CBIR SYSTEM BASED on INTEGRATION BETWEEN SURF AND GLOBAL FEATURES

M. H. Saad, H. I. Saleh, H. Konber, M. Ashour
Eng. Dept., NCRRT, EAEA, 3 Ahmed Al-Zomar, 8th District, Nasr City, Cairo,
E-mail: m.hassansaad@gmail.com

Abstract

Content-Based Image Retrieval (CBIR) is a challenging task which retrieves the similar images from the large database according to objective visual contents of the image itself. The content-based image retrieval has many application areas such as, education, commerce, military, searching, medical image and web image classification. This paper proposes a new image retrieval system, which uses color and texture information to form the feature vectors and Bhattacharyya distance and new similarity measure to perform the feature matching. The proposed CBIR system uses integration of the yc_bc_r color histogram, edge histogram and shape descriptor as global descriptor with surf salient point as local descriptor to enhance the retrieval results. The proposed technique is proper for precisely retrieving images even in deformation cases such as geometric deformations and noise. It is tested on a standard image databases such as Wang and UCID databases. Experimental work shows that the proposed approach improves the precision and recall of retrieval results compared to other approaches reported in recent literature.

Keywords- CBIR, image retireval, EHD, FCTH, MPEG7, YC_bC_r

1. Introduction

Content-Based Image Retrieval (CBIR) is a challenging task which retrieves the similar images from the large database. Most of the CBIR system uses the global features such as color, texture and shape to extract the features from the images. In Recent years the local feature (Interest points) are used to extract the most similar images with different view points and different transformations. Local feature describe pixels in an image through their local neighborhood content. They should be distinctive and at the same time robust to changes in viewing conditions.

SURF is an image local feature extracting and describing method. It finds and describes point correspondences between images with different viewing conditions. Unlike those global feature descriptors, such as MPEG-7 scalable color for color and MPEG-7 edge histogram for texture, which use a single global feature vector to represent an entire image. The local descriptors, such as SURF and SIFT, search for distinctive locations, that is, the so called interest points, in an image and then generate vectors to represent the interest points.

It is common to find hundreds or thousands of interest points in a medium sized image and the feature vectors generated are, therefore, of hundreds or thousands. Because of this ability to represent local features, local descriptors, such as SURF and SIFT, are able to find correspondences between two images in the cases of scaling, rotation and view point changing, whereas, in such circumstances, global descriptors will generally fail. Hence, those local descriptors have already found their applications in the areas of object recognition, 3D reconstruction, sub-image detection and content-based image retrieval [1-4]. It's more appealing than the earlier Scale-Invariant Feature Transform (SIFT) in terms of robustness and performance. This paper presents a scale- and rotation-invariant interest point detector called SURF (Speeded Up Robust Features).

Speeded-Up Robust Features is a fast and robust algorithm for local, similarity invariant image representation and comparison. SURF selects interest points of an image from the salient features of its linear scale-space, and then builds local features based on the image gradient distribution. The main interest of the SURF approach lies in its fast computation of approximate differential operators in the scale-space, based on Integral Image Representation and Box Filters, enabling real-time applications such as tracking and object recognition.

We provide a detailed analysis of the SURF descriptors for CBIR, and explore whether rotation invariant descriptors are helpful for image retrieval. The proposed approach allows matching images under partial occlusions, geometric deformation and even if they are not perfectly aligned or illuminated. To improve the performance of the system the SURF is combined with the global features (color, texture and shape) since SURF works only on gray scale images. Experimental results on the Wang and UCID database show that the proposed approach is robust and can outperform current generic approaches. Section 2 gives brief description about the proposed global feature. These descriptors are yc_bc_r color histogram, EHD and FD descriptors. Section 3 explains the SURF algorithm; moreover, it shows the effect of geometric deformation in the results. Section 4 illustrates the proposed CBIR system which integrates between the global and local features. Finally the methods of evaluating our system and proposed approach are overviewed.

2. Global Features

Color, texture and shape information have been the primitive image descriptors in content based image retrieval systems. The previous chapter presents a framework for combining all the three i.e. color, texture and shape information, and achieve higher retrieval efficiency. Global features related to color or texture or shape are usually used to provide a low level description of image content. The global features cannot capture all parts of the image having different characteristics. Therefore, local feature of image information is necessary. The combination of the color, texture and shape features with salient points provide a robust feature set for image retrieval.

2.1 Local Features

Local descriptors depict a pixel in an image through its local neighborhood content which should be distinctive and robust to changes in viewing environment or deformations or localization errors. Local Features find the corresponding pixel locations in images which compute the same amount of information about the spatial intensity patterns under different conditions [5-10].

Local descriptors have been extensively used in CBIR systems, where their robustness to geometric deformation and transformations allows the identification and recognition of a target object with great reliability. However their application to the retrieval of complex categories is challenging due to excessive computation time. In the retrieval and classification of images, the images may be described either by a single descriptor or a set of descriptors. In the former case, when a single descriptor computes the entire information of the image, it is a global descriptor. In the latter case, the descriptors are associated to different features of the image (regions, edges or small patches around points of interest) and are called local descriptors. Local descriptors have been initially proposed to solve problems in computer vision, from point matching in stereovision, to object detection.

3. Speeded up Robust Features (SURF)

The Speeded up robust features algorithm is a scale and rotation-invariant interest point detector and descriptor which are computationally very fast. The detector locates the interest points in the image, and the descriptor depicts the features of the interest points and constructs a distribution of Haar-wavelet responses within the interest point neighborhood as feature vectors of the interest points .

The performance of SURF is increased by using the integral image. This integral image is computed rapidly from an input image which led to speed up the calculation of the interest points. The major computational steps of SURF algorithm is as follows:

- Interest point detection,
- Interest point description,
- Feature matching.

3.1. Interest point detection

The use of integral images reduces the computation time extremely. Due to its good performance in accuracy, Hessian matrix approximation is used in the interest point detection.This paper briefly explains the concept of integral images. They allow for fast computation of box type convolution filters. The entry of an integral image $U\sum(X)$ at a location $X = (x, y)^T$ represents the sum of all pixels in the input image U within a rectangular region formed by the origin and X as shown in figure (1).

$$U\sum(X) = \sum_{i=0}^{i \le x} \sum_{j=0}^{j \le y} u(i, j) \tag{1}$$

After the integral image computation, three additions are performed to calculate the sum of the intensities over any upright rectangular area. Hence, the calculation time is independent of its size.

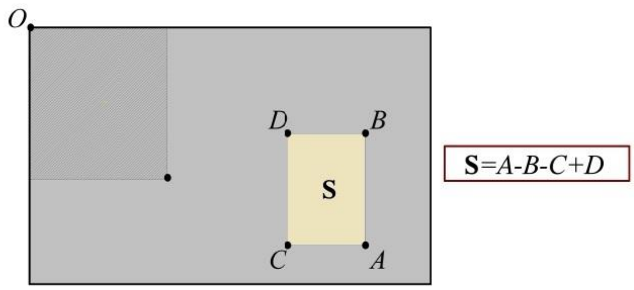


Figure 4 integral images

Given a point X =(x, y) in an image U, the Hessian matrix $H(X,\sigma)$ in X at scale σ is defined as follows

$$H(X,\sigma)=\begin{bmatrix} L_{xx}(X,\sigma) & L_{xy}(X,\sigma) \\ L_{xy}(X,\sigma) & L_{yy}(X,\sigma) \end{bmatrix} \tag{2}$$

Where, $L_{xx}(X,\sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$ with the image U in point x, and similarly for $L_{xy}(X,\sigma)$ and $L_{yy}(X,\sigma)$. Gaussians are optimal for scale-space analysis. However, they have to be discretizied and cropped in practical implementations which causes a repeatability loss in image under rotations around odd multiples of

$\pi/2$. In general, this drawback maintains for Hessian-based detectors. The approximation for the Hessian matrix with box filters is used as shown in figure (4.2). These approximate second order Gaussian derivatives (D_{xx}, D_{yy} , and D_{xy}) evaluated at a very low computational cost using integral images. Therefore, the computation time is independent of the filter size. The weights applied to the rectangular regions are kept straightforward for computational efficiency. This yield

$$\det\left(H_{approx}\right)=D_{xx}D_{yy}-(wD_{xy})^2 \tag{3}$$

Where, w ≈ 0.9. The relative weight w of the filter responses is used to balance the expression for the Hessian's determinant. Figure (4) shows an example of the detected interest points using the Hessian detector.

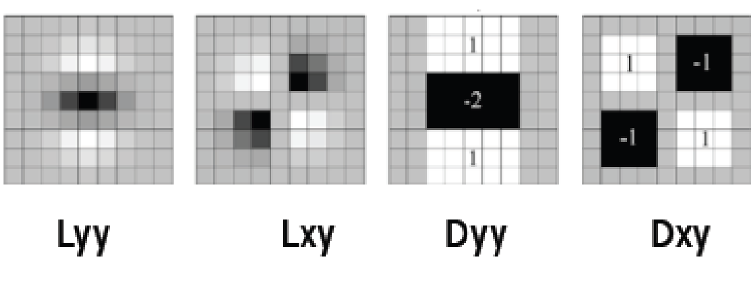
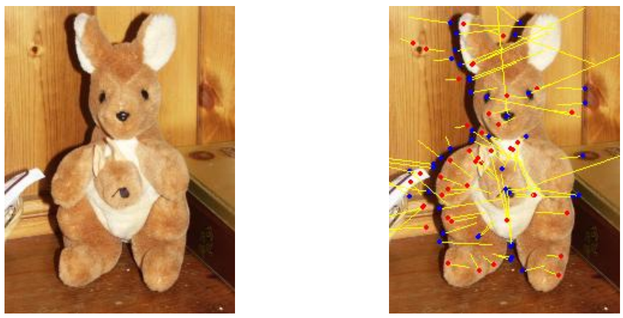


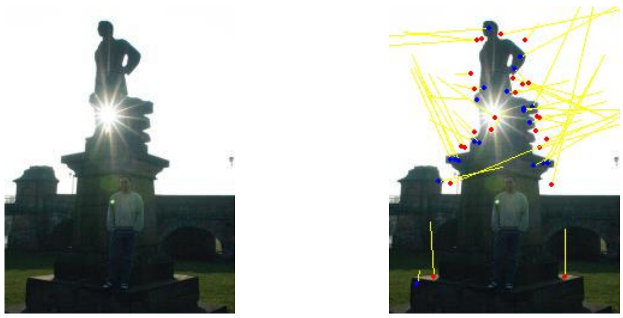
Figure 2. Hessian based key point detector

(a) horeses image

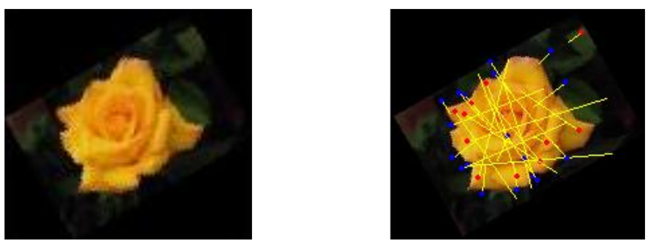
(b) bluiding image



(c) toyes image



(d) image having change of illumination



(f) image having geometric deformation

Figure 3. SURF Interest points for different images

3.2. Interest Point Description

The SURF descriptor depicts the distribution of the intensity content within the interest point neighborhood. The distribution of first order Haar wavelet responses in x and y directions is used taking advantage of integral images for speed. This reduces the time for feature computation and matching.

In order to be invariant to image rotation, a reproducible orientation of the interest points has been identified. For that purpose, the Haar wavelet responses are calculated in x and y directions within a circular neighborhood of radius 6s around the interest point, with s the scale at which the interest point was detected. The used filters are shown in Figure (4) only six operations are needed to compute the response in x or y direction at any scale.

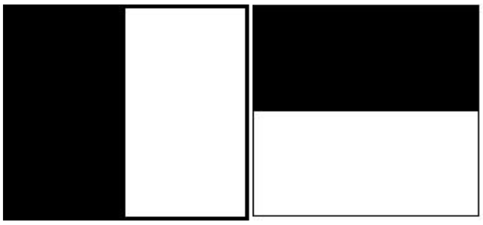


Figure 4. Haar wavelet filters to compute the responses in x (left) and y direction (right).

For the extraction of the descriptor, the first step consists of constructing a square region centered around the interest point and oriented along the orientation selected.

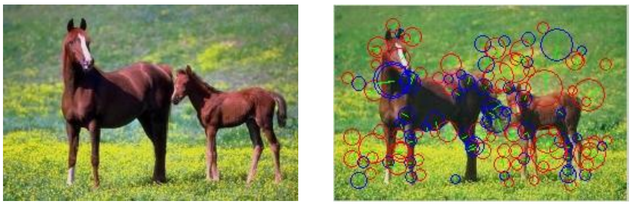
To preserve important spatial information, the region is split up regularly into smaller 4 x 4 square sub-regions. For each sub-region, the Haar wavelet responses are computed at 5x5 regularly spaced sample points. The Haar wavelet responses in horizontal and vertical directions are called d_x and d_y, respectively. To increase the robustness against geometric deformations and localization errors, the responses d_x and d_y are first weighted with a Gaussian standard deviated by σ =3.3s centered at the interest point.

Then, the summation of the wavelet responses d_x and d_y over each sub-region forms a first set of entries in the feature vector. To gain information about the polarity of the intensity changes, the absolute values of the responses is summed up. Hence, each sub-region has a 4D descriptor vector V for its underlying intensity structure and is defined as follows

$$V = \left(\sum d_x, \sum dy, \sum |d_x|, \sum |d_y| \right) \quad (4)$$

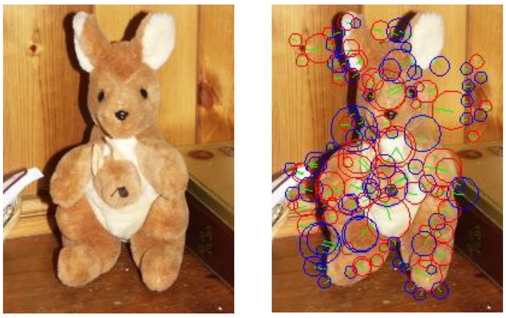
Concatenating each V for all 4 x 4 sub regions forms a descriptor vector of length 64. The wavelet responses are illumination-invariant. Contrast-invariant is achieved by turning the descriptor into a unit vector.

The properties of the descriptor for three distinctively different image-intensity patterns within a sub-region are shown in figure (5). Figure (6) demonstrates the SURF interest point descriptors for horses, building, toys, illuminated and deformed images.

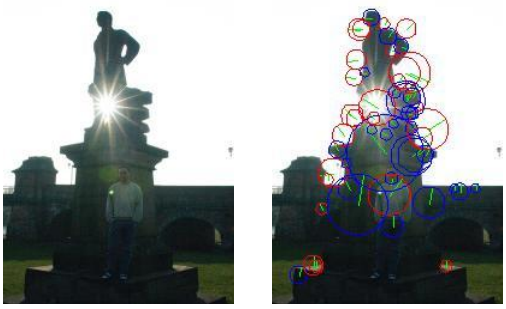


(a) horeses image

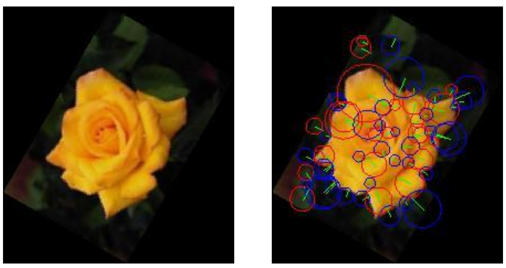
(b) building images



(c) toys image



(d) images having change of illumination



(e) image having geometric deformation

Figure 6 SURF interest points descriptor for different images

3.3. SURF Descriptors Matching

Matching the SURF descriptors of both images is done by a nearest neighbor (NN) algorithm which classifies objects based on closest training examples in the feature space. The NN algorithm is among the simplest machine learning algorithms. The NN works as follows: first, training process creates a database of the objects, for which we already know what the correct classification should be. Then, when the system is given a query, i.e., a new object to classify, the classification process simply finds the nearest neighbor of the query in the database, i.e., the database object that is the most similar to the query. Then, the system classifies the query as belonging to the same class as its nearest neighbor. Figures (7-10) illustrates the matching between pairs of original images and their deformed ones.

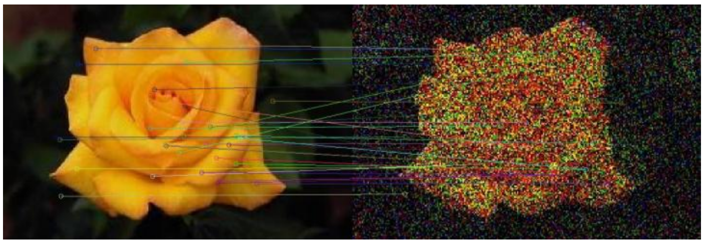


Figure 7. Matching between original image and noisy image

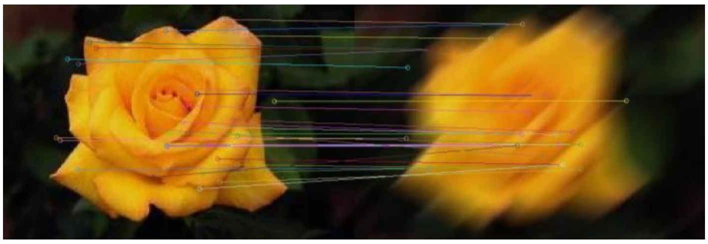


Figure 8. Matching between original image and blurred image

Figure 9. Matching between original image and cropped image

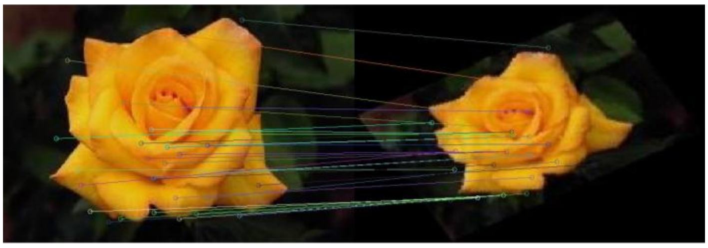


Figure 10. Matching between original image and scaled and rotated image

4. The proposed Image Retrieval System

The proposed approach combines the proposed CTSD descriptor which is applied globally on the whole image with the SURF descriptor which locally extracts salient points. The final score between two images is given by

$$finalScore = w_1 \times global\ score + w_2 \times local\ score \quad (5)$$

Where, w_1 and w_2 are the weights for global and local features. This score increases with the increase of similarity to the query image. In our experiments, w_1 and w_2 are equal to one each; giving an equal importance to the description of an image.

5. Performance Evaluation

Precision and Recall are metrics to calculate the ranking of the images returned by the retrieval system [10]. For a query q having a defined ground truth images over a database R(q), and let Q(q) be the retrieved result of images for that query. The precision of the retrieval is defined as the fraction of the retrieved images that are indeed relevant to the query [11].

$$\Pr ecision = \frac{|Q(q) \cap R(q)|}{|Q(q)|} \quad (6)$$

The recall is the fraction of the relevant images that is returned by the query [11].

$$\operatorname{Re}call = \frac{|Q(q) \cap R(q)|}{|R(q)|} \quad (7)$$

To evaluate a system over all the categories, Top N performance measurements [12] can be used. When submitting a query q to a CBIR system, the system returns N resulted list of images sorted based on similarity to the query image, where N is the number of top similar images. We denote $PR_N(q_i)$ as the precision of the top N returned sorted results. The aim of the user after submitting a query is to search for the most relevant images R(q). The precision
PR_j, j=1, 2, … of the top N results of a query q is defined as:

$$PR_N(q_i) = \sum_{i=1}^{N} \frac{\psi(p_k, R(q))}{N} \quad , \psi(x,y) = \begin{cases} 1; if\ x \in y \\ 0; if\ x \notin y \end{cases} \quad (8)$$

So the average precision for all queries performed on a CBIR system for a certain N number of returned results is defined as:

$$PR_N = \sum_{i=1}^{Total\_Query\_count} \frac{PR_N(q_i)}{Total\_Query\_count} \quad (9)$$

Similarly, the recall RE_j, j=1, 2, … N of the top N results of a query q is defined as:

$$RE_N(qi) = \sum_{i=1}^{N} \frac{\psi(p_k, R(q))}{\|R(q)\|} \quad (10)$$

And the average Recall for all queries is defined as:

$$RE_N = \sum_{i=1}^{N} \frac{RE_N(qi)}{\|R(q)\|} \quad (11)$$

6. Experiments results

The experiments were executed on the 1000 image Wang database [13] and the Uncompressed Color Image Database (UCID) [14]. The Wang database includes 10 categories; each category contains 100 images. For each query, a set of ground truth images that are relevant to the query were identified. Wang database have 20 queries each with a proposed ground truth. The UCID consists of 1338 uncompressed TIFF images of different topics related to indoors, outdoors and natural scenes, and man-made objects. UCID database have 162 queries each with a proposed ground truth. The results of the proposed approach have been compared with the results of the following MPEG7 descriptors: Scalable Color Descriptor (SCD), Color Layout Descriptor (CLD), and Texture Descriptor: Tamura Descriptor. Moreover, the results have been compared to HSV histogram. Also the results have been compared to the fuzzy color and texture histogram (FCTH) composite descriptor.
The evaluation of the top N precision results for the proposed approach compared to other approaches over the Wang database is shown in figure (10). The evaluation of the top N recall results for the proposed approach compared to other approaches over the Wang database as shown in figure (11). Figure (12) shows the

experimental results of the mean precision for each category in the Wang database. Figure (13) shows the experimental results of the mean recall for each category in the Wang database. It is shown that proposed technique is better accurate than the previous approaches for categories.

Figure (14-15) shows the evaluation of the top N recall results and the top N precision evaluated over the UCID. Consistent with the results the proposed technique is better accurate than the previous approaches. The proposed technique is compared with the HSV histogram, FCTH and the MPEG7 descriptors. The results prove that the proposed technique is better than the previous approaches

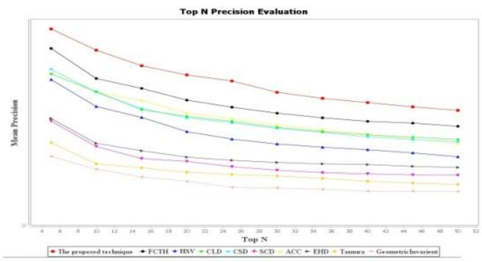


Fig.10. the evaluation of the top N precision results and the top N precision over Wang database

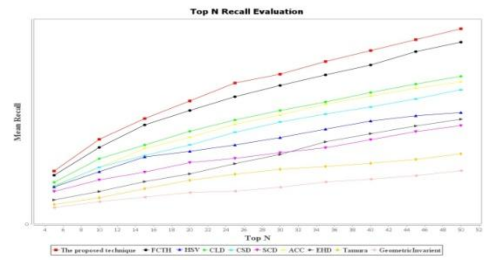


Fig.11. the evaluation of the top N recall results and the top N precision over Wang database

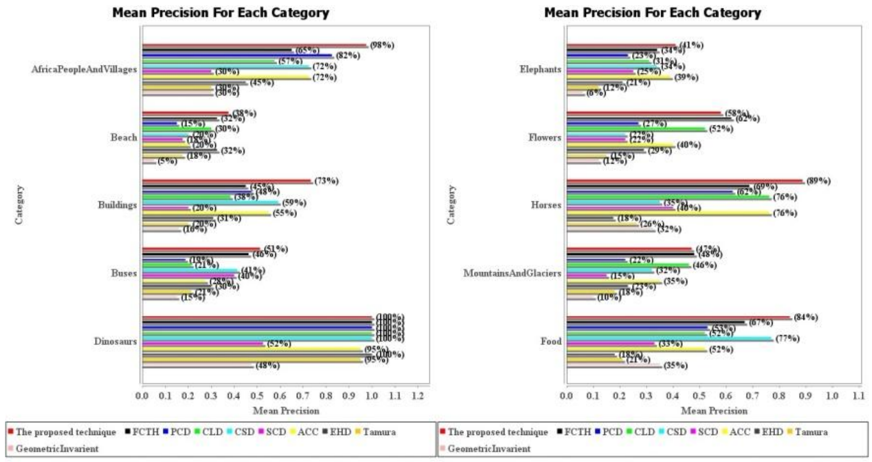


Fig.12. the mean precision for each category in the Wang database

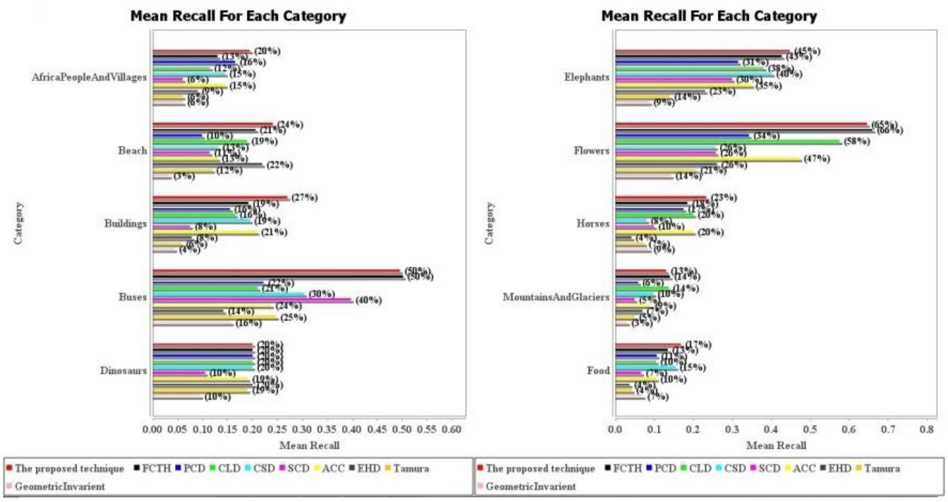


Fig.13. the mean recall for each category in the Wang database

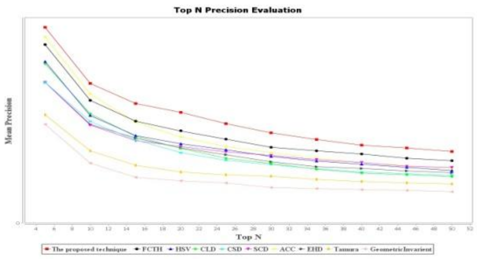


Fig.14. the evaluation of the top N precision results and the top N precision over UCID database

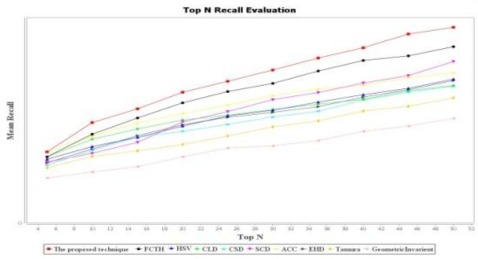


Fig.15. the evaluation of the top N recall results and the top N precision over UCID database

7. Conclusions

In this paper, we presented a new approach for image retrieval that integrates works in SURF (local feature), color, texture and shape features (global features). The new approach extracts image salient points to be integrated with the proposed global features. Agreedy graph matching algorithm worked as a similarity measure to detect the final image rank. The integration between global and local features has been explored by combining our approach with other three global feature techniques. The experimental results showed that this combination provide more accurate results than other compared techniques.

By extracting the most complex regions in images, the proposed approach is accurate in retrieving images even in the presence of geometric deformations or large occlusion or noise. The proposed integrated CBIR has improved the retrieval results.

8. Reference

[1] M. H. Saad, H. I. Saleh, H. Konbor, M. Ashour, " Image Retrieval based on Integration between YCbCr Color Histogram and Texture Feature ", published at international journal of computer theory and engineering (IJCTE) in Vol. 3, No. 5, 2011.

[2] M. H. Saad, H. I. Saleh, H. Konbor, M. Ashour, "Image Retrieval based on Integration between YCbCr Color Histogram and Shape Feature ", published at International Computer Engineering Conference ICENCO 2011.

[3] M. H. Saad, H. I. Saleh, H. Konbor, M. Ashour, "Image Retrieval based on Integration between Color and Geometrical Moment Features ", published at 10th international conference of nuclear sciences and applications, ESNSA 2012.

[4] Mohamed Ahmed Ebrahim Helala, Mazen Mohamed Selim, Hala Helmy Zayed, "An Image Retrieval Approach Based on Composite Features and Graph Matching," iccee, vol. 1, pp.466-473, 2009 Second International Conference on Computer and Electrical Engineering, 2009.

[5] Shamik Sural, Gang Qian, and Sakti Pramanik. Segmentation and histogram generation using the hsv color space for image retrieval. Proceedings of the 6th ACM international conference on Image and video retrieval, Amsterdam, The Netherlands, 2:II–589– II–592, 2002.

[6] Barcellos, R; Oliani , R. S.; Lorenzi, L. T.; Gonzaga, A. "Content Based Image Retrieval Using Color Autocorrelograms in HSV Color Space", In Proceedings Of Xviii Brazilian Symposium on Computer Graphics and Image Processing - Sibgrapi 2005, ISBN: 85-7669-036-5.

[7] Ji-quan Ma, "Content-Based Image Retrieval with HSV Color Space and Texture Features," wism, pp.61-63, 2009 International Conference on Web Information Systems and Mining, 2009

[8] Ka-Man Wong, Kwok-Wai Cheung and Lai-Man Po, "MIRROR: an interactive content based image retrieval system", Proceedings of IEEE International Symposium on Circuit and Systems, vol. 2, Japan, May 2005, pp. 1541-1544, 23-26.

[9] Jose M. Martinez, Mpeg-7 overview (http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm).

[10] Qasim Iqbal and J.K. Aggarwal. Combining structure, color and texture for image retrieval: A performance evaluation. 16th International Conference on Pattern Recognition (ICPR), 2:438–443, 2002.

[11] S. A. Chatzichristo_s and Y. S. Boutalis, "FCTH - Fuzzy Color and Texture Histogram - A low level Feature for Accurate Image Retrieval", 9th International Workshop on Image Analysis for Multimedia Interactive Services, IEEE Computer Society, Klagenfurt, Austria, 2008, pp. 191-196.
[12] Keinosuke Fukunaga. Introduction to statistical pattern recognition (2nd ed.) Academic Press., 1990.
[13] Sylvie Philipp-Foliguet, Julien Gony, and Philippe-Henri Gosselin. Frebir: An image retrieval system based on fuzzy region matching. Comput. Vis. Image Underst., 113(6):693–707, 2009.
[14] Zhang H. LongF. and Dagan Feng D. Fundamentals of content-based image retrieval,in multimedia information retrieval and management – technological fundamentals and applications. Springer-Verlag, pages 1–26, 2003.
[15] J. Stottinger, N. Sebe, T. Gevers, and A. Hanbury. Colour interest points for image retrieval. Vision Winter Workshop, 2007.
[16] James Z. Wang, Jia Li and Gio Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrate, Matching for Picture Libraries", IEEE transactions on pattern analysis and machine intelligence, Volume 23, no. 9, September 2001.
[17] G. Schaefer and M. Stich, "UCID -An Uncompressed Colour Image Database", Proc. SPIE, Storage and Retrieval Methods and Applications for Multimedia.