An Efficient Framework for Medical Image Retrieval System using Contribution Mechanism

Dr. B.P. Mallikarjunaswamy[1]

Mr. Shivamurthy R.C[2]

[1] *Prof.*: Dept of Computer Science & Engg
Sri Siddharatha Institute of Technology
Tumkur, Karnataka:572106, India

[2] *Research Scholar*: Dept of Computer Science & Engg.
Akshaya Institute of Technology
Tumkur, Karnataka: 572106, India

***Abstract*— With the advancement in image capturing device, the image data been generated at high volume. If images are analyzed properly, they can reveal useful information to the human users. Content based image retrieval address the problem of retrieving images relevant to the user needs from image databases on the basis of low-level visual features that can be derived from the images. Grouping images into meaningful categories to reveal useful information is a challenging and important problem. We apply the algorithm to content-based image retrieval and compare its performance with that of the k-means clustering algorithm. Unlike the k-means algorithm, our algorithm optimizes on both intra-cluster and inter-cluster similarity measures. It has three passes and each pass has the same time complexity as an iteration in the k-means algorithm. Our experiments on a bench mark image data set reveal that our algorithm improves on the recall at the cost of precision.**

Keywords-; Content based image Retreival System, clustering, k-means algorithm, medical image retrieval

I. INTRODUCTION

The rapid progress in computer technology for multimedia system has led to a rapid increase in the use of digital images. Rich information is hidden in this data collection that is potentially useful in a wide range of applications like Crime Prevention, Military, Home Entertainment, Education, Cultural Heritage, Geographical Information System (GIS), Remote sensing, Medical diagnosis, and World Wide Web [1, 2]. Rich information is hidden in these data collection that is potentially useful. A major challenge with these fields is how to make use of this useful information effectively and efficiently. Exploring and analyzing the vast volume of image data is becoming increasingly difficult. The image database containing raw image data cannot be directly used for retrieval. Raw image data need to be processed and descriptions based on the properties that are inherent in the images themselves are generated. These inherited properties of the images stored in feature database which is used for retrieval and grouping. The strategy for earlier image retrieval system focused on "search-by-query". The user provides an example image for the query, for which the database is searched exhaustively for images that are most similar. Clustering is a method of grouping data objects into different groups, such that similar data objects belong to the same group and dissimilar data objects to different clusters [3,4]. Image clustering consists of two steps the former is feature extraction and second part is grouping. For each image in a database, a feature vector capturing certain essential properties of the image is computed and stored in a feature base. Clustering algorithm is applied over this extracted feature to form the group.

We use the notion of 'contribution of a data point' for partitional clustering. The resultant algorithm requires only three passes and we show that the time complexity of each pass is same as that of a single iteration of the k-means clustering algorithm. While the k-means algorithm optimizes only on the intra-cluster similarity, our algorithm also optimizes on the inter-cluster similarity. Clustering has widespread applications in image processing. Color-based clustering techniques have proved useful in image segmentation [13]. The k-means algorithm is quite popular for this purpose. Clustering based on visual content of images is an area that has been extensively researched for

several years [14]. This finds application in image retrieval. Content-based image retrieval (CBIR) aims at finding images of interest from a large image database using the visual content of the images. Traditional approaches to image retrieval were text-based, where individual images had to be annotated with textual descriptions [8]. Since this is a tedious manual task, image retrieval based on visual content is very essential.

Organizing the retrieved search results into clusters is an intuitive form of content representation [14] and facilitates user's browsing of images [15]. Image clustering can also be used to optimize the performance of a CBIR system [16]. While the performance of a number of clustering algorithms in image retrieval have been analyzed in previous works [17, 18, 19, 20], we apply our proposed algorithm to CBIR and compare its performance with that of the k-means clustering algorithm.

The rest of the paper is organized as follows. In section 2 we give an overview of related work which identifies all the major research work being done in this area. Section 3 highlights about the medical image retrieval system. Proposed system is discussed in section 4 followed by result and performance analysis in Section 5 and finally in section 6 we make some concluding remarks.

II. Related Work

Jing Huang ed al [1, 2] discussed new feature called color correlogram for image indexing and comparison. These new features computed efficiently and show that performance very well. Sim, D. G., H.

K. Kim and R. H. Park [3,4,5]" the image retrieval scheme for JPEG formatted image is presented. Content based image retrieval for JPEG images has attracted many people's attention and a series of algorithms directly based on the discrete cosine transform domain. And to take full advantage of DCT coefficients and consider the color and texture information for the retrieval of JPEG formatted images. Here decompressing the images and then performing in the spatial domain. The feature vectors are computed from several DCT coefficients. And this operation is performed in the partial decoded domain. It can greatly decrease the retrieval complexity. M. Flickner et.al [6,7,8] proposed Color histograms are computationally efficient, and generally insensitive to small changes in camera position.

However, a color histogram provides only a very coarse characterization of an image, An images with similar histograms can have dramatically different appearances. Here, to describe a method which imposes additional constraints on histogram based matching. In histogram refinement, the pixels within a given bucket are split into classes based upon some local property. Split histograms are compared on a bucket by bucket basis, similar to standard histogram matching. Within a given bucket, only pixels with the same property are compared. Two images with identical color histograms can have different split histograms, split histograms create a finer distinction than color histograms. This is particularly important for large image databases, in which many images can have similar color histograms.

To describe a split histogram called a color coherence vector (CCV), which partitions each histogram bucket based on spatial coherence. A database with 15,000 images can be queried using CCV's in under 2 seconds. And to demonstrate that histogram refinement can be used to distinguish images A. P. Berman et.al [9,10] found that technique fairly integrates a diverse and expandable set of image properties (color, texture, and location) in a retrieval framework, and allows end users substantial control over their use. We propose a novel set of evaluation methods in addition to applying established tests for image retrieval; our technique proves competitive with state of art methods in these tests and does better on certain tasks. The Stairs algorithm can operate in a regional query mode with only a moderate increase in computational overhead. For certain queries this capability significantly increases the relevance of the images retrieved. Furthermore, it improves on many standard image retrieval algorithms by supporting queries based on subsections of images. The merits of drawing on different types of image features for Image retrieval are firmly established.

Our work capitalizes on this trend, providing a framework for fairly and consistently integrating diverse image properties into a description amenable to fast, reliable retrieval. J. Zhang et.al [11, 12, 13] suggest the image retrieval based on the textural information of an image, such as orientation, directionality, and regularity. Here, utilize texture orientation to construct the rotated Gabor transform for extraction of the rotation-invariant texture feature. The rotation-invariant texture feature, directionality, and regularity are the main features used in the proposed approach for similarity assessment. Using these features, we finally propose an efficient mechanism for CBIR and examine it through some applications. The system can now compare features of the query with features of images in the collection based on some matching criterions. Because three features are used in this work, three matching scores need to be computed. A weighted average of the matching scores is then calculated to get a final score for each image.

Finally, rank images based on these final scores and top-ranked images are displayed to the user as the result of retrieval. Haralick RM et.al [14,15]discussed the four image features are extracted by this system, which are color feature (HSV color histogram), texture feature (co-occurrence matrix), shape feature (moment invariant based-on threshold optimization), spatial relationship feature (based-on the Markov chains). According to the statistical analysis of the experiment results discover that the four visual features describe image characters variously. The retrieval precision based on color feature is better than based on texture feature. An image retrieval method combined color and texture features. According to image texture characteristic, a kind of image feature statistic is defined. By using feature weight assignment operators designed here, the method can assign weight to color and texture features according to image content adaptively and realize image retrieval based on combined image features. The retrieval results are more exact and efficient than other methods based on single feature and simple linear combined features of fixed weight, the retrieval results are more suitable to the human visual characteristic. The error matching is decreased and weight assignment is logical.

P.S.Hiremath ed al," [16,17,18,19] discussed four approaches such as multispectral Approach, HSV color space, YCbCr color space,and uses gray scale texture features for color texture analysis. the wavelet decomposed coefficient of image and its complements by using texture feature. Their experiments are carried out on Wang's dataset using JSEG for segmentation and compare the four different color space finally haar wavelet is more effective in texture feature compare with other wavelet so, the results are encouraging. P. S. Hiremath and Jagadeesh Pujari [20]discussed An integrated matching scheme based on higher priority of similar image and the adjacency matrix of a bipartite graph by using tiles of query shape information is computed by Gradient Vector Flow fields. This demonstration is efficiency compare with wavelet method. K.P.

Ajitha Gladis and K.Ramar [21,22,23,24] discussed mainly as the image can be represented on statistical properties, morphological features and fuzzy cluster features of the image in order to get more accurate results. He distance is measured through a back propagation network.

So, experimental results is quite effective in both performance and retrieval rate. Son Lam Phung and A. Bouzerdoum [25] proposed new feature called edge density. It differentiates objects from non-objects using image edge characteristics. This approach is based on a fast object detection method. The edge density, which measures the specific region of the object, that can be computed more efficiently.

Where each feature is the average edge magnitude in a specific subregion. The new feature capability compared to the Harr-like features[30]. Finally new feature show good discriminative capability. S.Nandagopalan ed al [26,27,28] discussed texture for texture co-occurrence matrix based entropy, energy, etc, and for edge density , Edge Histogram Descriptor (EHD).For retrieval of images, finally to reduce the computational complexity based on greedy stategy. So, its achieved better results for both local and global feature.

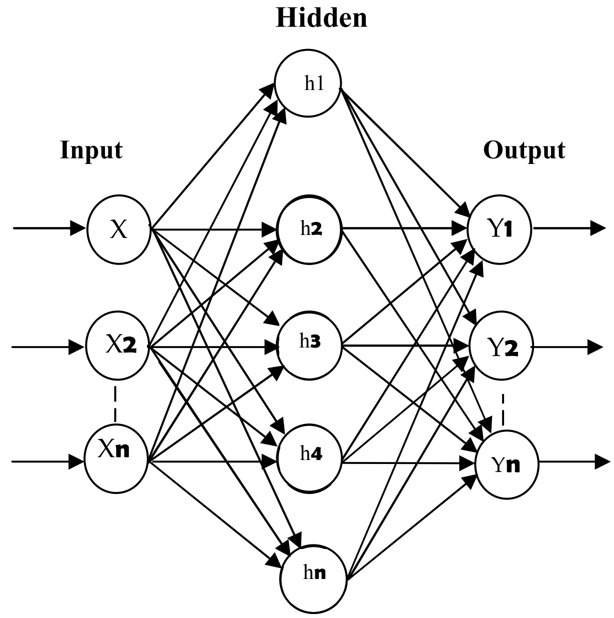


Fig 2: Back Propagation Network

Fig 3:Left: an image window. Middle: the edge magnitude.
Right: three edge density features

Mamta Juneja and Parvinder Singh Sandhu[29,30,31] proposed candy method for edge detection. Here to compare canny method with laplacian of Gaussian method .

So, the result shows canny's edge detection algorithms is performs better then laplacian method. Michele Saad [32.33] discussed to compare four color feature extraction algorithms such as

Operator	Canny	Lap of Gaussian
Canny	1	0.62386
Lap of Gaussian	1.602916	1
Prewitt	3.723412	2.322899

1) the conventional color histogram,

2) the fuzzy color histogram,

3) the color correlogram, and

4) a color/shape-based method and four texture feature extraction techniques such as

1) the steerable pyramid,

2) the contour let transform,

3) the Gabor wavelet transform, and

4) the complex directional filter bank.

Finally, the fuzzy color histogram and the Gabor wavelet transform were shows the highest color and texture retrieval results. In "J.Huang et.al" [34,35]proposed two stages for retrieving an image such as hierarchical clustering and then apply the clustered images to RBFN network.

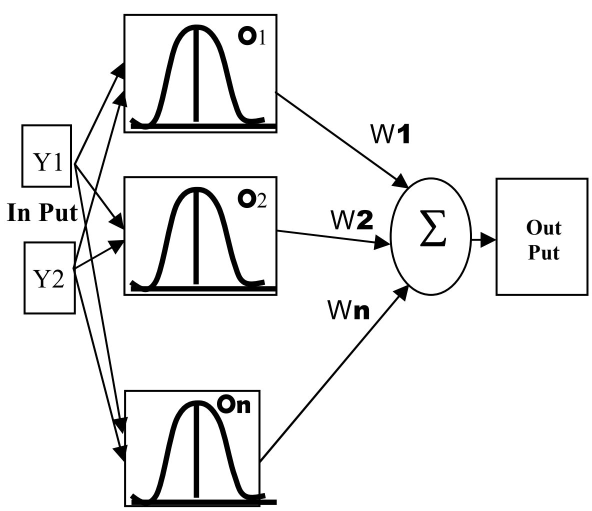


Figure 1 Give caption.

Hierarchical Algorithm is used to group similar images into clusters and RBFN Network which uses K-Means clustering and Gaussian function to retrieve the similar images. So that its get the better favored image results.N Gnaneswara Rao ed al [36,37,38] discussed the texture of an image is computed by using wavelet transformation, because its quite efficiently and also using clustering algorithm, to construct indexed image database based on the texture feature. Finally,clustering is to give the good matching and reduce the undesirable noice. P.AnandhaKumar and V.Balamurugan proposed two indexing technique such as Spatial assess method (SAM) and metric access method (MAM).

SAM providing good result on low dimensional feature. MAM-based balanced and dynamic indexing technique called feature based adaptive tolerance tree. Feature Based Adaptive Tolerance Tree (FATT), which brings effective solution and to increase efficiency of image retrieval. Rajshree S. Dubey et. al[39]discussed four techniques as Color Histogram, Color moment, Texture, edge histogram and it involves pattern recognition. Because it's most important tool for machine vision. Therefore, the combination of four techniques gives better result.

Kondekar V. H. et. al [40] discussed Image color quadratic distance for image histogram, Image Euclidian distance for image wavelet transform; image Hamming Distance for retrieval of an image. From these distance formulae is to increase the retrieval efficiency of an image. Ritu Shrivastava et. al [41,42] discussed to compare two clustering techniques such as K- mean and C-mean clustering for distance metric concept. Finally, K- mean algorithm is easy and fast to compute. C- Mean algorithm takes long computational time. Both converges but suffers from the problem of local minimum. In this survey paper candy method is easy and fast to compute the process. Image splitting and image compaction is to reduce the computation complexity by reducing feature vector size and Haar wavelets are used, since they are more effective compared to other wavelets. In each of the paper they provide several methods in that each method fulfills their works. The results are quite good for most of the query images and it is possible to further improve, to use genetic algorithm, cluster algorithm such as hierarchical clustering, Cure data Clustering, fusion algorithm and any other technique will including in CBIR, it will give the better and effective retrieval of an image.

III. Medical Image Retrieval System

Content-based image retrieval (CBIR) aims at finding images of interest from a large image database using the visual content of the images. Traditional approaches to image retrieval were text-based, where individual images had to be annotated with textual descriptions. Since this is a tedious manual task, image retrieval based on visual content is very essential. In this paper, we focus on the application of clustering to content-based image retrieval. A large collection of images is partitioned into a number of image clusters. Given a query image, the system retrieves all images from the cluster that is closest in content to the query image. We apply the proposed contribution-based clustering algorithm to image retrieval and compare its performance with that of the k-means algorithm.

Content Based Image Retrieval (CBIR) is a technique which uses visual contents, normally called as features, such as shape, color, texture, edge. etc…to search images from large scale image databases according to users' requests in the form of a query image. Content based retrieval of visual data requires a paradigm that differs significantly from both traditional databases and text based image understanding systems. The challenge in CBIR is to develop the methods that will increase the retrieval accuracy and reduce the retrieval time. Among them, Color feature is often broadly used to describe the images which are difficult to be segmented and needn't to consider space information. Texture is one of the most important ones, due to its presence in most real and synthetic world images, which makes it under high attention not only for CBIR but also for many other applications in computer vision, medical imaging, remote sensing, and so on .Finally the edge features that include five categories vertical, horizontal, 45 degree diagonal, 135 degree diagonal, and isotropic are added.

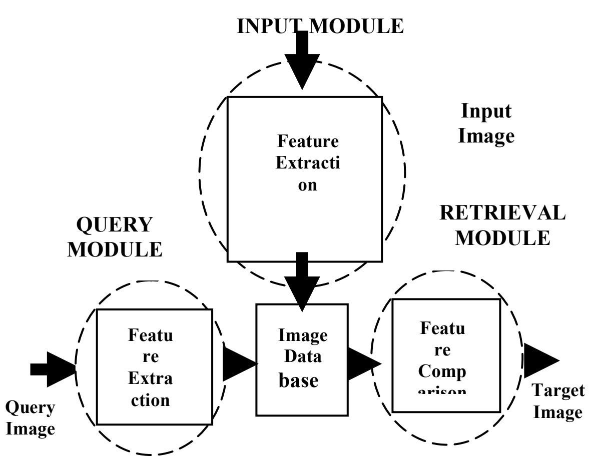


Fig 1: Block diagram of the content-based image retrieval system

In this paper, we focus on the application of clustering to content-based image retrieval. A large collection of images is partitioned into a number of image clusters. Given a query image, the system retrieves all images from the cluster that is closest in content to the query image. The overall system is shown in Fig. 1. We apply the proposed contribution-based clustering algorithm to image retrieval and compare its performance with that of the k-means algorithm.

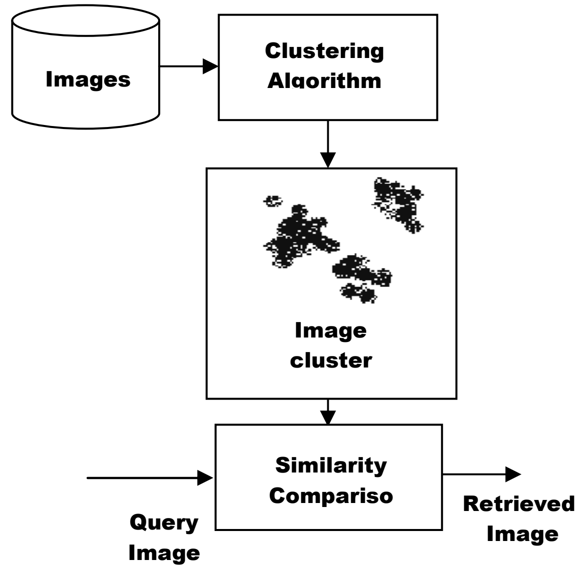


Fig. 2. Content-based Image Retrieval (CBIR) System.

Each image in the database is represented by a visual content descriptor consisting of a set of visual features [8]. A similarity/dissimilarity measure is then used to retrieve images whose features are closest to that of the query image. A common distance/dissimilarity metric is the Euclidean distance, which is used in our work. To represent the visual content of an image, we use a RGB color histogram. The color coordinates of the RGB color space are uniformly quantized into a number of bins. In our work, we use 8 bins each for the Red, Green and Blue coordinates, resulting in 512 bins/features.

IV. Proposed System

The proposed system highlights a novel method to perform contribution based medical image retrieval (MIR) where the term 'contribution' is empirically designed considering clustering of image pixels. The proposed system is based on partitional clustering that aims at partitioning a group of data points into disjoint clusters optimizing a specific criterion [2]. When the number of data points is large, a brute force enumeration of all possible combinations would be computationally expensive. Instead, heuristic methods are applied to find the optimal partitioning. The most popular criterion function used for partitional clustering is the sum of squared error function given by,

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} (x - m_i)^2$$

where k is the number of clusters, C_i is the i[th] cluster, x is a data point and m_i is the centroid of the i[th] cluster. A widely used squared-error based algorithm is the k-means clustering algorithm [2]. In this paper, the model is design using clustering algorithm similar to the k-means algorithm. The contribution is empirically defined as a data point belonging to a cluster as the impact that it has on the quality of the cluster. This metric is then used to obtain an optimal set of '*k*' cluster from the given set of data points. The notion of contribution has its origin in game theory [9]. A recent work by Garg [10] focuses on the merger of game theory and data clustering. Garg mapped cluster formation to coalition formation in cooperative games and used the solution concept of Shapely value to find the optimal number of clusters for a given set of data points. While his work uses the concept of contribution to find the optimal cluster number, we use it in a different manner for optimal partitioning of the data points into a fixed number of clusters. Given a cluster C_i with n points and centroid mi, the average intra-cluster dispersion is given by,

$$dispersion(C_i) = \frac{1}{n}\sum_{x \in C_i}(x - m_i)^2$$

The contribution of a point x , C_i is measured as

Contribution(x, C_i)=Dispersion(Ci-{x})-Dispersion(C_i)

Clearly, if the contribution of a data point is negative, it has an adverse impact on its cluster. On the other hand, a positive contribution indicates that the removal of the point from the cluster would degrade its quality. In our work, we propose a clustering algorithm that treats points with negative contribution different from those with positive contribution. In the case of a negative contribution point, the point is shifted to a cluster, where its contribution is the highest, possibly positive. On the other hand, for a positive contribution point, a multi-objective optimization criterion is taken, where we try to optimize both the intra-cluster and inter-cluster dispersion measures.

The proposed outline presents contribution-based clustering algorithm. It optimizes on two measures, namely the intra-cluster dispersion given by

$$a = \frac{1}{n}\sum_{x \in C_i}(x - m_i)^2$$

and the inter-cluster dispersion given by

$$\beta = \frac{1}{k}\sum_{i=1}^{k}(m_i - \overline{m})^2$$

where k is the number of clusters and $\overline{m}$ is the mean of all centroids. The algorithm tries to minimize α and maximize β. The three steps (passes) in the algorithm are described below.

Algorithm: Contribution based MIR

Input: Query Image

Output: Retrieved Similar Images

Start

1. Initialization
2. Randomly select k centroids ($m_1, m_2, ..., m_k$)
3. **For** each point x
4. Find $1 \le l \le k$ such that distance$_{(x,\, ml)}$ is minimum
5. Add x to cluster C_l and update centroid m_l.
6. **End For**
7. Negative Contribution Points
8. **For** each cluster C_l
9. **For** each point x C_l
10. **If** contribution$(x, C_l) < 0$
11. Move x to a cluster C_p such that contribution
12. (x, C_p) is maximum
13. Update centroid m_p
14. **End If**
15. **End For**
16. End For
17. Positive Contribution Points
18. **For** each cluster C_l
19. **For** each point x □ C_l
20. **If** contribution(x, Cl) ≥ 0
21. Move x to a cluster C_p such that $\frac{a - a_{new}}{a} + \frac{\beta_{new} - \beta}{\beta_{new}} +$ is maximum
22. Update centroid m_p
23. **End if**
24. **End for**
25. **End for**

Stop

Note that α_{new} and β_{new} are values of α and β after the point x is moved to cluster C_p.

V. RESULT AND PERFORMANCE ANALYSIS

The proposed framework for performing contribution based medical image retrieval system is evaluated on 777 medical images belonging to 22 categories of medical conditions pertaining to human body (e.g. spine, palm, MRI, skull, ankle etc) [AR]. Each category contained varying number of images. All the images contained a textual description mentioning the salient foreground objects. The images were clustered using our algorithm with the initial centroids chosen at random. The cluster whose centroid was closest in distance to the given test image was determined and the images belonging to the cluster were retrieved. The results were then compared with images retrieved using the k-means clustering algorithm with the same set of initial centroids. Some of the retrieved images for sample test images are given in Table I. The following performance measures were used to evaluate the performance of the algorithm,

Precision=(Total relevant images Retrieved)/(Total number of retrieved images)

Recall=(Total Number of Retrieved relevant images)/(Total number of relevant images in database)

For the purpose of performance comparison, the contribution based proposed MIR system is also compared with BTC Scheme [AR] and K-Means Algorithm [AR].

Scheme-1: Block Truncation Coding

Color Moment and Block Truncation Coding (BTC) are used to extract features for image dataset when the implementation was performed using BTC scheme. Steps in Block Truncation Coding Algorithm:

1. Split the image into Red, Green, Blue Components

2. Find the average of each component

Average of Red component

Average of Green component

Average of Blue component

3. Split every component image to obtain RH, RL, GH, GL, BH and BL images

RH is obtained by taking only red component of all pixels in the image which are above red average and RL is obtained by taking only red component of all pixels in the image which are below red average. Similarly GH, GL, BH and BL can be obtained.

4. Apply color moments to each splitted component i.e. RH, RL, GH, GL, BH and BL.

5. Apply clustering algorithm to find the clusters.

Scheme-2: K-means using DWT

It basically consists of 3 steps: feature extraction, image segmentation, and deciding similar images. The image has been partitioned into blocks of 4×4 pixels and a feature vector for each block consisting of 6 elements has been extracted. The LUV colour space has been used where L stands for luminance, U for hue and V for saturation, U and V contains colour information (chrominance). The first three of them are average of the values of the Luminance, Hue and Saturation, respectively of the 16 pixels present in the 4×4 blocks. For the other three features Haar (wavelet) transform [AR] has been used to L component of image. After a one-level wavelet transform, a 4×4 block is decomposed into 4 frequency bands of 2×2 block. The other three components of each feature vector are square root of second order moment of wavelet coefficients of the HL, LH and HH band, respectively because each of these bands provide the variations present in different directions. The k-means algorithm has been used to cluster the feature vectors into several classes with every class corresponding to one region in the segmented image. Then the K-means algorithm will do the three steps below until convergence – Iterate until stable (= set of centroids from the previous iteration equals the present set of centroids):

a) Determine the centroid coordinate.

b) Determine the distance of each object to the centroids.

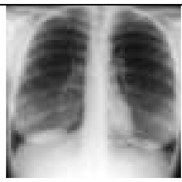
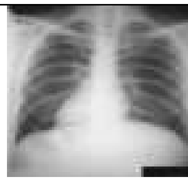
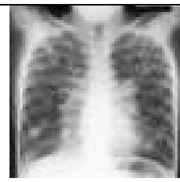
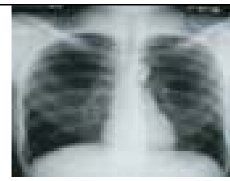
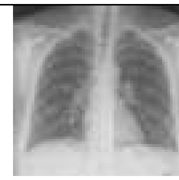
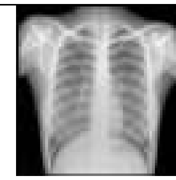
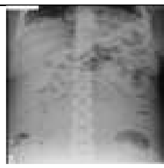
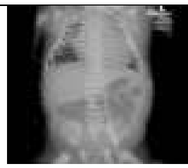
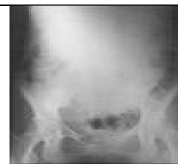
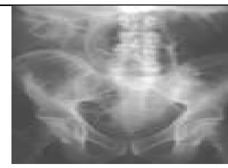
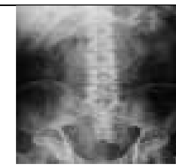
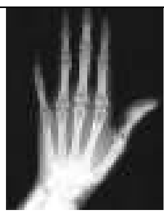
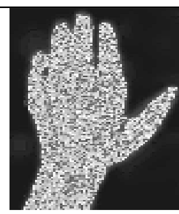
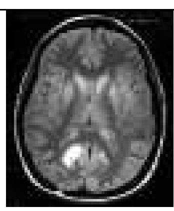
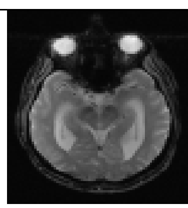
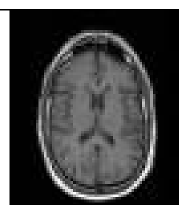
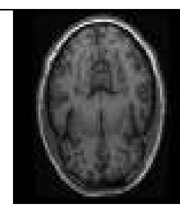
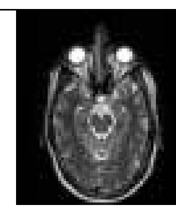
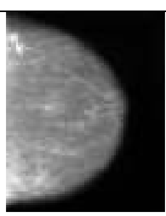
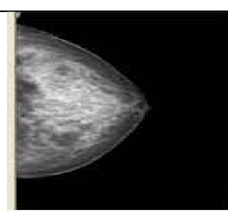
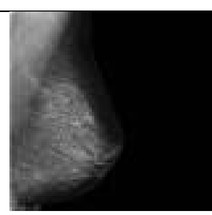
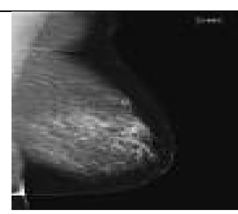
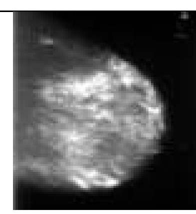
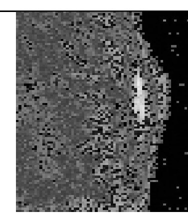
c) Group the object based on minimum distance.

An algorithm has been developed that determines the value of K (number of regions) required for the segmentation of the image, based on the variety of content present in the image.
1) First we start segmentation of feature vectors with K=2.
2) Then we find out the sum of the distances of each feature vector within a cluster from its centroid and then compare them to a experimentally determined critical value.
3) If there is more variety present, which means within a cluster the feature vectors differ considerably, then the value of this sum will be larger than the critical value.
4) If the value of the sum for each cluster present in the image is relatively higher than the critical value then we segment the feature vectors again by incrementing K by 1.
5) Steps 2 to 4 iterate until the sum becomes less than or equal to the critical value.

If two images have region sets $R_1 = \{r_1, r_2\}$ and $R_2=\{r'_1, r'_2, r'_3\}$, first the significance of each region is calculated using the area percentage scheme. Significance of a region depends on the fraction of image occupied by the cluster. We then apply the Most Significant Highest priority [1] algorithm to get the priority of the significance of all the combinations of components from R_1 to R_2. Then we get the final distance between the images by calculating the weighted sum of the components of the distance matrix where the weights are the components of the matrix obtained from Most Significant Highest priority algorithm.

Thus, using these techniques we can match the query image with each image in the database and sort the images present in the database according to the distance in increasing order of distance and hence decreasing order of similarity

Table 1 Results Accomplished

TEST IMAGE	RESULTS OF RETRIEVED IMAGES				
					
					
					
					
					
					

The cumulative results accomplished are shown in Figure 1. After performing the comparative analysis of the proposed system with considered BTC scheme [AR] and K-Means Scheme [AR], it can be seen that proposed system has outperformed as shown in Table 5.1.

Table 5.1 Comparative Analysis of various techniques Adopted

COMPARATIVE ANALYSIS			
SL_No	Techniques	Recall(%)	Precision(%)
1	Contribution	92.86	94.05
2	BTC	92.86	92.82
3	K-means	91.63	91.15

In image retrieval system, the content of an image can be expressed in terms of different features such as color, texture and shape. These low-level features are extracted directly from digital representations of the image and do not necessarily match the human perception of visual semantics. We proposed a framework of unsupervised clustering of images based on the color feature of image. Test has been performed on the feature database of color moments and BTC. K-means clustering algorithm is applied over the extracted dataset. Results are quite acceptable and showing that performance of BTC algorithm is better than color moments. The low recall in both cases is due to the fact that query output consists of images retrieved from a single cluster. Alternatively, the system could output a set of clusters ranked by relevance. This would improve the performance of the system and also, provide the user with a convenient interface for browsing through the retrieved images. It has to be noted that the motivation behind our work is not to improve the performance of existing CBIR systems, but instead to show that the proposed algorithm performs better at image clustering when compared to the popular k-means partitional clustering algorithm.

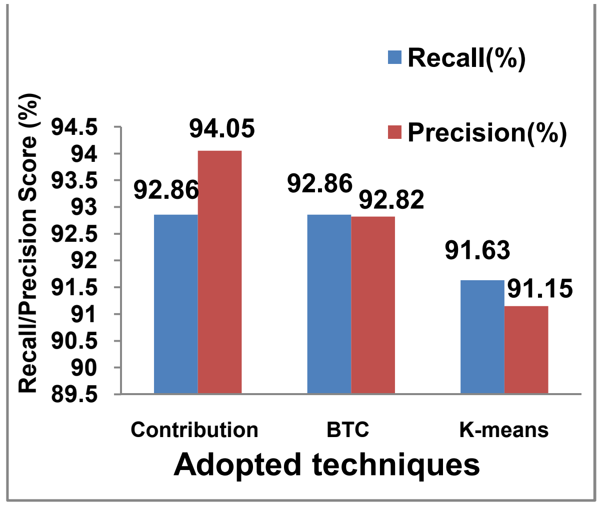


Figure 2. Results of performance Analysis of Proposed MIR System with BTC scheme [AR] and K-Means Scheme [AR]

VI. CONCLUSION

We have thus proposed a new partitional clustering algorithm based on the notion of 'contribution of a data point' Unlike the k-means algorithm, our algorithm optimized on both the intra-cluster and inter-cluster similarity measures and required fewer passes with each pass having the same time complexity as that of the k-means algorithm. We applied the clustering algorithm to content-based image retrieval and our experiments reveal that the algorithm improves on recall at the cost of precision. As with many other clustering algorithms, a limitation with our algorithm is that it requires the number of clusters to be known in prior. Various methods exist to determine the number of clusters for a given dataset [11] including the one based on game theory [10]. A problem with the k-means and k-medoids clustering algorithm is that they do not perform well when the clusters are non-spherical in shape. Whether the proposed algorithm suffers from such a limitation is yet to be investigated. Future lines of work would be to apply the concept of contribution to other clustering techniques such as hierarchical clustering. Our algorithm could also be extended to fuzzy partitioning of data points. The content-based image retrieval technique described in this paper uses only the RGB color histogram as the visual content descriptor of an image. The performance of the system with other visual features based on shape and texture and other distance metrics would have to be tested [8]. Also, learning through relevance feedback from the user could be incorporated in the proposed system [12].

References

[1] J. Han and K. Micheline, "Data mining concepts and techniques," Morgan Kauffman, 2006.

[2] R. Xu and D. Wunsch, "Survey of clustering algorithms," IEEE Transactions on Neural Networks, Vol.16, Issue 3, pp. 645– 678, May

[3] 2005.

[4] E. Forgy, "Cluster analysis of multivariate data: Efficiency vs. interpretability of classifications," Biometrics, vol. 21, pp. 768–780, 1965.

[5] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proc. 5th Berkeley Symp., vol. 1, 1967, pp.281–297.

[6] L. Kaufman and P. Rousseeuw, "Finding groups in data: An introduction

[7] to cluster analysis," Wiley, 1990.

[8] F. Höppner, F. Klawonn, and R. Kruse, "Fuzzy cluster analysis: Methods for classification, data analysis, and image recognition," New York, Wiley, 1999.

[9] P. Bradley, U. Fayyad, and C. Reina, "Scaling clustering algorithms to large databases," in Proc. 4th Int. Conf. Knowledge Discovery and Data Mining (KDD'98), 1998, pp. 9–15.

[10] F.H. Long, H.J. Zhang, and D.D. Feng, "Fundamentals of content-based image retrieval," in D.D. Feng, W.C. Siu, and H.J. Zhang (Eds), 'Multimedia information retrieval and management—technological fundamentals and applications', Springer-Verlag, New York, 2003, pp. 1–26

[11] M.J. Osborne, "An introduction to game theory," Oxford University Press, USA, 2007.

[12] V.K. Garg, "Pragmatic data mining: Novel paradigms for tackling key challenges," Project Report, Computer Science & Automation (CSA), Indian Institute of Science, 2009.

[13] D. Pelleg and A. Moore, "X-means: Extending k-means with efficient estimation of the number of clusters," in Proc. 17th Int. Conf. Machine Learning (ICML'00), 2000, pp. 727–734.

[14] Y. Liu, D. Zhang, G. Lu, and W. Ma, "A survey of content-based image retrieval with high-level semantics," Pattern Recogn., vol. 40(1), 2007, pp. 262-282.

[15] R. Datta, D. Joshi, J. Li, and J.Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," ACM Comput. Surv., 40(2), Apr. 2008, pp. 1-60.

[16] H. Zhang, J.E. Fritts, and S.A. Goldman, "Image segmentation evaluation: A survey of unsupervised methods," Comput. Vis. Image Underst., 110(2), May. 2008, pp. 260-280.

[17] D. Cai, X. He, Z. Li, W. Ma, and J. Wen, "Hierarchical clustering of WWW image search results using visual, textual and link information," in Proc. 12th Annual ACM International Conference on Multimedia (MULTIMEDIA '04), Oct. 2004, New York, NY, pp. 952-959.

[18] D. Kinoshenko, V. Mashtalir and E. Yegorova, "Clustering method for fast content-based image retrieval," Computer Vision and Graphics, 32, Mar. 2006, pp. 946-952.

[19] Y. Chen, J.Z. Wang, and R. Krovetz, "Content-based image retrieval by clustering," in Proc. 5th ACM SIGMM international Workshop on Multimedia information Retrieval, Berkeley, California, Nov. 2003, MIR '03, ACM, New York, NY, pp. 193-200.

[20] P.J. Dutta, D.K. Bhattacharyya, J.K. Kalita and M. Dutta, "Clustering approach to content based image retrieval," in Proc. Conference on Geometric Modeling and Imaging: New Trends (GMAI), 2006, IEEE Computer Society, Washington, DC, pp. 183-188.

[21] G. Liu and B. Lee, "A color-based clustering approach for web image search results," in Proc. 2009 International Conference on Hybrid information Technology (ICHIT '09), Daejeon, Korea, Aug. 2009, vol. 321, ACM, New York, NY, pp. 481-484.

[22] K. Jarrah, S. Krishnan, L. Guan , "Automatic content-based image retrieval using hierarchical clustering algorithms," in Proc. International Joint Conference on Neural Networks (IJCNN '06), Oct. 2006, Vancouver, BC pp. 3532 - 3537.

[23] "Object and concept recognition for content-based image retrieval,"[Online].Available:http://www.cs.washington.edu/research/ imagedatabase/groundtruth/.

Dr.B.P Mallikarjunaswamy: working as a professor in the Department of Computer Science & Engineering, Sri Siddhartha Institute of Technology, affiliated to Sri Siddhartha University. He has more than 20 years of Experience in teaching and 5 years of R & D. He is guiding many Ph.D scholars. He has published more than 30 technical papers in national and International Journals and conferences. His current research interests are in Pattern Recognition and Image Processing.

Mr.Shivamurthy R C received the BE degree from college of Engineering, Gulbarga University and received the M.Tech degree in Computer Science & Engineering from Malnad College of Engineering, Visvesvaraya Technological University, Belgaum. He served as a Bio Medical Engineer in AIMS Hospital & Cancer Research Center. He served as Assistant Professor and currently working as professor in the department of Computer Science at A.I.T, Tumkur, Karnataka and is also a Ph.D scholar in CMJ University, India.