JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

COMBINED HIERARCHICAL AND K-MEANS METHOD FOR EFFECTIVE RETRIEVAL OF IMAGES

[1]D.GAYATHRI, [2]K.SINDHUJA, [3]J.BOOMA

[1,2]Assistant Professor, Department of Electronics and Communication Engineering, PSNA Colleg of Engineering, and Technology, Dindigul, Tamil Nadu, India

[3]Associate Professor, Department of Electronics and Communication Engineering, Theni Kammavar Sangam College of Technology, Theni, Tamil Nadu, India

Emails:[1]deivagayathri@gmail.com[2]sindhudas91@gmail.com, [3]boomakumar2005@gmail.com

Received: 14 March 2020 Revised and Accepted: 8 July 2020

ABSTRACT: Image retrieval is one of the most important aspects in image processing. In recent years more number of approaches has been widely used for content based image retrieval. In content based image retrieval the images are retrieved depend on their matches from the dataset. The image features such as color, shape and size features are considered for effective retrieval. During image retrieval one of the features is identified and based on that search executes. Here for more number of images to reduce the search time first the images which are most relevant to the search image is grouped and the negative images (which are not closest to the search image) are removed. From the grouped images the retrieval is achieved.

KEYWORDS: Clusters, content based, K-Means, image retrieval

I. INTRODUCTION

The retrieval of images includes the task of finding the most significant images from the huge dataset. Due to the growth in multimedia, more image based data bases are used and it becomes so complex to retrieve the images. The significant task is to find out the closest match towards retrieval from a large group of dataset. For smaller data set it is able to search the image within the certain data but for larger image dataset it requires more advanced technique that are used to retrieve the image with exact perspective. Some systems use the query to retrieve the image where the retrieval of the image depends on the query. Some other system uses the keyword to retrieve the images. The file link is one more technique that take towards the required images. The retrieval depends on the identical result in terms of image color and other attributes such as shape, size etc., hence the retrieval system requires good technique which can perform well based on the dataset to retrieve. In research area the image retrieval improve the results with more accuracy.

When dealing with large scale images the retrieval depends on various dimensions for the accuracy. Images which are translated and also rotated to certain level have more complex issues in order to match it based on its content. Most of the applications that uses these type of retrieval is in the field of bio medicine, web image retrieval, iris recognition etc., Recent image retrieval simply uses the query and during the retrieval the specific query is matched towards the keyword and it simplify the task by matching the query with the keyword rather than matching directly with the images. Some of the considerations in this technique is that for large group of images it becomes unfeasible to match them and the particular features which are important in the image cannot be elaborated with the keyword. Hence it focuses on the content in the image and only the content can describe the image more than when we compare it with keyword. The content based image retrieval system identify various regions in the image with respect to the identical features such as color, shape and some more features concludes the result by how much they are close with the identified regions. The images with low level features are not included in the text based image retrieval system but in content based image retrieval it is extended and the retrieval accuracy is also achieved.

In image processing the concept of image classification can improve the process of retrieval speed in large data bases with better result accuracy. The similarity metrics is not directly used for image categorization but in case of image grouping it requires the similarity measures. But for retrieval the concept of image categorization and also the concept of grouping are used together for efficient retrieval. When some systems focuses on the arguments for

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

grouping but mostly its being an basic previous phase in content based image retrieval. The images grouped on query 'sky' is shown in Fig.2.

II. HIERARCHICAL AND K-MEANS GROUPING TECHNIQUES

The concept of hierarchical clustering and k-means is also used for the image retrieval towards large data bases. In case of pattern recognition these two techniques plays an important role to classify the images together various groups. In order to improve the retrieval speed the concept of hierarchical technique is used to form groups among images. For an iterative refinement the heuristic technique K-means is used which can make the process faster. The better grouping can be formed in images by executing the k-means technique repeatedly. Even from past days in computer field, color is mostly treated as an important feature where it produces good results, when the color feature is identical a maximum level of retrieval accuracy is obtained. For large number of images, the retrieval can depend on color where it can group based on color similarity. The important aspect in color feature extraction is the color space. It refers to an multidimensional space by representing the different color components together with various dimensions. There are always three dimensional color spaces available i.e., RGB. RGB is a color space which assigns the pixels on three element vector with their color intensity with three primary colors named as red, green, blue. By using the space spanned by the red, green and blue attributes with the representation of vectors by 3D RGB color space, the visible colors can be described. For identifying the color features in images the RGB color space is the beginning point.

III. PROPOSED APPROACH

In this paper the two grouping algorithms named as the hierarchical grouping and the most popular K-means algorithms are used to group the images into different categories by identifying the color present in the image. In case of hierarchical technique, it filters the images which are most relevant and followed by the filtration in relevant images, the grouping is achieved by k-means, where the most similar images only results at the final stage and the retrieved images based on this technique is shown in Fig.3.

IV. ALGORITHM

Step 1- Give the input image (I)
Step 2- Compare it with the databases images from $I=(I_0,I_1,\ldots I_{n-1})$
Step 3- Extract the color feature from the images
Step 4- Based on color, group the image into several sets
Step 5- Identify the groups which are similar and merge them to form a single parent group.
Step 6- Repeat Step 3 to Step 5 until it forms a single group.
Step 7- For the obtained single group from Step 6,apply the k-means clustering technique towards the group
Step 8- Enter the number of clusters (let it be M)
Step 9- By random guess M groups center locations.
Step 10- Individual data points determines which center it is closest
Step 11- The center locations 'owns' with set of coordinates
Step 12- Individual center points identify the Centroid of its own points
Step 13- The center moves towards the new Centroid
Step 14 - Continue Step 10 to Step 13

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

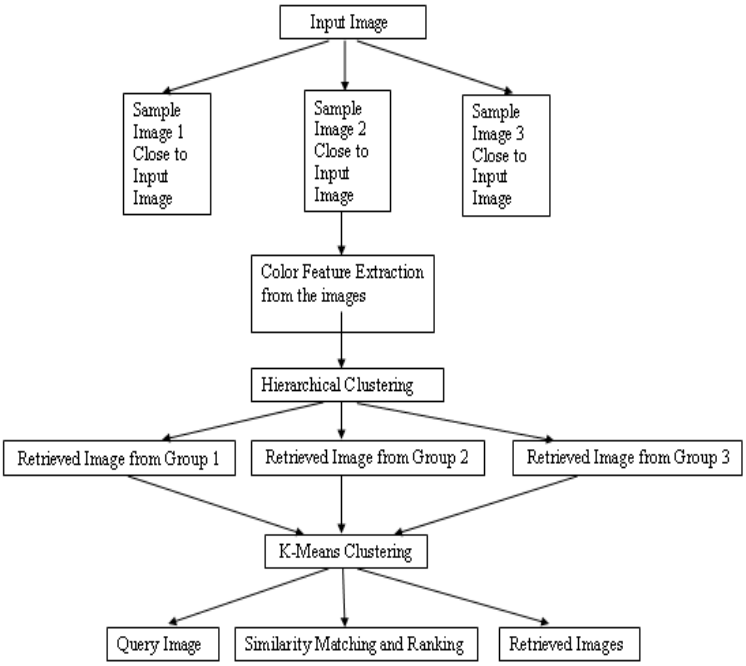

Fig.1. Image retrieval with combined Hierarchical and K-Means clustering

V. IDENTICAL IMAGES AND MEASURES

Content based image search from large database is most challenging process in case of large images databases. The distance measure is an important technique followed in images and wide number of applications in images performs this distance measure. The calculation of the similarity by using the query image with the data base images requires rank to sort them with their identical features. The retrieval depends on the two factors named as $[T_p]$, the time taken to determine the identity towards the query with images in databases and T_r represents the time required to rank the images depend on the identical images with the query. It can be represented by

$T_n = iT_p + O(n \log n)$ …….....(1)

where i represents the number of images in database.

T_p is the required time to calculate the similarity for the images and to sort the elements O (n log n) is used.

The retrieval time for grouping the images from the databases is defined as the sum of three times.

i. the required time to calculate the identical between the cluster center and query

ii. the required time to calculate the similarity between the query and the images towards the closest group

iii. the required time to rank all the groups of images.

Hence the total retrieval time is

$T_g = jT_p + eT_p + o(e \log n)$.....(2)

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

VI. ACCURACY ACHIEVED IN CLUSTERING

To the application domain view, one way of separating the image feature space of its feature vector is clustering. The nearest matches which are identical to the clusters is determined with more accuracy. In this paper the concept of hierarchical clustering and k-means are utilized. In hierarchical clustering concept the images are first filtered and k-means is implemented to retrieve the relevant search. When applying K-means the redefined images are grouped where it can obtain only positive images which are relevant to the content. It can help us to increase the computational speed of the retrieval process. Next step is to select the cluster centers, when the query image is analyzed and matched with the cluster centers. The cluster provides the rank for images which is given on the identical matches with query. The query image is next compared with all images from the clusters. This can reduces the comparisons to match exactly with the query image. The size of the cluster is considered and after all the number of clusters is examined, the total number of identical matches is identified.

Fig.2. Cluster of Images grouped based on Query "sky".

Fig.3. Retrieved image based on hierarchical Clustering.

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

Fig.4. Retrieved image based on Combined hierarchical and K-means clustering

The result of our proposed work groups the images and it is shown in Fig.4. In this result it classifies the images which are similar and group according to their similarity. The searching of images from large databases uses the clustered images based search to retrieval. For query based retrieval the images grouped for the query 'sky' are grouped and shown in the Fig.2. and the images retrieved by using hierarchical clustering is shown in Fig.3. Here the clustered images by hierarchical approach is applied to k-means approach which work with the input value H and support a set of 'i' objects to 'M' clusters, where the resulting cluster have most identical matches. If the cluster is similar to the search image, an object is assigned. This assignment of object is done depend on the distance between the object and the closest center. By doing this a new centroid is computed and it helps to identify the centroid with its own points until the criterion function converges this process continues. The end result produces the accurate retrieval and it performs well instead of using other algorithms.

VII. CONCLUSION

For content based image retrieval the approach of hierarchical and k-means are used for clustering the images, where the images are first grouped based on the identical color content and next step is to find the particular group by using K-means. For faster image retrieval hierarchical clusters helps and it finds the relevant matches which can be obtained from the large data base images. The next step is the optimized search, which produces the highest accuracy rate from image retrieval, where it is performed on the clusters by k-means. Most of the clusters have supports for similar images the user select the image set of his own choice and it can be further refined by implementing k-means technique. In this work the two proposed concepts i.e., the hierarchical concept and k-means technique not only supports for grouping the images, it will also produce accurate results of image retrieval with minimum search time.

FUTURE SCOPE

In our proposed work the images are retrieved based on hierarchical and k-means grouping. Instead of retrieving the images based on query and text based systems, the images that correspond to particular domain are grouped by the k-means algorithm. The images that are nearest to particular group is identified and the images that are present in the group are retrieved. Further work can be extended by implementing the recent optimization algorithm like crow search optimization algorithm to obtain the optimal results.

VIII. REFERENCES

[1] Santhana Krishnamachari, Mohamed Abdel Mottaleb, (1999). Hierarchical clustering algorithm for fast image retrieval. Part of the IS&T/SPIE Conference on Storage and Retrieval for Image and Video Databases" VII,San Jose, California: 427-435.

[2] Anelia Grigorova, Francesco G.B. De Natale, Charlie Dagli and Thomas S.Huang, (2007). Content based image retrieval by feature adaptation and relevance feedback. IEEE transactions on multimedia, Vol-9, No.6,

[3] Ritendra Datta, Dhiraj Joshi, Jia Li, James Z. Wang (2008). Image Retrieval: Ideas, Influences, and Trends of the New Age. ACM Computing Surveys, Vol. 40, No. 2,

[4] Mohamed Abdel-Mottaleb, Santhana Krishnamachari,Nicholas J. Mankovich (1998). Performance Evaluation of Clustering algorithms for scalable image retrieval. IEEE Workshop on Empirical Evaluation of Computer Vision Algorithms

[5] Murthy et al., (2010). Content Based Image Retrieval using Hierarchical and K-Means Clustering Techniques. International Journal of Engineering Science and Technology,Vol.2(3), 209-212.

[6] Xiang-Yang Wang, Hong-Ying Yang, Dong-Ming Li (2013). A New content-based image retrieval technique using color and texture information", Computers & Electrical Engineering, Volume 39, Issue 3, 746-761.

[7] Heng chen, Zhicheng zha. (2010). An effective relevance feedback algorithm for image retrieval. IEEE journal , 978-1-4244-6853.

[8] Yan He, Lei Yang, Yichun Zhang, Xiaoyu Wu and Yun Chang. (2014). The binary image retrieval based on the improved shape context. 7th international congress on image and signal processing.

[9] Da Pan, Ping Shi and Cuiying Li (2014). Sketch based image retrieval by using saliency.11th international conference on fuzzy systems and knowledge discovery

[10] Sugandha Agarwal, Ridhi Sharma and Rashmi Dubey (2016). Sketch based image retrieval using watershed transformation",Second international conference on computational intelligence & communication technology.

[11] Patrick P. K. Chan, Zhi-Chun Huang, Wing W.Y. NG and Daniel S. Yeung.(2011). Dynamic hierarchical semantic network based image retrieval using relevance feedback", International conference on machine learning and cybernetics, Guilin.

[12] T.Gokaramaiah, P.Viswanth and B.Eswara Reddy.(2010). A novel shape based hierarchical retrieval system for 2D images. International conference on advances in recent technologies in communication and computing.

[13] Vimina.E., Ramakrishnan.K, Navya Nandakumar and Poulose Jacob.K (2015). An efficient multi query system for content based image retrieval using query replacement.,IEEE Conference on image processing.

[14] Kamel Belloulata, Lakhdar Belallouche, Amina Belalia and Kidiyo Kpalma (2014). Region based image retrieval using shape adoptive DCT.IEEE Conference, Doi:978-1-4799-5403-2/14.

[15] Konstantinos A. Reftopoulos, Klimis S.Ntalianis, Dionyssios D.Sourlas and Stefanos D. Kollias (2013). Mining user queries with Markov Chains: application to online image retrieval.IEEE transactions on knowledge and data engineering, Vol-25, No-2.

[16] Hanjiang Lai, Pan Yan, Xiangbo Shu, Yunchao Wei and Shuicheng Yan(2016). Instance aware Hashing for multi label image retrieval", IEEE transactions on image processing, Doi: 10.1109/ TIP.2016.2545300.