PERFORMANCE EVALUATION OF CLUSTERING BASED CONTENT BASED IMAGE RETRIEVAL SYSTEMS

1K.SRINIVASA RAO 2Prof. I.Ramesh babu

1Associate Professor, Dept. of CSE, GIT, GITAM UNIVERSITY, Visakhapatnam - 530 045, 2 Professor, Dept. of CSE, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur(Dt)AP

Abstract- Retrieving images from available databases attracted researchers due to its ability to retrieve images in accurate and reliable way. Content based image retrieval scheme (CBIR) has lot of advantages over conventional META data based image retrieval scheme. The computer vision based application has huge importance in the field of digital image processing and content based image retrieval scheme (CBIR) yields good results based on three contents namely color, shape and texture. The color based content based image retrieval (CBIR) scheme is simple and efficient. Although tremendous progress has been made in the past years on image retrieval schemes but still acquiring images from large data sets based on various traditional image retrieval schemes is concerned area in the image processing domain. The proposed work presents content based image retrieval scheme (CBIR) based on three algorithms namely (a) Content based image retrieval scheme (CBIR) based on histogram properties (b) Content based image retrieval scheme (CBIR) based on histogram properties with K-Means clustering (c) Content based image retrieval scheme (CBIR) based on histogram properties with pillar K-means clustering. Finally the simulation results shows better accuracy and reliability in retrieving images in digital computers and the performance of the proposed system overcomes the disadvantages of the traditional systems.

KEYWORDS: Content based image retrieval scheme (CBIR), K-Means clustering, pillar –Means clustering, histogram properties

1. INTRODUCTION

Content based image retrieval scheme (CBIR) approaches reported in the literature are implemented mostly for the image processing and computer vision. The retrieved images have its own characteristics when it is retrieved from the collection of images with the significant size. The main applications of the image retrieval scheme are the digital image processing and computer vision and in that image processing covers compression, enhancement, transmission and interpretation. Conventional feature analysis mechanism suffers from accuracy and reliability while content based image retrieval scheme yields output fairly in terms of accuracy and performance oriented.

Accurate detection of faces in applications of crime, security has improved and automatic face recognition has helped in many applications. Among all available approaches content based image retrieval scheme (CBIR) is most successful in real time scenario. The content based image retrieval scheme (CBIR) automatically derived features based on three contents namely color, shape and texture and retrieved features derivation can be elaborated in semantic and as well as primitive. CBIR defers from the classic information retrieval in that the image data bases are essentially unstructured, since digitized images consists purely of arrays of pixel intensities, with no inherent meaning. One of the key issues with any kind of image processing is the need to extract useful information from the raw data( such as recognizing the particular shapes or textures) before any kind of reasoning about the image contents is possible.

2. BACKGROUND

1) A content based image retrieval scheme (CBIR) scheme based on color and gradient direction features is proposed by JIANLIN ZHANG, RAGHAVAN, V.V. in the year 2010 [1]. The work proposed in this paper initially divides the input image blocks then each block characteristics is extracted based on the color and edge direction features. Then usage of the clustering algorithm is observed to preserve all the extracted color information in the form of code book. Finally color code indexes are used to retrieve the image based on content based image retrieval scheme (CBIR)

2) In the year 1995 initial representation of the content based image retrieval scheme
K.SRINIVASA RAO and Prof.I.Ramesh babu

3. PROPOSED METHODS

This paper mainly consists of three algorithms; the performance of each algorithm is evaluated using standard image retrieval database.

(A) HISTOGRAM BASED CBIR

In this algorithm the images are retrieved by using the histogram properties of it. As stated in [4] the colour histogram features of an image can be denoted as below equations.

Without loss of generality the definition of histogram is

\[ P(g) = \frac{N(g)}{M} \]

Where \( P(g) \) is the probability of occurrence of gray level while \( N(g) \) is its intensity value and \( M \) is the total number of pixels. The following are the properties that used for CBIR

(i) Mean

\[ Mn = \sum_{g=0}^{L-1} g P(g) \]

Where \( Mn \) is the mean of gray level image \([0 \ L-1]\) \(([0 \ 255])\)

(ii) Standard Deviation

\[ \sigma_g = \sqrt{\sum_{g=0}^{L-1} (g - Mn)^2 P(g)} \]

(iii) Skewness

\[ Sk = \frac{1}{\sigma_g^3} \sum_{g=0}^{L-1} (g - Mn)^3 P(g) \]

The skew will be positive if the tail of the histogram spreads to the right (positive) and negative if the tail of the histogram spreads to the left (negative).

(iv) Energy

\[ E = \sum_{g=0}^{L-1} (P(g))^2 \]

(v) Entropy:

\[ En = -\sum_{g=0}^{L-1} P(g) \log(p(g)) \]

All these properties constitute the features for the images and similarity is measured using Euclidean or chessboard distance transforms. The experimental results obtained using this approach are shown in section IV

(B) K MEANS CLUSTERING BASED CBIR

The K-means [5] is one of the simplest unsupervised learning algorithms. The procedure...
follows an easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centroids, one for each cluster. The K-Means method is numerical, nondeterministic and iterative.

Step 1: Input the Image and perform Hierarchical clustering.
Step 2: Consider the Every point as its own cluster.
Step 3: Find Most Similar Pairs of Clusters.
Step 4: Merge those two points to one parent cluster.
Step 5: Repeat Step 3 to Step 5 until all points are merged into one cluster.
Step 6: Apply K-means clustering to the required image set obtained from Hierarchical clustering.
Step 7: Enter How Many Clusters (Let “k”).
Step 8: Randomly Guess K Cluster center Locations.
Step 9: Each Data point finds out which center it’s closest to.
Step 10: Thus Each Center “Owns” Set of Points.
Step 11: Each Center Finds the Centroid of its Own Points.
Step 12: Center now moves to the New Centroid.
Step 13: Repeat Step 9 to Step 12 until Terminate

Histogram is quantized to n-bins (8 or 16) for the clustered image and the respective properties form a feature vector for retrieval. The experimental results are shown in section IV

**IV RESULTS AND DISCUSSIONS**

The proposed algorithms are tested and performance is evaluated using Berkeley Image retrieval data base. [7] Which contains multiple colour natural images for retrieval and segmentation purpose. The results of each individual algorithm is presented below

(C) PILLAR K MEANS BASED CBIR

Because of initial starting points generated randomly, using K-means algorithm it is difficult to reach global optimum which will lead to incorrect clustering results. These obstacles in K-means have been addressed by specifying a procedure to initialize the cluster centers before proceeding with the standard k-means optimization iterations.[6]

Input: Dataset, D = {d1, d2, .......dn} //set of n data points. k //set of desired clusters
Output: A set of k clusters.
Step1: One centroid is uniformly chosen at random from among the data points.
Step2: For each data point z, the distance D(z), between z and the nearest chosen centroid is computed. Step3: One new data point is chosen at random as a new center, using a weighted probability distribution where a point z is chosen with probability proportional to D(z)^2.
Step4: The 2 and 3 steps are repeated until k centers have been chosen.
Step5: Now that the initial centers have been chosen, one can proceed using standard k-means clustering.

Histogram is quantized to n-bins (8 or 16) for the clustered image and the respective properties form a feature vector for retrieval. The experimental results are shown in section IV
Recall Rate = Number of relevant images retrieved/ Total number images in database
Precision Rate= Number of relevant images retrieved/ Total number relevant images in database

To evaluate the algorithm performance metrics like precision and recall are used [8,9]

REFERENCES


[7] https://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/

