Abstract—The Iris recognition has become as one of the promising biometrics feature in human identification system. In addition to conventional IRIS capture challenges, IRIS code searching and similarity matching of IRIS code has also become a challenging task in IRIS code VLOB (Very Large Object database). Hamming Distance (HD), a sequential similarity matching technique, proves inefficient for VLOB Iris code database objects. With technology growth the bit granularity has improved. This has resulted in more accuracy due to higher bit density. This high granularity leads to variable size IRIS codes rather than fixed sized IRIS codes hence adaptively is one major challenge. This paper proposes a novel, adaptive IRIS code database organization and searching algorithm proposing the improvement in template similarity matching process. The algorithm reduces the redundant IRIS code comparisons resulting in faster searching speed. For the IRIS that has accumulated the cataract, HD methods shows major failure results. The proposed IRIS code organization algorithm shows better performance over conventional HD method of template matching and IRIS code selection for template similarity matching.

Keywords- Database Organization, Iris codes Descriptors, Iris Code Searching , and Iris Code Similarity.

I. INTRODUCTION

With introduction of Internet applications and automation, the large information is searched far more quickly than our ability to process it. Iris recognition is gaining large-scale focus in biometrics recognition systems.

Objective: My contribution is to proposed and present Intelligence iris searching technique iris database organization to minimized redundant searches and hence improving the iris searching efficiency.

The IRIS recognition system has rapidly advanced from prototype stage to the patented algorithm in recent years due to higher accuracy and reliable recognition [1]. The conventional iris recognition system has steps that include: image capture, pre-processing, iris detection and feature extraction, iris code formation and database formation. This is followed by template matching for iris recognition. One of the popular template matching methods is using Hamming Distance (HD) or HD method [1][2][3]. The number of comparisons carried out for template matching depends first; upon the database size; and second, the bit comparisons needed for HD calculation [2].

Due to broad-spectrum reliability of iris recognition the ISO standard 19794- 6 has incorporated the iris data.

Wildes presented use of normalized correlation on iris image that can be used in identification process. Boles and Boash investigated two dissimilarity functions: learning and classification resulting in fast processing but high error rate. Tan et al. proposed nearest feature line with the constraint of eye lead occlusion problem. Ma et al. presented two different techniques such as XOR operator and weighted Euclidean and hamming distances but proved failure in the case of occluded iris. Lim et al. presented LVQ neural network for matching purpose but resulted in poor recognition rate [3][4][5][6]. The IRIS code is binary information. Number of theory exists for similarity matching of binary information. In information theory, the Hamming distance between two strings of equal length is the number of positions at which the corresponding symbols are different.

The correlation coefficient, (p) indicates the strength and direction of a linear relationship between two random variables. In information theory and computer science, the Levenshtein distance is a metric for measuring the amount of difference between two sequences. It can be considered a generalization of the Hamming distance, which is used for strings of the same length and only considers substitution edits. This research paper proposes a novel algorithm by organizing stored database information of IRIS codes that will reduce the redundant search iteration by method of elimination. The remaining paper is organized in four sections. The section II
gives the work outline and section III discusses the research algorithm in details. The section IV compares the algorithm with HD algorithm and section V presents the conclusions and future work. II. WORK OUTLINE

Conventional steps carried out by any iris recognition system are shown in Fig. 1[4]. The applications using iris recognition system such as identity verification system demand very large number of comparisons [3]. Equation (1) presents hamming distance calculations that are done for determining the deviation of iris template with the database iris codes.

\[
HD = \frac{\|code_A \otimes code_B \cap maskA \cap maskB\|}{\|maskA \cap maskB\|}
\]

With growth in technology and Internet data capture and storage has become distributed. The bit granularity of the capture has set new challenges of variable size IRIS codes. Variable size IRIS codes and voluminous size of database searching for IRIS code similarity matching are becoming major challenges rather than areas of the iris that are obscured by eyelids, or by eyelashes or reflection on iris are detected and prevented and applicable research results from influencing the iris comparison through bit-wise mask function [5][7][8][9]. The iris code contains the phase data that is XORed to detect bit agreement and deviations. The bits to be considered are first selected by (\) ANDing with mask functions associated with test iris (A) and the template iris (B). The proposed algorithm reorganizes the iris code useful for fast comparisons.

The iris database can contain training set of iris codes for a person resulting into huge number of iris codes in the database. The proposed algorithm applies method of elimination. This result in eliminating iris codes before comparison that are not favorite choices. The iris code set shows regional changes due to very common disease like cataract. This is resulting in failure of conventional iris code recognition. The proposed reorganization of iris code has application in recognition of iris with cataract. The selection of most favorable iris templates amongst number of available templates is the major of the research.

This is resulting in failure of conventional iris code recognition. The proposed reorganization of iris code has application in recognition of iris with cataract. The selection of most favorable iris templates amongst number of available templates is the major of the research. The research uses standard iris images from UBIRIS, with growth in technology and Internet data capture and storage has become distributed. The bit granularity of the capture has set new challenges of variable size IRIS codes. Variable size IRIS codes and voluminous size of database searching for IRIS code similarity matching are becoming major challenges rather than areas of the iris that are obscured by eyelids, or by eyelashes or reflection on iris are detected and prevented and applicable research results from influencing the iris comparison through bit-wise mask function [5][7][8][9]. The iris code contains the phase data that is XORed to detect bit agreement and deviations. The bits to be considered are first selected by (\) ANDing with mask functions associated with test iris (A) and the template iris (B). The proposed algorithm reorganizes the iris code useful for fast comparisons.

The iris database can contain training set of iris codes for a person resulting into huge number of iris codes in the database. The proposed algorithm applies method of elimination. This result in eliminating iris codes before comparison that are not favorite choices. The iris code set shows regional changes due to very common disease like cataract. This is resulting in failure of conventional iris code recognition. The proposed reorganization of iris code has application in recognition of iris with cataract. The selection of most favorable iris templates amongst number of available templates is the major of the research. The research uses standard iris images from UBIRIS.
The iris carries essential information required for biometrics recognition. It lies between the sclera and the pupil. The iris is separated from the eye image and the pupil is filtered using low pass filter. The iris inner and outer boundaries are located by finding the edge image using the Canny edge detector. The Canny detector works with three stages, such as 1) finding the gradient, 2) non-maximum suppression and 3) hysteresis thresholding [3].

The Wilfes proposed thresholding in a vertical direction only to reduce the influence due to the eyelids. This can reduce the pixels on the circular boundary. The Hough transform successfully localizes the boundary.

Using the gradient image, the peaks are localized using no maximum suppression. The next step, hysteresis thresholding, eliminates the weak edges below a low threshold. The threshold values were found by trial and error, and were obtained as 0.3 and 0.21. Edge detection is followed by the finding the boundaries of the iris and pupil. The Hough transform is used for detecting the parameters of geometric objects. The computations in Hough transform are reduced having eight way symmetric sample points on the circle for every search point and radius. The proposed algorithm uses it to find the circles in the edge image.

C. Image Normalization
After segmenting the iris, it is normalized. This is done to enable generation of the iris code organized into 8 parts or sub-bands of \( \Pi/4 \) radians each. Since the iris orientation changes from person to person, it is required to normalize the iris image. The normalization gives common boundaries of the image and makes easier to process. This process involves unwrapping the iris and converting it into its polar equivalent as shown in Fig 3. It is done using Daugmans Rubber sheet model. The center of the pupil is considered as the reference point and a remapping formula is used to convert the points on the Cartesian scale to the polar scale [3].

![Fig. 5. Iris code formation methodology](image)

D. Iris Adaptive Encoding
The final process is the generation of the iris code. For this, the most discriminating feature in the iris pattern is extracted. The phase information in the pattern only is used because the phase angles are assigned regardless of the image contrast. Amplitude information is not used since it depends on extraneous factors. The iris code generation is carried out on the texture-processed image for better performance as shown in Fig 4. Daugman has suggested the iris phase extraction information using 2D Gabor wavelets. It determines which quadrant the resulting phaser using the wavelet. Gabor filters are used to extract localized frequency information. The log-Gabor filters are more widely used for coding natural images. Statistics of natural images indicate the presence of high frequency components. Since the ordinary Gabor filters under represent high frequency components, the log-
filters become a better choice. The proposed iris extraction method is simpler than the LogGabor and efficient for frequency component processing. The iris image goes through the process shown in Fig 4. The features are sharpening before IRIS coding and sub banding is applied. The Fig. 5 shows the difference of Iris code generation output of the conventional iris recognition system and the proposed model.

E. Tagged Index IRIS code Organization

The iris codes are formed after sectoring or sub-banding the IRIS. Searching algorithms work effectively for byte aligned system. Hence IRIS code is divided in to 8 sub-bands. Since \(2^3 = 8\), three bits are used for tag formation process. Each tag is supported with a tag_flag_bit. IRIS code is a collection of binary bit sequence in the range 0, 1. Let \(B_k = i_0, i_1, i_2 \ldots \) be a sequence of binary bits in a \(k^{th}\) sector where \(k\) is in the bounds 0, 7 since 8 sub-band sectors are formed using rubber band model of IRIS data. The transition-state, \(i_{n-1}\) to \(i_n\) is in the binary range \([0, 1]\). The Table I show the transition states between two Consecutive bits of iris code. The iris code being binary \([0, 1]\) can take four states as shown in Table I. The transitions \(0 \rightarrow 0\) and \(1 \rightarrow 1\) show no change in state hence coded as 0 while transition state \(0 \rightarrow 1\) shows positive change or \(1 \rightarrow 0\) shows negative change hence interpreted as 1 and -1 respectively. This concept has grown further to be used with three consecutive bits as show in Table II. The quantization index is associated as a tag of the respective sector or sub-band. The IRIS codes are organized in the IRIS code table using grouping based on coding quantization index such that index has a range \([-2, -1, 0, 0, 1, 1, 2]\). The association of tag_flag_bit with tag value is shown in Fig 6.

This is resulted in to a B+ search tree balanced at \([-2, 2]\) having almost 8 levels depth as shown in Fig. 7. The TAG is selected for which Bit flag = 1. Each TAG has embedded TAGs and Bit flag is the basis of selection of next TAG. Hence each TAG is a container descriptor that holds description and the IRIS code data of the respective sector. When the IRIS code is generated, the coding quantization index is calculated and accordingly left branch or right branch is selected. This is resulted into 1 of 8 selections and elimination of 87.50% hence the improvement over the sequential code search per iteration of HD method. The algorithm searches only those IRIS codes having same Tagged index per IRIS sector and guarantee the IRIS code similarity matching within 8 searches in worst case and 3 searches in optimal cases.

\[ F(S_{n-1}, S_n) = \{0:0, 0:1 = 1, 1:0 = -1, 1:1 = 0\} \]

Let \(f\) be a function that operates on bit data such that the Function \(f\) returns code quantization index for each IRIS code sub-band as per (3). As shown in Fig. 5 the coding quantization index of IRIS code sub-band is used as tagged index and before going for HD analysis of sub-band. The tagged index is searched for most favorable IRIS sub-bands.

F. IRIS Code Sub-band Similarity Matching

Comparison of the bit patterns generated is done to check if the two irises belong to the same person. Calculation of Hamming Distance (HD) is done for this comparison. HD is a fractional measure of the number of
bits disagreeing between two binary patterns. Also, due to varying bit granular density and distributed IRIS data capture resulting in variable size IRIS codes varying from 48 bits to 128 bits. Conventional HD method suffers heavily due to fixed size codes. Since this code comparison uses the iris code data and the noisy mask bits, the proposed modified form of the HD algorithm is illustrated using A and B templates. Let A and B are the templates and associated masks for the iris code under test and the iris code from the data bank respectively. Let S be the sample space \( S = s_0, s_1, s_2... s_7 \). \( A(S) \) and \( B(S) \) are the sample spaces for iris codes A and B respectively. In general, if A and B are identical then the function in (3) will return success or else it will show the failure.

\[
F(A(S), B(S)) = \sum_{i=0}^{i=7} A(S_i) \otimes B(S_i) \tag{4}
\]

Practically, function \( f(A(s), B(s)) \) works till success and the worst case of success is all 8 samples are required for testing the iris code. Equation (3) further extended to the hamming distance calculations as shown in (4).

\[
HDS_i = \frac{\| codeA_{S_i} \otimes codeB_{S_i} \cap maskA_{S_i} \cap maskB_{S_i} \|}{\| maskA_{S_i} \cap maskB_{S_i} \|} \tag{5}
\]

Fig. 5 shows the difference in methodology of iris code sub-band organization for the iris under test. The iris code slice is searched using sub-band tags associated with circular link list as shown in Fig. 8. The iris code is tested slice wise. This helps eliminating the iris codes not eligible using (4). Table III shows the class descriptor used for IRIS code data. FOURCC is a 4 character data structure used to name the class. The”iris” is a container class which contains eight data classes in it, \( col_0 \), and \( col_1 \).... \( col_7 \) respectively which hold IRIS code sub-band data. The data classes hold IRIS codes to be searched. This class in algebraic terms of set theory is represented as

\[
S = \{ SIZE, "iris", \{ col_i \} | 0 \leq i \leq 7 \} \tag{6}
\]

The legacy methods for iris recognition mainly focus on feature representation and matching. These methods do not undertake research of selecting favorable iris code for matching to avoid searching delays during finding the match. Also IRIS code data is not stored as class descriptor to take advantages of objects oriented technology, the recent advances in technology. Hence these techniques suffer in various ways of storage efficiency, search efficiency and data communication efficiency.

These techniques do not focus on distributed data organization to accommodate data distributed all over the globe through Internet. Typically, data like IRIS needs global distributed attention for important applications like security. Therefore, we only analyze and compare the accuracy and efficiency of feature representation and matching of these methods. The methods proposed by Daugman [1][2], Wildes et al. [3], Boles and Boas hash [3][5] are probably the best-known. These characterize local details of the iris based on phase, texture analysis and zero-crossing representation respectively. The reorganization of CASIA Iris Database to support distributed
V. CONCLUSION & FUTURE CHALLENGES

The proposed method performs better searching of IRIS codes hence speedy iris recognition as compare to the conventional systems. A novel organization of iris code presents the faster searching of iris code over the conventional methods thereby saving on the searching times. Being defined as a class descriptor, it has become portable and any service on Internet can import it. In recent world, security has become a global challenge and speedy availability of secured data has become a prime challenge. IRIS recognition is one of such prime challenge. The proposed IRIS code organization and searching being a class descriptor supports future technologies which may include distributed cloud architecture “IRIS code cloud”. Being container class descriptor this data can be embedded in any MPEG-4 image providing high quality security and transportability. The destination application can identify the descriptor signature and encode the IRIS code information hence adaptive in nature.

The resulting in easy instantiation and availability of IRIS code information. The applications proposed by Daugman use millions of iris code comparisons in a system such a reservation system, airport security system with central database rather than a distributed database [1][2]. The proposed algorithm since faster in iris code similarity matching due to reduction in redundant searches is advocated for such time critical systems. Due to slice/TAG based organization of iris code, even a person developed cataract can be efficiently detected since the probability of failure of iris code sub-band slice is only 1 of 8 or 2 of 8 in worst case. Hence the proposed algorithm is useful in adverse cases of impurity in iris codes. The paper strongly advocates efficient searching of IRIS database using proposed organization.

REFERENCES


Mr. Pankaj Agarkar is working as Assistant Professor in Computer Engineering Dept of BSCOER, University of Pune (M.S.), E-mail: pmagarkar@rediffmail.com

Dr. S. N. Talbar is Professor and Dean with SGGS Nanded, SRTM University Nanded (M.S), India. He is now with the E&TC Department, India E-mail: sntalbar@yahoo.co.in