International Journal Of Computational Engineering Research (ijceronline.com) Vol. 3 Issue.1

An Histogram Based Approach for Content Based Image Mining

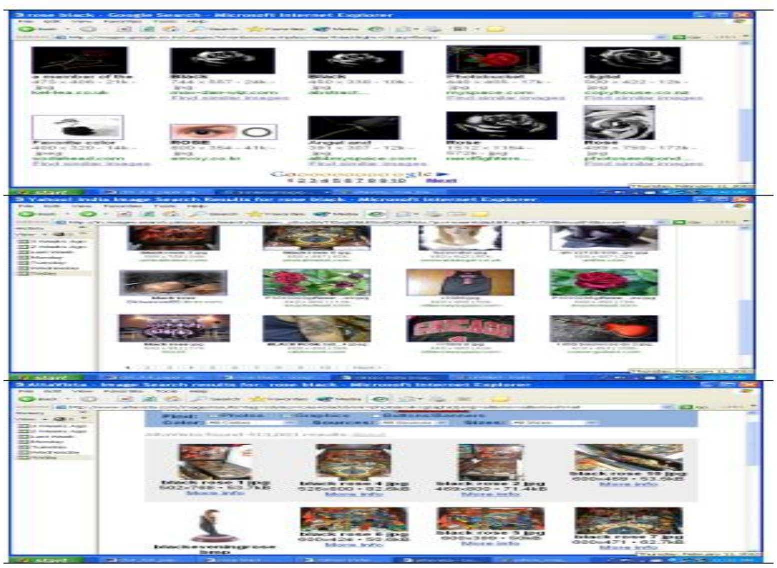
[1,]Mr. Ashish Bandre , [2,]Prof. Ramanand Singh , [3],Prof. S. G. Kerhalkar
[1,2,3,]Electronics & Communication Engg., Oriental Institute of Science & Technology, Bhopal

Abstract

Typical content-based image retrieval (CBIR) system would need to handle the vagueness in the user queries as well as the inherent uncertainty in image representation, similarity measure, and relevance feedback. We discuss how Histogram set theory can be effectively used for this purpose and describe an image retrieval system called HIRST (Histogram image retrieval system) which incorporates many of these ideas. HIRST can handle exemplar-based, graphical-sketch-based, as well as linguistic queries involving region labels, attributes, and spatial relations. HIRST uses Histogram attributed relational graphs (HARGs) to represent images, where each node in the graph represents an image region and each edge represents a relation between two regions. The given query is converted to a FARG, and a low-complexity Histogram graph matching algorithm is used to compare the query graph with the FARGs in the database. The use of an indexing scheme based on a leader clustering algorithm avoids an exhaustive search of the FARG database. We quantify the retrieval performance of the system in terms of several standard measures.

1. Introduction

Retrieval of required-query-similar images from abundantly available / accessible digital images is a challenging need of today. The image retrieval techniques based on visual image content has been in-focus for more than a decade. Many web-search-engines retrieve similar images by searching and matching textual metadata associated with digital images. For better precision of the retrieved resultant images, this type of search requires associating meaningful image-descriptive-text-labels as metadata with all images of the database. Manual image labeling, known as manual image annotation, is practically difficult for exponentially increasing image database. The image search results, appearing on the first page for fired text query rose black, are shown in Figure 1 for leading web search engines Google, Yahoo and AltaVista



Many resultant images of Figure 1 lack semantic matching with the query,
showing vast scope of research leading to improvements in the state-of-art-techniques. The need evolved two solutions – automatic image annotation and content based image retrieval. The content based image retrieval techniques aim to respond to a query image (or sketch) with query-similar resultant images obtained from the image database. The database images are preprocessed for extracting and then storing –indexing corresponding image features. The query image also gets processed for extracting features which are compared with features of database images by applying appropriate similarity measures for retrieving query similar-images.

2. Literature Survey:

The biggest issue for CBIR system is to incorporate versatile techniques so as to process images of diversified characteristics and categories. Many techniques for processing of low level cues are distinguished by the characteristics of domain-images. The performance of these techniques is challenged by various factors like image resolution, intra-image illumination variations, non-homogeneity of intra-region and inter-region textures, multiple and occluded objects etc. The other major difficulty, described as semantic-gap in the literature, is a gap between inferred understanding / semantics by pixel domain processing using low level cues and human perceptions of visual cues of given image. In other words, there exists a gap between mapping of extracted features and human perceived semantics. The dimensionality of the difficulty becomes adverse because of subjectivity in the visually perceived semantics, making image content description a subjective phenomenon of human perception, characterized by human psychology, emotions, and imaginations. The image retrieval system comprises of multiple inter-dependent tasks performed by various phases. Inter-tuning of all these phases of the retrieval system is inevitable for over all good results. The diversity in the images and semantic-gap generally enforce parameter tuning & threshold-value specification suiting to the requirements. For development of a real time CBIR system, feature processing time and query response time should be optimized. A better performance can be achieved if feature-dimensionality and space complexity of the algorithms are optimized. Specific issues, pertaining to application domains are to be addressed for meeting application-specific requirements. Choice of techniques, parameters and threshold-values are many a times application domain specific e.g. a set of techniques and parameters producing good results on an image database of natural images may not produce equally good results for medical or microbiological images.

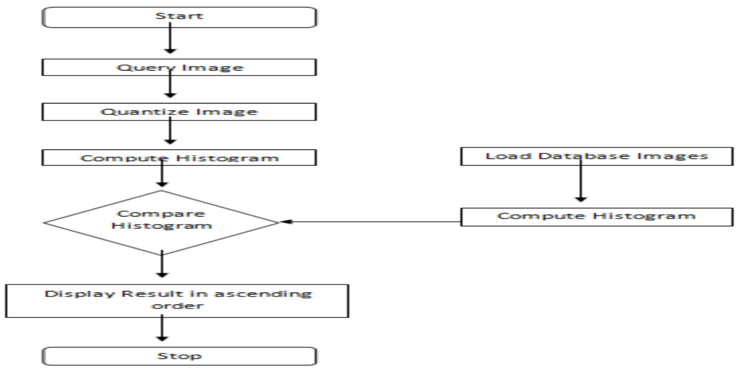
3. Methodology:

The color histogram for an image is constructed by counting the number of pixels of each color. In project, algorithms follow

(1) Selection of a color space
(2) Quantization of the color space
(3) Computation of histograms.

The approach more frequently adopted for CBIR systems is based on the conventional color histogram (CCH), which contains occurrences of each color obtained counting all image pixels having that color. Each pixel is associated to a specific histogram only on the basis of its own color, and color similarity across different color dissimilarity in the same is not taken into account. Since any pixel in the image can be described by three components in a certain color space histogram, i.e., the distribution of the number of pixels for each quantized color, can be defined for each component. By default the maximum number of colors one can obtain using the histogram function is 256. The conventional color histogram (CCH) of an image indicates the frequency of occurrence of every color in an image. The appealing aspect of the CCH is its simplicity and ease of computation.

4. Data Flow Diagram



The images data used in the experiment were taken from digital camera & few of the images were downloaded from a web site to create large database.

5. Quantization of the color space

Computing a color histogram for an image, the different color axes are divided into a number of so-called bins. A three dimensional 256x256x256 RGB histogram would therefore contain a total of 16777216 such bins. When indexing the image, the color of each pixel is found, and the corresponding bin's count is incremented by one.

6. Histogram Comparison

In order to compare histograms of two images, we first need to generate specific codes for all histogram bins. In this experiment, (r: 0-255, g: 0-255, b: 0-255) codes were generated for RGB histogram bins. When the images have been quantized into histograms, a method of comparing these is needed. Two Histogram are compared with an Equation

$$L1 = \sum_{i=1}^{n}(Q_i - I_i)$$

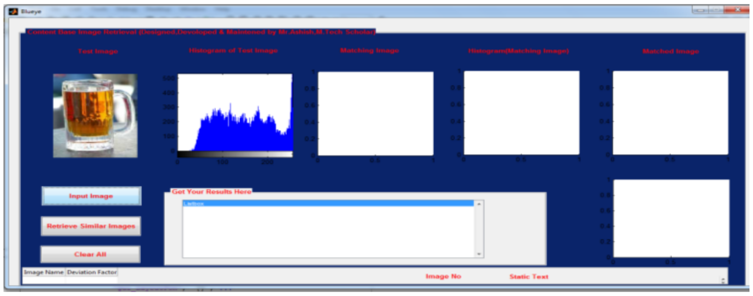
Where

Q_i is the value of bin i in the query image.

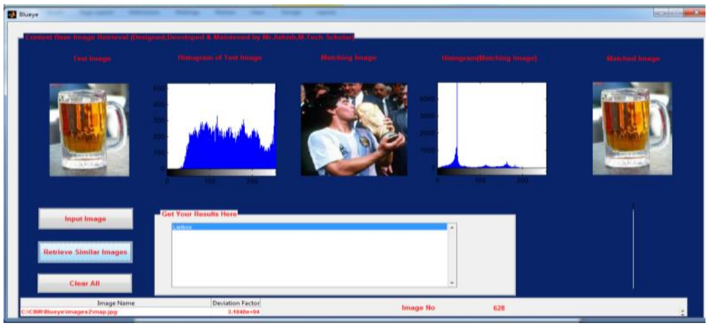
I_i is the corresponding bin in the database image.

Experiment Result

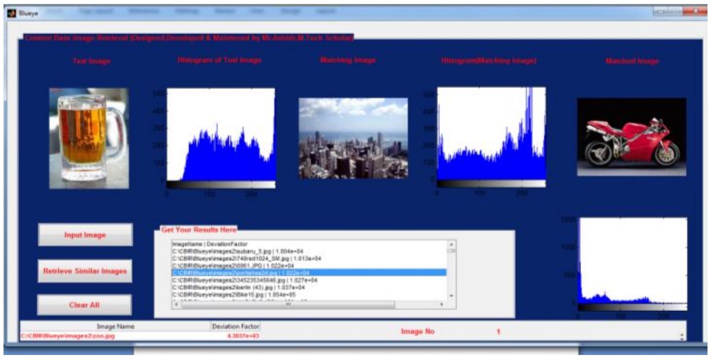
- **Project GUI**



Input Image With Histogram



Input Image With Exact Matched



Other Matched Images With respect to input Image

Test	1	2	3	4
Image Type	jpg	jpg	jpg	Jpg
Size	471 X 600	150 X 100	75 X 50	15 X 10
Rotate	No	No	No	No
Image Format	24 Bits RGB	32 Bits RGB	32 Bits RGB	32 Bits RGB
Mining Turn Around time	00:00:14:65	00:00:04:01	00:00:03:02	00:00:03:07
Overall Image Retrieval Time	00:00:53:682	00:00:06:58	00:00:03:95	00:00:03:14
Deviation Factor	0	9.26	11.49	48.52
Overall Mining Precision	100%	100%	100%	100%

Result Analysis with respect to parameters (After Resize the image)

Test	1	2
Image Type	Jpg	jpg
Crop	Yes	Yes
Size	167 X 77	142 x 84
Image Format	32 Bits RGB	32 Bits RGB
Mining Turn Around time	00:00:03:0484	00:00:03:054
Overall Image Retrieval Time	00:00:05:889	00:00:04:67
Deviation Factor	0	0
Overall Mining Precision	100%	100%

Result Analysis with respect to parameters (After Crop the image)

Test	Total Number of Relevant Images	Number of Relevant Images Retrieved	Total Number of Retrieved Images	Recall	Precision	Result By Srinivasa Kumar [9]	
						Recall	Precision
1	46	42	45	**91.30**	**93.33**	63.04	64.44
2	47	42	46	**89.36**	**91.34**	78.94	68.00
3	45	38	42	**84.44**	**90.37**	66.66	80.00
4	42	35	41	**83.33**	**85.36**	83.33	69.44
5	40	25	35	**59.52**	**71.42**	71.80	65.70

Comparison of Our Method with Shrinivasa Kumar Result

7. Conclusion

The main objective of the image mining is to remove the data loss and extracting the meaningful information to the human expected needs. The images are preprocessed with various techniques and the texture calculation is highly focused. Here, images are clustered based on RGB Components .Histogram is used to compare the images with some threshold constraints. This application can be used in future to classify the medical images in order to diagnose the right disease verified earlier.

8. Future Scope

- This system is useful in future to detect the diseases related with human.
- More effort to be taken to reduce the Image retrieval time of an given input Query Image
- In future this system is also implemented in the field of computer Vision which is concerned with the automated processing of images from the real world to extract and interpret information on a real time basis.
- In future these system is used in Astronomy to the study of celestial objects (such as stars, planets, comets, nebulae, star clusters and galaxies).

References

[1] Ji Zhang Wynne Hsu and Mong Li Lee, "Image Mining: Trends and Developments".
[2] A.Kannan,Dr.V.Mohan and Dr.N.Anbazhagan,"Image clustering and Retrieval using Image Mining Techniques", IEEE International Conference on Computational Intelligence and Computing Research, 2010.
[3] D.S. Zhang and G.Lu, "Content Based Image Retrieval Using Texture Features", In Proc. Of First IEEE Pacific-Rim Conference on Multimedia PCM'2000, pp.392-395.
[4] Yixin Chen, James Z.Wang and Robert Krovetz "Cluster Based Retrieval of Images by Unsupervised Learning". IEEE Transaction on Image Processing, August 2005, Vol 14, pp.1187-1199.
[5] Hewayda M.Lofty, Adel S.Elmaghraby, "CoIRS: Cluster Oriented Image Retrieval System", IEEE conference on tools with Artificial Intelligence, 2004.
[6] M. Datcu and K. Seidel," Image Information Mining: Exploration of Image Content in Large Archives", IEEE Conference on Aerospace, 2000, Vol.3.
[7] Zhang Y.J,"A Survey on evaluation methods for image segmentation," vol.29, no.8, 2000, pp.1335-1346.
[8] Ji Zhang Wynne Hsu Mong Li Lee," Image Mining: Issues, framework and techniques".
[9] Srinivasa Kumar Reddy "Content Based Image Retrieval System", Georgian Electronic Scientific Journal No.5, 2009.
[10] A. Kitamoto "Data Mining for Typhoon Image Collection", In Proceedings of the Second International Workshop on Multimedia Data Mining, August, 2001, San Francisco, CA, USA.
[11] M. Sasikalal and N. Kumaravel "Comparison of Feature Selection Techniques for Detection of Malignant Tumor in Brain Images", IEEE Conference, 11 -13 Dec 2005. Chennai, India.
[12] C.Ordonez and E.Omiecinski,"Discovering Association Rules Based on Image Content" Proceedings of the IEEE Advances in Digital Libraries Conference (ADL'99).
[13] W. Hsu, M. L. Lee and K.G.Goh," Image Mining in IRIS: Integrated Retinal Information System ", In Proceedings of ACM SIGMOD International Conference on the Management of Data, May 2000, Dallas, Texas.
[14] U. M. Fayyad, S. G. Djorgovski, and N.Weir," Automating the Analysis and cataloging of Sky Surveys" Advances in Knowledge Discovery and Data Mining, 1996, 471-493.
[15] O. R. Zaiane and J. W. Han," Mining Multimedia Data", Meeting of Minds, November 1998, pp 83-96, Toronto, Canada.