International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

An Efficient Image Retrieval System By Optimizing KNN

S Agalya*, S K V Jayakumar
Department of Computer Science, Pondicherry University, Puducherry, Tamil Nadu, India

ABSTRACT

This work presents a Content-Based Image Retrieval (CBIR) system embedded with a clustering technique to retrieve images similar to query image. In this work, extensive robust and important features were extracted from the images database and then stored in the feature repository. This feature set is composed of colour signature with the shape and colour texture features. Where the features are extracted from the given Query Image in the similar fashion. After that the number of cluster formed in a dataset. Cluster formation based on finding the Euclidean distance between each pair in a dataset. Consequently, a novel image retrieval using k nearest neighbor(KNN) classifier is achieved between the features of the Query Image and the features of the cluster images. Our proposed CBIR system is assessed by inquiring number of images(from the test dataset) and the efficiency of the system is evaluated by calculating precision-recall value for the results. The results were superior to other state-of-the-art CBIR systems in regard to precision.

Keywords: CBIR, Query Image(QI), k-Nearest Neighbor, Euclidean distance, Feature Extraction

I. INTRODUCTION

Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is image, like video frame or photograph and output may be image or characteristics associated with that image. Usually Image Processing system includes treating images as two dimensional signals while applying already set signal processing methods to them. It is among rapidly growing technologies today, with its applications in various aspects of a business. Image Processing forms core research area within engineering and computer science disciplines too. Image processing basically includes the following three steps Importing the image with optical scanner or by digital photography. Analysing and manipulating the image which includes data compression and image enhancement and spotting patterns that are not to human eyes like satellite photographs. Output is the last stage in which result can be altered image or report that is based on image analysis. The Purpose of Image processing is Visualization, Image sharpening and restoration, Image retrieval, Measurement of pattern, Image Recognition. The two types of methods used for Image Processing are analog and Digital Image Processing. Analog or visual techniques of image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. The image processing is not just confined to area that has to be studied but on knowledge of analyst. Association is another important tool in image processing through visual techniques. So analysts apply a combination of personal knowledge and collateral data to image processing.

Digital Processing techniques help in manipulation of the digital images by using computers. As raw data from imaging sensors from satellite platform contains

deficiencies. To get over such flaws and to get originality of information, it has to undergo various phases of processing. The three general phases that all types of data have to undergo while using digital technique are Pre- processing, enhancement and display, information extraction.

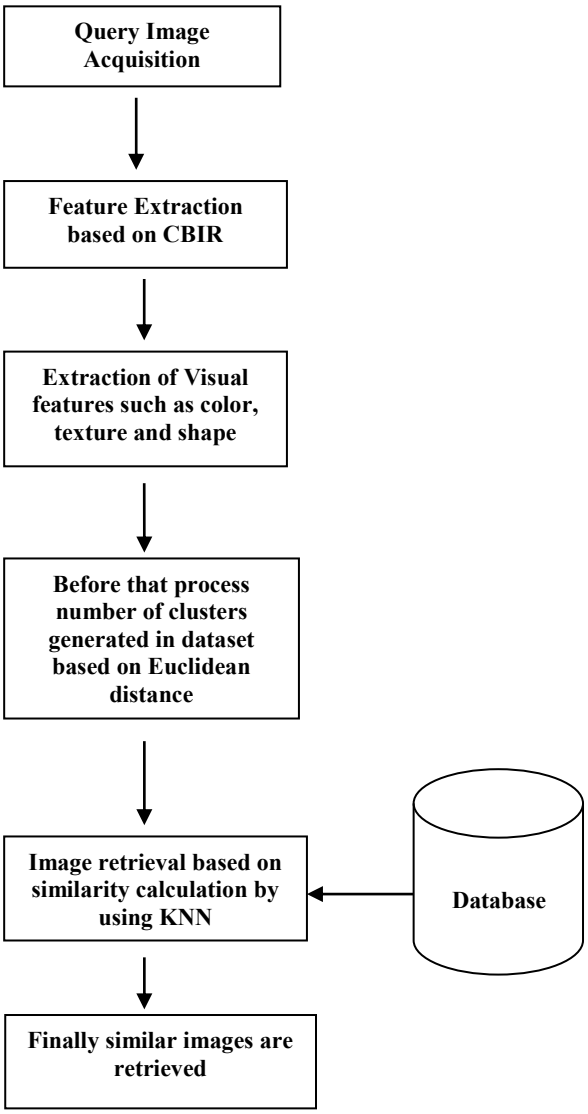


Figure 1. Block diagram of Image Retrieval

An image database contains a wide variety of images which are relevant to the query may be few. For better meet the user intent, the proposed system performs a search in relevant images only. This is different from traditional CBIR systems, which search the whole database for every feature. Relevance of the images are first established by comparing their color feature. Search based on texture and shape features is performed only on the images having color similarity with query image. This approach reduces the diversity of database by removing irrelevant images at each stage so that low level features can better represent the semantics of images. Experiments have shown that the proposed system produces desired results with greater accuracy. The present paper proposes a suggested image retrieval model based on extracting the most relevant features from the image dataset by using k-nn algorithm. The main purpose of texture- based retrieval is to find images or regions with similar texture.

Retrieval algorithms used in traditional CBIR systems search the whole database independently for different image features. Each of the features is represented by a point in the corresponding feature space. Some systems use several feature spaces to represent the same feature to improve retrieval accuracy. In this case, search in each feature space is performed independently, followed by data fusion methods to merge the retrieved sets (intermediate outputs) into one common output. An output is a ranked set of retrieved objects, which is an answer of the retrieval system to a given query. To merge the results of retrieval in different feature spaces, it is common to use linear combinations of the ranks of an element in each intermediate output as its rank in the common output.

II. LITERATURE SURVEY

Clustering is also used for image annotation, indexing, and object discovery. In clustering, groups of objects are formed with high intra-cluster and low inter-cluster similarities. Clustering algorithms are categorized in many ways. If human participation is required then it is supervised, otherwise unsupervised. In another one, an object can participate either in one cluster or many clusters. Clustering techniques are also classified as

hierarchical and partitional. Partitional algorithms form predefined number of clusters and cannot identify varying size clusters. In contrast, hierarchical algorithms output a hierarchy and are more informative. Hierarchical algorithms are either agglomerative or divisive. Agglomerative is bottom up, i.e., starts with one object per cluster and keeps on merging clusters until all objects are in one cluster. Inversely, the divisive starts with all elements in a single cluster, which is then divided successively. Conceptually, divisive is more complex as it requires another method for splitting, and thus agglomerative approach is used more often. Clustering in this manuscript is mainly used to enhance the precision of a CBIR system. An unsupervised clustering utilizing the notion of hierarchical agglomerative clustering is presented to determine visually separable clusters in a large image dataset. Each cluster has a representative image to depict the semantic of images it has. The concept of proportional reduction in information loss is used to keep track of the visual similarity of the image with cluster representative image before merging the images in the cluster. Inclusion of this clustering algorithm in the traditional CBIR system, as proposed in the next section, gives a cluster-based CBIR with improved efficiency.

III. PROPOSED WORK

Our proposed work utilizes a K nearest neighbor(KNN) to find the images that has the highest similarity with the QI from a database. During the CBIR process every image in the database and in QI colour, shape and texture features are extracted. After that the CBIR retrieves the most relevant images to the QI from the images database based on KNN. The Proposed work has following steps

- ✓ Set of images are collected from public database.
- ✓ Visual features such as color, texture and shape features are extracted from the database images. Finally dataset features are generated.
- ✓ After that the clustering process is explored in database. We proposed the number of clusters generation in database for efficient image retrieval. Cluster generation is achieved by using Euclidean distance.
- ✓ Same feature extraction is performed in query image.

Finally, the similar images are retrieved using KNN. The advantages of proposed system is the processing time will become less and accuracy will be high.

1) Processing of image using an Knn

A. Collection of images: The set of images are collected from publicly available database. Here we are used Corel dataset for our work. The Corel dataset consists of 10,908 different images with the size of 256 * 384 or 384 * 256 for each image. Therefore the results were reported using ten semantic sets, each set has 100 images. These groups of datasets are Buses, Mountains, Beach, Elephants, Food, Flowers Africa, Horses, Dinosaurs and Buildings.

B. Feature Extraction: The visual features are extracted from database images. In this stage image descriptor such as Content-based image retrieval method is used for feature extraction. Color, Texture and shape features are extracted in database images. Finally encoding the extracted information into feature vectors for further process. In colour based, colour hsv histogram, autocorrelation and color moments features are extracted. In texture based, energy, contrast, correlation and homogeneity features are extracted using GLCM method. In shape based, edge feature is extracted. Three types of feature included in this extraction they are

1. Colour Feature
2. Texture Feature and

3. Shape Feature

i)Colour Feature: Color feature is an essential component for image retrieval. For huge image databases, image retrieval using the color feature is very successful and effective. Although color feature is not a persistent parameter, because it is subjected to many non-surface characteristics for example, the taking conditions such as illumination, characteristics of the device, the device view point. The steps of the color feature extraction are shown below:

1) Color planes values RGB are separated into individual matrices namely Red, Green and Blue matrices.
2) For each color matrix color histogram should be calculated.
3) Variance and median of color histogram has to be calculated.
4) The summation of all row variances and medians are calculated.
5) The calculated features of all matrixes (R, G and B) are combined as feature vector.
6) The feature vectors are stored in the features database.

ii) Texture Feature: Texture features extraction using the GLCM method. The GLCM is a robust image statistical analysis technique. GLCM can be defined as a matrix of two dimensions of joint probabilities between pixels pairs, with a distance d between them in a given direction h. In our research we are using five features for the feature extraction. The Steps of the texture features extraction is shown below:

1) Filtering the input image using the 5*5 Gaussian Filter.
2) Filtered image is divided into 4 *4 blocks.
3) For each block Homogeneity, correlation, Contrast and Energy are calculated using GLCM, these features were calculated based on four directions.
4) The extracted features are stored in the feature database.

iii) Shape Feature: The shape feature extraction mainly aims to capture the properties of the shape of the image items. This eases the process of shape storing, transmitting, comparing against, and recognizing.

C. Formation of cluster: In this stage, group of visual features formation is explored. After feature extraction, a distance model containing similarity between each image pair is computed. Using this model, two representative images with maximum similarity are identified and clustered in bottom-up manner. Euclidean distance are used for cluster generation.

2) Retrieving Image From The Dataset

Our final stage is image retrieval. Similar images are retrieved based on features. Same feature extraction such as color, texture and shape features are extracted for query image. After that the similar images are retrieved by using KNN classifier model between cluster features and test features.

The Clustering Algorithm for grouping an image in a database includes following steps

Step 1) Collect the images from public database.
Step 2) Extract the visual features for database images.
Step 3) Get the size and number of image features.
Step 4) Randomly select the k cluster centre.
Step 5) Euclidean distance is calculated from the center of the source features to the center of each of the surrounding features. True Euclidean distance is calculated in each of the distance functions. Conceptually, the Euclidean algorithm works for each cell, the distance to each source cell is determined by calculating the hypotenuse with x_max and y_max as the other two legs of the triangle.

Step 6) Finally we get group of images called clusters within dataset.

Step 7) Cluster formation having 5 steps.

Stage 1:

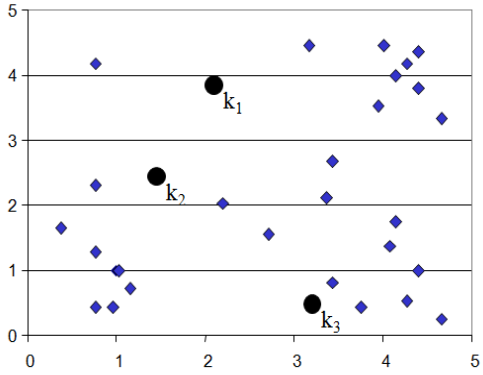


Figure2. Collection of images in a database, where k1,k2 and k3 are the pointer for grouping an images.

Stage 2:

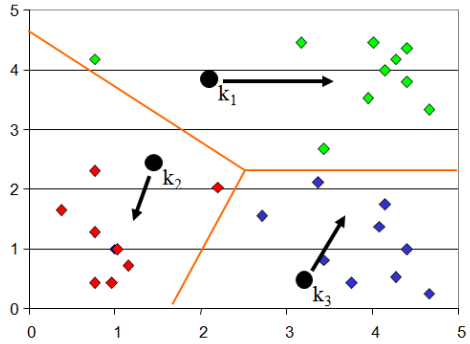


Figure3. k1,k2 and k3 found that the images which are closer to the pointer.

Stage 3:

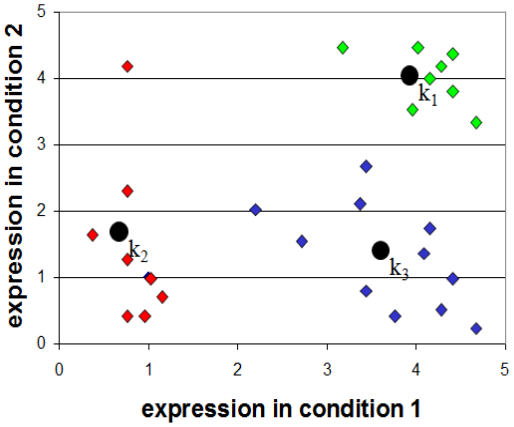
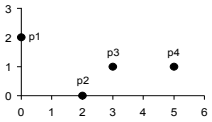


Figure 4. It is the final stage where the cluster is formed with respect to the distance between images.



point	x	y
p1	0	2
p2	2	0
p3	3	1
p4	5	1

	p1	p2	p3	p4
p1	0	2.828	3.162	5.099
p2	2.828	0	1.414	3.162
p3	3.162	1.414	0	2
p4	5.099	3.162	2	0

Figure 5. Feature extraction value of each image.

IV. EXPERIMENTAL RESULT

A)Euclidean Distance

The Euclidean Distance can be expressed by,

$$dist = \sqrt{\sum_{k=1}^{n}(p_k - q_k)^2}$$

Where n is the number of dimensions (attributes) and p_k and q_k are, respectively, the k^{th} attributes (components) or data objects p and q. Standardization is necessary, if scales differ.

B)k-Nearest Neighbor

The basic k-nearest neighbor classification includes Training method it saves the training examples at prediction time to find the k training examples $(x_1,y_1),\ldots(x_k,y_k)$ that are closest to the test example x and predict the most frequent class among those y_i's.

It improves Weighting examples from the neighbourhood for Measuring "closeness" and for finding "close" examples in a large training set quickly

Figure 5. Comparing the images from database with clustering and without clustering mechanism.

V. CONCLUSION

This work proposed an effective CBIR system using KNN to retrieve images from databases. Once the user inputted a query image, the proposed CBIR extracted image features like color signature, shape and texture color from the image. Then the database clustering model was explored by using Euclidean distance and using the KNN based similarity measure, images that are relevant to the QI were retrieved efficiently. The conducted experiments based on the Corel image database indicate that the proposed KNN algorithm has strong capability to discriminate color, shape and color texture features. Our proposed CBIR system was evaluated by different images query. The execution results presented the success of the proposed method in retrieving the similar images from the images database and outperformed the other CBIR systems in terms of average precision and recall rates. This can be represented from the precision and recall values calculated from the results of retrieval. In the future, filtering techniques and more efficient classifier will be employed to get more accurate results in the content based image retrieval system.

VI. REFERENCES

[1]. Prof KherfiML, Ziou D, Bernardi A. Image retrieval from the world wide web: issues, techniques, and systems. ACM Comput Surv (CSUR) 2004;36(1):35-67.

[2]. Alamdar F, Keyvanpour MR. Effective browsing of image search results via diversified visual summarization by clustering and refining clusters. Signal Image Video Process 2014;8(4):699-721.

[3]. Sternberg RJ. Cognitive Psychology. 6 edition. Wadsworth Cencage Learning; 2011.

[4]. Lin C-H, Chen R-T, Chan Y-K. A smart content-based image retrieval system based on color and texture feature. Image Vision Computi 2009;27(6):658-665.

[5]. ElAlami ME. A novel image retrieval model based on the most relevant features. Knowl Based Syst 2011;24(1):23-32.

[6]. Youssef SM. ICTEDCT-CBIR: integrating curvelet transform with enhanced dominant colors extraction and texture analysis for efficient content-based image retrieval. Comput Electr Engg 2012;38(5):1358-76.

[7]. Shrivastava N, Tyagi V. An efficient technique for retrieval of color images in large databases. Comput Electr Engg 2014.

[8]. Tuytelaars T, Lampert CH, Blaschko MB, Buntine W. Unsupervised object discovery: a comparison. Int J Comput Vis 2010;88(2):284-302.

[9]. Xu R, Wunsch D, et al. Survey of clustering algorithms. IEEE Trans Neural Netw 2005;16(3):645-78.

[10]. Chen Y, Wang JZ, Krovetz R. Clue: cluster-based retrieval of images by unsupervised learning. IEEE Trans Image Process 2005;14(8):1187-201.

[11]. Elkamel A, Gzara M, Ben-Abdallah H. A bio-inspired hierarchical clustering algorithm with backtracking strategy. Appl Intell 2014:1-21.

[12]. Kannan S, Devi R, Ramathilagam S, Hong T-P, Ravikumar A. Robust fuzzy clustering algorithms in analyzing high-dimensional cancer databases. Appl Soft Comput 2015;35:199-213.

[13]. Villalba LJG, Orozco ALS, Corripio JR. Smartphone image clustering. Expert Syst Appl 2015;42(4):1927-40.

[14]. Durduran SS. Automatic classification of high resolution land cover using a new data weighting procedure: the combination of k-means clustering algorithm and central tendency measures (KMC-CTM). Appl Soft Comput 2015;35:136-50.

[15]. Wang X-Y, Yu Y-J, Yang H-Y. An effective image retrieval scheme using color, texture and shape features. Comput Stand Interfaces 2011;33(1):59-68.

[16]. Sural S, Qian G, Pramanik S. A histogramm with perceptually smooth color transition for image retrieval.. In: JCIS; 2002. p. 664-7.

[17]. Manjunath BS, Ma W-Y. Texture features for browsing and retrieval of image data. IEEE Trans Pattern Anal Mach Intell 1996;18(8):837-42.

[18]. Pandey S, Khanna P. A hierarchical clustering approach for image datasets. In: Ninth international conference on industrial and information systems (ICIIS). IEEE; 2014. p. 1-6.

[19]. Rokach L, Maimon O. Clustering methods. In: Maimon O, Rokach L, editors. Data mining and knowledge discovery handbook. Boston, MA: Springer US; 2005. p. 321-52. doi: 10.1007/0-387- 25465- X _ 15.