ISSN: 2321-7782 (Online)
Volume 2, Issue 1, January 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Paper
Available online at: www.ijarcsms.com

Reverse Image Search Engine using Compact Composite Descriptor

Pushpa M. Chutel[1]
Department of Computer Science & Engg.
G.H. Raisoni college of Engineering
Nagpur - India

Apeksha Sakhare[2]
Department of Computer Science & Engg.
G.H. Raisoni college of Engineering
Nagpur - India

Abstract: ***With the massive growth of World Wide Web people are able to gain the access to massive amount of information. This information may contain data, images, flow charts, logos, maps etc. However locating and finding the relevant information is always a difficult task. A lot of textual-based search engines are available, but there are very rare image search engines through which desired images can be searched from the pool of images available on the web. Hence there is a need of such a image search engine using which the related and exact images can be searched. Reverse image search is such a content-based image retrieval query technique that involves providing the CBIR system with a sample image that it will then base its search upon. Reverse image search can be used to search either data related to the query image or the images related to that image or exact images.***

Keywords: Reverse Image Search; CCV; Image Feature Inverted Index; Compact Composite Descriptor.

I. INTRODUCTION

The World Wide Web contains a great quantity of information. The World Wide Web has grown very considerably in size, and is increasing enormously. Due to this enormous information retrieving the information of interest becomes very difficult. A lots of search engines are available for retrieving this information. This information can be text, images or visual information.

If focused on visual information, the World Wide Web contains several kinds of images and other visual information, such as videos, movies, and comic strips, in various formats such as JPG and GIF for still images, and MPG, AVI, and RAM for moving images. Photographs of people, museum collections and paintings, medical libraries, erotic, maps and charts, star photos and movie shots, advertising tapes, greeting cards, logos, sports images, humorous images, and comic strips are good examples of collections on the web.[1]

The most of the search engines available today are textual based i.e., they are all keyword based. One or more keyword is to be provided only then it can find the related web document. This web documents also contains images and most of the image search engine present today are primitive.[2]

The image search engines available on the web are basically categorized into two categories, one in which images are searched based on certain keywords. These keywords are used to describe which images or which type of images user is looking for. And based on those keywords similar images can be searched. But many a times, user finds their self in a situation when they don't find any keywords to describe it.

The other type of image search engines however searches the images based visual feature. They are based on the concept that the pictures speak a lot than thousands of words. For this, these engines are to be provided with a query image rather than keywords. This image search engines retrieves the images based on the query image. For searching the images based on query image many algorithms have been proposed.

© 2014, IJARCSMS All Rights Reserved

The searching process comprises of first indexing the images available over the web and then performing matching of the query image with the images in the web to retrieve the desired images .This paper lists all the available techniques for searching images on the web. And will try to find which technique is more feasible for searching the images based on query image and finally based on the retrieved results performance evaluation is done.

II. Reverse Image Search Engine

Reverse image search (RIS) is such a content-based image retrieval query technique that involves providing the CBIR system with a sample image that it will then base its search upon. Reverse image search is basically used to search either data related to the query image or the images related to that image or exact images.

Reverse image search allows users to discover content that is related to a specific sample image, popularity of an image, discover manipulated versions and derivative works, A regular image search on any search engine begins by entering a keyword, which is followed by a set of images related to that image. However, a lot of times, there is a need for searching with a given image and finding related images and also to find exact match of that image i.e. instead of giving a text as an input to search, an image is to be given as an input. This reverse image search based engine will not only find the similar images but will also find the exact match of the input image.

III. Review of Existing Approaches

Different approaches are available for searching the images on the web which makes use of different techniques. This section will survey all the existing approaches for searching the images using image search engines.

A. CCV (Color Coherence vector)

Color coherence vector is an image processing technique when applied to an image it exacts contents from that image. And stores content information in a coherence vector which can then be used for matching two images. CCV defines color coherence as degree to which the pixel of an image are members of large similarity colored region. If the query image is compared with the images stored in the database and the difference between them is found to be greater than 1000 then the difference is considered to be significant and the change is detected.[3]

However many a times this technique alone fails to find similarity between two images. Hence is not so effective.

B. Sobel Edge Detection and CCV

The technique presented in [3] [4] presented an idea that when a combination of sobel edge detection and CCV is used the percentage of similarity for matching query image with images in the web increases. Firstly it performs Sobel edge detection, contents from the uploaded image and image on the web are exacted .Then CCV matrix technique is used for the comparison which checks the content similarity between the images being compared for similarity.

C. New Wavelet Feature

Wavelet based features discussed in [5] were first introduced in Jacobs et al. it selects sixty four largest Haar wavelet coefficients in each of the 3 color band and stores them in feature vector as +1 or -1 along with their position in the transformation matrix. Low frequency coefficients tend to be more dominant than those of the high frequency coefficients and this makes this algorithm ineffective for images with sharp color changes. In addition to that, Haar wavelet basis is not suitable for natural images.[6].

© 2014, IJARCSMS All Rights Reserved

D. Discrete wavelet Transform

Wavelet transform has emerged as an effective tool for analysing texture features as it decomposes the image into various sub-bands which is multi-scale oriented. For computational convenience, the given input color image with size N÷N is converted to a sequence X(n) where n = N*N. The Discrete Haar wavelet transform described in [7] is adopted to decompose the input image by computing the approximations and details, which result in the low-pass and high-pass filters respectively. The decomposition process in [8] is performed till obtaining the optimum level i.e. fine level of the image.

E. RISE DCT Transform

RISE DCT transform is a technique in [9] was designed for the robust image search engine. The main idea of this technique is to develop an index for images database using the subset of JPEG coefficients in a compressed image. It creates an average of color components of each 8 x 8 pixel blocks in a JPEG image, which is the same as the DC-coefficient of each block after applying discrete cosine transform (DCT). The DC coefficient for each color component is used to develop an index. The main advantage of using average is that it can be applied to any image format. It is also more practical since it do not use the AC components of the DCT-transformed image [9].

F. Image Feature Inverted Indexing(FII)

Despite a lot of research on extraction of image feature, the design of feature index for high efficiency is yet to be worked out. The present content-based image retrieval engine takes the image feature into database so that more image features lead to more frequent modification of table field. Therefore, a new type of data indexing suitable for retrieval engine is needed. Combining image shape and texture feature with vector machining-supporting type, a new technique was designed named feature inverted indexing

This FII technique in [10] was used for image search engine Eva, it extracts image shape feature by using Zernike moments. Eva extracts image texture feature by using Gray level co-occurrence matrix algorithm depicting characteristic of texture. According this algorithm, Eva retrieves seven texture characteristics: uniform, contrast, correlation, variance, inverted-deviation-matrix, standard-average, standard deviation. Based on this features the query image is then compared with the images in the web

G. FCTH and CEDD

FCTH and CEDD techniques explained in [6] deals with the extraction of a new low level feature that combines, in one histogram, color and texture information. These features are named FCTH - Fuzzy Color and Texture Histogram and CEDD - Color and Edge Directivity Descriptor. FCTH results from the combination of 3 fuzzy systems and size is up to 72 bytes per image. CEDD size is up to 54 bytes per image and so these descriptors are suitable for large image databases. CEDD need less computational power for extraction than most of the MPEG-7 descriptors. And thus is efficient in retrieving distorted and transformed images. Compact Composite Descriptors (CCD) are able to capture both, color and texture characteristics, and so very useful in a very compact representation [11][12].

H. Content Oriented Image Retrieval (COIR)

COIR technique presented in [2] partitions the original image into several regions based on visual characteristics like color, edge, position, and texture using image processing techniques like edge detection[8][13], color analysis and region division. Each such region is considered as objects. The number of regions depends on the image contents.

The indexing engine calculates the visual attribute values for each region. The attributes assigned to each region are color, shape, texture, size, object location and object composition. The extracted attribute values are stored as metadata [14][15] and used during the query processing phase to determine similarity. Image index are stored as matrix and by comparing the matrix of the original image with that of in the web images are retrieved.

© 2014, IJARCSMS All Rights Reserved

Various techniques available for searching images in the web have been discussed. CCV extracts the contents from images but many times it fails to find similarity between images, it can be handled if used in combination with the Sobel edge detection [16]. New wavelet feature makes use of HAAR transform, but it is not suitable for natural images. It has discussed some techniques which are being adopted in many image search engines like RISE DCT transform in RISE, FII in Eva, COIR in AMORE [17], which are effective for those engine but they cannot search image in the web if query image is small or large or being rotated or scaled, or may have some color changes, neither it can search exact images to the query image [18]. However compact composite descriptors like FCTH and CEDD are more suitable for extracting exact images and are effective for searching images even if the searched image is small or large or being rotated or scaled, or may have some color changes, to the query image .

IV. PROPOSED SYSTEM AND ARCHITECTURE

A general architecture of the proposed RIS system is depicted in Fig 1. When user gives a query image, the block based low level feature from an image is exacted basically in terms of intensity and texture contrast and then clustering of this feature space is done to form meaningful patterns. Various algorithms are used for feature space clustering like CEDD, BCTH, CLD, FCTH etc. These algorithms automatically define the number of clusters.

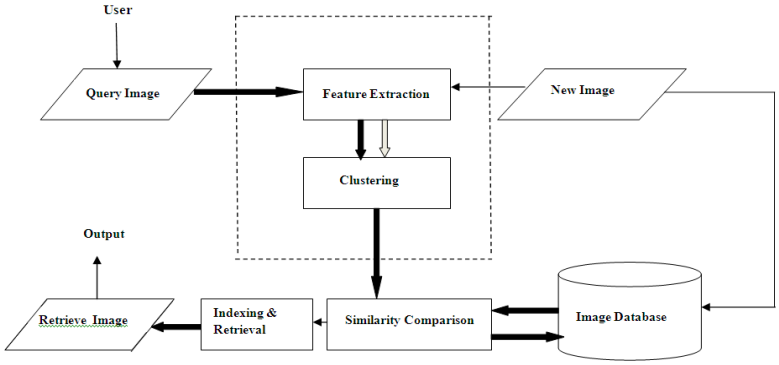

Fig 1. General Architecture of RIS system

V. IMPLEMENTATION AND EVALUATION

Reverse image search engine will retrieve both similar and exact image. For the initial implementation of the RIS, a system has been proposed. This RIS system is based on Compact Composite Descriptors(CCD), since they are global image features capturing both, color and texture characteristics, at the same time they are very useful in a very compact representation which is suitable for large image databases. This system will take an input image as a query image by browsing the image database folder which is shown in Fig 2. Once the image is selected, the applying algorithm is selected by which the system should retrieve the similar images. This applying algorithm includes CEDD, FCTH, BTDH, SCD etc. Based on these descriptors the performance of the system is evaluated and the results are also shown below.

A. Color And Edge Directivity Descriptor(CEDD)

The Color Edge Directivity Descriptor is a composite image descriptor that stores the color and texture information of an image in its histogram. CEDD size is many times restricted to 54 bytes per image, making this descriptor suitable for many large image databases. This descriptor is capable of retrieving accurate images even if the image has undergone much deformation such as noise, transformation, smoothing and various illumination changes. Hence, it makes CEDD suitable for being adopted by various popular search engines for retrieving exact images to the query image. Fig 3.Shows the results retrieved with CEDD descriptor with their deviation from query image.

© 2014, IJARCSMS All Rights Reserved

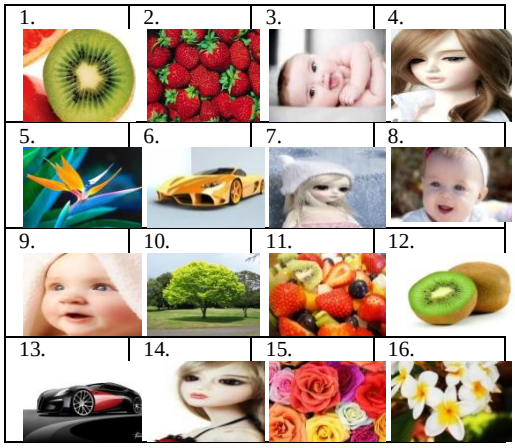

Fig 2.Sample images in image database folder

B. Fuzzy color and Text Histogram(FCTH)

The Fuzzy Color and Texture Histogram is also a composite descriptor which is similar to CEDD descriptor that stores image color and shape and texture information in its histogram. But unlike CEDD, it captures the texture information through the Haar Transform. So usually the results from both the descriptors are somewhere similar.

C. Color Layout Descriptor (CLD)

The Color Layout Descriptor is used to extract color of any region in the image through spatial distribution. It first divides the input image into a grid of 64 blocks, and then through the medium of color the dominant color of each block is extracted.

D. Edge Histogram Descriptor(EHD)

Edge represents an important content of an image. These descriptor stores the variation of frequency and brightness of an image into its histogram hence called as Edge Histogram Descriptor (EHD).

E. Scalable Color Descriptor (SCD)

Scalable Color Descriptor is a Color Histogram which is encoded by a Haar transform with a uniform quantization of the HSV space to 256 bins. The bin values are non-uniformly quantized to a 11-bit value. This descriptor is useful for finding visually similar images and percentage of similarity between two images.

F. Brightness and Texture Directionality Histogram (BCTH)

Brightness and Texture Directionality descriptor extracts the brightness, texture characteristics and spatial distribution into compact 1D vector and stores it into its histogram. The most important characteristic of this descriptor is that its size adapts according to the storage capabilities of the application that is using it. Thus it is suitable for large image databases.

Table1. Shows the results generated when using the respected descriptors for retrieving the images from image database folder.

© 2014, IJARCSMS All Rights Reserved

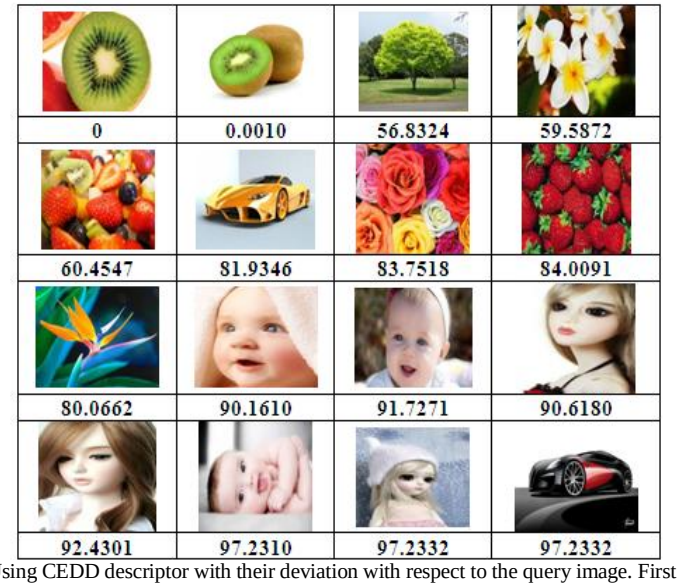

Fig 3.Image retrieved Using CEDD descriptor with their deviation with respect to the query image. First image is query image.

	CEDD	FCTH	BTDH	CLD	EHD
1.	**0**	**0**	**0**	**0**	**1.0257**
2.	84.0091	86.1636	17.7390	81.1755	158.4569
3.	97.2310	85.1994	26.5752	80.1866	284.6459
4.	92.4301	75.9961	18.5312	74.5857	174.5727
5.	80.0662	91.3275	45.4957	161.3544	223.0667
6.	81.9346	73.4104	25.7135	77.6321	279.7159
7.	97.2332	86.1944	14.8225	97.8880	154.7465
8.	91.7271	76.0339	16.2835	71.6125	176.3498
9.	90.1610	75.9961	13.6908	58.2786	235.7550
10.	47.9864	48.4870	7.9374	70.4321	243.0155
11.	60.4547	51.7857	9.0672	58.3744	143.5225
12.	**0.0010**	**0.0012**	**0.0115**	**0.1124**	**2.2579**
13.	97.2332	88.7594	61.7879	95.7787	300.9562
14.	90.6180	73.1944	19.9355	83.5157	268.8175
15.	83.7518	59.9118	6.3976	71.6705	184.8005
16.	59.5872	50.5672	36.8059	57.4905	126.9289

Table1. Image retrieval results of given descriptors based on Query image

The results shown in table shows that for any compact composite descriptor selected we get the result for all the images stored in the image database folder which is in terms of deviation i.e., the difference of dissimilarity between query image and images being used to check for similarity. It has been seen cedd, fcth, bcth and cld are able to match the exact images as their deviation is 0 and hence can be used for matching the exact and similar images as they are resistant to many deformation and illumination changes.

VI. Conclusion and Future Work

In this paper, various techniques available for image retrieval have been discussed. These traditional techniques faces many difficulty when exact images are to be retrieved, hence a concept of reverse image search has been proposed and designed. And thus the results generated from this system have been discussed. This system mainly aims at retrieving the exact as well similar images from web based on a query image. According to the concept of this system till now it is able to retrieve the exact images from a given image database folder by mean of compact composite descriptors. And will try to implement this system on large image databases which can then be used for detecting unauthorized use of brands and copyright images.

References

1. M. L. Kherfi, D. Ziou, A. Bernardi, “Image Retrieval From the World Wide Web:Issues, Techniques, and Systems”, ACM Transaction on Computing Surveys, Vol. 36, No. 1, March 2004, pp. 35–67.

2. Sougata Mukherjea, Kyoji Hirata and Yoshinori Hara, “AMORE: A World Wide Web image retrieval engine”, ACM, Journal of World Wide Web , Volume 2, Issue 3, 1999.

© 2014, IJARCSMS All Rights Reserved

3. Divya venkata, divakar yadav." Image query based search engine using content based image retrieval". 2012 14th International Conference on Modelling and Simulation.

4. Divya Ragadha, Deepika Kulshreshtha and Divakar Yadav "Techniques for refreshing Images on Web Document ", 2011 International Conference on Control, robotics and Cybernetics.

5. A. Lakshmi Subrata Rakshit "New Wavelet features for image indexing and retrieval". 2010 IEEE 2nd International Advance Computing Conference.

6. P.Praveen Kumar, Aparna, Dr K Venkata Rao PhD." Compact Descriptors for Accurate Image Indexing And Retrieval: Fcth And Cedd". International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 8, October – 2012 ISSN: 2278-0181

7. Ting Yao, Chong-Wah Ngo, Tao Mei." Circular Reranking for Visual Search". IEEE Transaction on image processing, vol. 22, no. 4, April 2013.

8. Dr. K. Seetharaman, M. Kamarasan "A Smart Color Image Retrieval Method Based on Multiresolution Features", Computational Intelligence & Computing Research (ICCIC), 2012 IEEE International Conference on. A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

9. Goswami, D, Bhatia, S.K. "RISE: A Robust Image Search Engine" , 2006 IEEE International Conference on Image Processing

10. Tan Hao, Chen Yu, Qiu Hang ,"Implementation Of FII-based image retrieval engine", Apperceiving Computing and Intelligence Analysis, 2009. ICACIA 2009. International Conference on

11. Konstantinos Zagoris, Savvas A. Chatzichristofis, Nikos Papamarkos and Yiannis S. Boutalis, "img (Anaktisi): A Web Content Based Image Retrieval System" , 2009 Second International Workshop on Similarity Search and Applications.

12. Harsh Kumar Sarohi, Farhat Ullah Khan ,"Image Retrieval using Perceptual Hashing", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 9, Issue 1 (Jan. - Feb. 2013), PP 38-40.

13. Piyush Kansal and Vinay Krishnamurthy,"Reverse Image Search".

14. Xinmei Tian, Yijuan Lu, Linjun Yang." Query Difficulty Prediction for Web Image Search". IEEE Transaction on multimedia, vol. 14, no. 4, August 2012.

15. Yin-Hsi Kuo, Wen-Huang Cheng, Hsuan-Tien Lin, Winston H. Hsu." Unsupervised Semantic Feature Discovery for Image Object Retrieval and Tag Refinement". IEEE Transaction on multimedia, vol. 14, no. 4, August 2012.

16. Liu Dong-Fei, Chen Wei, "Research and Application of the Image Search Algorithm Based on LIRE", 2012 2nd International Conference on Computer Science and Network Technology.

17. David Edmundson and Gerald Schaefer. "Recompressing images to improve image retrieval performance ". Department of Computer Science, Loughborough University, Loughborough, U.K 2012 IEEE International Conference.

18. Boulbaba Guedri, Mourad Zaied and Chokri Ben Amar. "Indexing and Images Retrieval by Content ". 2011 IEEE 10th International Conference on Frontiers of Information Technology.

AUTHOR(S) PROFILE

Pushpa Chutel received the Bachelors degree in Computer Science and Engineering from Anjuman College of Engineering and Technology, Nagpur in 2009. She has 2 years of teaching experience from Jhulelal Institute of Technology. Her main area of interest includes Image processing, network security and pattern recognition. She is now pursuing Masters in Technology in Computer science and Engineering from Raisoni College of Engineering, Nagpur as a GATE student.

Apeksha Sakhare is currently working as an Assistant professor in G.H. Raisoni College of Engineering. She has completed Masters in Technology in Computer science and Engineering from Raisoni College of Engineering, Nagpur in 2010. She has completed Bachelors degree in Computer science and Engineering from Bapurao Deshmukh College of engineering, Wardha in 2008. She has one national paper, one international paper and six international journal publication to her credit.

© 2014, IJARCSMS All Rights Reserved