IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 3, March 2015.
www.ijiset.com
ISSN 2348 – 7968

An Effective Image Annotation using Content Based Image Clustering

Ms.Jeenatunnisa.J.H[1], Ms.Ranjani.R[2]

[1]UG Student, Computer science and Engineering, IFET College of Engineering, Villupuram

[2]Assistant Professor, Computer Science and Engineering, IFET College of Engineering, Villupuram

Abstract

Clustering is the form of unsupervised classification which aims for grouping the datasets based on their similarities and its characteristics. The advanced growing technology in image capturing devices leads to large collection of images in the web. CBIR systems are an active research area developed for effective image retrieval from the large collection of images. Clustering algorithms optimizes the image search in the large databases and improves the efficiency of image retrieval system. For each clustering technique the retrieval accuracy is computed by comparing the results of retrieval without clustering against results of retrieval with clustering.

Keywords—CBIR, image clustering, retrieval accuracy, Image Annotation,Precision and Recall.

1. INTRODUCTION

Images are convenient medium for representing the information in the field of multimedia and object identification. Every day in the development of computer technology the more number of images are produced in the web via social networks. So the images are properly organized for fast and better retrieval. Image retrieval is the process of searching and retrieving of images from the large collection of images. In present scenario, the advancement in image capturing devices the collection of images in the web are growing larger and become more diverse content. From the diverse content of the images locating an exact image it is possible for small set of images but for a large collection of images is a challenging problem. So the clustering concept is used to retrieve the images which minimizes the searching time of images. Clustering is the special type of classification, which retrieves the images based on similarity features of images. Image search engines are becomes very popular due to increase in various images.

Content-based image retrieval (CBIR) is a research area which provides an effective result to search and retrieve the images from the huge collection of images. CBIR search engines have been developed for efficient image retrieval from the collection of images in the database. The main aim of CBIR systems are addresses the problem of retrieving semantically

relevant images and reducing the searching time of images. Instead of text based image retrieval, the Content based image retrieval method retrieves the images based on extracted low level visual features to ensure the query given as the input by using the color, shape and textures. It matches the similarities between the query image and images in the database. The content based image retrieval method is achieved by two step process. In the first step the feature extraction is done based on the color, shape and textures to extent the image content. In the second step the extracted features are matched with the database images to yield the results to the user. Then the similar images are clustered to reduce searching time of the images and improve accuracy of retrieval. Clustering the images are based on the principle of minimizing the intra class similarity and maximizing the inter class similarity.

2.RELATED WORK

Earlier searching techniques are not mainly based on the visual features of images but it based the text based approach. It is difficult to annotate the images in the large database. It is a challenging problem to find the exact results. The Daubechies4 wavelet transform [1] is used to extract the features. Based on these extracted features image classification has been done using Fuzzy K means algorithm. Then the images are organized and stored. Euclidean distance [2] method used for similarity measurement and to retrieve the similar images from the database. Due to the progressive increase in the database images the scalability of the image database increases, so the performance of the retrieval process decreases.

The graph theoretic approach [3] which adapted in the image retrieval by post-processing step to find the image similarity. In this method initially the feature extraction is done for the images in the image database, and then the clustering algorithm is applied to analyze images to form an image clustering database. Then the query image features are extracted it is compared with the image clustering database and matching results are retrieved. Subsequently the image features are another problem, so determining the image features are significant to find the similarities. Test based image retrieval is not enough to accuracy, hence the direction moves towards the CBIR research. Re-Ranking [4] of images improves the performance to find the effective results. Using color histograms the features are extracted and the images are plotted in the feature space. The distance calculation is done for similarity measures. Then Re-Ranking of images are performed by using the K means algorithm and hierarchical algorithm. K means algorithm takes long time to perform

ISSN 2348 – 7968

clustering, so the features like textures and edges are used for Re-Ranking. Accordingly the different Re-Ranking system is used to improve the efficiency. Hierarchical clustering approach [5] which forms image clusters in the image database based on the color features. The centroids of these clusters are calculated using K means clustering. The query image features such as color, textures, shape and spatial feature are extracted also the clusters are extracted in the image database. The features are compared with the clusters center images which are having more similarity to the query image. The relevance feedback analysis performed on the retrieved images based on the user's feedback. Some object recognition techniques are included to overcome the semantic gap.

3.PROPOSED WORK

CBIR system concerns that indexing and retrieving the images based on their content. A content based image retrieval system considers some similarity measures between the query images and the target images. It ranks the images based on the highest similarity to the query image and discards the similarities between the target images. Clustering algorithms minimizes the searching time of images. Image clustering comprises two step processes. The first step is feature extraction and second step is forming groups. In the proposed system the RGB color model is used as the feature extraction. Initially the image features are extracted then the clustering algorithm is applied to analyze the similarities of images to form the image clustering database. Then the images are clustered based on the accuracy as a similarity measures, and resultant images are returned to the users. Then the annotations of the clusters are performed. The retrieval accuracy of clustering are compared with the resultant images with clustering and the resultant images without clustering. The top ranked images are clustered to improve the retrieval accuracy.

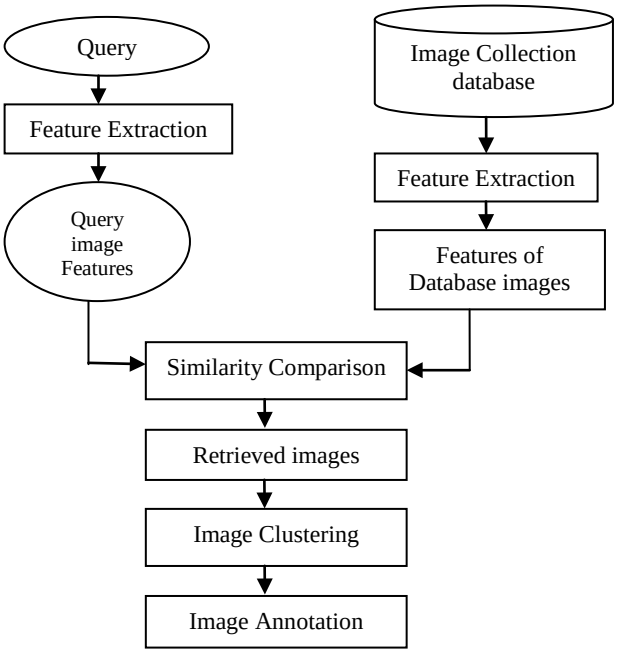

Fig.1. Proposed System Architecture

Precision and recall values are used as comparative analysis for image clustering. The Proposed System of CBIR concerns the following steps:

- Initially the input is given as query image, and its features are extracted.
- The features of the image database are extracted and stored in the feature database.
- The comparative analysis based on the similarity measures is done to find the similar images.
- The clustering is performed to form the image clusters based on the accuracy values.
- Then the annotation is performed on the image clusters and stored.

3.1 Feature Extraction

Feature Extraction is the ground work for the content based image retrieval. It identifies the common signatures, termed as feature vector for each image based on the pixel values. The images are formed using the pixels. Visual features such as color and shape optimal values are extracted. The input query image is resized and the feature extraction is done using the RGB color model. The RGB pixels of the images are calculated and stored in the feature database. When the input query image is given the RGB values are calculated. The average values of Red, Green, and Blue are calculates as follows:

$$\text{Red average}=\frac{\text{No. of Red pixels in the image R (I)}}{\text{No. of pixels in the image I}}$$

$$\text{Green average}=\frac{\text{No. of Green pixels in the image G (I)}}{\text{No. of pixels in the image I}}$$

$$\text{Blue average}=\frac{\text{No. of Blue pixels in the image B (I)}}{\text{No. of pixels in the image I}}$$

Where R (I) =RED pixels in the image

G (I) =Green pixels in the image

B (I) =Blue pixels in the image

I =No. of pixels in the image

The features of the image represent the characteristics of the images. The CBIR system concentrating on the image features to enable the query to retrieve the similar images. RGB values of the database images also calculated and stored in the feature database. The average values of red, green, and blue are compared with each other to find the maximum values of the image. Based on the intensity values the images are clustered. The shape optimal value is found out to identify the images.

3.2 Similarity Comparisons

The given query image is preprocessed and features are extracted. When the average values of red, green, and blue are calculated the query image values are compared with the stored values. The image feature which is closest to the query image will be returned to the user. The similarity measurement

ISSN 2348 – 7968

is based on the color, and shape. The query image is compared with the each image in the database. At the end these step is performed one after another to retrieve the similar images from the database. The relevant images are displayed.

3.3 Content based image clustering

Clustering minimizes the searching time of the images. The query image is compared with the database images and relevant images are returned to the users. Then the top ranked images are clustered based on the accuracy values of the images. The performance of the clustering is measured in terms of precision and recall values. The content based image clustering clusters the images based on the image content. The given query image is preprocessed and features are extracted. The clusters are formed based on the similarity to the query image.

a) Precision

Precision is defined as the relationship between the numbers of relevant images retrieved to the total number of images retrieved.

$$\text{Precision} = \frac{\text{No. of relevant images retrieved}}{\text{Total No. of images retrieved}}$$

b) Recall

Recall is defined as the relationship between the numbers of relevant images to the total number of images in the database.

$$\text{Recall} = \frac{\text{No. of relevant images retrieved}}{\text{Total no. of images in the database}}$$

c) Accuracy

Accuracy is defined as the average of the precision and the recall values.

$$\text{Accuracy} = \frac{\text{Precision} + \text{Recall}}{2}$$

Precision measures the ability of the system to retrieve only the images that are relevant. Recall measures the ability of the system to retrieve all the images that are relevant. For retrieval efficiency the precision and recall parameters are calculated from the image database. Based on the accuracy values the images are clustered. Based on the fixed threshold value for the accuracy the image clusters are formed. The image clusters are formed mainly with respect to the similarity of the query image rather than the database images. In this method the relations among the retrieved images are taken into the consideration for clustering, and it may also provide the additional information for ranking the images. The content based image clustering returns the image clusters that are closely adopted to the search query image. The image clusters are retrieved for the query image instead of set of ordered images which reduces the searching time of images and improves the retrieval process.

3.4 Image Annotation

Currently the image annotation plays a vital role in image retrieval. In clustering phase the image clusters are returned that are most relevant to the search query image. Then the annotation of the image clusters are performed and stored in the database. The image annotation greatly achieves the high performance without user interaction. The image annotation for the image clusters used for fast retrieval of images and progressively improve the efficiency.

4. Experimental Results

The Proposed system uses the Wang database [6] for evaluation. It has 10 categories of images totally 1000 images. Each image in the database are colored image and of size 384×256 pixels. To measure the retrieval effectiveness the standard parameters precision, and recall values are used.

TABLE I. PRECISION AND RECALL VALUES OF RGB MODEL

SI .No	Evaluation Measures		
	Categories	*Precision*	*Recall*
1	Flowers	0.85	0.55
2	Beach	0.7	0.45
3	Dinosurs	0.90	0.65
4	Buses	0.50	0.35
5	Mountains	0.75	0.40
6	Buildings	0.34	0.22
7	Horses	0.54	0.45

Here Table 1. Shows the precision and recall values of the image categories that are retrieved using the RGB model. Here only the five categories are considered for evaluation.

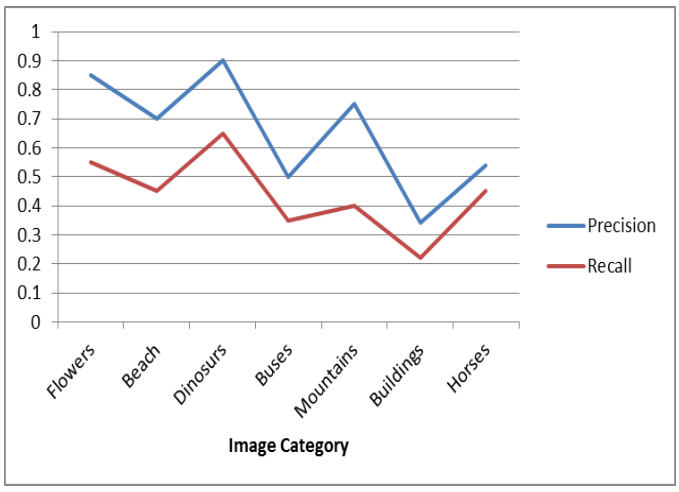

Fig.2. Precision-Recall Graph

From the Figure 2. Graph is plotted using the values precision and recall values for each image category. It is analyzed that precision value increases then the recall value decreases for each image category. Because the precision

ISSN 2348 – 7968

values depends only on the relevant images retrieved for the given query image, but the recall values depends on the database images compared with the query image. So the number of comparisons increases in similarity comparisons.

Then the accuracy values of the proposed system compared with the K means algorithm. The accuracy values are plotted for each image category.

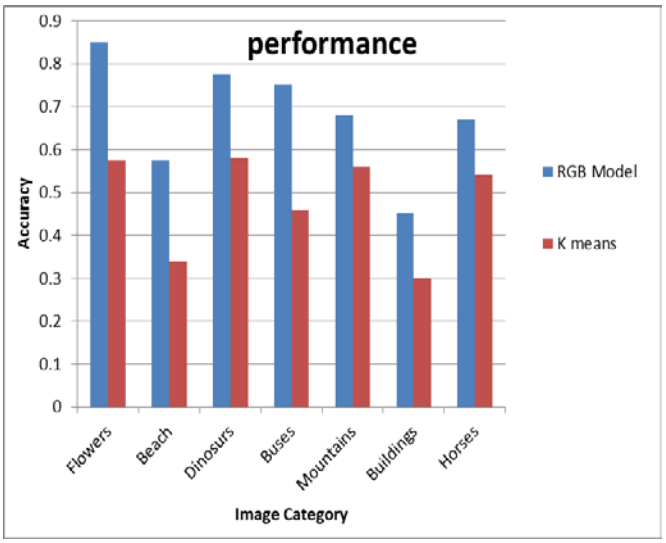

Fig 3 Comparisons of Accuracy values with K means algorithm

From the Figure 3 graph it is noticed that the graph is plotted based on the accuracy values different categories of images comparing K means and proposed algorithm. The proposed RGB model gives the efficient results rather than the K means algorithm. It takes long time for searching images. Based on the color moments and shape based results retrieves good results.

5.CONCLUSION

Now a days, the sharing of images increases a lot of time for searching of images. Hence the images in the database increase every day. This paper focus on improvement in image search using the clustered image annotation. Content based image clustering is used to cluster the similar images with respect to the characteristics of the query image. By the use of clustering the total time needed to search the images was reduced. Content based clustering technique provides dynamic and local visualization of database images. The image clusters are annotated to improve the effective search results to the users. This method progressively increases the performance and produces the needed results to the user.

In the future by using some object recognition technique the semantic gap can be resolved for image retrieval system.

References

[1] E. Saravana Kumar,Dr. A. Sumathi, K. Latha “Performance and Accuracy related issues of Content-Based Image Retrieval” International Journal of Power Control Signal and Computation(IJPCSC) Vol3. No1. Jan-Mar 2012 .

[2] Wei Bian and Dacheng Tao,”Biased Discriminant Euclidean Embedding for Content Based Image Retrieval” IEEE Transactions on Image Processing, vol. 19, no. 2, Feb. 2010.

[3] Hong Liu,and Xiaohong Yu,“Application Research of k-means Clustering Algorithm in Image Retrieval System”Proceedings of the Second Symposium International Computer Science and Computational Technology(ISCSCT ’09) *26-28,Dec. 2009, pp. 274-277*

[4] Mayuri D. Joshi , Revati M. Deshmukh, Kalashree N.Hemke, Ashwini Bhake “Image Retrieval using Re-Ranking Algorithms” International Journal Of Core Engineering & Management(IJCEM) Volume 1, Issue 1, April 2014

[5] Priyanka Gupta, Umesh Kumar, Ankur Chawla “Impact of clustering in image retrieval techniques ” International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 4, April 2013

[6] Wang, J.Z., “Wang Database,” (online). Available at: http://wang.ist.psu.edu/.

[7] Dayong Wang, Steven C.H. Hoi, Member, IEEE, Ying He, and Jianke Z “Mining Weakly Labeled Web Facial Images for Search-Based Face Annotation” IEEE transactions on knowledge and data engineering, vol. 26, no. 1, January 2014.

[8] K. Haridas, Dr. Antony SelvadossThanamani “Efficient Content Based Image Retrieval System in Visual Words, Color and Edge Directive Descriptors and Fuzzy Color and Texture Histogram” International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 10, October 2014

[9] Mrs. Mallika, Mr. S. Anantha Krishnan” An Efficient Content Based Image Retrieval Using HierarchicalClustering Approach” vol.2 Issue.9,pg:119-126.

[10] Dr. Sanjay Silakari , Dr. Mahesh Motwani and Manish Maheshwari “Color Image Clustering using Block Truncation Algorithm” IJCSI International Journal of Computer Science Issues, Vol. 4, No. 2, 2009

[11] Simardeep Kaur and Dr. Vijay Kumar Banga “Content Based Image Retrieval: Survey and Comparison between RGB and HSV model ” International Journal of Engineering Trends and Technology (IJETT) - Volume4Issue4- April 2013

[12] A.Kannan Dr.V.Mohan Dr.N.Anbazhagan “Image Clustering and Retrieval using Image Mining Techniques” 2010 IEEE International Conference on Computational Intelligence and Computing Research

[13] Mrs Monika Jain, Dr. S.K.Singh “A Survey On: Content Based Image Retrieval Systems Using Clustering Techniques For Large Data sets” International Journal of Managing Information Technology (IJMIT) Vol.3, No.4, November 2011

[14] Mahip M.Bartere, Prashant R.Deshmukh “Cluster Oriented Image Retrieval System” Emerging Trends in Computer Science and Information Technology -2012(ETCSIT2012)

[15] Raviraj Kasture, Dr. A. M. Dixit” Internet Image Search Based On User Intention” International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 6, June 2014.

[16] R.Ranjani, S.Anitha Elavarasi,Akilandeshwari,Categorical Data Clustering Using Cosine Based Similarity for Enhancing the Accuracy of Squeezer Algorithm,IJCA , vol 45, pp41-45,May2012.