International Journal of Applied Engineering and Technology ISSN: 2277-212X
An Open Access, Online International Journal Available at http://www.cibtech.org/jet.htm
2022, Vol.12, pp.313-320/Sathiya et al.
Research Article

DEEP LEARNING BASED GAIT IMAGE RECOGNITION

K. Sathiya Bharathi, B.K. Akshaya, G. Divyashri, R. Haritha, R. Neerthana
Department of Electronics & Communication Engineering, Anjalai AmmalMahalingam Engineering College, Kovilvenni, Thiruvarur
**Author for Correspondence: sathiya2188@gmail.com*

ABSTRACT
Machine learning (ML) techniques such as (deep) artificial neural networks (DNN) are solving very successfully a plethora of tasks and provide new predictive models for complex physical, chemical, biological and social systems. However, in most cases this comes with the disadvantage of acting as a black box, rarely providing information about what made them arrive at a particular prediction. The present paper studies the uniqueness of individual gait patterns in clinical biomechanics using DNNs, we take a Multi-view Gait DNN (MvDNN) to generate fake gait samples to extend existing gait dataset. The development of deep learning has promoted cross-view gait recognition performances to a higher level. However, performances of existing deep learning based cross-view gait recognition methods are limited by lack of gait samples under different views. By measuring the time-resolved contribution of each input variable to the prediction of ML techniques such as DNNs, our method describes the first general framework that enables to understand and interpret non-linear ML methods in (biomechanical) gait analysis and thereby supplies a powerful tool for analysis, diagnosis and treatment of human gait. The experimental results on CASIA-B and OUMVLP dataset demonstrate that fake gait samples generated by the proposed MvDNN method can improve performances of existing state of-the-art cross-view gait recognition methods obviously on both single-dataset and cross-dataset evaluation settings.

INTRODUCTION
Gait recognition is a biometric method for recognizing persons using features extracted from their walking style. Different from many biometric modalities gait recognition has some remarkable characteristics: a gait feature, which has traits for discriminating individuals, can be acquired even from the unconscious gait of an uncooperative subject, at a good distance from the camera, and from relatively low image resolution (e.g., a person with 30-pixel height in an image). Gait recognition can therefore be usefully applied to surveillance or crime investigation using CCTV footage. However, because gait features of uncooperative subjects may contain covariates that influence the gait itself and/or the appearance of a walking person, robustness to such covariates is quite important for accurate gait recognition. Covariates include, but are not limited to views, walking speeds , clothing , and belongings . Among the covariates, a change in view occurs frequently in real situations and has a large im pact on the appearance of the walking person. Matching gait across different views (cross-view matching) is therefore one of the most challenging and important tasks in gait recognition. Two different families of approaches for gait recognition have been proposed: appearance-based and model-based methods. Appearance-based approaches use captured image sequences directly to extract gait features, while model-based methods extract model parameters from the images.

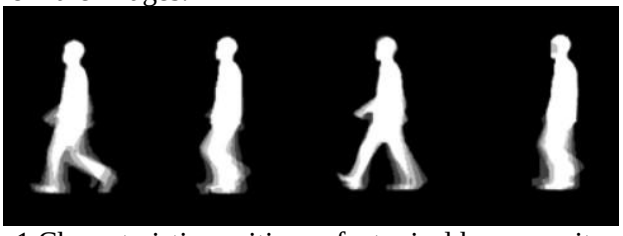

Fig. 1.Characteristic positions of a typical human gait cycle

In particular, 3D model-based approaches are preferable for cross-view matching owing to their view-invariant nature. A gait cycle represents the fundamental temporal unit of processing in gait recognition, and corresponds to a periodic cycle that transits from Rest to Right-Foot-Forward (RFF) to Rest to Left-Foot-Forward (LFF) to Rest position [1]. This basically encompasses the entire range of possible positions that a human body passes in the overall course of walking. Fig. 1.shows the five characteristic positions of a typical human gait cycle. As the subject walks across the Field-of-View (FoV) of the camera, multiple gait cycles are usually captured depending upon the FoV of the camera and the gait dynamics of the individual. The gait cycle, being the basic unit in gait-based image processing, contains information about dynamic motion and relative motion among all the body parts as the individual moves. In other words, the dynamics and periodicity of the gait cycle characterizes the motion of the individual, along with the static features like height, width etc. The gait cycles are repetitive in nature and as the number of acquired gait cycles increases, there is a consequent increase in information redundancy.

LITERATURE SURVEY

Xin Chen[1] et al proposed a generating fake gait samples under different views across different gait datasets based on Multi-view Gait Generative Adversarial Network (MvGGAN). MvGGAN includes a single generator which generates fake gait samples under several walking conditions and a discriminator which realizes adversarial training and preserves person identity information. By adding generated fake gait samples to the original gait datasets and performing domain alignment between real and fake samples, the performances of deep learning based gait classification networks can be improved obviously.

Sudeep Sarkar, P. Jonathon Phillips [2] researched about Identification of people by analysis of gait patterns extracted from video has recently become a popular research problem. To provide a means for measuring progress and characterizing the properties of gait recognition, we introduce the HumanID Gait Challenge Problem. The challenge problem consists of a baseline algorithm, a set of twelve experiments, and a large data set. The baseline algorithm estimates silhouettes by background subtraction, and performs recognition by temporal correlation of silhouettes.

John N. Carter, Mark S. Nixon [3] designed a new method for viewpoint independent gait biometrics. The system relies on a single camera, does not require camera calibration, and works with a wide range of camera views. This is achieved by a formulation where the gait is self-calibrating. These properties make the proposed method particularly suitable for identification by gait, where the advantages of completely unobtrusiveness, remoteness, and covertness of the biometric system preclude the availability of camera information and specific walking directions.

Qiang Wu, Jian Zhang [4] recognized that gait is an important biometric feature to identify a person at a distance, e.g., in video surveillance application. In the learning processes, sparse regression based on the elastic net is adopted as the regression function, which is free from the problem of overfitting and results in more stable regression models for VTM construction. Based on widely adopted gait database, experimental results show that the proposed method significantly improves upon existing VTM- based methods and outperforms most other baseline methods reported in the literature.

Haruyuki Iwama, Mayu Okumura [5] designed The “OU-ISIR Gait Database, Large Population Dataset”—and its application to a statistically reliable performance evaluation of vision-based gait recognition. Whereas existing gait databases include at most 185 subjects, we construct a larger gait database that includes 4007 subjects (2135 males and 1872 females) with ages ranging from 1 to 94 years.

D. Muramatsu, A. Shiraishi, Y. Makihara, and Y. Yagi [6] discussed about that Camera-based gait recognition is a useful method for authenticating a person from a distance, even if the resolution of the acquired images is not high. However, different views of the compared gallery and probe decrease the recognition accuracy. To solve this problem, we propose a gait based authentication method that uses an arbitrary view transformation scheme. The proposed method constructs a transformation matrix associated with the view of the set of gallery and probe using a 3D gait database composed of non-target multiple subjects' visual hulls.

Y. Makihara and Y. Yagi [7] proposed an iterative scheme of spatio-temporal local color transformation of background and graph-cut segmentation for silhouette extraction. Given an initial background subtraction, spatio-temporal background color transformation is processed for fitting modeled background colors to input background ones under a different illumination condition. An approach, named SM-VTM (Systematic Mapping based on Visual Text Mining), to support categorization and classification stages in the systematic mapping using Visual Text Mining (VTM), aiming at reducing time and effort required in this process.

I. PROPOSED SYSTEM

We develop a model for gait classification using deep learning methods. Separate gait databases are used in our experiment. The gait signatures are extracted from gait energy image (GEI) using convolutional neural network (CNN). The classifiers we used for classification of human gait is MobileNET as shown in Fig. These models are tested on standard gait dataset CASIAA, OUISIR database. The experimental results will be collected on both the datasets.

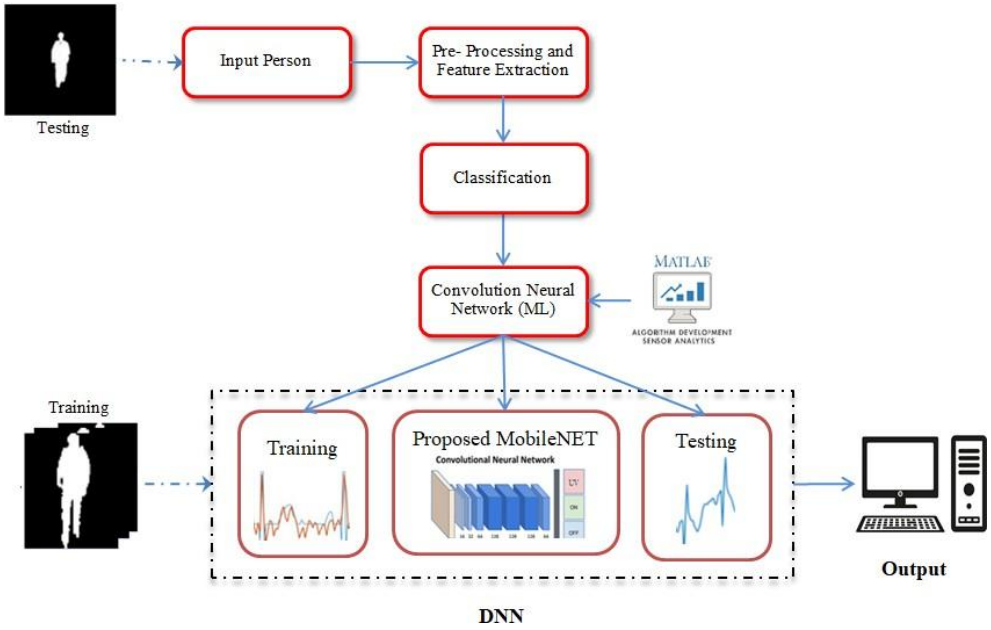

Fig.2 Proposed Architecture

1. Methodology

The experiment overview is shown in Figure 2. The major steps involved consisted of data collection, data acquisition, classifier design, training and testing. The CNN was implemented using MATLAB R2021a. Training was performed on a laptop with a Graphics Processing Unit (GPU) with 2 GB of memory.

2. Data Collection

The data were obtained. It contains three-dimensional skeletal joint points of 23 subjects recorded by a depth camera. Samples were collected under indoor environment where each subject was required to wear standardized outfits. Ten walking sequences per subject was captured in oblique view, where the subjects were required to walk in a 45˚ angle from the upper side of the camera. There were ten walking sequence that consisted of 15,600 recorded joint points for each sample.

3. Data Acquisition

The gait data of each walking sequence was set to the corresponding subjects and transposed to 60 by 26 three-dimensional arrays to fit the input requirement of the CNN. The column represents all the joint points at a time frame, while the row represents the joint data across time. Once the input was reshaped,

the gait data was then split and distributed into training and testing data sets in 50:50 ratio. The input was distributed randomly to improve generalization.

4. Classifier -MobileNET Architecture

In this section we first describe the core layers that MobileNET is built on which are depthwise separable filters. We then describe the MobileNET network structure and conclude with descriptions of the two model shrinking hyper parameters width multiplier and resolution multiplier.

The MobileNET model is based on depthwise separable convolutions which is a form of factorized convolutions which factorize a standard convolution into a depthwise convolution and a 1×1 convolution called a pointwise convolution. For MobileNET the depthwise convolution applies a single filter to each input channel. The pointwise convolution then applies a 1×1 convolution to combine the outputs the depthwise convolution. A standard convolution both filters and combines inputs into a new set of outputs in one step. The depthwise separable convolution splits this into two layers, a separate layer for filtering and a separate layer for combining. This factorization has the effect of drastically reducing computation and model size. Figure 3 shows how a standard convolution 2(a) is factorized into a depthwise convolution 2(b) and a 1 × 1 pointwise convolution 2(c).

A standard convolutional layer takes as input a DF × DF × M feature map F and produces a DF × DF × N feature map G where DF is the spatial width and height of a square input feature map1 , M is the number of input channels (input depth), DG is the spatial width and height of a square output feature map and N is the number of output channel (output depth). The standard convolutional layer is parameterized by convolution kernel K of size DK ×DK ×M×N where DK is the spatial dimension of the kernel assumed to be square and M is number of input channels and N is the number of output channels as defined previously. The output feature map for standard convolution assuming strides one and padding is computed as:

$$\mathbf{G}_{k,l,n} = \sum_{i,j,m} \mathbf{K}_{i,j,m,n} \cdot \mathbf{F}_{k+i-1,l+j-1,m}$$

Standard convolutions have the computational cost of:

$$D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F$$

Where the computational cost depends multiplicatively on the number of input channels M, thenumber of output channels N the kernel size Dk × Dk and the feature map size DF × DF .

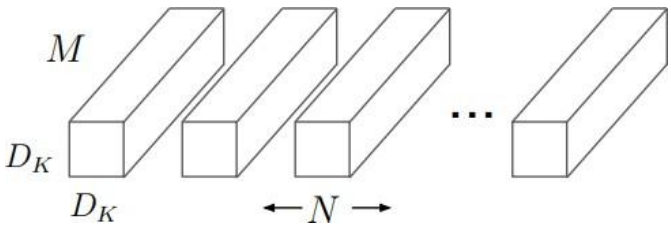

Figure Standard Convolution Filters

The MobileNet structure is built on depthwise separable convolutions as mentioned in the previous section except for the first layer which is a full convolution. By defining the network in such simple terms we are able to easily explore network topologies to find a good network. The MobileNet architecture is defined in Table 1. All layers are followed by a batchnorm [13] and ReLU nonlinearity with the exception of the final fully connected layer which has no nonlinearity and feeds into a softmax layer for classification. Figure 3 contrasts a layer with regular convolutions, batchnorm and ReLU nonlinearity to the factorized layer with depthwise convolution, 1 × 1 pointwise convolution as well as batchnorm and ReLU after each convolutional layer. Down sampling is handled with strided convolution in the

depthwise convolutions as well as in the first layer. A final average pooling reduces the spatial resolution to 1 before the fully connected layer. Counting depthwise and pointwise convolutions as separate layers, MobileNet has 28 layers.

It is not enough to simply define networks in terms of a small number of Mult-Adds. It is also important to make sure these operations can be efficiently implementable. For instance unstructured sparse matrix operations are not typically faster than dense matrix operations until a very high level of sparsity. Our model structure puts nearly all of the computation into dense 1 × 1 convolutions. This can be implemented with highly optimized general matrix multiply (GEMM) functions. Often convolutions are implemented by a GEMM but require an initial reordering in memory called im2col in order to map it to a GEMM. For instance, this approach is used in the popular Caffe package [15]. 1×1 convolutions do not require this reordering in memory and can be implemented directly with GEMM which is one of the most optimized numerical linear algebra algorithms. MobileNet spends 95% of its computation time in 1 × 1 convolutions which also has 75% of the parameters as can be seen in Table 2. Nearly all of the additional parameters are in the fully connected layer.

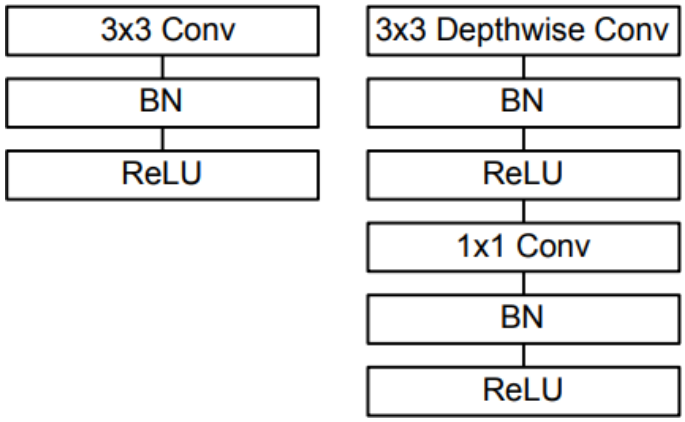

Figure 3 Left: Standard convolutional layer with batchnorm and ReLU. Right: Depthwise Separable convolutions with Depthwise and Pointwise layers followed by batchnorm and ReLU

However, contrary to training large models we use less regularization and data augmentation techniques because small models have less trouble with over fitting. When training MobileNets we do not use side heads or label smoothing and additionally reduce the amount image of distortions by limiting the size of small crops that are used in large Inception training. Additionally, we found that it was important to put very little or no weight decay (l2 regularization) on the depthwise filters since their are so few parameters in them.

RESULTS & DISCUSSION

In order to extract walking characteristics of a person for classification and forthcoming identification, a complete gait cycle is analyzed and a sequence of images are produced. Subsequently dataset is created in form of binary image frames. This dataset is stored for oversight situation wherever there is no previous information about the object. Furthermore, indoor silhouettes from the identifiable video clips and the outdoor silhouettes from the three different sets of open gait CASIA database are used37. Video clips of individuals are captured from different viewing angles and each clip is divided into 25 frame per second. All image frames (more than 16,821 images) for the participants are divided into 2 disjoint sets; the first set for training, the second set for validation and testing. In the proposed experiment, a subset of 80% of the data is presented to the network during the training and the network is adjusted according to its error. A subset of 20% of the data is used for the network validation and for providing an independent measurement and testing of the MobileNet performance. It is considered that the data for the same participant does not exist in both of the training and validation and testing sets.

To evaluate the proposed method, the analysis in MATLAB R2021a on a 64-bit Windows PC with Intel®2.8 GHz x-64 based processor and 16 GB RAM is conducted. As Fig. Demonstrated the smoothed curves of training and validation accuracy percentage.

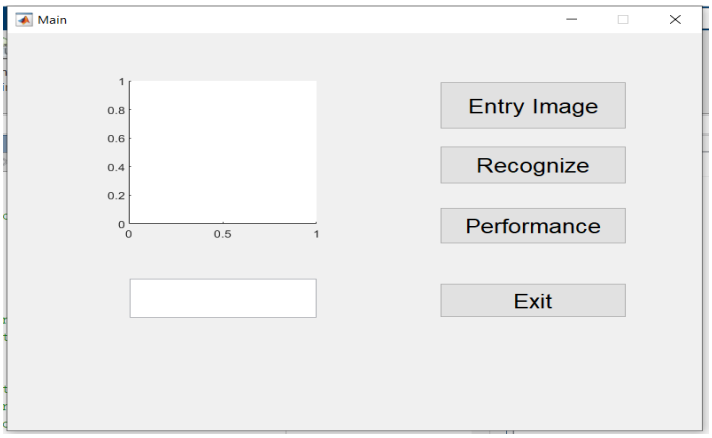

Figure.4: GUI design

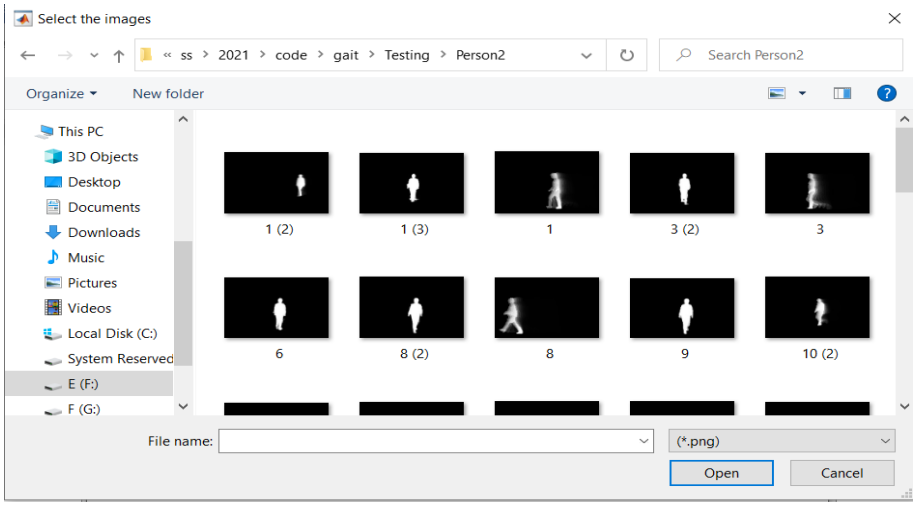

Figure 5 : Trained images

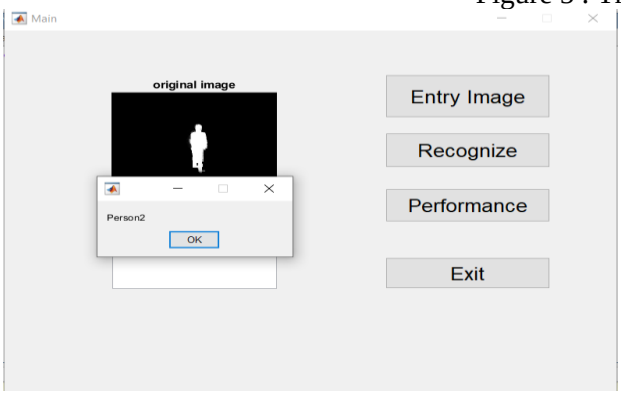

Figure 6: Recognition

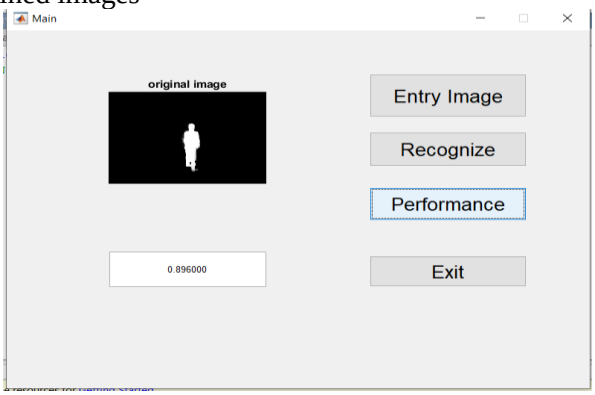

Figure 7: Performance analysis

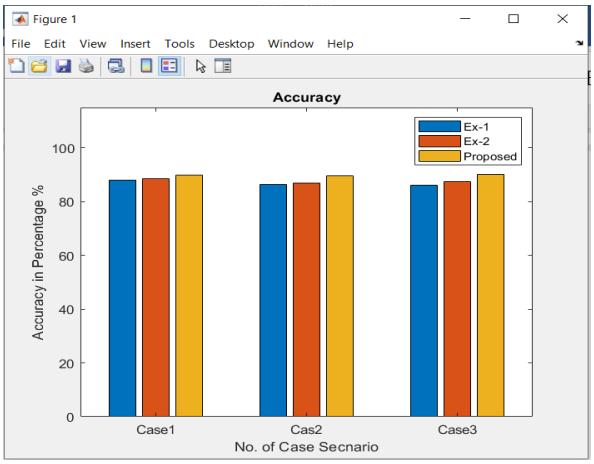

Figure 8: Accuracy analysis

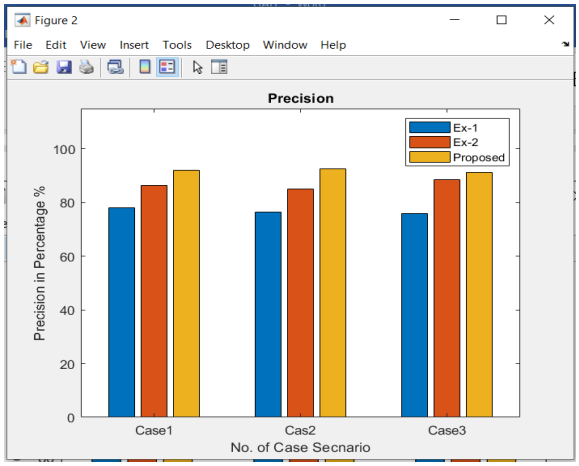

Figure 9: Precision analysis

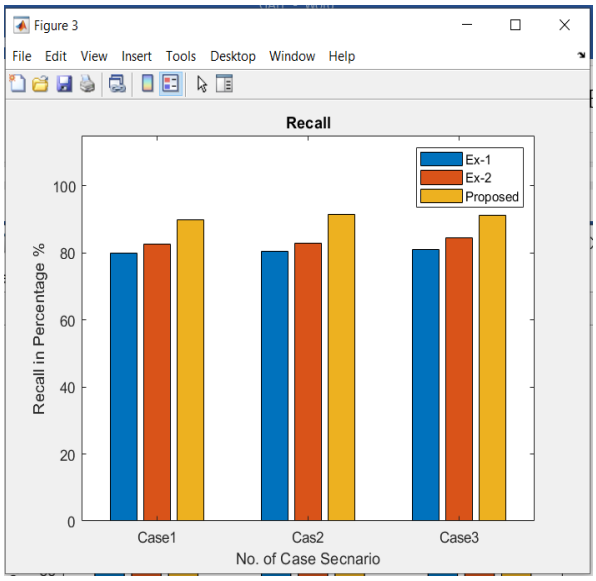

Figure 10: Recall rate

CONCLUSION

The proposed gait recognition approach achieves impressive results in terms of training/validation accuracy and mean square errors. The conducted experimental outcomes report competitive performance as compared to many traditional machine learning methods and previous deep gait models specifically for the case of low-image resolutions and large-scale dataset of input images. It can be concluded that gait analysis is not only to increase the level of security and safety in the community level but also to classify age and gender. In the recent future, gait will be deployed for human being recognition in conjunction with other biometrics and in many other applications. The proposed method compared to a broader variation view and in conjunction with a new appropriate database will be confirmed.

REFERENCES

1. Chen, Xin, Xizhao Luo, Jian Weng, Weiqi Luo, Huiting Li, and Qi Tian. "Multi-view gait image generation for cross-view gait recognition." IEEE Transactions on Image Processing 30 (2021): 3041-3055.
2. Sarkar, Sudeep, P. Jonathon Phillips, Zongyi Liu, Isidro Robledo Vega, Patrick Grother, and Kevin W. Bowyer. "The humanid gait challenge problem: Data sets, performance, and analysis." IEEE transactions on pattern analysis and machine intelligence 27, no. 2 (2005): 162-177.
3. Goffredo, Michela, Imed Bouchrika, John N. Carter, and Mark S. Nixon. "Self-calibrating view-invariant gait biometrics." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 40, no. 4 (2009): 997-1008.
4. Kusakunniran, Worapan, Qiang Wu, Jian Zhang, and Hongdong Li. "Gait recognition under various viewing angles based on correlated motion regression." IEEE transactions on circuits and systems for video technology 22, no. 6 (2012): 966-980.
5. Iwama, Haruyuki, Mayu Okumura, Yasushi Makihara, and Yasushi Yagi. "The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition." IEEE Transactions on Information Forensics and Security 7, no. 5 (2012): 1511-1521.
6. Muramatsu, Daigo, Akira Shiraishi, Yasushi Makihara, and Yasushi Yagi. "Arbitrary view transformation model for gait person authentication." In 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 85-90. IEEE, 2012.
7. Makihara, Yasushi, and Yasushi Yagi. "Silhouette extraction based on iterative spatio-temporal local color transformation and graph-cut segmentation." In 2008 19th International Conference on Pattern Recognition, pp. 1-4. IEEE, 2008.
8. Lu, Jiwen, and Yap-Peng Tan. "Uncorrelated discriminant simplex analysis for view-invariant gait signal computing." Pattern Recognition Letters 31, no. 5 (2010): 382-393.
9. Liu, Nini, Jiwen Lu, and Yap-Peng Tan. "Joint subspace learning for view-invariant gait recognition." IEEE Signal Processing Letters 18, no. 7 (2011): 431-434.
10. J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Target dependent score normalization techniques and their application to signature verification," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 35, no. 3, pp. 418–425, Aug. 2005.
11. S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanID gait challenge problem: Data sets, performance, and analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 2, pp. 162–177, Feb. 2005.
12. M. Goffredo, I. Bouchrika, J. N. Carter, and M. S. Nixon, "Selfcalibrating view-invariant gait biometrics," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 40, no. 4, pp. 997–1008, Aug. 2010.
13. W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Gait recognition under various viewing angles based on correlated motion regression," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 6, pp. 966–980, Jun. 2012.
14. H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, "The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition," IEEE Trans. Inf. Forensics Security, vol. 7, no. 5, pp. 1511–1521, Oct. 2012.
15. D. Muramatsu, A. Shiraishi, Y. Makihara, and Y. Yagi, "Arbitrary view transformation model for gait person authentication," in Proc. IEEE 5th Int. Conf. Biometrics, Theory, Appl. Syst., Sep. 2012, pp. 85–90.