

An Unsupervised Cluster-based Image Retrieval Algorithm using Relevance Feedback

Jayant Mishra¹, Anubhav Sharma² and Kapil Chaturvedi³

¹Department of Computer Application, UIT, RGPV, Bhopal
rgtu.jayant@gmail.com

²Department of Computer Science, RITS, Bhopal
ac.anubhav@gmail.com

¹Department of Computer Application, UIT, RGPV, Bhopal
kplchaturvedi@yahoo.com

ABSTRACT

Content-based image retrieval (CBIR) systems utilize low level query image feature as identifying similarity between a query image and the image database. Image contents are plays significant role for image retrieval. There are three fundamental bases for content-based image retrieval, i.e. visual feature extraction, multidimensional indexing, and retrieval system design. Each image has three contents such as: color, texture and shape features. Color and texture both plays important image visual features used in Content-Based Image Retrieval to improve results. Color histogram and texture features have potential to retrieve similar images on the basis of their properties. As the feature extracted from a query is low level, it is extremely difficult for user to provide an appropriate example in based query. To overcome these problems and reach higher accuracy in CBIR system, providing user with relevance feedback is famous for provide promising solution.

KEYWORDS

Content-based image retrieval, Color Histogram, Relevance Feedback, k-means.

1. INTRODUCTION

Content-based image retrieval [1,18] has been an active research area in recent years. There has been an experimental increase in the amount of non-text based data being generated from various sources. In particular, images have been gaining popularity as an alternative and some time more viable option for information storage [6]. Through the increase in storage and transmission abilities more visual information is being made available on-line. These need to search and well manage large volumes of Multimedia information [3]. However, while this presents a wealth of information, it also causes a great problem in retrieving appropriate and relevant information from databases.

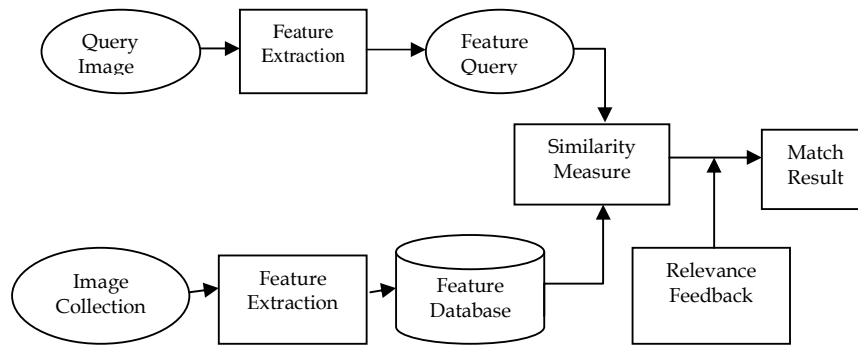


Fig. 1 General model of CBIR

CBIR extracts low-level features which is Inbuilt in the images to present the contents of images. Each image has features such as categories into three main classes: color [2, 11], texture [9, 11] and shape [2, 10] features. This has resulted in a growing interest, and great active research, into the extraction of relevant information from non text-based database. One of the most common data in multimedia systems is image. Therefore, the capability of managing images, allocating their quick and efficient indexing and retrieval are repeatedly followed in multimedia database management systems. Researchers have been focusing on different ways of retrieving information from database or objects based on their contents. Especially, content-based image retrieval (CBIR) systems have attracted significant research and commercial interest. In field of medicals, academic world, government and huge collections of images are being produced. Many of these groups are the product of digitizing existing collections of analogue photographs, diagrams, drawings, paintings, and prints.

Color is most common feature of an image such as used in Content-Based Image Retrieval [2, 9, 13]. These features are able to identify objects [10] and retrieve similar images on the basis of their contents. These methods do work very efficient in object recognition and Web searching [14]. An efficient query reformulation is necessary for finding the relevant images from the database. Relevance feedback (RF) is an interactive process which filters the retrieval results to a particular query by utilizing the user's feedback on previously retrieved images [5, 8].

II. Feature Extraction

Feature (content) extraction is the basis of content-based image retrieval. In sense, features may include both visual features (color, texture, shape) and text-based features (key words, annotations).

Color

Color feature is one of the most reliable and easier visual features used in image retrieval. It is robust to background complication and is independent of image size and orientation [3]. A lot of techniques available for retrieving images on the basis of color similarity from image database [1]. Each image included to the collection is analyzed to compute a color histogram which shows the proportion of pixels of each color within the image. The color histogram for each image is then stored in the database. Color, texture and shape information have been the primitive image descriptors in content based image retrieval systems [7].

Texture

Textures are characterized by differences in brightness with high frequencies in the image spectrum. They are useful in distinguishing between areas of images with similar color similar color (such as sky and sea, or water, grass). A variety of methods has been used for measuring texture similarity; the best- established depend on comparing values of what are well-known as second-order statistics estimated from query and stored images. Essentially, these estimate the relative brightness of picked pairs of pixels from each image. From these it is possible to measure the image texture such as contrast, coarseness, directionality and regularity [16] or periodicity, directionality and randomness [11].

Shape

Shape is an essential feature for perceptual object recognition and classification of images in content-based image retrieval. Shape representation is significant concern both in object recognition and classification. That means an image has to be segmented before extracting most shape features. The shape representation typically is divided into two categories; boundary-based and region-based. Fourier descriptors are a representative for the boundary-based method while moment invariants are characteristic of the region-based method [20].

III. Related Work

Color is common and important feature in image retrieval system. Color comparison between two images would however be time consuming and difficult problem to overcome this problem they introduced a method of reducing the amount of information. One way of doing this is by quantizing the color distribution into color histograms [12]. Using color histogram easier way for color distribution or they used histogram divide in to different classes for matching.

K-means algorithm is one of the most widely used clustering algorithms in spatial clustering analysis [17]. It is easy and efficient. But is also having limitations: It is sensitive to the initialization. It doesn't perform well in global searching and is easy to get into local optimization and the improved K-means is based on the classical method to make the process of optimization more dependent [4, 15].

Many of image retrieval applications are based on color feature and shape feature [3]. Estimating local texture based on pixels of the intensity image and a fuzzy index to point out the presence of major colors [11]. It is based on the texture co-occurrence matrix.

K-means clustering for the classification of feature set obtained from the histogram refinement method [4]. Histogram refinement provides a set of features for proposed for Content Based Image Retrieval (CBIR) [2]. They used global histograms for image retrieval, because of their effectiveness and insensibility to minor changes, are broadly used for content based image retrieval.

IV. Proposed Work

Each image has three features Color, Shape and Texture. For fast and improve Image retrieval performance we are using color feature extraction. Using color feature extraction firstly we converted color image into grey level, this is containing values from 0 to 255.

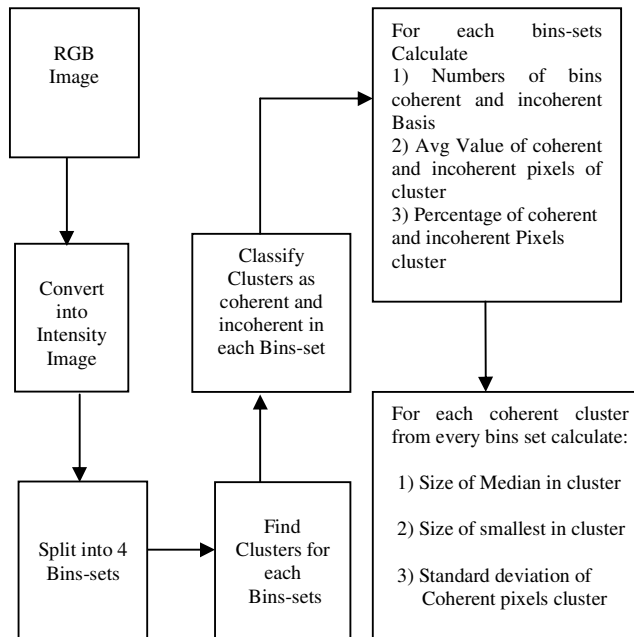


Figure1. Block diagram of algorithm.

Algorithm:

1. Input as RGB Image.
2. RGB image convert into gray image or Intensity Image:

$$I = \frac{1}{3} (R + G + B)$$

Intensity or gray image contains values from 0 to 255.

3. Gray Image converts into histogram bins sets

For i = 64 to 256

{

h(i) = g(x_i, y_i);

{

Distance = pdist(h(i));

Clust_distance = linkage [B, 'centroid'];

%%measure Distance between clusters

Clust = cluster (c, 'maxclust', 2);

%%Let (clust == 1) be the coherent and (clust == 2)

Incoherent cluster.

```

If (clust = = 1)
    Call Calculate_num (n);
    Call Calculate_Percent (Per,n);
    Call Calculate_Avg (a,n);
    Call Calculate_Median_bin(m,n);
    Call Calculate_Smallest_bin(s,n);
    Call Calculate_Std(t,n);
Else If (clust = = 2)
    Call Calculate_num (n);
    Call Calculate_Per(Per,n);
    Call Calculate_Avg(a,n);
Else
    "No cluster found"
End

```

6. Retrieval using query.

V. Result and Analysis

In this part, several tests are performing in order to analysis the performance of histograms and here we used coral image database. To test the proposed method. Firstly RGB image converted to intensity image. Then the image were split into bins sets and the features described in section IV were calculated which is based on coherent and incoherent pixels. Followed these feature set; images were grouped in similar clusters using k-means clustering method. Histogram refining method more refines the histogram by dividing the pixels in a given container into a number of classes based on color coherence vectors. Several features are calculated using Matlab[19] for each of the cluster and these features are further classified using the k-means clustering method.

Table1. Example of parameter values for Incoherent pixels.

Binset	(X _i)	(CX _i)	(μX _i)	(Y _i)	(CY _i)	(μY _i)
1	52	1	51262	48	61	747
2	53	1	52065	47	127	364
3	51	1	50382	49	191	251
4	47	1	46207	53	255	204

We denoted Percentage of Coherent pixels (X_i), Number of Incoherent pixels clusters (CX_i), Average of Incoherent pixels cluster (μX_i), Percentage of Incoherent pixels (Y_i), Number of Incoherent pixels clusters (CY_i), Average of Incoherent pixels cluster (μY_i). After calculation of above features, we consider some additional features based on coherent bins cluster only. At this stage, incoherent Pixels cluster is ignored. These features are preferred.

Similarly, we have additional parameter values related with the coherent bins in clusters including size of median bins in cluster in each bin (MX_i), size of smallest bins in cluster in each bin (SX_i) and standard deviation of coherent bins in clusters (DX_i). for each of the largest, median and smallest bin cluster value.

Table2. Additional features among the coherent Pixels cluster.

	Binset1	Binset2	Binset3	Binset4
(MX_i)	324	161	91	76
(SX_i)	0	0	2	0
(DX_i)	4227700	913210	541820	386960

Two of them are based on the size of the Cluster. These intensity values median, standard deviation, mean, variance, various sizes of objects are considered as appropriate features for retrieval.



(a) Query Image



(b) Retrieval Result

VI. Conclusion

This algorithm is based on color and texture features of static image. The intensity values median, standard deviation and various sizes of the intensity values are considered as appropriate features for retrieval. The proposed technique present very little amount of memory for features storage and a prominent rate of computation and give good results in terms of accuracy. We have shown that k-means clustering is fairly useful for appropriate image retrieval queries. The K-means clustering algorithms to group the images content into different clusters based on the color feature and k-means clustering algorithms have been often used in the pattern recognition. Animated images performance behaviour is different based on this method.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere thanks to my guide Prof. Nishchol Mishra his wisdom and abundance of encouragements from the very early stage of this research as well as giving me constant support throughout the work and I also give special thanks to Prof. Sanjeev Sharma Head, SOIT, RGPV for his great support and motivating me to complete my work. I would also like to give special thanks to my friends. If anyone is to be acknowledged first, they are my family member's for their inseparable support and prayers My Papa, Mommy and Both Sister's. I express my deep sense of gratitude to them for their support and continuous encouragement.

References

- [1]. Bing Wang et al."A Semantic Description For Content-Based Image Retrieval" College Of Mathematics And Computer Science, Hebei University, Baoding 071002, China -2008.
- [2]. Youngeum An et al. "Classification of Feature Set Using K-Means Clustering From Histogram Refinement Method" (Korea Electronics Technology Institute, South Korea)- --2008.
- [3]Waqas Rasheed et al. "Sum of Values of Local Histograms for Image retrieval" - Chosun University, Gwangju, South Korea-2008.
- [4]. Tapas Kanungo et al. "An Efficient K-Means Clustering Algorithm: Analysis AND Implementation" Senior Member, Ieee-2002.
- [5]. Samar Zutshi et al. "Proto-Reduct Fusion BASED Relevance Feedback IN Cbir", (Monash University)—2009.
- [6].James Z. Wang et al."SIMPLIcity: Semantics-sensitive Integrated Matching for Picture Libraries," IEEE Trans. on Pattern Analysis and Machine Intelligence- 2001.
- [7]. P. S. Hiremath et al.-"Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement" -Dept. of P.G. Studies and Research in Computer Science, Gulbarga University,Gulbarga, Karnataka, India-2007.
- [8]. P. Suman Karthik et al. "Analysis Of Relevance Feedback In Content Based Image Retrieval" Center For Visual Information Technology International Institute Of Information Technology Gachibowli, Hyderabad-India, -2006.
- [9]. Zhi-Gang Fan et al "Local Patterns Constrained Image Histograms For Image Retrieval" Advanced R&D Center Of Sharp Electronics (Shanghai) China -2008.
- [10]. Tang li et al. "Developing a Shape-and-Composition CBIR Thesaurus for the Traditional Chinese Landscape" University of Maryland College of Information Science College Park, MD, United States Library Student Journal, July 2007
- [11]. Sanjoy Kumar Saha et al. "CBIR Using Perception Based Texture And Colour Measures"CSE Department; CST Department Jadavpur Univ., India; B.E. College, Unit ISI, Kolkata, India - -2003.

- [12]. Greg Pass Ramin Zabih et al. "Histogram Refinement for Content-Based Image Retrieval" Computer Science Department Cornell University-1996.
- [13]. Mohamad Obeid' Bruno et al. "Image Indexing & Retrieval Using Intermediate features" Equipe MIRE ENIC, Telecom Lille , Cite Scientifique, Rue G. Marconi, 59658 Villeneuve d'Ascq France-2000.
- [14]. M.S. Lew et al. "Content-Based Multimedia Information Retrieval: State Of The Art And Challenges" *ACM Transactions On Multimedia Computing*- 2006.
- [15]. Zhe Zhang et al. "Improved K-means Clustering Algorithm"-(College of Automation, Northwestern Polytechnic University, Xi'an 710072) -2008.
- [16]. Tamura et al. "Texture Features Corresponding to Visual Perception"- IEEE Trans on system, Man and cyber 8-460-472-1978.
- [17]. Prof. Dr. K. Sakthivel et al. "Performance Enhancement In Image Retrieval Using Modified K-Means Clustering Algorithm"- Journal of Mathematics and Technology, Rangasamy College of Technology, Tiruchengode, Tamilnadu ISSN: 2078-0257, February, 2010.
- [18]. Morteza Analoui et al." Content-based Image Retrieval Using Artificial Immune System (AIS) Clustering Algorithms"-proceeding of the international Multi conference of Engineers and computer Scientists 2011 VolII, IMECS – India 2011.
- [19]. www.mathworks.com, MATLAB and Simulink for Technical Computing , 2008.

Authors

1. Jayant Mishra, Asst. Prof. UIT, RGPV, Bhopal. He has obtained M. Tech. Degree in Computer Application & Technology branch from School Of Information technology, RGPV, Bhopal in 2009. He is also done MCA from Technocrats Institute of Technology, Bhopal in 2007. He has worked as Asst. Prof. In Department of Computer Science, Acropolis Institute Research & Technology- Bhopal, Since Dec 2009 to March 2011. He's working as Asst. Prof. in UIT, Rajiv Gandhi Technical University, Bhopal (M.P.) Since March 2011. His one **international Journal** has been published in Dec'2009. His research area of interests are Multi Media Data mining & Digital Image Processing.



2. Anubhav Sharma. He has obtained BE from JIT Borawan (M.P.) in 2007. He's perusing M. Tech. (final sem) Degree in Computer Science branch from RITS-RGPV, Bhopal in 2011. His research area of interests is Digital Image Processing.



3. Prof. Kapil Chaturvedi. He has obtained MCA from SATI (Degree), Vidisha (M.P) in 2008. He's perusing Ph.D. from RGTU Bhopal. He has worked as lecturer In Department of Computer Applications, SATI (Degree), Vidisha (M.P.) Since Aug 2008 to March 2011 He's Working as Assistant Professor in Department of Computer Applications, UIT, Rajiv Gandhi Technical University, Bhopal (M.P.) Since March 2011. My Area of Interest is Data Mining & Digital Image Processing.

