Efficient Classifying and Indexing for Large Iris Database Based on Enhanced Clustering Method

Emad Taha KHALAF[1], Muamer N. MOHAMMAD[2],

Kohbalan MOORTHY[1], Ahmad Taha KHALAF[3]

[1] Soft Computing & Intelligent System Research Group, Faculty of Computer Systems & Software Engineering, Universiti Malaysia Pahang, 26300, Kuantan, Pahang, Malaysia,
[2] The State Company for Internet Services, Ministry of Communications of Iraq, Baghdad, Iraq
[3] Faculty of Medicine, SEGi University, No. 9, Jalan Teknologi, Taman Sains Selangor, Kota Damansara, PJU 5, 47810 Petaling Jaya, Selangor, Malaysia
pcc13010@stdmail.ump.edu.my (*Corresponding author*),
muamer@ump.edu.my, kohbalan@ump.edu.my, tahakhalaf@segi.edu.my

Abstract: Explosive growth in the volume of stored biometric data has resulted in classification and indexing becoming important operations in image database systems. A new method is presented in this paper to extract the most relevant features of iris biometric images for indexing the iris database. Three transformation methods DCT, DWT and SVD were used to analyse the iris image and to extract its local features. The clustering method shouldering on the responsibility of determining the partitioning and classification efficiencies of the system has been improved. In the current work, the new Weighted K-means algorithm based on the Improved Firefly Algorithm (WKIFA) has been used to overcome the shortcomings in using the Fireflies Algorithm (FA). The proposed method can be used to perform global search and exhibits quick convergence rate while optimizing the initial clustering centers of the K-means algorithm. From the experimental results, the proposed method was indeed more effective for clustering and classification and outperformed the traditional k-mean algorithm. The Penetration Rates underwent reductions and reached the levels of 0.98, 0.13 and 0.12 for three different databases. Also, the Bin Miss Rates decreased to 0.3037, 0.4226 and 0.2019 for the investigated databases**.**

Keywords: Iris Biometric, Clustering, K-means Algorithm, Fireflies Algorithm, Computational Intelligence.

1. Introduction

Among the many biometric features, iris is deemed as important because it is used to verify a person by using distinct texture patterns, it is a rich in unique texture features to every individual [20][11]. The extracted features during enrolment process are stored in a high dimensional data, it becomes challenging to identify a person in large-scale biometric systems e.g. In United Arab Emirates a national border crossing security initiative has launched [26]. A large scale Indian UID system is undertaken to issue a unique ID card to every person using iris and fingerprint [18]. in United Kingdom biometric personal identification system based on iris is used to enter the automated barriers In certain airports [12]. Scalability is one of major problem facing biometric system, where the number of people to be enrolled into biometric system increases to billions this causes a challenge to identification system in terms of low response time, retrieval efficiency and high search, besides to accuracy [10]. However, Despite the rapid proliferation of large-scale databases, researchers focused only on accuracy within small databases. while they neglecting the scalability and speed issues to large-scale applications. where number of false acceptances grow geometrically with the increasing size of the database [26]. Meanwhile, time taken to claim identification is depends on the database size. So, there is a stringent need to partition the database to reduce the searching space. Clustering is an important for classifying and indexing approach because it is shouldering on the responsibility of determining the partitioning and classification success, where clusters are groups of data elements partitioned based on a distance function. The aim of clustering is to find an optimal partitioning for a group of unknown data. In this paper an efficient approach of classifying and indexing large visual databases based on an enhanced clustering method is proposed. The organizing of this paper as follow: Section 2 includes related work. In Section 3 the proposed modules are given along with the section references. Section 4 includes the experimental results with the discussion of the presented approach for three different kinds of databases. Last section presented the Conclusions.

2. Related Work

The hierarchical methodology was proposed in [16] to retrieve an iris image. An innovative method was used for indexing iris database. This method exploited two distinct features of iris image to extract the contents. In order to form the basis for indexing, the iris color was used. The texture feature of image was used for the

retrieving process from the indexed database. Nevertheless, the color indexing scheme is problematic as it is computationally complex (due to high dimensionality features). Also, most of the existing texture indexing and partitioning schemes use insufficient number of features for partitioning and classification.

Two methods have been proposed by Mukherjee and Ross [28]. In the first method, the IrisCode features were calculated by using the Gabor wavelet [9]. The dimensions of IrisCode features were large (up to 2048). It have been proposed row/column/block averaging with a principle component analysis (PCA) to reduce the dimensionality of features. The second method was called the signed pixel level difference histogram (SPLDH) analysis [28].

The Scale Invariant Feature Transform (SIFT) was used to index iris data by using texture features of iris image [27]. The number of key points as well as the dimension of features (124 dimensions) for every key point was found. Then, the key points based on geometric hashing had been indexed. the indexing method has been done using a large number of key points (around 100 key points) as well as a high dimensional vector of features. IrisCode (averaging of column/block/row, and PCA), SIFT, and SPLDH techniques are suffering from problems such as high computational load and unreliable statistics. Additionally, the method of indexing that depends on partitioning biometric data sets [28] is unable to separate the data sets uniformly, which might affect the searching outcome. Hence, these approaches have been found to be neither efficient nor accurate enough for practical applications.

An indexing technique was developed for large multimodal biometric database [15]. Iris was one of the traits focused in [15] and the feature vectors were normalized and the space is reduced to lower dimensional. The inner boundary of the iris was extracted using Circular Hough transformation [19] and the outer boundary was recovered using the intensity variation approach [22]. Then, the iris feature was transformed into rectangular block and the features were extracted using the Haar Wavelet Transform [30]. Later, feature normalization was done and the feature vectors were used to form the kd-tree. In order to find similar matching, the tree was traversed starting from the root. The

Euclidean distances (based on the search results) were then computed to identify the correct match.

An indexing approach based on the histogram of energy derived from the texture of iris was proposed [24]. The energy histogram was used for indexing iris database. The penetration rate was 0.63% and the bin miss rate was 43.6%.

Dey and Samanta [10] proposed a retrieval approach based on indexing method for iris biometric templates. The proposed method was developed based on the Gabor energy from different scales and orientations of the iris image it has been generated a 12-dimensional index key texture features. The values of index keys of all individuals were used to create the index space. From the index space and the index key of the query image, a candidate set was retrieved. After that, the retrieved candidates were ranked according to the occurrences.

Mehrotra proposed two approaches for indexing iris database [25]. In the first method, the Energy Histogram of DCT coefficients were used to form a B-tree. The second indexing method was based on Geometric Hashing of SIFT keypoints. This method used the energy histogram of re-ordered DCT coefficients. However, by using simple features (inadequate description of iris texture), false indexing and partitioning might be expected.

3. Proposed Method

It is anticipated that the response time of a query should not be based on the total number of stored templates in the database but on the similar templates to the query template. The database has to partition logically in order to reduce the searching space [8]. The proposed method has combined DCT, DWT and SVD techniques for selecting the local texture features of iris image. The proposed WKIFA method was developed for clustering the extracted data. The procedures of the proposed method are:

3.1 Iris Image Pre-processing

The stage of isolating the actual iris region in iris recognition is called pre-processing stage. For segmentation process, both Canny edge detection operator and circular Hough transformation were applied [9]. For normalization, the Daugman's rubber sheet model [14] was used where the

Iris disk was converted to a rectangular region with prefixed size. The image was enhanced by using the Contrast Limited Adaptive Histogram Equalization CLAHE algorithm [23][14].

3.1.1 Feature selection

In this method, the most discriminating features of iris were selected. Note, in the existing methods [10][24], only the total energy value is used for partitioning and indexing the iris database. The proposed algorithm of extracting the local features for each image is illustrated in Figure 1. The processes are detailed below:

3.1.2 Dividing image

The first step of feature selection is to divide the image I into 8x8 blocks. Due to the fact that analyzing textural features is easier in detailed sub-images [1], each block was treated individually in the proposed method in order to isolate the most relevant features of each block. Analyzing small blocks is useful for image processing and image compression [1].

3.2 Analysing image using three transformation methods (DCT, DWT and SVD)

Each block of image was transformed using three transformation methods, i.e. DCT, DWT and SVD. Both DCT and DWT methods were combined in order to secure the advantages of both algorithms. SVD was used to reduce the data dimension of the analyzed image while selecting the most important features. The steps of feature extraction are:

First step: DCT transformation was applied on each image (8x8 blocks). DCT is one of the best image transformation methods. It is an orthogonal transformation that transforms all image blocks from space domain into frequency domain. It also helps to decompose each block into spectral sub-bands of different importance (the image's visual quality is respected) [7][2]. Here, the 2-D DCT has been used and the general formula for a 2-D (8x8 blocks) DCT is defined as:

$$F(u,v) = \frac{C(u)C(v)}{4} \sum_{x=0}^{7}\sum_{y=0}^{7} f(x,y) \qquad (1)$$

$$cos\left[\frac{(2x+1)u\pi}{16}\right] cos\left[\frac{(2y+1)v\pi}{16}\right] \qquad (2)$$

where:

$$C(u), C(v) = \frac{1}{\sqrt{2}} \; for \; u,v = 0$$
$$C(u), C(v) = 1 \; otherwise \qquad (3)$$

Next step: By applying the 2-level DWT on each transformed 8x8 blocks, these blocks were decomposed to produce seven sub-bands [2], i.e. LL 2, HL 2, LH 2, HH 2, HL 1, LH 1 and HH 1. The benefit of using DWT for block size 8×8 is to reduce the computational complexity. Also, multi-resolution decomposition provides useful discrimination between textures. The wavelet transform is expressed mathematically as:

$$F(a,b) = \int_{-\infty}^{\infty} f(x)\Psi^*_{(a,b)}(x)dI \qquad (4)$$

Final step: was applied to a set of blocks with different sizes for BDCT (for all 2nd-level sub bands). The two orthogonal unitary matrices were obtained. A diagonal matrix constitutes the singular values of the matrix of sub-bands. In an image, the important feature can be expressed by a series of singular values along the diagonal direction. In the proposed indexing model, the 8×8 image blocks were decomposed into 7 sub-bands by the 2-level DWT. For uniform saving format, the local features of each sub-band of image can be expressed by applying the first two singular values. A is the matrix, rows is m, columns is n, and r is the rank ($r \le n \le m$) three matrices can be factored [29]:

$$A = U \sum V^T \qquad (5)$$

where U is the Matrix of $m \times m$ orthogonal matrix and V is the matrix of $n \times n$ orthogonal matrix, while $\sum$ is an $m \times n$ diagonal matrix and singular values (SV) on the diagonal with non-negative numbers.

The important features $F_{i,j}$ were selected. The seven sub-bands can be expressed as {*S1, S2, S3, S4, S5, S6* and *S7*}, where the sizes of *S1, S2, S3* and *S4* are 2×2. On the other hand, the sizes of *S5, S6* and *S7* are 4×4. In the existing method, only one energy value can be assigned for each sub-image [10][24]. Hence, it is inaccurate to express features of energy value for the 4×4 sub-bands of *S5, S6* and *S7*. In the proposed indexing model, the first two singular values (σ_1, σ_2) from the $\sum$ matrix were used to represent each sub-band, where the singular vector expresses the features for all sub-

band images. Note, the most important feature of an image can be expressed by the biggest singular vector value [31]:
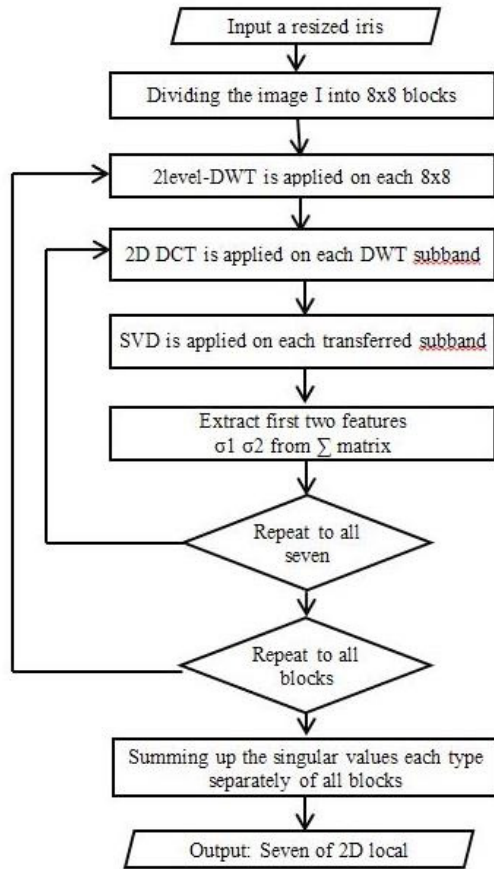

Figure 1. Block diagram of extracting local features for iris image

3.2.1 Create feature vector for iris image

Upon extracting the features of all blocks of each image, each block now contains two features from each sub-band. To produce the final feature vector (F) of an image, the singular value $\sigma 1$ of all blocks were summed up [24]. The similar procedure was applied for the singular value $\sigma 2$:

$$F = \sum S_i\left(x, y\right) \qquad (6)$$

The outcome was seven 2-D features $F_{i,j}$. Each feature pair represents a key used for partitioning process.

3.2.2 Partitioning and Classification Database

The next process is partitioning the database into number of groups in order to speed up the searching process. Therefore, the database was logically divided, where images having similar

texture were indexed together. Each sub-band was divided into groups.

3.3 Proposed Clustering Method (WKIAF)

Clustering is an important part in the proposed approach, shouldering on the responsibility of determining the partitioning and classification success. where clusters are groups of data elements partitioned based on a distance function. The aim of clustering is to find an optimal partitioning for a group of unknown data.

In cluster analysis, the weighted K-means clustering algorithm based on the improved Firefly Algorithm (WKIFA) method was used to partition the input data set into k partitions (clusters) [21]. A cluster is a set of objects that are „similar" between them and „dissimilar" to other objects. Clustering using distance functions (distance based clustering) is a most popular technique which uses to determine if objects are "close" to each other [34].

The clustering performance has been improved by the proposed method besides the convergence rate of the algorithm also improved. The original K-means clustering algorithm is uses the distance function for division. However, the Euclidean distance division criterion has some limitations, leading to weak clustering effect such as noise and emergence of abnormal points [3].

The idea of this proposed method is given in Figure 2. The firefly algorithm was used to obtain the optimal solution. As the K-means algorithm is based on the initial clustering center, it takes into account the impact of different sample data. In this algorithm, the Euclidean distance with weights was introduced, thus increasing the distinction between data attribute levels and reducing the impact of abnormal points. The firefly movement is essentially random. The perturbation method has been improved to achieve global optimization and to improve clustering accuracy and stability.

• Weight Calculation: With n sample data, X = {x_1, $x_2, x_3 ... x_n$} $\in R_m$ is the data to be clustered, where $x_i= (x_i,1, x_i,2, x_i,3 ... x_i,m)^T$ are m-dimensional vectors, In the data, the influence of each component is different, and the weight is defined:

$$\omega_{id} = x_{id} / \frac{1}{n}\sum_{d=1}^{n}x_{id} \qquad (7)$$

Where x_{id} represents the first component in the first sample, $\frac{1}{n}\sum_{d=1}^{n}x_{id}$ indicates sample the average of the sum of the dth components of each data object in the data, ω, shows the overall distribution of the sample data.

3.3.1 Weighted objective functions

The traditional K-means algorithm is to find a set of clustering centers $V=\{v_1,v_2,v_3 \ldots v_k\}$ which minimizes the objective function as:

$$J\left(X,V\right)=\sum_{j=1}^{k}\sum_{x_i \in G_j}d(x_i,v_j) \tag{8}$$

where: G_j on behalf of the first j a collection of samples in a category v_j right G_j within all samples x_i of the cluster center $d(x_i,v_j)$ indicates sample data x_i and clustering center v_j between the Euclidean distance, defined as follows:

$$d\left(x_i,v_j\right)=\left\|x_i-v_j\right\|=\sqrt{\sum_{l=1}^{m}\left|x_{il}-v_{jl}\right|^2} \tag{9}$$

$$v_j=\frac{1}{n_j}\sum x_i \tag{10}$$

where: v_j denotes j the center of the class, where $j = 1,2,3 \ldots k$, $i = 1, 2, 3 \ldots n,$ and n_j is a class v_j In the number of sample data x_i on behalf of v_j in the sample data.

For the traditional Euclidean calculation method, the weight is introduced ω, the relationship between each sample data and cluster centers is:

$$d_{\grave{u}}\left(x_i,v_j\right)=\left\|x_i-v_j\right\|_{\grave{u}}=\sqrt{\sum_{j=1}^{m}{}_{id}\left|{}_i-{}_j\right|^2} \tag{11}$$

The objective function is shown in formula 16.

$$J_c\left(X,V\right)=\sum_{j=1}^{k}\sum_{x_i \in G_j}\sqrt{\sum_{j=1}^{m}\omega_{id}\left|x_i-v_j\right|^2} \tag{12}$$

After introducing the attribute weights in formula 9, the traditional K-means algorithm is not changed, and the distance between the normal data and the clustering center is smaller, While the distance becomes larger between the center of the class and abnormal data, and makes the original distribution is not obvious, and the data samples are not easy to be classified improved objective function become more prominent and more suitable for the clustering. After the improvement not only improve the clustering accuracy, but also can effectively reduce the number of iterations algorithm to improve efficiency.

• Improved firefly algorithm calculation of attractiveness and disturbing way: The improved degree of attraction is calculated as:

$$\beta\left(r\right)=\frac{\beta_0}{1+\gamma \times r_{ij}^2} \tag{13}$$

When the firefly j attracts the firefly i to move it, the distance between them will be getting smaller and smaller. Using the equivalent infinitesimal replacement principle in higher mathematics, the simple fraction of the right side of the equal sign is used instead of the exponential function on the left side of the equal sign. The computational complexity is smaller and easier to implement, $e^{-yxr_{ij}} \approx \frac{1}{1+\gamma \times r_{ij}^2}$. The attractiveness of the firefly is calculated by formula 13, and the formula for updating the firefly position is shown in 14:

$$x_{i+1} = x_i + \frac{{}_0^2}{1+{}^3 \times r_{ij}^2} \times \left(x_i - v_0\right) + a \times \left(\text{rand} - 0.5\right) \tag{14}$$

where: v0 is the current optimal clustering center, β, γ, α, rand, the meaning of the same as before.

$$v_0 = \frac{1}{n_i}\sum_{y \in \Gamma_i}y \tag{15}$$

where n_i is the first Γ_i the number of data in the cluster, y representing the data value.

In the traditional firefly algorithm, the fireflies are attracted to the relatively bright firefly, where the disturbance items $\alpha \times(rand\text{-}0.5)$ can increase the algorithm search area. This could avoid premature fall into the local optimum and increase the algorithm's local search ability; however, it would decrease the convergence speed and the stability. When the objective function value is large, the perturbation effect is not apparent, and fluctuation can be found near the local optimal value.

3.3.2 Location of fireflies update

After introducing the perturbation operator $\alpha \times rand \times(x_i\text{-}v_0)^2$, the position update formula of the firefly can be expressed as:

$$x_{i+1} = x_i + \frac{{}_0^2}{1+{}^3 \times r_{ij}^2} \times \left(x_i - v_0\right) + a \times \text{rand} \times \left(x_i - v_0\right)^2 \tag{16}$$

Formula 16 is based on the optimal class center is also a random disturbance. The brightest firefly moves according to formula 17.

$$x_{*+1} = x_* + a \times rand \times \left(x_i - v_0\right)^2 \qquad (17)$$

Where x_* represents the location of the brightest fireflies currently, v_0 same as above. In the firefly algorithm, according to Formula 16 improved, can effectively avoid the most Bright firefly random movement, easily lead to the algorithm in the local optimal value or the most global and the convergence speed is slow and the optimization accuracy is reduced.

• Basic steps of weighted K-means algorithm based on improved firefly:

Step 1: Parameters such as the number of clusters k, the population size N, the maximum attractiveness, the light intensity β_0, the absorption coefficient γ, the step factor α, the maximum number of iterations T_{max}, the iteration stop threshold ε are initialized.

Step 2: A point k is randomly selected as the initial position of the firefly. The distance from each sample point to each cluster center is calculated from formula 11;

Step 3: According to the calculated distance, the sample points are sequentially divided into the category with the closest cluster center;

Step 4: According to the initial clustering results obtained in Step 3, the brightness of the firefly is calculated from formula 12.

Step 5: If $I_i > I_j$ the firefly's objective function j is small, indicating that the firefly is in good location and the firefly j would be attracted. The size of the movement is determined from formula 13. Formulae 16 and 17 are used to update the location of the firefly;

Step 6: Update the location, re-clustering, and update the firefly brightness, re-record the brightest firefly brightness, location, clustering results;

Step 7: Reaches the maximum number of iterations or reaches the threshold, the algorithm is stopped, otherwise go to step 4;

Step 8: Output the results. Figure 2 shows the basic Flowchart of the algorithm.

E. Proposed Searching and Retrieving technique: The searching and retrieving are enhanced, which could turn into a bottleneck for the biometric system, especially when its response time is affected. The proposed method is based on Half-Searching method, B-tree and a parallel search technique to improve the search process within each features group.
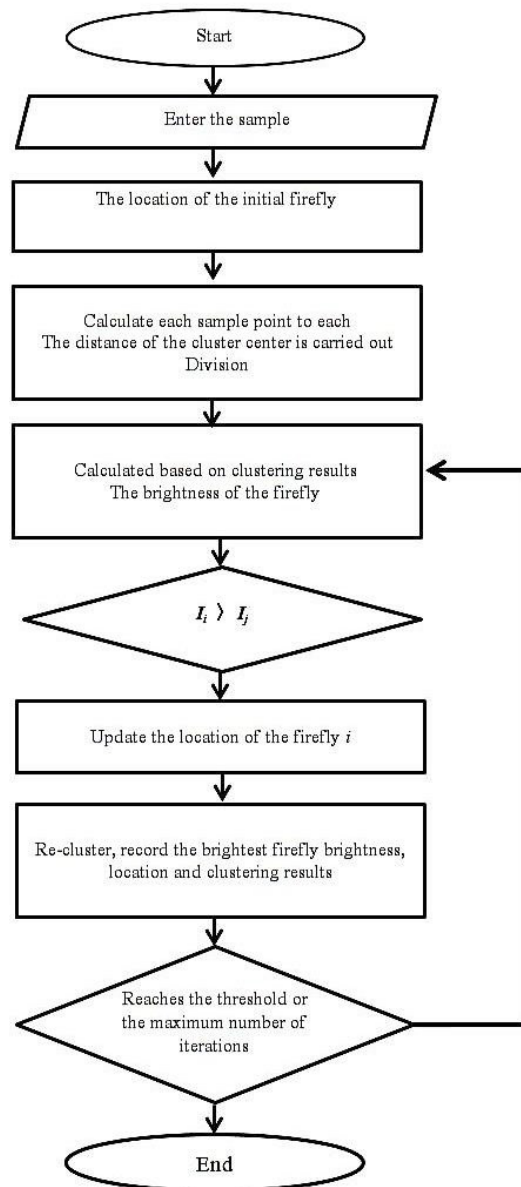

Figure 2. The clustering method (WKIFA)

3.3.3 The proposed database arrangement method

The steps as follow:

Step 1: For each image identify a global key, the key values are consist of the groups numbers which is include the subbands of the image, key

values combined together according to Morton order traversal which is more accurate [24][2].

Step 2: In order to preparing the database for searching processes, each group is ranking according to ascending order based on the first singular value σ_1 (from smallest to largest).

Step 3: After ranking all the groups, each group is divided into two Bins (Bin1 include σ_1 features & Bin2 include σ_2 features).

Step 4: Distribute the database into two B-tree structures.

Step 5: Traverse the two B-trees using the generated global key.

Step 6: Store the template at the leaf node of the two B-trees.

After the database is arranged the searching can be start. The basic steps of searching and retrieving method are as follow:

Input query image.

Step 1: Extract the local features.

Step 2: Appling K-means to find global key.

Step 3: Traverse at same time the two B-trees using the global key to find the specific two groups.

Step 4: Apply Half-Searching to find the candidate features inside the groups. (The first feature value uses for matching and the second one for correction).

Step 5: Retrieve all candidates.

Step 6: Perform comparison between the query image with the retrieved candidates.

Step 7: Find the probable match.

3.4 Experimental Results

To verify the validity and feasibility of the clustering algorithm, the experiment mainly includes two parts: using the multi-peak function of [33] to test the validity and feasibility of the firefly algorithm to find the initial clustering center. The WKIFA algorithm with the traditional K-means, FA algorithm and [34] Algorithm in the standard UCI dataset on the simulation experiment, through the experimental results compared to the advantages of this algorithm. The

composition of the experimental dataset is shown in Table 1. For easy and efficient operation, the UCI dataset is first standardized and normalized: the normalized formula is as follows (where the mean is used):

$$x^* = \frac{x_i - \ddot{x}_i}{S_i} \tag{18}$$

where $\ddot{x}_i$ the mean of the data, S_i is the standard deviation of the data.

Table 1. Composition of the experimental sample data set

Dataset	Num. of classes	Properties dimension	Dataset size
iris	3	4	150
wine	3	13	178
seed	3	7	210
Glass	6	9	214
Hayes-Roth	3	4	132
New-thyroid	3	5	215

The results of the proposed method show more effective for clustering and classification stage of the system and outperforms to the traditional k-mean. The initial clustering centers of traditional K-means algorithm has been optimized using the FA which has power ability of global search and quick convergence rate. As the same time, the defects made by noise data and other uncertainties have been reduced using a kind of weighted Euclidean distance, this led improve in the clustering processing which shouldering on the responsibility of determining the success and efficiency partitioning and classification of the system.

Table 2. Mean clustering results of the algorithm

Dataset	K-means	Yu, et al. Alg.	FA Alg.	WKIFA Alg.
Iris	87.93	90.56	91.13	92.16
Wine	56.85	70.23	71.36	72.15
Seed	86.97	88.07	88.89	90.46
Glass	54.05	57.18	57.46	63.12
Hayes-Roth	77.32	81.06	79.25	82.35
New-thyroid	72.34	79.63	78.12	80.28

Table 3. Average number of iterations of the algorithm

Dataset	K-means	Yu, et al. Alg.	FA Alg.	WKIFA Alg.
Iris	7	4	3	3
Wine	10	9	8	7
Seed	9	7	7	6
Glass	11	9	10	7
Hayes-Roth	9	6	8	5
New-thyroid	10	7	8	6

In order to take all possibilities into consideration such as illumination, scaling and rotation noise, different types of databases have been used to test the performance of the proposed iris biometric indexing system. The used database is CASIA-IrisV3-Interval (CASIAV3I) [5], number of subjects 249 in the database with 2655 of images from right and left eyes, and CASIA-IrisV4-Thousand (CASIA V4T) [6], the total of images is 54,607, it is contains six kinds of subsets: over than 1,800 genuine subjects in addition to 1,000 virtual subjects. The BATH University Database [4] consists of 2000 iris images from 50 subjects each from right and left eye. The database is collected by IITK [13] comprised of over 1800 iris images taken from 600 subjects. The performance of a proposed partitioning, classification and indexing systems are usually measured in terms of efficiency or Penetration Rate (PR), accuracy or Bin Miss Rate (BM), and speed or computational complexity [17]. For performance indicators each kind of biometric database gives quite different measurements. So, the comparison for performance between different algorithms should be based on the same database. The Penetration Rate (PR) is the percentage on an average of total database to be scanned on every search. Lower PR mean efficient the system. BM is obtained by counting the number of genuine biometric templates which have been miss-placed in a wrong class [32], as higher accuracy and efficiency are measured by lower error rate and lower penetration rate [17]. The final results have been evaluated and compared with the existing results, the comparison of the final results of the indexing system using traditional k-mean algorithm and WKIFA algorithm in term of Bin Miss Rate and Penetration Rate as shown in Table 4, the results in the table show the PR, BM and the retrieved candidates (classes) when the number of subbands are changing for the databases CASIA, subbands is changing number of classes at leaf node also changes. Figure 3, Figure 4 and Table 4 show the comparison of the final results of the indexing system using traditional k-mean algorithm and WKIFA algorithm in term of BM and PR for the databases CASIA, BATH and IITK.

4. CONCLUSION

The system has been tested on publicly available databases, the experimental results showed efficiency of the indexing system by a considerably low penetration rate 0.98%, 0.13% and 0.12% and lower bin miss rate of 0.3037%, 0.4226% and 0.2019% compared to the existing schemes for CASIA, BATH and IITK iris databases respectively, where an efficient classification and indexing method have low miss rate and low penetration rate. The results of the new clustering algorithm WKIFA showed that the proposed method is more effective for clustering and classification stage of the system and outperforms to the traditional k-mean, where the performance has been improved by reduced the Penetration Rate into 0.98, 0.13 and 0.12. Besides improving the accuracy by reduce the Bin Miss Rate into 0.3037, 0.4226 and 0.2019 for the mentioned database.

Table 4. The performance of system using WKIFA algorithm for CASIA database

| Key length	Indexing method using traditional K-means			Indexing method using WKIFA		
	Class	BM	PR	Class	BM	PR
1	4	0.0122	58.04	3	0.00231	51.95
2	11	0.0128	20.41	9	0.0097	18.01
3	28	0.0920	9.34	26	0.0742	8.107
4	55	0.1231	7.22	43	0.106	6.381
5	98	0.2191	3.84	90	0.1981	4.104
6	142	0.2700	2.28	138	0.237	1.067
7	199	0.3037	0.98	178	0.2604	0.131
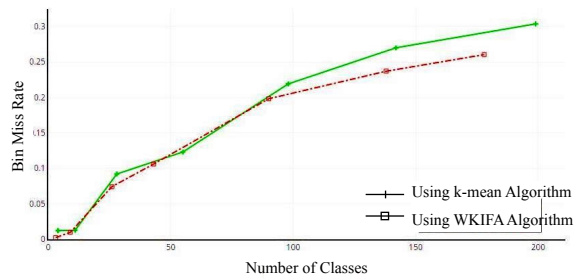
(a)
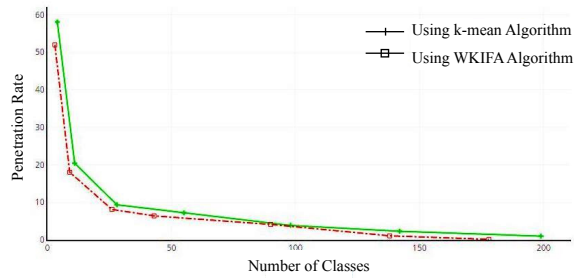
(b)

Figure 3. Comparison the performance when using the two algorithm for CASIA
(a) Bin Miss Rate, (b) Penetration Rate
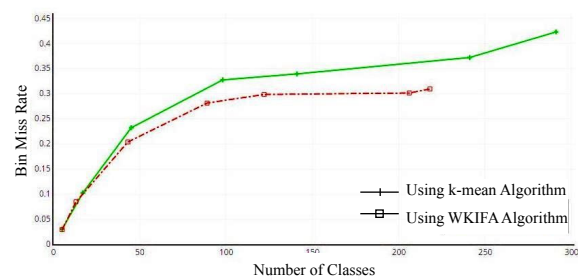
(a)
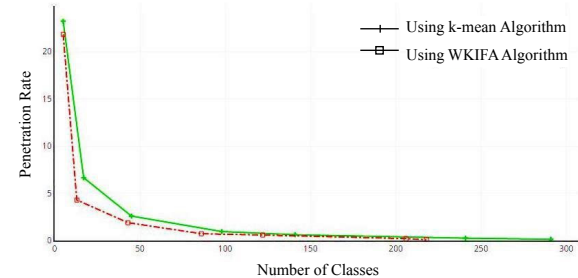
(b)

Figure 4. Comparison the performance when using the two algorithm for BATH
(a) Bin Miss Rate, (b) Penetration Rate

ACKNOWLEDGMENTS

REFERENCES

1. Abyoto, R. W., Wirdjosoedirdjo, S. J. & Watanable, R. G. (1998). Unsupervised texture segmentation using multiresolution analysis for feature extraction, *J Tokyo Univ. Inform. Sci.*, *2*(9), 49-61.

2. Ahmed, N., Natarajan, T. & Rao, K. R. (1974). Discrete Cosine Transform, *IEEE Trans. Computers*, 90-93.

3. Amaricai, A. (2017). Design Trade-offs in Configurable FPGA Architectures for K-Means Clustering, *Studies in Informatics and Control*, *26*(1), 43-48.

4. BATH Iris Database. (2004). Retrieved 2 April 2015, from: <http://www.bath.ac.uk/elec-eng/pp./sipg/>.

5. Center for Biometrics and Security Research. (2002). *Cbsr.ia.ac.cn*. Retrieved 4 May 2017, from: <http://www.cbsr.ia.ac.cn/IrisDatabase.htm>.

6. Center for Biometrics and Security Research. (2010). *Cbsr.ia.ac.cn*. Retrieved 5 May 2017, from: <http://www.cbsr.ia.ac.cn/ china/ Iris%20Databases%20CH.asp>.

7. Chen, M., Zhang, Y. & Lu, C. (2017). Efficient architecture of variable size HEVC 2D-DCT for FPGA platforms, *AEU - International Journal of Electronics And Communications*, *73*, 1-8.

8. Claramunt, C., Schneider, M., Wong, R. et al. (2015). *Advances in spatial and temporal databases*, 1st ed., 522.

9. Daugman, J. G. (1993). High confidence visual recognition of persons by a test of statistical independence, *IEEE Transactions on Pattern Analysis & Machine Intelligence, 15*(11), 1148-1161.

10. Dey, S. & Samanta, D. (2012). Iris Data Indexing Method Using Gabor Energy Features, *IEEE Transactions on Information Forensics And Security*, *7*(4), 1192-1203.

11. Gadde, R., Adjeroh, D. & Ross, A. (2010). Indexing Iris Images Using Burrow Wheelers Transform. In *Proc of IEEE International Workshop on Information Forensics and Security*, Dec 2010.

12. Gorodnichy, D., Yanushkevich, S. & Shmerko, V. (2014). Automated border control: Problem formalization. In *Proceedings of the IEEE Symposium on Computational Intelligence in Biometrics & Identity Management* (pp. 118-125).

13. Iris Database. *Cse.iitk.ac.in*. Retrieved 24 April 2017, from: <cse.iitk.ac.in/users/rajeshb>.

14. Ismail, A. I., Ali, H. S. & Farag, F. A. (2015). Efficient enhancement and matching for iris recognition using SURF. In *2015 5th National Symposium (NSITNSW)* (pp. 1-5). IEEE.

15. Jayaraman, U., Prakash, S. & Gupta, P. (2009). An efficient technique for indexing multimodal biometric databases, *International Journal of Biometrics*, *1*(4), 418.

16. Jayaraman, U., Prakash, S. & Gupta, P. (2012). An efficient color and texture based iris image retrieval technique, *Expert Systems With App.*, *39*(5), 4915-4926.

17. Jiang, X. D. (2009). Fingerprint Classification, *Encyclopaedia of Biometrics*, 439-446. Springer.

18. Kavati, I., Prasad, M. V. & Bhagvati, C. (2017). Search Space Reduction in Biometric Databases, *Developing Next-Generation Countermeasures for Homeland Security Threat Prevention*, 236-262.

19. Kerbyson, D. J. & Athertonm, T. J. (1995). Circle detection using Hough transform filters. In *International Conference on Image Processing and its Applications* (pp. 370-374).

20. Khalaf, E. T., Mohammad, M. N., Moorthy, K. & Abdulkareem, R.. (2017). Biometric Template Protection based on Hill Cipher Algorithm with Two Invertible keys. In *5th International Conference on Software Engineering & Computer Systems*, Langkawi, Malaysia (pp. 22-24).

21. K-means clustering. (2017). *En.wikipedia.org*. Retrieved 30 April 2017, from: <http://en.wikipedia.org/wiki/Kmeans _clustering>.

22. Ma, L., Tan, T. N., Wang, Y. H. & Zhang, D. (2004). Local intensity variation analysis for iris Recognition, *Pattern Recognition*, *37*(6), 1287-1298.

23. Martíne, P. & Ramos, P. (2014). Feature Extraction Using SIFT with a Preprocessing by Adding CLAHE Algorithm to Enhance Image Histograms. In *International Conference on Mechatronics. Electronics and Automotive Engineering*.

24. Mehrotra, H., Srinivas, B. G., Majhi, B. & Gupta, P. (2009). Indexing Iris Biometric Database Using Energy Histogram of DCT Subbands, in Ranka S. et al. (eds.), *Contemporary Computing. IC3 2009. Communications in Computer and Information Science*, 40. Berlin, Heidelberg, Springer.

25. Mehrotra, H. (2010). Iris Identification using Keypoint Descriptors and Geometric Hashing, *Master's thesis*. NIT Rourkela.

26. Mehrotra, H. & Majhi, B. (2013). Local feature based retrieval approach for iris biometrics, *Front. Comput. Sci.*, *7*, 767–781.

27. Mehrotra, H., Majhi, B. & Gupta, P. (2010). Robust iris indexing scheme using geometric hashing of SIFT keypoints, *Journal of Network And Computer Applications*, *33*(3), 300-313.

28. Mukherjee, R. & Ross, A. (2008). Indexing iris images. In *International Conference on Pattern Recognition (ICPR 08)* (pp. 1-4).

29. Nair, S. & Aruna, P. (2015). Comparison of DCT, SVD and BFOA based multimodal biometric watermarking systems, *Alexandria Engineering Journal*, *54*(4), 1161-1174.

30. Rafael, C. Gonzalez & Richard, E. Woods (2001). *Digital Image Processing*, 220-243.

31. Sadygov, R. (2014). Use of singular value decomposition analysis to differentiate phosphorylated precursors in strong cation exchange fractions, *ELECTROPHORESIS*, *35*(24), 3498-3503.

32. Wayman, J. (1999). Error rate equations for the general biometric system, *IEEE Robotics & Automation Magazine*, *6*(1), 35-48.

33. Yang, X. S. (2010). *Nature-inspired metaheuristic algorithms*. Luniver press.

34. Yu, H., Cheng, X., Jia, M. & Jiang, Q. (2013). Optimized k-means clustering algorithm based on artificial fish swarm. In *Proceedings of the International Conference on Mechatronic Sciences* (pp. 1783-1787).