e-ISSN : 0976-5166
p-ISSN : 2231-3850

AN ENHANCED GAIT RECOGNITION SYSTEM BASED ON THE FEATURES FUSION METHODOLOGY WITH RECURRENT NEURAL NETWORK (RNN)

Jaychand Upadhyay

Assistant Professor, Department of Information Technology, Xavier Institute of Engineering, Mumbai University, Mahim, Mumbai, 400016, India
jaychand.u@xavier.ac.in

Tad Gonsalves

Full Professor, Department of Information and Communication Sciences, Sophia University, 7-1 Kioicho, Chiyoda City, Tokyo 102-8554, Japan
t-gonsal@sophia.ac.jp

Abstract

In computer vision, recognition and identification of individuals from the gait video sequence have recently become one of the demanding research problems. Due to various aspects, like changes in walking speed, variations in viewing angle, clothing, and carrying capacity, robust gait recognition is still challenging. Various methodologies have been developed with deep learning approaches, which intend to improve the robustness of gait recognition systems. However, the existing techniques are inefficient in recognition due to the varying illumination effects and clothing. This research work aims to develop an efficient framework for gait recognition, where the Spatio-Temporal Open Pose Gait (ST- OPG) and Recurrent Neural Network (RNN) techniques are deployed. The four different features such as normalized 2D joint pose features, body parts, angular trajectory, and temporal displacement are extracted and fused, improving the recognition performance. Data augmentation has also been performed by using the RNN technique, which is simple and effective in gait recognition. The widely used CASIA B dataset has been utilized for evaluating the performance of the system. In addition, the obtained results are compared with the state-of-the-art methods for proving the superiority of the proposed technique.

***Keywords*: Biometric Authentication; Deep Learning; Recurrent Neural Network (RNN)**

1. Introduction

Biometric systems (Bernadelli and Silva (2021); De Lima and Schwartz (2019); Hua et al. (2021); Sheng and Li (2020)) are widely used for authenticating and recognizing people based on their characteristics like iris, face, hands, retina, and other body parts. Compared to the other recognition schemes, Gait recognition has gained significant attention in many application systems, due to its behavioural biometric modality. It is a kind of biometric recognition system (Liu et al. (2019); Z. Wang et al. (2022)), which is mainly used for identifying a person based on the walking posture. It is a non-invasive technique when compared to other biometric systems. The key benefit of using the gait recognition (Tran et al. (2021); F. Wang et al. (2018)) system is that it offers good recognition results at various viewing angles. Due to this fact, it has been widely used in many application systems (Rida et al. (2019)) like video surveillance, forensics, and criminal activities investigation processes. However, the performance of gait recognition systems (Uddin et al. (2017); Chi, Wang, and Meng (2017)) highly depends on the appearance of people like their clothing, things that they are carrying, and environmental variations. For resolving this problem, different types of machine learning (Bhargavas et al. (2017)) and deep learning models are used in the conventional works, which mainly intend to obtain outstanding performance outcomes in gait recognition (El-Alfy et al. (2017); Khan et al. (2020); Nattee and Khamsemanan (2019); Singh et al. (2018)). Yet, the major challenges and difficulties faced by the existing methodologies (Kang, et al. (2017); Kwolek et al. (2019); Xu et al. (2019)) are as follows: it is more subject to the intra-class variations of people's appearance, varying illumination effects, and viewing angle.

The major objectives behind this research work are as follows:
•To efficiently extract the features of the input gait video sequence, the Spatio-Temporal Open Pose Gait (ST-OPG) algorithm is utilized.
•To fuse the features of normalized 2D joint pose features, body parts, angular trajectory, and temporal displacement for obtaining an increased recognition performance.
•To train the model for obtaining stabilized and improved performance results, data augmentation has been performed with the use of the Recurrent Neural Network (RNN) classification technique.
•To analyse the results of the proposed mechanism by using the popular CASIA B datasets during evaluation with training and testing.
The remaining part of this study is structured as follows: Section II investigates related works in gait recognition with its advantages and disadvantages. Section III illustrates the working of the proposed methodology for Body Pose Estimation and Gait Recognition. Section IV shows the experimental results and its comparison with the state-of-the-art. Finally, Section V concludes the study with its achievements and shows a direction for future work.

2. Related works

This section discusses the recent works related to body pose estimation, gait recognition, appearance-based methods, and model-based methods. It also describes the strengths and weaknesses of each mechanism based on its working functions and operations.

2.1. Pose Estimation and Gait Recognition

The pose-based gait recognition techniques involve 3D key points for drawing link co-ordinate diagrams of the human body for estimating the body pose. On a similar line, Lan et al. (2022) have used 3D key point coordinates and built a spatial-temporal graph. Further, they utilized the graph for feature extraction in gait recognition. An, et al. (2018) intended to analyse the performance of deep learning mechanisms such as Long Short Term Memory (LSTM) and Convolutional Neural Network (CNN) for body pose recognition. Here, the 3D body pose features have been extracted from the given set by splitting them into a varying number of frames. After that, the features of both LSTM and CNN could be concatenated for obtaining the gait feature based on the center loss and softmax loss computations. The main reason for using these deep learning models was to extract the temporal and spatial features of the given video based on the fusion of functions. The advantages of this work were improved recognition rate and reduced computational complexity. Li, et al. (2020) implemented a Skinned Multi-Person Linear (SMPL) model for developing a framework for gait recognition. The stages involved in this framework were pre-processing, feature extraction, and recognition, where the shape and pose features were extracted based on the loss function computation. Liao, et al. (2017) came up with an enhanced Pose-based Temporal Spatial Network (PSTN) for tuning the performance of gait recognition. Here, the cross-entropy loss and contrast loss have been estimated for reliably extracting the temporal and spatial features. Also, the image normalization was performed based on the distance between two joints such as neck coordinate, position, and center of the hip position.

Zhang, et al. (2019) presented a new encoder framework for extracting the appearance features and disentangling pose by using the representation learning methodology. Here, the cross reconstruction loss and gait similarity loss functions have been computed for estimating the similarity between two videos. In addition to that, the aggregation process was also performed with the help of LSTM technique for illustrating the pose features. The performance of this recognition model could be assessed by using various gait datasets with respect to the measures of runtime, accuracy, and recognition rate. Sokolova, et al. (2019) deployed a deep convolution learning model for extracting the motion estimation by using the gait descriptors. Here, the data pre-processing was performed to train the features for body representation, and temporal representation. Then, the data augmentation could be carried out for uniformly segregating the number of samples, which helps to extract the patches that comprise the body parts.

2.2. Appearance-based Recognition Models

The appearance-based methods for gait recognition deals with the analysis of human body silhouette. Generally, these methods rely on gait energy images (Han and Bhanu (2006)). Yao, et al. (2021) employed a multi-stage CNN model with skeleton extraction for developing a robust gait recognition system. The objective of their study was to estimate the association between the body part candidates with respect to the candidate part positions. Here, the gait period was also estimated for constructing the waveform based on the gait features. Moreover, the transformation model was also used for resolving the view change problem, which comprises the processes of 3D gait feature extraction, view-invariant gait extraction, and view-angle computation. Moreover, the weight adjustment was mainly carried out by integrating the skeleton and gait features. The advantages of this work were increased recognition, ensured robustness, and accuracy. Babaee, et al. (2019) utilized a Generative Adversarial Networks (GAN) for enhancing the gait-based recognition process by identifying the real and fake images. Here, an identity-based discriminator has been used for reconstructing GEI by computing the loss functions. The objective of the study was to efficiently extract the gait features with increased recognition accuracy. De Lima, et al. (2021) recommended a simple and appearance-based gait recognition methodology for accurate body pose estimation. In this system, the features were extracted based on the histograms movements, and body part signals with the help of the PoseDist model. Moreover, the Sequence Dynamic Time Warping (SDTW) based normalization technique was deployed for reducing the matching cost by estimating the squared distance.

2.3. Model-based Recognition Methods

The model-based gait recognition systems try to model the individuals' body to perform recognition. Choi, et al. (2019) employed a model-based framework for gait recognition. It was suitable for minimizing the effects of noise patterns with more robustness. The main objective of this work was to enhance the skeleton quality of body features by constructing the cost matrix based on the input and registered frames. The majority weighted voting scheme was employed for avoiding inaccurate skeleton computations based on frame-level discriminative features. Moreover, the two-stage linear matching was performed based on the quality-adjusted cost matrix, which helps to enhance the weighted majority voting module. Zhang, et al. (2019) developed an Angle Center Loss (ACL) estimation model with LSTM framework for improving the gait recognition. The key consideration of this paper was to pull out both temporal and spatial features based on the loss function, which ensures increased robustness, effectiveness, and gait recognition. From this paper, it was observed that the head and upper legs estimation were more essential parts for improving the cross-view gait recognition. Here, some other discriminative feature learning models were also assessed based on the estimation of the loss function and Spatio-temporal feature learning process. Jinnovart, et al. (2020) suggested a Recurrent Neural Network (RNN) technique incorporated with the multi-level feature fusion methodology for increasing the robustness of gait-based recognition. The aim of the study was to develop the model-based 2D gait recognition framework by analysing the features of body shape and appearance. Elharrouss, et al. (2021) introduced a new gait recognition approach with the multitask CNN technique for designing the human re-identification system. This work examined some of the most important features used for improving the recognition system, which includes the types of gait energy image, sequence, motion data, walking speed, and joint relationship. In order to verify the increased detection accuracy of this framework, background subtraction method has been incorporated with the stages of angle estimation, and gait verification. Fan, et al. (2020) utilized a Micro-Motion Capture Module (MCM) for enhancing the fine-grained learning features used for gait recognition. Also, the Focal Convolution (FConv) methodology was used for slicing the features with respect to the input feature map. For this purpose, the frame level feature extractor has been utilized, which separates the features into various blocks. Luo and Tjahjadi (2020) recommended a model based gait recognition methodology for detecting the accurate 2D and 3D body poses. Here, the structural gait semantic image has been constructed based on the semantic folding mechanism, which helps to extract the body parameters by constructing the 2D metric. Moreover, the human body parsing was performed for computing the 3D gait features, which were further used by the sequence learner for accurate recognition.

After going through the related studies on gait recognition, it is analysed that most of the studies are predominantly concentrating on improving the process of gait recognition based on the feature learning model. Still, it faces major challenges related to the increased time consumption for learning the features, inefficient recognition, and reduced accuracy. To solve these problems, this study proposes a framework for gait recognition based on the amalgamation of features.

3. Proposed Methodology

This section illustrates the detailed working of the proposed methodology by using an algorithmic flow graph. The key intention of this paper is to develop an efficient gait recognition system with increased robustness based on the fusion of features. Fig. 1 represents the detailed functioning of the proposed system. It involves the stages of input gait video obtainment, body pose estimation, feature extraction, the fusion of features, and recognition. Initially, the input gait video is obtained from the dataset, and 2D and 3D body poses have been extracted based on Spatio-Temporal Open Pose Gait (ST-OPG) algorithm. In this design, four different Spatio-temporal features such as normalized 2D, joint angular trajectory, temporal displacement, and body parts are extracted for enhancing the robustness and accuracy of the gait-based recognition system. Consequently, these feature vectors have been fused for obtaining improved performance results. Finally, the RNN-based deep learning model is implemented for accurately predicting the subject classes. The stages involved in this system are as follows:

- Input video obtainment
- 2D body pose feature extraction
- Preprocessing
- Data augmentation

• Post-processing

3.1. *Feature Extraction*

At first, the 2D body pose features are extracted by using the Spatio-Temporal Open Pose Gait (ST-OPG) algorithm, which is mainly used for extracting four different types of features such as:

• Normalized 2D joint pose features
• Body parts
• Angular trajectory
• Temporal displacement

Typically, all joint parts of the human body are not required for improving the recognition rate of gain patterns. Hence, this work considered around 25 numbers of joint parts that are having rich gait representation, which is estimated by using the ST-OPG technique. Based on the literature, it is identified that change in the human leg is one of the essential features used for improving the gait recognition. This study mainly considers the knee joint, because it provides increased robustness compared to the other joints. For instance, some of them can put their hands into the pocket of their coat, but they cannot do the same during normal walking. Hence, this type of situation may change the coordinates of joints. This work considers six different and essential joints such as Left Knee (LK), Right Knee (RK), Left Big Toe (LBT), Right Big Toe (RBT), Left Ankle (LK), and Right Ankle (RK). For these 6 joints, there are 12-dimensional feature vectors extracted for a single frame as shown below:

$$P_{FE} = [a_1, b_1, a_2, b_2 \dots a_6 b_6]^T \quad (1)$$

For obtaining an improved performance outcome, it is necessary to normalize the sequence data with different poses based on the position, size, and walking speed. Here, the origin of coordinates O_{Jc} has been extracted for each subject at the left, right and middle hip joints, and then the average of these coordinates is estimated. Then, the Euclidean distance *D* is computed for normalizing the different coordinates of body parts from the hip joint to the neck joint. It is represented as follows:

$$O_{Jc} = \frac{(O_{LeftHip} + O_{RightHip} + O_{MiddleHip})}{3} \quad (2)$$

$$D = ||O_{Jc} - O_{neck}||_2 \quad (3)$$

$$O_i^N = \frac{(O_i - O_{Jc})}{D} \quad (4)$$

Where, O^N represents the new coordinate of joint O_i of the particular pose. After extracting the joint pose features, the joint angles are estimated based on the temporal information for analysing the dynamics of human gait motion. In this stage, the discriminative features have been extracted with the lower limbs of trajectories, and are estimated as follows:

$$\delta = tan^{-1} \frac{O_{2,x} - O_{1,x}}{O_{2,y} - O_{1,y}} \quad (5)$$

$$\gamma = tan^{-1} \frac{O_{3,x} - O_{1,x}}{O_{3,y} - O_{1,y}} \quad (6)$$

$$\alpha = \delta_1 + \delta_2 \quad (7)$$

Here, the joints O_1, O_2, and O_3 form the angular trajectory pattern, which comprises five sets of angular patterns of the lower limb of human body. Each trajectory feature is considered a gait feature and is computed as follows:

$$T_{FE} = [\alpha_1, \delta_1, \gamma_1, \alpha_2, \delta_2, \gamma_2 \dots \alpha_5, \delta_5, \gamma_5]^T \quad (8)$$

After that, the temporal displacement feature has been extracted, which comprises the displacement information two adjacent frames. Consider the adjacent frames of *n* and *n+1*, where the coordinates of any two adjacent frames are used to estimate the normalized difference value as shown below:

$$\Delta c_1^n = \frac{c_1^{n+1} - c_1^n}{\sum_{i=8}^{8} ||O_1^{n+1} - O_1^n||_2^2} \quad (9)$$

$$\Delta d_1^n = \frac{d_1^{n+1} - d_1^n}{\sum_{i=8}^{8} ||O_1^{n+1} - O_1^n||_2^2} \quad (10)$$

$$D_{FE} = [\Delta c_1, \Delta d_1, \Delta c_2, \Delta d_2 \dots \Delta c_8, \Delta d_8]^T \quad (11)$$

Where, O_i^n indicates the 2*D* coordinate of i^{th} body joint in n^{th} frame of video, and $(\Delta c_1^n, \Delta d_1^n)$ represents the displacement of coordinates. In this stage, the 16-dimensional feature vector has been obtained with the selected 8 number of joint coordinates. Consequently, the static gait features are also extracted in this algorithm. For instance, the length of body parts is extracted with respect to the position of joint coordinates. Typically, the spatial gait features are most essential for gait recognition, because it increases the robustness of recognition against the carrying and clothing changes. Finally, all these extracted features are fused for obtaining an improved performance result. In this work, the multiple features extracted from the same frame are integrated together before augmentation. The constructed feature matrix with respect to the different number of frames and multiple features is represented in Fig 2.

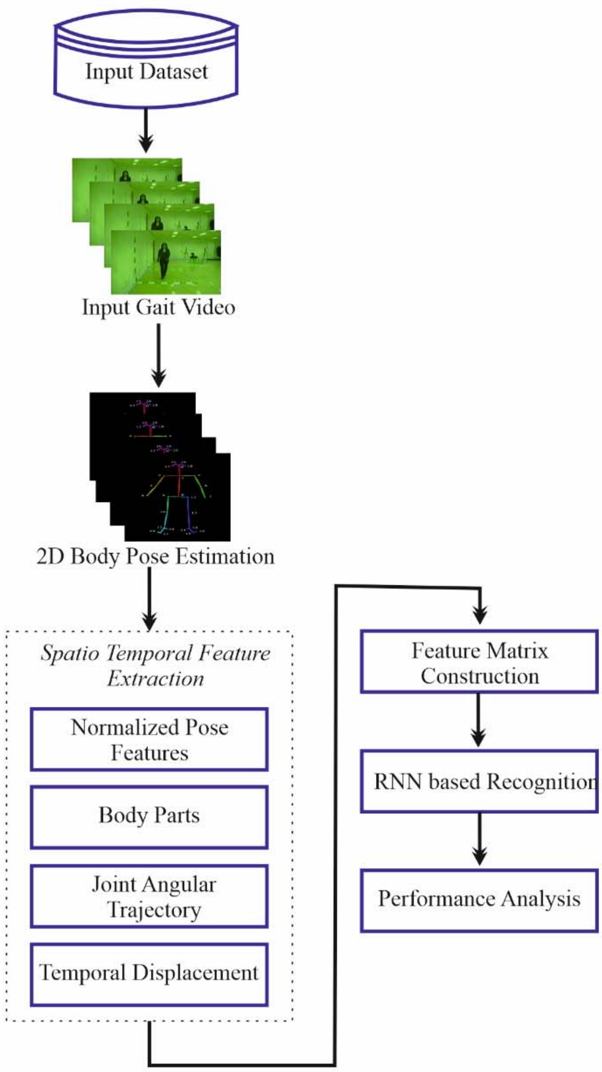

Figure 1. The overall flow of the proposed gait recognition system

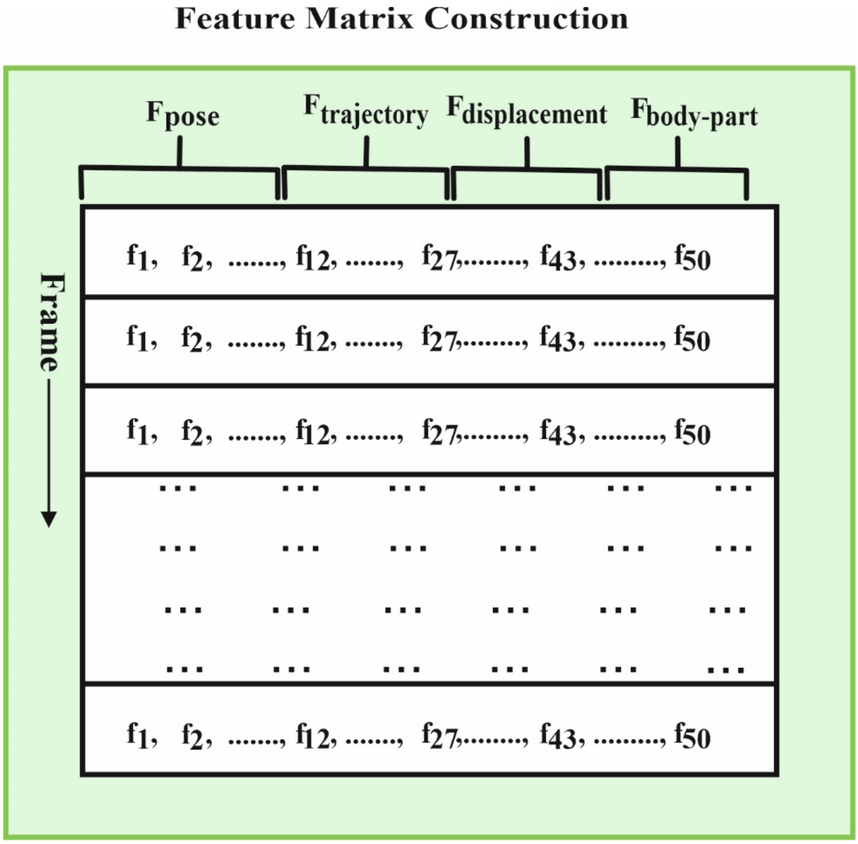

Figure 2. Feature matrix construction

3.2. *Pre-processing of Extracted Features*

Generally, the feature pre-processing is responsible for handling the missing data problem by eliminating the occlusions of frames. The following processes are performed for pre-processing the feature vectors:

(1) Due to the missing values of hip joints, it is difficult to estimate the origin of coordinates; hence the frame has been excluded.

(2) Due to the minimum extracted information, more than one body joints are lost between the portions of knee and ankle joints; hence this type of frame has been excluded.

(3) In some part, it is difficult to locate the individual joints in the frame, hence it has been excluded.

3.3. *Data Augmentation*

In this work, the data augmentation is mainly performed to train the model for obtaining the stabilized and improved performance results. Here, the CASIA B gait dataset has been utilized for evaluation that allows training the model with four normal walking samples. Thus, this work intends to augment the data for improving the training model of CASIA B dataset. Also, overlapping the video clips is one of the suitable solutions for enhancing the amount of training model. For this purpose, this work splits the input video frame into varying number of video clips, where for every video clip, the varying number of images has been overlapped with the previous clip. The most extensively used deep learning model named as, RNN is utilized in this work for obtaining robust gait recognition. Typically, this architecture contains 2 BiGRU layers and 1 softmax layer in which each BiGRU comprises around 80 cells with the batch normalization and softmax layer correspondingly. For this network, the extracted 50-dimensional features are fed into the input, and the batch normalization layer is used to normalize the extracted features. Then, the standardized activations are fed to the out- put softmax layer, which comprises different output neurons for producing the predicted results with subject ID. The architecture of the proposed RNN-based gait recognition process is shown in Fig 3. During the training of classifier, the set of parameters are tuned with the sequence by using the Adam optimization technique. The overfitting problem is resolved by adding the batch normalization later, which helps to improve the learn capacity of network.

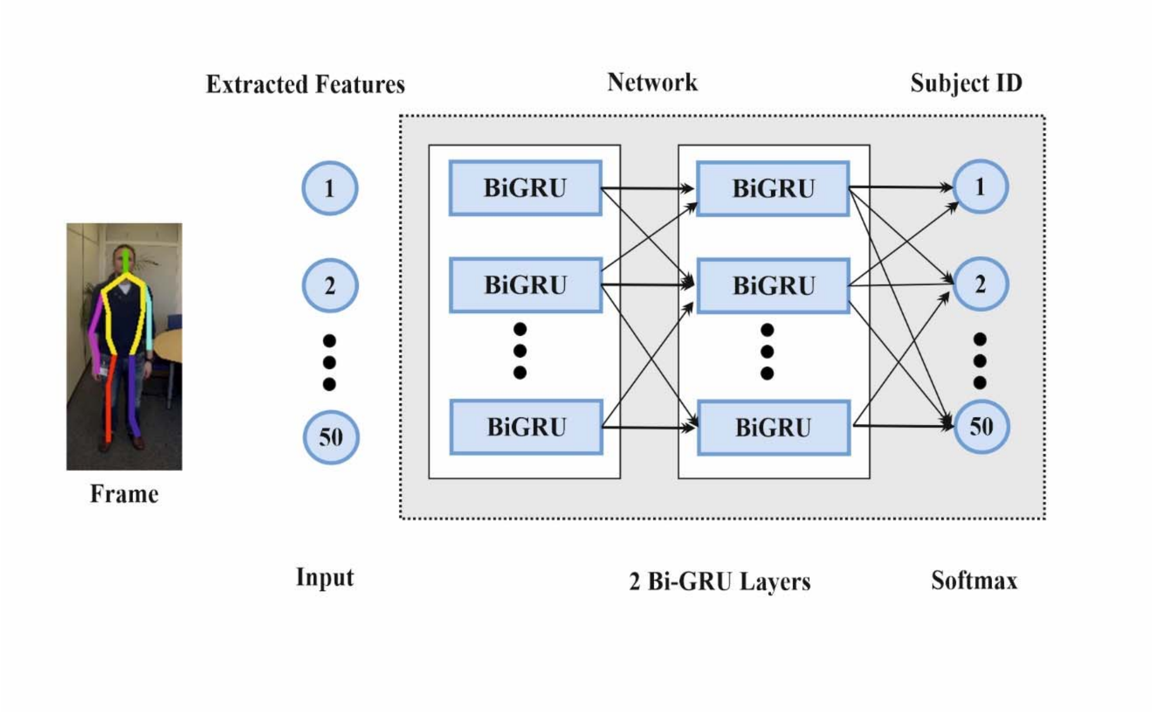

Figure 3. Architecture of RNN based gait recognition system

3.4. *Post Processing*

After training the model, the testing has been accomplished that requires the subject ID for obtaining the complete gait video sequence. For this purpose, the majority voting mechanism has been utilized in this work, which is mainly used for accurately predicting the subject ID based on the increased count of votes at all-time steps. Fig. 4 illustrates the output prediction of the proposed model. Here the separate video sequences are taken as the input, and classes of probabilities are produced as the output. Then, the obtained output has been further processed by the voting scheme for predicting the subject ID, which is illustrated as follows:

$$V^{ts} = [V_1, V_2 \dots V_{ns}]^T \quad (12)$$

$$OV^{ts} = [O_1, O_2 \dots O_{ns}]^T \quad (13)$$

$$OV_i^{ts} = P(V_i | M^{ts}) \quad (14)$$

Where, V_{ns} indicates the vector with the number of subjects, ts is the time stamp of gait video sequence, $M^{ts} \varepsilon R^{28\times50}$ defines the feature matrix, and O_v is the output vector. Furthermore, the subject ID has been predicted by using the majority voting scheme as shown in below:

$$V = argmax\{V_i^{ts} | 1 < i \leq ns\} \quad (15)$$

$$V = argmax_{i\varepsilon(1,2\dots n)} \sum_{ts=1}^{NT} V_i^{ts} \quad (16)$$

Where, NT indicates the total count of time steps involved, and V is produced as the final predicted result.

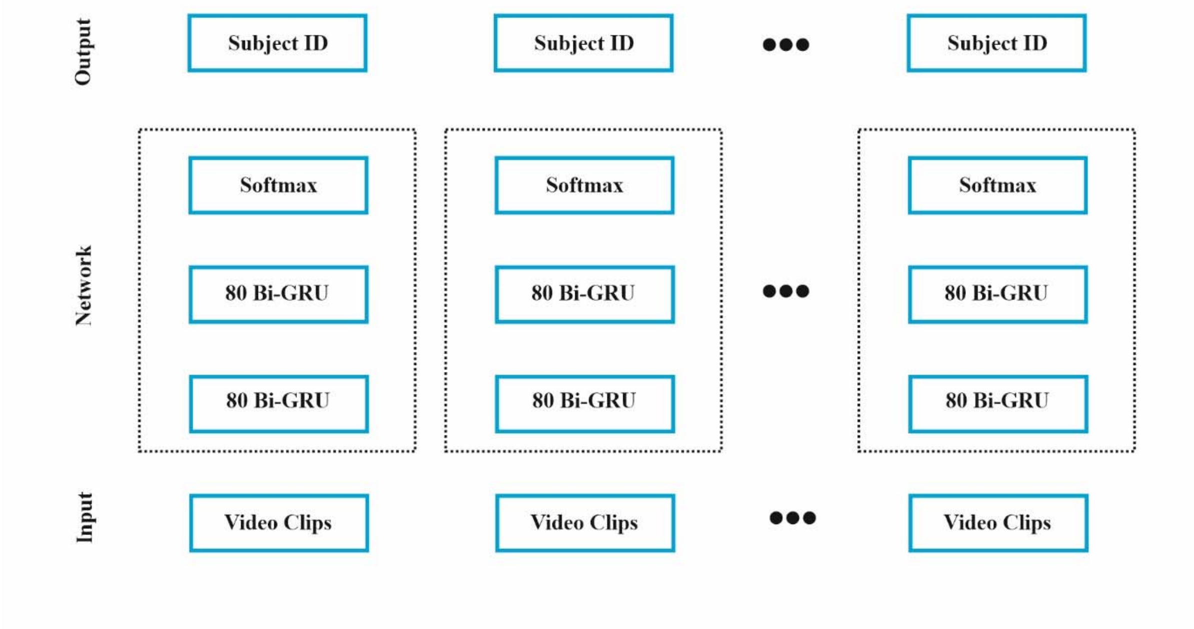

Figure 4. Output prediction of RNN

4. RESULTS AND DISCUSSION

This section compares the results of benchmarking works with the proposed method. For assessing the performance of the proposed methods, standard dataset has been utilized in this work, which comprises different RGB colour video frames.

4.1. Dataset

We have found original gait frames from the CASIA B gait dataset. Hence, we utilized CASIA B dataset (The Institute of Automation, Chinese Academy of Sciences (CASIA) CASIA B dataset http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp (2005)) in this work for result verification. The CASIA B dataset comprises 124 subjects (including both males 93 and females 31, where each subject includes different sequences. The CASIA B walking sequences are categorized into normal walk (NM) – 6 sequences, walking with bag carrying (BG) – 2 Sequences, and, walking with coat (CL) – 2 sequences. These sequences are captured at the same time with eleven viewing angles {0°, 18°, 36°, 72° … 180°}. Fig. 5 represents the different sequences of images obtained from the CASIA-B dataset.

Figure 5. Different walking sequences of CASIA B [36]

4.1. *Experimental setting on CASIA B*

This study categorized the CASIA B subjects for training and testing of the proposed method. The training phase considers subject id: 01 to 62 whereas the testing phase considers subject id: 63 to 124. The training phase involves all sequences of subjects (id: 01 to 62 with 6 NM sequences, 2 BG Sequences, and 2 CL sequences). The testing phase is divided into gallery set and probe set. The gallery set considers 4 NM sequences (NM01–NM04) whereas probe set considers the remaining 2 NM sequences (NM05–NM06), 2 BG sequences, and 2 CL sequences.

4.2. *Gait Recognition on CASIA B*

We have evaluated the performance of gait recognition on CASIA B; the outcome of the experiment is enlisted in Table 1– Table 3. We have considered 4 NM sequences (NM01–NM04) in the gallery set. Table 1 illustrates the performance of gait recognition with the probe set containing last 2 NM sequences (NM05–NM06). Table 2 illustrates the performance of gait recognition with a probe set containing 2 BG sequences (BG01–BG02). Table 3 illustrates the performance of gait recognition with a probe set containing 2 CL sequences (CL01–CL02). We know that one sequence has 11 different viewpoints in the gallery set. Similarly, there are 11 different viewpoints in the probe set as well. This means for each probe set, this experiment will have 121 different outcomes for the recognition rate.

Table 1. Performance of gait Recognition (% accuracy) considering NM #5-6 as the probe data.

Probe Angle NM # 5-6		0	18	36	54	72	90	108	126	144	162	180
	0	97.95	89.76	62.34	46.79	28.37	27.61	28.42	31.62	48.53	66.34	64.72
	18	90.62	97.75	96.23	84.15	55.92	36.61	42.28	38.32	61.53	66.47	65.62
	36	70.28	96.23	96.97	94.38	77.72	67.24	60.26	52.63	59.92	60.21	46.24
Gallery	54	39.83	82.47	96.23	97.74	92.93	84.17	76	60.18	55.12	45.37	39.74
Angle	72	33.28	55.12	76.85	89.74	96.97	97.74	80.87	71.93	60.69	38.21	25.27
	90	28.36	40.56	67.92	82.42	92.23	98.54	91.34	71.93	67.21	42.17	23.67
NM	108	28.36	38.93	63.94	72.78	84.09	90.57	98.54	92.17	82.56	60.69	32.59
#1-4	126	39.73	45.32	54.27	67.15	71.23	76.86	93.76	95.36	92.95	80.87	44.57
	144	43.83	51.83	59.86	63.92	54.25	65.57	84.89	95.38	97.74	88.96	61.52
	162	63.23	61.47	52.63	42.97	38.12	32.46	54.28	72.78	90.54	98.56	81.69
	180	69.67	64.72	50.23	32.48	23.59	21.24	22.83	36.53	63.87	90.34	98.65

Table 2. Performance of gait Recognition (% accuracy) considering BG #1-2 as the probe data.

Probe Angle BG # 1-2		0	18	36	54	72	90	108	126	144	162	180
	0	75.21	57.47	42.19	28.45	17.18	15.57	17.97	21.98	26.05	38.92	35.69
	18	57.47	76.84	73.59	59.07	35.67	30.08	20.37	29.24	33.27	34.89	34.86
	36	42.96	71.19	78.46	75.21	57.48	43.76	41.35	38.91	36.49	36.49	30.08
Gallery	54	31.67	61.46	74.41	77.64	66.34	57.46	50.21	43.76	38.21	31.67	22.76
Angle	72	23.57	43.76	55.87	66.34	70.36	63.93	53.42	46.17	39.72	24.38	19.56
NM	90	21.97	29.27	52.63	60.69	67.98	71.19	59.09	55.86	45.36	23.59	15.54
#1-4	108	20.37	27.62	42.95	44.57	60.69	67.18	71.98	66.36	58.27	32.47	17.18
	126	22.78	29.27	33.89	43.76	46.98	51.83	66.34	70.37	66.34	44.57	25.21
	144	34.09	30.67	33.08	36.47	38.13	49.41	51.02	61.48	75.21	62.31	34.09
	162	37.31	42.97	30.07	24.42	23.59	29.31	30.08	33.25	59.85	66.34	51.08
	180	43.75	36.47	22.78	17.14	15.54	16.36	17.21	22.76	31.67	53.44	61.42

Table 3. Performance of gait Recognition (% accuracy) considering CL #1-2 as the probe data.

Probe Angle CL # 1-2		0	18	36	54	72	90	108	126	144	162	180
	0	47.79	34.89	22.76	14.72	15.54	16.34	13.96	24.41	26.03	29.26	25.18
	18	34.89	49.38	53.44	33.28	27.63	17.97	21.98	22.78	30.87	26.07	23.57
Gallery	36	27.62	43.75	57.28	51.82	46.98	32.44	33.27	36.44	32.44	29.24	22.76
	54	18.76	24.41	57.47	62.28	52.67	48.54	38.92	38.92	31.62	22.76	13.89
Angle	72	15.54	21.17	48.59	52.62	59.06	51.82	46.17	43.73	32.44	22.76	13.12
	90	10.62	20.37	38.12	56.67	55.05	57.43	59.08	51.04	35.67	21.98	10.89
NM	108	9.88	14.76	34.09	40.34	49.37	45.37	60.64	48.54	42.98	21.98	11.53
#1-4	126	14.72	16.35	24.41	30.08	34.08	41.33	62.28	55.83	47.76	30.06	18.76
	144	22.76	26.84	29.24	30.05	27.62	31.67	49.41	52.62	56.67	44.56	21.98
	162	30.05	23.54	23.54	22.76	14.73	14.73	29.24	31.67	43.76	59.08	38.92
	180	30.83	24.41	13.94	10.69	13.92	14.73	15.54	21.19	24.38	37.23	40.26

4.3. *Result Comparison*

This study compares the performance of gait recognition with other benchmarking work based on the average recognition rate. We have considered the following benchmarking works: GEI+PCA (Han and Bhanu (2006)), SPAE (Yu et al. (2017)), GaitGANv1 (Yu et al. (2017)), GaitGANv2 (Yu et al. (2019)), PoseGait (Liao et al.

(2020)). From table 4-6 and Fig. 6-8, it can be observed that the performance of the proposed system is superior to other studies considered (mentioned above) for comparison under every probe set. Based on the results obtained in the experiment, it can be derived that the proposed study improves the gait recognition performance under viewing angle variation. The outcome of the experiment under BG sequence and CL sequences authenticates the robustness of a system against various occlusion of gait like bag carrying and wearing a long coat. It can also be seen that the performance of the proposed method is superior to other benchmarking works even if the variation between probe angle and gallery angle is more.

Table 4. Comparative analysis of proposed techniques with other benchmarking methods under NM probe type.

Probe Type	**GEI+PCA**	**SPAE**	**GaitGANv1**	**GaitGANv2**	**PoseGait**	**Proposed Method**
Probe NM	29.09	62.82	60.96	66.34	63.78	**64.88**

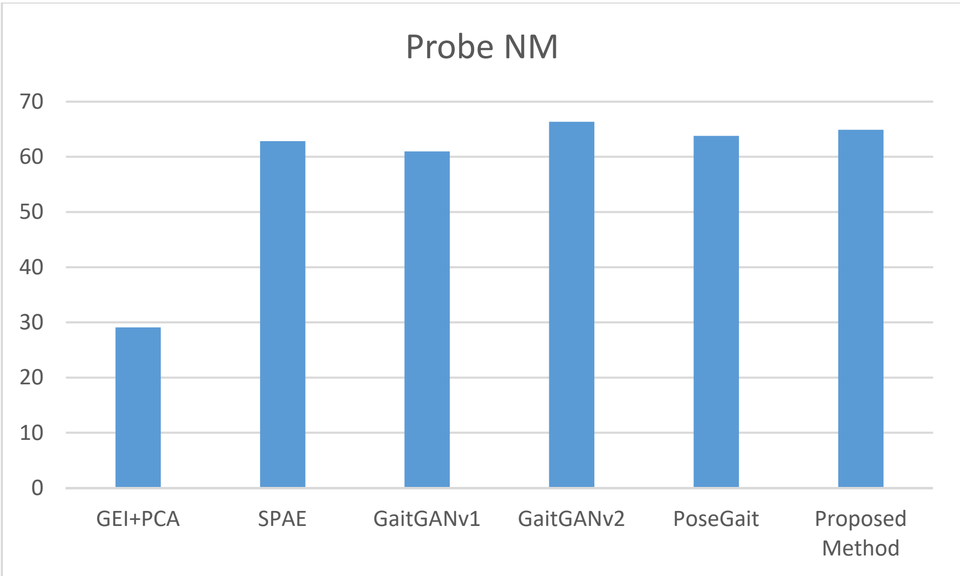

Figure 6. Comparative analysis of the performance of average recognition rate under viewing angle variations based on NM probe set

Table 5. Comparative analysis of proposed techniques with other benchmarking methods under BG probe type.

Probe Type	**GEI+PCA**	**SPAE**	**GaitGANv1**	**GaitGANv2**	**PoseGait**	**Proposed Method**
Probe BG	17.3	40.38	39	46.17	42.52	**43.31**

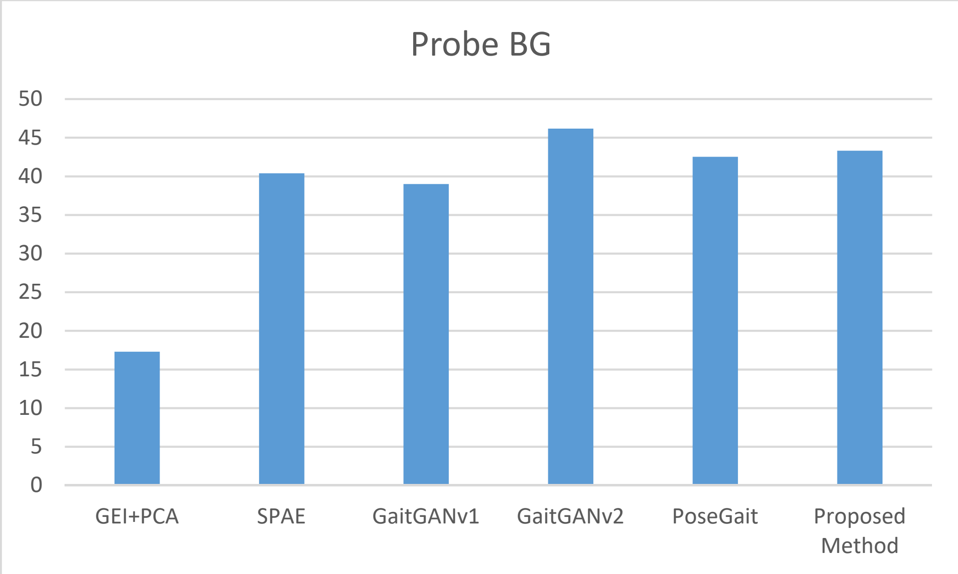

Figure 7. Comparative analysis of the performance of average recognition rate under viewing angle variations based on BG probe set.

Table 6: Comparative analysis of proposed techniques with other benchmarking methods under CL probe type.

Probe Type	GEI+PCA	SPAE	GaitGANv1	GaitGANv2	PoseGait	Proposed Method
Probe CL	6.55	26.05	22.31	25.91	31.98	**32.96**

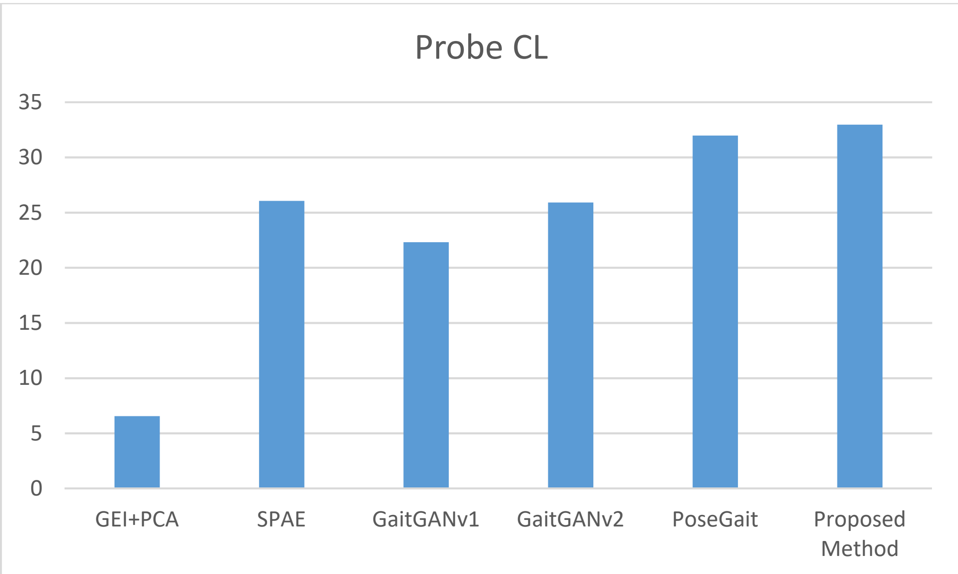

Figure 8. Comparative analysis of the performance of average recognition rate under viewing angle variations based on CL probe set.

4.4. *Result analysis based on body features on CASIA B*

In the proposed recognition system, the features like body pose, angle, limb, and motion have been extracted and fused for improving the average recognition rate. Here, the recognition rate is computed for all probe sequences corresponding to the extracted and fused set of features-refer to Fig.9 and Table 7. From the performance assessment, it is analysed that the proposed gait recognition mechanism yields an enhanced recognition rate by efficiently extracting the feature coordinate points.

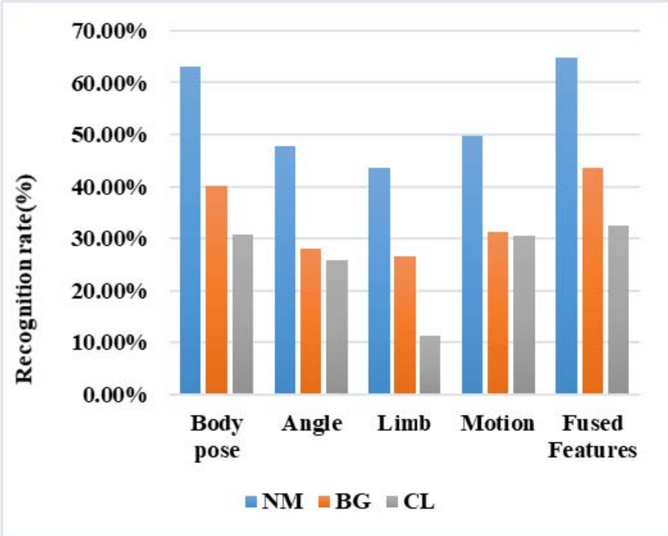

Figure 9. Recognition rate of the proposed mechanism under all sequences of CASIA B dataset with four different extracted features

Table 7. Analysis of recognition rate with respect to varying features

Features	Probe Set		
	NM	BG	CL
Body pose	62.98%	40.22%	30.75%
Angle	47.89%	27.99%	25.75%
Limb	43.65%	26.66%	11.25%
Motion	49.68%	31.25%	30.45%
Fused Features	64.82%	43.65%	32.46%

5. CONCLUSION

This paper presented an enhanced framework based on ST-OPG with RNN models for an efficient gait recognition system. The motive of this study is to extract multiple features of the gait sequence obtained from the input datasets for obtaining an improved recognition rate. The major advantages of this system are increased robustness, simplicity, effectiveness, and high recognition rate. Here, the human body pose information is considered as the gait features such as normalized 2D joint pose features, body parts, angular trajectory, and temporal displacement. This feature extraction methodology provides an increased recognition rate under varying clothing and carrying conditions. Also, the RNN technique is utilized for performing the data augmentation, which is computationally inexpensive in comparison to other deep learning methods. Throughout the process of performance assessment, the proposed recognition mechanism results are verified using CASIA B datasets. From the results, it is learned that the proposed technique outperforms the other benchmarking works with increased robustness and recognition rate. However, in the case of gait occlusion, the performance of the proposed system declines drastically to cross view angles. Thus, there is research direction to enhance the performance of the proposed system under large view variations.

The vision-based system finds difficulties in perceiving the joint pose features under gait occlusion, especially in the case of a person wearing a long coat. To improve the recognition rate, we can increase the number of viewpoints for perceiving more gait features and perform extensive training of a system.

Conflicts of interest: The authors have no conflicts of interest to declare.

References

[1] An, W., Liao, R., Yu, S., Huang, Y., & Yuen, P. C. (2018). Improving Gait Recognition with 3D Pose Estimation. *Biometric Recognition*, 137–147. https://doi.org/10.1007/978-3-319-97909-0_15
[2] Babaee, M., Zhu, Y., Kopuklu, O., Hormann, S., & Rigoll, G. (2019). Gait Energy Image Restoration Using Generative Adversarial Networks. *2019 IEEE International Conference on Image Processing (ICIP)*. https://doi.org/10.1109/icip.2019.8803236
[3] Bernadelli, R., & Silva, P. R. D. (2021). Dynamic Time Warping in Iris Biometric Recognition Process. IEEE Latin America Transactions, 19(01), 42-49.
[4] Bhargavas, K. W. G., Harshavardhan, Mohan, A. N., Sharma, C., & Prathap. (2017). Human identification using gait recognition. International Conference on Communication and Signal Processing, 1510-1513.
[5] Chi, W., Wang, M. Q., & Meng, H. (2017). A gait recognition method for human following in service robots. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 48, 1429-1440.
[6] Choi, S., Kim, Kim, C., & Kim. (2019). Skeleton-based gait recognition via robust frame-level matching. IEEE Transactions on

Information Forensics and Security, 14, 2577-2592.
[7] De Lima, V. C., Melo, V. H., & Schwartz, W. R. (2020). Simple and efficient pose-based gait recognition method for challenging environments. *Pattern Analysis and Applications*, *24*(2), 497–507. https://doi.org/10.1007/s10044-020-00935-z
[8] De Lima, V. C., & Schwartz, W. R. (2019). Gait Recognition Using Pose Estimation and Signal Processing. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, 719–728. https://doi.org/10.1007/978-3-030-33904-3_68
[9] El-Alfy, H., Mitsugami, I., & Yagi, Y. (2018). Gait Recognition Based on Normal Distance Maps. *IEEE Transactions on Cybernetics*, *48*(5), 1526–1539. https://doi.org/10.1109/tcyb.2017.2705799
[10] Elharrouss, O., Almaadeed, N., Al-Maadeed, S., & Bouridane, A. (2020). Gait recognition for person re-identification. *The Journal of Supercomputing*, *77*(4), 3653–3672. https://doi.org/10.1007/s11227-020-03409-5
[11] Fan, C., Peng, Y., Cao, C., Liu, X., Hou, S., Chi, J., Huang, Y., Li, Q., & He, Z. (2020). GaitPart: Temporal Part-Based Model for Gait Recognition. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr42600.2020.01423
[12] Han, J., & Bhanu, B. (2006). Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(2), 316–322. https://doi.org/10.1109/tpami.2006.38
[13] Hua, C. H., Kim, K., Huynh-The, T., You, J. I., Yu, S. Y., Le-Tien, T., Bae, S. H., & Lee, S. (2021). Convolutional Network With Twofold Feature Augmentation for Diabetic Retinopathy Recognition From Multi-Modal Images. *IEEE Journal of Biomedical and Health Informatics*, *25*(7), 2686–2697. https://doi.org/10.1109/jbhi.2020.3041848
[14] Jinnovart, T., Cai, X., & Thonglek, K. (2020). Abnormal Gait Recognition in Real-Time using Recurrent Neural Networks. *2020 59th IEEE Conference on Decision and Control (CDC)*. https://doi.org/10.1109/cdc42340.2020.9304106
[15] Kang, Z., Deng, M., & Wang, C. (2017). Frontal-view human gait recognition based on Kinect features and deterministic learning. 36th Chinese Control Conference (CCC), 10834-10839. https://doi.org/10.23919/ChiCC.2017.8029085
[16] Khan, M. H., Farid, M. S., & Grzegorzek, M. (2020). A non-linear view transformations model for cross-view gait recognition. *Neurocomputing*, *402*, 100–111. https://doi.org/10.1016/j.neucom.2020.03.101
[17] Kwolek, B., Michalczuk, A., Krzeszowski, T., Switonski, A., Josinski, H., & Wojciechowski, K. (2019). Calibrated and synchronized multi-view video and motion capture dataset for evaluation of gait recognition. Multimedia Tools and Applications, 78, 32437-32465.
[18] Lan, T., Shi, Z., Wang, K., & Yin, C. (2022). Gait Recognition Algorithm based on Spatial-temporal Graph Neural Network. *2022 International Conference on Big Data, Information and Computer Network (BDICN)*. https://doi.org/10.1109/bdicn55575.2022.00018
[19] Li, X., Makihara, Y., Xu, C., Yagi, Y., Yu, S., & Ren, M. (2021). End-to-End Model-Based Gait Recognition. *Computer Vision – ACCV 2020*, 3–20. https://doi.org/10.1007/978-3-030-69535-4_1
[20] Liao, R., Cao, C., Garcia, E. B., Yu, S., & Huang, Y. (2017). Pose-Based Temporal-Spatial Network (PTSN) for Gait Recognition with Carrying and Clothing Variations. *Biometric Recognition*, 474–483. https://doi.org/10.1007/978-3-319-69923-3_51
[21] Liao, R., Yu, S., An, W., & Huang, Y. (2020). A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognition*, *98*, 107069. https://doi.org/10.1016/j.patcog.2019.107069
[22] Liu, Y., Jiang, X., Sun, T., & Xu, K. (2019). 3D Gait Recognition Based on a CNN-LSTM Network with the Fusion of SkeGEI and DA Features. *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. https://doi.org/10.1109/avss.2019.8909881
[23] Luo, J., & Tjahjadi, T. (2020). Gait Recognition and Understanding Based on Hierarchical Temporal Memory Using 3D Gait Semantic Folding. *Sensors*, *20*(6), 1646. https://doi.org/10.3390/s20061646
[24] Nattee, C., & Khamsemanan. (2019). A deep neural network approach for model-based gait recognition. Thai Journal of Mathematics, 17, 89-97.
[25] Rida, I., Almaadeed, N., & Almaadeed, S. (2018). Robust gait recognition: a comprehensive survey. *IET Biometrics*, *8*(1), 14–28. https://doi.org/10.1049/iet-bmt.2018.5063
[26] Sheng, W., & Li, X. (2020). Siamese denoising autoencoders for joints trajectories reconstruction and robust gait recognition. *Neurocomputing*, *395*, 86–94. https://doi.org/10.1016/j.neucom.2020.01.098
[27] Singh, J. P., Jain, S., Arora, S., & Singh, U. P. (2018). Vision-Based Gait Recognition: A Survey. *IEEE Access*, *6*, 70497–70527. https://doi.org/10.1109/access.2018.2879896
[28] Sokolova, A., & Konushin, A. (2019). Pose-based deep gait recognition. *IET Biometrics*, *8*(2), 134–143. https://doi.org/10.1049/iet-bmt.2018.5046.
[29] The Institute of Automation, Chinese Academy of Sciences (CASIA). (2005, January). *Center for Biometrics and Security Research* (CASIA B) [Gait Dataset]. http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp
[30] Tran, L., Hoang, T., Nguyen, T., Kim, H., & Choi, D. (2021). Multi-Model Long Short-Term Memory Network for Gait Recognition Using Window-Based Data Segment. *IEEE Access*, *9*, 23826–23839. https://doi.org/10.1109/access.2021.3056880
[31] Uddin, M. Z., Khaksar, W., & Torresen, J. (2017). A robust gait recognition system using spatiotemporal features and deep learning. *2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. https://doi.org/10.1109/mfi.2017.8170422
[32] Wang, C., Zhang, F., Zhu, X., & Ge, S. S. (2022). Low-resolution human pose estimation. *Pattern Recognition*, *126*, 108579. https://doi.org/10.1016/j.patcog.2022.108579
[33] Wang, C., Li, Z., & Sarpong, B. (2021). Multimodal adaptive identity-recognition algorithm fused with gait perception. *Big Data Mining and Analytics*, *4*(4), 223–232. https://doi.org/10.26599/bdma.2021.9020006
[34] Xu, C., Makihara, Y., Li, X., Yagi, Y., & Lu, J. (2019). Speed-Invariant Gait Recognition Using Single-Support Gait Energy Image. *Multimedia Tools and Applications*, *78*(18), 26509–26536. https://doi.org/10.1007/s11042-019-7712-3
[35] Yao, L., Kusakunniran, W., Wu, Q., Zhang, J., Tang, Z., & Yang, W. (2021). Robust gait recognition using hybrid descriptors based on Skeleton Gait Energy Image. *Pattern Recognition Letters*, *150*, 289–296. https://doi.org/10.1016/j.patrec.2019.05.012
[36] Yu, S., Chen, H., Reyes, E. B. G., & Poh, N. (2017). GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. https://doi.org/10.1109/cvprw.2017.80
[37] Yu, S., Chen, H., Wang, Q., Shen, L., & Huang, Y. (2017). Invariant feature extraction for gait recognition using only one uniform model. *Neurocomputing*, *239*, 81–93. https://doi.org/10.1016/j.neucom.2017.02.006
[38] Yu, S., Liao, R., An, W., Chen, H., García, E. B., Huang, Y., & Poh, N. (2019). GaitGANv2: Invariant gait feature extraction using generative adversarial networks. *Pattern Recognition*, *87*, 179–189. https://doi.org/10.1016/j.patcog.2018.10.019
[39] Zhang, Y., Huang, Y., Yu, S., & Wang, L. (2020). Cross-View Gait Recognition by Discriminative Feature Learning. *IEEE Transactions on Image Processing*, *29*, 1001–1015. https://doi.org/10.1109/tip.2019.2926208
[40] Zhang, Z., Tran, L., Yin, X., Atoum, Y., Liu, X., Wan, J., & Wang, N. (2019). Gait Recognition via Disentangled Representation Learning. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2019.00484

e-ISSN : 0976-5166
p-ISSN : 2231-3850

Authors Profile:

Jaychand Upadhyay is PhD candidate at Information & Communication Sciences Department, Sophia University, Tokyo. He received his B.E and M.E. degree from Mumbai University, India, in 2005 and 2011, respectively. He has published 20 papers in international conference and journals. His research interests include gait recognition, computer vision and deep learning.

Tad Gonsalves received the B.S. degree in theoretical physics, the M.S. degree in astrophysics, and the Ph.D. degree in information systems from Sophia University, Tokyo, Japan. He is a Full Professor with the Department of Information and Communication Sciences, Faculty of Science and Technology, Sophia University. He has published 100 papers in international conferences and journals. He is the author of the book Artificial Intelligence: A Non-Technical Introduction (Sophia University Press, Tokyo, Japan, 2017) and co-author of Artificial Intelligence for Business Optimization: Research and Applications (CRC press, London, 2021). His research interests include bio-inspired optimization techniques and the application of deep learning techniques to diverse problems, such as autonomous driving, drones, digital art and music, and computational linguistics. His research laboratory (https://www.gonken.tokyo/) in Tokyo specializes in applications of deep learning and multi-GPU computing.