

A Hybrid Domain Feature Convolution Approach for Facial Human Recognition

Jyothi Kiran G C¹, Ramachandra A C², Jagadeesh H S³, K B Raja⁴

¹ Assistant Professor, Department of Electronics and Communication Engineering, BMS College, Bengaluru, India.

Email: jyothirkk1@gmail.com

² Department of Electronics and Communication Engineering, NMIT, Bengaluru, India.

Email: ramachandra.ace@gmail.com

³ Department of Electronics and Communication Engineering, Sir. MVIT, Bangalore.

⁴ Department of Electronics and Communication Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bengaluru, India. Email: raja_kb@yahoo.com

Received: 28th Feb, 2026 | **Revised:** 14th Mar, 2026 | **Accepted:** 4th Apr, 2026 | **Available Online:** 20th Apr, 2026

ABSTRACT

Individual variations in facial-emotions, facial-rotation, and similarities in facial structure make face recognition in biometrics challenging. We propose a hybrid domain feature convolution approach for facial human recognition here. The benchmarked face photos are considered, resized to a uniform 128x128 size, and then converted to grayscale. The Discrete Wavelet Transform (DWT) is used to alter quality of face pictures. The Histogram of Oriented Gradients (H-OG) is used to extract properties that are independent of changes. The LL bands of the D-WT and H-OG are connected in a tumble by uniting the LL band's output to the H-OG's input. The final features are created by merging the in and out data of H-OG. The Euclidean Distance (E-D) is used to compute the findings when matching test and database items. It has been observed that the proposed system performs better in terms of outcomes than the existing ones.

Keywords: Wavelet Transform, Histogram, Multitask Convolution Network, FAR, FRR.

How to cite this article: Jyothi Kiran GC, Ramachandra AC, Jagadeesh HS, Raja KB. A Hybrid Domain Feature Convolution Approach for Facial Human Recognition. *Int J Drug Deliv Technol.* 2026;16(30s):807-816. DOI: 10.25258/ijddt.16.30s.80

Source of support: Nil.

Conflict of interest: The authors declare no conflict of interest.

I. INTRODUCTION

Humans may be challenging to recognize in a face photo due to their diverse appearance and range of positions. Face Detection is an efficient procedure for Man-Machine interface in the current global pandemic era, enabling social distance to be maintained in a variety of environments, including public transportation, offices, and hospitals. The primary requirement is a attribute corpus that allows effective human split. Since ancient times, biometrics have been developed as the most sophisticated form of computerized human identification, ushering in a new era of international protection. Biometric credentials are determined by both behavioral traits like a signature, keystroke, or stride, and physical behaviours. The time-based domain implies are converted into other-domain to extract features for biometric and to obtain extra evidence for other purposes. Fourier Transform, Discrete Fourier Transform, Fast Fourier Transform, and Discrete Cosine and Sine Transforms are readily available transformations that extract the frequency components of the signal without supplying time information by converting the time

domain/spatial domain into the frequency domain. The incapacity of the transformations to produce frequency components that correlate to time and location is resolved by the Wavelet Transform. In this work, a combined approach using Discrete Wavelet Transform (DWT) and Histogram of Oriented Gradients (HOG) is proposed for

extracting initial facial features. The final feature set is obtained by applying convolution to the LL sub-band of DWT along with the HOG feature matrix, as these components capture the most significant and reliable characteristics of facial images for human identification. The effectiveness of the proposed approach is demonstrated through database comparisons and feature matching using the Euclidean Distance (ED) metric.

II. LITERATURE SURVEY

Xiaochao et al., [1] proposed a Receptive Field Enhanced Multi-Task Cascaded CNN network by taking the advantage of the Inception V2 block and receptive field block to enhance the feature discriminability and

robustness for small targets and performance of our network is improved on the AFW, Pascal Face, Fddb and on the three sub-datasets of the Wider Face benchmark in comparison with MTCNN respectively. Lin et al., [2] proposed a guided cascaded face super-resolution network consists of a pose deformation module converts guide image's posture into same as low-resolution face image based on 3D fitting and 3D morphable model, LR image and deformed guide image are used as input of super-resolution network are formed by a cascade of two layers of networks, which extract different features, which are superior to other state-of-the-art methods such as SRResNet, SRGAN, VDSR, DBPN. Huiyu et al., [3] presented a multi-task cascaded convolutional networks algorithm-based accelerator to detection and alignment for multiple faces using multiply-accumulate, clustering non-maximum suppression, efficient pipeline architecture to resize module, intersection over union computation and eliminate the hardware-inference sorting process in NMS and to implement a complexity and memory-intensive proposal network respectively. Manjur et al., [4] utilized face detection using three-layered edge computing architecture to build a deep learning framework of multi-task cascading to recognize the face on latency and face detection load on three-layer and cloud computing architectures. Zuolin et al., [5] proposed a video oriented cascaded intelligent face detection algorithm, which builds deep learning network by cascading multiple features, from edge features, contour features, local features to semantic features, and advances layer by layer and simulation results has good detection performance for single face and multi face images, strong robustness for rotating face.

Ahmed et al., [6] defined a hybridized and enhanced eye detection method uses a faster region-based convolutional neural network with Gabor filters and naive Bayes model to address the problems associated with eye detection schemes with four components: convolution layers, a region proposal network, a detection network, and a decision model based with the accuracy in terms of eye detection is 100%. Ning et al., [7] proposed and designs a light weight convolutional neural network for detecting facial emotions in real-time and to achieve a better classification effect with system which employs multi-task cascaded convolutional networks to complete face detection and transmit the obtained face coordinates to the facial emotions classification model and then it accomplishes the task of emotion classification. Jun et al., [8] introduced to prevent collusion attacks using Video face detection algorithm based on multi-task Cascaded Convolutional Networks is designed and

implemented to solve the problems of pose, light and occlusion in the virtual currency trading system which algorithm utilizes the inherent correlation between detection and calibration to improve the detection performance under the framework of deep cascading tasks.

Ej et al., [9] proposed a distance estimation method for thermal camera applications based on the coordinates of the facial key points extracted using multi-task cascaded convolutional neural network which the method exhibits higher accuracy and the least CPU and memory consumption in comparison with conventional methods. Yanwei et al., [10] developed a diverse high-resolution dataset that contains many high-resolution images and labelled objects in total with a large variance in scale and appearance which has a rich diversity in season variance, illumination variance, and weather variance along with four different detectors namely one-stage RetinaNet, anchor-free FCOS, two-stage FPN, and Cascade R-CNN includes object detection and pedestrian detection. Ruijin et al., [11] projected a cascaded anchor-free architecture to estimate bounding box to avoid ambiguous keypoint associations and reuses existing features to reduce the number of parameters and TBox with One subnetwork uses a stacked hourglass network to detect each vehicle as a pair of corners without using anchors learns corner affinity fields, enabling it to perform robust corner grouping. Junping et al., [12] extracted a novel two-stage defect detection method one is in the localization stage called TOL-Framework is constructed to reduce the background and realize tighter oriented localization cascades a regression module inside basic framework and adds an external postprocess network, which is adversarial trained by standard insulators to refine the localization.

II. PROPOSED SYSTEM MODEL

This section discusses human recognition using facial images by cascading earlier techniques like DWT and HOG. The experimental setup for using convolution to fuse the input and outputs of the HOG algorithm to extract robust features is built in the following phases.

A Face Figure database:

The available database used are as given JAFFE, L-SPACEK, YALE, and ORL, further the dataset is split into test and dataset.

a. Japanese-Female Face Expression (JA-FFE)

There are 200 image samples total—ten distinct people, each with twenty different photos. Every picture has a 256 * 256 grayscale resolution. The frontal, erect attitude

was used to take the database shots. All the tiff-formatted images of a single person with seven different emotional facial expressions are shown in Figure 1.

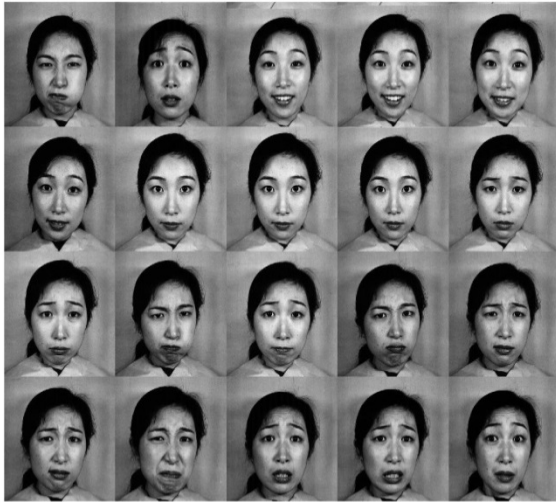


Figure 1. Images of JAFFE data set

b. L-Spacek- Databases

Each of the 152 individuals in the collection has 20 pictures, for a total of 3040 images. The images are 200*180 of size with green background. The Figure 2 shows samples of a single person's face picture.



Figure 2. L- Spacek image of a person

c. YALE-DB

There are different images of 152 subjects in the collection, for a total of 3040 photos. The only variations are in the face expressions; all the pictures are 200*180 size images. Examples of a single person's face are as in Figure 3.



Figure 3. Images of Yale database

d. ORL-DB

A typical face database contains ten different facial expressions for a single individual and different lighting circumstances. Similarly, the faces of forty persons were captured on camera in a variety of locations. A shaded background, an upright forehead, and a small head tilt are considered when taking photos. Each of the 400 photos, which are all in PGM format and measure 92 by 112, features 40 individuals. Ten pictures of a single person are shown in Figure 4.



Figure 4. O-RL Images of a single person

B Preprocessing

The RGB face images are converted to grayscale for rapid and simple image processing. The Grayimages from multiple marked face files with different features are lowered to a normal attribute of 128*128 to obtain a fixed number of topographies for all types of face files.

C Discrete-Wavelet-Transform (DWT)

Relevance includes retrieving biometric trait information, de-noising signals, compressing images, and locating pure frequencies. This approach can be used for both de-noising and compressing face images. This linear transformation shifts and scales the basis function wavelet. The wavelet is a minorupsurge with a finite interval and an mean estimate of zero. The WT overcomes the limitations of conventional transformations by

integrating time and frequency data. The wavelet's window width changes, and its power is uniformly spread in both progressive and adverse directions. Figure 5 shows its band pass characteristics in the frequency domain using L-PF, H-PF, and down sampling of image. The 2D-DWT signals for computerized images are executed via sub-band coding. Sub-band analysis is used to extract images in estimated forms from both headings. To complete the rise of discernable scaling and wavelet duties in n1 tests in horizontal headings and n2 tests in vertical headings, the 2-D signal analysis employs taking over 2-D filter tasks based on the following circumstances.

The substitute-sampler mesh is augmented by two, with each sub-band representing 1/4 of the input factors. They are denoted by L-L, L-H, H-L, and H-H, respectively.

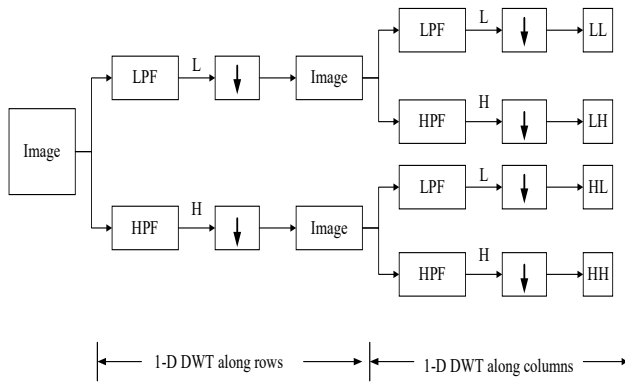


Figure. 5 2D-DWT Block Diagram

Low frequency components make up the LL band, an approximation band that contains a significant amount of information about the original image. High frequency components are represented by the detailed bands LH, HL, and HH, which, in turn, supply all possible edge information that enhances an original image. Four sub-bands, each comprising 64 * 64 pixels, are created from the original 128 * 128 image. We only utilize 64*64 LL band coefficients in our method.

D Histogram of Oriented Gradient (HOG)

It is resistive to modifications in occurrence and may be identified by people in face photographs because it operates with local matrix cells. The following grains, dx and dy , use both upright and straight slopes to calculate the H-OG of the L-L Sub-band.

$$dx = [-1 \ 0 \ 1]$$

$$dy = [-1 \ 0 \ 1]^T$$

The slope level and paths are estimated using the Equations shown.

$$\text{Magnitude of gradient} = \sqrt{(I_x^2 + I_y^2)} \quad (7)$$

$$\text{Angle of slope} = \arctan(I_y/I_x) \quad (8)$$

To determine the L-L band matrix's H-OG, the L-L band matrix is split into 8x8 sections. The slopes go from 0 to 180 instead of 0 to 360 degrees for an 8*8 matrix, and they are unspecified slopes since their minus values are represented by identical numbers. shows how the size and orientation of the HOG gradient for the 8*8 region are determined. The bin is selected based on the direction, and the value is filled in with the magnitude. The 9-bin histogram that results from inclusion all the factors in the 8*8 cells represents directions 0, 20, 40, 60, 80, 100, 120, 140, and 160. The angle dictates which bin is selected, while the magnitude dictates the value which enters the bin. The blue-circled coefficient value, which has an angle of 80 degrees and a matching magnitude of 2, is in the fifth bin. The slope at the red-circled factor has an slope of 10° and a scale of 4. The plotting splits equally between the first and second bins because 10° is intermediate between 0 and 20.

E Convolution.

To make the system unique and hybrid the features from different methods are extracted and combined using convolution. Which produces a different set of features for biometric detection by reducing noise and increasing edge detection and sharpening. By collaborating the products of the L-L and H-OG coefficients of the L-L band after one is reversed and shifted, it delivers a new result with a distinctive output. Next, the sum is calculated for each shift. The L-L band L (M1, N1) of size 64*64 and the L-L band matrix H (M2, N2) of size 42*42 are mixed.

F Matching.

The ED is the calculation of gap amid two points in vectors to link matches and changes between the vectors and is given in Equation.

$$D1(p, q) = \sqrt{\frac{1}{M} \sum_{i=1}^M (p_i - q_i)^2} \quad (10)$$

In this case, M represents feature vectors' attribute, p_i represents the DB feature vector's factors, and q_i represents the test vector's coefficients.

III. MODEL IMPLEMENTATION AND ANALYSIS

The mixing between two sets of features yields a generally notable third set of appearance for biometric recognition by decreasing noise and improving edge discovery and honing. By uniting the creations of the L-L coefficients and H-OG coefficients of the L-L after one is backed and moved, it creates a new result with a

distinctive output. The sum is calculated for every shift value.

A Performance Metrics.

a. False Accept Rate (FAR)

The pictures in the data-set pair the test-image, which is not in data-set, it is computed using the below equation.

$$FAR = \frac{\text{Number of Imposter persons accepted as genuine}}{\text{Total number of persons outside the database}} \quad (11)$$

b. False Rejection Rate (FRR)

The test image from the data set is not matched with the images in the data set and can be calculated using Equation.

$$FRR = \frac{\text{Number of genuine persons in the database rejected}}{\text{Total number of persons inside the database}} \quad (12)$$

c. Equal Error Rate. (EER)

It is the connection value of F-RR and F-AR in the chart.

d. Recognition Rate. (RR)

The total of rightful person effectively equalled in the data-set and computed using the below Equation 13.

$$TSR = \frac{\text{Number of genuine persons recognized correctly}}{\text{Total number of persons inside the database}} \quad (13)$$

B Implementation and Experimental Result Analysis.

Performance measures are used to assess the proposed model on the face datasets JA-FFE, L-SPACEK, YALE, and O-RL. The dataset's whole population is split to groups, Persons in Database (P-ID) and Persons not in Database (POD), to calculate F-RR and F-AR for the evaluation of % RR.

a. Result Investigation using JAFFE Dataset.

This dataset has ten individuals in total, each of which has twenty photos. Plotting of the performance indicators, including proportion F-RR, F-AR, and T-SR, is done for variations in threshold values. The variations of F-RR, F-AR, and T-SR drawn for P-ID variations with constant P-OD for sequences of 4:4, 5:4, and 6:4, respectively.

Figures 10 and 11 display the differences of F-RR, F-AR, and T-SR derived for constant P-ID with P-OD deviations for 5:4 and 6:4 sequences, respectively. It is noted that while the values of F-AR and T-SR rise with threshold values, the values of F-RR fall as threshold values rise.

The values of percentage EER, OTSR and MTSR for dissimilar PID and POD mixtures are prepared at Table 1. It is witnessed that, values of the OTSR percentage rise with rise in values of PID and percentage MTSR values constant at 100.

Table 1. Deviations of result METRICS with PID and POD for JAFFE Database.

PID	POD	%EER	%OTSR	%MTSR
5	5	17	76	100
6	5	15	81	100
7	5	13	83	100
5	6	16	76	100
5	7	13	76	100

b. Result Investigation using L-SPACEK Dataset

There are 152 individuals in total in this dataset, each of them has twenty photos. The parameters—F-RR, F-AR, and T-SR—are computed for deviation using threshold values and shown. Figures 6, 7, and 8 display the variations in F-RR, F-AR, and T-SR produced with variances in P-ID and constant P-OD for combinations of 40:60, 50:60, and 60:60, respectively. Figures 9 and 10 display the % F-RR, F-AR, and T-SR deviations for constant P-ID with P-OD variances for 60:40 and 60:50 sequences. While the F-AR and T-SR values increase as threshold values climb, the stated F-RR values decrease as threshold values grow. TSR values decline from 100% to 0% for threshold values less than 0.6, while they remain constant at 100% of the threshold values between 0.6 and 1.2. The T-SR values stay constant for higher threshold values, however the F-AR values rise from 0% to 100%, which is a disadvantage.

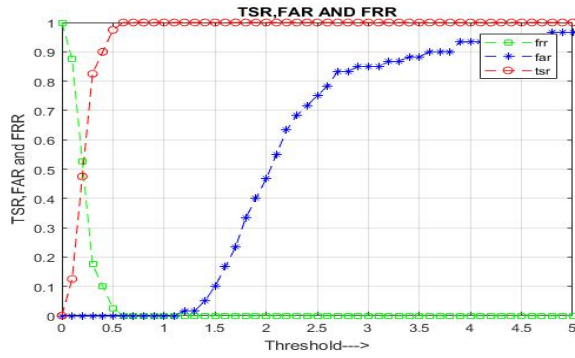


Figure. 6 The result deviations for 40:60 combination of PID and POD

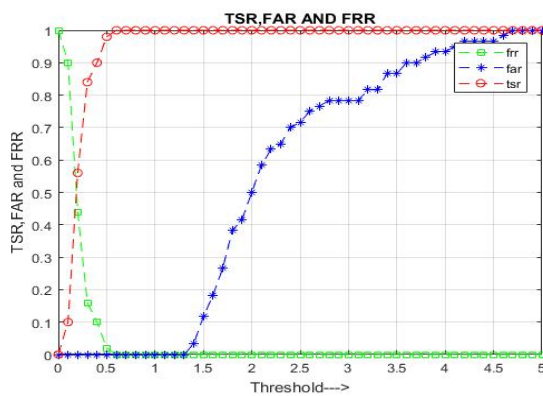


Figure.7 The result deviations for 50:60 combinations of PID and POD

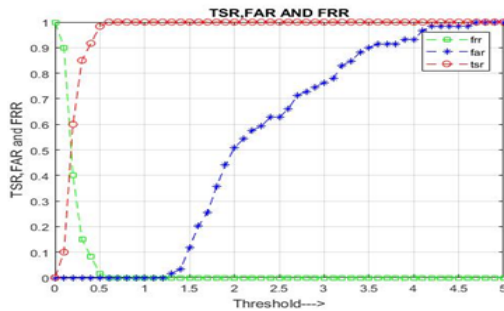


Figure. 8 The result deviations for 60:60 combinations of PID and POD

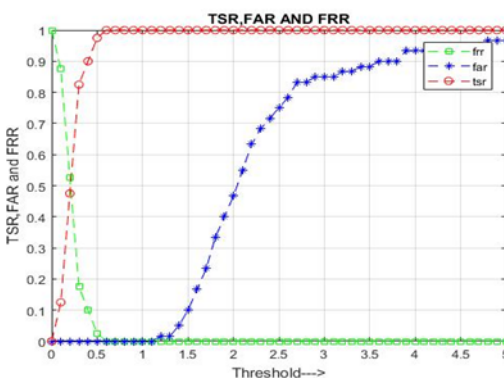


Figure.9 The result deviations for 60:40 combinations of PID and POD

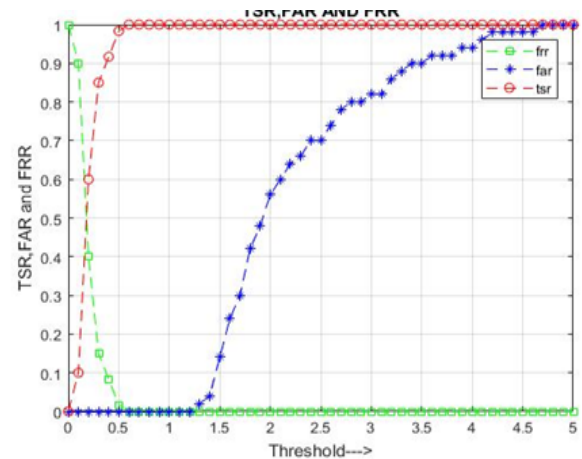


Figure.10 The result deviations for 60:50 combination of PID and POD

The changes in the quantities of E-ER, O-TSR and M-TSR for different P-ID and P-OD incorporations are established in Table 2.

Table 2. The result constraints with different PID and POD combinations for L-SPACEK.

PID	POD	%EER	%OTSR	%MTSR
40	40	0	99	100
40	50	0	99	100
40	60	0	98	100
50	60	0	99	100
60	59	0	99	100

For basestandard ranging 0.6 and 1.2, the calculated E-ER, O-TSR, and M-TSR percentage values for the various P-ID and P-OD combinations stay close to constant at 0, 99, and 100. The ratio M-TSR is 100% for better limit values, which is beneficial; yet, the most hostile condition is when the F-AR rises to 100%.

c. Findings Analysis using Y-ALE Dataset.

The charts show the processed findings of proportion F-RR, F-AR, and T-SR for change over in limit quantities. Figures 11, 12, and 13 display the F-RR, F-AR, and T-SR variations for in P-ID with constant P-OD for the corresponding combinations of 8:5, 9:5, and 10:5. For sequences of 8:6 and 8:7, correspondingly, Figures 14 and 15 present the deviations of F-RR, F-AR, and T-SR

foundpointed at regular P-ID with deviations in P-OD. It is viewed that when edge values grow, F-RR values decrease, however F-AR and T-SR values increase as levelstandardsmount. When the edge value is larger than 2, the T-SR moralsstayregular at 100%; when the edge value is less than 2, the T-SR values fall from 100%.

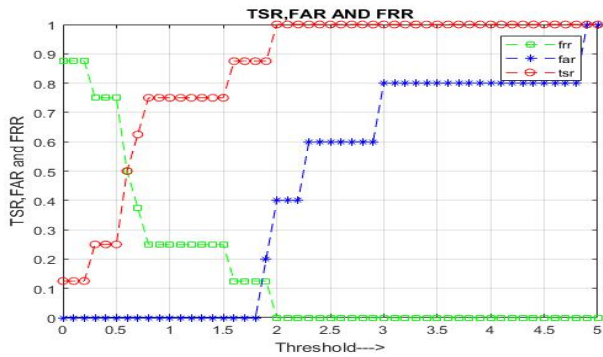


Figure.11 The result deviations for 8:5 combination of PID and POD

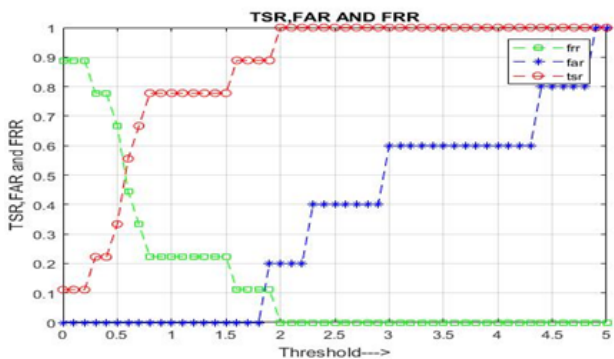


Figure.12 The result deviations for 9:5 combination of PID and POD

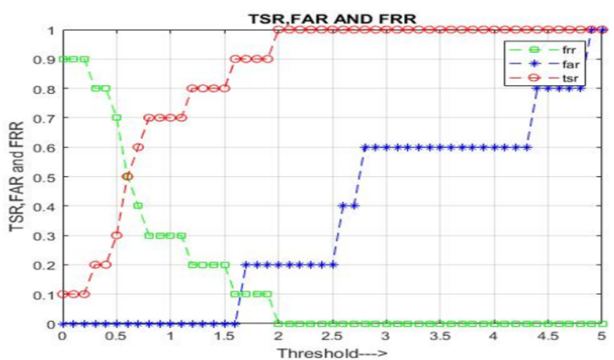


Figure.13 The result deviations for 10:5 combination of PID and POD

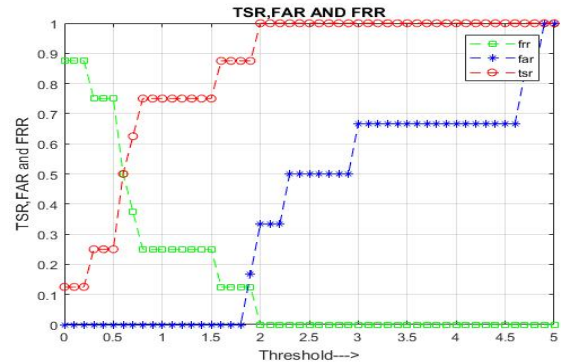


Figure. 14 The result deviations for 8:6 combination of PID and POD

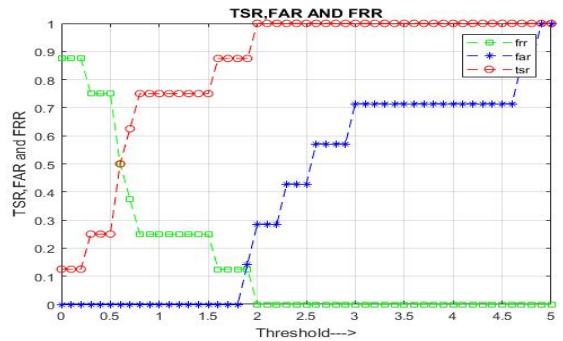


Figure. 15 The result deviations for 8:7 combination of PID and POD

Table 3 arranges the quantities of E-ER, O-TSR, and MT-SR for different P-ID and P-OD sequences. It is viewed that E-ER and O-TSR values remainsame for P-OD deviations and regular P-ID sequences of limit values closer to 1.9. The M-TSR is 100% for limit greater than 2, which is beneficial; however, the F-AR raises as the limit goes to 100%, which is the most undesired condition.

Table 3. The result constraints with different PID and POD combinations for YALE

PID	POD	%EER	%OTSR	%MTSR
8	5	12	88	100
8	6	12	88	100
8	7	12	88	100
9	5	11	89	100
10	5	10	90	100

d. Result Investigation using ORL Dataset

Every 40 rarepersons in the model has ten distinct photos. The working metrics, namely F-RR, F-AR, and T-SR, are computed for threshold value deviations and plotted in the diagrams. Figures 16, 17, and 18 display the

F-RR, F-AR, and T-SR variations for variances in P-ID with constant P-OD for the corresponding combinations of 10:10, 20:10, and 30:10. Figures 19 and 20 display the F-RR, F-AR, and T-SR deviations for constant P-ID with POD deviations for 10:20 and 10:30 sequences, respectively. While the F-AR and T-SR values improve as edgestandardshike, the F-RR values decrease as threshold raise. When the edge value is less than 2.1, the T-SR value reduces from 100% to 0%, but it continuessame at 100% of the edge values starting at 2.1. The T-SR values stay constant at higher threshold values, however the F-AR values rise from 0% to 100%, which is a drawback.

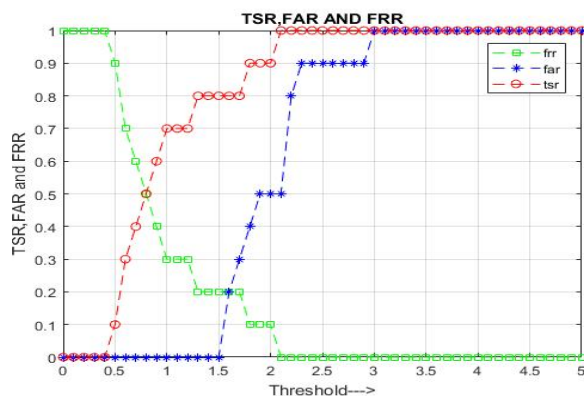


Figure.16 The result deviations for 10:10 combination of PID and POD for ORL dataset

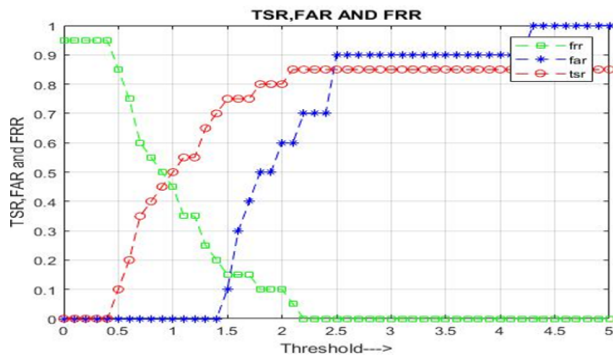


Figure.17 The result deviations for 20:10 combination of PID and POD for ORL dataset

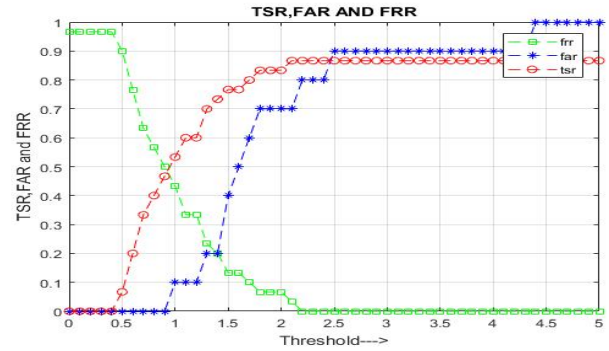


Figure.18 The result deviations for 30:10 combination of PID and POD for ORL dataset

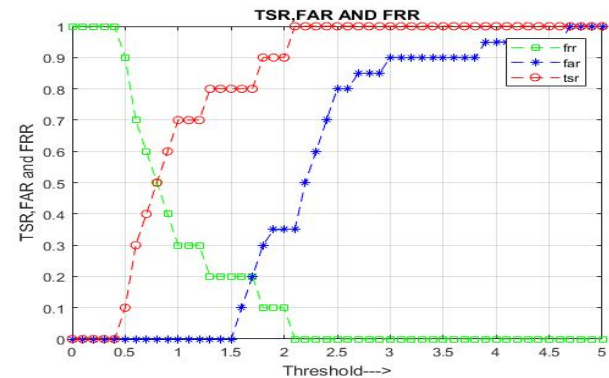


Figure.19 The result deviations for 10:20 combination of PID and POD for ORL dataset

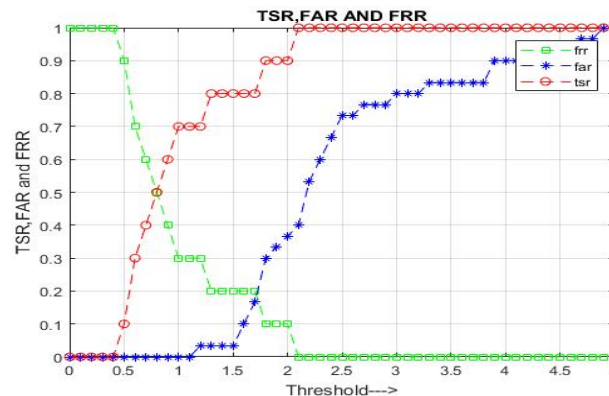


Figure.20 The result deviations for 10:30 combination of PID and POD for ORL dataset

Table 4 records the changes in the values of E-ER, O-TSR, and M-TSR for separate P-ID and P-OD mixes. For changes in P-OD and continuous P-ID sequences of limitstandards about 1.7, the % E-ER and O-TSR estimates are closelycontinuous at 20 and 80, respectively. The M-TSR is 100% for edge values more than 2.2 with persistent P-ID at 10, which is helpful; but, the F-AR raises as the edge goes to 100%, which is the most unsoughtconsideration.

Table 4. The result constraints with different PID and POD combinations for ORL

PID	POD	%EER	%OTSR	%MTSR
10	10	20	80	100
10	20	20	80	100
10	30	18	80	100
20	10	15	75	85
30	10	20	74	86

C Comparison of results using different face datasets

Table 5 shows the execution measures E-ER and O-TSR in relative to edge values via the L-Spacek, Y-ALE, JA-FFE, and O-RL face data-sets. The L-Spacek dataset's E-ER and O-TSR are zero and 99%, correspondingly, because there are exceptionally few alterations in each image. It is noted that at minorthreshold, the test images are paired with dataset face images. For face datasets to match query and dataset photos, high limit values are required. Since all person's view and other facial image variances are high, the average E-ER values are high while the average O-TSR values are low.

Table 5. Comparison of results for different datasets

Metrics	Face Datasets			
	L-Spacek	YALE	JAFFE	ORL
Average Threshold	0.5	1.8	1.4	1.75
% Average EER	0	11.4	13.5	19
% Average OTSR	99	89	77.5	77.8

D Comparison of Projected Technique with Current Techniques

Table-6 exhibits the M-TSR utilizing the estimated method's O-RL, Y-ALE, and JA-FFE data-sets in relation to the current methods available. The study confirmed that when assessed to present methods, the calculated method makes better the M-TSR.

Table 6. %MTSR Comparison of Projected Technique with Current Approaches

Methods	%MTSR		
	ORL	YALE	JAFFE

	dataset	dataset	dataset
Erhu Zhang et al.,	74.06	80.77	-----
Shailaja and Anuradha	87	92.8	-----
Yubo Wang et al.,	----	-----	92.4
Projected method	94.2	100	100

IV. CONCLUSION

A hybrid domain feature convolution for human recognition from face images is proposed. The 128x128 uniform form is created by converting the face photos from various face databases with varying dimensions. The D-WT is used to reduce and reject noise from face photographs by taking into account the LL band's low frequency coefficients. The L2-Norm from the LL band was used to normalize the HOG, resulting in a cascade connection between DWT and HOG. The L-L matrix of size 64x64 is fused with the H-OG remodelled medium of size 42x42 to construct efficient and trustworthy final features of face images. To measure outcome limitations, the ED methodology is retained. It is surveyed that the recommended approach beats the current facial recognition techniques. In the future, ED may be replaced by deep learning methods for comparing big image datasets.

REFERENCES

- [1] X. Li, Z. Yang and H. Wu, "Face Detection Based on Receptive Field Enhanced Multi-Task Cascaded Convolutional Neural Networks," in *IEEE Access*, vol. 8, pp. 174922-174930, 2020.
- [2] L. Cao, J. Liu, K. Du, Y. Guo and T. Wang, "Guided Cascaded Super-Resolution Network for Face Image," in *IEEE Access*, vol. 8, pp. 173387-173400, 2020.
- [3] H. Mo et al., "A Multi-Task Hardwired Accelerator for Face Detection and Alignment," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 4284-4298, Nov. 2020.
- [4] M. Kolhar, F. Al-Turjman, A. Alameen and M. M. Abualhaj, "A Three Layered Decentralized IoT Biometric Architecture for City Lockdown During COVID-19 Outbreak," in *IEEE Access*, vol. 8, pp. 163608-163617, 2020.
- [5] Z. Dong, J. Wei, X. Chen and P. Zheng, "Face Detection in Security Monitoring Based on Artificial Intelligence Video Retrieval Technology," in *IEEE Access*, vol. 8, pp. 63421-63433, 2020.
- [6] A. K. Nsaif et al., "FRCNN-GNB: Cascade Faster R-CNN With Gabor Filters and Naïve Bayes for Enhanced Eye Detection," in *IEEE Access*, vol. 9, pp. 15708-15719, 2021.

- [7] N. Zhou, R. Liang and W. Shi,"A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection," in *IEEE Access*, vol. 9, pp. 5573-5584, 2021.
- [8] J. Wei,"Video Face Recognition of Virtual Currency Trading System Based on Deep Learning Algorithms," in *IEEE Access*, vol. 9, pp. 32760-32773, 2021.
- [9] E. M. F. Caliwag, A. Caliwag, B. -K. Baek, Y. Jo, H. Chung and W. Lim,"Distance Estimation in Thermal Cameras Using Multi-Task Cascaded Convolutional Neural Network," in *IEEE Sensors Journal*, vol. 21, no. 17, pp. 18519-18525, 1 Sept.1, 2021.
- [10] Y. Pang, J. Cao, Y. Li, J. Xie, H. Sun and J. Gong,"TJU-DHD: A Diverse High-Resolution Dataset for Object Detection," in *IEEE Transactions on Image Processing*, vol. 30, pp. 207-219, 2021.
- [11] R. Liu, Z. Yuan and T. Liu,"LearningTBox With a Cascaded Anchor-Free Network for Vehicle Detection," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 321-332, Jan. 2022.
- [12] J. Zhong, Z. Liu, C. Yang, H. Wang, S. Gao and A. Núñez,"Adversarial Reconstruction Based on Tighter Oriented Localization for Catenary Insulator Defect Detection in High-Speed Railways," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 1109-1120, Feb. 2022.