Journal of Computer Applications ISSN: 0974 – 1925,
Volume-5, Issue EICA2012-1, February 10, 2012

BTC with K Means Classifier Using Color Image Clustering

Samathal S
PG Scholar,
St.Peters University, Chennai,
E-mail: samathalcse@gmail.com

Mohanraj N
M.Tech.,
St.Peters University, Chennai

Abstract - We can not retrieves the images from the image database, based on color, shape, texture or any other information accurately, to over come this problem Content based image retrieval (CBIR) technique proposed. It can be derived from the image itself. CBIR addresses the main problem is that retrieving images relevant to the user needs from image database on the basis of owl eve visual features. Grouping images in to meaningful categories to reveal the useful information is a challenging and important problem. Clustering is a data mining concept to group a set of unsupervised data based on the conceptual clustering principal: Maximizing the intra class and minimizing the interclass similarity .Color moment and Block Difference of Inverse Probability (BDIP) and Block Variation of Local Correlation Coefficient (BVLC) are the two methods used to extract features for image dataset. K means clustering algorithm is used to conducted to group the image dataset for different clusters. For various types of input classes precision and recall rate have been calculated

Keywords: *Image feature, Clustering, Color moments, Precision and recall.*

I. INTRODUCTION

The rapid increase in the digital images for multimedia system is used for hiding this data collection that is potentially used in the wide range of applications like Geographical Information System (GIS), remote sensing, and Crime Prevention, Military, Home Prevention and World Wide Web (WWW) .Analyzing the vast volume of image data is becoming increasingly difficult. The raw image cannot be directly used for retrieval from the image database. Raw image data need to be processed and descriptions based on the properties that are inherent in the images themselves are generated. The images stored in the feature database are used for retrieving and grouping.

Clustering is a method of grouping data objects into different groups. The similar data objects are belongs to the same group and dissimilar data objects are belongs to the different groups. Image clustering consists of two steps are feature extraction and grouping. A feature vector capturing certain essential properties of the image is computed and stored and computed in feature base for each image in a database. To form the group clustering algorithm is applied to form the group.

II. EXISTING SYSTEM

Interacting with images and performs extraction of meaningful information of images with descriptions based on properties that are inherent in the images itself it is the process of feature extraction. In image retrieval color information is most intensively used feature because of its strong correlation with the underlying image objects. Color Histogram is the commonly and very popular color feature used in many image retrieval system for representing the distribution of color in an image. It is widely used in RGB and HSV. The mathematical foundation and color distribution of images can be categorized by color moments. Color Coherence Vector (CCV) have been proposed to incorporate spatial information in to color histogram representation. It is used to store the coherent versus incoherent pixels. Color Histogram is consisting of lack of spatial information and high sensitivity to noise interference the lightening intensity changes and the quantization occurs.

Texture refers to the presence of spatial pattern that has some properties of homogeneity. Textures are replications, symmetries and combinations of various basic patterns or local functions, usually with some random variation. There are a number of texture features which have been used frequently liked Tamura Texture feature, Simultaneous Auto Regressive (SAR) models, Gabor texture features and wavelet transform features. Classifying image by content is an important way to mine valuable information from large image collection Based on low level visual features the image can be grouped in to meaningful categories. The concept of fuzzy Iterative dichomiser 3 (ID3) is used to generate decision tree is discussed in image retrieval. ID3 is a decision tree based on Shannon's information theory.

A sample data set is described by a set of attributes and an outcome, ID3 produces a decision tree, which can classify the outcome value based on the given attributes like Color, Texture and Spatial Location. Image dataset were defined in 10 classes (concepts): grass, forest, sky, sea, sand, sunset, firework, flower, tiger and fur. The attribute with smallest entropy is selected from those attributes not yet used as the most significant for decision making at each level of ID3 decision tree.

The proposed work of this paper is to cluster the images based on color feature by using data mining approach. Concept of color moment is extended to obtain the features and k_means algorithm is applied to cluster the images. The previous work related to image retrieval and mining. The proposed work introduces the concept of 3.1 Color moments, 3.2 Block Truncation Coding (BTC) Algorithm, 3.3 K-means clustering algorithm. Large images are partitioned into a number of smaller image tiles. Those individual image tiles are used to extract the feature images.

III. PROPOSED SYSTEM

An image can be spatial representation of an object and it is represented by a matrix of intensity value. The sample points are represented as pixels which can be differentiate by color intensity. The three layer basic color image can be described with each layer as a Red, Green, and Blue as shown in Figure 1.

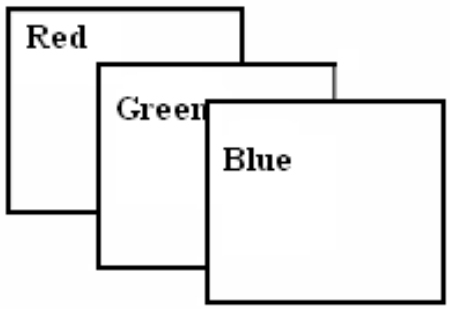

Figure 1: Image Components

A. Color Feature Extraction

Color moments are measures that can be used to differentiate the images based on their color feature. The basis of color moments lays in the assumption that the distribution of color in an image can be interpreted as a probability distribution. A number of unique moments can be characterized by probability distribution consider the mean and variance can be differentiate by using Normal distribution. If the color in an image follows a certain probability distribution, that moments of distribution can be used to identify that image based on the features on color. Stricker and Orengo use three central moments of an image's color component of the ij-image pixel and P is the height of image, and Q is the width of the image. They are Mean, Standard deviation and Skewness.

Mean

Mean can be understood as the average color value in the image.

$$E_k = \frac{1}{PQ} \sum_{i=1}^{P} \sum_{j=1}^{Q} p^k ij$$

Standard Deviation

The standard deviation is the square root of the variance of the distribution.

$$SD_k = SQRT\left(\frac{1}{PQ} \sum_{i=1}^{P} \sum_{j=1}^{Q} (p^k ij - E_k)^2\right)$$

Skewness

Skewness can be understood as a measure of the degree of asymmetry in the distribution.

$$S_k = \left(\frac{1}{PQ} \sum_{i=1}^{P} \sum_{j=1}^{Q} (p^k ij - E_k)^3\right)^{1/3}$$

B. Block Truncation Coding Algorithm

BTC is a type of lossy image compression technique for grey scale images. It divides the image in to blocks.

Steps in BTC Algorithm:

1. Split the images in to 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 each component in to RH, RL, GH, GL, BH, 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, BL can be obtained.
4. RH, RL, GH, GL, BH, BL apply color moments to each splitted component.
5. To find the clusters apply clustering algorithm.

K means Clustering Algorithm

K – means is one of the simplest unsupervised learning algorithm in which each point is assigned to only one particular cluster. The procedure follows a simple, easy and iterative way to classify a given data set through a certain number of clusters fixed a priori. The procedure consists of the following steps,

Step 1: Set the number of cluster k
Step 2: Determine the centroid coordinate
Step 3: Determine the distance of each object to the centroids
Step 4: Group the object based on minimum distance
Step 5: Continue from step 2, until convergence that is no object move from one group to another.

IV. EXPERIMENTS

An image database consists of 1000 images including 10 classes to perform in proposed scheme. Each class has 100 images. Each image is of size 384*256 pixels. Unichrome feature values are extracted from a single color layer of Red, Green, and Blue.

Original Image (Bus Image)

Red Green Blue

Figure 2: RGB Components of Sample Image

The first part of evaluation computes color moments for each of the three color components. Each color component yields a feature vector of three elements as discussed in section 3.1 i.e. mean, standard deviation and skewness. Thus total nine feature vectors are calculated for one image.

In the second part apply Block Truncation Coding Algorithm discussed in section 3.2 over RH, RL, GH, GL, BH and BL. Thus total 18 feature vectors are calculated for one image.

Table 1: Recall and Precision using Color Moments

Classes	Recall	Precision
African People and Villages	30 25.43	25.43
Beaches	47	40.17
Buildings	33	23.57
Buses	37	35.24
Dinosaurs	100	92.59
Elephants	32	35.56
Flowers	56	90.32
Horses	50	54.95
Mountains and glaciers	23	34.33
Food	21	20.79

Figure 3: Sample Dinosaurs Clusters using Color Moments

Two statistical measures were computed to assess system performance namely Recall and Precision. Recall consists of the proportion of target images that have been retrieved among all the relevant images in the database.

$$\text{Recall} = \frac{\text{Number of Relevant Images Retrieved}}{\text{Total Number of Relevant Images}}$$

$$\text{Precision} = \frac{\text{Number of Relevant Images Retrieved}}{\text{Total Retrieved Images}}$$

Table 2: Recall and Precision using BTC

Classes	Recall	Precision
African People and Villages	44	33.58
Beaches	42	42.42
Buildings	8	7.92
Buses	52	44.83
Dinosaurs	99	97.06
Elephants	39	44.83
Flowers	80	94.12
Horses	50	58.82
Mountains and glaciers	35	43.21
Food	25	22.32

Figure 3 & 4 showing the sample images in Dinosaurs and Flowers are clustered using color moments and BTC algorithm.Table1 & 2 shows the values of recall and precision of each class using color moments and BTC algorithms. For maximum Figure 4 Sample Flowers Clusters using BTC Classes recall and precision of BTC is better than color moment feature extraction.

V. CONCLUSION

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 images. Proposed a framework of unsupervised clustering of images based on the color feature of images. 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 quiet acceptable and showing that performance of BTC algorithm is better than color moments.

REFERENCES

1. S.Silakari Motwani and M Maheswari (2009), “Color Image Clustering using Block Truncation Algorithm” IJCSI(International Journal of Computer Science),Vol. 4, Issue no. 2, pp.31-35.
2. Ying Liu1, Dengsheng Zhang1,Guojun Lu1 , Wei-Ying Ma2(2005), “Deriving High-Level Concepts Using Fuzzy-Id3 Decision Tree for Image Retrieval”, IEEE , pp 501-504.
3. Gholamhosein Sheikholeslami, Wendy Chang, Aidong Zhang(2002), “SemQuery: Semantic Clustering and Querying on Heterogeneous Features for Visual Data”, IEEE Trans. Knowl. Data Eng. 14(5): pp.988-1002
4. Alex Pentland, Rosalind W. Picard, Stan Sclaroff(1999), “Photobook: Tools for Content-Based Manipulation of Image Databases”, Storage and Retrieval for Image and Video Databases (SPIE): pp.34-47
5. Krzysztof Koperski, Giovanni Marchisio, Selim Aksoy et al.(2002), "Applications of Terrain and Sensor Data Fusion in Image Mining", IEEE, pp 1026-1028