ISSN No. 0976-5697

Volume 3, No. 7, Nov-Dec 2012

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Utilization of Hierarchical and flat clustering in Content Based Image Retrieval

Priyanka Gupta
Dept. Of Computer Science & Electronics
AIM & ACT Banasthali University, Tonk,
Rajasthan, India
priyanka.gupta23pg@gmail.com

Abstract: In this paper we are going to describe an image retrieval system whose input query is images and as output retrieves the images related to the image content similar to the query ignoring the similarities among the database images. Content based image retrieval means search will analyze the actual content of image rather than metadata for example attributes as tags, keywords and descriptions associated with the image. Here Actual content of the image may be considered as colors, textures, shapes and other information that can be derived from image itself. Main unique objective of this system is the utilization of hierarchical and k-means clustering techniques. Here proposed techniques consisting two stages-First in hierarchical clustering we filter most of the images in an unstructured large database and then applying clustered images to K-means such that system can return more better accurate results

Keywords: CBIR, hierarchical clustering, K-means, similarity

I. INTRODUCTION

Retrieval of images is process of searching and retrieving images from a huge database of images. Collection of the images is growing on web. Retrieving images from this collection is a great challenge. For searching the images user has to use query term attributes as keyword, image link and by clicking on any image and retrieval system will return images those matches to query. For matching of search criteria these attributes can be used as meta-tagging, color, shape, region information. Where textual and keywords are used for searching is called image meta searching and where for image retrieval similarity between images can be measured based on some attributes textures, colors and shapes.

Before content based image retrieval system search-engines on web need to depend on meta-data information. Human has to enter keywords for images manually in a large database. This technique can produce much garbage (unnecessary stuff) in searched result. So this technique may be inefficient, expensive due to involving manual annotation by human labor and may not be able to capture all keywords to describe the image. Because of limitations in meta-based system, for large database textual information and human involving is impossible and not sufficient for large data base. In this case such type of images can be missed those use the synonyms in their description. There are several disadvantages with this technique as

a) Due to huge collection of digital images, it is not optimal and feasible to annotate manually.
b) All the rich features cannot be described for image retrieval by using manually annotated keywords.

Due to all these drawbacks of text based retrieval CBIR (Content Based image Retrieval) a new technology has been called [1].

It's a technology for retrieving desired results based on some syntactic image features. In CBIR features extracted from images exist in high dimensional space so there is high complexity in calculating for similarity retrieval, indexing and searching so similarity measure concept is used to calculate the degree of similarity among images. This technology differentiate different regions present in image based on similarity in different feature as colors, shape, pattern, texture and then closeness of the regions can be used to determine the similarity between two images. Thus we can overcome the difficulties present in meta-text based image retrieval. Low level image features can be automatically extracted from images by using CBIR and describe any image in more descriptive manner text based image retrieval.

Image categorization or classification is the preprocessing steps used for speeding up retrieval process of image and for automatic image annotation in large database. Image clustering inherently depends on similarity metrics while in image categorization using the methods that does not depend on similarity measures. Image categorization is followed by similarity measurement. Categorization and similarity match steps conduct retrieval process both together. Clustering is a fundamental "early" step in image retrieval [1].

Hai [2] proposed to understand the images through the analysis of semantic links existing among web pages. In this area where the ever-increasing number of images acquired through the digital world, it makes the brute force searching almost impossible.

Here we are using Hierarchical and K-means clustering techniques which are being used for desired image retrieval in a large database. Both techniques are being used for pattern reorganization. Hierarchical approach is used to make group the similar images into cluster to fast retrieval process. K-means is an iterative refinement heuristic approach because it can be run several times to find best clustering. K-means is very sensitive to noise so better result is obtained when we try k-clustering similarity. Some low level features can be explained which can only be extracted using CBIR—

A. Color Retrieval- Color is most widely used visual feature for retrieval of image similarity and it is independent of image size and it's orientation. Color retrieval provides the best result and it is same as those

which is derived by human visual processing system. One factor of color feature extraction is the color space which is multi dimensional space in which multiple dimensions represent multiple components of color. For example three dimensional color space model RGB assigns each pixel three element vector (color intensities of primary colors).RGB color space provides a useful starting point for representing color features of images [3].

B. ***Texture Retrieval-*** It is also a visual feature of image for separating the regions in the image. But it's not having more utility. It is useful only in differentiating the regions of image having same color for example (water and sea or garden and grass) .Different methods are used for calculating texture similarity. These queries can be converted into color queries by providing a query image as example or by selecting favored texture example from palette [3].

C. ***Shape Retrieval-*** It is also a visual feature for retrieval of images based on their similarity. This feature plays an important role for recognizing the objects and related similar images based on their content. Mainly used for determining a set of consecutive levels and aspect ratio of images [3].

II. PROPOSED METHOD

To group the images into clusters based on having color content, clustering strategies hierarchical and K-means clustering are being used here. Both clustering techniques are used in pattern recognition,

In hierarchical clustering filtering of most images is done and then apply K-means to clustered images.

Then desired image result is obtained [4].

A. Hierarchical and K-means combined algorithm:

This algorithm can be clearly explained as—

B. Take input image as query:

Image can be taken as from anywhere or any folder can be selected from any large database from which we have to search any image for input query.

C. Hierarchical clustering:

After selecting image searching of image is done based on resemblance of color, texture and shape feature (similarity judgment standards). Anyone of these features can be selected for image retrieval and then images are hierarchically grouped based on these feature similarities.

D. K-means clustering:

Here images are processed for final image retrieval .First most similar images are clustered al-together and then arranged in sorted order in decreasing order of their similarity. Thus images those are most resembled to image query are found and then image results retrieved after applying these clustering are displayed.

III. IMAGE SIMILARITY MEASUREMENT

Searching images in a large database is a challenging task. First similarity (matching) between query image and all images present in database is calculated. Then ranks are assigned by sorting their similarities .So total image retrieval time can be calculated as summation of time to calculate the similarities between query image and all images in database and the time to rank all the images present in database based on sorting order of similarities.

When the images in the database are clustered, the retrieval time is the summation of three times for example (a)the time to calculate the similarity between the query and the cluster centers.(b)time to calculate the similarity between the query and the images in the nearest clusters (c)time to rank the images[5].

IV. RETRIEVE ACCURACY AND EFFICIENCY WITH BOTH TYPE OF CLUSTERING

Aspect of this retrieval system is proper utilization of hierarchical and K-means clustering, here filtering of most images is done in hierarchical clustering and then K-means is applied to clustered images obtained from previous clustering technique. Then get the desired image search result. After this clustering process and selecting cluster centers, query image is compared with all the images present in database. And then clusters are ranked according to sorted order of similarities. Thus numbers of comparisons are reduced [6].Here query image is not need to be compared to all the images of data base. There is no need for user to search through a large data base, he has to consider only clustered image results. Now again we apply K-means on clustered images obtained from hierarchical clustering which needs to take input parameter, k and partition of n objects into k clusters so that intra-cluster similarity is maximum. This clustering process is repeated until criterion function converges. Thus image retrieval becomes accurate and efficient after applying these clustering approaches.

V. CONCLUSION AND FUTURE WORK

In this paper we have presented an approach for content based image retrieval based on 2 clustering where images are grouped having same color content and selected groups are clustered in K-means technique. This image retrieval technique gains the usefulness of Query By Example methodology in which user has to only submit the sample query image to system as input and system returns most similar matched images to query. And this system optimizes the relevance of the result that was obtained from traditionally retrieval system. Hierarchical clustering provides faster image retrieval and allows the user to search most relevant image in a large data base, and K-means is used to improve the performance to produce the accurate and refined desired result.

Accuracy and efficiency can be improved more further by including some more feature as ROI for low level visual-description as gap between user's expectations and low level visual descriptions (inherently difficult for visual content to capture them, gap between the object in the world and information in a description extracted from recording a scene) can be overcome by considering some more features.

VI. REFERENCES

[1]. Ritendra Datta, Dhiraj Joshi, Jia Li and James Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," ACM Computing Surveys, vol. 40, no. 2, article 5, April 2008, pp. 5:1-60.

[2]. Z. Hai, Retrieve images by understanding semantic links and clustering image fragments, Journal of Systems and Software 73, 2004, pp. 455–466.

[3]. John Eakins, Margaret Graham,"Content-based Image Retrieval", JISC Technology Application Report 39, October 1999.

[4]. V.S.V.S Murthy et al. "Content based image retrieval using Hierarchical and K-means clustering techniques" IJEST Vol. 2(3), 2010, pp. 209-212.

[5]. Santhana Krishnama chari, Mohamed Abdel –Mottaleb, 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, January 1999 pp 427-435.

[6]. Mohamed Abdel-Mottaleb, Santhana Krishnamachari, Nicholas J. Mankovich," Performance Evaluation of Clustering algorithms for scalable image retrieval", Appeared in IEEE Workshop on Empirical Evaluation of Computer Vision Algorithms, CVPR 1998.