International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:10, No:10, 2016

Image Retrieval Based on Multi-Feature Fusion for Heterogeneous Image Databases

N. W. U. D. Chathurani, Shlomo Geva, Vinod Chandran, Proboda Rajapaksha

***Abstract*—**Selecting an appropriate image representation is the most important factor in implementing an effective Content-Based Image Retrieval (CBIR) system. This paper presents a multi-feature fusion approach for efficient CBIR, based on the distance distribution of features and relative feature weights at the time of query processing. It is a simple yet effective approach, which is free from the effect of features' dimensions, ranges, internal feature normalization and the distance measure. This approach can easily be adopted in any feature combination to improve retrieval quality. The proposed approach is empirically evaluated using two benchmark datasets for image classification (a subset of the Corel dataset and Oliva and Torralba) and compared with existing approaches. The performance of the proposed approach is confirmed with the significantly improved performance in comparison with the independently evaluated baseline of the previously proposed feature fusion approaches.

***Keywords*—**Feature fusion, image retrieval, membership function, normalization.

I. Introduction

THE rapid development of multimedia and network technology has led to the wide use of image databases. Therefore, efficient and accurate image retrieval techniques are essential to satisfy user needs. As a result, content-based image retrieval (CBIR) has become an active research area in multimedia information retrieval. A number of CBIR techniques have been proposed [1]-[9] during past years.

These techniques have addressed different stages of CBIR process like feature selection [1], [4], [7]-[9], representation [3], [5], [10], indexing and searching [2], [6] in variety of ways to improve the retrieval quality. When considering feature selection, image features are vital to describe images in CBIR. So, there is a plethora of image features that has been found to describe image properties like colour, texture and shape of an image. Different features reflect the different characteristics of the image. Therefore, a single feature type may describe an image only in a specific angle.

A single image feature type is not adequate to differentiate images with the increasing size and variability of image databases. To overcome the shortcomings of single feature vector recognition algorithm, feature fusion such as a combination of colour, texture, and shape features was introduced to CBIR [1], [5] to cover a heterogeneous image dataset. Multi-feature fusion is one of the ways to improve retrieval performance in CBIR among other different techniques. The general solution for this technique can be obtained by combining two approaches:

(i) Feature engineering using distance combination, weightings and normalization of features.
(ii) Using trained classifiers with the derived features to optimise performance with training data.

This paper targets the improvement of the solution by focusing on feature engineering.

A simple feature fusion approach is, to combine all the features to generate a single feature vector, or to obtain a summation of distances over different features [1]. But this simple approach assumes that all features carry equal importance for retrieval. However, each feature has its own significance in image retrieval and in order to obtain effective outcome, the varying degree of importance in each feature needs to be captured. Some systems achieve this by using methods such as weighting schemes [3], [4], different distance measures [4], [5] or feature normalization [1].

Feature fusion has significant impact on CBIR and thus performance of feature fusion is highly dependent on features dimensions and ranges. As features have different variability, appropriately selected distance measures for each feature help to improve the retrieval performance. Feature normalization maps feature into fixed range and feature distribution must be appropriately normalised.

Five existing late feature fusion methods as shown in Table I (where result-lists from individual descriptors are fused during query time) have been compared in [2]. It was found that the addition of all scores per image with normalization (Z-score + CombSUM) outperform the other methods. Normalization is done using Z-score (mean and standard deviation) in their method which is different from the proposed distance normalization approach which is described in detail in Section II in this paper.

In this paper we describe a novel, simple yet effective approach to achieve linear feature fusion with combination of feature weights (significance) and distance normalization (distance distribution), which can be applied to any combination of features. This general approach is invariant to the distance measures, dimensions and ranges of the features, as a pre-defined membership function is used. Empirical evaluation is performed on a subset of the standard Corel dataset to validate the performance of this proposed approach and it is compared against other implemented and

N. W. U. D. Chathurani, Shlomo Geva, and Vinod Chandran are with the Department of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, Australia (e-mail: d.nanayakkara@hdr.qut.edu.au, s.geva@qut.edu.au, v.chandran@qut.edu.au).

Praboda Rajapaksha is with the Department of Computer Engineering, University of Uva Wellassa, Badulla, Sri-Lanka (e-mail: proboda @uwu.ac.lk).

independently evaluated approaches. This approach is further validated using Oliva and Torralba dataset.

TABLE I
SOME LATE FUSION METHODS COMAPARED IN [2]

CombSUM	Addition of all scores per image, without any normalization
BC+ CombSUM	Borda Count [11] originates from social theory in voting. An image with the highest rank on each of the feature similarity ranking lists gets highest votes. Votes across ranked-lists are naturally combined with CombSUM.
Z-score+ CombSUM	Z-score is a linear normalization per query which maps each score to its number of standard deviations above or below the mean score. Present the results with CombSUM.
IRP	The Inverse Rank Position [12] merges ranked lists. It is the inverse of the sum of inverses of the feature similarity rank scores for each individual feature for a given image from relevant feature similarity ranking lists.
HIS+ multiplication	HIS [6] is a non-linear normalization which maps each score to the probability of a historical query scoring a collection image below that score. Those probabilities combined with multiplication.

The rest of the paper is organized as follows. Section II describes an overview of the proposed feature fusion approach. Section III provides details of the experiment and the results obtained. Conclusions drawn from the research findings and future work are included in Section IV.

II. PROPOSED APPROACH

The image content can be described by colour, texture, shape, etc. These features describe images in different ways. The features that need to be used are selected according to the dataset which is going to be addressed. The features that we have used to represent images are based on colour, texture and shape, as we are considering general image databases. All these features are selected from an experimental evaluation. Then related feature weight and membership function is defined for each feature. Finally, when querying, the similarity is calculated by using these measures.

A. Image Features

Colour is the most straightforward visual feature used in CBIR systems. Colour can be represented in many ways. The more popular colour descriptors, namely colour histogram (ch) and colour moment (cm) were selected for this system as they have shown good retrieval performance in the literature [7]. Colour features are relatively robust and simple to represent. The colour histogram is efficient and insensitive to small changes in camera view point. It was adopted from [7] as it achieved better retrieval results using the YCbCr colour space, providing a closer match with human perception. The first order (mean), the second order (variance), and the third order (skewness) moments were shown to be effective and efficient in representing colour distribution and it helps to overcome the quantization effect in colour histogram.

Notwithstanding the fact that texture is not well defined, it is very helpful in describing real world images. We use Garbor wavelets (gabor) and Edge histogram descriptor (ehd) to capture texture feature. Gabor wavelets, as proposed by [8], have proven to be very useful texture analysis and re widely adopted in the literature. Four scales and six orientations were selected from experiments for the Gabor wavelet and the rotation and scale invariance property is achieved by the simple circular shift operation proposed in [8]. The mean and the standard deviation of each filter are used as the feature vector. The Edge histogram descriptor effectively describes heterogeneous textures. It captures the spatial distribution of edges and helps to extract different textures using five filters including vertical, horizontal, 45 degree, and 135 degree diagonal edges. If there is an arbitrary edge without any directionality, then it is classified as a non-directional edge [9].

Invariant moments (im) are used to describe the shape feature. Invariant moments are a compact representation on pixel distribution of a shape of an image which is invariant to translational, rotational and scale. Moments are limited to seven by the calculation, as the use of higher order moments result in being sensitive to noise and hence cause hindrance to accuracy.

All these features are selected as they have shown good individual performances in the literature [4], [7]-[9], as well as being further validated through preliminary experimental evaluation. The performance of feature fusion does not depend on the individual performance of features, but depends on diversification of the features, as well as the inter-relation of features. Therefore, some feature combinations may degrade the retrieval quality and adversely affect the performance of the individual features when used in isolation. A suitable combination of features has to be selected. We have tested other features such as Generic Fourier descriptor [13], and Discrete wavelet transform [14] using cross validation. However, experimental results were not promising with the combination of other features. We achieved the best performance with the combination of the five features described above from the separate experiments of sequential forward selection (add one feature in) and sequential backward selection (take one feature out) of features. Mean Average Precision (MAP) is used as performance measure to identify the retrieval quality.

B. Membership Function

Global image representation, as well as local representation, are used to validate the proposed late feature fusion method. The results demonstrate that the proposed feature-fusion achieves better retrieval results. Features are extracted from the full image for the whole dataset for global representation. Grid representation is used as local representation. Firstly the image is subdivided into nine non-overlapping blocks and then four overlapping blocks are generated by combining the sub-images by assuming that the main object of the image is generally located at the centre of the image. Each block is generated by merging four blocks that are generated from nine blocks. Five individual feature vectors are used to represent each image. The performance of the proposed fusion technique is not affected by the range and the length of feature vectors. So it maintains the normal form with absolute values.

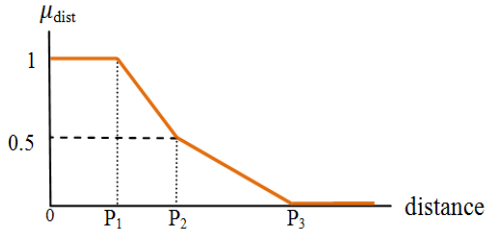

Fig. 1 Distance regions generated by piecewise linear function

C. Weights Calculation

Weight assignment for features is important in multiple feature fusion as different features have different significance. Each image in the database can be represented as follows and Table II gives notations for the features:

$$F_I=[\ f_1, f_2, f_3, f_4, f_5]$$

TABLE II
NOTATION OF FEATURES

Feature Index	f_1	f_2	f_3	f_4	f_5
Feature Name	ch	cm	dwt	ehd	gfd

w_i is the weight related to i^{th} feature f_i where (i=1:5) and w_i is considered as 1 (same significance) for each feature in simple feature fusion. Different weights are used in weighted feature fusion according to their relative single feature performance. The calculated feature weights are as below:

$$w_{f_1}=0.266,\ w_{f_2}=0.159,\ w_{f_3}=0.218,\ w_{f_4}=0.233,\ w_{f_5}=0.124$$

where; $\sum_{i=1}^{5} w_{f_i}=1$

Precision is used as the performance measure which is described in evaluation section. Related weight values are calculated according to the MAP values. MAP is calculated for each feature using training set of images as queries. Higher MAP for a feature means better the ability to retrieve correct images, higher the weight related to it. Here is a general solution for weight calculation. If we have well categorized specific dataset to improve results further, we can assign different weights for different categories for one feature, but that solution will be specific for the selected dataset.

D. Membership Function

A simple piecewise linear function is used to generate rules to define four regions of similarity for each feature. It is easily implemented and all the values are mapped to the interval [0, 1]. Regions are defined for each feature according to the distance measure in this function, as shown in Fig. 1 (0-P_1, P_1-P_2, P_2-P_3, P_3<). Four regions are selected by defining three points according to the ranked distance in ascending order as best (first), average (middle) and worse (last).

Average distance of the first five images, middle five images and last five are calculated individually, from the listed n number of images according to the similarity measure, which are used as training set and average is calculated (This n is described later in this section). Then these calculated averages of all the image categories are used to calculate point P_1, P_2, P_3 (as shown in Fig. 1), respectively, for each feature. As an example, calculation of P_1 for feature f_i ($P_{1\text{-}i}$) is as shown in (1):

$$P_{1\_i}=\frac{\frac{1}{N}\sum_{n=1}^{N} dist_n\_f_i}{C} \tag{1}$$

where N = number of images considered (N=5), C = number of categories in the data set (C=10 and C=8 for selected datasets), $dist_r\_f_i$ is the r^{th} ranked distance related to the feature f_i.

When searching images, a membership value must be computed for each feature vector by using the calculated distance. Equation (2) is used to map the distance to the value in the range of [0, 1] (least similar to most similar). Random n number of images are taken out from each class (half of the each class used as training set i.e. n=50 for Wang dataset) to generate this distance membership function and there were 500, 1,344 images altogether in the training set for Wang and Oliva and Torralba datasets respectively. The n numbers of images were selected randomly and the training set was changed from time to time by selecting different image sets to confirm that the performance of the proposed approach does not vary with the selected training set, which means performance is not heavily dependent on the selected dataset and not optimized for a particular training set. Finally, the average is taken in to consideration.

However, piecewise linear function must be defined for each dataset before using it. Each time training set must be used to define membership function. As this process is offline it will not affect the searching time, but it simplifies searching process because it helps to be invariant to feature's dimension, ranges, feature normalization and distance measure as it maps the distance values to the value in the 0-1 range.

$$\mu_{dist\_f_i}=\begin{cases}1 & if\quad dist\le P_{1\_i}\\ \left(\frac{0.5}{P_{1\_i}-P_{2\_i}}\right)dist+\frac{1-0.5*P_{1\_i}}{P_{1\_i}-P_{2\_i}} & P_{1\_i}<dist\le P_{2\_i}\\ \left(\frac{0.5}{P_{2\_i}-P_{3\_i}}\right)dist+\frac{0.5-0.5*P_{2\_i}}{P_{2\_i}-P_{3\_i}} & P_{2\_i}<dist\le P_{3\_i}\\ 0 & dist>P_{3\_i}\end{cases} \tag{2}$$

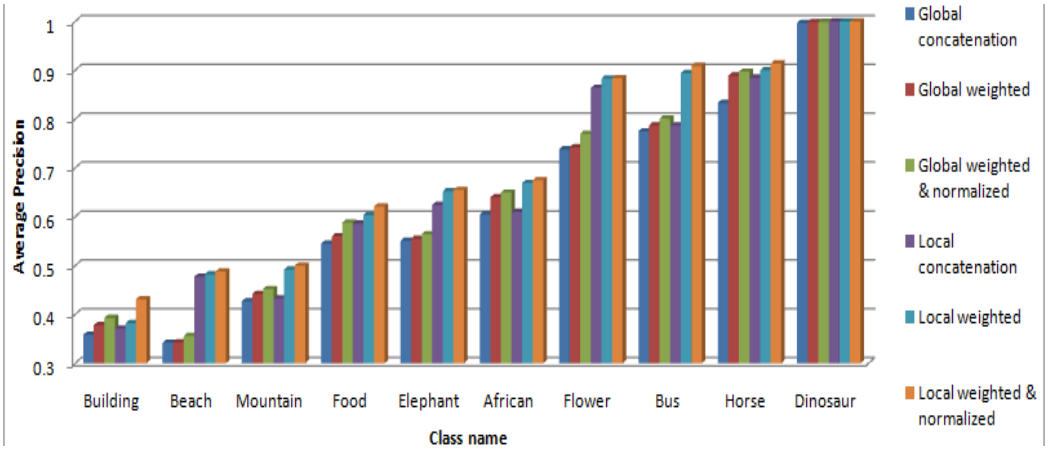

Fig. 2 Performance comparison of linear feature fusion for each class in Wang dataset

E. Similarity Measure

The performance of a CBIR system mainly depends on the particular image representation and similarity matching function employed. Different colour, texture and shape features are extracted for this image representation as we are targeting on general images. Related feature weights and their membership values related to the distance of the query are used in the similarity measure. The proposed approach is simple and easy to adopt and it is described as below.

Similarity is calculated by using defined weights and membership values for each feature in the proposed approach. Different features have different significance, and hence, the significance of each feature is considered for multi-features based retrieval. Euclidean distance is used as the distance measure. Similarity between image *Q, and I* can be calculated as below;

Step 1. First, the distance $Dist(f_{iQ}, f_{iI})$ is measured between image Q and image I for feature f_i , and computed $dist_{f_{i\_QI}}$.

Step 2. Membership value $\mu_{dist\_f_i}$ is computed for feature f_i from the distance membership function by using the $dist_{f_{i\_QI}}$ in (2).

Step 3. Weight of the feature f_i is considered as w_i. These weights are as shown in the Section II.*C* weight calculation.

Step 4. Repeat the steps 1 to 3 for each feature f_i that is used in the system (here 5) and computed the membership value $\mu_{dist\_f_i}$ and weight value w_i.

Step 5. Similarity measure *Sim*(*Q,I*) is computed, fusing all the feature measures using (3).

$$Sim(Q,I) = \sum_{i=1}^{N} w_i * \mu_{dist\_f_i} \quad (3)$$

Step 6. Repeat steps 1-5 for the whole dataset and list them all. Then rank the list according to *Sim* (*Q,I*) and retrieve the top ranked images.

III. Evaluation

To evaluate the effectiveness of the proposed approach, experiments were performed on two general purpose image datasets

A. Datasets

The first dataset is composed of 1,000 manually selected images from Corel image database (Wang dataset) [15]. This database contains diverse images of 10 classes with 100 images in each category namely African people, Beaches, Buildings, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains, and Food. These images are JPEG with the resolution of 384x256 or 256x384.

For further validation Oliva and Torralba dataset [16] is used. It includes 2688 images classified into eight categories with different class sizes namely Coast and beach (360), Open country (410), Forest (328), Mountain (374), Highway (260), Street (292), City center (308), and Tall buildings (356). Images are in JPEG format with the resolution of 256x256. As these datasets are well classified, it is possible to quantitatively evaluate and compare the performance.

B. Experimental Setting

The performances of global individual features are considered to calculate relative weights and the membership function. Since the goal of feature fusion is to achieve better retrieval results than any single feature, the best result of single feature (global ch) performances is used as the baseline.

The most common evaluation measure in information retrieval is precision and it is the ratio between the number of relevant images retrieved and the total number of images retrieved. Fusion-based similarity measures are compared based on average precision by evaluating top 20 retrieval results. Furthermore, average precision at N is calculated (AP@N). A retrieved image is considered a correct match if and only if it is in the same category as the query image. It is assumed that the results can be improved further by using a large training set. In these experiments 1,000, 2,688 images are used, with half of the images used for training and other half used for testing. It may be noted that this method solves

the problem of high dependency on feature dimensions and ranges in feature fusion.

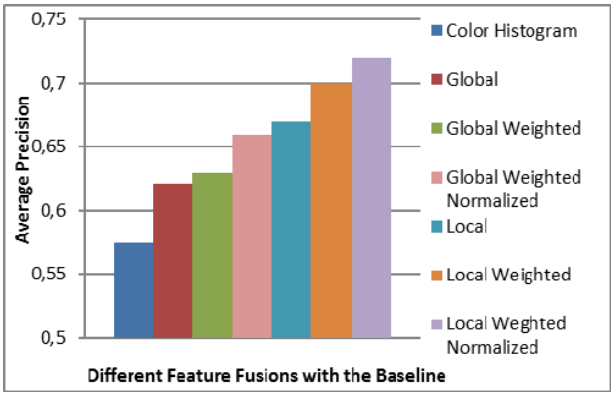

Fig. 3 Performance comparison of feature fusion with the baseline on Wang dataset (AP@20)

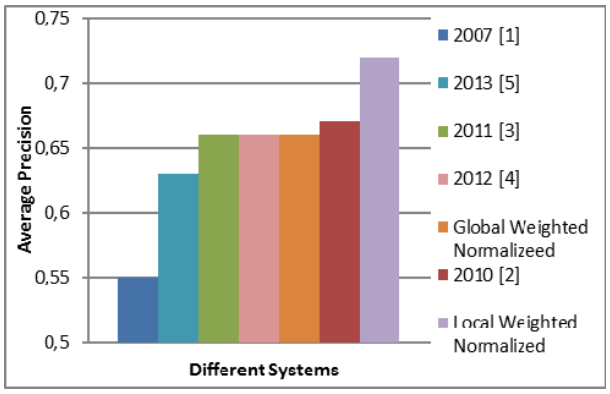

Fig. 4 Performance comparison of different systems on Wang dataset (AP@10). 2010 [2] is (Z-score + CombSum) the best late fusion method from the compared methods in Table I

C. Results

Fig. 2 shows the performance comparison of feature fusion with the baseline for each class (AP), where performance of colour histogram is considered as the baseline. Different feature fusion methods given below are compared.

i. Simple global and local feature fusion (concatenation) by considering each feature with equal significance for retrieval.
ii. Weighted global feature fusion, weighted local feature fusion.
iii. Global feature fusion with weights and distance normalization, local feature fusion with weights and distance normalization.

Fig. 3 further elaborates the performance. Local representation (grid) shows higher performance (MAP=0.67, MAP=0.7, MAP=0.72, for case i, ii and iii) compared to the performance of global representation (MAP=0.62, MAP=0.63, MAP=0.66, for case i, ii and iii). Weighted feature fusion shows higher performance (MAP=0.7 in local and MAP=0.63 in global) than the performance of simple feature concatenation (MAP=0.67 in local and MAP=0.62 in global). Performance (MAP=0.72 in local and MAP=0.66 in global) of weighted feature fusion in combination with distance normalization gives the highest performance in both global and local representation. According to the results obtained, the proposed approach shows higher performance.

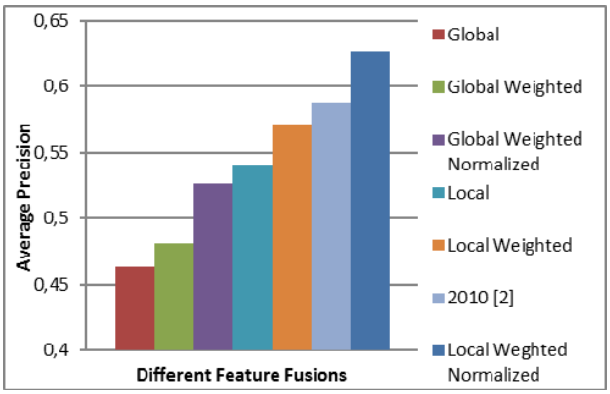

Fig. 5 Performance comparison of feature fusion for Oliva and Torralba dataset with Z-score + CombSum fusion (AP@20)

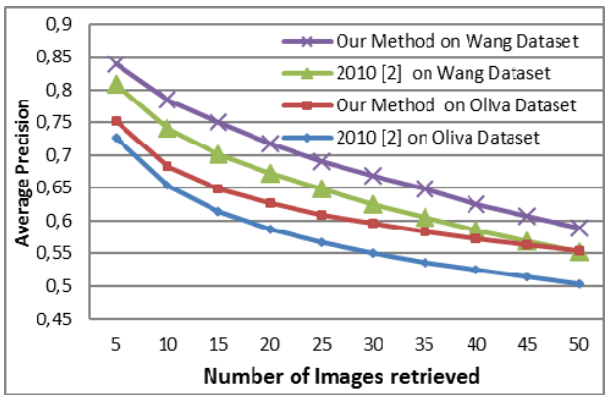

Fig. 6 Average precision at N (AP@N)

Fig. 4 shows the performance comparison with other systems which have been used the Wang dataset for evaluation. Our system outperforms the other systems by obtaining MAP of 0.72. All the other systems (addition of all scores or merging features [1], [3], Weighted distance [4], [5]) show MAP less than 0.66 except one which was proposed in [2]. In [2], the authors have tested five feature fusion methods, as shown in Table I, and found that the addition of all scores per image with normalization (Z-score + CombSum) achieves the best performance. Please refer to [2] for a detailed description of these methods, as we consider only the best performed one from [2]. Z-score + CombSum method is tested on Wang dataset and achieved only MAP of 0.67 for local feature fusion (2nd best performance of compared performances). Z-score + CombSum is the best among five feature fusion approaches that had been compared and our proposed approach is superior to that best late fusion method described in [2]. The proposed approach shows superior performance in both local and global representation.

The proposed approach was further validated using Oliva and Torralba general purpose image dataset. Fig. 5 shows the performance comparison of different fusions, as in Fig. 3, the proposed feature fusion method shows an improvement in performance on this dataset as well, seeing that our method achieves 0.63 MAP while Z-score + CombSum achieves 0.59 MAP. So it is confirmed that the proposed approach can be applied to any database and it is not optimized for one dataset. While this approach has its advantages as mentioned, the main drawback of the approach is to be trained in the beginning, which is an off-line process.

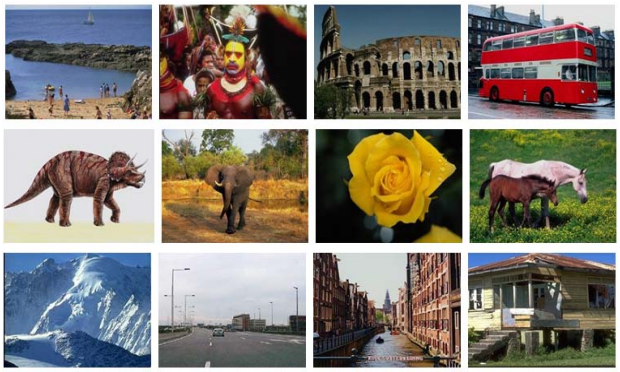

Fig. 7 Sample images covering Wang and Oliva and Torralba datasets

Retrieval quality of the proposed approach is assessed by calculating MAP@N (N is the number of images retrieved at a time) on both the datasets. Fig. 6 shows MAP @ N for Z-score + CombSum [2] and our method (weight + normalization). From that it can be seen that our late fusion method is better than the other methods and has good retrieval performance.

Fig. 7 shows some images from different classes of Wang and Oliva and Torralba datasets (First two rows from Wang and third row from Oliva and Torralba dataset).

The authors have shown that the proposed approach is better to obtain precise CBIR results than existing methods using two general purpose datasets.

IV. Conclusion and Future Work

First, the best feature combination is found from different low level features using cross validation. Then using those features a simple fusion-based similarity matching approach is proposed based on a weighted combination of similarity measures of different features according to their relative performance and distance normalization. A simple membership function is used to normalize the distances to [0, 1] interval to remove the effect of biasing due to the length of the feature vectors, and the large values of distance. The proposed approach can be easily adopted in any feature combination. The proposed approach is tested on global representation, as well as local representation and observed the improvement in retrieval quality. Moreover, the proposed system shows superior performance in retrieval quality relative to the existing feature fusion approaches. Dynamic feature weighting will be done according to the given query image in the future to improve retrieval quality further.

References

[1] P. S. Hiremath, J. Pujari, "Content based image retrieval based on colour, texture and shape features" In 15th International Conference on Advance Computing and Communications, 2007 pp. 780-784.
[2] S. A. Chatzichristofis, A. Arampatzis, "Late fusion of compact composite descriptors for retrieval from heterogeneous image databases", In 33rd International Conference on Special Interest Group on Information Retrieval SIGIR, 2010, pp.825-826.
[3] X. Yuan, J. Yu, Z. Qin, T. Wan, "A SIFT-LBP image retrieval model based on bag-of-features". In Proc. IEEE ICIP 18th International Conference on Image Processing, 2011, pp. 1061-1064.
[4] M. H. Saad, H. I. Saleh, H. Konbor, M. Ashour, "Image retrieval based on integration between YCbCr colour histogram and shape feature", In Proc. ICENCO 7th International Computer Engineering Conference, 2012, pp. 97-102.
[5] N. S. Mansoori, M. Nejati, P. Razzaghi, S. Samavi, "Bag of visual words approach for image retrieval using colour information", In Proc. ICEE 21st Iranian Conference on Electrical Engineering, 2013, pp. 1-6.
[6] A. Arampatzis, J. Kamps, "A signal-to-noise approach to score normalization", In ACM International Conference on Information and Knowledge Management CIKM, 2009, pp797–806.
[7] G. Qiu, "Indexing chromatic and achromatic patterns for content-based", Journal of pattern recognition, B2002, pp. 1675-1686.
[8] M.H. Rahmana, M.R. Pickering, M.R. Frater, "Scale and Rotation Invariant Gabor Features for Texture Retrieval", In Proc. DICTA International Conference on Digital Image Computing Techniques and Applications, 2011, pp. 602-607.
[9] S. Agarwal, A.K. Verma, P. Singh, "Content Based Image Retrieval using Discrete Wavelet Transform and Edge Histogram Descriptor", In Proc. ISCON International Conference on Information Systems and Computer Networks, 2013, pp. 19-23.
[10] V. Takala, T. Ahoen, M. Pietikainen, "Block-based methods for image retrieval using local binary patterns", In Proc. of the 14th Scandinavian Conference on Image Analysis, 2005, pp. 882-891.
[11] J. A. Aslam , M. Montague. "Models for Metasearch" , Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2001, pp. 276–284.
[12] M. Jovic, Y. Hatakeyama, Yutaka, F. Dong, K. Hirota, " Image Retrieval Based on Similarity Score Fusion from Feature Similarity Ranking Lists ", Fuzzy Systems and Knowledge Discovery, 2006, pp. 461-470
[13] Y. Minaqiang, K. Kidiyo, R. Joseph, "A survey of shape feature extraction techniques", Journal of Pattern Recognition, 2008, pp. 43-90.
[14] R. Manthalkar, P. K. Biswas, B. N. Chatterji, "Rotation Scale invariant Texture Features Using Discrete Wavelet Packet Transform", Journal of Pattern Recognition, 2003, pp. 2455-2462.
[15] J. Li, J. W. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach", IEEE transaction on Pattern Analysis and Machine Intelligence, vol. 25(9), 2003, pp. 1075-1087.
[16] A. Oliva, A. Torralba, "Modeling the shape of the scene: A Holistic representation of the spatial envelope, International Journal of Computer Vision, vol. 42(3), 2001, pp. 145-175.