Computer Vision in Contactless Biometric Systems

Farukh Hashmi[1], Kiran Ashish[2], Satyarth Katiyar[3], and Avinash Keskar[4]
[1]Department of Electronics and Communication Engineering, National Institute of Technology, India
[2]Computer Vision Engineer, Viume, India
[3]Department of Electronics and Communication Engineering, Harcourt Butler Technical University, India
[4]Department of Electronics and Communication Engineering, Visvesvaraya National Institute of Technology, India

Abstract: *Contactless biometric systems have increased ever since the corona pandemic outbreak. The two main contactless biometric systems are facial recognition and gait patterns recognition. The authors in the previous work [11] have built hybrid architecture AccessNet. It involves combination of three systems: facial recognition, facial anti-spoofing, and gait recognition. This work involves deploying the hybrid architecture and deploying two individual systems such as facial recognition with facial anti-spoofing and gait recognition individually and comparing the individual results in real-time with the AccessNet hybrid architecture results. This work even involves in identifying the main crucial features from each system that are responsible for predicting a subject. It includes extracting few crucial parameters from gait recognition architecture, facial recognition and facial anti-spoof architectures by visualizing the hidden layers. Each individual method is trained and tested in real-time, which is deployed on both edge device NvidiaJetsonNano, and high-end GPU. A conclusion is also adapted in terms of commercial and research usage for each single method after analysing the real-time test results.*

Keywords: *AccessNet, gait patterns, facial recognition, contactless biometric systems, crucial features, NvidiaJetsonNano.*

1. Introduction

The field of behavior biometrics has inspired a re-examination of many previously overlooked human characteristics. One such characteristic that has been overlooked is the gait biometric. Gait biometrics refers to the individual aspects of human locomotion. Gait can also be defined as the walking posture of a person, involving a systematic movement trend and variations present at joints of the upper limbs and lower limbs during walking. Gait biometric can be exploited for recognizing a person according to the walking posture. Derawi *et al*. [5] suggested that there are 24 distinct components that can uniquely identify an individual's gait. Most of the biometrics presently used require that the subject be physically close before the authentication system can work such as iris scanner and fingerprint scanner. Gait biometric systems are also non-intrusive as they do not require the cooperation of the subject and they do not force the subject to behave in a certain way. The main problem with gait recognition is that there doesn't exists any device, which is able to capture the entirety of motions and attributes that form the human gait. The accuracy of gait recognition system also gets affected when human appearance changes greatly because of viewpoint, clothing and carrying objects.

Two primary research techniques have been used to perform recognition via computer vision: appearance based, and model based. The appearance technique is the silhouette-based technique and it involve obtaining a human silhouette by separating a moving person from a static background in each frame of a video. The changes in the shape of the human silhouette over time are analyzed [33]. In this methodology, the false silhouette may be obtained because of the clothing or the carried objects. The other technique for gait recognition is the model-based technique. Out of the two techniques, it is seldom used and involves fitting a mathematical model to human movement using the features extracted from gait sequences. A significant amount of computational power is required in this approach [39]. Soft biometrics defined by Jain *et al*. [17] involves a set of characteristics that provide some information for recognizing individuals but are not capable of distinguishing between individuals. Ross *et al*. [25] suggested that for large-scale deployment of biometric systems, one biometric is not enough. With the use of multi-biometrics systems, the resources required for computation are also reduced as the multi-biometrics systems help in narrowing down the space of potential identities of the subject in the recognition process. The gait biometric can be used to pre-select a group of possible candidates, which can further go through some other recognition system. Thus, several biometrics can be fused to make the overall process more robust. Face spoof attack is an attack in which a fraudulent user tries to deceive the facial recognition system by pretending to be a registered user and attempt to be authenticated as the genuine user.

The previous work done by the Hashmi *et al*. [12] is

an analysis of a combined accuracy of all three individual methods for giving an access to a test subject. This paper includes the extension of understanding real-time test results of each individual method and thereby comparing with the AccessNet [12] which is the previous work done by the authors. The extension includes, training with more subjects included in the training dataset for Gait recognition. An intense gait recognition testing is done by deploying camera in different angles and validating the model meter-wise.

This paper includes:

1. Related work explaining about other state-of-the-methods on all three methods (gait recognition, facial recognition and facial anti-spoofing).
2. Technical approach that involves architectural design flow for all three methods individually and AccessNet [12].
3. Experimental results and discussion explains about computational complexity during training and testing in real-time on CPU, GPU and edge device NvidiaJetsonNano. A comparison table is drawn comparing each individual accuracy with AccessNet.

2. Related Work

As stated, earlier gait recognition techniques using computer vision can be broadly classified into two categories: appearance-based techniques and model-based techniques. [1, 2, 3, 23, 39, 43] are model based technique. Zhao *et al.* [43] proposed a system in which video sequences captured by multiple cameras are used as input, and then a human 3D model was created. Lengths of key segments and the motion trajectories of lower limbs were extracted and used as static parameters and dynamic features respectively. Ariyanto *et al.* [1] also used temporal 3D data in their paper. They used articulated cylinders with 3D Degrees of Freedom at each joint to model the human lower legs. They were able to extract both gait structural and dynamics features using their approach. Luo *et al.* [23] proposed a self-occlusion optimized simultaneous sparse representation model to achieve 10 high robustness in limited gait frames. Yam *et al.* [39] combined primitive shapes to represent body structure. Bouchrika *et al.* [3] used elliptic Fourier descriptors to model the motion of human joints. However, it is difficult to accurately determine model parameters from low quality image sequences, which are captured by the cameras. Model based techniques are computationally expensive compared to appearance-based techniques. Several advancements have been made in the appearance-based techniques also [21, 24, 33]. Makihara *et al.* [24] extracted frequency-domain features of the volume by Fourier analysis and used a view transformation model to transfer gait features from one view to another. Liao *et al.* [21] proposed a pose-based temporal-spatial network to extract the temporal-spatial features, which gait recognition system less susceptible to clothing and carried objects. Wang, Liang *et al.* [33]

used Eigen space transformation based on principal component analysis to reduce the dimensionality of the input feature space. Hu *et al.* [16] proposed an incremental framework based on the optical flow that improved the usability of gait traits in surveillance videos. Some deep learning-based gait recognition systems have also been proposed to extract the gait feature [4, 6, 32, 36, 41]. Castro *et al.* [4] used Convolutional Neural Networks (CNN) to learn optical flow components for gait recognition. Feng *et al.* [6] worked with Long Short-Term Memory (LSTM) recurrent neural network in their paper, which can preserve the preserve temporal information in a gait sequence. Zhang *et al.* [41] proposed a Siamese neural network-based gait recognition framework to automatically extract discriminative gait features for human identification. Tran *et al.* [32] used homogeneous architecture with small 3x3x3 convolution kernels to capture features from both the spatial and temporal dimensions in a video. In [40], GAN was used to generate side-view gait images in without the carried object in normal clothing. Yuqi *et al.* [42] proposed a joint CNN based framework for both gait recognition and gait based soft biometrics such as age, gender, etc.

Many algorithms and techniques have been proposed to solve the problems related to face recognition. Deep Learning based approaches [18, 28, 29, 31, 35] have been very successful for face recognition. Sun *et al.* [28] used CNN for face recognition and the face identification task was used to increase the inter-personal variations while the face verification task reduces the intra-personal variations. Wen *et al.* [35] and Khan *et al.* [18] showed that deep CNN can be used to learn the age invariant deep face features for face recognition. Cheng *et al.* [41] proposed a two-layer CNN, which learned the high-level features and achieved an improvement in the discriminative power of the face recognition system. Face spoof attacks can easily fool the face recognition system [8] and therefore several authors have worked on anti-spoofing techniques. Fourati *et al.* [7] used image quality assessment and motion cues to distinguish between genuine and fake face-appearances to prevent face spoof attack on face recognition system. Xu *et al.* [38] proposed LSTM-CNN architecture to utilize temporal information for binary classification of a true and a fake face image. Liu *et al.* [22] used a CNN-RNN model to estimate the face depth with pixel-wise supervision.

3. Technical Approach

3.1. Gait Patterns Extraction

In this section, we will discuss about the crucial features, which are to be extracted from each pattern. Given different subjects in the CASIA-B dataset [44], which are binary silhouettes of walking persons,

features from each frame would be different from the previous frame as well as different from all angles. However, the values would lie close to each other for each frame and angles. These patterns are identified based on the classifier ensembles. This work is extension to the previous work done by the authors on Gait patterns in the field of Defence [9].

The gait patterns can be easily extracted from binary format images, using RGB images can dilute the weights and irrelevant features being captured for training. Hence, binary color mode is used to train gait patterns. The training dataset CASIA-B preserves information about the motion from each frame and outline shape, which further defines the knee angle, pressure, and footstep size. As the dataset is captured while subjects are in motion, noise can be found which can prune the accuracy levels while calculating crucial parameters such as knee angle, foot pressure, and concise step size. This sensitive information can only be obtained through the stable pieces and high-intensity pixels. After careful analysis, it can be concluded that most of the motion part is captured in the border areas. Hence, most of the motion related information can retrieved from these areas. The pre-processing of the dataset includes separating of the foreground from the background.

In order to eliminate the noise, each frame was binarized and morphological operations are used. All the binary images are resized into 224x224-image resolution. To avoid complex computation and unrelated Wang [34] used Principal Component Analysis (PCA) for dimensionality reduction. We are using Dropout [19] feature in our custom layers to take off the unrelated features, which in turn produced more accuracy.

One of the key things in identifying the differences between subjects is to compute difference between patterns of the lower silhouettes using Euclidean distance. The main disadvantage of just relying on this method is the view angle and the distance. If the subject is more than 7m away from the camera and the view angle is top, false negatives are be reported. Another major concern is of parameters extracted during the legs overlapped during the motion. Hence, considering multiple factors such as parameters from lower silhouettes and applying a powerful feature extraction on a whole image using ResNet152 [14] architecture as a base with ImageNet [26] pre-trained weights, given to our custom build architecture for training the images based on their unique ID (class name).

Representing the computing of silhouettes, we generally take the full-length gait period, denoted with 'N'. The parameters of the lower half of the silhouettes are known when both the legs are farther apart from each other. These are the frame when crucial parameters are extracted from each subject, i.e., the distance between pixels reaches a maximum when the legs are far away

from each other. Logically, the time interval between two consecutive maximums is each step. With given a total frame sequence as 'S' of binary silhouettes of the subjects, gait period length as 'N', we divide the total sequence into sub-sequences Si and the average of total sequence based on total sub-sequences as:

$$AS(i) = \frac{1}{N}\sum_{k=1}^{N} S^i(k) \tag{1}$$

Where 'k' in $S^i(k)$ is the image in the sequences 'i', where 'i' ranges from $i = 1, \dots\dots\dots\dots\dots\left[\frac{T}{N}\right]$, where 'T' is the total of images in a full-length sequence.

The extracted features are adopted with residual learning with stacked layers. The output linear equation is represented as:

$$y_{t(n)} = F\left(x_{t(n)}, W_i\right) + x_{t(n)} \tag{2}$$

With given input $x_{t(n)}$ and variable 't' denotes the subject and 'n' the sequence frame, W_i represents the filter-matrix. Output $y_{t(n)}$ is the dependent on variables input and the conv-filter matrix. The residual learning, which acts as a base, is further given to our custom build network which is tuned into a sequential model followed by custom layers which we will be discussing in the Experiments section. The biases are cut-off through LeakyReLu [37] activation function and Dropout function to slice the unwanted features, which may dilute the feature learning's. Overall proposed approach for gait patterns recognition is shown in Figure 1.

3.2. Facial Recognition

The second layer in the access authentication after gait pattern recognition is the facial recognition. We are using 512-dimensions facial points to train a subject. The main advantage of having 512D than 128D or 212D which computes faster than 512D is the accuracy and dataset size. Extracting 512 facial landmarks ensures not much data is needed from each subject. To avoid further complications, the facial recognition is not trained with any intense network since we have already extracted 512 facial landmarks from each face. Hence, KNN with Euclidean distance is plotted between each subject. Experiments were made for KNN using Euclidean distance and Cosine similarity and Euclidean distance gave the best accuracy.
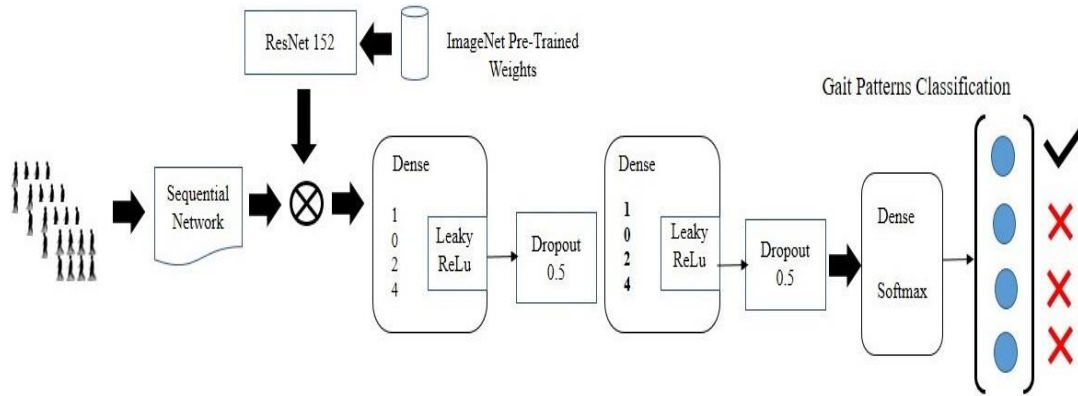
Figure 1. Technical approach of Gait Patterns recognition.

From Table 1, although the error rate is less for 128D, often-false positives are observed with increase in subjects. Hence, more facial dimensions are required to plot accurately.

Table 1. Comparison between facial dimensions and error rates computed with euclidean distance vs. cosine similarity.

Facial Dimensions	Cosine Similarity Error Rate	Euclidean Distance Error Rate
128D	1.9%	0.6%
212D	2.3%	1.3%
512D	6%	2%

3.3. Facial Anti-Spoofing Mechanism

The major drawback of facial recognition is the spoof attacks. The datasets used for training anti-spoof attacks was mostly custom collected through capturing real-faces through various mobile phones (both Android & iPhone), camera (720p & 1080p) with different lighting conditions and the same faces on photos, laptop screens, faces on YouTube videos (while in motion), faces on mobiles. This custom dataset plays a crucial role due to its variance of pixel intensities while compared to different environment setup. ROSE lab's Liveliness detection [20] was too used in addition to our custom collected dataset.

The technical approach for anti-spoofing is to just identify the continuous variations in the pixel values, which will be classified as spoof and a plain linear curve for real face. Since, in both the classes, faces will be present, as well as motion, the algorithm first understands the common features and patterns from both the patterns. This would be crucial part to identify what not to consider. Many state-of-the-art methods fail to do this. Our novel approach, which uses ResNet 101 [14] as base and custom designed layers concatenating with the residual learning, first discards the common features from both the classes. The major constituencies for this are: faces and motion. The main pattern to identify is the pixel intensities variations from each frame. The pixel levels have an irregular variation for spoofed faces, either shown in mobiles, photo frames, face masks or any other means. Proposed technical approach for facial anti-spoofing is shown in Figure 2.
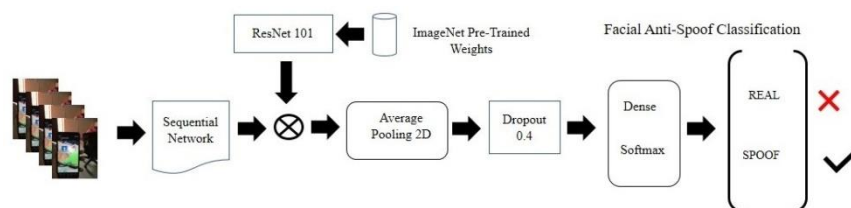

Figure 2. Technical approach of facial anti-spoofing.

4. Experimental Results and Discussion

4.1. Network Construction

This section describes the construction and experimental results of each individual method and at last comparison with the AccessNet [12], which is a combination of all three methods. Each method is evaluated in real-time.

4.1.1. Gait Pattern Recognition

For training to classify the subjects based on their gait patterns, we initially used CASIA-B dataset to understand the feature learning of the model. With CASIA-B containing different view angles of single subject and a full-length frame motion video, all possible parameters from multi-point view are extracted. The data is preprocessed with maintaining image size as 224x224 with color mode as binary. The dataset ratio is divided into 80-20 split for train-test split. ResNet 152 [14] architecture with ImageNet pre-trained weights is used as a base layer to avoid training from the scratch. The custom network is built with initializing Sequential mode, followed by ResNet 152 base model, where the last layer is frozen and Dense layer with 1024 filters with LeakyReLu activation function, output given to Dropout with value 0.5. Following the same layer pattern as previously, Dense with 1024, LeakyReLu activation function and output

given to Dropout function with 0.5. The last layer is the Dense with number of classes nodes and Softmax activation function. Adam optimizer is used with learning rate 0.0001. The training was done with 101-epoch size; with batch size as 128.The, overall trainable parameters were 4.6M.
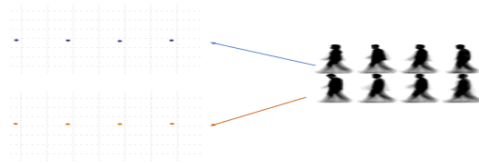

Figure 3. Extraction of walking motion and analyzing the uniformity stepwise.

Figure 3 describes a sample of one of the crucial gait patterns "uniformity of walking motion" which analyses the step-size motion of the subject. Other crucial gait patterns are knee-pressure, knee angle, hand movement speed etc., these gait patterns will determine the prediction of each subject. Each gait pattern is extracted from each subject and trained with respect to their class names.

Real-time testing is done on 50-trained subjects for gait recognition alone. The results are shown in Table 5. When the subject is at least 10 meters away from the camera with camera angle placed between 72^0 and 108^0, the accuracy is high when compared to other extreme right and left angles. As the distance decreases and increases beyond 10 meters, the accuracy declines because of the overlapping challenge of both the legs and hands with the body. The testing is done by two ways, deploying the trained model on edge-device NvidiaJetsonNano and another on high-end GPU server, with camera resolution of 1080p. The FPS on high-end server is greater with 180 FPS than edge-device FPS with 60 FPS.

Table 5. Accuracy with respect to probe angles of the camera and the distance between the camera and the subject.

Probe angle Distance from Camera	0^0	18^0	36^0	54^0	72^0	90^0	108^0	126^0	144^0	162^0	180^0
2m	0.18	0.26	0.31	0.35	0.39	0.41	0.38	0.32	0.29	0.25	0.10
5m	0.36	0.40	0.51	0.53	0.59	0.63	0.65	0.59	0.48	0.41	0.35
10m	0.54	0.57	0.61	0.64	0.78	0.91	0.86	0.74	0.70	0.63	0.33
15m	0.59	0.53	0.51	0.58	0.65	0.72	0.76	0.81	0.71	0.68	0.41
20m	0.29	0.33	0.34	0.40	0.46	0.49	0.47	0.43	0.40	0.34	0.29

4.1.2. Face Recognition and Anti-Spoofing

Facial recognition was trained on KNN with each subject having 4-7 images. 512D facial landmarks extracted from each image under the subject ID, is stored in a dictionary file. This is then given for Euclidean distance to compute the nearest distance between test cases. Spoof dataset with mostly custom collected dataset and dataset collected from various sources, motion videos are converted into image frames and are resized into

224x224. Color mode is RGB. ResNet 152 architecture with ImageNet pre-trained weights are used to train the model. With having ResNet 152 architecture as based model, 3 layers are used given to the last layers of ResNet architecture. Global Average Pool with 2D, input given as ResNet base model, following is the Dropout with 0.4. The final layer is Dense with number of classes as output nodes and Softmax activation function. The optimizer used as RMSProp with learning rate as 0.001. The training was done with 101 epochs and batch size as 128.

Facial anti-spoofing is validated on several architectures and tested in real-time with each trained model. The custom collected dataset is unique and different from ROSE lab and YouTube videos; hence, more features are available to learn while training.

Table 2. Accuracy table of evaluation of facial anti-spoofing on various architectures.

Architecture	Accuracy
VGG16 [27]	0.78
VGG19 [27]	0.81
ResNet50 [14]	0.85
ResNet101 [14]	0.93
ResNet152 [14]	0.93
InceptionV3 [30]	0.87
MobileNet [15]	0.86

Table 2 describes the real-time testing of facial anti-spoofing model trained on various architectures. ResNet101 and ResNet152 surpassed the rest of the architectures in real-time. Faces were extracted from each face available in each image from the training dataset, and the cropped faces are stored from each class, i.e., real and spoof. These cropped faces are trained and the similar manner of extracting face and classifying into real and spoof is followed in the real-time testing. The real-time prediction in testing for each face was ~10ms using GPU and ~55ms using CPU.

The main features of classifying the classes into real and spoof are the pixel patches. Analyzing the hidden layers, pixels from replay attacks are 50% higher in density than any normal real face in real-time. Pixel patches from masks and print like photos are 10% more than any real face in real-time. The main reason for high pixel intensity is the lightening effect on the faces.

4.1.3. AccessNet [11]

AccessNet [12] comprises combination of all three methods. The access authentication is given only if the overall accuracy for each subject is more or equal to 0.9 (90%). With having subject 7m away from the camera, motion feed is fed to the gait network and data preprocessing is done by converting them into binary silhouettes. In parallel, when the subject is 3m away from the camera, facial recognition is initiated alongside with anti-spoofing network, where the input for both is common. Once the face is revealed and anti-spoof network classifies the subject, all the accuracies

are together summed up for getting an average value. The threshold is initially set to 0.9. Technical architecture of AccessNet is shown in Figure 4.

The authors have trained each method on 50 individual and tested in real-time with all three individual and with AccessNet [12]. The comparison table is drawn in Table 3.
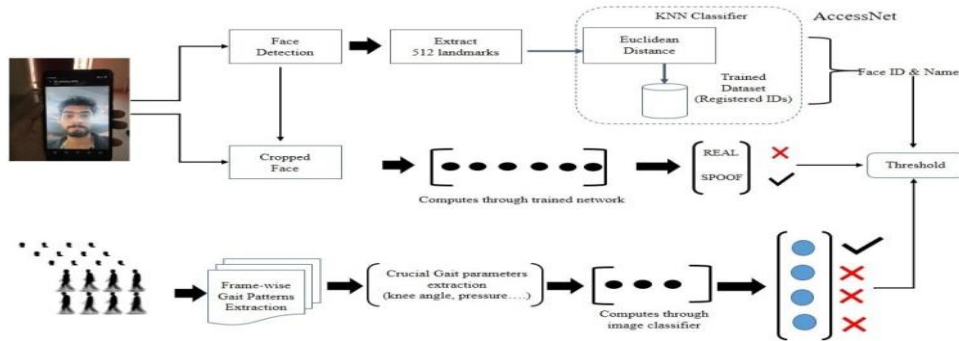

Figure 4. AccessNet architecture [11].

Table 3. Accuracy comparison from real-time testing on 50 individual subjects.

Method	Accuracy
Gait Patterns Recognition	87%
Facial Recognition	94%
Facial Anti-Spoofing	97%
AccessNet [12]	90%

4.2. Performance Metrics

All the datasets and networks are trained with Nvidia GPU 1080x. The training patch accuracy is noted at 90.55% for Gait patterns recognition network and 92.33% for facial anti-spoofing network. The training performance graphs are plotted in Figure 5.
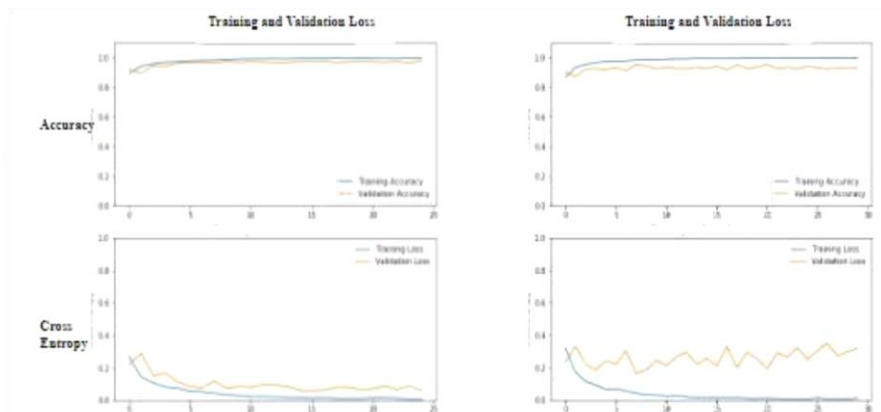

Figure 5. Training Performance of gait and facial anti-spoofing networks.

The performance of the architecture is measured by F1 score [10, 11], which is calculating the success of precision and recall rates. Table 4 describes the average F1 score of the test samples taken 50 test samples.

The training time for facial recognition on 50, individuals are approximately 5 minutes using GPU, facial anti-spoofing was approximately 4 hours, and gait patterns on 50 individuals was 6 hours. Hence, the computational complexity of gait patterns is more than the rest two methods.

Table 4. Average F1 score for 50 test cases using AccessNet architecture.

Network	Precision	Recall	F1 Score
Gait Patterns	0.91	0.92	0.91
Facial Recognition	0.98	0.97	0.97
Facial Anti-Spoofing	0.95	0.94	0.94

5. Conclusions and Future Work

From Table 4 accuracy figures, we can say that

implementation of facial recognition with facial anti-spoofing algorithms is best suited for all corporate offices, shopping complexes, public and private institutions. It is light weight and can be deployed in any standalone edge device. The computation power drawn from both the architectures are 4GB and inference speed in edge device like NvidiaJetsonNano is ~30 FPS. Hence, in real-time deployment, classes can be increased and decreased as well for facial recognition and train on new classes. Facial anti-spoofing is constant and re-training with new dataset is not necessary as it deals with pixels of face not identifying class name of a face.

Gait patterns is much complex and is mostly suitable for implementing in all military complexes for advanced surveillance systems. Training of new classes consumes more time and hence, is not suitable for any commercial deployment.

AccessNet network can't be used in real-time if the

subjects from training dataset are different from the testing dataset. Each time numbers of subjects are added, their initial motion, face photos are required to train the model. The model training takes less than an hour for having a smaller number of subjects, however, increases with having more subjects and data. The network works accurately if the camera is kept at the top side angles, and accuracy may decrease a bit if kept in straight angles since the backward legs motion couldn't be captured if kept in top straight angle. The results prove that we can get highest accuracy levels from facial recognition and anti-spoofing. However, gait patterns might give accuracy less than 90% due to various parameters, therefore summing all the three outputs for each subject, can ensure if the subject is registered and authenticated. After several experiments of subjects, threshold is fixed at 0.9 or 90%.

The future extension of this paper is in analyzing emotions from users in commercial complexes, especially in shopping malls for scene understanding between the customer's emotion towards each product and each offer on the product. The authors have implemented a part of it in [13].

References

[1] Ariyanto G. and Nixon M., "Model-Based 3D Gait Biometrics," *in Proceedings of IEEE International Joint Conference on Biometrics*, Washington, pp. 1-7, 2011.

[2] Bodor R., Drenner A., Fehr D., Masoud O., and Papanikolopoulos N., "View-Independent Human Motion Classification Using Image-Based Reconstruction," *Image and Vision Computing*, vol. 27, no. 8, pp.1194-1206, 2009.

[3] Bouchrika I. and Nixon M., "Model-Based Feature Extraction for Gait Analysis and Recognition," *in Proceedings of International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications*, Berlin, pp. 150-160, 2007.

[4] Castro F., Marín-Jiménez M., Guil N., and De La Blanca N., "Automatic Learning of Gait Signatures for People Identification," *in Proceedings of International Work-Conference on Artificial Neural Networks*, Cham, pp. 257-270, 2017.

[5] Derawi M., Bours P., and Holien K., "Improved Cycle Detection for Accelerometer Based Gait Authentication," *in Proceedings of IEEE 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Darmstadt, pp. 312-317, 2010.

[6] Feng Y., Li Y., and Luo J., "Learning Effective Gait Features Using LSTM," *in Proceedings of IEEE 23rd International Conference on Pattern Recognition*, Cancun,pp. 325-330, 2016.

[7] Fourati E., Elloumi W., and Chetouani A., "Anti-Spoofing in Face Recognition-Based Biometric Authentication Using Image Quality Assessment," *Multimedia Tools and Applications*, vol. 79, no. 1, pp. 865-889, 2020.

[8] Ghouzali S., and Marie-Sainte S., "Face Identification Based Bio-Inspired Algorithms," *The International Arab Journal of Information Technology*, vol. 17, no. 1, pp. 118-127, 2020.

[9] Hashmi M., Farukh B., Ashish K., and Keskar A., "GAIT Analysis: 3D Pose Estimation And Prediction in Defence Applications Using Pattern Recognition," *in Proceedings of the 12th International Conference on Machine Vision*, Shenzhen, 2020.

[10] Hashmi M., Ashish B., Keskar A., Bokde N., and Geem Z., "FashionFit: Analysis of Mapping 3D Pose and Neural Body Fit for Custom Virtual Try-On," *IEEE Access*, vol. 8, pp. 91603-91615, 2020.

[11] Hashmi M., Ashish B., Keskar A., Bokde N., Yoon J., and Geem Z., "An Exploratory Analysis On Visual Counterfeits Using Conv-Lstm Hybrid Architecture," *IEEE Access*, vol. 8, pp. 101293-101308, 2020.

[12] Hashmi M., Ashish K., Katiyar S., and Keskar A., "AccessNet: A Three Layered Visual Based Access Authentication System for Restricted Zones," *in Proceedings of the IEEE 21st International Arab Conference on Information Technology*, 6th of October city, pp. 1-7, 2020.

[13] Hashmi M., Ashish B., Sharma V., KeskarG., Bokde N., Yoon J., and Geem Z., "LARNet: Real-Time Detection of Facial Micro Expression Using Lossless Attention Residual Network," *Sensors*, vol. 21, no. 4, pp. 1098, 2021.

[14] He K., ZhangX., RenS., and Sun J., "Deep Residual Learning for Image Recognition," *in Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, pp. 770-778, 2016.

[15] Howard A., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., and Adam H., "Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:*1704.04861, 2017.

[16] Hu M., Wang Y., Zhang Z., Zhang D., and Little J., "Incremental Learning for Video-Based Gait Recognition with LBP Flow," *IEEE Transactions on Cybernetics*, vol. 43, no. 1, pp. 77-89, 2012.

[17] Jain A., Dass S., and Nandakumar K., "Can Soft Biometric Traits Assist User Recognition," *in Proceedings of The Biometric Technology for Human Identification*, Dhanbad, pp. 561-572, 2004.

[18] Khan S., Javed M., Ahmed E., Shah S., and Ali S., "Facial Recognition Using Convolutional Neural Networks and Implementation on Smart Glasses," *in Proceedings of the IEEE*

International Conference on Information Science and Communication Technology, Karachi, pp. 1-6, 2019.

[19] Lengerich B., Xing E., and Caruana R., "On Dropout, Overfitting, and Interaction Effects in Deep Neural Networks," *arXiv preprint arXiv:*2007.00823, 2020.

[20] Li H., Li W., Cao H., Wang S., Huang F., and Kot A., "Unsupervised Domain Adaptation For Face Anti-Spoofing," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 7, pp. 1794-1809, 2018.

[21] Liao R., Cao C., Garcia E., Yu S., and Huang Y., "Pose-Based Temporal-Spatial Network (PTSN) for Gait Recognition with Carrying and Clothing Variations," *in Proceedings of Chinese Conference on Biometric Recognition*, pp. 474-483, Cham, 2017.

[22] Liu Y., Jourabloo A., and Liu X., "Learning Deep Models For Face Anti-Spoofing: Binary or Auxiliary Supervision," *in Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake, pp. 389-398. 2018.

[23] Luo J., Tang J., Tjahjadi T., and Xiao X., "Robust Arbitrary View Gait Recognition Based on Parametric 3D Human Body Reconstruction and Virtual Posture Synthesis," *Pattern Recognition*, vol. 60, pp. 361-377, 2016.

[24] Makihara Y., Sagawa R., Mukaigawa Y., Echigo T., and Yagi Y., "Gait Recognition Using A View Transformation Model in The Frequency Domain," *in Proceedings of European Conference on Computer Vision*, Berlin, pp. 151-163, 2006.

[25] Ross A., Nandakumar K., and Jain A., *Handbook of Multibiometrics*, Science and Business Media, 2006.

[26] Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., and Fei-Fei L., "ImagenetLarge Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211-252, 2015.

[27] Simonyan K. and Zisserman A., "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv:* 1409.1556, 2014.

[28] Sun Y., Wang X., and Tang X., "Deep Learning Face Representation by Joint Identification-Verification," *in Proceedings of Advances in Neural Information Processing Systems*, Montreal, pp. 1988-1996, 2014.

[29] Sun Y., Wang X., and Tang X., "Hybrid Deep Learning for Face Verification," *in Proceedings of The IEEE International Conference on Computer Vision*, Sydney, pp. 1489-1496. 2013.

[30] Szegedy C., Vanhoucke V., Ioffe S., Shlens J., and Wojna Z., "Rethinking The Inception Architecture for Computer Vision," *in The Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, pp. 2818-2826, 2016.

[31] Taigman Y., Yang M., Ranzato M., and Wolf L., "Deepface: Closing the Gap to Human-Level Performance in Face Verification," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, pp. 1701-1708, 2014.

[32] Tran D., Bourdev L., Fergus R., Torresani L., and Paluri M., "Learning Spatiotemporal Features with 3d Convolutional Networks, "*in Proceedings of The IEEE International Conference on Computer Vision*, Santiago, pp. 4489-4497, 2015.

[33] Wang L., Tan T., Ning H., and Hu W., "Silhouette Analysis-Based Gait Recognition for Human Identification, "*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp.1505-1518, 2003.

[34] Wang Q., "Kernel Principal Component Analysis and Its Applications in Face Recognition and Active Shape Models," *arXiv preprint arXiv*: 1207.3538, 2012.

[35] Wen Y., Li Z., and Qiao Y., "Latent Factor Guided Convolutional Neural Networks for Age-Invariant Face Recognition," *in Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4893-4901. 2016.

[36] Wu Z., Huang Y., and Wang L., "Learning Representative Deep Features for Image Set Analysis," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp.1960-1968, 2015.

[37] Xu B., Wang N., Chen T., and Li M., "Empirical Evaluation of Rectified Activations in Convolutional Network," *arXiv preprint arXiv*: 1505.00853, 2015.

[38] Xu Z., Li S., and Deng W., "Learning Temporal Features Using LSTM-CNN Architecture for Face Anti-Spoofing," *in Proceedings of the IEEE 3rd IAPR Asian Conference on Pattern Recognition*, Kuala Lumpur, pp. 141-145, 2015.

[39] Yam C. and Nixon M., "Model-based Gait Recognition in: Encyclopedia of Biometrics," *Springer*, pp.1082-1088, 2009.

[40] Yu S., Chen H., Garcia Reyes E., and Poh N., "Gaitgan: Invariant Gait Feature Extraction Using Generative Adversarial Networks," *in Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, pp. 30-37, 2017.

[41] Zhang C., Liu W., Ma H., and Fu H., "Siamese Neural Network Based Gait Recognition for Human Identification," *in the Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Shanghai, pp. 2832-2836, 2016.

[42] Zhang Y., Huang Y., Wang L., and Yu S., "A Comprehensive Study on Gait Biometrics Using A Joint CNN-Based Method," *Pattern*

Recognition, vol. 93, pp. 228-236, 2019.

[43] Zhao G., Liu G., Li H., and Pietikainen M., "3D Gait Recognition Using Multiple Cameras, "*in the Proceedings of IEEE 7th International Conference on Automatic Face and Gesture Recognition*, Southampton, pp. 529-534, 2006.

[44] Zheng S., Zhang J., Huang K., He R., and Tan T., "Robust View Transformation Model for Gait Recognition," *in Proceeding of 18th IEEE International Conference on Image Processing*, Brussels, pp. 2073-2076, 2011.

Farukh Hashmi received his B.E in Electronics and Communication Engineering from R.G.P.V Bhopal University in 2007. He obtained his M.E. in Digital Techniques & Instrumentation in 2010 from R.G.P.V Bhopal University. He received Ph.D. at VNIT Nagpur under the supervision of Dr. AvinashKeskar. He has a teaching experience of 11 years. He is currently an Assistant professor at Department of Electronics and Communication Engineering, National Institute of Technology, Warangal. He has published up to 65 papers in National/International Conferences/ Journals. His current research interests are Image Processing, Internet of Things, Embedded Systems, Biomedical Signal Processing, Computer Vision, Circuit Design, and Digital IC Design etc. Mr. Mohammad F. Hashmi is a member of IEEE, LMIETE, and LMIAENG.

Kiran Ashish is currently working as Computer Vision Engineer at Viume, an AI based startup at Barcelona, Spain. He completed his Bachelor's degree in ECE from Anurag Group of Institutions, Hyderabad in 2019. He is having 2 years of industry experience in the computer vision domain and has worked in several verticals. He has published six papers in National/International Conferences/Journals. His current field of interests include Deep Learning, Neural Networks and Computer Vision, Biomedical Imaging, Facial; recognition etc.,.

Satyarth Katiyar is studying Bachelors in Technology in Electronics Engineering, Harcourt Butler Technical University (HBTU), Kanpur (U.P.) India from 2017. He will complete B.Tech. Degree in the June 2021. He is working as a research Intern under the supervision of Dr. Md. Farukh Hashmi. He has published three papers in National/International Conferences/Journals. His current field of interests include Deep Learning, Neural Networks and Computer Vision, Biomedical Imaging, Facial; recognition etc..

Avinash Keskar completed his B.E. from VNIT, Nagpur in 1979 and received gold medal for the same. He completed his M.E. from IISc, Bangalore in1983, receiving the gold medal again. He also completed his Ph.D. from VNIT Nagpur in1997.The author is a life member of IAENG. He has 32 years of teaching experience and 7 years of industrial experience. He is currently a Professor and Head of the Department at Department of Electronics Engineering, VNIT Nagpur. His current research interests include Image Processing, Computer Vision, Soft Computing, and Fuzzy Logic etc. Prof. Keskar is a senior member of IEEE, FIETE, LMISTE, FIE.