Innovative Computational Framework Applied With DRL, Encryption and KNN for Biometric Face Authentication

Shivani Kumari, C S Raghuvanshi , Ankita Jain
Faculty of Engineering & Technology, Rama University, Kanpur, Uttar Pradesh, India
[1]shivani.kumari2k20@gmail.com, [2]drcsraghuvanshi@gmail.com, [3]ankitajain1691@gmail.com
*Corresponding Author: ankitajain1691@gmail.com

Abstract. Security framework based on Biometric recognition is highly acknowledged in the research community. However, security challenges have been seen in the existing research related to recognition failure and security issues. The proposed model endorses a paradigm of biometric recognition where the focus is to maintain the perseverance of the recognition accuracy under any unpredictable circumstances. The model utilizes a set of non-symmetrical dynamic video inputs in which facial subjects have been traced and analyzed. The aim of the work is to provide an effective and consistent recognition rate and preserve the security of the biometric characteristics. The proposed model uses the Haar-Cascade algorithm to trace the facial subject from an input video frame. This algorithm utilized the haar-plotting method to obtain the facial portion accurately. Then, the model applied preprocessing operations to normalize the variation and noise. The histogram equalization approach and bilateral filter are the techniques that have been utilized under normalization to eliminate the unwanted illumination effect. Furthermore, the model brings a deep reinforcement learning technique for feature extraction in terms of unique binary codes. Biometric features are sensitive and permanent identities of an individual must be preserved. Therefore, this model uses a hybrid combination of encryption techniques. In encryption technique the extracted features are encrypted through shuffling and chaotic methods. Further, a K-nearest Neighbor (KNN) algorithm has been utilized for the classification and recognition tasks. The proposed model outperforms the state of the art technique by obtaining 99.80% accuracy.

Keywords: Facial biometric recognition, Pre-processing, Haar-Cascade, deep reinforcement learning (DRL), Shuffling & Chaotic, K- nearest Neighbor (KNN).

1. Introduction

As with increasing online transactions, the security aspect is one of the high-priority considerations of any research. Real-time video recorders can record pictures of faces with a variety of stances and distractions, such as noise, darkness, etc. The encryption of the biometric features is provisioned to avoid any unpredictable worst-case misuse of biometric data from the server [1].

The encrypted biometric features are also unique and stable. The proposed model is conducting the scenario of classification and recognition under a secure environment in which encrypted features are used instead of actual sensitive feature points [2]. Here, achieving high identification accuracy for face data under various disruptions is the main goal. Without heading towards a conventional approach where the researcher uses the extracted feature directly for training and testing purposes. The video input images are reliable for face recognition and authentication. It is because of multiple frames of facial identity in the video input. If a system fails to recognize the facial identity of one single frame, then the system will look for other frames of the same facial identity to make the correct recognition. However, the major disadvantage of video input facial identity is its large size as compared to any static input image [3]. In Video input, instead of giving a single facial image, sequence frames are fed to the system so the system may require a large processing time to process all the frames of the video. The aim of the model is to minimize the processing time for face recognition from video input. Therefore, the face detection part is successfully executed by the Haar-Cascade method (Viola Jones) [4]. The proper segmentation is also followed after detection. The entire captured frame contains side details including background objects. So, the Haar-Cascade has been applied to masked-out the facial portion where the biometric features reside. Then, the model performs the normalization of the segmented facial image in which the histogram equalization technique [6] is applied to avoid illumination variation. All of the coefficients of features are improved, including the poor-quality pixels with the help of the normalization procedure. Other kinds of noise are also suppressed using contrast enhancement. The segmentation image is pre-processed using morphological operation in which inappropriate pixel alignment is adjusted into fixed ratio scaling. The pre-processed normalized image is utilized by the initial level nodes of the deep reinforcement learning model for feature extraction. This algorithm scans input normalized images repeatedly to find unique local binary patterns [7]. Deep reinforcement learning learns the feature extraction from its feedback using a proportion network and agent. There is an agent in this algorithm that collects the reward in every iteration of feature extraction. The purpose of the reinforcement learning model [8] is to increase the rewards by collecting a greater number of hidden local patterns. These local features are unique in each facial identity and also stable against any variation. In other words, the local binary features are the least effective unit in facial subjects which possibly stays un-deviated under unpredictable variation. The proportional network sends errors in feature extraction back to the earlier nodes so that the reinforcement model updates the weight of its nodes and computes local patterns again. The reinforcement model maps the fetched features into a binary tree or codes. These binary codes are created using the histogram of oriented gradients (HOG) technique [9]. The proposed algorithm successfully locates the distinct HOG characteristics on the face of the user to identify it from other faces and provide appropriate binary codes. There can be more than one binary tree formed belonging to one facial identity based on the variation. The extracted features are then encrypted and shuffled using a chaotic algorithm. The chaotic encryption [10] R, G, and B components of the facial image separately. This encryption technique minimizes the correlation among these components. This encryption technique is applied to binary-coded local binary features. After, the proposed model applies a shuffling technique [11] on the encryption features in which scrambling of encrypted bits is done. This dual encryption hybrid model uses some specific key for encryption to satisfy the high security of the features. The classification and recognition part is separately done by the KNN algorithm. The encrypted features are utilized for training and testing purposes rather than using actual biometric features. The training of the model has been done by reading the encrypted features by the nodes of the network. The KNN forms classes for each type of encrypted feature during the training phase. To create a different encrypted code during the testing phase, the test biometrics photos have been encrypted using the same key. Only the mapping of the two encrypted codes will be used by the model for its final recognition. Thus, this approach will not only remove the possibility of fraudulent authentication but will also remove the recognition dependency on vital biometric information that, once copied or lost, cannot be recovered. The security architecture for an online attendance system or any other cloud-based system access can be supported by this scenario.

In contrast to previous works, the proposed approach makes use of a real-time video input dataset [5] that consists of frames of randomly clicked photographs of a distinct person. These pictures are not taken at a certain distance or in a particular setting. The proposed approach eliminates this by using refined and local binary pattern-based features that are retrieved for non-symmetrical images and are not predicted to change in the presence of noise.

The generalized approach of the proposed method is shown in Figure 1 in which hash codes of the normalized features are mapped by the suitable classifier to establish the recognition process.

The main contribution of the paper is as follows:

- The proposed approach introduces an innovative and effective way to identify the facial region of an image uniquely by applying an altered version of the Haar-Cascade method. The method creates a rectangular border around the facial area by first plotting the Haar features of the face and then identifying the shapes of the mouth, eyes, nose, and other features. As a result, in the situation of a crowd, every face feature can be segmented out with ease and error-free.

- Unique features from the biometric facial pre-processed images are captured by DRL method and then security features to the facial biometric traits are applied by using shuffling and chaos techniques. These techniques are used to encrypt the biometric traits and generate a hash code which is then utilized to map the correlations for the recognition purpose. Finally, KNN algorithm is applied which is a resilient classifier that updates its learning and develops into a powerful classifier that can process any huge and complicated hash feature with ease and differentiate each feature in classes.
- The proposed approach obtained 99.80% accuracy in consistently identifying images with unpredictability.

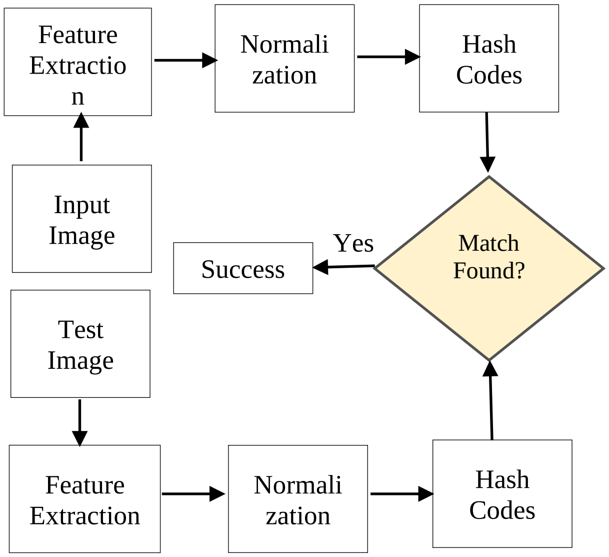


Fig.1. Generalization of the Model

The arrangement of the additional modules is as follows. In section 2, the related work is illustrated. In section 3, the proposed methodology is presented. Section 4 contains the experimental results. The conclusion is summarized in section 5.

2. Related Work

The primary distinction between the proposed approach and other methods [10] [11] is not just the level of precision but also the kind of dataset and attributes used. Du et al. [12] proposed review in the field of biometrics and security which shows that the researches happened in the last decade turns obsolete when it comes to dealing with variations.. Singhal et al. [13] utilized the symmetrical dataset, which consists of symmetrical photographs clicked in particular environmental conditions, have been conducted using other current methodologies. The limitation of this work is lack of generalization which leads to increase testing phase error for any new case. Ali et al. [14] proposed the identification consistency that might not be strong enough to handle photos with a lot of random fluctuations. The concerning challenge leads to increase false acceptance rate. These current models can only function if they receive the required training. However, repeatedly training a machine learning model for diverse datasets is such a burden. The primary constraint of the Teoh et al. [15] could be its processing overhead when dealing with photos that have a lot of variation. These photos might have been taken in a crowded area or might have a dramatic lighting effect. When clicking on non-symmetric random images, the likelihood of inaccuracy is substantial.

3. Proposed Methodology

The methodology uses a video dataset called the choke point dataset [5] from which video frames are taken and used for the face recognition task. The proposed methodology is depicted in Figure 2. First, input video frames are captured from the surveillance dataset. Then, we applied the Haar-cascade classifier (Viola Jones) algorithm to extract the facial region of the subject from the frame. Then, we applied pre-processing operations to normalize variations in the features. Further, Deep reinforcement learning has been applied to obtain the unique features. After, these extracted features are encrypted into unique hash codes by using chaotic and shuffling algorithms. Finally, we applied the KNN model for the classification of unique hash codes into the identity classes.

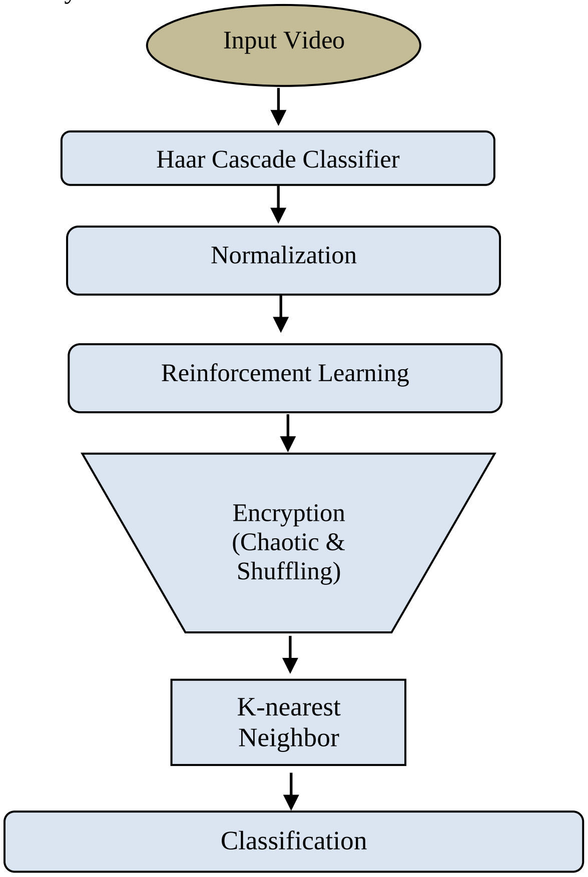


Fig 2. Flowchart of Proposed Methodology

The proposed solution gathers video footage from a camera recorder, which captures different individuals entering the space. A single person has been captured in several photos across multiple input video frames. After then, each face of the user is identified by analysis of these video frames. Figure 3 shows the sample of input video frames.

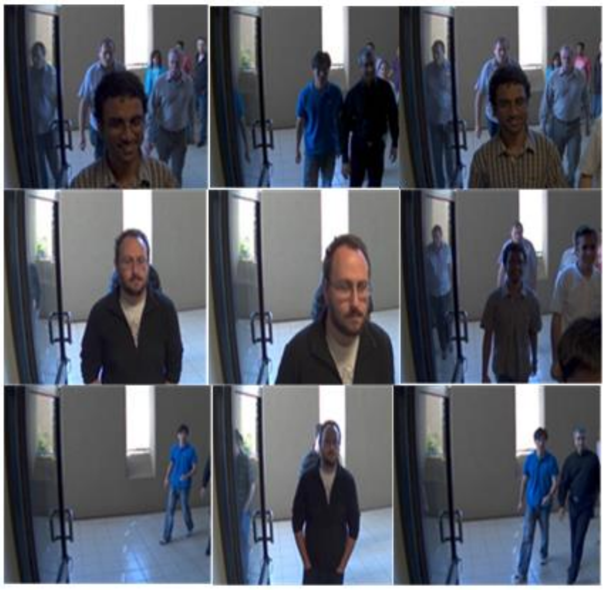


Fig. 3. Input video frame

In Figure 3, several persons are shown in the video frames. Multiple images of the same individual can appear in video frames. The dimensions of each video frame are 552 by 480 pixels. These video frames undergo a pre-processing procedure in which any undesired noise variations are removed.

3.1 Haar-Cascade Method

For the working of proposed system, a real-time object identification technology that can extract features from the surveillance footage is required. To find facial regions in video frames, the model uses the Haar-cascade method, which finds haar characteristics. These qualities are distinct and resistant to alteration, which is to be expected in environments with noise, poor lighting, erratic facial expressions, and posed subjects [15]. The Haar-cascade algorithm tracks the facial region of a subject by enclosing it in a rectangular perimeter and searches through all of the detected feature points in video frames. The method fits the features inside the rectangular region which is unique for each individual. Haar characteristics have been calculated by taking the area of a rectangular region in which the sum of the values of characteristics is obtained. The characteristics have been transformed by the addition of all the pixel values residing in the rectangular region. Under a given threshold, the assessment of the generated sum is compared many times. For the face of an individual, only the characteristic points that fall above the threshold are nominated; the remaining feature points are eliminated. This approach removes the non-facial data that contains side information such as background, other body parts, etc. Figure 4 shows the results obtained after applying the Haar-cascade classifier. It has been seen that all the facial regions are extracted well from the input video frames whose samples are shown in Figure 3.

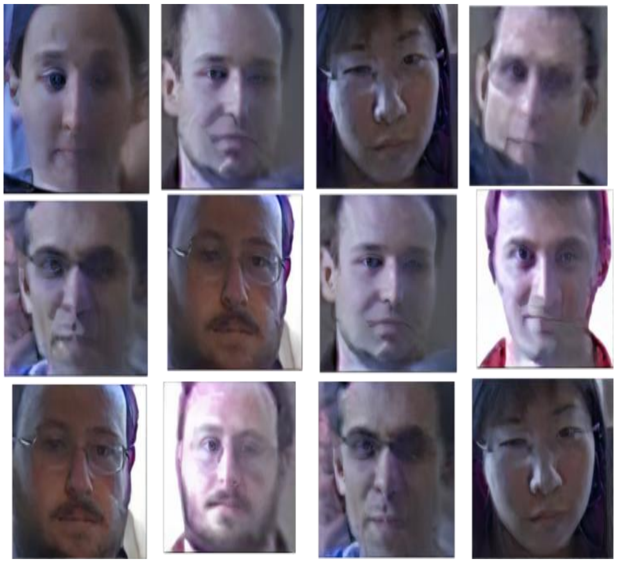


Fig. 4. Extracted Facial Regions from the video frame

3.2 Normalization

The normalization process is applied to enhance the feature points of the segmented images. The normalization process targets the variations of lighting, blurring or any other random noise that may negatively affect the consistency of the recognition.

3.2.1 Histogram equalization

A method called histogram equalization is used for contrast enhancement of the facial subject. Histogram equalization method performs [16] the analysis of the pixel intensity values of the image.

$$\mathrm{H}(X_n) = M_n \quad (1)$$

Here, H is the histogram function of the digital image. X_n is the n^{th} intensity level. M_n is the number of input pixel of an image. The histogram method preserves the intensity of the pixel values as uniform. Besides using histogram method, the proposed work uses several morphological operations in order to remove noise affect from the image. It makes the description of region shape, boundaries, intensity etc. more useful for the model. The normalization basically targets the following operations:

- Locating and cropping the facial region by using a rectangle window according to face model.
- Locating the center position of the two eyes to a fixed position
- Scaling the image size according to the requirement of model.
- Rotating the image frame to line up the eye coordinate.

The goal of the preprocessing work is to enhance the feature patterns. To guarantee the quality level, improvements have been made to the video frames. This aids in the extraction of the high-quality characteristics necessary for precise face identification. The proposed model uses histogram equalization and bilateral filters during the pre-processing stage to enhance video frame quality and maintain characteristic point quality.

3.2.2 Bilateral filter

To improve the overall component value, the proposed model pre-processes the video frames. In order to provide smoothness to video frame images while maintaining edge integrity, this filter seeks to combine non-linear neighboring pixels. It gathers the weighted average data of the local region of interest with intensity values. One can compute the weighted average (w(p)) of a pixel as:-

$$w(p) = \frac{\sum_{p' \in P} Im(p')P(P - p')T(Im(P) - Im(p'))}{\sum_{p' \in P} P(p' - P)Im(Im(P) - Im(p'))} \qquad (2)$$

Here $P(P - p')\ and\ T(Im(P) - Im(p')$ are the spatial and total weights of p' pixel information. P is the set of pixel information and 'Im' is the intensity value. Then Gaussian function is applied on pixel and its intensity to smooth its properties.

$$p'(P) = \frac{1}{\sigma_p \sqrt{2\pi}} e^{-p'^2/2\sigma^2} \qquad (3)$$

$$I(P) = \frac{1}{\sigma_{Im} \sqrt{2\pi}} e^{-p'^2/2\sigma^2} \qquad (4)$$

3.3 Deep Reinforcement Learning (DRL)

DRL allows an algorithm to learn from its observations. After accepting the retrieved facial information produced by the Haar-Cascade method, the proposed model extracts the HOG features, which are then converted into a binary tree. Various occurring characteristics from different facial photos of the same individual are combined to create equivalent binary codes. All matched features face subjects mapped into binary trees with similar mappings. A single individual may appear in the video frames more than once according to the proposed model. In this scenario, a single binary tree for each of the numerous frames that have one individual arriving several times is constructed using DRL [17]. DRL aims to produce several binary tree structures for various subjects that are recorded by a video camera. A single linked face subject that is distinct from the binary codes of other facial subjects is identified by one binary code segment. On the retrieved photos, DRL continuously finds the hog features and builds its binary codes. It sends back the error in finding the hog features via a backward propagation method. It enhances learning for different facial subjects in this way. DRL consists of two main parts: the environment and the agent.

DRL Algorithm

DRL Input : C_t, H_t
Output provided by DRL : Binary codes

Step 1:- Initialize value of global Reward R.
Step 2:- Initialize sequence
R (C_t, H_t)
C_1 = {x1} and preprocessed sequenced φ1 = φ (C_1)
C_t is the global cost function and H_t is the global weight corresponding to state

Step 3:-

R(C_t, H_t)= R(C_t, H_t)+α.[r+ϒmax(R(C'_t, H'_t)- R(C_t, H_t))] (5)

Loss function = - r(C_t, H_t) logπ(C_t , H_t)

Here, α is learning rate, r is local reward at an instance of a time, C'_t and H'_t are the local cost and local wright respectively that are associated with r.

A policy network seen in deep reinforcement learning builds the binary tree via trial-and-error interactions using the performance of the agent in a simulation environment. A state in DRL includes hog features of the face subjects. The model trains itself repeatedly with the help of a back-propagation network to maximize the output and minimize the loss. It tries to reduce the gap between the predicted output and the actual output. Figure 4 shows the working of DRL.

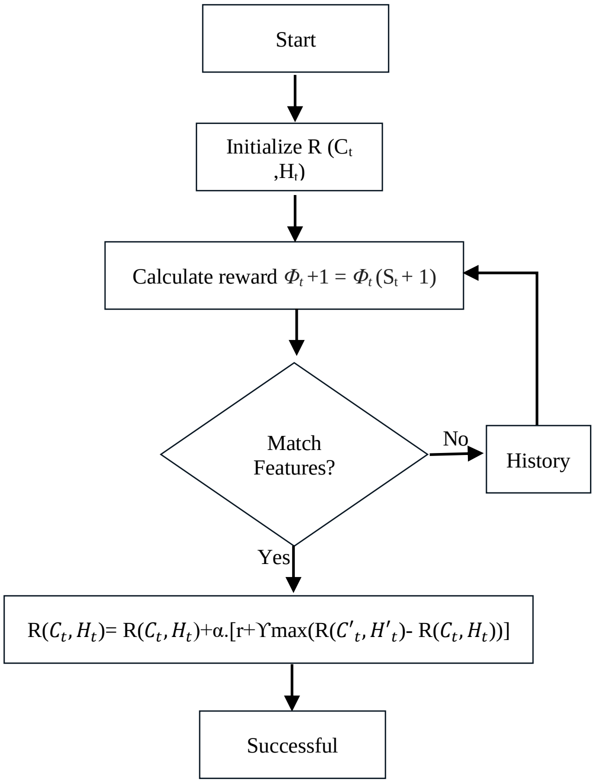

Start
Initialize R (C_t ,H_t)
Calculate reward $\Phi_t+1 = \Phi_t(S_t + 1)$
Match Features?
No
History
Yes
R(C_t, H_t)= R(C_t, H_t)+α.[r+ϒmax(R(C'_t, H'_t)- R(C_t, H_t))]
Successful

Fig. 5. Working of DRL

Figure 5 shows the working of DRL in which the features of each input image are extracted in terms of reward and it is compared with the actual standard value. The actual standard value is fixed based on the average feature value of the biometric image. The local points may have different feature values. The comparison of each local pattern is carried out with a actual standard value. If the matching is not done properly, then its feedback is sent back in terms of history to the previous nodes. Then, the weights of input local features are modified accordingly to obtain robust features that may be further used in the classification process. The DRL generates binary codes of a feature point repeatedly until its matching gets satisfied with the standard parameter. And so, it increases the value of the reward function to get rich feature extraction.

3.4 Encryption

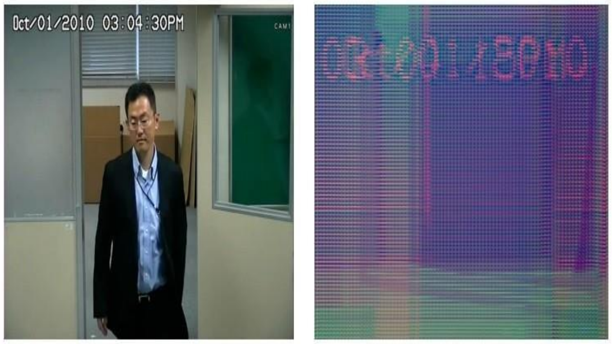
The proposed model applies encryption techniques to encrypt the unique biometric traits generated by DRL algorithm in the form of unique binary trees. The purpose of the encryption is to generate unique hash codes of the features in order to provide security to these sensitive features while storing it in the model database for the authentication purpose. A hybrid encryption model has been applied in the proposed work in which the first layer of encryption will be done by as Shuffling technique and the second layer of encryption will be done by Chaotic technique.

3.4.1 Shuffling Method

Shuffling algorithm is applied on the extracted binary codes of the biometric local features. Shuffling technique of image pixel values has been demonstrated to be really potent in security analysis. The binary coded feature information is shuffled in which additional swapping of pixels has been performed within the image file. The succeeding component shifting has improved the security of image from all potential threat available presently. In the Shuffling phase, the column coordinate and row coordinate of the shuffled image pixels are established by 2 logistic maps. Without deviation of generality, the size of the plain image f is given by m x n, where f (X, Y) represents the value of the pixel at the Yth column and Xth row of the plain image. The logistic map is described as shown below:

$$X_i+1 = UX_i\ (1 - X_i),\ X_n \in [0, 4] \qquad (6)$$

Whereas X_n represents an independent variable;
U represents the control variable of the logistic map; i =0, 1, 2. Nevertheless, there exists empty windows in the chaotic region for the logistic map at the time U [3.82, 3.85]; may suffer lack of randomness and also it is not secure if they are utilized in encryption. Thus, the variable U is close to 6 when it is really utilized. Figure 6 contain the sample of original and shuffled image.



(a) (b)

Fig. 6. **(**a) Original frame (b) Shuffling frame

There are three steps which are discussed as follows:

Step one.
Chaotic sequence X_i is produced by Eq. 6. Uo and Xo are the variables as beginning secret keys, in which i=1, 2, 3,….,m. The value sum represents the sum of all plain image pixel amounts. Clearly, X_i *(*0, *1*). Currently floor (X) rounds of components of X to the closest integers toward minus infinitude. Mod returns the remnant after division. Sequence p is utilized to mark

the row coordinate of pixels of the disarranged image. Clearly, pi ∈ [1, *m*].

$$pi = floor\ (mod\ ((Xi * 108+sum/(m * n * 256) *108),m) + 1 \qquad (7)$$

Step two

Yi, chaotic sequence, sis produced by Eq. (6). *Yo* as well as *Ui* are the variables as beginning secret keys, in which i=1, 2, . . ., n. series q is utilized to mark the column coordinate of pixels of the disarranged image. Clearly, *qi* ∈ [1, *n*].

$$qi = floor\ (mod\ ((Yi *108 +sum/\ (m * n * 256) * 108), m) + 1 \qquad (8)$$

Step three

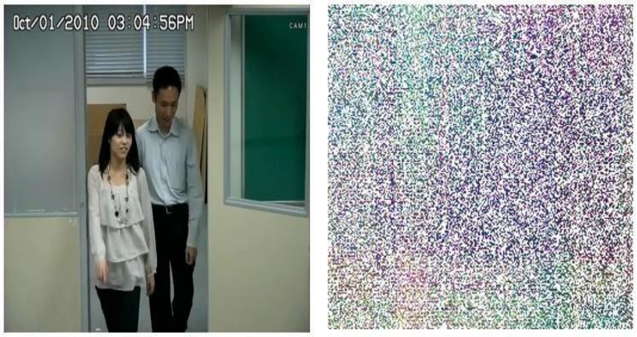
In respect to series p as well as q, an indirect matrix is produced and the position of pixels in the image is decided by the proposed exchange model. The study proposed that the plain image p size is m x n = 4 x 6. The values in the illustration 2 exhibit the scan sequence of the virtual coordinate and the plaintext matrix. The amount of i is ranged from 1 – m and the j are ranged from 1 – n. The F1 is the shuffled image, is extracted by interchanging the pixel amounts of f (i, j) as well as f (p (i + j -1) mod m, qj) every iteration by Eq. (9):

$$F1\ (i, j) = f\ (i, j)\ (pi+j-1 \bmod m, qj) \qquad (9)$$

The output image is produced as a shuffled image till the end of interchanges. Then shuffled image is then forwarded into chaotic encryption to preserve dual layer security.

3.4.2 Chaotic Encryption

The shuffled image is now introduced to the second layer of hybrid decomposition in which chaotic encryption has been applied. In chaotic encryption, the R, G, and B components are independently encrypted, and therefore the correlation is minimized among them which ensures robust encryption. The permutation and diffusion state of chaotic encryption helps to reduce the correlation between the intensity values of the color elements of the picture. The pixels of the image are permuted several times concerning a key. The diffusion process has been performed after every two rounds of permutation which reflects the change in the image histogram significantly. Chaotic codes are synchronized with each other and so they are deterministically generated. This encryption does not affect any kind of data loss. Chaotic encryption is used to spread particular data in an entire space. Chaotic encryption works on a finite set of integers. The pixels of the image are rearranged and XOR-ed with some specific key. The same key is utilized in decryption also. The chaotic encryption conducts the Chen chaotic method utilizing the 4th sequence Runge-Kutta algorithm [18]. Chen's system equations are alike as compared to the Lorenz system. Figure 7 contains a sample of the original frame and a chaotic frame.



(a) (b)

Fig. 7. **(**a) Original frame (b) Chaos frame

As seen in figure 7, original frame is given which is taken from the dataset and the corresponding chaotic frame is shown which is generated by applying chaotic encryption.

3.5 K-nearest Neighbor (KNN)

DRL characteristics have been classified using the KNN technique. The KNN technique is based on the Euclidean distance [19] principle, which calculates the distance that exists among feature values to determine whether a facial object belongs in a class. This approach identifies the neighboring property that has been used to carry out facial subject identification. The identity of an individual is uniquely identified by each class using feature points. Identity of every person is saved in a distinct class that can accurately differentiate from other classes during training. The distance metric between local characteristics is used by the KNN to compare similarities between them. DRL has already recovered all of the local and global features from a facial image into the binary codes. Two forms of local characteristics are found on face subjects: speeded-up robust features (SURF) [20] and invariant feature transformation (SIFT) [19]. Additionally, the global characteristics include global subject data for the face. The KNN algorithm uses the extracted characteristics to map similarity in order to complete the identification task. Considering f_x is the feature of an image I_x and feature f_y of image I_y . The calculated distance is given in equation 10:

$$\text{Distance } (f_x , I_y) = \frac{d(f_x,\ TT_2(f, I_y))}{d(f_x,\ TT_1(f_x, I_y))} \qquad (10)$$

Here, f_x and $TT_1(f_x, I_y)$ are considered features for matching. If the ratio Distance (f_x , I_y) is smaller than fixed threshold then it is considered that the feature f_x matches to an image I_y.

$$C\left(f_x , I_y\right) = \begin{cases} 1 \text{ or ture,} & if\ D\left(f_x , I_y\right) < threshold \\ 0 \text{ or false,} & otherwise \end{cases}$$

The match function in this case is called "C," and the threshold value of 0.8 permits 8% of mismatched or overlapping characteristics. Stated differently, the matching percentage of the local features f_x and f_y is used to determine the correct resemblance between the two pictures, I_x and I_y. It is further explained by equation given below:

$$S^C\left(I_x, I_y\right) = \frac{1}{I_x}\sum_{f_x=I_x} C\left(f_x , f_y\right) \qquad (11)$$

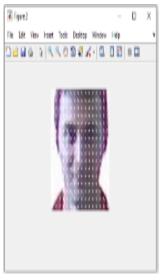
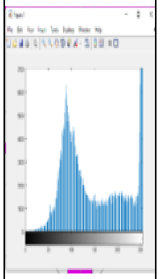
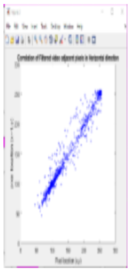
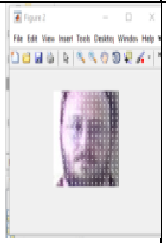
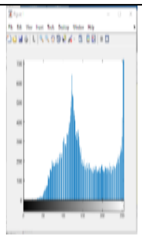
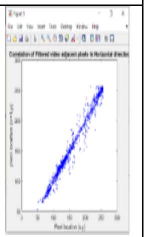
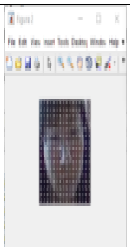
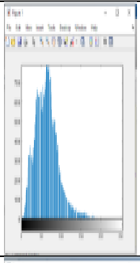
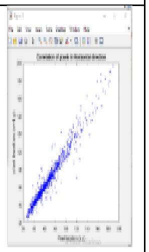
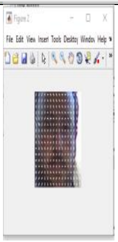
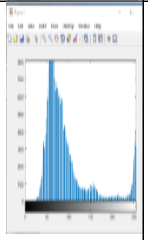
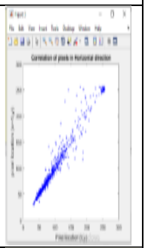
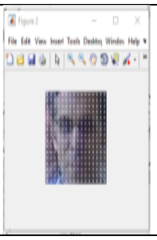
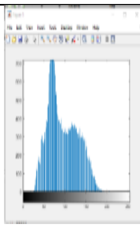
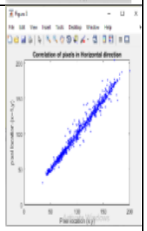
Here, $C\left(f_x , I_y\right)$ is 1 if f_x has a match in I_y and 0 if not.

4 EXPERIMENTAL RESULTS

For testing and training, the dataset was split into 70 to 30 ratios. Approximately 48 video films and 64204 captured images of people walking around a room are used in the proposed model, which is based on the typical data set [5]

The various individuals captured in the video frames is displayed in Figure 3. A total of 125 frames of video, representing five different people entering through the video recorder, have been recorded. Every individual is recorded in five distinct frames. Following pre-processing, these images are sent to the Haar-Cascade method, which extracts the faces from the pictures. Next, DRL was used to hold the extraction of features. These extracted features are encrypted in the form of hash codes that are unique and their testing and training is successfully carried out by KNN model. The recovered face portion, HOG representation, neighborhood histogram representation, and correlation are all shown in Table 1.

TABLE 1: Extracted face, HOG characteristics, Histogram and correlation coefficient.

INPUT FACE	HOG Visualization	Histogram	Correlation Coefficient
			
			
			
			
			

Based on binary codes produced by DRL, each of the five individuals that were recorded in 125 video frames is identified as a single person. The KNN method is used for categorization after generation of feature vector. Figure 8 illustrates about KNN recognizing faces while live capturing, identifying the area surrounding the face of individual in a rectangle box with a yellow tint.

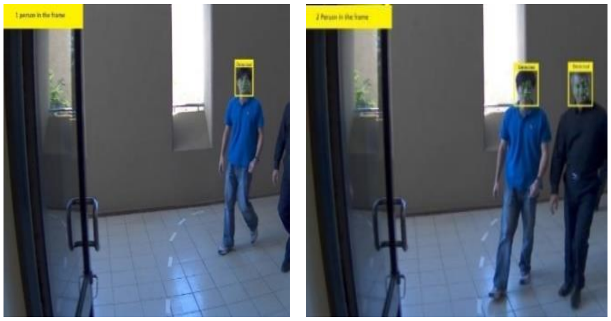


Fig. 8. Face identity labeling in real time by KNN model

The KNN algorithm concludes the identification of face subjects in the live footage, as shown in Figure 8. When a person appears again, the model detects patterns in the local characteristics and recognizes it as one individual. It is also clear from figure 8 that the model can distinguish between different identities when many people are shown collectively in a frame of video. Further, the correlation among the feature vector within the image has also been measured. The correlation among two-adjacent pixels of the input video frame is tested. The study arbitrarily chosen one thousand pairs of adjoining pixels, that is, in vertical, diagonal, and horizontal direction from an image prior encryption and other succeeding the encryption. Figure 9 shows the correlation matric that shows the similarity among the features before and after the encryption.

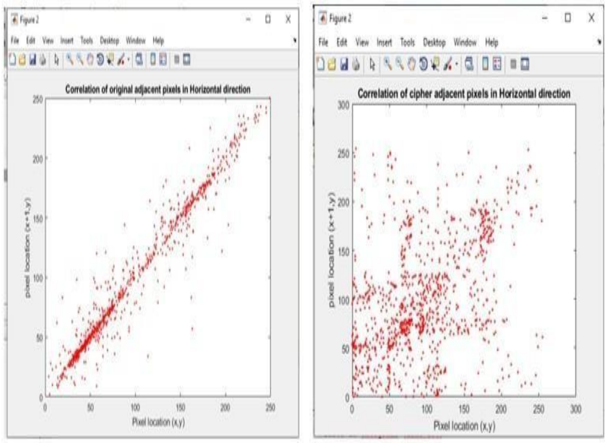


Figure 2
File Edit View Insert Tools Desktop Window Help
Correlation of original adjacent pixels in Horizontal direction
pixel location (x+1,y)
Pixel location (x,y)
0
50
100
150
200
250
Figure 2
File Edit View Insert Tools Desktop Window Help
Correlation of cipher adjacent pixels in Horizontal direction
pixel location (x+1,y)
Pixel location (x,y)
0
50
100
150
200
250
300

(a) (b)

Fig. 9. (a) Correlation coefficient of original frame. (b) correlation coefficient of encrypted frame

As we can see in figure 9 (a), there is a large correlation among the characteristics of the original image. Figure 9 (b) shows that there is a very less correlation among the features after the encryption. Hence, the shuffling and chaotic encryption technique help to reduce the correlation among the characteristics vector which enhances the security of the biometric details. The grey-scale values of 2-adjoining pixels from the image are represented by X and Y. The correlation of 2-vertically adjoining pixels in the plain-image, as well as encrypted image, is shown in figure 9. With 300 examples of footage frames, the proposed model has an average identification accuracy of 99.80% for images. The ROC curve for the proposed method of identifying photos is displayed in Figure 10.

Figure 10 indicates that the proposed model for face identification from input video frames is effective under several distortion and diversity in dataset. The false-positive rate is 0.2% which signifies the low failure of recognition of some test encrypted features. Table 2 contains a state-of-art comparison of the proposed model.

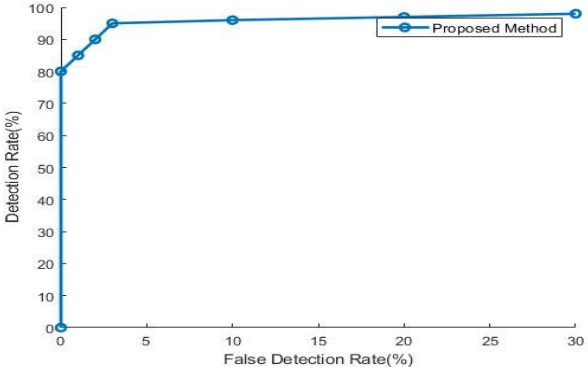
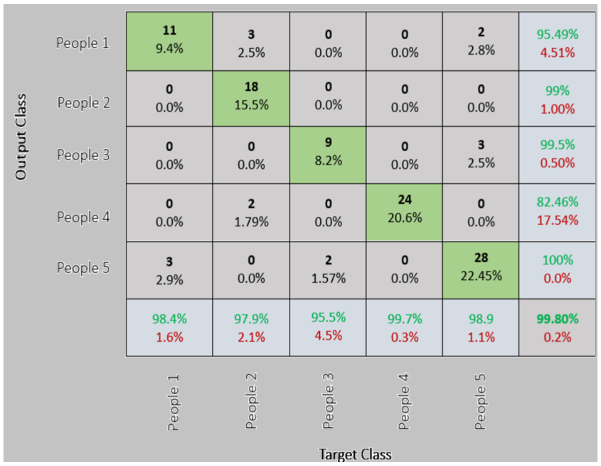


Fig. 10. ROC curve



Output Class	People 1	People 2	People 3	People 4	People 5	
People 1	11 9.4%	3 2.5%	0 0.0%	0 0.0%	2 2.8%	95.49% 4.51%
People 2	0 0.0%	18 15.5%	0 0.0%	0 0.0%	0 0.0%	99% 1.00%
People 3	0 0.0%	0 0.0%	9 8.2%	0 0.0%	3 2.5%	99.5% 0.50%
People 4	0 0.0%	2 1.79%	0 0.0%	24 20.6%	0 0.0%	82.46% 17.54%
People 5	3 2.9%	0 0.0%	2 1.57%	0 0.0%	28 22.45%	100% 0.0%
	98.4% 1.6%	97.9% 2.1%	95.5% 4.5%	99.7% 0.3%	98.9 1.1%	99.80% 0.2%

Target Class

Fig 11. Confusion Matrix

Figure 11 shows the results of the efficiency evaluation using the confusion matrix, which was computed during the gathering of video data with five distinct person counts.

TABLE 2: A state-of-art comparison of proposed model with other models in term of accuracy

S. No	Face Representation	Accuracy (%)
1.	Gabor and intensity [6]	70.10
2.	Bone density, dental structure or face.[7]	86.54
3.	Gabor-LBP features [8]	72.53
4.	CNN, VGG-16 Architecture [9]	83.00
5.	Gabor filter and PCA [10]	97.33
6.	**The proposed technique**	**99.80**

Table 2 makes it evident that, in comparison to several other current facial identification algorithms, the proposed model is effective. The proposed approach can be applied to numerous visual identification programs, including internet forensics, investigation of crimes, and real-time attendee tracking.

5 Conclusion

The effective and reliable techniques for face recognition from surveillance video were examined in the proposed work. The proposed model significantly incorporated pre-processing tasks by utilizing Histogram equalization and bilateral filter techniques. Further, the Haar-Cascade algorithm has been applied to obtain a segmentation of the region of interest. Then, DRL has been applied for feature extraction which is then encrypted by applying Shuffling & Chaotic techniques. And finally, the KNN algorithm has been applied to get a good standard of recognition results during noise and fluctuations. An accuracy of 99.80% has been attained for real-time video frame captures. The proposed model also shows consistent accuracy with security features which makes the model more robust against various image processing cyber threats. The intended photographs demonstrate a strong resemblance to the original and validate its accurate identification. As a result, the model generates an excellent recognition rate for photos that have been modified in the natural environment. However, the performance of the proposed model suffers in handling of high dimensional real time video dataset. We will address this issue in our future work.

References

[1] Rusia, M. K., & Singh, D. K. (2023). A comprehensive survey on techniques to handle face identity threats: challenges and opportunities. *Multimedia Tools and Applications*, *82*(2), 1669-1748.

[2] Hangaragi, S., Singh, T., & Neelima, N. (2023). Face detection and Recognition using Face Mesh and deep neural network. *Procedia Computer Science*, *218*, 741-749.

[3] Khan, S. S., Sengupta, D., Ghosh, A., & Chaudhuri, A. (2024). MTCNN++: A CNN-based face detection algorithm inspired by MTCNN. *The Visual Computer*, *40*(2), 899-917.

[4] Rodriguez, A. M., Geradts, Z., Worring, M., & Unzueta, L. (2024). Improved likelihood ratios for face recognition in surveillance video by multimodal feature pairing. *Forensic Science International: Synergy*, 100458.

[5] The chokepoint dataset is sponsored by NICTA. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy, as well as the Australian Research Council through the ICT Centre of Excellence program, http://arma.sourceforge.net/chokepoint/.

[6] Qian, Y. (2024). Face recognition technology for video surveillance integrated with particle swarm optimization algorithm. *International Journal of Intelligent Networks*.

[7] Gupta, N. S., Ramya, K. R., & Karnati, R. (2024). A Review Work: Human Action Recognition in Video Surveillance Using Deep Learning Techniques. *Информатика и автоматизация*, *23*(2), 436-466.

[8] Karpagam, M., Jeyavathana, R. B., Chinnappan, S. K., Kanimozhi, K. V., & Sambath, M. (2023). A novel face recognition model for fighting against human trafficking in surveillance videos and rescuing victims. *Soft Computing*, *27*(18), 13165-13180.

[9] Saleem, S., Shiney, J., Shan, B. P., & Mishra, V. K. (2023). Face recognition using facial features. *Materials Today: Proceedings*, *80*, 3857-3862.

[10] Singh, R., Saurav, S., Kumar, T., Saini, R., Vohra, A., & Singh, S. (2023). Facial expression recognition in videos using hybrid CNN & ConvLSTM. *International Journal of Information Technology*, *15*(4), 1819-1830.

[11] Tariq, S., Jeon, S., & Woo, S. S. (2023). Evaluating trustworthiness and racial bias in face recognition apis using deepfakes. *Computer*, *56*(5), 51-61.

[12] Du, H., Shi, H., Zeng, D., Zhang, X. P., & Mei, T. (2022). The elements of end-to-end deep face recognition: A survey of recent advances. *ACM Computing Surveys (CSUR)*, *54*(10s), 1-42.

[13] Singhal, P., Srivastava, P. K., Tiwari, A. K., & Shukla, R. K. (2022). A Survey: Approaches to facial detection and recognition with machine learning techniques. In *Proceedings of Second Doctoral Symposium on Computational Intelligence: DoSCI 2021* (pp. 103-125). Springer Singapore.

[14] Ali, W., Tian, W., Din, S. U., Iradukunda, D., & Khan, A. A. (2021). Classical and modern face recognition approaches: a complete review. *Multimedia tools and applications*, *80*, 4825-4880.

[15] Teoh, K. H., Ismail, R. C., Naziri, S. Z. M., Hussin, R., Isa, M. N. M., & Basir, M. S. S. M. (2021, February). Face recognition

and identification using deep learning approach. In *Journal of Physics: Conference Series* (Vol. 1755, No. 1, p. 012006). IOP Publishing.

[16] Fuad, M. T. H., Fime, A. A., Sikder, D., Iftee, M. A. R., Rabbi, J., Al-Rakhami, M. S., ... & Islam, M. N. (2021). Recent advances in deep learning techniques for face recognition. *IEEE Access*, *9*, 99112-99142.

[17] Zheng, J., Ranjan, R., Chen, C. H., Chen, J. C., Castillo, C. D., & Chellappa, R. (2020). An automatic system for unconstrained video-based face recognition. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *2*(3), 194-209.

[18] Cevikalp, H., & Dordinejad, G. G. (2020). Video based face recognition by using discriminatively learned convex models. *International Journal of Computer Vision*, *128*(12), 3000-3014.

[19] Singh, A., Prakash, S., Kumar, A., & Kumar, D. (2022). A proficient approach for face detection and recognition using machine learning and high-performance computing. *Concurrency and Computation: Practice and Experience*, *34*(3), e6582.

[20] Kumar, A., Yadav, R. K., & Saini, D. J. B. (2023). Create and implement a new method for robust video face recognition using convolutional neural network algorithm. *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, *5*, 100241.