RESEARCH ARTICLE OPEN ACCESS

Query Image Searching With Integrated Textual and Visual Relevance Feedback for Web Image Retrieval

Dr. Raja Murali Prasad[1], Dr. G.A.E. Satish Kumar[2]
Professor, ECE Dept., Vardhaman College Of Engineering, Hyderabad, Telangana, India[1,2]

ABSTRACT
There are many researchers who have studied the relevance feedback in the literature of content based image retrieval (CBIR) community, but none of CBIR search engines support it because of scalability, effectiveness and efficiency issues. In this, we had implemented an integrated relevance feedback for retrieving of web images. Here, we had concentrated on integration of both textual features (TF) and visual features (VF) based relevance feedback (RF), simultaneously we also tested them individually. The TFRF employs and effective search result clustering (SRC) algorithm to get salient phrases. Then a new user interface (UI) is proposed to support RF. Experimental results show that the proposed algorithm is scalable, effective and accurated.

Index Terms:content based image retrieval (CBIR), relevance feedback (RF), search result clustering (SRC), web image retrieval and integrated TVRF

I. INTRODUCTION

Recent years there is a rapid growth in searching engines such as Bing image search: Microsoft's CBIR engine (Public Company), Google's CBIR system, note: does not work on all images(Public Company), CBIR search engine, by Gazopa (Private Company), Imense Image Search Portal (Private Company) and Like.com (Private Company), image retrieval has become a challenging task. The interest in CBIR has grown because of the retrieval issues, limitations and time consumption in metadata based systems. We can search the textual information very easily by the existing technology, but this searching methods requires humans to describe each images manually in the database, which is not possible practically for very huge databases or for the images which will be generated automatically, e.g. images generated from surveillance cameras. It has more drawbacks that there is a chance to miss images that use different equivalent word in the description of images.

The systems based on categorizing images in semantic classes like “tiger” as a subclass of “animal” can debar the miscatergorization problem, but it will requires more effort by a use to identify the images that might be “tigers” , but all of them are categorized only as an “animal”. Content-based image retrieval (CBIR) is an application of methods of acquisition, pre-processing, analyzing, representation and also understanding images to the image retrieval problem, that is the problem of exploring for digital images from large databases.

The CBIR system is opposed to traditional approaches, which is known on concept based approaches i.e., concept based image indexing (CBII) [1].

II. RELATED WORK

In the past decades several CBIR systems have been proposed, and still the researchers are focusing on developing extended CBIR systems with more effective results. The letter proposed in [4] gives a comparison of different approaches of CBIR based on similarity measures and image features to identify the similarity between the images, which provides accurate information for retrieving the relevant images from large database. Wan Siti*et.al* proposed in [5] compares the several medical image retrieval systems based on the feature extraction and to improve the effectiveness of the CBIR system for medical images such as magnetic resonance (MR) images and computed tomography (CT) images [10]. The major concept proposed in [5] is to help in the diagnosis such as to find the similar disease and monitoring of patient s progress continuously. B. S. Manjunath*et.al* presented in [6] is the combination of color, texture with inclusion of edge compactness for Motion Picture Expert Group (MPEG)-7 standards. Another approach proposed in [7] used different color spaces such as HSV and YCbCr explains a similar approach based on color and texture analysis. The work proposed in [8] introduces a new retrieval system which has done by using wavelet transformation with both color and texture features together and will perform better than existed state of art algorithms.

Recently, retinal image retrieval system called CBIR for retinal and blood vessels extraction [9] has been analyzed by the histogram features of RGB color components. The multi resolution analysis has applied to the image to acquire the texture information. In addition to improve the performance, morphological operations are applied to study the shape of object. Swati Agarwal has proposed a new CBIR system in [11], which is by using discrete wavelet transform and edge histogram descriptor (EHD). Here the retrieval is based on color and texture features not by using color information in the image, input image first decomposes the input query image into several sub bands i.e., approximation coefficients and detail coefficients, where detail coefficients consists of horizontal (LH), vertical (HL) and also the diagonal information (HH) of the image. Afterwards, EHD is used to gather the information of dominant edge orientations. This mixture of 3D-DWT and EHD will improve the efficiency of the CBIR system. In this paper, we proposed an integrated textual and visual relevance feedback (ITVRF) for web image retrieval to improve the CBIR system efficiency, accuracy with reduced time.

III. PROPOSED SYSTEM

For image retrieval, classification and indexing both color and texture have been used widely in various applications. Histogram of a image is a graphical analysis of a image, which represents the color information of image. It is a first order statistical measure. The major drawback of this histogram based approaches is that the spatial distribution and local variations will be ignored. Local spatial variation of pixel intensity is commonly used to capture texture information in an image. The images collected from several photo forum sites have rich metadata. These images constitute the dataset evaluation for the proposed RF framework. For example, a picture from the database has the following data. We are denoting it by *Q*,for later citation of this picture.

- Title: *early morning*
- Category: *landscape, nature, rural*
- Comment: *I found this special light one early morning in Pyreness along the Vicdessosriver near our house. . . .*
- One of the critiques: *wow. . . I like this picture very much…I guess the light has to do with everything ... the light is great on the snow and on the sky (strange looking sky by the way)... greatly composed ...nice crafted border… a beauty.*

Above mentioned metadata is used for the construction of textual space. There are two variables to build the textual space. One is directly by using the above metadata and second is, search result clustering (SRC) algorithm.

To represent the TF, space model of vector with TF-IDF weighting scheme is adopted. More specifically, the TF of an image *I* is a vector of *L*dimension and it can be given by

$$\overrightarrow{F^T} = (w_{1,\dots,}w_L)$$
$$w_i = tf_i.\ln\left({}^{N}/_{n_i}\right)$$

Where:

- $\overrightarrow{F^T}$is the TF of an image *I;*
- $w_i$is the weight of the i_{th}term in textual space of *I;*
- *L* is the number of textual space distinct terms of all images;
- $tf_i$is the i_{th}term frequency in textual space of *I;*
- *N* is the total number of images;
- $n_i$is the number of images whose data contains the i_{th}term.

To illustrate the straightforward approach where all metadatais utilized to construct the textual space, we use the photo *Q* introduced at the beginning of this section as an example. Given the query "*early morning*," we have 151 resulting images including photo *Q*. Based on those resulting images, we collect all distinct terms from the metadata which results in totally 358 distinct terms. For *Q*, it has 48 distinct terms, which consist of *early, morning, landscape, nature, rural, I, found, this, special, light, one, in, Pyrenees, along, the, Vicdessos, river, near, our, house, wow, like, picture, very, much, guess, has, to, do, with, everything, is, great, on, snow, and, sky, strange, looking, by, way, greatly, composed, nice, crafted, border, a*, and *beauty*.

Given N=151, L=358 and 48 distinct terms of *Q*, then we can calculate the n_i and tf_i with respect to *Q*.As a result, we can get the$w_i$according to eq. (2). Finally, the TF can be obtained by eq. (1)

To visually represent an image, a 64-dimensional feature was extracted. It is a combination of three features: six-dimensional color moments, 44-dimensional banded auto correlogram, and 14-dimensional color texture moments. For color moments, the first two moments from each channel of CIE-LUV color space were extracted. For correlogram, the HSV color space with inhomogeneous quantization into 44 colors is adopted. For textual moments, we operate the original image with templates derived from local Fourier transform and obtain characteristic maps, each of which characterizes some information on a certain aspect of the original image. Similar to color moments, we calculate the first and second moments of the characteristic maps, which represent the color texture information of the original image. The resulting visual feature of animage is a 64-dimensional vector $\overrightarrow{F^V} = (f_{1,\dots,}f_{64})$. Each feature dimension is normalized to [0, 1] using Gaussian normalization for the convenience of further computation.Rocchio's algorithm is used to perform RF in textual space, which has been developed in mid-60's and it has been proven to be one of the most effective RF algorithms in information retrieval.

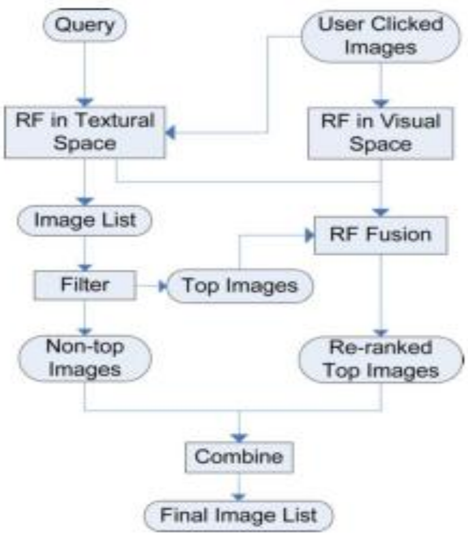


Fig1. Flow chart of Proposed CBIR model

Optimal query features can be defined as follows:

$$\overrightarrow{F_{\text{opt}}} = \overrightarrow{F_{\text{ini}}} + \frac{\alpha}{N_{\text{Rel}}} \sum_{I \in \text{Rel}} \overrightarrow{F_I} - \frac{\beta}{N_{\text{Non}-\text{Rel}}} \sum_{J \in \text{Non}-\text{Rel}} \overrightarrow{F_J}$$

Where:

$\overrightarrow{F_{\text{ini}}}$ =initial query vector;

$\overrightarrow{F_I}$ = relevant image vector

$\overrightarrow{F_J}$= non-relevant image vector

Rel= set of relevant images

N_{Rel} = number of relevant images

$N_{\text{Non}-\text{Rel}}$ =number of non-relevant images

$\alpha$is the parameter that controls the relative contribution of initial query and relevant images;

$\beta$is the parameter that controls the non-relevant images and the initial query contribution. In this case, we have only relevant images, so we set β=0 and α=1 in our experiments.To perform RF in visual space, Rui's algorithm is used.Assume clicked images to be relevant, both an optimal queryand feature weights are learned from the clicked images. More specifically, the feature vector of the optimal query is the meanof all features of clicked images. The weight of a feature dimension is proportional to the inverse of the standard deviation of the feature values of all clicked images. Weighted Euclidean distance is used to calculate the distance between an image and the optimal query. Although Rui's algorithm is used currently, any RF algorithm using only relevant images could be used in the integrated framework.

3.1. Multimodal fusion

There has been some work on fusion of relevance feedback in different features of spaces such as linear combination, support vector machine (SVM) based non-linear combination and super-kernel fusion algorithms. All of them are incapable for a system, which offers only relevant images.Since textual features are more semantic-oriented and efficient than visual features while visual features have finer descriptive granularity than textual features, we combine the RF in both feature spaces in a sequential way. The flow chart of the RF of our unified framework is shown in Fig. 1. First, RF in textual space is performed to rank the initial resulting images using the optimal query learned in above section. Then, RF in visual space is performed to re-rank the top images. The re-ranking process is based on a dynamic linear combination of the RF in both visual and textual spaces.

The similarity metric used to re-rank a top image I using integrated TVRF is defined as follows:

$$S = \beta \cdot S^V + (1-\beta) S^T$$

$$\beta = \alpha \cdot \exp(-\lambda \cdot D_{\text{ave}})$$

$$D_{\text{ave}} = \sum_{i=1}^{n} \left\| \overrightarrow{F^V}_i - \overrightarrow{F^V}_{\text{opt}} \right\| / n$$

$$\overrightarrow{F^V}_{\text{opt}} = \sum_{i=1}^{n} \overrightarrow{F^V}_{\text{i}} / n$$

$$S^V = 1 - D^V$$

Where:

- $S$is the metric of similarity in both textual and visual spaces;
- $S^V$is the similarity between visual features of I and $\overrightarrow{F^V}_{\text{opt}}$;
- S^T is the cosine similarity between textual features of I and $\overrightarrow{F^V}_{\text{opt}}$;
- β is the dynamic parameter of linear combination for similarity metric in both textual and visual spaces;
- $\alpha$and λ are the controlling parameters of relative contribution of RF in visual space;
- D_{ave} is the clicked image deviation in visual space;
- $\overrightarrow{F^V}_i$is the clicked image visual feature vector
- $\overrightarrow{F^V}_{\text{opt}}$ is the optimal query feature vector in visual space;
- $D^V$is the weighted Euclidean distance between visual feature of I and $\overrightarrow{F^V}_{\text{opt}}$

3.2. SRC-Based Textual Space

We have used the SRC algorithm for constructing an accurate and low dimensional textual space for the resulting web images. The author re-formalizes the clustering problem as a salient phrase problem of ranking. Given a query and the search result ranked list, it first parses the entire list of titles and snippets then all possible phrases extracted from the contents and five properties of each phrase will be calculated. Those consists of phrase frequency/inverted document frequency (TFIDF), length of phrase (LP), similarity of intra cluster (CSI), entropy of cluster (EC) and independence of phrase (INDP). These five properties are supposed to be relative to the phrases score of salience. In our case, snippets are comments and critiques. In the following, the current phrase is denoted as ω, and the document set that contains ωas $D(\omega)$. Then, the five properties can be given by

$$\mathrm{TFIDF} = f(\omega) \cdot \log \frac{N}{|D(\omega)|}$$

$$\mathrm{LP} = n$$

$$\mathrm{CSI} = \frac{1}{|D(\omega)|} \sum_{\mathrm{d_i} \in \mathrm{D}(\omega)} \cos(\mathrm{d_i}, \mathrm{c})$$

$$c = \frac{1}{|D(\omega)|} \sum_{\mathrm{d_i} \in \mathrm{D}(\omega)} \mathrm{d_i}$$

$$\mathrm{EC} = -\sum_{t} \frac{|D(\omega) \cap D(t)|}{|D(\omega)|} \log \frac{|D(\omega) \cap D(t)|}{|D(\omega)|}$$

$$\mathrm{INDP} = \frac{\mathrm{INDP_l} + \mathrm{INDP_r}}{2}$$

$$\mathrm{INDP_l} = -\sum_{t=l(W)} \frac{f(t)}{\mathrm{TF}} \log \frac{f(t)}{\mathrm{TF}}$$

Where f is a calculation of frequency

We use a single formula to combine them and calculate a single salient score for each phrase by using the above five properties. In our case, each term can be a vector and it is represented as

$$x = (\mathrm{TFIDF}, \mathrm{LP}, \mathrm{CSI}, \mathrm{EC}, \mathrm{INDP})$$

Therefore, in our experiments, we used linear regression model and is given by

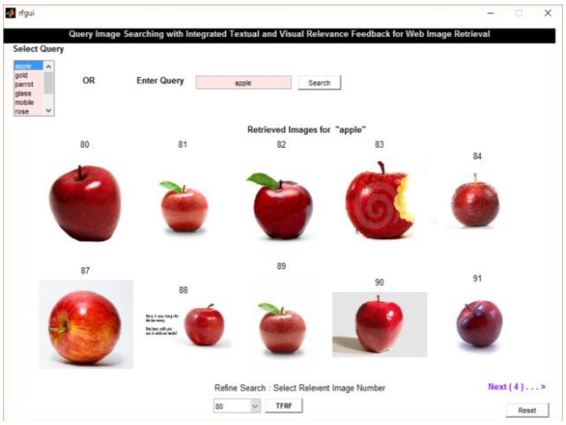
$$y = b_0 + \sum_{j=1}^{p} b_j x_j + e$$

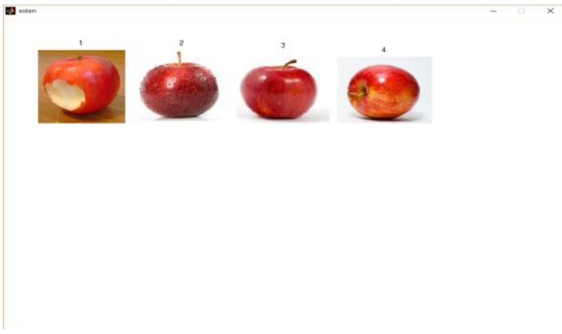
Where:

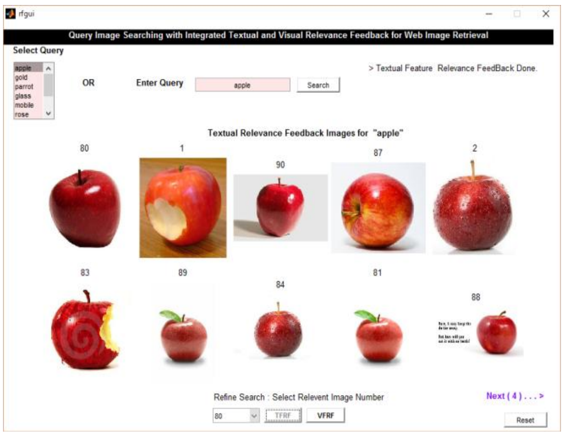
- e is a zero mean random variable;
- b_j is a coefficient defined by the condition that the square residuals sum is as small as possible.

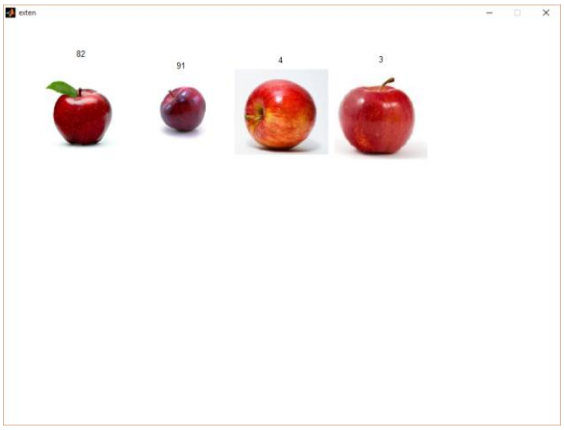
IV. Simulation Results

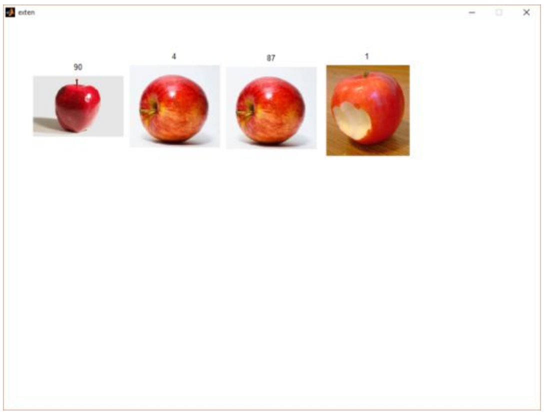
Experiments have been done in MATLAB 8.3 version environment with 4GB RAM and *i3* processor. We had considered a metadata set which has been taken from various photo forum sites. The images are '*apple', 'rose', 'gold', 'glass', 'knife', 'sun', 'sky'* and *'parrot'*etc., then for each query, we tested it with existing relevance feedback, TFRF, VFRF, TVRF and proposed ITVRF algorithms for retrieving the relevant images from given metadata base. All the experimental results have shown that the proposed algorithm has performed out well with improved precision and efficiency. Fig2 shows that the apple image retrieval with conventional RF then after textual, visual and proposed relevance feedback algorithms outputs have been displayed. Later, we had shown the relevant images of *sun*. Alsogiven the precision of various images with the conventional and proposed relevance feedback CBIR systems, in which we had achieved almost 99% of accuracy with an improved efficiency. Finally, we can conclude that the proposed algorithm is more robust among previous RF methods with improved precision, efficiency and even accuracy.

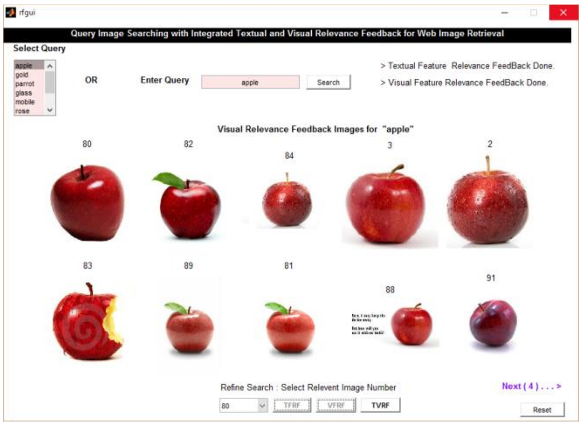


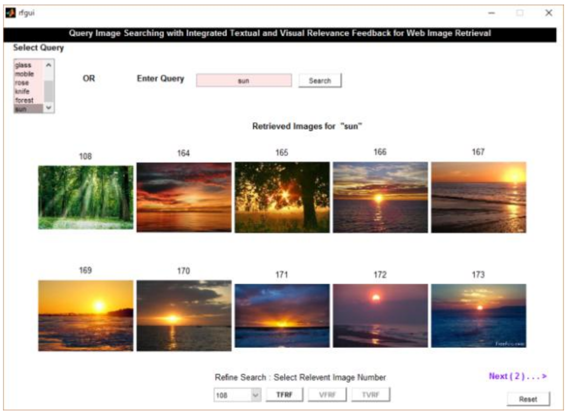


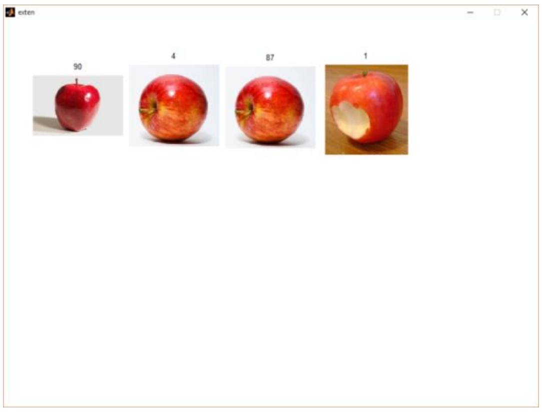


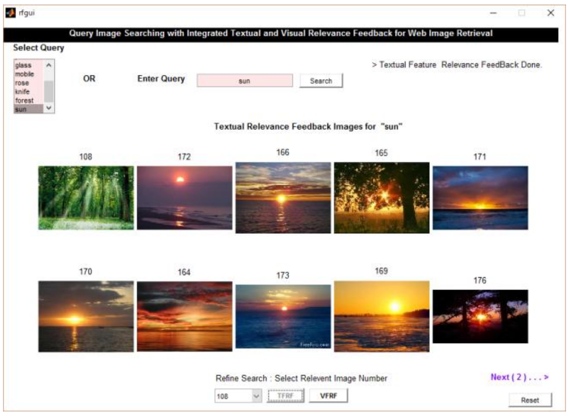


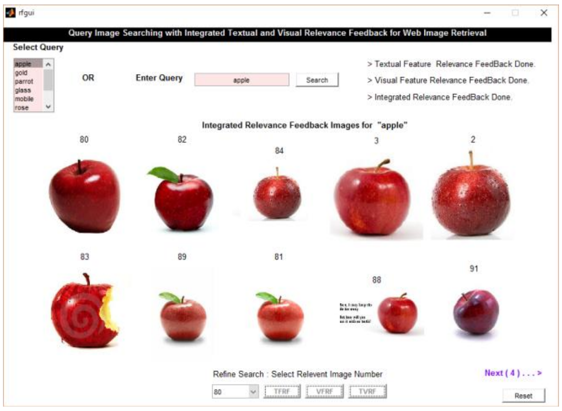


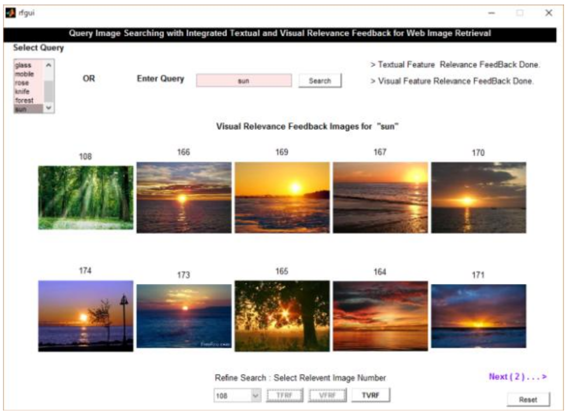

Query Image Searching with Integrated Textual and Visual Relevance Feedback for Web Image Retrieval
Select Query
OR
Enter Query
apple
Search
> Textual Feature Relevance FeedBack Done.
> Visual Feature Relevance FeedBack Done.
Visual Relevance Feedback Images for "apple"
Refine Search : Select Relevent Image Number
Next (4) . . . >
TFRF
VFRF
TVRF
Reset

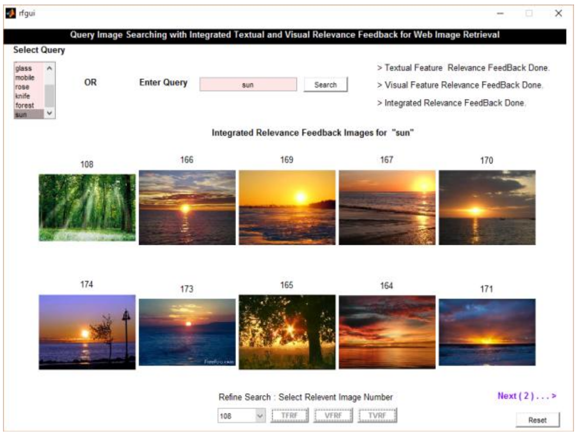

Query Image Searching with Integrated Textual and Visual Relevance Feedback for Web Image Retrieval
Select Query
OR
Enter Query
sun
Search
Retrieved Images for "sun"
Refine Search : Select Relevent Image Number
Next (2) . . . >
TFRF
VFRF
TVRF
Reset




Query Image Searching with Integrated Textual and Visual Relevance Feedback for Web Image Retrieval
Select Query
OR
Enter Query
sun
Search
> Textual Feature Relevance FeedBack Done.
Textual Relevance Feedback Images for "sun"
Refine Search : Select Relevent Image Number
Next (2) . . . >
TFRF
VFRF
Reset


Query Image Searching with Integrated Textual and Visual Relevance Feedback for Web Image Retrieval
Select Query
OR
Enter Query
apple
Search
> Textual Feature Relevance FeedBack Done.
> Visual Feature Relevance FeedBack Done.
> Integrated Relevance FeedBack Done.
Integrated Relevance Feedback Images for "apple"
Refine Search : Select Relevent Image Number
Next (4) . . . >
TFRF
VFRF
TVRF
Reset


Query Image Searching with Integrated Textual and Visual Relevance Feedback for Web Image Retrieval
Select Query
OR
Enter Query
sun
Search
> Textual Feature Relevance FeedBack Done.
> Visual Feature Relevance FeedBack Done.
Visual Relevance Feedback Images for "sun"
Refine Search : Select Relevent Image Number
Next (2) . . . >
TFRF
VFRF
TVRF
Reset



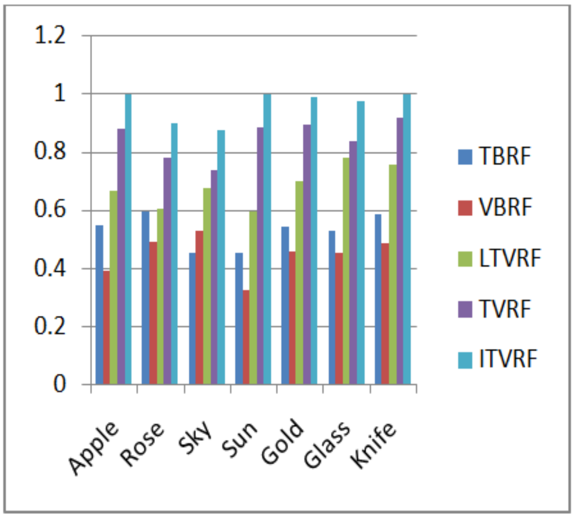


Fig.2 Comparison of proposed and existing CBIR techniques

V. CONCLUSION

In this letter we had proposed an adaptive CBIR scheme for large database systems using an integrated textual and visual relevance feedback (ITVRF). The performance of the CBIR system had improved in terms of more relevant images with good accuracy over existing relevance feedback systems also to reduce the computational complexity while improving the system efficiency. The proposed system has proven that this approach has got superior performance than the existing CBIR schemes.

REFERENCES

[1]. Chua, T.-S., Hung-ICengPung; Guo-Jun Lu; Hee-Sen Jong, "A concept based image retrieval system" Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences, Jan. 1994.

[2]. Eakins, John; Graham, Margaret, "Content-based Image Retrieval", University of Northumbria, New castle

[3]. X Wang, T S Chua and Al-Hawamdeh, "Probabilistic and semantic based retrieval in hypertext", Proc. of the South-East Asia Regional Computer Conf., pp. 25.01 -25.17 1992

[4]. Khan, S.M.H., Hussain, A. ;Alshaikhi, I.F.T., "Comparative Study on Content Based Image (CBIR)," International Conference in Advanced Computer Science Applications and Technologies (ACSAT), 2012.

[5]. Wan Siti, H. Munirah, W. Ahmad, M. Faizal and A. Fauzi, "Comparison of Different Feature Extraction Techniques in Content Based Image Retrieval For CT Brain Images," 10Th IEEE workshop on Multimedia Signal Processing, pp. 503-508, 2008.

[6]. B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, A.Yamada, "Colour and Texture Descriptors," IEEE Transactions on Circuits and Systems for Video Technology, 1998

[7]. P. S. Hiremath, S. Shivashankar, J. Pujari, "Wavelet Based Features for Color Texture classification with Application to CBIR," International Journal of Computer Science and Network Security, Vol. 6, No.9A, September 2006

[8]. TianYumin, Mei Lixia, "Image Retrieval Based on Multiple Features Using Wavelet," 5th IEEE International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'03), 2003.

[9]. M. R. Zare, R. N. Ainon, W. C. Seng, "Content-based Image Retrieval for Blood Cells," Third Asia International Conference on Modeling & Simulation, 2009

[10]. Prasad, B.G., Krishna, A.N., "Statistical Texture Feature Based Retrieval and Performance Evaluation of CT Brain Images" 3rd International Conference on Electronics Computer Technology (ICECT), Vol. 2, April 2011

[11]. Swati Agarwal, A.K. Verma, Preethvanti Singh, "Content Based Image Retrieval using Discrete Wavelet Transform and Edge Histogram Descriptor" International conference on Information Systems and Computer Networks, 2013

[12]. D. Ashok Kumar, J. Esther, "Comparative Study on CBIR based by Color Histogram, Gabor and Wavelet Transform," International Journal of Computer Applications (0975 – 8887) Volume 17– No.3, March 2011.

Author's Biography

Dr. Raja Murali Prasad received his bachelor's degree from the Institution of Engineers in 1989 and M. Tech in ECE from Pondicherry Engineering College in 1993. He worked in various engineering colleges as faculty member. Presently, he is working as faculty member in the Department of Electronics and Communication Engineering, Vardhaman College of Engineering, Hyderabad. He completed PhD at JNT University Anantapur. His research interests include digital communications, control systems and wireless communications.

Dr.G.A.E.Satish Kumar received his B.Tech degree in Electronics & Communications Engineering from Sri Krishnadevaraya University in 1995.He then received his M.E in Communication Systems from Gulburga University in 1999 and Ph.D.in Signal Processing from JNT University, Hyderabad in 2009.He entered in to teaching field in1998 as a Lecturer and latter promoted as Associate Professor & Professor. Presently he is working as Professor in ECE department in Vardhaman College of Engineering (Autonomous),Hyderabad (Telangana, India).He has Published 20 Research papers in National/International Journal/Conferences. His research interests include Signal & Image processing and Communications. He has guided around 14 M. Tech projects and guiding 04 Ph.D. Scholars.He is a member of IEI and life member of ISTE. He has 19 years of teaching experience.