International Journal of Natural and Engineering Sciences 4 (1): 95-101, 2010
ISSN: 1307-1149, www.nobel.gen.tr

Robust and Effective Frame Work for Image Retrieval Scheme Using Shift Invariant Texture and Shape Features

Tohid Sedghi[1]* Mehdi C. Amirani[1] Majid Fakheri[1]
[1]Department of Electrical Engineering, Urmia University, Iran

Corresponding author
E-mail: st_t.sedghi@urmia.ac.ir

Received: November 01, 2009
Accepted: Januraary 29, 2010

Abstract

We aim to take advantage of both shape and texture properties of images to improve the performance of the image indexing and retrieval algorithm. Beside that a framework for partitioning the image into non-overlapping tiles of different sizes which results in higher retrieval efficiency is presented. In new approach, texture features are obtained with use of Shift Invariant dual-tree *complex wavelet transform* (SIDT CWT) method. According to the new algorithm, image is divided into different regions, each of them extracting features of Energy, standard deviation of each of the SIDT CWT which serve as local descriptors of texture. Invariant momentums are then used to record the shape features. The combination of the texture features and shape features provide a robust feature set for image retrieval. The most similar highest priority (MSHP) principle is provided for matching the images. Experimental result show that the proposed method yield higher retrieval accuracy than some conventional methods even though its feature vector length and feature generating time of query image is less than those of other approaches.

Keywords: Complex Transform, local descriptors, Regioning, tile features, Image Retrieval.

INTRODUCTION

Content Based Image Retrieval (CBIR) is a technique used for retrieving similar images from an image database. The most challenging aspect of CBIR is to bridge the gap between low-level feature layout and high-level semantic concepts. CBIR systems have used local color and texture features [1-5] for increasing the retrieval accuracy, but nevertheless they have reached to the desirable goal.

In region based image retrieval (RBIR) systems [4-5], the image is segmented into regions based on color and texture features. The regions are close to human perception and are used as the basic building blocks for feature computation and similarity measurement. RBIR systems have been proven to be more efficient in terms of retrieval performance. In [4], images are compared based on individual region to region similarity. Precise image segmentation has still been an open area of research; for example the integrated region matching (IRM) algorithm of [4] proposes an image-to-image similarity based on combining all the regions between the images. In [4] and our approach, every region is assigned significance worth based on its size in the image. In [8], fuzzy features are used to capture the shape information. Shape signatures are computed from blurred images and then global invariant moments are used as the shape features. The retrieval performance is shown to be better than a few of RBIR systems such as those in [3-4] [8-9]. The studies mentioned above clearly indicate that, in CBIR, local features play a significant role in determining the similarity of images along with the shape information of the objects. Precise segmentation is not only difficult to achieve but also is so critical in object shape determination. In [10], a windowed search over location and scale is shown to be more effective in object based image retrieval than the methods based on inaccurate segmentation.

The objective of this paper is to develop a technique which captures local texture descriptors in a coarse segmentation framework of grids. The new method has a shape descriptor in terms of invariant moments computed on the edge image. The image is partitioned into different sizes of non-overlapping tiles. A new framework is used for texture analysis. The features computed on these tiles serve as local descriptors of texture. Invariant moments are used to serve as shape features. The combination of these features forms a robust feature set in retrieving applications. Then, an integrated matching procedure based on the adjacency matrix of a bipartite graph between the image tiles is provided, similar to the one discussed in [4], which yields image similarity. Our

method is similar to IRM, but so simple and less time consumer since all of the RBIR systems are complicated due to using different kinds of complicated algorithms which makes them be time consumption. We note that in any CBIR system, fast retrieval is the main objective. The experimental results are compared with the methods of [3-4], [8-9], [13-14], and [15]. The results indicate that the new method performs better.

The rest of the paper is organized as follows. In Section 2, we explain the proposed system. In Section 3, we present experiments and discuss about the results, and finally Section 4 concludes the paper.

MATERIAL and METHODS

Fig. 1 shows the comprehensive block diagram of the proposed CBIR system. In the following, we describe different parts in detail.

Regioning and Blocking

Each image is partitioned into 9 non-overlapping tiles. These tiles will serve as the local texture descriptors for the image. Features extracted from the transformed image tiles are used for texture similarity. Feature extraction method will be completely discussed in Section 2.4. Two different databases were used for experiments which comprise of images of sizes either 256 x 384 or 384 x 256. After partitioning, the sizes of each individual tile will be 192*64, 96*128, and 192*128 which is shown in Fig. 2a. here, we have supposed that the most important part of image is its centre, i.e. the centre contains the most valuable information compared to the side tiles. Hence, we apply region processing in such a way that image centre appears in a large tile and the sides of image in small tiles Fig. 2a. We assign weights to each tile in order to demonstrate the significance of tiles.

Integrated Im age MatchingIn our method, a tile from the query image is allowed to be matched to any tile in the target image. A tile will participate in the matching process only once. A bipartite graph of tiles for the query image and the target image is built. The labelled edges of the bipartite graph indicate the distances between the tiles.

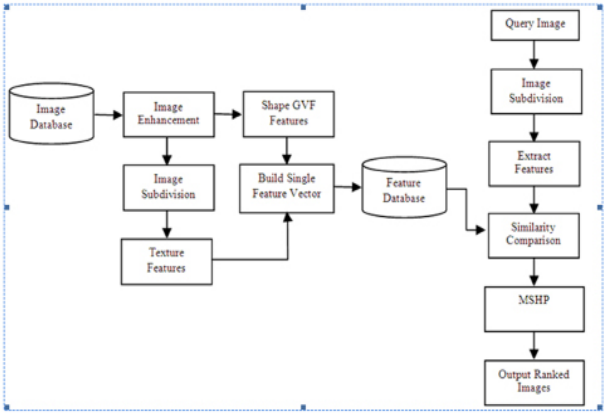


Fig. 1. System block diagram

A minimum cost matching is performed for this graph. Since, this process involves too many comparisons; hence the method has to be implemented efficiently. To do so, we have designed an algorithm for finding the minimum cost matching based on MSHP rule using the adjacency matrix of the bipartite graph. Here, the distance matrix is considered as the adjacency matrix. The minimum distance d_{ij} of the matrix is computed between tile i of the query image (define query image) and tile j of target image (define target image). The distance is recorded and the row corresponding to tile i and column corresponding to tile j are blocked. This will prevent tile i of query image and tile j of target image from further participating in the matching process. The distances between tile i of query image and other tiles of target image and also the distances between tile j of target image and other tiles of query image are ignored (because every tile is allowed to participate in the matching process only once). This process is repeated until every tile finds a matching. Therefore, the complexity of matching procedure is reduced from V(n^2) to V(n), where n is the number of tiles involved.

The integrated minimum cost match distance between images is now defined as:

$$D_q = \sum_{i=1,n} \sum_{j=1,n} d_j \quad (1)$$

where D_q is the distance between q (query) and t (target) images and d_{ij} is the best-match distance between tile i of query image q and tile j of target image. Best match means that the smallest distance between one tile of query image and target images.

Feature set

The feature set comprises texture and shape descriptors derived as follows:

Texture, Shift Invariant Dual-Tree Complex Wavelet Transform

Basis function of DT-CWT it is orthogonal and can be computed faster than Gabor filter. The main benefits of the SIDT CWT over the real DWT are that the complex wavelets are approximately shift invariant and have separate sub-bands for positive and negative orientations. Note that the conventional separable real wavelet suffers from the lack of shift invariance, provide just three orientations while having a poor directional selectivity and also cannot distinguish between 45^o and -45^o directions. The CWT attains these properties by replacing the tree structure of the conventional wavelet transform with a dual tree. At each scale, one of the trees produces the real part of the dual-tree complex wavelet coefficients, while the other tree produces the imaginary part [7][11]. The extra redundancy allows a significant reduction of aliasing terms and causes the complex wavelets to become approximately shift invariant. Translation causes large changes to the phase of the wavelet coefficients, but

their magnitude (and hence energy) is much more stable. By using even and odd filters alternately in these trees, it is possible to achieve overall complex impulse responses with symmetric real parts and anti symmetric imaginary parts. Further study about SID CWT can be found in [11]. In present work, all images of both Image sets are decomposed using SIDT-CWT(up to third level of decomposition), Gabor wavelets and standard wavelets, two kind of feature sets are produced feature set 1(FS1) for image set 1and feature set 2 (FS2) for image set2. The energy and standard deviation (STD) were computed on each sub-band separately by using this two parameters values. The Energy (E_l) and standard deviation(σ_l) of the *l*th sub-band is computed as follows:

$$E_l = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} |W_l(i,j)| \qquad (2)$$

$$\sigma_l = \sqrt{\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} |W_l(i,j) - \mu_l|^2} \qquad (3)$$

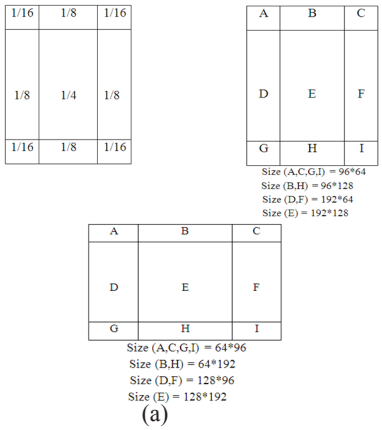
Where $W_l(i,j)$ is the *l*th wavelet decomposed sub-band, M*N is the size of wavelets-decomposed sub-bands and μ_l is the mean value of the *l*th sub-band. Our main feature vector is constructed by using sub features that related to different tiles of image. A sub feature vector is now constructed using σ_l and E_l resulting sub feature for *n* number of total sub-bands and for *k*th tile of regioned image as follows:

$$\overline{f}^{k}_{\sigma E} = [\sigma_1^k, \sigma_2^k, ..., \sigma_n^k, E_1^k, E_2^k, ..., E_n^k] \qquad (4)$$

Dimension of feature vector will be (n*2), so each tile of image has a sub feature vector with size of 40 elements. Next, we assign different weights for each of sub feature vectors, as depicted in Fig. 2. It is observed that the larger partitions are assigned larger weights. Finally, we combine 9 weighted vectors of 9 regions as a vector shown in the following:

$$F = (1/16)\ \overline{f}^{k}_{\sigma E} + (1/8)\ \overline{f}^{k}_{\sigma E}. + (1/16)\ \overline{f}^{k}_{\sigma E}. + (1/8)\ \overline{f}^{k}_{\sigma E}. + (1/4)\ \overline{f}^{k}_{\sigma E}. + (1/8)\ \overline{f}^{k}_{\sigma E}. + (1/16)\ \overline{f}^{k}_{\sigma E}. + (1/8)\ \overline{f}^{k}_{\sigma E}. + (1/16)\ \overline{f}^{k}_{\sigma E}. \qquad (5)$$

Appling sub features combination technique has advantages:



(a)

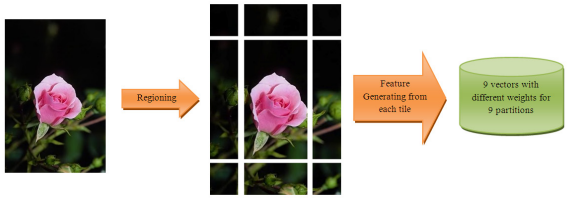


Fig. 2 a) Image Regioning for extracting textural features weights of partitions,
b) Textural feature generating.

1) It reduced dimension of main feature vector, it causes to use less memory usage.
2) Central tile of image has a specific importance in image retrieval.

For creation of feature set above, procedure is repeated for all the image of image sets and theses feature vectors are stored in feature database. Table1 demonstrate comparison of averaged retrieval accuracy for both image sets when only textural features have been generated.

Shape feature

As table1 Shows retrieval accuracy is mediocre, gaining higher retrieval percentage requiring to use other features, because by utilizing only textural features, textural information of an image are capture and other rich information is missed so this motivated to take advantage of some beneficent features such as shape features.The gradient vector flow (GVF) field gives excellent results on concavities supporting the edge pixels with opposite pair of forces, obeying force balance condition in one of the four directions (horizontal, vertical and diagonal) unlike the traditional external forces which support either in the horizontal or vertical directions only. In addition, [6] has used for shape extraction features and simulation results show better performance to other approaches such as hue and Zernike moments.

The algorithm for edge image computation is given below:

1. Convert the input colour image to gray scale one.
2. Blur the image using a Gaussian filter.
3. Compute the gradient map of the blurred image.
4. Compute GVF [6].
5. Filter out only strong edge responses. where σ is the standard deviation of the GVF.
6. Converge onto edge pixels till edge image is yielded.

After Edge computation, Translation, rotation, and scale invariant one-dimensional normalized contour sequence moments are computed on the edge of image [12]. The gray level edge images of *R*, *G* and *B* individual planes are taken and the shape descriptors are computed as follows:

$$F_1 = \frac{(\mu_2)^2}{m_1} \quad ; \quad F_2 = \frac{\mu_3}{(\mu_2)^{3/2}} \quad ; \quad F_3 = \frac{\mu_4}{(\mu_2)^2} \quad ; \quad F_4 = \overline{\mu_5} \qquad (6)$$

Where

$$m_r = \frac{1}{N} \sum_{i=1}^{N} [z(i)]^r \quad ; \quad \mu_r = \frac{1}{N} \sum_{i=1}^{N} [z(i) - m_1]^r \quad ; \quad \overline{\mu_r} = \frac{\mu_r}{(\mu_2)^{r/2}} \qquad (7)$$

Table1. Average Retrieval Accuracy of both image sets only for textural features (100 images

Feature Measures	Gabor Wavelets	Standard wavelets	SIDT-CWT
FS1	42.2%	40.1%	52.3%
FS2	75.7%	70.3%	80.1%

In the above, $z(i)$ is the set of Euclidian distances between the centroid and all N boundary of the digitized shape. A total of 12 features result from the above computations (4 features for each of R, G and B). In addition, moment invariants to translation, rotation, and scale are obtained on R, G and B planes individually considering all the pixels [12]. These transformations are summarized as below:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \quad \text{(which p and q you use)} \quad (8)$$

where

$$\gamma = \frac{p+q}{2} + 1 \text{ (Central moments)} \quad ; \phi = \eta_{20} + \eta_{02}$$

(Moment invariant) (9)

The above computations will yield 3 additional features amounting to a total of 15 features. Canberra distance measure is used for similarity comparison in all the cases. It allows the feature set to be in unnormalized form. The Canberra distance measure is given by:

$$canbdist = \sum_{i=1}^{d} \frac{|x_i - y_i|}{|x_i| + |y_i|} \quad (10)$$

where x and y are the feature vectors of image sets and query image, respectively, of length d.

RESULTS and DISCUSSION

In this section, we evaluate the performance of the proposed method. For this purpose, we consider 2 Image sets and apply the new method and different methods mentioned before on them. We compare their performances from different points of view.

Image sets

Image set 1: The first used database is downloaded from the website http://wang.ist.psu.edu/docs/related, which contains 1000 Corel images with ground truth. The image set includes 100 different images in each of 10 categories. The images are of the sizes 256 x 384 or 384 x 256. Fig. 3 depicts some image selected from database 1 and Table 2 shows the 10 categories.

Image set 2: The second database images are randomly picked up from the websites http://www.mcsh.kh.edu.tw and http://co25.mi.com.tw. Also some other images were downloaded from http://wang.ist.su.du/IMAGE. Fig. 4 shows some of the query and database

Table 2. Ten classes of image set 1

Classes	Semantic name
1	African people and village
2	Beach
3	Building
4	Buses
5	Dinosaurs
6	Elephants
7	Flowers
8	Horses
9	Mountains
10	Food

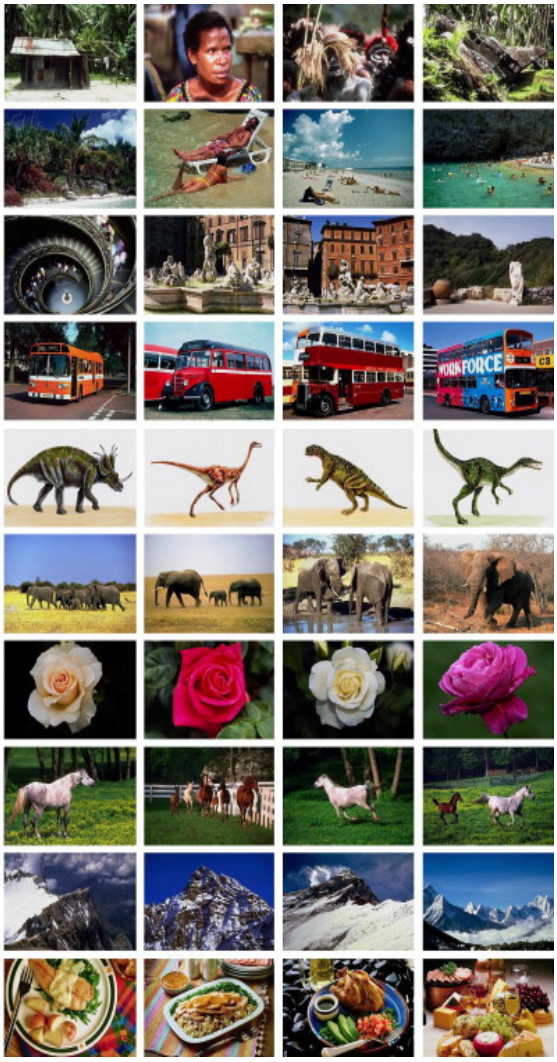


Fig. 3. Sample images from the Wang's databases, where the images are randomly taken from image set 1.

Fig. 4. Some examples of image set 2.
a) Some query image
b) the database images corresponding to the images in (a).

images. Further, we have used Lin [13] complete Image sets, where there are two sets of images SetD={ $I_1^d, I_2^d, I_3^d, ..., I_{1051}^d$ } and SetQ={ $I_1^q, I_2^q, I_3^q, ..., I_{1051}^q$ }.

Each set contains 1051 full color images. The images in SetD are employed as the database images and those in SetQ are used as the query images.

3.2 Experimental results on Image sets

3.2-1 Performance of the proposed system on image set1

The results are benchmarked with the other systems as in [3-4], [8-9], and [13] using the same database. The quantitative measure is the average precision [3] as explained below:

$$p(i) = \frac{1}{100} \sum_{1 \le j \le 100,\ r(i,j) \le 100,\ ID(j)=ID(i)} 1 \quad (11)$$

where $p(i)$ is the precision of query image i, $ID(i)$ and $ID(j)$ are category IDs of image i and j, respectively, which are in the range of 1 to 10. $r(i, j)$ is the rank of image j (i.e. position of image j in the retrieved images for query image i, which is an integer between 1 and 1000). This value is the percentile of images belonging to the category of image i in the first 100 retrieved images. The average precision $p(i)_t$ for category $t (1 \le t \le 10)$ is given by

$$p(i)_t = \frac{1}{100} \sum_{1 \le j \le 100,\ D(i)=t} 1 \quad (12)$$

The comparison of experimental results of the proposed method with the other retrieval systems reported in the literature [3-4], [8-9], and [13] is presented in Table 3. It is obvious that the proposed method achieves more average precision of various images than the other five methods. The FIRM [3] and SIMPLIcity [4] algorithms are both segmentation based methods. Since in these methods, textured and non textured regions are treated differently with different feature sets, their results are claimed to be better than histogram based method [9]. Further, edge based system [8] is at par or at times better than SIMPLIcity [4] and FIRM [3]. In comparison

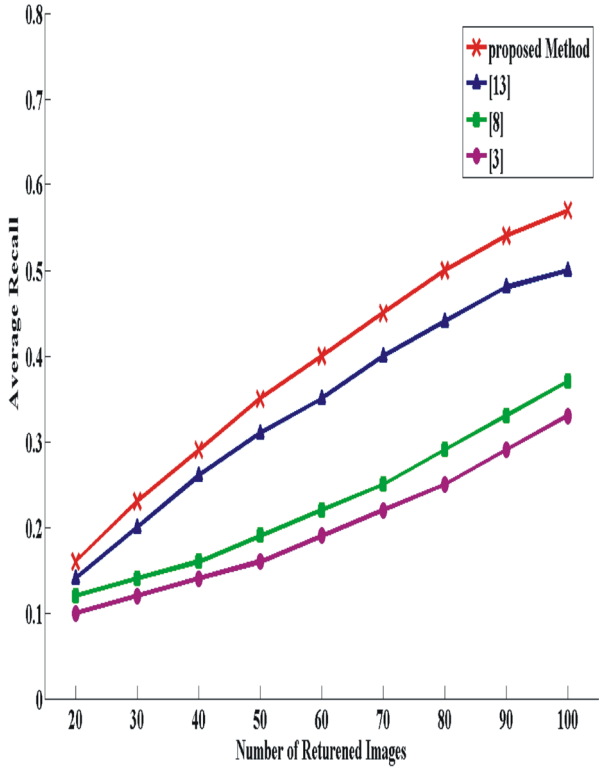


Fig. 6. The average recall of these methods on image set 1.

between the proposed method and that of [13], we observe that in most of the categories, our proposed method performs better than these systems. Further, the recall of each retrieval image and the average recall of the various images were calculated. The average precision of the retrieval results of the various images with the number of returned images, and the average recall are demonstrated in Fig. 5 and Fig. 6. The experimental results clearly reveal that for the first 20–100 returned images of the 1000 ten-category image set1, the new method is significantly superior In the average recall experiment, to the methods of [3-4], [8-9], and [13].

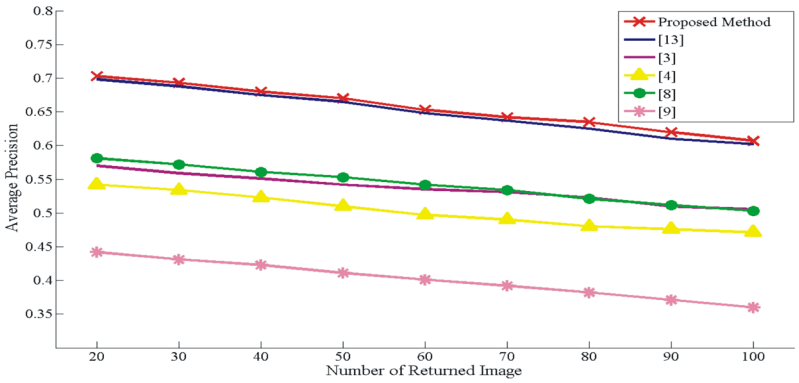


Fig. 5. Average precision of different methods on image dataset

Table 3. Comparison of average precision (%) of the proposed method and other standard retrieval systems [3-4], [8-9], [13] when 100 images of dataset 1 are retrievable.

Semantic name of Class	FIRM [3]	SIMPLIcity [4]	Edge Based [8]	Histogram Based [9]	Horng Lin Method [13]	Present Method
Africa	47%	48%	45%	30%	48%	**55%**
Beaches	35%	32%	35%	30%	44%	40%
Building	35%	35%	35%	25%	36%	**40%**
Bus	60%	36%	60%	26%	69%	68%
Dinosaur	95%	95%	95%	90%	96%	**99%**
Elephant	25%	38%	25%	36%	55%	**62%**
Flower	65%	42%	65%	40%	89%	79%
Horses	65%	72%	65%	38%	70%	**77%**
Mountain	30%	35%	30%	25%	42%	**45%**
Food	48%	38%	48%	20%	53%	**55%**
Mean	50.5%	47.1%	50.3%	36%	60.2%	**61.7%**

Performance of the proposed system on image set2

In this part, the performances of system are evaluated considering image set 2. As mentioned, there are two sets of images, SetD and SetQ where each one contains 1051 full colour images. The images in SetD are employed as the database images and those in SetQ are used as the query image, where each image pair (I_i^d, I_i^q) is randomly chosen from the same animation. In each experiment, each I_i^q is used as the query image. For each query image, the system responds to the user, L database images with the shortest image, matching distances to I_i^q. If I_i^d exists among the L database images, we say that the system has correctly found the expected image. Otherwise, the system has failed to find the expected image. Accuracy rate of system is explained with accuracy of retrieved images (ACC, %) [13]. Experiment is to compare the retrieval accuracy of our system with [13-15]. Table 4 shows that the proposed method is much superior compared to the methods of [3] [8][13]. Retrieval precision increases with the increasing number of returned images.

Computational Cost

Table 5 presents that, proposed method has the same number of multiplication and comparisons as that with scheme [13]. However the number of additions is reduced by about 1.4. We also measure CPU times of the two methods which were executed on 1.7 GHz Corei12 CPU and 4 GB primary memory. Experimental results show that the proposed retrieval gives 38ms and that with [13] 44ms for Image set1. We can see that our approach is 1.2 times faster than [13]. However the CPU time of the retrieval approach is expected to be much shorter than that of [13] retrieval approach for huge image sets. It is interesting to know that in our approach main feature vector dimension is the least one comparing to Gabor, standard wavelets and [13]. We note that increasing dimension of feature vector will produce better results but volume of memory is increased. Table 6 provides CPU times for feature extracting search of image set for given query image and feature vector length for different feature sets. Feature extraction time for query image, in Gabor wavelet is most expensive.

Table 4. Comparison of average precision (%) of the proposed method and other standard retrieval systems [3-4], [8-9], [13] when different number of images are retrieved.

	1	2	3	4	5	10	20	30	50	100
Huang and Dais[14]	65.3	72.2	74.7	77.0	78.1	83.5	86.2	88.4	92.0	94.7
Jhanwar etal.[15]	62.4	70.7	74.8	76.6	79.0	84.0	87.7	90.2	92.3	94.6
Lin[13]	85.5	90.5	92.3	93.2	93.6	95.1	96.7	97.7	98.9	99.2
Present Method	88.3	92	93.9	94.4	94.6	95.7	97	98.1	99	99.3

Table 5. computational complexity per pixel for image set1.

Methods	Addition	Multiplication	Comparison
Ref [13]	47.9	18.6	15.1
Present method	38.2	18.6	15.1

Table 6. Feature Vector length, Feature generation and seeking time for a query image.

	Gabor Wavelets	Standard wavelets	Ref [13]	SID-CWT Moments & MSHP
Feature vector length	42(6 orientation, 7 scale)	40	Variable from 50 to larger lengths	35
Feature Generation time(s)	3.54	0.61	1~3	0.59
Seeking time(s)	0.08	0.08	0.08	0.08

CONCLUSION

Evaluation of novel method for image retrieval in optimized framework is presented. Images are partitioned into non-overlapping tiles. Texture Features drawn from transformed using Shift Invariant dual-tree complex wavelet transform from each of non-overlapping tiles computed energy and standard. An integrated matching scheme between image tiles was implemented for image similarity. Invariant moments were used to describe the shape features. A combination of these texture and shape features provides a robust feature set for image retrieval. It is interesting to know that in our approach dimension of feature vector is the least one comparing to other approaches. We note that increasing dimension of feature vector will produce better results but volume of memory is increased. Further, the computational time of the proposed method is less than the previously presented methods which is an advantage in CBIR systems. The experiments using 2 image sets demonstrate the efficiency of the proposed method in comparison with the existing methods in literatures, that is, the new method achieves more retrieval accuracy.

REFERENCES

[1] Amirani, M., Sadeghi Gol, Z., Beheshti, A., "Efficient Feature Extraction for Shape-based Image Retrieval," Journal of Applied Science, pp. 2378-2386, 2008.

[2] Sedghi, T. Composition of texture and shape features for content based image retrieval, M. S. Dissertation, Urmia University, Iran, 2010.

[3] Y. Chen and J. Z. Wang, "A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval," in *IEEE Trans. on PAMI*, vol. 24, no.9, pp. 1252-1267, 2002.

[4] J. Li, J.Z. Wang, and G. Wiederhold, "IRM: Integrated region matching for image retrieval," in *Proc. of the 8th ACM Int. Conf. on Multimedia*, pp. 147-156, Oct. 2000.

[5] Z. Rui-zhe, Y. Jia-zheng, H. Jing-hua, W. Yu-jian; B. Hong "A Novel Generalized SVM Algorithm with Application to Region-Based Image Retrieval ," in, International Forum on Information Technology and Applications, pp. 280 - 283, 2009.

[6] Chenyang Xu, and J. L. Prince, "Snakes, shapes, and gradient vector flow", *IEEE Transactions on Image Processing*, vol. 7, no 3, pp. 359-369, Mar. 2005.

[7]N.Kingsbury, Complex wavelets for shift invariant analysis and filtering of signals,J.Appl.Comput. HarmonicAnal.10(3)(2001) 234–253.

[8] Hiremath P. S, Jagadeesh Pujari, "Content Based Image Retrieval based on Texture and Shape features using Image and its complement ", IJCSS, Volume2, No. 1, 2008.

[9] Y. Rubner, L. J. Guibas, and C. Tomasi, "The earth mover's distance, multi-dimensional scaling, and color-based image retrieval", in *Proceedings of DARPA Image understanding Workshop*, pp. 661-668, 1997.

[10] D. Hoiem, R. Sukhtankar, H. Schneiderman, and L.Huston, "Object-based image retrieval using statistical structure of images", *In Proc CVPR*, 0,

[12] D. Zhang, Guojun Lu, "Review of shape representation and description techniques", *Pattern Recognition*, vol. 37, pp. 1-19, 2004.

[13] C. Lin, R. Chen, and Y. Chan, "A smart content-based image retrieval system based on color and texture feature", *Image and Vision Computing*, vol. 27, pp. 658–665, 2009.

[14] P.W. Huang and S.K. Dai, "Image retrieval by texture similarity", *Pattern Recognition*, vol. 36, no. 3, pp. 665–679, 2003.

[15] N. Jhanwar, S. Chaudhurib, G. Seetharamanc, and B. Zavidovique, "Content based image retrieval using motif co-occurrence matrix", *Image and Vision Computing*, vol. 22, pp. 1211–1220, 2004.