrieval, person identification in TV broadcast and visual concept recognition [16-18].The framework developed with relevance feedback mechanism and TBIR is given in Figure-1.

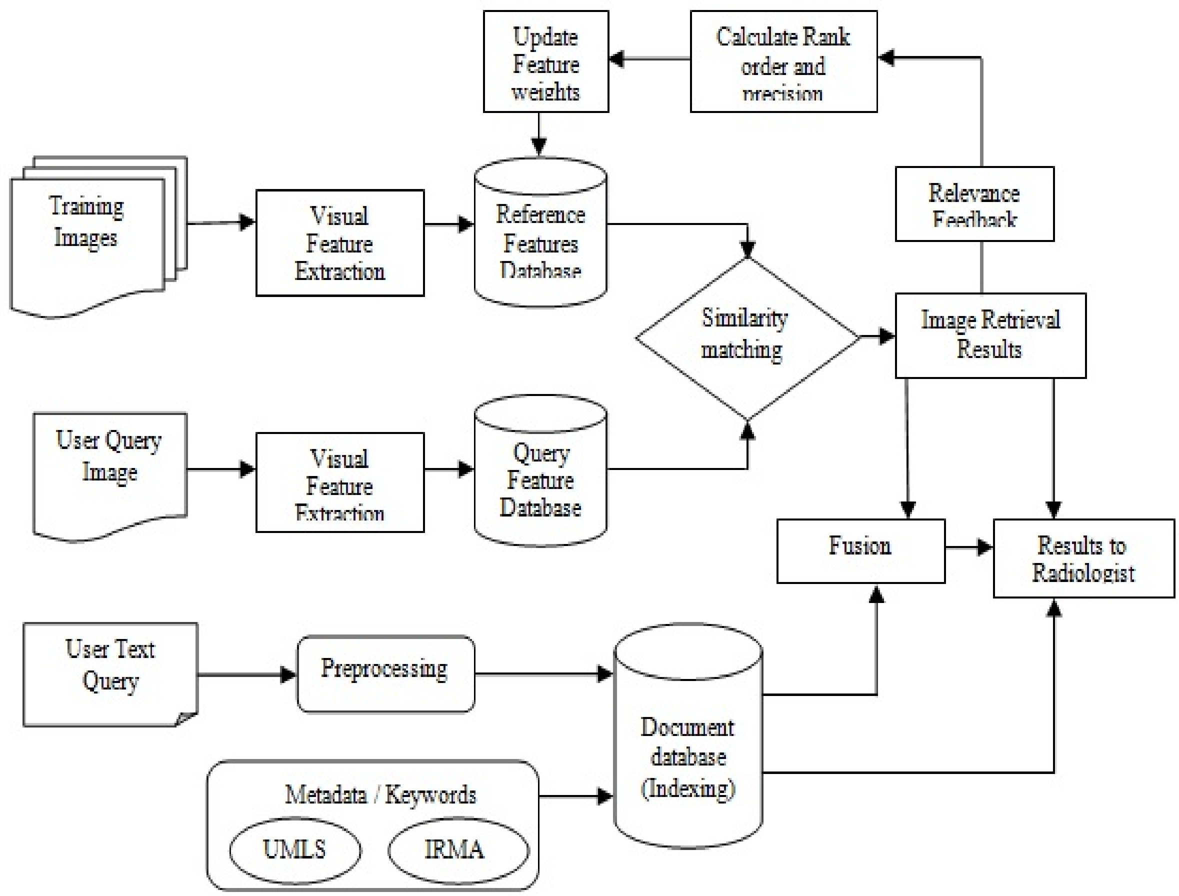

Figure 1. Image retrieval using relevance feedback mechanism and text based query.

3. METHODOLOGY

In this work, the image database consisting of 12000 training/reference and 2000 testing/query images obtained from RWTH, Germany [19] are considered for retrieval process.

3.1 Feature extraction

Features of both homogeneous and heterogeneous dataset images are extracted initially. Totally 199 global and local features such as Tamura (3), Gabor Features (48), Region based features (08), Wavelet moments (64), Global textures (13), Histogram of Gradients (HOG) (81), moment invariants (07) are extracted from an image. The features are extracted from training and testing images separately. To find the closeness between the query (or) testing images, various similarity measures are used. In this work, Euclidean distance (L2), Manhattan or City block (L1), Relative Deviation, Mahalanobis, Cosine, Chebyshev ($L\infty$) and Spearmen distances are used to find the association between the query and training images. Based on the features extracted and similarity measures, the retrieved images based on the given query for the heterogeneous datasets are shown in fig. The retrieval is performed by Euclidean distance. The precision, recall and F- score of this retrieval process is 0.55, 0.13 and 0.21 respectively. The computational time taken by the retrieval system is 1.8 second. Similarly the various similarity measures applied for homogeneous datasets and its performance measures are available in [20]. The images retrieved from the heterogeneous database without relevance feedback are given in Figure-2. shows the number of relevant images from the retrieved images is very poor.

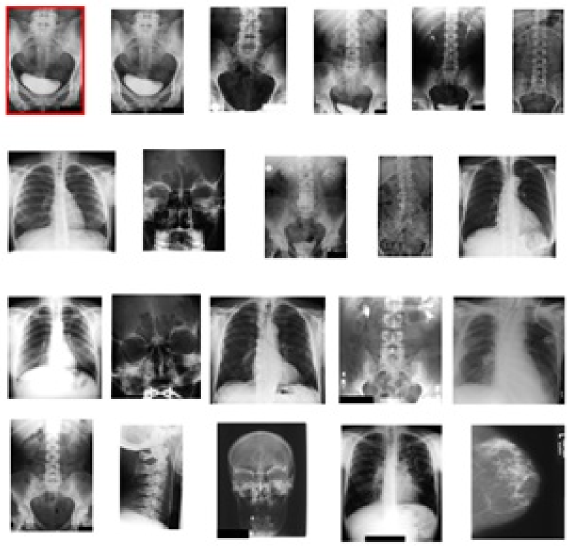

Figure-2. Retrieved images for the given X ray – hip - image as a query (the image represented by red outer line) without relevance feedback.

3.2 Feature based image similarity matching

A generalized method to retrieve images for all types of queries is difficult. Feature descriptors which are extracted from the training and testing images are at different levels. Data fusion method is very much useful to find the scores of similarity matching between different features. This fusion method can be adopted in CBIR with predetermined weights to increase the efficiency of the system. The weights are normalized based on the accuracies of the features subject to $0 \leq \omega^F \leq 1$ and $\sum \omega^F = 1$. In this approach the query image and images in the database are denoted by Q_I and I respectively. The similarity between them are represented by

$$Sim(Q_I, I) = \sum_F \omega^F Sim^F(Q_I, I) \qquad (1)$$

Here F € {Tamura, Gabor, Region, wavelet, texture, HOG, moments} and ω^F are the weights of different image representations.

3.3 Weighted relevance feedback

Using this relevance feedback mechanism, the feature weights are updated at each iteration by calculating the precision and order of the rank of relevant images from retrieved images in individual list. In order to update the weights of the features the following algorithm is used.

Algorithm

```
1 Initialize equal weighting of each feature
2 Perform initial similarity matching between the features
3 {
4       Relevance feedback starts
5       {
6               Do updation
7               {
8               Calculate the efficiency of returned images
9               E = (Σ_{i=1}^{k} Rank(i)) / (k/2) . P(k)   // effectiveness of top k returned images
                Where Rank (i) =0; for non relevant images in i position
                Rank (i) = (k-i) / (k-1); for relevant images
10              Calculate the precision value
                P(k)=Rk/k; // Precision at top k
                Where Rk the number of relevant images in the top K retrieved results
11              }
12              Update feature weights by É=ώF.
13              Find similarity matching between Q_I and I by
14              Sim(Q_I, I) = Σ_F ώ^F Sim^F(Q_I, I)
15      }
16      Find top K returned images
17 }
18 Stop the retrieval process when user finds relevant images.
```

VOL. 10, NO. 9, MAY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

From the above algorithm, as rank order and precision value does not depend each other, the effectiveness of the measure is calculated by the dot product between rank order and precision. If the returned top K images are relevant, the performance score will be higher. This will yield the effectiveness of the system to be 1. In order to normalize the scores the total score is modified by $É=ώ^F$.

3.4 Textual feature extraction

In order to improve the retrieval system performance, text based query is added as an additional feature with CBIR. In this text based query, the weights to the relevant text from a document has to be identified. Then the relevant texts can be fused to the top K retrieved images.

To assign the weights to the terms in a document, some preliminary actions must be taken such as removing unimportant words and porter stemming [21]. The resulting terms in the document are called as indexed terms. The query also undergoing the same preprocessing terms.

The final weights to the terms in the document is calculated as follows,

$$w(t,d) = \frac{g(t,d)}{(1-\lambda).c+\lambda.n_1(d)} \tag{2}$$

$n_1(d)$ is number of words that appear only once (singleton), c is average number of singletons in all documents, $\lambda \in \{0 \rightarrow 1\}$ and g(t,d) is defined as,

$$g(t,d) = \begin{cases} \frac{[1+\log n(t,d)]}{[1+\log \bar{n}(d)]} & \text{if } t \in d \\ 0 & \text{else} \end{cases} \tag{3}$$

Here n (t,d) is the frequent occurrence of term in the document d and $\bar{n}$(d) is average term frequency in d. For weighting the documents in the query, w(t,q) is calculated as follows,

$$w(t,q)=[1+\log n(t,q)].idf(t) \tag{4}$$

and for weighting the terms in the query, inverse document frequency is calculated as

$$idf(t) = \log\left[\frac{k}{n(t)}\right] \tag{5}$$

where k is a term, k is the number of documents and n(t) is the number of documents in which t occurs.

The calculated ranking value (Retrieval Status Value – RSV) is the relevance of a document d for a query q as described below

$$RSV(q,d) = \sum_{t \in T} w(t,q).w(t,d) \tag{6}$$

T is the set of all documents.

3.5 Mixed retrieval

The similarity between a mixed query $Q=(Q_I,Q_T)$ $\{Q_I \rightarrow$ image(s); $Q_T \rightarrow$text$\}$ and a couple of image with the associated medical report (t,d) is given b

$$\lambda(Q,I,d) = \alpha \frac{\lambda_v(Q_I,I)}{\max_{z \in D_I} \lambda_v(Q_I,I)} + (1-\alpha)\frac{\lambda_T(Q_T,d)}{\max_{z \in D_T} \lambda_T(Q_T,z)} \tag{7}$$

Where z is the image in the database, $\lambda_v(Q_I,I)$ is the score of the visual similarity between the query image Q_I and images in the database. $\lambda_T(Q_T,d)$ is the similarity between the textual query Q_T and document. D_I is the image database and D_T is the text database. Usually α varies from 0 to 1. If α is 0, then the text based similarity is possible and for α is 1, the visual similarity is performed.

4. RESULTS AND DISCUSSIONS

The relevance feedback method is applied to heterogeneous images which consist of various modality images. Based on the weight updation in each iteration, the precision and recall is increased about 20% in different similarity matching's. The precision and recall values can be calculated from the simplified Table-1.

Table-1. To determine precision and recall.

	Relevant	Not relevant
Retrieved	A (Correctly retrieved)	B (Incorrectly retrieved)
Not retrieved	C (Missed)	D (Correctly rejected)

$$Precision = \frac{A}{(A+B)};\ Recall = \frac{A}{(A+C)}$$

The similarity performed for hip image as a query. As relevance feedback mechanism is very much useful in increasing the retrieval efficiency, still the gap exists between the semantic and syntactic visual features.

VOL. 10, NO. 9, MAY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

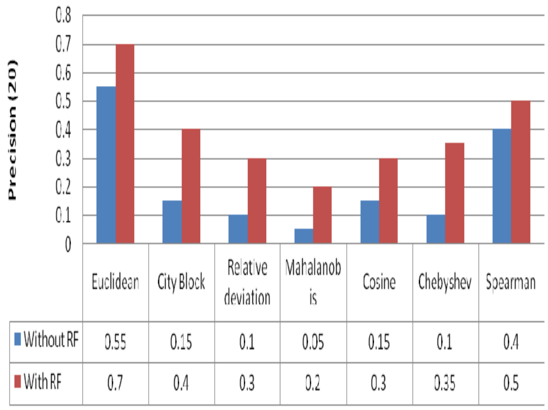

Figure-3. comparative values of precision obtained from various similarity measures using relevance feedback.

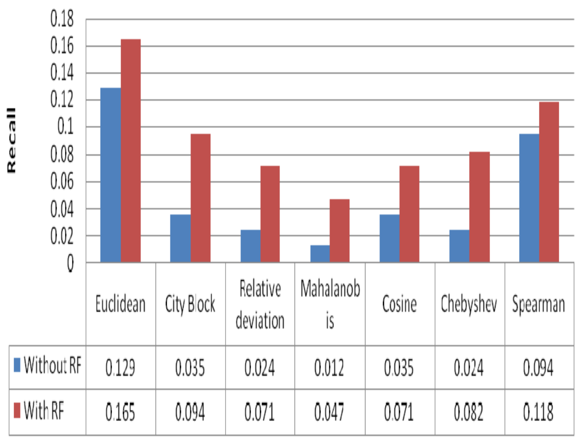

Figure-4. comparative values of recall obtained during retreival for the given hip as query image.

Table-2. Comparative values of various algorithms.

Methodology	MAP@20	Recall
Without RF	0.514	0.11
With RF	0.683	0.178
Mixed	0.935	0.33

The above values are obtained for the hip as query image. In order to provide the meaningful comparative statement, we have chosen Euclidean distance to find the similarity matching. Using the proposed architecture, the results can be obtained only using image based query or text based query or the combination of both. Out of this various methods, the fusion of text and features based on weighted method make the overall system performance to be increased. But still the semantic gap exists. In order to fill the gap, we can employ feature selection method after feature extraction. As various methods are available like filter and wrapper methods suggested by researchers, using optimization method to remove redundant features will improve the efficiency of the system. As weight updation is given to all feature descriptors, it is very difficult task to identify the useful features for different queries. Hence better optimization can be employed.

REFERENCES

[1] Rice RE. 2005. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform. Vol. 75 pp. 8–28.

[2] Rui Y, Huang TS and Chang SF. 1999. Image retrieval: past, present and future. J Visual Comm Image Represent 10: pp. 1–23.

[3] Müller H., Michoux N., Bandon D. and Geissbuhler A. 2004. A review of content-based image retrieval systems in medical applications - clinical benefits and future directions. Int J Med Inform 73: pp. 1–23.

[4] M. Flickner, H. Sawhney, H. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, “Query by image and video content: The QBIC system,” IEEE Computer, vol. 28, No.9, pp. 23–31, 1995.

[5] Smeulders A. W. M., Worring M., Santini S., Gupta A. and Jain R. 2000. “Content-based image retrieval: The end of the early years”. IEEE Trans. on Pattern Analysis and Machine Intelligence 22 pp. 1349–1380.

[6] M. Guillaumin, J. Verbeek and C. Schmid. 2010. “Multimodal semisupervised learning for image classification,” in CVPR.

[7] M. Law, N. Thome and M. Cord. 2012. “Hybrid pooling fusion in the bow pipeline,” ECCV.

[8] M. S. Kankanhalli, J. Wang and R. Jain. 2006. “Experiential sampling in multimedia systems,” Trans. Multi., Vol. 8, No. 5, pp. 937–946, October.

[9] C. Gui, J. Liu, C. Xu and H. Lu. 2010. “Extended CBIR via learning semantics of query image,” Advances in Multimedia Modeling, pp. 782-785.

[10] S. Clinchant, J. Ah-Pine and G. Csurka. 2011. “Semantic combination of textual and visual information in multimedia retrieval,” In Proceedings of the 1st ACM International Conference on Multimedia Retrieval, April, pp. 44. ACM.

[11] W. H. Adams, G. Iyengar, M. R. Naphade, C. Neti, H. J. Nock and J. R. Smith. 2003. "Semantic indexing of multimedia content using visual, audio and text cues," EURASIP, Vol. 2, pp. 170– 185.

[12] X. Wang and M. Kankanhalli. 2010. "Portfolio theory of multimedia fusion," MM.

[13] I. Kim, S. Oh, B. Byun, A. G. A. Perera and C. H. Lee. 2012. "Explicit performance metric optimization for fusion-based video retrieval," ECCV2.

[14] S. T. Strat, A. Benoit, H. Bredin, G. Quenot and P. Lambert. 2012. "Hierarchical late fusion for concept detection in videos," ECCV.

[15] R. Yan, J.n Yang and A. G. Hauptmann. 2004. "Learning query class dependent weights in automatic video retrieval," MULTIMEDIA '04, pp. 548–555, ACM.

[16] X. S. Hua and H. J. Zhang. 2004. "An attention-based decision fusion scheme for multimedia information retrieval," 2004, PRCM.

[17] H. Bredin, J. Poignant, M. Tapaswi, G. Fortier, V. Le,T. Napoleon, H. Gao, C. Barras, S. Rosset, L. Besacier, J. Verbeek, G. Quenot, F. Jurie and H. Ekenel. 2012. "Fusion of speech, faces and text for person identification in tv broadcast," ECCV.

[18] N. Liu, E. Dellandrea, C. Zhu, C. E. Bichot and L. Chen. 2012. "A selective weighted late fusion for visual concept recognition," ECCV.

[19] http://ganymed.imib.rwth-aachen.de/irma/onlinedemos_en.php.

[20] Vijay Jeyakumar and Bommannaraja Kanagaraj. 2014. "Evaluation of Similarity measures in a medical Image Retrieval System", International Journal of Applied Engineering Research, Vol. 9, No. 21 pp. 11039-1105.

[21] M. F. Porter. 1980. An algorithm for suffix stripping, July. Programm.