Comparative study on Content-Based Image Retrieval (CBIR)

Sumaira Muhammad Hayat Khan, Dr.Ayyaz Hussain
sumairafd@yahoo.com, ayyaz.hussain@iiu.edu.pk
Department of Computer Science.
International Islamic University Islamabad
Islamabad, Pakistan

Dr.Imad Fakhri Taha Alshaikhli
imadf@iium.edu.my
Department of Computer Science
International Islamic University Malaysia
Selangor, Malaysia

***Abstract*- The process of retrieving desired images from a large collection is widely used in applications of computer vision. In order to improve the retrieval performance an efficient and accurate system is required. Retrieving images based on the content i.e. color, texture, shape etc is called content based image retrieval (CBIR). The content is actually the feature of an image and is extracted through a meaningful way to construct a feature vector. Images having the least distance between their feature vectors are most similar. This paper gives comparison of three different approaches of CBIR based on image features and similarity measures taken for finding the similarity between two images. Results have shown that selecting an important image feature and calculating that through a meaningful way is of great importance in image retrieval. All the important features must be considered while constructing a feature vector and a proper similarity measure should be used for calculating the distance between two feature vectors. These parameters play very crucial role in deciding the overall performance of the any CBIR system. Some future direction were identified and under our future work.**

Keywords- image retrieval; CBIR; feature vector; feature extraction; similarity measures.

I. INTRODUCTION

It is a need to create large data sets because of the advances in data storage and image acquisition technologies. The need for keeping these databases is growing because of the increasing amount of digitally produced images in areas like medicine, journalism, and private life. An efficient way is required to manage these databases [18]. The process of retrieving desired images from a large collection of images (image database) is of great importance in computer vision. Its application in almost every field is increasing for the ease and convenience of users. The process of image retrieval is based on the features that can be automatically extracted from the images themselves. These systems are thus named as CBIR (Content Based Image Retrieval) systems and have received great attention in the literature of image information retrieval. QBIC system from IBM [22] is one of the image retrieval systems that were available in the beginning. Photobook system from Massachusetts Institute of Technology (MIT) [23] is also one of such systems. Another image retrieval system is Blobworld [24] developed at UC Berkley. SIMBA, CIRES, SIMPLIcity, Image Retrieval in Medical Applications (IRMA) and FIRE also use the same approach for image retrieval. Image retrieval system that follows the discrete approach is VIPER/GIFT [25]. This system describes the local and global properties of the image by incorporating color and texture features [18].

CBIR is of great importance in the medical field. Doctors have to access large amounts of images every day. These images being kept in large databases require a quick and efficient method to deal with them [19]. People have developed their own databases where they have kept thousands of images [20]. Journalism is another field of life in which image saving and retrieving is of great importance. Journalists need to search images by new techniques that give more efficient and accurate results [21]. There are also certain websites that share photos. It gives ease to the user to store their images and others to access and view the available images. Flickr and Google Picassa are such sites where photo sharing can be done. Different methods have been proposed to calculate these features in different ways.

Three research papers are selected here for discussion and comparison. Section II describes the theoretical background of CBIR. Three different approaches of CBIR are presented in section III. Section IV gives a comparison of these approaches and section V concludes the discussion and presents suggested future work.

II. THEORETICAL BACKGROUND

In CBIR, feature vector is generated for each image describing the content of image and is stored as an index in feature database. Hence images are automatically indexed by this process. For retrieval of image from a database, similarity of the feature vector of the query and database image is measured. A threshold is specified and the distance is measured between the query image and the database image. This distance must be less than the specified threshold [1].

In CBIR, there are two main methods to define similarity between query image and any target image in the image database. This similarity comparison is either performed globally or locally. Global feature descriptor take the whole image to find its different features like color, texture, shape, Gabor filter features etc. But global features have some

978-0-7695-4959-0/13 $25.00 © 2013 IEEE
DOI 10.1109/ACSAT.2012.40

drawbacks associated with them. First is that they lack important information about spatial feature distribution. Second is the sensitivity to intensity variations and distortion. Spatial limitations have been overcome by color coherence vector [11], color correlogram [12], spatial color histogram [13] and spatial chromatic histogram [14].Local approaches have also been proposed to extend the capability of CBIR. Region based image retrieval methods have been widely used among various local feature based approaches. In region based image retrieval each image is divided into smaller regions and features of region are extracted. The similarity of two images is then calculated based on the corresponding region based features. Segmentation based color, texture, shape and spatial location are further utilized to search and retrieve similar regions from the database [15]. Segmentation algorithm plays a vital role in feature extraction at local level. Segmentation of an image can be achieved by a number of ways. Each method segments an image based on certain properties [8, 17].Visual features of an image are used to describe image's content in CBIR. These features can either be local or global. Typical global features include shape, color and texture features while in local features extraction is done on local part which can be get from segmentation. Color is one of the most widely used visual features represented through some color model. RGB, HSV and YCbCr are the most commonly used color models. HSV and YCbCr are calculated from RGB by applying different formulas. Color is represented by histogram. Histogram is invariant to rotation, translation and scaling of an object but histograms lack semantic information. Two images with same histograms may have different contents. Color distribution in an image is known through color moments [1]. Color indexing is one of the processes by which images in a database are retrieved on the basis of color content. The system requires several important objectives such as automated extraction of color, efficient indexing and effective retrieval to be satisfied [7, 16]. Texture is another important characteristic used for the classification and recognition of objects. Texture representation methods are basically structural and statistical. Co-occurrence matrix is formed and from this matrix five features are extracted. Feature is constructed by applying the gabor filter and finding out its magnitude by taking mean, standard deviation and skew [1]. Shape is also used to measure similarity between the query image and database images. It is an important and primitive visual feature for image content description. Shape descriptors can be divided into two categories, region-based and contour-based methods. Former uses the whole area. Later one uses only the information present in the contour of an object. Circularity, aspect ratio, discontinuity angle irregularity, length irregularity, complexity, right-angleness, sharpness and directedness are features from the contour image. Graphical rough sketches can be used to describe the important features of image i.e. edge-based features of shape and texture. Effective extraction of edge-based features of shape is difficult. Work has been done in automatic extraction of shift, scale and rotation invariant image features from an object's edge and application of these features to content based image retrieval. Image features of the edge shape of objects are extracted using the relative positions of edge pixels. Local features are calculated in an image on certain interest point. For this purpose we use different types of interest point detectors like affine covariant features, difference of Gaussian etc [3]. Appearance-based feature extraction techniques such as principal component analysis (PCA), linear discriminant analysis (LDA) and independent component analysis (ICA) transform images into a lower dimensional subspace [4]. With the progress of features used in image retrieval a work has been done in combining different features in CBIR. Feature can be fused together by linearly combining different similarities [5].

Image retrieval by using these measures is known as content based image retrieval. In these systems users can query easily [6].

III. DIFFERENT APPROACHES OF CBIR

CBIR is of great importance in many fields of life and therefore several techniques have been proposed time to time for its efficiency and accuracy. These techniques are different from each other depending upon the number and type of image features and the distance measures being used to find the similarity between the query image and one of the images from the database. Three of the techniques are selected here depending upon the type of image feature and the distance measure they use. Two of the techniques uses color, texture and edge features while the third one uses shape features. Hence, all of the main image features are considered during the selection of theses techniques. Also there results are good for consideration. Therefore they are selected for comparison to find out which method gives better results and what the reason for their good or bad result is.

A. A Novel Fusion Approach (NFA)

This approach considers feature extraction at local level but it also incorporates global level features up to some extent. For local feature extraction image is first segmented into meaningful parts and features are extracted from each segment separately. The method can be divided into the following steps.

- Image segmentation- Image is segmented into coherent regions by applying K-Means algorithm.
- Color feature extraction- Color feature f^c of each region is extracted by finding the K-Means cluster centre of this region.
- Texture feature extraction- Texture feature f^t for each region is computed through five steps.

Step 1: Generate a "texture template" image by keeping all the pixels in region *j* intact and setting all the pixels outside region *j* as white.
Step 2: Convert this texture template to the gray scale image.
Step 3: Apply a two-level Haar wavelet transform to the gray-scaled "texture template" image obtained from Step 2.
Step 4: Calculate the average energy in each high frequency band (i.e., low–high, high–low, and high–high band) of level 1 and level 2 wavelet decompositions.
Step 5: Construct a texture feature vector j for region j by concatenating the six average energy values in an appropriately scaled measurement unit.

- Feature fuzzification- Each regional color and texture feature is then fuzzified and fuzzy region matching is done on these fuzzified features.
- Edge histogram calculation- The global and semi-global edge histogram descriptors (EHDs) have been utilized to decrease the impact of segmentation
- Local similarity- local similarity of image regions is calculated by incorporating region position and area difference between two matched regions. Overall region based similarity score is calculated which is supposed to be 1 for two identical regions.
- Global similarity- Manhattan distance measure is used for the finding global and semi global similarity.
- Combined similarity- A combined similarity is calculated by assigning weights to each of them. Regional or local similarity is given higher values of weights and semi global similarity is assigned lowest weight value.

Experimentation is performed on 5000 images from Corel database. Five query images are selected with different semantics. Results are compared with UFM method. A more qualitative evaluation is performed by using 10 distinct image categories and randomly selecting 15 images from each category as query images. Average precision of each category is calculated by evaluating top 20 retrieved results. Different variants of the same method are used and the same data is applied on each of them [10].

TABLE 1 COMPARISON OF AVERAGE RETRIEVAL PRECISION OF EACH CATEGORY BY USING SEVEN DIFFERENT METHODS. REFERENCE (XIAOJUN QI, YUTAO HAN, "A NOVEL FUSION APPROACH TO CONTENT-BASED IMAGE RETRIEVAL", PATTERN RECOGNITION 38, (2005) 2449 – 24)

	Proposed	UFM	IRM	HSV 32 bins	HSV 64 bins	Color Indexing	EHD
Beach	0.2800	0.2533	0.2800	0.1967	0.2567	0.2233	0.1067
Building	0.5467	0.4667	0.4133	0.2667	0.3400	0.4133	0.1367
Vehicle	0.5600	0.3167	0.1367	0.1933	0.2067	0.1633	0.4700
Flower	0.8233	0.6867	0.4933	0.5433	0.5967	0.3833	0.3100
Horse	0.8900	0.7633	0.8033	0.7400	0.8033	0.8367	0.4033
Food	0.5067	0.3000	0.2367	0.2033	0.2567	0.3133	0.1167
Dinosaur	0.7567	0.6100	0.7600	0.7267	0.8033	0.6567	0.6700
People	0.2433	0.2233	0.1433	0.1333	0.1133	0.1467	0.1400
Sunset	0.7100	0.3967	0.4333	0.3533	0.4100	0.3733	0.4500
Office	0.3333	0.2600	0.3367	0.1767	0.2633	0.3067	0.1433
Average	**0.5650**	**0.4277**	**0.4037**	**0.3533**	**0.4050**	**0.3817**	**0.2947**

B. A Universal Model (UM)

The method proposed here considers local features of an image by first dividing the image into different segments and then calculating image features from each of the segment. Feature vectors of these segments are compared by finding Euclidean distance between them.

- Image segmentation- Images are first segmented and only dominant segments are used for feature extraction.
- Color feature extraction- Color feature is calculated by constructing a color histogram for each region which is independent of the image size and orientation.
- Texture feature extraction- Texture feature is extracted by finding the values of energy, entropy, contrast and homogeneity from the co-occurrence matrix.
- Edge histogram calculation- Edge histogram descriptor represents the local edge distribution in the image which is obtained by subdividing the whole image into 4x4 sub images and constructing edge histogram by finding five different edges in each sub image.
- Similarity calculation- Euclidean distance measure is used for finding the similarity between the two images i.e. the query image and the database image.
- Image retrieval- Semantically closer images are retrieved after comparison.

Corel image database is used in experiments. 1000 natural color images, 100 in each category. Precision and recall for four query images are computed using standard formulas [2].

TABLE 2 PRECISION AND RECALL VALUES IN %. REFERENCE (S. NANDAGOPALAN, DR. B. S. ADIGA, AND N. DEEPAK, A UNIVERSAL MODEL FOR CONTENT-BASED IMAGE RETRIEVAL, WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY 46 2008)

Query Image	Color	Texture	EHD	All
1	21.8	50.0	23.6	35.2
	28.0	15.0	34.1	60.0
2	100.0	75.0	87.0	100.0
	98.0	62.0	68.0	78.0
3	74.6	20.0	65.0	42.8
	59.0	10.0	37.0	90.0
4	91.7	75.0	85.6	92.0
	24.0	33.0	34.9	28.0

C. *A Genetic Programming Framework (GPF)*

Feature extraction is performed at global level. Only shape feature is considered in this framework. The concept of composite descriptors is proposed here in which simple descriptors are combined through genetic programming. Each simple descriptor is an image feature having its own similarity measure.

- Shape descriptors- Beam angle statistics (BAS), multiscale fractal dimensions, Fourier descriptors and moment invariants are extracted.
- Composite descriptor- GP is used to combine these simple descriptors to find an optimum composite descriptor
- Similarity measures- BAS is measured through optimum correspondent subsequence (OCS) and other descriptors are measured through Euclidean distance
- These simple descriptors are combined to form composite descriptors.
- Initially, the population starts with individuals created randomly.
- A fitness function is used to assign fitness for each individual.
- Genetic operators are applied to this population aiming to create more diverse and better performing individuals
- Best individuals is determined in the end from final generation

Two different databases have been used in experiments. Fish shape database contains 1000 images created by using 100 fish contours is one of the database used in experiments. Each of the 100 classes has 10 shapes. MPEG-7 is another database of 1400 images used as main part of the core experiments. It has 70 classes of various shapes and each class has 20 images [9].

TABLE 3 AVERAGE PRECISION AFTER 10 IMAGES ARE RETURNED CONSIDERING THE GP-BASED DESCRIPTORS USING BAS40 AND BAS60 DESCRIPTORS. REFERENCE (RICARDO DA S. TORRESA, ET AL. "A GENETIC PROGRAMMING FRAMEWORK FOR CONTENT-BASED IMAGE RETRIEVAL", PATTERN RECOGNITION 42 , (2009) 283-292)

Descriptor	MPEG-7 precision@10		Fish shapes precision@10	
	Sample 1	Sample 2	Sample 1	Sample 2
BAS60	66.27	65.37	93.25	92.30
GP with PAVG@10	73.13 (10.35%)	72.04 (10.20%)	96.12 (3.07%)	93.04 (0.80%)
GP with FFP1	**73.36 (10.69%)**	72.40 (10.75%)	98.03 (5.12%)	96.04 (4.05%)
GP with FFP2	73.30 (10.60%)	72.30 (11.50%)	**98.03 (5.12%)**	**96.04 (4.05%)**
GP with FFP3	73.19 (10.44%)	72.08 (10.26%)	98.03 (5.12%)	96.04 (4.05%)
GP with FFP4	72.96 (9.94%)	71.30 (9.07%)	98.03 (5.12%)	95.95 (3.95%)
GP with CHK	70.08 (10.27%)	69.10 (5.71%)	98.03 (5.12%)	96.00 (4.00%)
GP with LGM	73.2 (10.45%)	**73.37 (12.24%)**	97.03 (4.05%)	95.30 (3.25%)
GA	69.37 (4.68%)	68.30 (4.48%)	93.40 (0.16%)	92.55 (0.27%)

IV. COMPARISON OF THREE DIFFERENT APPROACHES

Three different approaches discussed above are compared on the basis of image features, distance measure and precision of results. Results of these approaches show that local level feature extraction is more important than global level feature extraction. Also each image feature if calculated with a separate distance measure can improve the result.

A. *With Respect to Image Features*

The first two techniques discussed above used the same image features for the construction of feature vectors. These features are color, texture and edge histogram descriptors. However the third technique discussed above used only shape feature for constructing a feature vector for image. For NFA and UM if shape and spatial features are added they will give more discrimination power to their systems. While the GPF should incorporate the color and texture features which are very important features in image retrieval. GPF has been tried on a different dataset however its results are not promising as the precision is not more than 12 %. The method worked on shape feature only and is applied on the dataset which considers only the shape of objects contained in the images.

B. *With Respect to Distance Measure*

NFA calculated the similarity between two regions considering the region area and region position while similarity between EHD's are found through Manhattan distance measure. So it used two different similarity measures. UM has used only one distance measure for all of the features i.e. Euclidean distance measure while GPF has used different distance measure for different types of image features. It has used OCS for BAS and Euclidean distance for multiscale fractal dimensions. Each feature with its own distance measure can give better results. It is found that if the same similarity measure is used for all of the features the results are not very good however different image features calculated

through different similarity measures have better performance [9].

C. With Respect to Precision

From the above results we can see that the first method i.e. NFA has better results as compared to the UM technique as both have used the same image features and the same dataset. For vehicle NFA gave 56% precision while UM gave 35%. Dinosaur 76% and 100%, horse 89% and 42%, flower 82% and 92%. NFA has been tried on 10 different groups of images while UM on just four images. The overall precision of NFA method was found to be 56%. The average precisions for the four images tried on both methods came out to be 76% for NFA and 67% for UM. From the results we can see that the more image is simple i.e. having minimum number of objects the more it will be easy to retrieve. Um method has not been tried on more complex images having great number of objects like beach images, people images, building images etc while NFA has been subjected to all these kinds of images.

V. Conclusion and Future Work

Image features play an important role in retrieving relevant images from the database. NFA incorporated both local and global level features and has fused them for retrieving images. So it contained properties of both local image descriptors as well as global image descriptors. UM method used only local features to construct a feature vector. While the GPF considered global image features only. Local level feature extraction can improve the retrieval accuracy for GPF. Also incorporation of color and texture feature will add efficiency to this system. NFA has given better results compared to the other two methods and can be improved even more if some more image features are added and only important features are selected through any selection algorithm like PCA. Also it can be improved by using automatic fuzzy membership function which can adjust its parameters depending upon the type of image given to it. Segmentation also plays an important role in feature extraction at local level. A good segmentation technique should be tried for promising results. We are working to find a good segmentation technique for better feature extraction at local level. Also we are working on image features to find as many of them as possible and use important features among them by selecting them through PCA. Feature fusion can also improve results to a great extent therefore we are considering different combinations of features to find a fused feature. We also need to find a good combination of image feature with its own similarity measure.

REFERENCES

1. Ryszard S. Chora´s, “Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems,” International Journal Of Biology And Biomedical Engineering, Issue 1, Vol. 1, 2007.
2. S. Nandagopalan et al. “A Universal Model for Content-Based Image Retrieval”, World Academy of Science, Engineering and Technology 46 2008
3. Mohamed Aly, Peter Welinder Mario, and Munich Pietro Perona, “Automatic discovery of image families: global vs. local features”, IEEE, 2009.
4. Vitomir Struc and Nikola Pavesic, “A case study on appearance based feature extraction techniques and their susceptibility to image degradations for the task of face recognition”, IEEE, World Academy of Science, Engineering and Technology 54 2009
5. Zhi Lijia, Zhang Shaomin and Zhao Dazhe, “Learning Based Combining Different Features for Medical Image Retrieval”, Fifth International Conference on Image and Graphics, 2009
6. G.Ohashi and Y.Shimodaira, “Edge-Based Feature Extraction Method And Its Application To Image Retrieval”, semantics, sybernatics and informatics, volume 1, number 5
7. John R.Smith and Shih-Fu Chang, “Tools And Techniques For Color Image Retrieval”, storage and retrieval for image and video databases, 1995.
8. S.Farid, F.Ahmed, “Application of Niblack method on Images”, International Conference on Emerging Technologies, ICET, 2009.
9. Ricardo da S. Torresa et al. “A genetic programming framework for content-based image retrieval”, Pattern Recognition 42 , (2009) 283-292.
10. Xiaojun Qi and ,Yutao Han, “Anovel fusion approach to content-based image retrieval”, Pattern Recognition 38, (2005) 2449 – 2465.
11. G. Pass and R. Zabith, “Histogram refinement for content-based image retrieval”, IEEE Workshop on Applications of Computer Vision, 1996, pp. 96–102.
12. J. Huang et al. “Image indexing using color correlogram”, IEEE International Conference on Computer Vision and Pattern Recognition, Puerto Rico, June 1997, pp. 762–768.
13. A. Rao, R.K. Srihari and Z. Zhang, “Spatial color histograms for content-based image retrieval”, IEEE International Conference on Tools with Artificial Intelligence, 1999, pp. 183–186.
14. L. Cinque et al., “Color-based image retrieval using spatial-chromatic histogram”, Image Vision Comput. 19 (2001) 979–986.
15. W.Y. Ma and B.S. Manjunath, “NeTra: a toolbox for navigating large image databases”, Multimedia Systems 7 (1999) 184–198.

16. T. Wang, Y. Rui and J.G. Sun, “Constraint based region matching for image retrieval”, Int. J. Comput. Vision 25 (1–2) (2004) 37–45.
17. Rafeal C.Gonzalez, Richard E.Woods and Steven L.Eddins: Digital Image Processing Using MATLAB.
18. Thomas Deselaers, Image Retrieval, Object Recognition, and Discriminative Models, Doctoral dissertation, december 2008.
19. Muller, W. Muller, S. Marchand-Maillet, and D. M. Squire, “Strategies for positive and negative relevance feedback in image retrieval”, International Conference on Pattern Recognition, volume 1 of Computer Vision and Image Analysis, pages 1043{1046, Barcelona, Spain, Sept. 2000a.
20. Sun, H. Zhang, L. Zhang, and M. Li. “Myphotos a system for home photo management and processing”, In ACM Multimedia Confernce, Juan-les-Pins, France, Dec. 2002, pages 81-82.
21. H. Armitage and P. G. Enser. “Analysis of user need in image archives”, Journal of Information Science, Apr. 1997, 23(4):287-299.
22. Faloutsos et al. “ Effcient and effective querying by image content”. Journal of Intelligent Information Systems, July 1994, , 3(3/4):231-262.
23. Pentland, R. Picard, and S. Sclaro. “Photobook: Content-based manipulation of image databases”, International Journal of Computer Vision, 1996, 18(3):233-254.
24. Carson et al. “ Blobworld: Image segmentation using expectation maximization and its application to image querying”. IEEE Transaction on Pattern Analysis and Machine Intelligence, Aug. 2002, 24(8):1026-1038.
25. M. Squire et al. “Content-based query of image databases, inspirations from text retrieval: Inverted, frequency-based weights and relevance feedback”, In Scandinavian Conference on Image Analysis, Kangerlussuaq, Greenland, June 1999, pages 143-149.