ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

Content Based Image Retrieval Using Improved Particle Swarm Optimization – K-Means Clustering With Support Vector Machine Algorithm

Prameela Devi. Chillakuru, Dr. Kumanan .T, Sarada Devi. CH

Research Scholar, Meenashi Academy of Higher Education and Research, Chennai, India

Professor, Dept. of CSE, Meenashi Academy of Higher Education and Research, Chennai, India

Research Scholar, Meenashi Academy of Higher Education and Research, Chennai, India

ABSTRACT: The feature extraction process is done by using IPSOKMC algorithm which is used to improve the extraction of important features from the given image database through best objective function. Image retrieval is performed through efficient SVM algorithm which is used to provide relevant image results. Experimentally it is proven that the proposed IPSOKMC with SVM is working efficiently with various aspects like precision, accuracy, recall, and F-measure.In this proposed work, image retrieval is done by using IPSOKMC with SVM algorithm. The proposed algorithm helps to detect most similar images for query image. In this work it is also shown comparison between various techniques with various aspects.

I. OVERVIEW OF CBIR

The CBIR systems create machine-interpretable descriptions of an image's physical characteristics such as color, texture, shape etc. These descriptions, known as extracted features, can then be compared by a measure of similarity. The similarity between a given query image and every image in the image archive is then computed by CBIR system, and finally ranks the images in the collection according to their degree of relevance to the user's query.

Even after designing several compact and efficient features vector for indexing each image, still there remains a wide gap between actual human perceptions model with that of feature based model used for CBIR [1]. This is mainly due to the fact that the design of feature vector is solely based on low level visual information (like color, texture, shape etc.) without taking into account the semantic information like spatial organization, context, events etc. To overcome this shortcoming and in trying to incorporate certain amount of visual perception characteristic in the feature vector design, people have proposed the relevance feedback mechanism in CBIR design with active participation of human observers. The typical CBIR system with relevance feedback is shown in Fig.1.1, where the building blocks without the dotted line box describe the general CBIR system and the building blocks with the dotted line describe the CBIR system with relevance feedback model.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

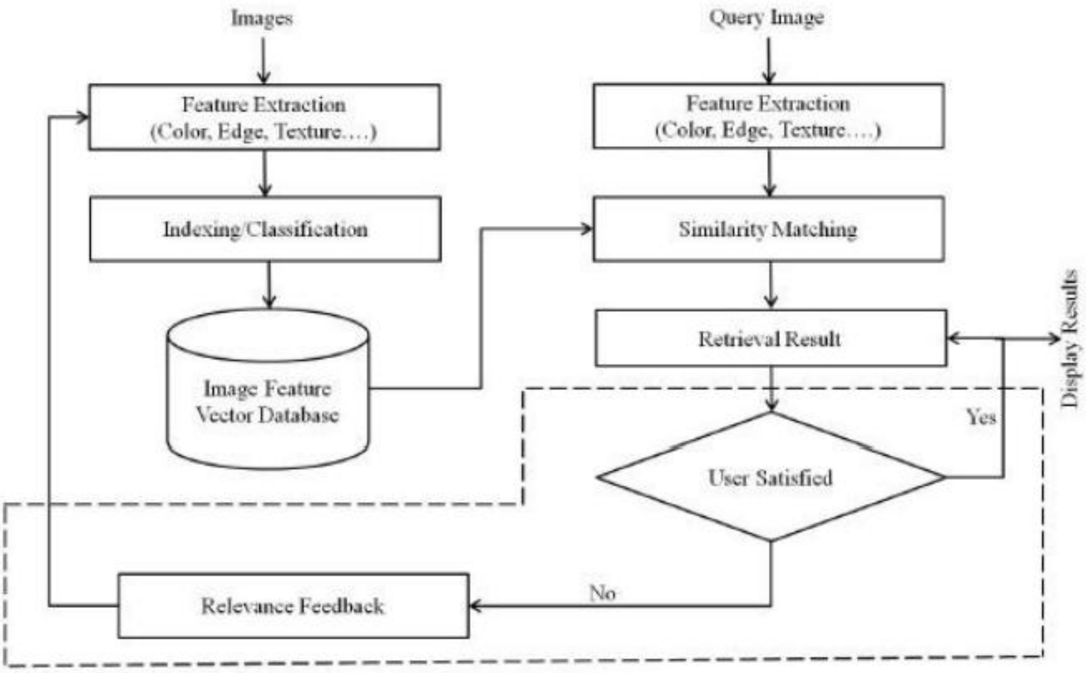

Figure 1.1 Generic Framework for Content-Based Image Retrieval

Images are being generated at an ever-increasing rate by sources such as defence and civilian satellites, military reconnaissance and surveillance flights, fingerprinting and mugshot-capturing devices, scientific experiments, biomedical imaging, and home entertainment systems. Application areas in which CBIR has a principal activity are numerous and diverse: art galleries and museum management, architectural and engineering design, law enforcement and criminal investigation, interior design, trademark and copyright database management, remote sensing and management of earth resources, scientific database management, weather forecasting, retailing, fabric and fashion design, medical imaging[2] and picture archiving and communication systems. With the recent interest in multimedia systems, CBIR has attracted the attention of researchers across several disciplines [3].

II. FEATURE EXTRACTION IN CBIR SYSTEM

One of the major building blocks of a CBIR system is the feature extraction block. In the feature extraction block, an image is represented by features which allow searching for images similar to a given image. A feature is defined to capture a certain visual property of an image. Color, texture, shape, etc. are the most widely used features in CBIR applications. Several CBIR algorithms have been developed using only single features [4]. A single feature normally describes a specific aspect of the image content; therefore multiple visual features are necessary for the general purpose CBIR [5]. Recent research work uses multiple visual features, which can improve the retrieval performance effectively. The most commonly used features in designing CBIR system are color, texture, shape, and key point descriptor and, brief discussion about each of which will be presented in the following paragraph. Besides that, descriptions about some commercially available CBIR system will also be presented.

Different color features in CBIR systems: The color signature of object in the foreground and also that of background makes it as one of important vision queue. Different texture features in CBIR systems: Although texture plays an important role in image processing and computer vision task, but unfortunately, there is no precise definition of the notion texture. It define texture is the variation of data at scales smaller than the scales of interest [10]. Generally speaking, textures are complex visual patterns composed of entities, or sub-patterns that have characteristic brightness,

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

directionality, coarseness, randomness, fineness, smoothness, granulation, slope, size, etc. Thus texture can be regarded as a similarity grouping in an image [11, 12]. However, the computation complexity and retrieval accuracy are the main challenging drawbacks of texture-based CBIR systems.

Different shape features in CBIR systems: Like color and textures, the shape is another fundamental unit of perception which carries semantic information of an image.

Shape analysis is a well-explored research area with many shape representation and similarity measurement techniques. Shape features can be extracted using various methods such as Fourier descriptors, B-splines, Curvature Scale Space (CSS), etc.

Different key point's descriptors in CBIR systems: It describes local information using key points of some image parts such as region, object of interest, edges, or corners. Recently, key point descriptors have shown their superiority for various types of applications in computer vision, as well as in CBIR paradigm.

.

III. PROPOSED METHODOLOGY

Initially the noises in the images are removed by histogram equalization. In the next step, feature extraction is done by Improved Particle Swarm Optimization – K-Means Clustering (IPSOKMC) algorithm, and retrieval of the medical images using Support Vector Machine (SVM) approach is proposed.

1.3.1 Noise Reduction using Histogram Equalization

Histogram equalization is as a contrast enhancement technique with the objective to obtain a new enhanced image with a uniform histogram. This can be achieved by using the normalized cumulative histogram as the grey scale mapping function. Modification to enhance the image quality using histogram can be categorised as three basic functions such as Histogram Equalization, Histogram stretching and non-linear mapping technique [13]. The idea here to enhance image quality by accessing each pixel(x, y) based on neighbouring pixels. During histogram representation the image produces contrast intensities that are not well distributed. In this step few of the adjustments have been made on the image so it produce a better contrast image. In histogram equalization the intensity values are distributed effectively. This helps areas on the image with low contrast to have a better or higher contrast before filtering steps.

Image histogram equalization actually deals with averaging and reduction of noise by adding certain other noisy images also. Image denoising can be done in the many ways includes one of the strong techniques called image filter. Filters are commonly used to adjust the rendering of an image, a background, or a border.

Once using histogram, image contrast is increased it can be passed for the filtering process. First average masking is used here for removing the blur of the image as filter. Median filter is used here for removing the noise as it includes salt-and-pepper noise. After median filter input image is passed to the high pass filter for denoising and sharpening. Generally in a blurred the slop at the edge is small with compared to the sharpened image. Therefore by increasing the slope can makes the image more sharpen.

1.3.2 Feature Extraction using Improved Particle Swarm Optimization-K Means Clustering Algorithm

Feature extraction includes CBIR system and meaning features extracted include text-based features such as keywords, annotations, and visual features such as color, texture, and shape. CBIR comparison low-level features of image extracted. In color feature, important features of images possibly recognized by humans are clear. As such, color is a property, i.e., light reflection aids information processing in the brain. Color implementation expressed variance in objects or places during the daytime. Therefore, color is extensively a visual feature used in CBIR that moderates the robust and simple processing.

Energy: It is also called angular second moment, homogeneity, and uniformity, which are image consistency measures. The energy value is small when gray-level intensities are close to each other. When all the matrix elements are more irregular, the value of energy is high

$$\text{Energy} = \sum_i \sum_j C_{\theta,d}^2(i,j) \qquad (1)$$

Entropy: It is the opposite of energy; thus, it has a lower value when the matrix is irregular. It has its highest peak when the matrix is uniform.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

Entropy $=-\sum_i \sum_j C_{\theta,d}(i,j) log C_{\theta,d}(i,j)$ (2)

Contrast: It is also called inertia and measures the difference moment of the matrix. The value will be high if the image has high local variation.

Contrast = $\sum_i \sum_j C_{\theta,d}(i-j)^2 C_{\theta,d}(i,j)$ (3)

Step 1: Choosing k initial cluster centers i.e., Z1 (1), Z2 (2)…Zk (k) that are randomly selected and are usually selected as the first K samples of the given sample set.

Step 2: Distributing the samples {x} by the k th iterative phase between the K cluster domains, using the following relation:

$x \in S_j(k) \quad if \quad \|x - z_j(k)\| < \|x - z_i(k)\|$ (4)

For all i=1, 2, 3… K; i≠j, Sj (k) denotes the sample set of cluster by Zj (K)

Step 3: Computing the new cluster centers, i.e. $z_j(k+1), j = 1,2..,k$ such that the sum of the squared distances of all points Sj (k) to the new cluster center is reduced. The new cluster center zj (k+1) computed for the performance index minimization. Minimization of performance index in simply the mean of Sj (k) gives the new cluster center as follows

$z_j(k+1) = \frac{1}{N_j}\sum_{x \in S_j(k)} x, \qquad i = 1,2,\dots k$ (5)

Where Nj is the number of Sj(k). Obviously, k-mean derivations from the clusters are sequentially updated mode.

Step 4: If zj(k+1)=zj(k) for j=1,2,…,K, the algorithm occurs while the procedure is terminated. Otherwise, one should return to step 2.

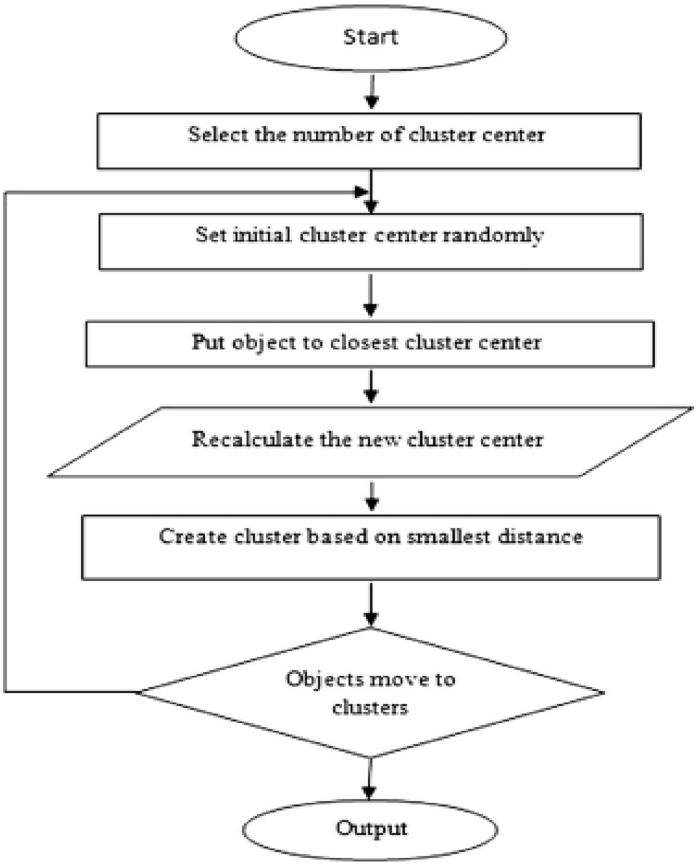

Figure 1.2 Flowchart of K-Means Clustering Algorithm

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

However, k-mean algorithm behaviour influenced the number of cluster centers specified, the initial cluster centers choice, the sample order taken as well as the data geometrical properties [15]. Even if this algorithm has no proof of merging, it produces acceptable results when the data exhibit specific is rather far from each other. This algorithm requires K values as well as different starting configurations. Flowchart of k-means algorithm is shown in Figure 1.2.

Step 1: Initialization: the velocity and position of every particle are randomly set within predefined ranges.

Step 2: Velocity updating: iteration of each agent and the velocities of them are updated with the model given below:

$$V_i^{k+1} = wV_i^{k+1} + c_1R_1(\text{pbest-}P_i^k) + c_2R_2(g\text{best-}P_i^k) \quad (6)$$

where P_i^k and P_i^k represent the velocity and position of particle i, exclusively. Step 3: Position updating: let a unit time interval between successive iterations, the positions of every particle updated with the given model below:

$$P_i^{k+1} = P_i^k + V_i^{k+1} \quad (7)$$

After the updating, Pi ought to be checked and limited to the allowed range.

Step 4: Memory updating: modernize pbesti and gbest when the condition is standard.

$$pbest_{i+1} = P_i \quad if \quad f(P_i) < f(pbest_i) \quad (8)$$

$$pbest_{i+1} = pbest_i \quad if \quad f(P_i) > f(pbest_i) \quad (9)$$

Step 5: Termination checking: the algorithm procedures repeat steps 2 to 4 until certain disorders such as a certain iteration number failure to make progress has been met. Therefore, after termination, the algorithm reports the values of gbesti and f (gbesti) as the solution.

IPSOKMC Clustering Algorithm

PSO is quick and effective to solve some of optimization problems, in the document clustering research area; it is potential to view the clustering problem as an optimization problem that locates the optimal centroids of the clusters rather than an optimal partition finding problem. This view offers us a chance to apply PSO optimal algorithm on the clustering solution. The PSO clustering can be divided into two stages, a global searching and local searching. The step of this algorithm can be reviewed as follows:

Step 1: Initial value for the position and velocity of particle is selected randomly from image dataset. Each particle is a potential solution for clustering. In the context of clustering, a single particle signifies the centroid of clusters. So, therefore, it particle initialized is given as

$$P_i = z_{i1}, z_{i2}, \dots, z_{ik} \quad (10)$$

Step 2: For each particle, a) Evaluate a particle fitness based on clustering criteria (distance between query image x and cluster center z). Thus, swarm definition of particle fitness as follows:

$$f(i) = \frac{\sum_{j=1}^{k} \sum_{\forall x_p \in S_J} (x_p - z_{ij})^2}{N_p} \quad (11)$$

Step 3: Repeat step 2 until one of the following termination conditions is achieved:

1. Iteration number exceeds a predefined maximum.
2. Once the change in the cluster centroids is negligible.
3. Once there is no change of cluster membership.

In IPSOKMC clustering algorithm PSO influences as a technique but when combined with other computational intelligence technique yields efficient and effective result. The motivation behind this approach is benefitting to both advantages of PSO and k-means algorithms. This algorithm used the ability of PSO in global search at the initial stage and rapid convergence of k-means to the local searching to find the final result.

The proposed algorithm is to retrieve images similar to the query image as the following step:

Input: query color image and dataset of images

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

Output: n images retrieved similar to the query image

Step 1: Select one color image as a query.

Step 2: Extract color and texture features from images in the dataset during offline stage and from query image through online stage.

Step 3: Create the feature vector that represents the content of images and stores it in the database.

Step 4: Cluster the feature vector of images in the database using proposed PSO-k-means clustering algorithm into different groups.

Step 5: Calculate the distance between the query image and the centroid of each cluster to find the smallest distance.

Step 6: Retrieve top n images belonging to the best cluster that similar to query image

1.3.3 Image Retrieval using SVM Algorithm

SVM works by the mapping of input space to the feature space. Feature space is defined as the space which is kept for the purpose of calculating similarity by usage of the kernel function. The main feature of this space is that here the linear separation is much easier [17]. Here the transformation of the unprocessed data is done into predetermined extent sample vectors.

For enhancing the performance of this classification it is possible to retrieve the nearby feature of the image as it will be helpful to attain the accuracy in the results [18]. The retrieval of all features of image on the basis of visual characteristics of image to the query is the primary aim of CBIR.

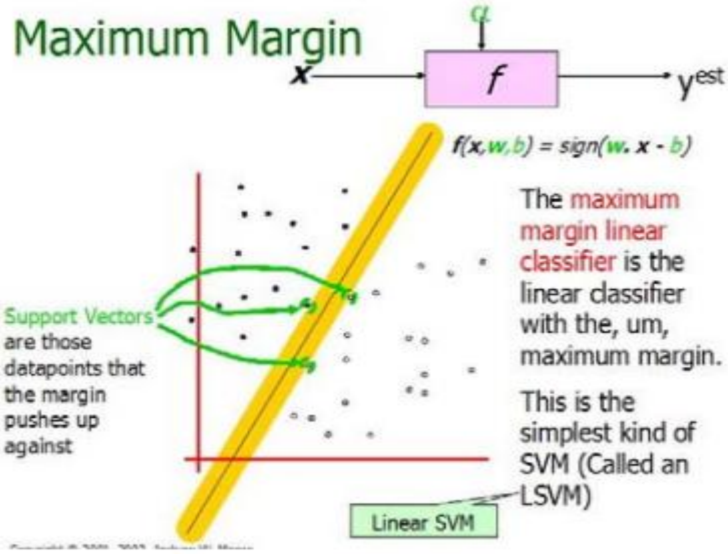

Figure 1.3 Illustration of Linear SVM

The above illustration 1.3 is the maximum linear classifier with the maximum range. In this context it is an example of a simple linear SVM classifier. Now we try to express the SVM mathematically and for this tutorial we try to present linear SVM. The goals of SVM are separating the data with hyper plane and extend this to non-linear boundaries using kernel trick.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

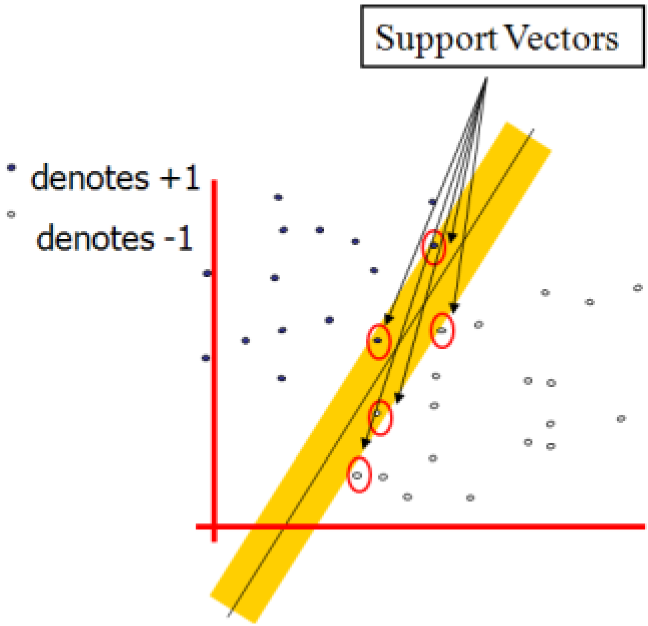

Figure 1.4 Representation of Support Vectors

Given a set of labelled training data and a Mercer kernel K, there is a set of hyperplanes that separate the data in the induced feature space F. This is a set of consistent hyperplanes or hypotheses known as version space. SVM image retrieval system employs a multi-resolution image representation. The main image features are color, shape, texture and similarity matrix.

V. EXPERIMENTAL RESULT

In this section let have a comparative analysis on existing HSV and EDBC technique with the proposed IPSOKMC with SVM algorithm with various aspects like precision, recall F-measure accuracy and error in terms of percentage.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

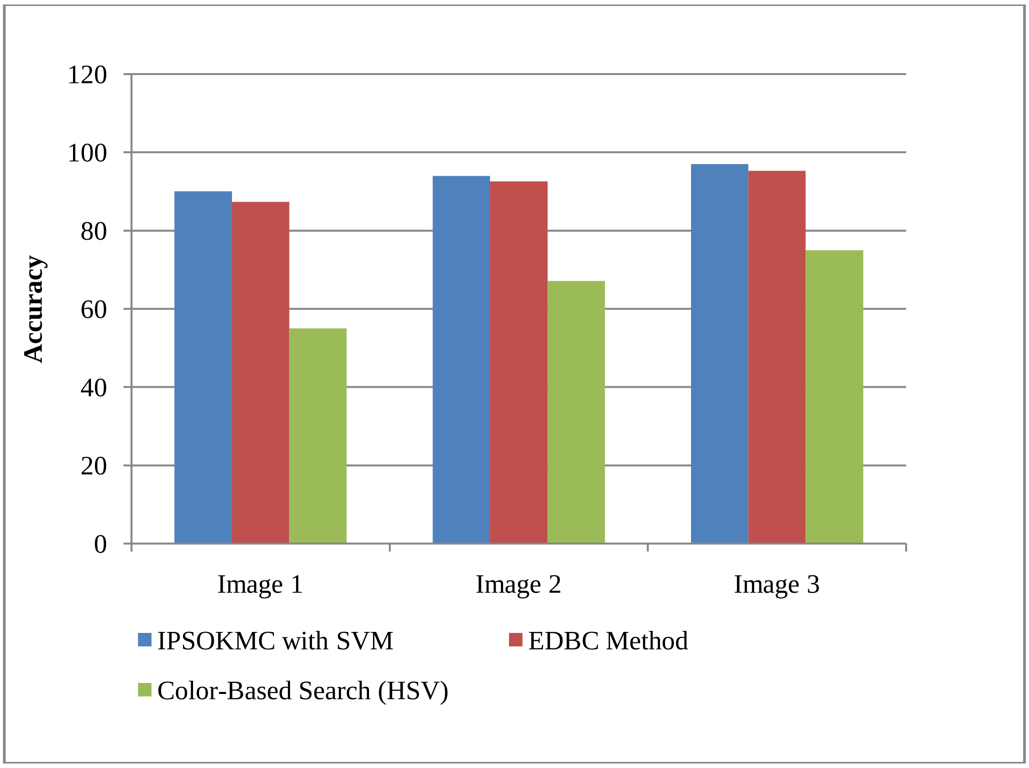

Figure 1.5 Accuracy

Figure 1.5 illustrates the experimental outcomes regarding accuracy for image 1, image 2 and image 3. The proposed IPSOKMC with SVM technique shows 93.66 % accuracy whereas existing EDBC and Color Based Search technique shows 91.75 and 65.72% respectively from the database.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

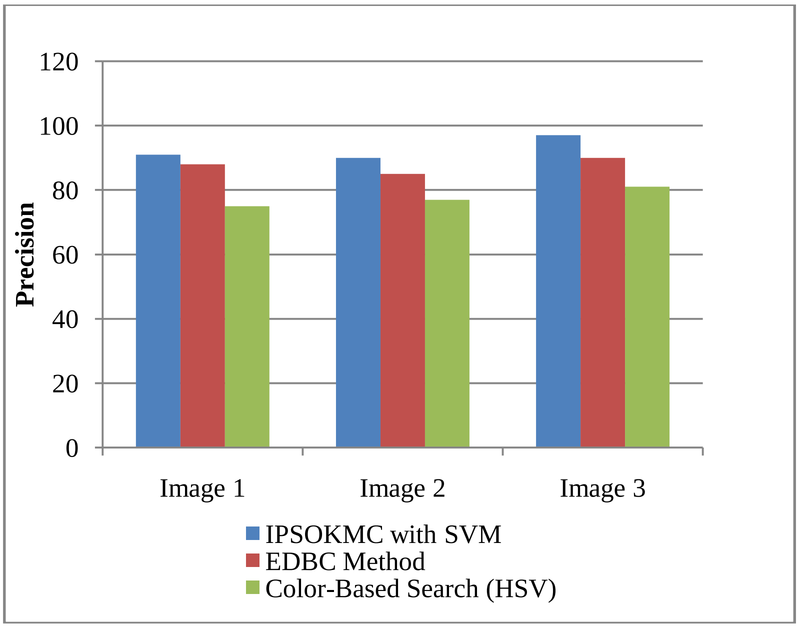

Figure 1.6 Precision

Figure 1.6 illustrates the experimental outcomes regarding precision for image 1, image 2 and image 3. The proposed IPSOKMC with SVM technique shows 92.66 % precision whereas existing EDBC and Color Based Search technique shows 87.6 and 77.65% respectively from the database.

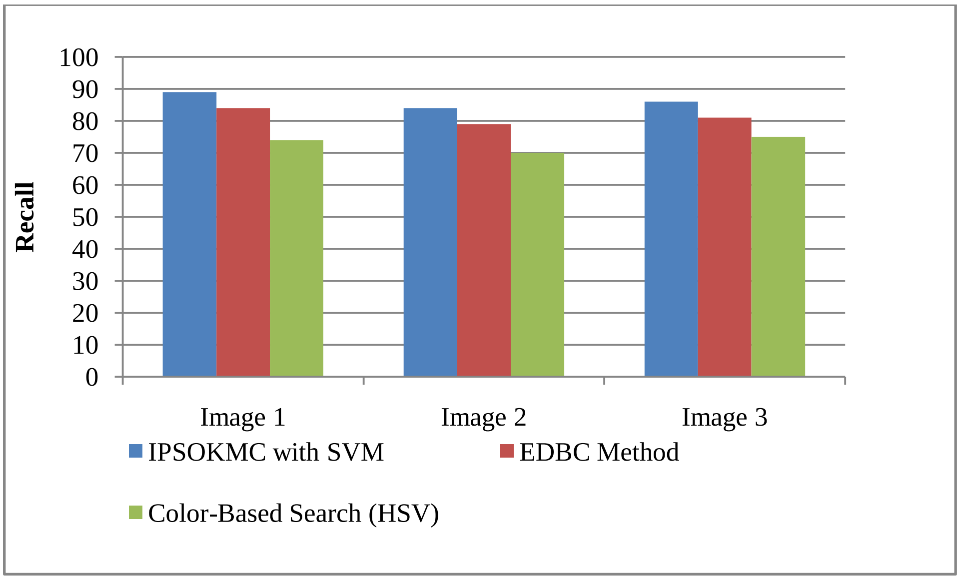

Figure 1.7 Recall

Figure 1.7 illustrates the experimental outcomes regarding recall for image 1, image 2 and image 3. The proposed IPSOKMC with SVM technique shows 86.33% recall whereas existing EDBC and Color Based Search technique shows 81.3 and 73.51% respectively from the database.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

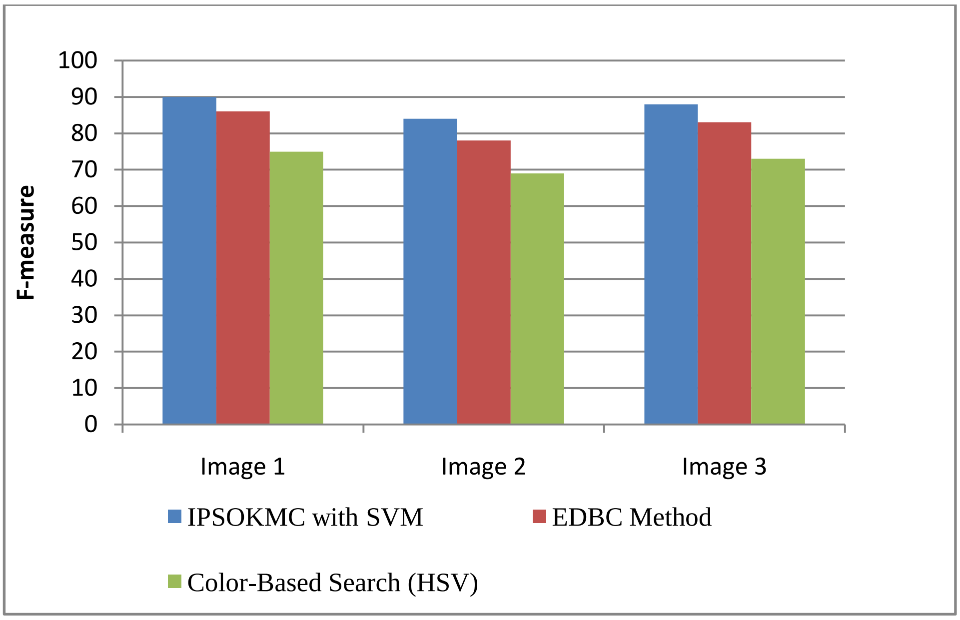

Figure 1.8 F-measure

Figure 1.8 illustrates the experimental outcomes regarding F-measure for image 1, image 2 and image 3. The proposed IPSOKMC with SVM technique shows 87.3% F-measure whereas existing EDBC and Color Based Search technique shows 82.3 and 72.31% respectively from the database.

REFERENCES

1. KWAK, DONG- MIN, et al. "Content- Based Ultrasound Image Retrieval Using a Course to Fine Approach." *Annals of the New York Academy of Sciences* 980.1 (2002): 212-224.
2. Lehmann, Thomas Martin, et al. "Content-based image retrieval in medical applications for picture archiving and communication systems." *Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation*. Vol. 5033. International Society for Optics and Photonics, 2003.
3. Murala, Subrahmanyam, R. P. Maheshwari, and R. Balasubramanian. "Directional local extrema patterns: a new descriptor for content based image retrieval." *International journal of multimedia information retrieval* 1.3 (2012): 191-203.
4. Zhang, Qi, et al. "Content-based image retrieval using multiple-instance learning." *ICML*. Vol. 1. 2002.
5. Datta, Ritendra, Jia Li, and James Z. Wang. "Content-based image retrieval: approaches and trends of the new age." *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*. ACM, 2005.
6. Omer, Ido, and Michael Werman. "Color lines: Image specific color representation." *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*. Vol. 2. IEEE, 2004.
7. Deng, Yining, et al. "An efficient color representation for image retrieval." *IEEE Transactions on image processing* 10.1 (2001): 140-147.
8. Liu, Guang-Hai, and Jing-Yu Yang. "Content-based image retrieval using color difference histogram." *Pattern recognition*46.1 (2013): 188-198.
9. Han, Ju, and Kai-Kuang Ma. "Fuzzy color histogram and its use in color image retrieval." *IEEE Transactions on image Processing* 11.8 (2002): 944-952.
10. Mirmehdi, Majid, and Maria Petrou. "Segmentation of color textures." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.2 (2000): 142-159.
11. Gómez, Walter, W. C. A. Pereira, and Antonio Fernando C. Infantosi. "Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound." *IEEE transactions on medical imaging* 31.10 (2012): 1889-1899.
12. Zhang, Jianguo, et al. "Local features and kernels for classification of texture and object categories: A comprehensive study." *International journal of computer vision*73.2 (2007): 213-238.
13. J.Alex Stark, "Adaptive Image Contrast Enhancement using Generalization of Histogram Equalization", IEEE transactions on Image Processing Vol. 9, No. 5, May 2000
14. Lam, Yau-King, Peter Wai-Ming Tsang, and Chi-Sing Leung. "PSO-based K-Means clustering with enhanced cluster matching for gene expression data." *Neural Computing and Applications* 22.7-8 (2013): 1349-1355.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

15. Dhanachandra, Nameirakpam, Khumanthem Manglem, and Yambem Jina Chanu. "Image segmentation using K-means clustering algorithm and subtractive clustering algorithm." *Procedia Computer Science* 54 (2015): 764-771.
16. YANG, H. C., ZHANG, S. B., DENG, K. Z., & DU, P. J. (2007). Research into a feature selection method for hyperspectral imagery using PSO and SVM. *Journal of China University of Mining and Technology*, *17*(4), 473-478.
17. Ramakrishna Reddy.Eamani and G.V.Hari Prasad, "Content-Based Image Retrieval UsingSupport Vector Machine in digital imageprocessing techniques", International Journal of Engineering Science and Technology (IJEST), ISSN: 0975-5462 Vol. 4 No.04 April 2012.
18. Ja-Hwung Su, Wei-Jyun Huang, Philip S. Yu, "Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns " IEEE, and Vincent S. Tseng, Member, Vol. 23, No. 3, March 2011