International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Design and Analysis of CBIR System using Hybrid PSO and K-Mean Clustering Methods

Shailesh Kumar* **and Arun Kumar Shukla**

Department of Computer Science and Information Technology, SIET, SHUATS, Naini, Allahabad, India

Accepted 10 March 2017, Available online 14 March 2017, **Vol.7, No.2 (April 2017)**

Abstract

Images have always been considered an effective medium for presenting visual data in many applications of industry and academia. With the development of technology, a large amount of images are being generated every day. Therefore, managing and indexing of images become essential in order to retrieve similar images effectively. In conventional systems, images are generally indexed with textual annotations. However, as the database grows larger, the use of keywords based methods to retrieve a particular image becomes inefficient. Content-based Image Retrieval (CBIR) systems demonstrate excellent performance at computing low-level features from pixel representations but its output does not reflect the overall desire of the user. In this paper we Proposed Hybrid Approach which is combination of PSO and K-Mean clustering. Therefore the image contents are evaluated using the edge for shape feature extraction; grid color movement analysis is performed for color feature evaluation. Consequently the query image is also extracted for comparison and accurate image extraction. Proposed implementation is provided using MATLAB configuration which is show improved system performance outcomes and accurate image retrieval of the query image.

Keywords: *CBIR, PSO, K-Mean Clustering, query image, RGB, histogram.*

1. Introduction

The emergence of multimedia technology and the rapidly expanding image collections on the Internet have attracted significant research efforts in providing tools for effective retrieval and management of visual data. The need to find a desired image from a large collection is shared by many professional groups, including journalists, design engineers and art historians. Difficulties faced by text-based image retrieval brought the researchers to develop new solutions to represent and index visual information (Aman Chadha and Sushmit Mallik *et al*, 2016).

A. Content based Image Retrieval (CBIR)

The term CBIR refers to the process of retrieving similar images from a large collection of image database. The image retrieval is done on the basis of similarity matching between query image and database images. Content-based image retrieval (CBIR) is a technique for retrieving images on the basis of automatically-derived features such as color, texture and shape. The architecture of a CBIR system can be understood as a basic set of modules that interact within each other to retrieve the database images according to a given query. In typical content-based image retrieval system, the visual contents of the images in the database are extracted and described by multi-dimensional feature vectors. The feature vectors of the images in the database form a feature database. To retrieve images, users provide the retrieval system with query images or sketched figures. The figure 1 shows general block diagram of CBIR system (Sonit Singh, Aman Saini and Shamneesh Sharma *et al*, 2013).

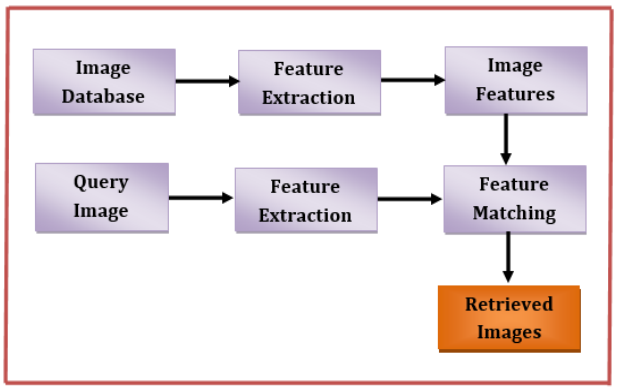

Figure 1: Block Diagram of CBIR System

In on-line image retrieval, the users submit a query image to the CBIR system in search of desired images.

*Corresponding author **Shailesh Kumar** is a Research Scholar and **Arun Kumar Shukla** is working as Assistant Professor

The system represents this query image with a feature vector. The similarities between the feature vectors of the query example and those of the images in the feature database are then computed and ranked. Retrieval is computed by applying an indexing scheme to provide an efficient way of searching the image database. Finally, the system ranks the retrieval results and then returns the images that are most similar to the query images (PritiMaheswary, Dr. Namita Srivastava *et al*, 2009).

The content-based approach can be summarized as follows:

- Computer vision and image processing techniques in are used to extract content features from the image.
- Images are represented as collections of their prominent features. For a given feature, an appropriate representation of the feature and a notion of similarity are determined.

Image retrieval is performed based on computing similarity or Dissimilarity in the feature space, and results are ranked based on the similarity measure (P. S. Hiremath and Jagadeesh Pujari *et al*, 2007)

B. Image Retrieval

An image retrieval system can be defined as searching, browsing, and retrieving images from massive databases consisting of digital images. Although Conventional and common techniques of retrieving images make use of adding metadata namely captioning keywords so as to perform annotation of words. However image search can be described by dedicated technique of search which is mostly used to find images. For searching images user provides the query image and the system returns the image similar to that of query image (Michele Saad *et al*, 2008). Image Retrieval has been adopted in most of the major search engines, including Google, Yahoo!, Bing, etc. A large number of image search engines mainly employ the surrounding texts around the images and the image names to index the images.

C. Feature Extraction

Feature (content) extraction is the basis of content-based image retrieval. In a broad sense, features may include both text-based features (key words, annotations) and visual features (color, texture, shape, faces).Visual feature extraction is the basis of any content-based image retrieval technique. In a broad sense, features may include both text-based features (key words, annotations) and visual features (color, texture, shape, etc.). Within the visual feature scope, the features can be further classified as low-level features and high-level features. The selection of the features to represent an image is one of the keys of a CBIR system. Because of perception subjectivity and the complex composition of visual data, there doesn't exist a single best representation for any given visual feature. Multiple approaches have been introduced for each of these visual features and each of them characterizes the feature from a different perspective (Datta, R., Joshi, D., Li, J. and Wang, J. Z. *et al*, 2008).

The paper is composed as takes after. Section II quickly depicts the related work of image, retrieval. In Section III, we give a depiction of the Content Based Image Retrieval model which is based on K-means and PSO. Result summary are introduced in Section IV. At last, Sections V talk about our conclusion and future works.

2. Literature Survey

Particle swarm optimization (PSO) is recent approach that can be employed in a wide range of applications. It is an evolutionary computing method based on colony aptitude which is a better parallel searching algorithm. Image segmentation is a low level vision task which is applicable in various applications such as object recognition, medical imaging, document analysis, just to name a few. PSO itself is a very powerful technique and when combined with other computational intelligence technique results in a truly affected approach. In this paper, (Amanpreet Kaur *et al*, 2012) have reviewed how PSO can be combined with various other methodologies such as neural networks, rough sets, clustering, Thresholding, genetic algorithm, wavelets and fuzzy systems.

In web-scale image retrieval, the most effective strategy is to aggregate local descriptors into a high dimensionality signature and then reduces it to a small dimensionality. Thanks to this strategy, web-scale image databases can be represented with small index and explored using fast visual similarities. However, the computation of this index has a very high complexity, because of the high dimensionality of signature projectors. In this work, (Romain Negrel *et al*, 2014) propose a new efficient method to greatly reduce the signature dimensionality with low computational and storage costs. Our method is based on the linear projection of the signature onto a small subspace using a sparse projection matrix. They report several experimental results on two standard datasets and with 100k image distracters. They show that given method reduces both the projectors storage cost and the computational cost of projection step while incurring a very slight loss in mAP (mean Average Precision) performance of these computed signatures.

Rabab M. Ramadan *et al*, 2009 presents a novel feature selection algorithm based on particle swarm optimization (PSO). PSO is a computational paradigm based on the idea of collaborative behavior inspired by the social behavior of bird flocking or fish schooling. The algorithm is applied to coefficients extracted by two feature extraction techniques: the discrete cosine transforms (DCT) and the discrete wavelet transform (DWT). The proposed PSO-based feature selection algorithm is utilized to search the feature space for the optimal feature subset where features are carefully

selected according to a well-defined discrimination criterion. Evolution is driven by a fitness function defined in terms of maximizing the class separation (scatter index). The classifier performance and the length of selected feature vector are considered for performance evaluation using the ORL face database. Experimental results show that the PSO-based feature selection algorithm was found to generate excellent recognition results with the minimal set of selected features.

In this paper, (Shikha Nirmal *et al*, 2009) review recent advances in image retrieval. The two fundamental components of a retrieval system, representation and learning, are analyzed. Each component is decomposed into its constituent building blocks: features, feature representation, and similarity function for the representation; short term and long term procedures for learning. Authors identify a series of requirements for each of the sub areas, e.g. optimality, invariance, perceptual relevance, computational tractability, and point out various approaches proposed to satisfy them. Several open problems are also identified.

In this paper (Gurpreet Kaur and Mnupreet Kaur *et al*, 2014) briefly discuss the similarity measures based on which matches are made and images are retrieved. Another important issue in content-based image retrieval is effective indexing and fast searching of images based on visual features. Dimension reduction and indexing schemes are also discussed. For content-based image retrieval, user interaction with the retrieval system is crucial since flexible formation and modification of queries can only be obtained by involving the user in the retrieval procedure. Finally Relevance feedback is discussed which helps in improving the performance of a CBIR system.

3. Proposed Work

This section provides the understanding about the proposed system which is required to develop for the demonstration of the performance study of IR system.

A. Flow Chart

CBIR is basically a two-step process which is Feature Extraction and Image Matching (also known as feature matching). Feature Extraction is the process to extract image features to a distinguishable extent. Information extracted from images such as colour, texture and shape are known as feature vectors. The extraction process is done on both query images and images in the database. Image matching involves using the features of both images and comparing them to search for similar features of the images in the database. Using multiple feature vectors to describe an image during retrieval process increases the accuracy when compared to the retrieval using single feature vector.

In this proposed work we simulate IR using k-mean clustering approach along PSO based approach. K-means perform grouping of image features of vectors of feature whereas PSO modified centroid of images. On this basis we calculate different image constraints and apply k-mans and PSO method

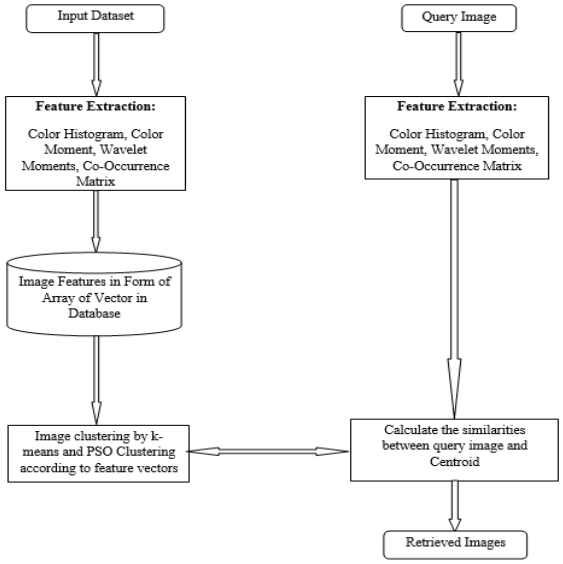

Figure 2: Flow Work of Proposed System

B. Proposed Algorithm

The given section provides the detailed description of the proposed implemented color and edge based image retrieval system. The entire system processes for improving the basic CBIR system, is summarized using the algorithm steps as given in table 1:

Table 1: Proposed PSO+K-Means CBIR Algorithm

Input: Image Dataset (D), Text Image (T), $temp\ (t_1)$, $temp\ (t_2)$ Output: Retrieved Images (I)
Process: 1:t_1 = readDataset (D), t_2 = readTextimage (T) 2:Calculate different image constraints $Histogram_1$ = Color histogram (temp (t_1)) $Histogram_2$ = Color histogram (temp (t_2)) $Color_1$ = Color Movment (temp (t_1)) $Color_2$ = Color Movment (temp (t_2)) $Wavelet_1$ = Wavelet Transfoem ((temp (t_1)) $Wavelet_2$ = Wavelet Transfoem ((temp (t_2)) $Co-occurrence_1$ = Co − occurrance Matrix ((temp (t_1)) $Co-occurrence_2$ = Co − occurrance Matrix ((temp (t_2)) 3:$Features_1$ = Mearge ($Histogram_1, Color_1, Wavelet_1, Co-occurrence_1$) 4:$Features_2$ = Mearge($Histogram_2, Color_2, Wavelet_2, Co-occurrence_2$) 5:k − Means Centroid = k − Means ($Features_1$) 6:updatedCentroid = PSOclustering (k − Means centroid) 7:retrievedImage = imageMatching ($udatedCentroid, Features_1$) 8:I = Retrieved Image 9:return (I)

The proposed system is able to deliver enhancement of the IR to maintain accuracy of the image quality and their features. The proposed algorithm of PSO+ K-means is given table 1 in this algorithm the user input image and database image list is required to produce for searching the relevant image. The input query image is evaluated using the color histogram, colour

movement, co-occurrence matrix and wavelet movements. After calculating after this feature of input query image and input database image, firstly pass input dataset image values to the array of vector database and perform image clustering using PSO and k-Means. Secondly, we have calculated similar values for the input query image. After that, we calculate similarities between query image and centroids. Finally, club the similarity matching and the image clustering to produce retrieved image.

4. Result Analysis

A. Precision

In any data retrieval or search applications the precision is a fraction of search results which is most relevant to the input query. The provided precision of the proposed content based image retrieval system are given using figure 5.1. This can be evaluated using the user feedback basis and can be evaluated by the following formula.

$$\text{Precision} = \frac{\text{Relevant Images} \cap \text{Retrieved Images}}{\text{Retrieved Images}}$$

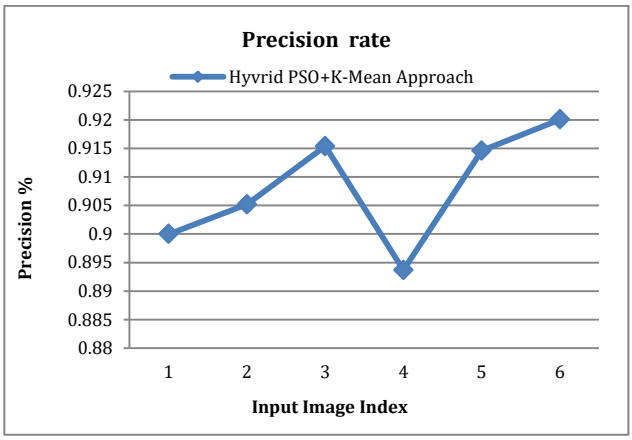

Figure 3: Precision Rate

The precision rate of the implemented system is described in the figure 3, the computed precision values are demonstrated using the Y axis of the given figure and the X axis shows the image value. It shows the amount of training images in the database. According to the obtained results the performance of the proposed system is increases as the amount of data in database is increases. In addition of the precision rate is growing continuously as the similar kinds of images are also increases in data base.

A. Recall

In data retrieval application or the search application recall values are measured for accuracy measurement in terms of relevant document retrieved or relevant data obtained according to the input user query. This can be evaluated using the following formula.

$$\text{Recall} = \frac{\text{Relevant Images} \cap \text{Retrieved Images}}{\text{Relevant Images}}$$

The figure 4 shows the recall values of the proposed image retrieval application. In order to represent the performance of the proposed image retrieval system the X axis contains the amount of images in database and the Y axis reports the obtained recall rate of the implemented system. According to the obtained results the performance of the proposed system is enhances as the amount of data is increases in the database. The retrieval accuracy with the increasing amount of data is also increases thus the proposed concept is adoptable for the image search applications. Therefore the performance of the proposed system is much efficient for different types of query images.

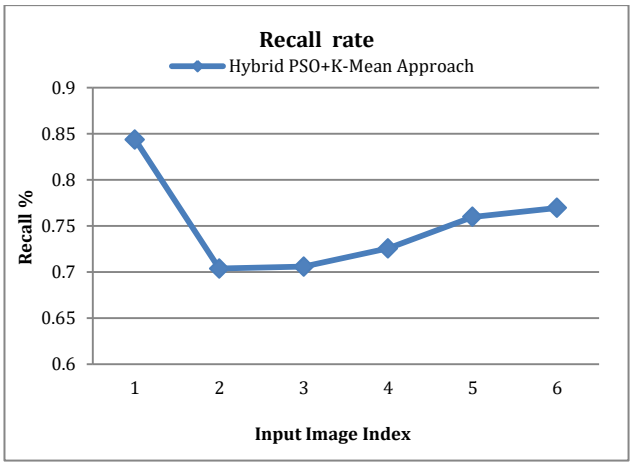

Figure 4: Recall Rate

B. F-measures

The f-measures of the system demonstrate the fluctuation in the computed performance in terms of precision and recall rates. The f-measures of the system can be approximated using the following formula.

$$\text{F} - \text{Measures} = 2.\frac{\text{Precision X Recall}}{\text{Precision+Recall}}$$

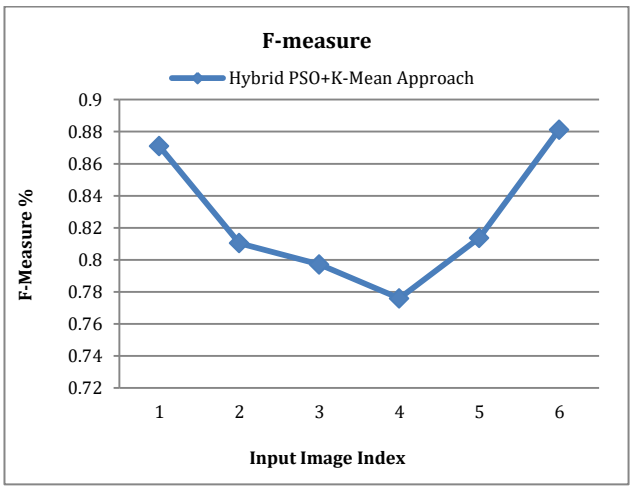

Figure 5: F-Measures

The figure 5 shows the performance of CBIR systems in terms of f-measures. To demonstrate the performance of the system the X axis shows the amount of data is placed in storage during experiments and the Y axis shows the obtained performance in terms of f-measures. According to the obtained results the performance of the proposed system is much stable and enhancing approach lies on combination of PSO and clustering method. In addition of that the results are in more progressive manner as the amount of data base is increases. Thus the obtained results are adoptable and efficient for the image retrieval applications.

Conclusion and Future Work

Content based image retrieval has been popular now a day, as it improves the image search. Images are retrieved on the basis of its contents. Searching images on the basis of its title and metadata is not efficient. The need to have a versatile and general purpose content based image retrieval (CBIR) system for a very large image database has attracted focus of many researchers of information-technology-giants and leading academic institutions for development of CBIR techniques. So we have proposed new technique for image retrieval. In this paper the content based technique is studied in detail and using the available image features i.e. shape, and color distribution the images are searched. That technique helps to group the similar image contents in a group, in addition of the proposed technique is works on the basis query by image technique thus that make more promising outcomes from the retrieval.

Future Work

- The computational complexity in terms of time complexity is required to enhance because the time complexity of the system is increases with the amount of data
- The presented work is only works for the image based query for similarity computation that can also be implemented for text based query processing
- Enhancing the developed application to web-based CBIR system.

References

Aman Chadha and Sushmit Mallik (2012), Comparative Study and Optimization of Feature-Extraction Techniques for Content based Image Retrieval, *International Journal of Computer Applications (IJCA)*, Volume 52– No.20.

Sonit Singh, Aman Saini and Shamneesh Sharma (2013), Evaluation of Feature Extraction Techniques in Content Based Image Retrieval (CBIR) System, *International Journal of Electronics & Communication Technology (IJECT)*, Vol. 4, Issue Special - 2, PP. 81-83.

PritiMaheswary, Dr. Namita Srivastava (2009), Retrieval of Remote Sensing Images Using Colour & Texture Attribute, *International Journal of Computer Science and Information Security*, Vol. 4, No. 1&2.

P. S. Hiremath and Jagadeesh Pujari (2007). Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement, *International Journal of Computer Science and Security*, Volume (1): Issue (4).

Michele Saad (2008), Low-Level Color and Texture Feature Extraction for Content-Based Image Retrieval, *Final Project Report, EE 381K: Multi-Dimensional Digital Signal Processing*, Pages: 20-28.

Datta, R., Joshi, D., Li, J. and Wang, J. Z. (2008), Image retrieval: Ideas, influences, and trends of the new age, *ACM Computing Surveys*. 40(2), 1-60.

Kaur, Amanpreet, and M. D. Singh (2012), An overview of PSO-based approaches in image segmentation*, International Journal of Engineering and Technology,* 2.8: 1349-1357.

Romain Negrel, David Picard, Philippe-Henri Gosselin (2014), Dimensionality Reduction of Visual Features Using Sparse Projectors For Content-Based Image Retrieval, *IEEE International Conference on Image Processing,* Paris, France, PP. 2192-2196.

Ramadan, Rabab M., and Rehab F. Abdel-Kader (2009), Face recognition using particle swarm optimization-based selected features, *International Journal of Signal Processing, Image Processing and Pattern Recognition,* 2.2 : 51-65.

Nirmal, Shikha (2009), Content Based Image Retrieval Techniques, *Proceedings of the 3rd National Conference, Computing For Nation Development*.

Gurpreet Kaur and Mnupreet Kaur (2015), Content Based Image Retrieval using BPNN and K-Mean Algorithm, *International Journal of Science and Research (IJSR)*, Volume 4, Issue 1.