Received: April 29, 2023. Revised: June 26, 2023. 124

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Image Based Person Re-Id Using GAN, DWT, and MLTP with ANN Classifier

Arun kumar Doddanahalli Ramalingegowda[1]* Krishna Alabujanahalli Neelegowda[1]
Bangalore Gnanamurthy Prasad[2]

[1]Department of Computer Science and Engineering, SJB Institute of Technology, Affiliated to Visvesvaraya Technological University, Karnataka, India
Affiliated to Visvesvaraya Technological University, Karnataka, India
[2] Dayananda Sagar College of Engineering, Karnataka, India
* Corresponding author's Email: arunkumardr1987@gmail.com

Abstract: The recognition of an individual from images/videos captured from different cameras in real-world surveillance scenarios is the Person's re-identification. We propose image-based person reidentification (Re-id) using GAN, DWT, and MLTP with ANN classifiers. The central server database contains four benchmarked face databases by considering six images per person. The generative adversarial network (GAN) is used to convert side-angled test images into frontal images for an effective person Re-id. The discrete wavelet transform (DWT) and modified local ternary pattern (MLTP) techniques are used on images of the central server database and generated frontal test images to extract initial features. The Final features are computed by using the convolution of the DWT-LL band and MLTP features. The artificial neural network (ANN) is used to categorize and recognize test images. The existing research based on HOG, Hu Invariant moments, CNN, and SVM is presented and published, having a recognition accuracy of 97%. However, our projected technique uses GAN, which improved recognition accuracy to 99% compared to existing techniques having low recognition accuracy.

Keywords: ANN, DWT, GAN, MLTP, Person re-identification.

1. Introduction

An individual is effectively identified based on images captured from several closed-circuit television (CCTV) is known as person Re-identification (Re-Id). The query person-of-interest is to be determined whether that person has appeared at another different time taken by different cameras or even the same camera at a dissimilar time instant. The Re-Id is in very much demand for public safety from bad incidents and cuts down crimes. It plays an important role in identifying the interested person and tracking that person from different cameras. The person Re-Id approaches are divided into two classes, such as image-based and video-based. These approaches are based on appearance, gait, or feature extraction techniques. The increasing demand for personal Re-Id with generative adversarial networks (GAN) [1], hybrid features, and ANN classifiers have gained attention in the research of computer vision. Person Re-Id is possible with CCTV footage only when a person stands straight in front of a CCTV, whereas the images of a person with different pose angles captured by CCTV lead to identification difficulty. It quickly improved attention in computer vision and pattern recognition of its reputation for numerous applications, viz., video surveillance, human-computer interaction, robotics, content-based video retrieval, *etc*., [2].

Re-identification is inspiring due to several lookouts, low resolutions, illumination variations, unrestricted postures, occlusions, heterogeneous modes, composite camera settings, background clutter, untrustworthy bounding box generation, etc. These tasks are considered unsolved problems for real-time applications [3]. The significant applications are video surveillance plus pedestrian hunting, multi-camera pursuing, and behaviour investigation [4].

The guidelines for person reidentification tasks are the feature learning method and the metric learning technique. The methods of robust feature extraction are the ensemble of localized features [5], symmetry-driven accumulation of local features [6], the local maximal occurrence representation [7], color & texture features, and the fusion net of CNN features and manual features [8]. The weighted histogram feature of middle dense and significant blocks is based on the Gauss mixture model [9]. The simple and effective metric algorithm [10] is the metric learning scheme that measures the changes among samples by likelihood ratio and acquires the Mahalanobis metric, which replicates the log-likelihood ratio property.

Contribution: The proposed research contributions are as follows

1. The GAN converts captured test images with different angles to frontal images for better performance of a proposed algorithm.
2. The features of the central server database and generated test images are obtained by convolution of DWT and MLTP techniques for the identification of persons.
3. The ANN Classifier is used for the classification of images for person Re-id and to test the performance of the proposed algorithm.

Research paper organization: The paper is arranged as a literature survey is deliberated in section 2, and in section 3, the projected new method is pronounced. The investigational result analysis is conferred in section 4. The conclusion with future work is given in section 5.

2. Literature survey

The research techniques presented by various authors on person re-identification are described in this section. Mang Ye *et al.,* [11] conducted a wide-ranging summary of person Re-ID from three dissimilar viewpoints, deep feature representation learning, deep metric learning, and ranking optimization. They introduced an evaluation metric (mINP) for person Re-ID, representing the cost of outcome in all the right contests, which delivers norms to assess the Re-ID system for actual applications. The limitation of this paper is; it conducted a comprehensive survey with an in-depth analysis of person Re-ID and open-world settings has given to address various practical challenges. X. L. Liao *et al.,* [12] formulated person re-ID based on a deep structured prediction with the structural interactions among person images by using neural-style-transfer-based structure sampling and fully parameterized energy networks. The limitation is that the re-ID models the structural interactions of person images in the output label space.

Le An *et al.,* [13] proposed a reference-based process for person Re-Id from different cameras. A subspace is formed in which the correlations of the reference data from cameras placed at different locations are exploited using regularized canonical correlation analysis (RCCA). The probe data and the gallery data are projected onto the RCCA subspace, and the reference descriptors (RDs) of the probe and gallery are produced by calculating the comparison among them for person reidentification. It is used for single-shot person re-identification only and not for multi-shot person re-identification.

S. Tan *et al.,* [14] proposed the dense invariant features (DIFs) to model the differences in images across camera sights. The concept of maximizing the margin and minimizing the error on the fused features is used to obtain optimal space for ranking. The limitation is that not much attention to learning aligned patches from images of the probe view is considered. Li *et al.,* [15] planned a low-dimensional metric space feature extracted method from different views of the same person. The model simultaneously learns a uniform metric space for several features that is practical for multi-feature fusion. The limitation of this method is that; action recognition and scene recognition in person reidentification are not possible. Liu *et al.,* [16] projected a person Re-Id based on feature transformation and metric learning technique. The local models are from subsets of training samples by the global model. The learned local models improve the strength of discriminative and generalization aptitude. The limitation is that the algorithm is not sensitive to the number of local models within a wide range.

Yang *et al.,* [17] proposed recurrent auto encoder (RAE) framework to unify two approaches. The three modules of RAE are a feature-transfer (FT) module, a pixel-transfer (PT) module, and a Fusion module. An encoder is used in the FT module to map source and target images to a shared feature space. The decoder is used to reconstruct original images with its features in the PT module. The bilinear pooling layer fuses both high and low-level knowledge with the two proposed module's knowledge. The limitation is that the accuracy is a little low and the time consumption process. Zhang *et al.,* [18] proposed a part-based nondirect coupling embedded generative adversarial network (GAN) system to obtain common features of dissimilar positions of the identical individual. The network features are robust for position deviations of an individual, and no auxiliary posture info and computational cost are

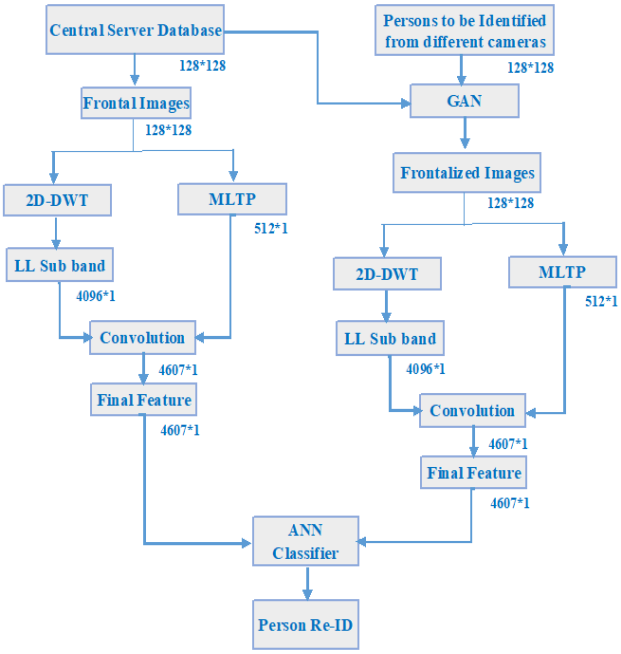


Figure. 1 Proposed Re-Id model

Figure. 2 The ORL database image samples of four persons

Figure. 3 The Indian female dataset image samples of four persons

essential in the test stage. The limitation is that a decoder that combines the bilinear interpolation feature maps and the deconvolution feature maps leads to an increase in complexity. Tanveer and Habib Mahmood [19] proposed an unsupervised method for the person Re-Id based on the GAN. The deep convolutional generative adversarial network for feature learning from image sequences is used. The limitation is that they do not use any labels so this leads to reduce identification issues. Haghighi and Taheri [20] projected a GAN based on bipartite graphs to transfer images to precise canonical postures in order to overlook undesired side possessions of the postures. The person's Re-Id is based on feature extraction networks, with different undisputed postures. The limitation is that it transfers images to only specific canonical poses; hence the accuracy of identification reduces.

3. Proposed method

The image-based person Re-Id using GAN, DWT, MLTP, and ANN classifier is presented in this section for a better recognition rate, and the block diagram is shown in Fig. 1. Person Re-Id lies in face image frontalization and personal identification in terms of the association among training and testing classes in ANN.

3.1 Central server database

The four benchmarked databases, such as ORL, IM, IF, and CMU-PIE is considered to test the proposed method. The six images of each person with frontal and slight variations in front angles are considered in every database. The total number of persons is 180, with the total number of images being 1080 are considered to test the algorithm.

3.1.1. Olivetti research laboratory (ORL) face database:

It has forty people, with ten images of each person ensuing in a total of 400 images [21]. The images have dissimilarities in posture and expressions, with the size of every image 112X92 with the grayscale format. Fig. 2 shows the ORL database sample images of four people. The six images per person with frontal and slight front angle variations *i.e.*, 6x40=240 images, are stored in the central server database.

3.1.2. The indian female face database:

It has twenty people with twelve images per person resulting in a total of 240 images [22]. Each image has a size of 480X640 with facial orientations and expressions with smiles, laughter, sadness, disgust, and neutrality. The Indian female dataset sample face images of four people are shown in Fig. 3. The six images per person with frontal and slight front angle variations *i.e.*, 6x20=120 images, are stored in the central server database.

3.1.3. The Indian male face database:

It has twenty persons with eleven images of each person resulting in a total of 220 images [22]. Every

Figure 4. The Indian male dataset face image samples of four persons

Figure 5. The CMU Multi-PIE Dataset of four persons

image has a size of 480X640 with various facial orientations and expressions. The Indian Male dataset sample images of four people are revealed in Fig. 4. The six images per person with frontal and slight front angle variations i.e., 6x20=120 images are stored in the central server database.

3.1.4. The CMU Multi-PIE face database

It has 337 persons, and images are captured during four sessions over a five-month period with 15 angles, 19 illumination conditions, and a variety of facial expressions resulting in a total of 7,50,000 images [23]. Every face image in the database has a size of 128X128. Fig. 5 shows the CMU Multi-PIE dataset samples of four persons, and the image has a size of 128X128. The six images per person with frontal and slight front angle variations ie., 6x100=600 images, are stored in the central server database.

3.2 Test section

The persons to be identified from different cameras in real-time are simulated to test the proposed method by considering four side-angled images per person from ORL (4 IPP X 40 = 160 images), Indian female (4 IPP X 20 = 80 images), Indian male (4 IPP X 20 = 80 images), and the CMU multi-PIE (4 IPP X 100 = 400 images) face databases. The total number of persons considered is 180 with the total number of images being 720 in the test section.

3.3 The preprocessing of face images

The different face image sizes in the typical face datasets and color images are transformed into grayscale images and resized to 128x128. All images are of .png format.

3.4 The generative adversarial network (GAN) model architecture

It is used for face frontalization from side-angled face images and is the most recent important algorithm in the machine-learning research area based on a deep-learning generative model proposed by Ian J. Goodfellow in 2014 [24]. It contains two neural networks such as the generator model and the discriminator model, which oppose each other to achieve more accurate estimations. The goal of the generator model is to create new possible images that can fool the discriminator while the goal of the discriminator model is to recognize whether the image is real from the original data or fake generated by the generator model.

The face frontalization GAN architecture consists of one generator and one discriminator. The side-angled images to be tested are fed to the generator to generate the frontal images then the discriminator takes the generated frontal (GF) image and real frontal (RF) image from the central server database to classify which one is the real and which one is the fake image as shown in Fig. 6.

3.4.1. The generator:

The generator has two parts viz., encoder and decoder. The encoder converts images into eigenvectors; the decoder reconstructs the frontal face with eigenvectors. The structure of the generator is shown in Fig. 7.

The original side-angled profile image size of 128x128 is fed to the encoder of the generator and the traditional deep convolution structure is used to obtain 512-dimensional eigenvectors from 3 channels of RGB as it is the color image. The dimensions of the structure of the encoder are given in Table 1.

The encoder consists of seven layers and 1 to 5 layers contain 2D convolution, batch normalization, and Rectified Linear Unit (ReLU). Layer 6 has only 2D convolution, and layer 7 has only max pooling. The color side angled profile image with a size of 128x128x3 is fed to convolution layer 1 of convolution neural network (CNN), which has a kernel size of 3x3, stride 1x1, and padding 2x2 to obtain the output of size 130x130 with 16 channels. The size of the output is the size of the next input layer and is calculated using Eq. (1).

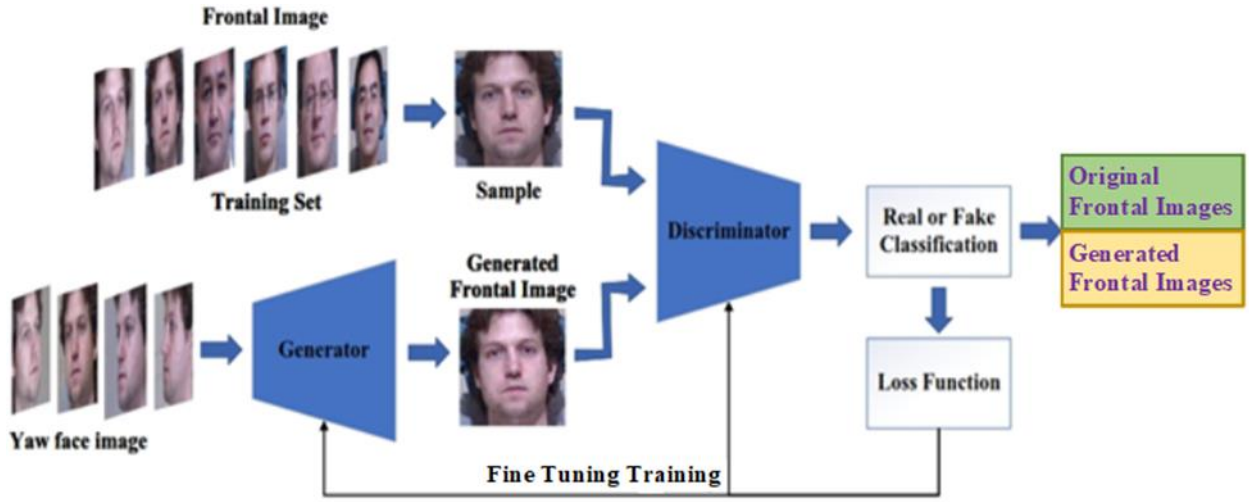


Figure. 6 The face frontalization GAN system

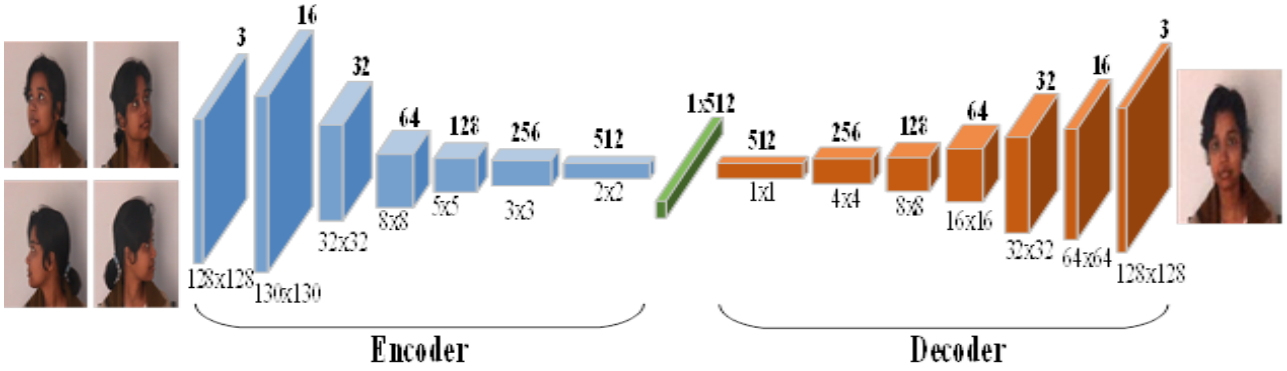


Figure. 7 The Generator structure

Table 1. The structural dimensions of the encoder

Layer	Input		Kernel Size/ Stride/ Padding	output	
	size	channel		size	channel
1	128x128	3	3/1/2	130x130	16
2	130x130	16	8/4/2	32x32	32
3	32x32	32	8/4/2	8x8	64
4	8x8	64	4/2/2	5x5	128
5	5x5	128	4/2/2	3x3	256
6	3x3	256	4/2/2	2x2	512
7	2x2	512	2/2/0	1	512

$$Size_{output} = \left(\frac{Size_{input} + 2P - K}{S}\right) + 1 \qquad (1)$$

where $Size_{input}$ = input image size, P is padding size, K is kernel size, and S is the size of the stride.

The kernel is used to obtain the features of the images and moves over the input data to perform the dot product with the sub-region of input data. The kernel size is the size of the matrix of the filter and the stride is the step size of the kernel when sliding through the image. The stride of 1x1 is the kernel slides through the image pixel by pixel. The number of pixels added to an image is padding, if the zero padding is set to one then a pixel value of zeros is added to the image. The illustration of the working process of kernel, stride, and padding on the sample matrix values is shown in Fig. 8.

The decoder input is the 512-dimensional eigenvectors obtained from the encoder and is reconstructed to achieve the color face image of 128X128 by using deconvolution. Table 2 shows the structure of the decoder.

The decoder consists of six layers. In layers 1 to 6, each layer contains 2D transposed convolution, batch normalization, and Leaky ReLU but layer 6 has an extra Tanh activation function.

Kernel Filter:

From Table 2, the matrix size of 64x64x16 is fed to transposed convolution layer 6, which has a kernel size of 4x4, stride 2x2, and padding 1x1 to obtain output as the generated color frontal image at a size of 128x128 with 3 channels. The output size is the input size of the next layer. It is calculated from Eq. (2).

$$Size_{output} = \left[\left(Size_{input} - 1\right) \times S\right] - 2P + K \qquad (2)$$

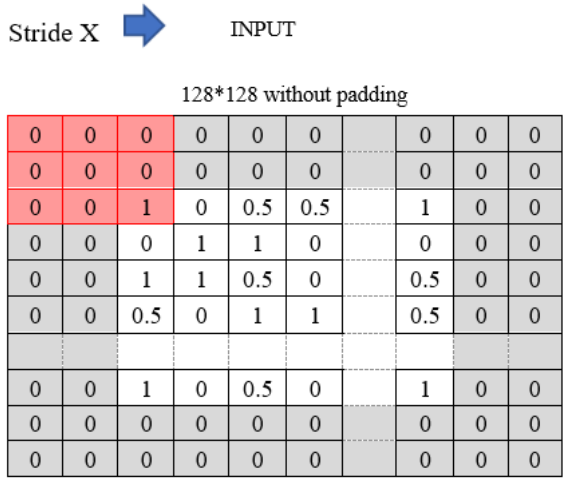


Figure. 8 The illustration of kernel, stride, and padding

Table.2. The decoder structure dimensions

Layer	Input		Kernel Size/ Stride/ Padding			output	
	size	channel				size	channel
1	1	512	4	1	0	4x4	256
2	4x4	256	4	2	1	8x8	128
3	8x8	128	4	2	1	16x16	64
4	16x16	64	4	2	1	32x32	32
5	32x32	32	4	2	1	64x64	16
6	64x64	16	4	2	1	128x128	3

where $Size_{input}$ is the input image size, *P is* padding size, *K* is kernel size, and *S is* the size of stride.

3.4.2. The discriminator:

It contains only the encoder part which is used to resolve whether the input image is a real original frontal face or a fake generated frontal face. The encoder structure of the discriminator is revealed in Fig. 9.

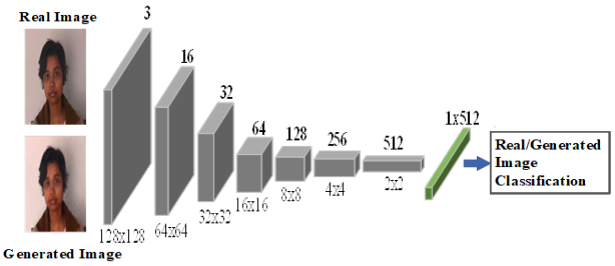


Figure. 9 Structure of discriminator

Table 3. The discriminator structure

Layer	Input		Kernel Size/ Stride/ Padding			output	
	size	channel				size	channel
1	128x128	3	4	2	1	64x64	16
2	64x64	16	4	2	1	32x32	32
3	32x32	32	4	2	1	16x16	64
4	16x16	64	4	2	1	8x8	128
5	8x8	128	4	2	1	4x4	256
6	4x4	256	4	2	1	2x2	512
7	2x2	512	4	2	1	1	512

The real frontal and the generated frontal image is fed to the encoder in the discriminator which is the traditional deep convolution structure used to obtain 512-dimensional eigenvectors to classify the real or generated frontal image. The details of the encoder structure are revealed in Table 3.

The encoder in the discriminator consists of seven layers. Layer 1 has 2D convolution and Leaky ReLU with a negative slope of 0.2. Layers 2 to 6, contain 2D convolution, batch normalization, and Leaky ReLU with a negative slope of 0.2, and Layer 7 has only 2D convolution and Sigmoid activation function.

3.4.3. The training process in face frontalization GAN model

The training process in the GAN model means the way to find the appropriate weights of each neural node which obtain the lowest values of the loss function. The loss function is a way to assess how effectively the algorithm models the dataset. It is one of the important parameters of neural networks which used as a prediction error of neural networks. If the predictions are completely incorrect, the loss function will return a higher value otherwise it will return a lower number of loss functions when the predictions are fairly good so loss functions are related to model accuracy. The one cycle of tuning the algorithm to try and improve the model through the full training dataset is called an epoch. The popular loss functions

Figure. 10 Six Samples of each person connected to the discriminator for training GAN

used in this section are Mean Absolute Error (MAE) and Mean Square Error (MSE). The MAE or L1 Loss function minimizes the sum of the absolute differences (S) between the existing target value (y_i) and the estimated values ($f(x_i)$) as shown in Eq. (3).

$$L1 = \sum_{i=1}^{n} |y_i - f(x_i)| \qquad (3)$$

where *xi* denotes the feature set of a single sample and *n* denotes the samples of absolute differences.

The MSE or L2 Loss function minimizes the sum of the square of the differences (S) between the existing target value (Y_i) and the estimated values ($f(x_i)$ where x_i denotes the feature set of a single sample and n denotes the samples of absolute differences as shown in Eq. (4).

$$L2 = \sum_{i=1}^{n} (y_i - f(x_i))^2 \qquad (4)$$

The L2 error is larger in the case of outliers compared to L1 from equations 3 and 4, as the variance among an incorrectly forecasted target value and the original target, value is quite large and squaring it leads to even larger values. Hence the L1 loss function is more robust and not affected by outliers. The L2 loss function tries to regulate the model according to the outlier values, even at the expense of other samples. Hence, the L2 loss function is extremely sensitive to outliers in the dataset. An ideal model would have L1 and L2 errors of 0 which means the estimated images are completely the same as the target image. The datasets used for the training process are shown in Fig. 10.

The L2 loss function tries to regulate the model according to the outlier values, even at the expense of other samples. Hence, the L2 loss function is extremely sensitive to outliers in the dataset. An ideal model would have L1 and L2 errors of 0 which means the estimated images are completely the same as the target image. The datasets used for the training process are shown in Fig.10.

3.4.4. The new database of face frontalization GAN model

The main goal of the face frontalization GAN system is to increase the recognition rate in the person

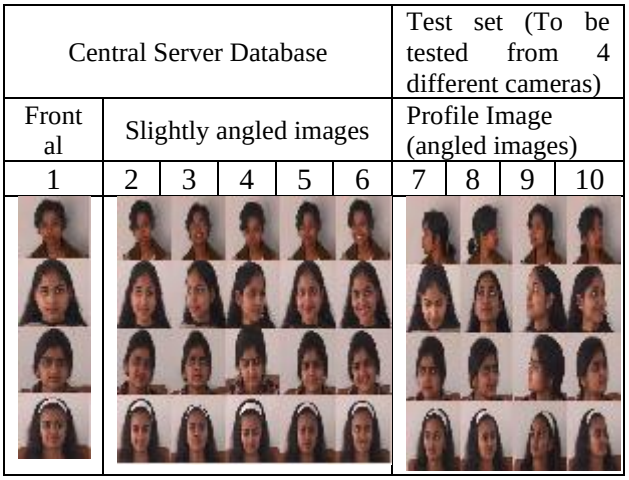


Figure. 11 The original database before using GAN

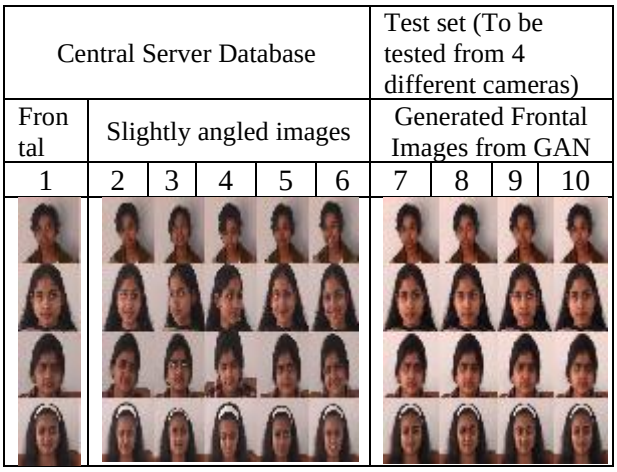


Figure. 12 The new database after using GAN

Re-id system by eliminating the side-angled profile images to the frontal images. After the training process on the GAN system, the new frontal images are created from the test side-angled profile images for the person Re-Id. The images from 1 to 6 are frontal and little angled face images which are deposited in the central database for reference and the last four images from 7 to 10 of each person to be identified are captured from four different CCTV with pose angles as shown in Fig. 11. The GAN is used to generate four frontal images from four side-angled images as shown in the new database given in Fig. 12. The comparison of images in the original database and the new database is shown in Fig. 11 and Fig. 12, respectively.

3.5 Discrete wavelet transform (DWT)

It is a tool to convert spatial domain images into frequency domain [25, 26] consisting of low-frequency and high-frequency coefficients using low and high pass filters with decimation by 2. The output of DWT has four bands of equal size corresponding

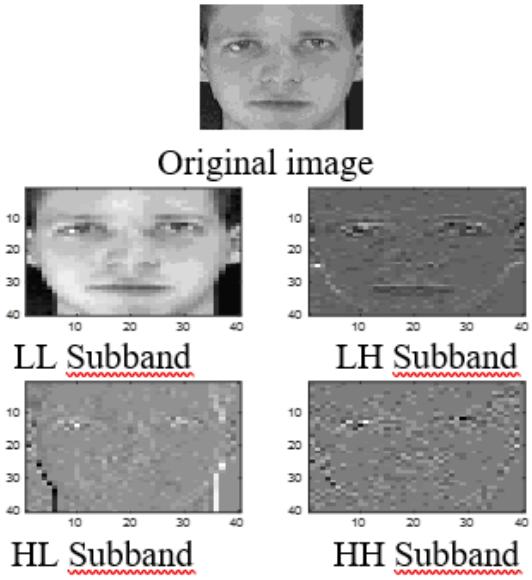


Figure. 13 DWT decomposition

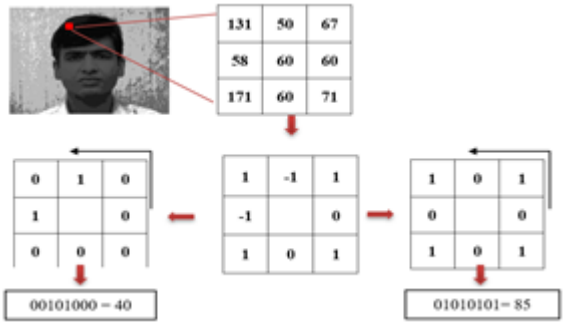


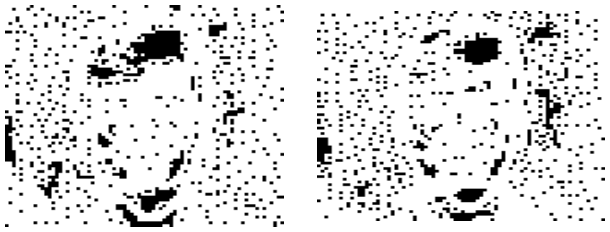
Figure. 14 The illustration of MLTP

to three high-frequency bands and one low-frequency band. The significant data of the original image is available in the low-frequency band. The insignificant edge data of the image exists in three high-frequency bands corresponding to horizontal, vertical, and diagonal edges. In our method, the low-frequency band coefficients are considered by ignoring high-frequency band coefficients results in low dimensional features for high-speed computation. The coefficients of low-frequency band low- low (LL), and high-frequency bands viz., low- high (LH), high-low (HL), and high-high (HH) are acquired based on Eq. (5-9) for the image matrix of size 2X2. DWT decomposition is as shown in Fig. 13.

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{5}$$

$$LL = \frac{a+b+c+d}{2} \tag{6}$$

$$LH = \frac{a+b-c-d}{2} \tag{7}$$



(a) (b)

Figure. 15 The LSP and RSP image of MLTP: (a) LSP image and (b) RSP image

$$HL = \frac{a-b+c-d}{2} \tag{8}$$

$$HH = \frac{a-b-c+d}{2} \tag{9}$$

Where a, b, c, and d are the constants of the 2X2 matrix.

3.6 Modified local ternary pattern (MLTP) [27]

It is the modified version of the local ternary pattern (LTP), which is applied to images of size 128x128 to get effective features by setting zero threshold values in LTP. The three shades of binary values are obtained by comparing neighbour pixel intensity values with the centre pixel intensity of the 3X3 block as given in Eq. (10).

$$b_i(x_c, y_c) = \begin{cases} -1, & P_n < P_c \\ 0, & P_n = P_c \\ 1, & P_n > P_c \end{cases} \tag{10}$$

Where $b_i\ (x_c, y_c)$ = Binary value of neighbor pixels, P_n = value of neighbor pixel intensity, P_c = value of center pixel intensity.

The binary $b_i\ (x_c, y_c)$ is split into two segments viz., left side pattern (LSP) and right-side pattern (RSP). The negative binary values of the 3X3 matrix are considered as positive binary values and positive binary values as zeros in LSP. The negative binary values of the 3X3 matrix are considered as zero binary values and positive binary values as positive binary values in RSP. The binary values of LSP and RSP are transformed into decimal values and assigned to center pixel of 3X3 using Eq. (11).

$$MLTP = \sum_{i=0}^{7} b_i(x_c, y_c) 2^i \tag{11}$$

The computation of MLTP on 3X3 matrix is illustrated in Fig. 14. The neighbouring pixel intensity values are related with the centre pixel intensity value and converted into binary based on Eq. (11). The LSP is obtained by assigning zeros to 1's

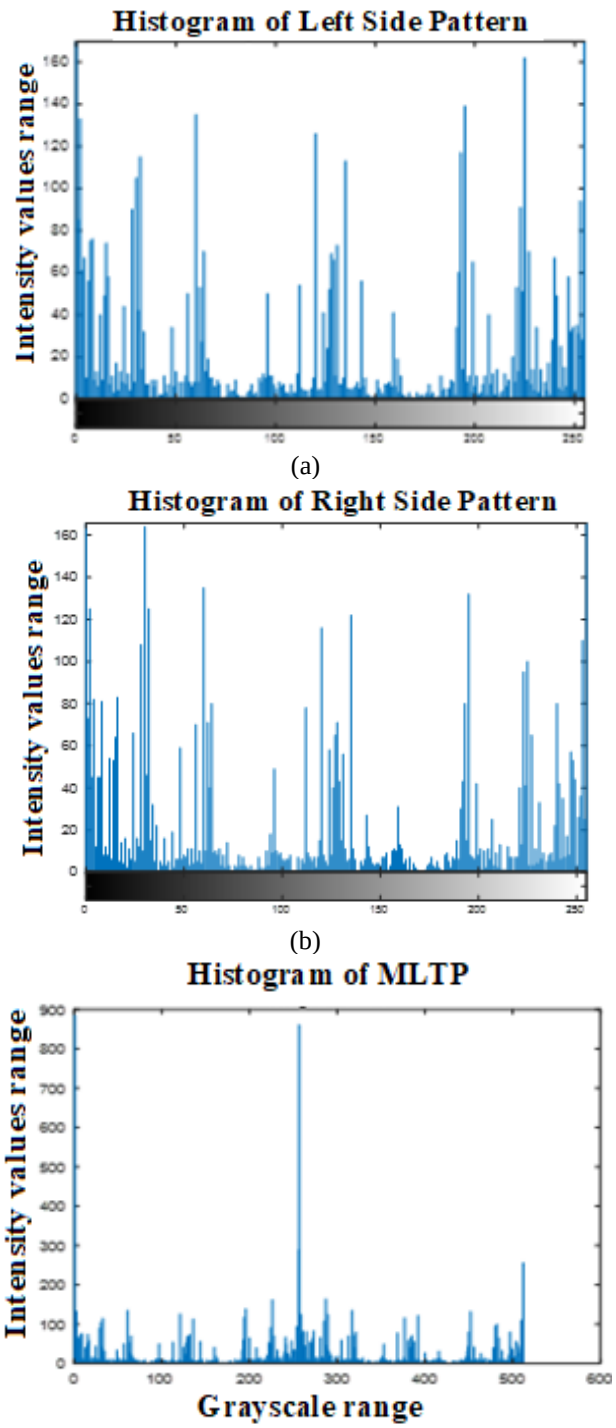


Figure. 16 Histogram on: (a) LSP, (b) RSP, and (c) concatenation

and 1's to -1's. The RSP is obtained by assigning zeros to -1's. The LSP and RSP binary values are converted into decimal values.

The images of size 128X128 are considered with pixel intensity values matrix and applied MLTP to convert into LSP and RSP as shown in Fig. 15. It is noticed that the images are not clear compared to the original image but the coefficient values are effective in person Re-Id efficiently. Then, the histogram is applied to LSP and RSP to obtain histogram coefficients of length 1X256 as shown in Fig. 16. The histograms are applied on LSP and RSP to decrease the number of features of LSP and RSP to 256 only. The concluding features of MLTP are obtained by concatenation of LSP and RSP histogram features as shown in Fig. 16c. The length of final MLTP features is only 512 for each image which is compressed compared to original image dimension of 128x128=16384, hence the computation time to compare test images with stored database images is less.

3.7 Convolution:

The features extracted by DWT and MLTP are fused using linear convolution technique for effective final features. The number of DWT features=64x64=4096, The number of MLTP features=256+256=512, and the final number of features=4096+512-1=4607.

3.8 ANN Classifier

It uses images and computational blocks accomplished by machine learning and pattern recognition for person Re-Id [28, 29]. The ANN used in our method has 3 layers such as input, hidden, and output. The final features from trained images are accepted by the input layer, which feeds them to the network. The input layer neurons are the number of final features, which is 4607. The action of each hidden component is computed by the hidden layer and the input units' activities weights on the contacts between the input and hidden unit. The hidden layer has 1 to 100 neurons to obtain the optimum model for classification performance. The output layer generates output units based on the hidden units' actions as well as the weights among the hidden and output units. The output layer neurons are equal to the number of classes and corresponding nodes in the output layer, in this case, are 40, 20, 20, and 100 based on the number of people in the ORL, Indian female, Indian male, and CMU- PIE face databases respectively.

4. Result analysis

The persons are reidentified using a face recognition system with the help of a percentage recognition rate (PRR). The recognition accuracy is enhanced as side-angled face images to be identified are transformed into frontal images using GAN. The final effective features are obtained by fusing MLTP and DWT using convolution.

Table 4. The PRR values using three techniques with variations in Hidden Layer neurons for the ORL database

No. of Neurons in the Hidden layer	PRR		
	DWT	MLTP	Proposed Fusion method
92	91.25	68.25	96.5
96	88	68.75	96.75
59	87.25	72	97
64	87.25	71.75	97
77	88.25	70.75	97
49	90	69.75	97.5
62	93.5	70.5	97.5
67	90.25	68.75	97.5
88	88.75	70.25	97.75
71	89	71.25	99

Table 5. The PRR values by varying neurons in Hidden Layer with Indian Male Face database.

No. of Neurons In the Hidden layer	PRR		
	DWT	MLTP	Proposed Fusion method
99	91.5	71	97.5
22	96	72.5	98
71	96.5	75.5	98
77	95.5	73	98
60	96	76	98.5
91	93.5	73	98.5
95	94	76.5	98.5
35	96.5	77.5	99
74	93.5	73.5	99
85	93.5	75	99

Table 6. The PRR variations with Hidden Layer neurons for the Indian Female Face database

No. of Neurons In the Hidden layer	PRR (%)		
	DWT	MLTP	Proposed Fusion method
95	92	77	97.5
35	96	75.5	98
60	95.5	76	98
79	95	71.5	98
97	94.5	75	98
19	91	69.5	98.5
31	93.5	76	98.5
36	93	73.5	98.5
85	92.5	71	98.5
74	93.5	75.5	99

Table 7. The PRR values with hidden layer neurons for CMU-PIE Face database

No. of Neurons In the Hidden layer	PRR (%)		
	DWT	MLTP	Proposed Fusion method
96	92.8	63.8	95.9
33	93.8	62.6	96
71	91.8	64.4	96
100	92.9	64.1	96
49	92.8	63.8	96.1
37	93	64.5	96.2
35	93.5	64.1	96.4
85	93	64.5	96.4
81	91.4	62.7	96.6
97	93.7	62.9	96.6

4.1 Result examination using ORL, IM, IF, CMU-PIE face database

The values of PRR are computed by varying the number of neurons in the hidden layer using DWT, MLTP, and the proposed fusion method is tabulated in Tables 4-7. The graphical representation of PRR with the hidden layer neurons is revealed in Figs. 17-20 for the four benchmarked face databases. The larger number of Hidden Layer neurons does not affect much on PRR values using all three methods. The proposed fusion method has higher PRR values compared to traditional frequency (DWT) and spatial domain (MLTP) techniques.

It is detected that the number of hidden neurons does affect the values of PRR with too few hidden neurons less than 5, as it does not have the capacity to learn enough of the fundamental patterns to discriminate. Hence the PRR values become minimum or zero. The PRR values start increasing for hidden layer neurons from 5 and attain maximum PRR with around 10 neurons for all the three techniques. The increased number of hidden neurons from 10 does not increase significantly PRR values

The proposed fusion technique has better PRR values compared to techniques based on DWT and MLTP with GAN. The proposed model performance using PRR values is related to the existing methods, and the PRR values are specified in Table 8. It is

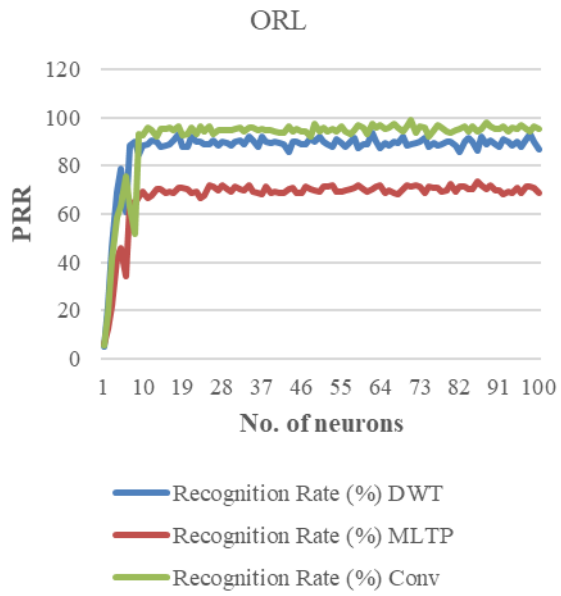


Figure. 17 PRR variations with the hidden layer neurons

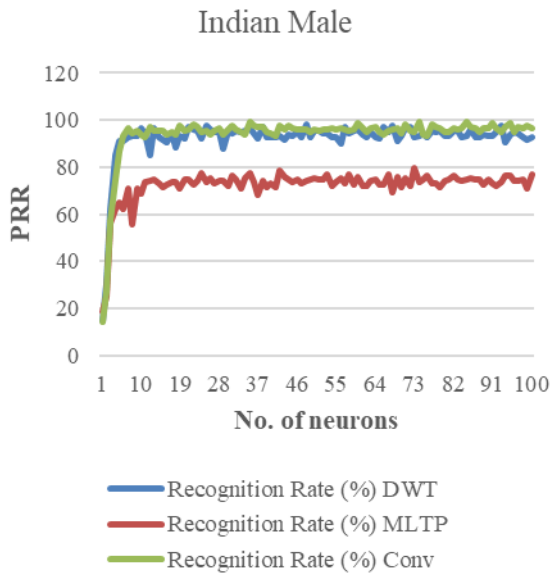


Figure. 18 PRR variations with the hidden layer neurons

exposed that; the projected method achieves better accuracy compared to the current approaches. In the existing methods presented by Wibowo et al., [30], Mohamed A. El-Sayed and M. A. Elsafty [31], and Palaniswamy and Suchitra [32] used techniques such as HOG, Hu Invariant moments, Gabor filter, histogram of HSV, and deep learning pose illumination invariant emotion recognition feature extraction. The classifiers SVM and CNN are used for the classification of features for person re-identification in the existing methods. The limitations of existing methods are (i) the database images are considered directly for person reidentification

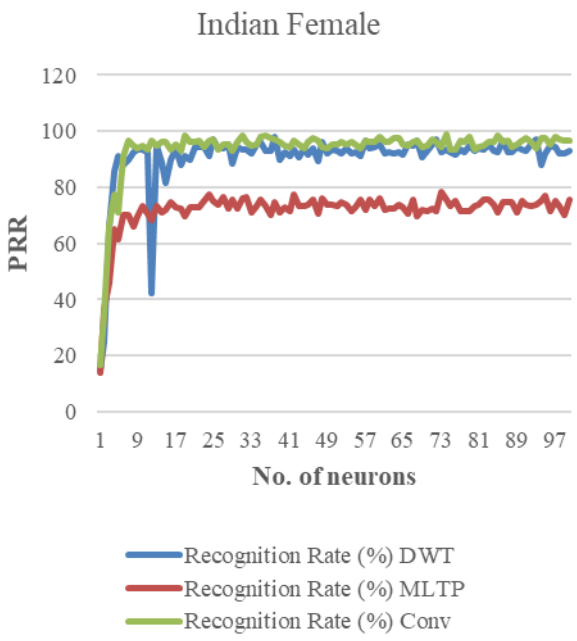


Figure. 19 PRR variations with number of neurons in the hidden layer

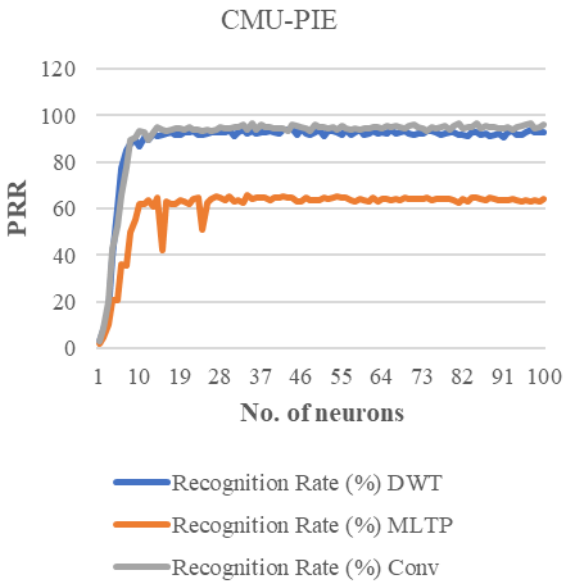


Figure. 20 PRR variations with the hidden layer neurons

without any modifications on images ie., images with frontal images, side-angled images, etc. (ii) the features are extracted directly from database images without any modifications on side-angled images. (iii) The values of PRR are less in the existing methods as the features of frontal and side-angled images are divergent. However, in the proposed method, all side-angled images are converted into frontal images using GAN resulting in similar features for different images of same person. Hence

Table 8. The proposed method comparison with current methods

Databases	Authors and Methods	Accuracy (%)
ORL	Wibowo et al., [30] HOG + Hu Invariant Moments + SVM	97.22
	Proposed Model	**99.00**
Indian Faces	Mohamed A. El-Sayed and M. A. Elsafty [31] Gabor + Histogram of HSV + Multi SVM	96.20
	Proposed Model	**99.00**
CMU Multi-PIE	Palaniswamy and Suchitra [32] DPIIER + CNN	96.55
	Proposed Model	**96.60**

the values of PRR are better in the proposed method compared to existing methods.

The enhancement in the projected method is defendable for the following reasons.

i. The GAN is used to convert oriented test images into frontal images for better identification of an individual
ii. The initial compressed noiseless features are extracted by the frequency domain technique DWT and spatial domain technique MLTP.
iii. The final effective features are obtained by the convolution of two initial features, and ANN is used for classification.

5. Conclusions:

Person Re-id is a challenging task as the captured images from different cameras have various angles and intensities. In this paper, the existing problem of captured images having different angles is solved by converting every image into a frontal image using GAN. The image features are obtained using DWT, MLTP, and convolution. The images are classified using ANN for Person re-identification effectively.

The proposed method is verified using PRR and compared with the existing methods to validate the superiority of the projected model. In the future, the technique will be used with deep learning methods for larger image databases and also with video sequences.

Conflicts of interest

"The authors declare no conflict of interest."

Author contributions

The First author has taken care about the Conceptualization, methodology, investigation, resources, data curation, writing-original draft preparation; The Second author has involved in the software, validation, review and editing, formal analysis, visualization and supervision. The Third author has involved in the review process.

Acknowledgments

We would like to thank anonymous authors for the guidance in the preparation of the article.

References

[1] I. J. Goodfellow, J. P. Abadie, M. Mehdi, B. Xu, David W. Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets", In: *Proc. of ACM International Conference on Neural Information Processing Systems*, pp. 2672–2680, 2014.

[2] K. Wang, H. Wang, M. Liu, X. Xing, and T. Han, "Survey on person re-identification based on deep learning", *CAAI Transactions on Intelligence Technology, IET Journal*, Vol. 3, Issue 4, pp. 219–227, 2018.

[3] W. Wei, W. Yang, E. Zuo, Y. Qian, and L. Wang, "Person re-identification based on deep learning - An overview", *Elsevier Journal of Visual Communication and Image Representation*, Vol 82, 2022.

[4] X. Wang and R. Zhao, "Person Re-identification: System Design and Evaluation Overview", In: Gong, S., Cristani, M., Yan, S., Loy, C. (eds) Person Re-Identification. *Advances in Computer Vision and Pattern Recognition*, Springer, London, 2014, doi: 10.1007/978-1-4471-6296-4_17.

[5] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking", In: *Proc. of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2007)*, pp. 356–368, 2007.

[6] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person Re-Identification by Symmetry-Driven Accumulation of Local Features", In: *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2360–2367, 2010.

[7] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning", In: *Proc. of*

IEEE Conference on Computer Vision and Pattern Recognition, pp. 2197–2206, 2015.

[8] S. Wu, Y. C. Chen, X. Li, A. C. Wu, J. J. You, and W. S. Zheng, "An enhanced deep feature representation for person re-identification", In: *Proc. of IEEE Winter Conference on Applications of Computer Vision*, pp. 743–758, 2016.

[9] Y. L. Wei and C. H. Lin, "Efficient weighted histogram features for single-shot person re-identification", *Journal of Signal Processing Systems*, Vol. 90, No. 4, pp. 477–491, 2018.

[10] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large Scale Metric Learning from Equivalence Constraints", In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2288–2295, 2012.

[11] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. H. Hoi, "Deep Learning for Person Re-Identification: A Survey and Outlook", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 6, pp. 2872-2893, 2022.

[12] X. L. Liao, C. Zhang, M. Dong, and X. Chen, "Person Reidentification by Deep Structured Prediction - A Fully Parameterized Approach", *IEEE MultiMedia*, Vol. 26, No. 3, pp. 42-55, 2019.

[13] L. An, M. Kafai, S. Yang and B. Bhanu, "Person Reidentification with Reference Descriptor", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 26, No. 4, pp. 776-787, 2016.

[14] S. Tan, F. Zheng, L. Liu, J. Han, and L. Shao, "Dense Invariant Feature-Based Support Vector Ranking for Cross-Camera Person Reidentification", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 28, No. 2, pp. 356-363, 2018.

[15] S. M. Li, C. Gao, J. G. Zhu, and C. W. Li, "Person Reidentification using Attribute-Restricted Projection Metric Learning", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 28, No. 8, pp. 1765-1776, 2018.

[16] Z. Liu, H. Lu, X. Ruan and M. H. Yang, "Person Reidentification by Joint Local Distance Metric and Feature Transformation", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 30, No. 10, pp. 2999-3009, 2019.

[17] Y. Yang, G. Wang, P. Tiwari, H. M. Pandey and Z. Lei, "Pixel and Feature Transfer Fusion for Unsupervised Cross-Dataset Person Reidentification", *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1-13, 2021.

[18] Y. Zhang, Y. Jin, J. Chen, S. Kan, Y. Cen, and Q. Cao, "PGAN: Part-Based Nondirect Coupling Embedded GAN for Person Reidentification", *IEEE MultiMedia*, Vol. 27, No. 3, pp. 23-33, 2020.

[19] W. Tanveer and M. H. Mahmood, "Unsupervised Person Re-Identification based on Feature Learning of Deep Convolutional Generative Adversarial Network", In: *Proc. of IEEE International Conference on INnovations in Intelligent Systems and Applications*, pp. 1-6, 2021.

[20] A. B. Haghighi and M. Taheri, "Person Re-Identification using Ensemble of Networks on Pose Transferred Images", In: *Proc. of IEEE International Conference on Pattern Recognition and Image Analysis*, pp. 1-5, 2021.

[21] AT&T Laboratories Cambridge (1994) "The ORL Database of Faces", Available: *http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html*.

[22] IIT Kanpur campus 2002, '*Indian Face Databse'*, Available*:* *http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/*.

[23] R. Gross, I. Matthews, J. Cohn, T. Kanade and S. Baker, "Multi-PIE", In: *Proc. of IEEE International Conference on Automatic Face & Gesture Recognition*, Vol. 28, Issue 5, 2010,

[24] G. Ian, P. A. Jean, M. Mehdi, X. Bing, W. F. David, O. Sherjil, C. Aaron, and B. Yoshua, "Generative Adversarial Nets", In: *Proc. of International Conference on Neural Information Processing Systems* (NIPS 2014), pp. 2672–2680, 2014.

[25] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, Issue 7, pp. 674–693, 1989.

[26] G. V. Sagar, S. Y. Barker, K. B. Raja, K. S. Babu, and K. R. Venugopal, "Convolution based Face Recognition using DWT and feature vector compression", In: *Proc. of IEEE International Conference on Image Information Processing*, pp. 444-449, 2015.

[27] P. Rangsee, K. B. Raja and K. R. Venugopal, "Modified Local Ternary Pattern Based Face Recognition using SVM", In: *Proc. of IEEE International Conference on Intelligent Informatics and Biomedical Sciences*, pp. 343-350, 2018.

[28] S. Haykin, "Neural Networks and Learning Machines", *Pearson, 3rd Edition*, 2008.

[29] C. Panjaitan, A. Silaban, M. Napitupulu, and J. W. Simatupang, "Comparison K-Nearest

Neighbors (K-NN) and Artificial Neural Network (ANN) in Real Time Entrants Recognition", In: *Proc. of IEEE International Seminar on Research of Information Technology and Intelligent Systems*, pp. 1-4, 2018.

[30] E. P. Wibowo, S. A. Harseno, and R. K. Harahap, "Feature Extraction Using Histogram of Oriented Gradient and Hu Invariant Moment for Face Recognition", In: *Proc. of IEEE International Conference on Informatics and Computing (ICIC)*, pp. 1-5, 2018.

[31] M. A. E. Sayed and M. A. Elsafty, "Metaheuristic Algorithms for Verification based on Biometric Features", *International Journal of Applied Engineering Research*, ISSN 0973-4562, Vol. 13, Issue 10, pp. 7564-7569, 2018.

[32] S. Palaniswamy and Suchitra, "A Robust Pose & Illumination Invariant Emotion Recognition from Facial Images using Deep Learning for Human-Machine Interface", *IEEE International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, pp. 1-6, 2019.