# S.H.A.D.E: System for Human Authentication in Dynamic Environment

Muskan Pandey
MIT Academy of Engineering
Pune, India

Rajkumar Yadav
MIT Academy of Engineering
Pune, India

Aditi Pandey
MIT Academy of Engineering
Pune, India

***Abstract*— Security and surveillance is becoming an emerging field. With many advances happening and different systems deployed, we propose a system that combines facial as well as gait recognition. The current systems deployed either require human-computer interaction or are not dynamic enough. Our system is completely dynamic and extracts maximum features for dynamic human recognition. Facial recognition is one of the most widely used methods for the identification of humans. A lot of work is being done in this field. Our system uses few-shot learning and has cross-age recognition. To add to the accuracy, we combine gait recognition, thus human recognition. Gait is biometry that differentiates individuals by the way they walk. Research on this topic has gained evidence since it is unobtrusive and can be collected at distance, which is desirable in surveillance scenarios. Most of the systems for gait recognition are silhouette based which are not robust and have many intrinsic limitations. Our system combines Subsequence Dynamic Time Warping to compare signals from probe and gallery, ranking based on minimum matching distance cost, and secondly computes the Euclidean distance of the movement histograms to define the person from the gallery that is closest to the probe. Finally, combines the score fusion to provide the identity of a subject. A combination of these two models leads to a robust, error-free system that recognizes humans in real-time scenarios.**

***Keywords*— Security & surveillance, dynamic human recognition, facial recognition, gait recognition, dynamic time warping, cross-age recognition***

## I. INTRODUCTION

Over the past few decades, the security of a system is seen as of ultimate priority because of which authentication programs in a system became an issue of high concern. Authentication technology provides access control to a system by checking the user's credentials with the already existing database and accordingly giving authority to the user. Human authentication system has grown its roots into various sections of the human recognition system. Few of them include fingerprint, iris, signature, retinal, vein, voice recognition, etc. In most of these systems, there is a need for extra attention from the user end, reducing the dynamicity of the system. In the wake of digitalization, the dynamicity of a system is of the utmost importance. The speed of a system is as important as the efficiency, keeping both in mind, face and gait recognition seems to be the best fit for this scenario, as, there is no engagement with the system from the user's side.

In this paper, we address the problem dynamic human recognition-based authentication for a given entrance gate a hardware-software solution to identify every unique person who enters or exit the gate, with a log of all previous entry/exit time, photo/videos recorded. That means there will not be a previous history of an individual in the first entry. The system should immediately alert the security if it detects a new person and the security will decide to allow/restrict that person entering the premises. Whereas, the system would learn from its previous history of videos/images dynamically and will allow a known person. For a given size of the gate, the number of cameras with the optimal resolution will be worked out as part of the system. The solution would be scalable.

## II. PREVIOUS WORK

Face recognition system is a way of recognizing humans with the help of their faces using various techniques. The technological flow of its market has a beginning from 2014, the Gaussian Face algorithm, developed by the Chinese University of Hong Kong's researchers. In facial recognition, they achieved a score of 98.52% which was then compared with the score achieved by humans, which was 97.53%. This efficiency was achieved despite them having the vulnerability of memory capacity and calculation times. Further in 2014, Facebook launched its Deep Face program that determined whether two faces in photographs belong to the same person. An excellent score of 97.25% was achieved which was just 0.28% below when tested with humans. In June 2015, Google shaped this idea into the algorithm as Face Net. Face Net used Labelled Faces in the Wild (LFW) dataset, this algorithm set the new record of 99.63% (0.9963 ± 0.0009). A company from Mountain view an integrated artificial neural network with its algorithm achieving almost perfect results. In May 2018, Amazon started promoting its cloud-based face recognition service named Recognition to law enforcement agencies.

Gait is defined as "a manner of walking" from the Webster Collegiate Dictionary. Hence human recognition based on the information like the coordinates of various body parts, pose estimation, etc. is known as gait recognition. There are various implementation techniques in the market for gait recognition. The most inaction technique includes tracking the pose of a person, with the help of the position of the body parts with the help of point features. These statistical, parametric and Hidden Markov Model are fitted to the tracking data and probability tests are used for recognition. In mixed state statistical models for motion representation with the help of particle filters for estimation and recognition. In the linear Gaussian model,

recognition is achieved by designating a metric on the space of models. Zelnik-Manor et al [9], proposed a technique to determine the distance between video sequences with the support of the Spatio-temporal gradient at multiple temporal scales. Some techniques for recognition were on the basis of the periodic movement of the identified point features.

## III. FACE

Face recognition is one of trending research work going on in the field of computer vision and deep learning. One of the many applications of face recognition systems lies in domain related to security and surveillance systems. The problem is divided into two parts: face detection and face recognition. Face detection involves searching out of regions in the image containing facial features. Face recognition consists of many feature variants to look after like age, pose, color, emotion, and illumination. Some of the methods discovered till date are:

### *A. Linear appearance-based methods:*

It involves mechanisms like principal component analysis (PCA), independent component analysis (ICA), linear regression classifier (LRC), etc. which fail inadequately due to diversity in expression, illumination, and other factors. Hence a nonlinear approach was proposed.

### *B. Nonlinear extensions*:

These methods present with kernel PCA, kernel LDA or locally linear embedding (LLE) and many other nonlinear methods like these use kernel-based methods where the general idea is to map facial images to higher-dimensional space. Nonlinearity helps deal with complexity better than a linear approach to some extent but has made no severe impact.

### *C. Face descriptor-based methods:*

This method involves generating local feature-based descriptors that provide a global descriptor feature. Firstly, after face detection, each of the local descriptors is evaluated then aggregated to give global descriptor. It has unavoidable computational expenses. This method is advantageous with varying illuminance and expression changes.

### *D. Gabor wavelets:*

The Gabor wavelets provide salient visual properties and spatial frequency. These wavelets have significantly high dimensionality of Gabor features. It is impractical for real-time applications.

### *E. 3D-based face recognition:*

Face recognition system using this method are usually most expensive because of the collection of a huge set of 3d points of the human face which helps tackle various pose variations.3d points gives depth information of each face view. This can achieve better robustness than 2d based ones.

### *F. Neural network-based recognition system:*

With the increasing availability of digitized data and computational power along with the development of deep learning algorithms, today face recognition has reached human-level accuracy with little constraints. In traditional methods feature vector corresponding to each of the variants discussed above was discovered separately, whereas in deep learning approach a single feature vector tackles all the variants present in the dataset used in building face recognition model.

In real-time surveillance systems using CCTV cameras, we have to deal with the varying resolution of facial region inputs which adds to requirements of another set of feature space which helps the recognition model to extract resolution invariant features. Research work is done [15] for getting resolution invariant feature involved building a ResNet with two structural parts with first part trained for normal recognition with use of triplet loss and second part consisting of two branch networks trained with high resolution and corresponding low-resolution images.

Many architectures were worked with to improve upon the model's performance like Alex Net, VGGNet, GoogleNet, etc. most of these were introduced as the baseline model for face recognition. In face-recognition, there are many scenarios where intra-class variation is more than inter-class variation hence it becomes very necessary to pick the best loss function. The survey [10] mentioned presents different loss methods like Euclidean-distance-based loss, angular/cosine-margin-based loss, SoftMax loss, etc.

Taking into consideration the problems related to face recognition as mentioned above we designed a ResNet model with subdivisions in-network at the further end which is trained with triplet loss to resolve resolution issues as discussed earlier. The CASIA A face database is used for this model that has images of each individual with significant age gaps, pose and all other variants. While training each set of input is used to generate multiple resolution frames, which are further paired for input to the model trained using triplet loss. The recognition works on cross-age samples with few-shot training.

## IV. GAIT

Gait is the manner in which a person walk.

### *A) Extracting Features*

Pose estimator proposed by Cao et al. [2] is proposed to extract the body coordinates. The pose estimator returns coordinates of 18 body parts: neck, nose and both ears, wrists, elbows, hips, knees, ankles, shoulders and eyes. For each frame *f* on gait sequence *i* and body part indexed *b*, $1 \leq f \leq F^i$, where $F^i$ is the total number of frames of gait sequence *i*. $P^i_{b,f}$ is the (x,y) coordinate for each *b* in $i^{th}$ gait sequence. The coordinates are whole numbers except when there is an invalid coordinate due to occlusion or if the part is not detected, then the coordinates are (-1, -1). The invalid coordinates are noise which might disrupt the results; therefore, a tracking of invalid parts is performed. A threshold is defined and noise index $N^i$ is maintained for each $i^{th}$ sequence. If $N^i$ crosses the threshold, it is used for the classification of “noisy” body parts on the Subsequence Dynamic Time Warping and the Euclidean Distance calculation.

$$N^{i} = \{ b: \frac{\#\{ P^{i}_{b,f}.x = -1 \ \forall f \in [1,2,\ldots,F^{i}]\}}{F^{i}} > \gamma \},$$

$\boldsymbol{\gamma}$ is a parameter representing the noise threshold value.

### B) Signals from features

To get the information dynamically from every gait sequence i, signal S is generated. Ears, eyes, and nose are not considered because their positions do not help in gait recognition. Therefore, from 2 to 13, each b will generate Si which will have two values: x and y coordinate. Si will have 12×2 lines and Fi columns.

$$S^{i}_{2b-3,f} = \left\{ -1, \frac{P^{i}_{b,f}.x - P^{i}_{1,f}.x}{max_j \, P^{i}_{j,f}.y - P^{i}_{1,f}.y} \right\}, \qquad (2)$$

If $P^{i}_{1,f}.x$ = -1, otherwise

$$S^{i}_{2b-3,f} = \left\{ -1, \frac{P^{i}_{b,f}.y - P^{i}_{1,f}.y}{max_j \, P^{i}_{j,f}.y - P^{i}_{1,f}.y} \right\}, \qquad (3)$$

If $P^{i}_{1,f}.y$ = -1, otherwise

## V. Gait Recognition

Fusion of two methods is used for recognition. A cost function is defined, and the person whose value minimizes the value of this function is classified as the result.

### A) Subsequence Dynamic Time Warping

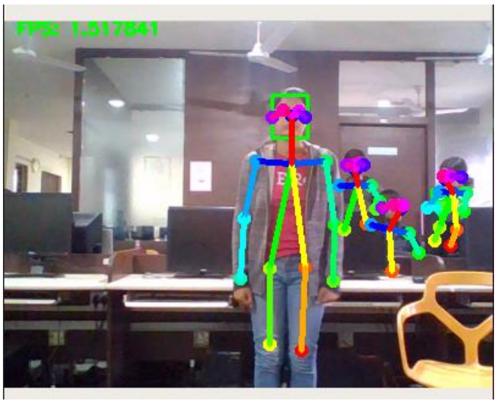
Dynamic Time Warping is a non-linear dynamic time normalization technique. Warping means bending or adjusting as a result of some external factors. A warping function is used to map the time axis from samples in the gallery. This technique has two major advantages: it is robust to walking speed and the signals which are compared don't have the same number of samples. Gait sequences that have low minimum matching cost are likely to be of the same person.

### B) Euclidian Distance

Euclidian distance is also applied to compare the values of the cost function. This distance is used to rank the individuals according to the distance from the sample.

### C) Score Fusion

Fusion of features is done to improve recognition accuracy. $\alpha$ is used to weight the fusion of scores. Apt value of $\alpha$ is chosen based on observations of trial and error methodology.



## VI. RESULTS

The experiment will be conducted on real-time videos obtained. The dataset contains only frontal data with occluded samples also. It is from a residential complex where the system is designed to be deployed. The use of Euclidean distance improves the accuracy from 92.5% to 98.75%.

The system will first detect and recognize face and then have a verification with gait recognition results. This overcomes the shortcomings of both biometrics separately and hence improves the robustness and accuracy. It also eliminates spoofing possibilities.

## ACKNOWLEDGMENT

We would like to acknowledge Dr. Avinash Bhute who guided us and mentored us in this project. Sahil Tah, Kaustubh Farfad, Om Avhad, Arti Bhalodia, Vaibhav Chaudhari, Aditya Veer who helped us in surveying and collecting data for the experiment.

## REFERENCES

[1] de Lima V.C., Schwartz W.R. (2019) Gait Recognition Using Pose Estimation and Signal Processing. In: Nyström I., Hernández Heredia Y., Milián Núñez V. (eds) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2019. Lecture Notes in Computer Science, vol 11896. Springer, Cham

[2] Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: Realtime multi-person 2d pose estimation using part affinity fields. In: CVPR. vol. 1, p. 7 (2017).

[3] Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. vol. 4, pp. 441–444. IEEE (2006).

[4] G Johansson. Visual motion perception. Scientific American, (232):76–88, 1975.

[5] T.Starner and A. Pentland. Real-time American sign language recognition from video using hmm. In Proc. of ISCV 95, volume 29, pages 213–244, 1997.

[6] M. E. Brand and A. Hertzmann. Style Machines. ACM SIGGRAPH, pps 183-192, July 2000

[7] C. Bregler. Learning and Recognizing Human Dynamics in Video Sequences. In Proc of CVPR, pp. 568-574, 1997

[8] R. Cutler and L. Davis Robust real-time periodic motion detection, analysis, and applications. In IEEE Trans. PAMI, 22(8), August 2000.

[9] L. Zelnik-Manor and M. Irani Event-based video analysis. In Proc of CVPR, 2001.

[10] I. Masi, Y. Wu, T. Hassner and P. Natarajan, "Deep Face Recognition: A Survey," 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Parana, 2018, pp. 471-478.

[11] Saez Trigueros, Daniel & Meng, Li & Hartnett, Margaret. (2018). Face Recognition: From Traditional to Deep Learning Methods.

[12] Zhao, Jian & Cheng, Yu & Cheng, Yi & yang, yang & Lan, Haochong & Zhao, Fang & Xiong, Lin & Xu, Yan & Li, Jianshu & Pranata, Sugiri & Shen, Shengmei & Xing, Junliang & Liu, Hengzhu & Yan, Shuicheng & Feng, Jiashi. (2018). Look Across Elapse: Disentangled Representation Learning and Photorealistic Cross-Age Face Synthesis for Age-Invariant Face Recognition. 10.13140/RG.2.2.24847.23204.

[13] Bashbaghi, Saman & Granger, Eric & Sabourin, Robert & Parchami, Armin. (2018). Deep Learning Architectures for Face Recognition in Video Surveillance.

[14] A. Vinay, Vinay.S. Shekhar, K.N. Balasubramanya Murthy, S. Natarajan, Face Recognition Using Gabor Wavelet Features with PCA and KPCA - A Comparative Study, Procedia Computer Science, Volume 57, 2015, Pages 650-659, ISSN 1877-0509

[15] Z. Lu, X. Jiang and A. Kot, "Deep Coupled ResNet for Low-Resolution Face Recognition," in IEEE Signal Processing Letters, vol. 25, no. 4, pp. 526-530, April 2018.