Journal of Mathematics and Informatics
Vol. 11, 2017, 155-162
ISSN: 2349-0632 (P), 2349-0640 (online)
Published 11 December 2017
www.researchmathsci.org
DOI: http://dx.doi.org/10.22457/jmi.v11a20

Journal of
Mathematics and
Informatics

Content Based Image Retrieval System Using Clustering Technique

***S.Vadivel Murugan*[1] and R. *Chithra*[2]**

Department of Computer Science, PEVR College, Trichy-23, Tamilnadu, India
[1]email: s.vmurugan@yahoo.com; [2]email: chithra_viswanathan@yahoo.co.in

Received 28 November 2017; accepted 7 December 2017

Abstract. Content Based Image Retrieval (CBIR) is a technique for retrieving images on the basis of automatically-derived features such as color, texture and shape. In this proposed work CBIR system presented that uses Wavelet Transform and the color feature as a visual feature to represent the images. Wavelet Transform is proposed as a pre-processing step to make the image enhancement operations. Then the K-means clustering algorithm is used to cluster the images according to their features.

Keywords: K-means algorithm, metadata, clustering technique, query by image, retrieval, similarity.

AMS Mathematics Subject Classification (2010): 91C20, 94A08

1. Introduction

Imaging has played an important role in our life. "Content-based" means that the search analyzes the contents of the image rather than the metadata such as keywords, tags, or descriptions associated with the image. The term "content" in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. CBIR is desirable because searches that rely purely on metadata are dependent on annotation quality and completeness. Humans manually annotate images by entering keywords or metadata in a large database which can be time consuming and may not capture the keywords desired to describe the image. The evaluation of the effectiveness of keyword image search is subjective and has not been well-defined. Image retrieval is a challenging topic that has been a research focus from many years. The retrieval system presents similar images. The user should define what the similarity between images has to be. Color histogram (GCH) is used for representing images by their histograms, and the similarity between two images will be determined by the distance between their color histogram. Furthermore, this approach is sensitive to intensity variations, color distortions, and cropping.

2. Preliminaries

Tayade et al. this paper investigates Search Based Face Annotation (SBFA) framework with the help of mining weakly labeled web facial images which are freely available on World Wide Web (WWW). The drawback as a problem in SBFA is how to perform effectively annotation by considering ordered list of most similar facial images which are

weakly labeled that are often noisy and incomplete[1]. Prasanth et al. the face annotation has many real world applications. The challenging part of search based face annotation task is management of most familiar facial images and their weak labels. To tackle this problem, different techniques are adopted. The efficiency and performance of annotating systems are improved tremendously by using these methods. Here this paper proposes a review on different techniques used for this purpose and check the pros and cons of each technique [2]. Wang et al. proposes an effective Unsupervised Label Refinement (ULR) approach for refining the labels of web facial images using machine learning techniques. To further speed up the proposed scheme, we also propose a clustering-based approximation algorithm which can improve the scalability considerably [3]. Hongmei He et al proposed a framework of a multi-engine system for facial recognition configurable in image types, watch sizes and engines based on performance matrices. The value of each cell in a performance matrix presents a confidence level for facial recognition [4]. Xiao Zhang et al. proposes in this paper, a face annotation system to automatically collect and label celebrity faces from the web. With the proposed system, we have constructed a large-scale dataset called “Celebrities on the Web,” which contains 2.45 million distinct images of 421 436 celebrities and is orders of magnitude larger than previous datasets. Collecting and labeling such a large-scale dataset pose great challenges on current multimedia mining methods [5]. Wang et al we mainly focus on tackling the second challenge of the retrieval-based face annotation paradigm. To improve the annotation performance, a novel Weak Label Regularized Local Coordinate Coding (WLRLCC) algorithm was proposed, which effectively exploits the principles of both local coordinate coding and graph-based weak label regularization [6]. Rajshree et al The Image Retrieval is based on the color Histogram, texture. The perception of the Human System of Image is based on the Human Neurons which hold the 1012 of Information; the Human brain continuously learns with the sensory organs like eye which transmits the Image to the brain which interprets the Image. Thus, in conclusion, the Histogram Intersection-based image retrieval in HSV color space is most desirable among the retrieval methods mentioned in considering both computation time and retrieval effectiveness [7]. Hoi et al, a few studies consider a human name as an input query, and mainly aim to refine the text-based search results by exploiting visual consistency of facial images, which is closely related to automated image re-ranking problems [8]. Ricardo. Torres a et al addressed the problem by presenting a Genetic Programming framework to the design of combined similarity functions. We conclude that the new framework is flexible and powerful for the design of effective combination functions. The effectiveness results demonstrate that the GP framework can find better similarity functions than the ones obtained from the individual descriptors [9]. Xu et al presented a new technique to acquire a direct optimal solution for MFA without resorting to objective function modification as done in many previous algorithms. They conduct comprehensive experiments on the USF Human ID gait database and the Corel image retrieval database[10]. Wang et al. introduced a modification to incorporate the constraint that a face can only appear once in an image. Unlike these studies of filtering the text-based retrieval results, some studies have attempted to directly annotate each facial image with the names extracted [11]. Zhang et al examined the research issues in image mining, current developments in image mining, particularly, image mining frameworks, state-of-the-art techniques and systems. We will also identify some future research directions for

image mining[12]. Smeulders et al proposed a new local density score to represent the importance of each returned image [13].

3. Methodology

The proposed CBIR system consists of Database (DB) creation and Image Retrieval. Figure 3.1 shows the diagram of DB creation and Figure 3.2 shows the flow diagram of DB evaluation.

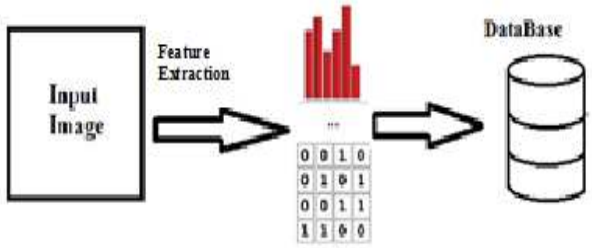

Figure 3.1: DB Creation

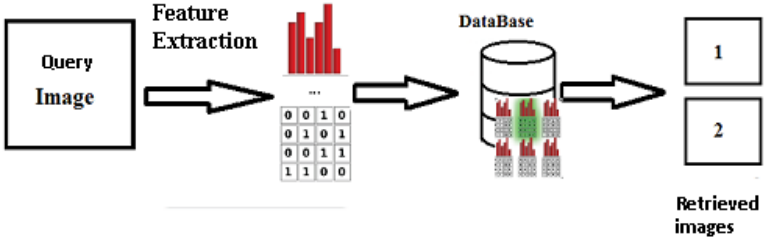

Figure 3.2: DB Evaluation

The various steps used in proposed algorithm works as follows:

Step-1: Collect the images from which one wants to retrieve the images

Step-2: Change the original RGB color space to grayscale;

Step-3: Find Image histogram. An image histogram is a chart that shows the distribution of intensities in an indexed or intensity image

Step-4: Read the images from database one by one and extract the color and shape features from the image. Prepare the feature database that contains the features of all images.

Step-5: Apply k-means clustering algorithm to the feature database.

Step-6: Read the query image and extract the same feature with same method as used for database images.

Step-7: Compute similarity between query image features and clustered database by Euclidean distance and find the cluster closest to query image.

Step-8: Display the required no. of images from the selected cluster.

K-means Clustering Algorithm

The K-means algorithm takes the input parameter, k, and partitions a set of n objects into k clusters so that the resulting intra cluster similarity is high but the inter cluster similarity is low. Cluster similarity is measured in regard to the mean value of the objects in a cluster, which can be viewed as the cluster's centroid. It is fast, robust and easier to understand. It gives best result when data set are distinct or well separated from each other. The algorithm for most popular K-means algorithm as

Algorithm

Given the data set X, choose the number of clusters $1 < c < N$.
Initialize with random cluster centres chosen from the data set.
Repeat for $l = 1, 2, \ldots$

Step 1: Compute the distances

$$D_{ik}^{2} = \left(x_{k} - v_{i}\right)^{T} \left(x_{k} - v_{i}\right), \qquad 1 \le i \le c, \qquad 1 \le k \le N.$$

Step 2: Select the points for a cluster with the minimal distances, they belong to that cluster.

Step 3: Calculate cluster centers

$$v_{i}^{(l)} = \frac{\sum_{j=1}^{N_i} x_{i}}{N_{i}}$$

Until

$$\prod_{k=1}^{n} \max \left| v^{(l)} - v^{(l-1)} \right| \neq 0$$

Ending calculate the partition matrix.

4.Experimental results

The own database of 100 images of ten categories is used to evaluate the performance of CBIR system. Table4.1 shows that the categories of images like Suriya, Sachin, Beach, House, Flower, Food, River, African, Horse and Mountain which are studied in this research work. All images in database are colored images and of size 384 × 256 or 256 × 384 pixels.

Table 4.1: Categories of own image dataset

S. No.	Categories	Images	S. No.	Categories	Images
1.	Suriya		6.	Food	
2.	Sachin		7.	River	
3.	Beach		8.	African	
4.	House		9.	Horse	
5.	Flower		10.	Mountain	

The evaluation of CBIR system is implemented in MATLAB using the above proposed algorithm. The image retrieval system combining clustering algorithm is shown as Figure 3.1. First, extract the image features of each image in image database and apply the clustering algorithm to analysis the similarities of images in the database for constructing the images clustering database. Figure 4.1 shows the output of the proposed CBIR system for the input image using histogram equalization. Figure 4.1 and Figure 4.2 show the retrieval results with and without K-means clustering algorithm. The upper left image is the query image. The right part is the retrieved images. Figure 4.1 displays the retrieval results without K-means clustering algorithm, while Figure 4.2 displays the results with K-means clustering algorithm that images similar to each other and to query.

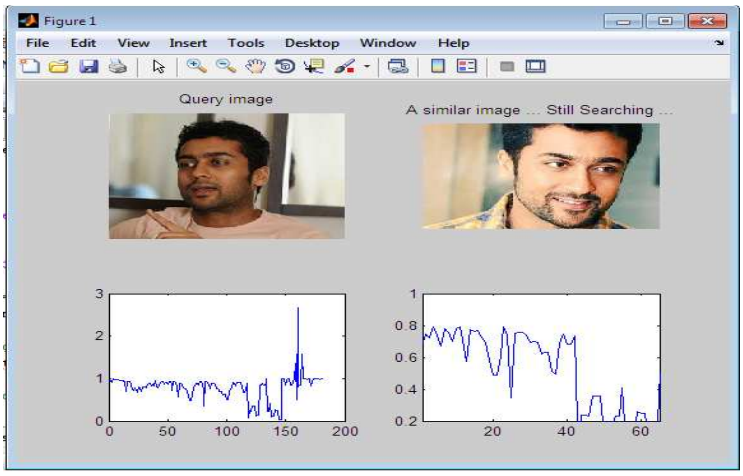

Figure 4.1: Analyzing the input

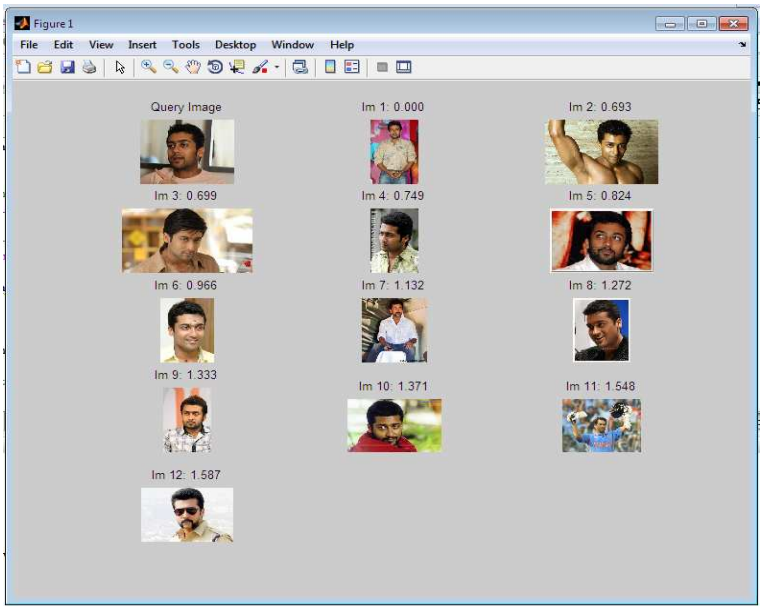

Figure 4.2: Retrieval Results for Suriya Query image

Table 4.2: Accuracy of CBIR with 3-D color image

S.No.	Query Image	Images in Database	Matching Image	Un Matching Image	Accuracy (%)
1.	Suriya	12	11	1	92%
2.	Foods	12	10	2	83%
3.	Houses	12	8	4	67%
4.	Mountain	12	8	4	67%
Average					77%

In the Table 4.2, the accuracy value of the proposed method is 77% which is much greater than the 3-D image histogram based method.

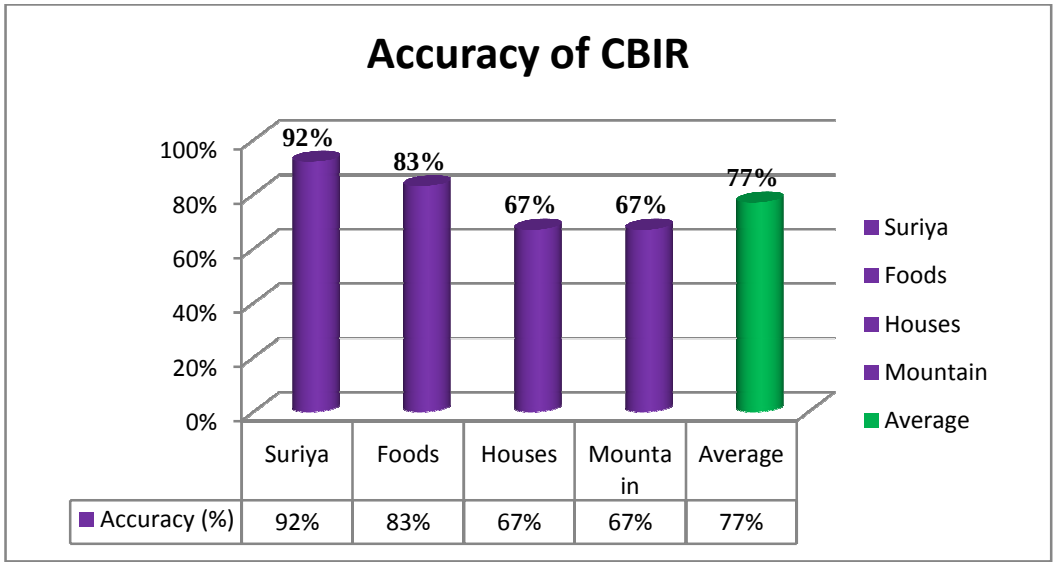

Figure 4.3: Performance Evaluation of CBIR method based on Accuracy

The pictorial representation of Table 4.2 is shown on the Figure 4.3. It shows the efficiency of proposed method based on accuracy.

5. Conclusion

In this work, an attempt has been made to retrieve the similar images from the database based on content based image retrieval method by supplying a query image. Clustering technique is working in a real sense as it is again reducing the time, the color feature is used to form the clusters here K=2 is specified for this current CBIR system. The results are improved by using the K-means clustering concept. The average accuracy value of this method is 77% tested on different images.

Acknowledgement. I would like to thank my professor for supporting me research throughout writing this thesis and my life in general.

REFERENCES

1. Dipak D. Tayade and Dipak Pardhi, Search based face annotation Approach for Mining Weakly Labeled Web facial Images, *International Journal of Modern Trends in Engineering and Research*, 2(2015) 2393-8161.
2. Krishna Prasanth I B and S.Anoop, A Review on Content Based Image Retrieval and Search Based Face Annotation on Weakly Labeled Images, *International Journal of Computer Science and Information Technologies*, 5(6) (2014) 8269-8271.
3. Dayong Wang, Steven C.H.Hoi, Ying He and Jianke Zhu, Mining Weakly Labeled Web Facial Images for Search-Based Face Annotation, *IEEE Transactions On Knowledge And Data Engineering*, 26 (2014) 166-179.
4. Hongmei He and Richard M Guest, A Framework of Configurable Multi-Engine Systems Based on Performance Matrices for Face Recognition, *IEEE International Conference on Technologies for Homeland Security*,12-14 (2013) 1-50.

5. Xiao Zhang, LeiZhang, Xin-JingWang and H.Y.Shum, Finding Celebrities in Billions of Web Images, *IEEE Transactions on Multimedia*, 14 (2012) 995-1007.
6. D. Wang, S.C.H.Hoi, Y.He and J. Zhu, Retrieval-Based Face Annotation by Weak Label Regularized Local Coordinate Coding, *Proc. 19th ACM Int'l Conf. Multimedia(Multimedia)*, 1 (2011) 353-362.
7. Rajshree S. Dubey, Niket Bhargava and Rajnish Choubey, Image Mining using Content Based Image Retrieval System, *International Journal on Computer Science and Engineering*, 02 (2010) 2353-2356.
8. S.C.H. Hoi, R.Jin, J.Zhu and M.R. Lyu, Semi Supervised SVM Batch Mode Active Learning with Applications to Image Retrieval, *ACM Trans. Information Systems*, 27 (2009) 1-29.
9. Ricardo da S. Torres, Alexandre X.Falcãoa, Marcos A.Gonçalves b, João P.Papaa, Baoping Zhangc, Weiguo Fanc and Edward A. Foxc, A genetic programming framework for content-based image retrieval, *Pattern Recognition(ELSEVIER)*, 42 (2009) 283-292.
10. Dong Xu, Dacheng Tao and Stephen Lin, Marginal Fisher Analysis and Its Variants for Human Gait Recognition and Content-Based Image Retrieval, *IEEE Transactions on Image Processing*, 16 (2007) 2811-2821.
11. X.J. Wang, L.Zhang, F.Jing and W.Y. Ma, Anno Search: Image Auto-Annotation by Search, *Proc. IEEE Computer Science Conference Computer Vision and Pattern Recognition(CVPR)*, 06 (2006) 1483-1490.
12. Ji Zhang, WynneHsu and Mong Li Lee, Image Mining: Trends and Developments, *Journal of Intelligent Information Systems*, 19 (2002) 7-23.
13. A.W.M. Smeulders, M.Worring, S.Santini, A.Gupta and R.Jain, Content-Based Image Retrieval at the End of the Early Years, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22 (2000) 1349-1380.