VGG-16 based Gait Recognition using Skeleton Features

Irshad Ali
Department of ECE
National Institute of Technology
Rourkela, India 769008
aliirshad845@gmail.com

Iman Junaid
Department of ECE
National Institute of Technology
Rourkela, India 769008
imanjunaid71@gmail.com

Samit Ari
Department of ECE
National Institute of Technology
Rourkela, India 769008
samit.ari@gmail.com

***Abstract*—Gait recognition is a new technology that can identify people with various walking patterns. Gait of a person is the manner in which they walk. A person's walking stride is distinctive due to their body movements. Due to the popularity of the Kinect, human gait can be recognised to using 3D skeletal information. In this paper, Skeleton Gait Energy(SkeGEI) is image-based approach for gait recognition and VGG-16 with multilayer perceptron is proposed to effectively exploit raw depth information collected by the Kinect sensor. To restore as much gait data as possible, fine-tuned VGG-16 is used to extract space - time deep feature data from SkeGEI. Multilayer perceptron is then used to ascertain the connection between the corresponding subject and the gait features. Softmax is used for classification. Experiments on three different datasets show that our technique outclasses the majority of gait recognition methods.**

***Index Terms*—Gait Recongnition, VGG-16, SkeGEI, Multilayer perceptron**

I. Introduction

Gait recognition is the method of recognising person from the way they move. The important element that makes gait detection easier is the way a person's body changes when they are walking. Gait recognition issues are resistant to distantly captured, low-resolution photos. Additionally, gait is challenging to mimic because everyone's walking gait is unique. However, the participation of changes, such as viewing angular position, carrying capacity and apparel, can impact gait recognition process. Handcrafted solutions was suggested in the beginning to address the aforementioned problem. Model free methods and model based method and are two types of handcrafted methods. Model-based methodologies for understanding gait characteristics use a walking of human skeletal design. Model-free methods, on other hand, use gait silhouettes of human to extract features. To recognise gait features, model based methods use a walking human skeletal framework. Deep learning techniques are widely used by researchers because they can teach themselves the most discriminant information. Deep learning models retrieve huge amount of difficult features, which affects performance significantly. The usage of pre trained CNN models had attracted attention among deep learning. This is because a pre trained prototype was started learning from a huge database in order to obtain important attributes. AlexNet [3], DenseNet [2], ResNet [1], and VGG [4] are the most commonly used pre trained designs for human recognition. Fine tuning and transfer learning methodologies are used to integrate pre trained designs, greatly improving accuracy. As a result, this paper has proposed pre trained design and multi-layer perceptron (MLP) to boost accuracy under a wide range of conditions. The SkeGEI is calculated by first averaging the image frame by one gait cycle. The obtained SkeGEI's gait features are then retrieved using a fine tuned VGG-16 design. To determine, connection between features and subjects, a multilayer perceptron is used. At last, subjects are categorised using the classification layer. Table I contains a literature survey of works related to various approaches.

II. Methodology

Figure 1 depicts the process flowchart for the proposed algorithm. By transforming human gait skeleton coordinates into skeletal gray-scale frames, we can retrieve skeletal images of the human gait cycle and then find SkeGEI of the gait patterns. SkeGEI is fed into the VGG-16 model, which has been fine tuned. VGG-16 design generates extracted features which it replicate the SkeGEI's high- and low-level representations. Following that, MLP learns the connections between both the extracted feature as well as the different class. Finally, Class probability distributions value are returned to the classification stage. Final class label represents the most probable classes. Overall method was tested and trained using three datasets.

A. Pre-processing

Get binary skeleton gait image: Scale the foot and head skeleton coordinates to ensure that all skeletal are within the same height. Given that the neck is one of the most stationary joints during walking, location of the neck on the image is fixed at a stationary point and taken as origin, after which other joints are converted to corresponding coordinates. Lastly, on the canvas, draw the skeleton joints and limbs to create a binary skeleton gait image.

TABLE I
LITERATURE SURVEY

Approaches	Authors	Work Done
Model-based approaches	Kastaniotis *et al.* [5]	It is proposed that a structure initially designed for human activity recognition be modified and applied to human walking recognition. The new proposal enables us to the create sophisticated representations of walking patterns, allowing us to express the dynamic feature of walking more effectively.
	Fernandez *et al.* [6]	Method based on graph neural networks is introduced. The body joint coordinates and adjacent matrix indicating the skeleton joints are fed into the architecture. In order to generate a rectified output of the input feature, a residual relation is also used.
	Chen *et al.* [7]	It is proposed that each participant be segmented using hypergraph partitioning and recognised using a multi-linear canonical correlation analysis algorithm (UMCCA).
Model free approaches	Raida *et al.* [8]	To decrease intra-class variation and enhance recognition accuracy, the most discriminative human part of the body was chosen based on group Lasso of motion.
	Nandy *et al.* [9]	It was proposed to use a different statistical structure analysis methodology based on Gait Energy Image (GEI) decomposition, that has broken down into 3 separate structure categories: grid resolution, horizontal and vertical.
Deep learning	Wang *et al.* [10]	CNN-based method for integrating various GEI angles into a single network.
	Yu *et al.* [11]	CNN-LSTM networks were developed to retrieve space - time deep feature data by DA and SkeGEIs characteristics.

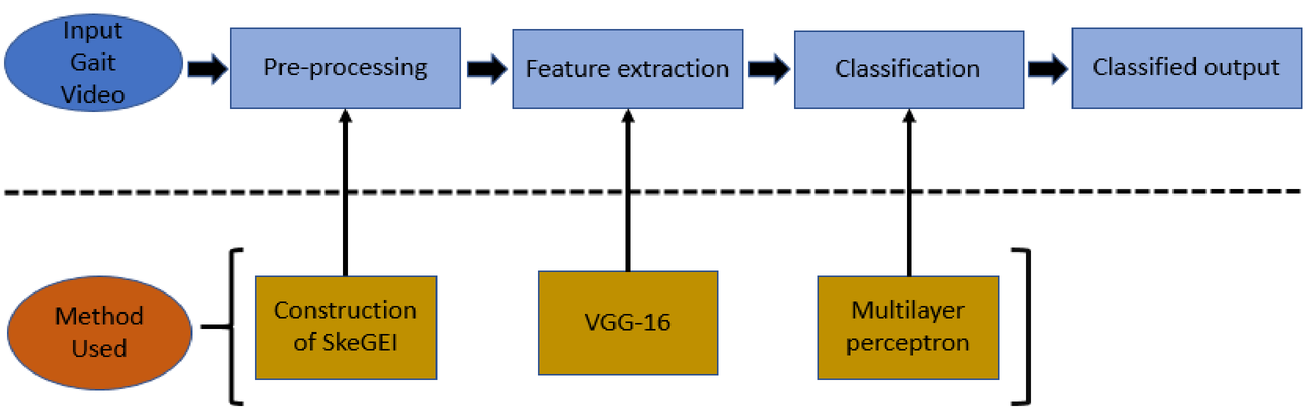

Fig. 1. Overview of praposed methodology

Extraction of the gait cycle: The gait cycle must be accumulated in order to reject noise frame and ensure that skeGEI's include the entire human walking procedure. The gait cycle is determined in this paper by the Euclidean length between both legs as in longitudinal perspective. Taking into account the length walked for every gait cycle window:

$$l_{wd} = \sqrt{(m_r - m_l)^2 + (n_r - n_l)^2} \tag{1}$$

(m_r, n_r) are the location point of the right leg joint in the vertical view, and (m_l, n_l) are the location point of the left leg joints. For every highest value seems to be when spacing between foot is highest, and for every minimum value seems to be when foot intersect vertically. The frame number differences between every two consecutive odd maximum frames are used to estimate the average differences as human gait cycles in video frame.

Obtain SkeGEI: GEIs [12] is an a kind of average sum technique that uses a gray-scale image of human gait frame to display the energy changes that occur in the body of human during one full human gait cycle. SkeGEI's is calculated using GEI from skeletal frame of human walking cycle in order to convert human gait skeleton coordinates into skeletal gray - scale frames. This is how the SkeGEI is described:

$$F(m,n) = \frac{1}{X}\sum_{r=1}^{X} A_r(m,n) \tag{2}$$

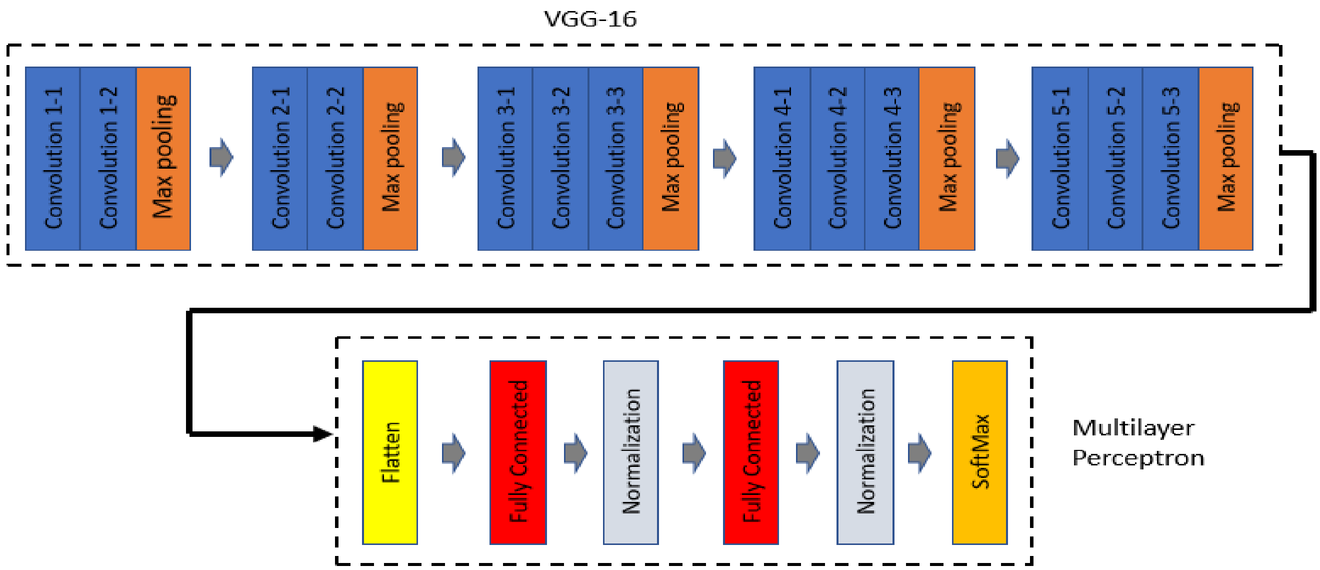

Fig. 2. The proposed VGG16-MLP design's architecture.

for which m and n are pixel locations, X is complete number of images in the human gait cycles, r is sequence frame count number, and $A_r(m, n)$ is binary skeleton gait image in a single frame.

B. Feature extraction

We have used VGG-16 design to extract features. Figure 2 depicts the proposed network's design. The network is built with a finely tuned pre-trained VGG-16 design and an MLP. The deep gait characteristics have been retrieved using a fine-tuned pre-trained VGG-16 design. MLP is used in the network to express the connection between leaned features extracted and related classes in greater detail.

Fine Tunining of VGG-16: Transfer learning refers to the adaptation of acquired knowledge from a pre-trained prototype from one issue to another. Due to the fact that the design had been pre-trained on a huge dataset, transfer learning is an important deep learning technique because it requires significantly fewer computer complexity. By transmitting the learned features of pre-trained designs, the method is typically used to solve downstream work. To make sure that the pre-trained design capable of adapting to downstream work effectively, It is necessary to fine-tune. The VGG-16 design is an improvement of AlexNet [3] prototype that includes additional convolution operations and the lowest kernel size window. The VGG-16 design consists of thirteen convolution stages, five max pooling layer, and two full connected layers divided into five convolutional and max-pooling layer sets. Two convolution layers are next followed by a max pooling layer in the initial two components. The following three sets are made up of three convolution layers and max pooling layer. The design differs from the other pre-trained designs in that it uses kernel size window of 3×3 and we have used padding one and stride one. Using the lowest kernel size window significantly decreases the number of parameters required. Using smallest kernel size window also kept system from ever become overfit. Stride two employed to apply Max pooling to a 2×2 pixel window. As a result, the extracted features' dimension is cut in half. Rectified linear unit function was applied to all convolutional layers since it is more computation accurate and diminishes the gradient vanishing error.

C. Classification

We used a multilayer perceptron for classification. The map of the final max pooling layer is vectorized by flattened and supplied to MLP. MLP consists of fully - connected layers, two batch normalisation layers, and one classifier layer. Each fully-connected layer has five hundred twelve neurons. The link between the collected feature and class labeling is formed in the full connected layer. Leaky rectified linear unit function is used in both layers that are full connected. Leaky rectified linear unit function improves on the rectified linear unit function which can represent both positive and negative values. As a result, layer is more optimised and balanced, which leads to a faster training program. Activation function of a leaky ReLU is as follows:

$$g(m) = \left\{ \begin{array}{ll} m & m > 0 \\ \alpha m & m \leq 0 \end{array} \right\} \quad (3)$$

When m is a negative number, it must be multiplied by α. Following the addition of both fully connected layers, batch normalisation layers are added. The mean and standard deviation of each mini batch are subtracted and divided during batch normalisation. Normalisation makes sure that the training is performed effectively. Because the human gait recognition problem involves several classes, the Soft-max activation function is employed for classifying. Soft-max activation function returns likelihood that every input belongs to a particular category. The following is how the Soft-max activation function is determined:

$$P(m_j) = \frac{\exp(m_j)}{\sum_{i=1}^{c} \exp(m_i)} \tag{4}$$

where m_j represents the Softmax funtion for class j and c is number of class in classifications. During training, Adam optimizer used to boost network concurrence. Avoid over-training network, the model employs early stopping mechanism. When performance no longer enhances, the early-stopping mechanism is crucial for preventing the training process. Because gait recognition requires multi class categorization for classes, the VGG-16 design employs logarithmic loss function. Following is logarithmic loss function for categorical:

$$loss(logarithmic) = -\sum_{j=1}^{c} \hat{m_j} \, . \log m_j \tag{5}$$

for which $\hat{m_j}$ represents the true class classification, m_j represents the Soft-max activation function for class j, and c represents the class count.

III. EXPERIMENTS AND RESULTS

A. Preparation of Data

Kinect based human gait datasets are used to assess the capability of proposed approach.

Dataset from Kinect Sensor Gait Biometry [13]: It's a single direction gait dataset made up of walking patterns recorded by the Kinect V1 sensor on 160 different subjects. For every participant was instructed to walk five times in a semicircular path ahead of Kinect sensor, and Kinect sensor could rotate to follow the subject in order to keep the particular subject as in central point.

Gait data from SDU [14]: This dataset includes 1040 walking patterns recorded by 2 Kinect V2 sensor systems from 52 subjects. Each participant learned 20 different walking sequences from six fixed directions (0°, 90°, 135°, 180°, 225°, and 270°) and two random directions.

CILgait dataset [15]: The Kinect V2 sensor was used to collect walking sequences from 12 subjects for this dataset. Each subject had 22 patterns recorded and 16 are sequences with directions (0°, 45°, 90°, 135°, 180°, 225°, 270° and 315°), while remaining are unusual patterns.

B. Hyperparameter Tuning

Tuning hyperparameters is crucial for optimising deep learning model performance. The tuning takes into account four hyperparameters shown below. Table II shows optimal hyperparameter values for the proposed method.

C. In comparison to Previous Human Gait Recognition Work

Inside this segment, we make comparisons the recognition accuracy of our approach to state-of-the-art methodologies, including skeleton-based and other types. In three different datasets, our method produced the same or superior results.

TABLE II
THE PROPOSED VGG-16AND MLP MODEL'S PARAMETER TUNING SUMMARY AND IT'S VALUE

Hyperparameters	Value
Size of batch	32
Value of dropout	0.29
Rate of learning	0.0001
Optimizer	Adam

(1) Dataset from Kinect Sensor Gait Biometry : Table III compares our approach's accuracy to that of other approaches in experiments upon Kinect Gait Biometry Dataset [16]. When compared to method [11], which uses fine-tuning VGG-16 to extract pattern characteristics from gait data, our method outperforms it by 1.73%.

TABLE III
THE PROTOCOL WAS EMPLOYED TO EVALUATE THE ACCURACY OF EXISTING APPROACHES. [17](%)

Appaorch	Dataset for Kinect Gait Biometry
Nirattaya [17]	97.00
And and Ar [13]	87.7
Ke and Yan [18]	95.40
Jie and Li [16]	96.56
Yu Liu [11]	97.39
Ours	99.12

(2) SDU gait dataset: Figure 3 compares accuracy on SDU gait dataset to existing approaches that use [15] protocol. SDU-arb datasets, our approach's accuracy is 4.89% higher than the skeleton-based approach's highest accuracy.

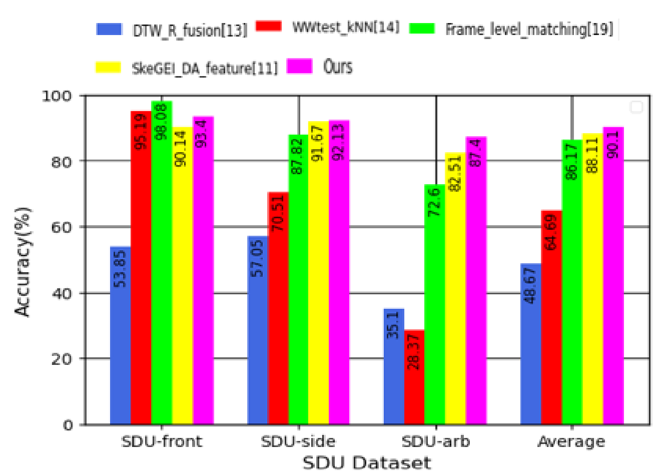

Fig. 3. The protocol was used to compare the accuracy of existing techniques on the SDU dataset [15].

(3) Gait dataset of CIL : Figure 4 depicts the recognition accuracy on CIL gait dataset, as well as existing methods is based on the [15] protocol. On the datasets CIL-SC and CIL-S, our approach outperforms highest performance of SkeGEI and DA feature by 1.29% and 4.64%, respectively.

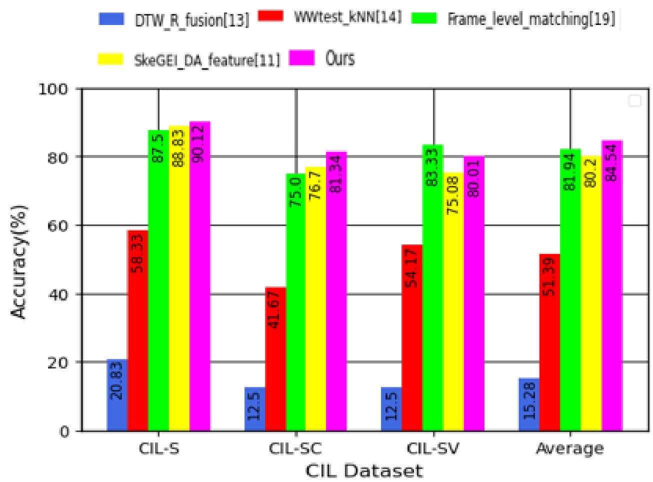

Fig. 4. The protocol [15] was used to compare the performance of previous approaches on the CIL dataset.

IV. CONCLUSION

In gait recognition process, parameters like viewing angle present significant challenges. The work presented, incorporates the fine tuned VGG-16 design and MLP for classification. To retrieve deep gait characteristics of SkeGEIs, the VGG-16 network is used. The MLP's fully connected layers, batch normalisation, and classifier described the relationship between map of feature and the corresponding class. So when walking style is varied, the results show that the proposed method performed admirably on huge datasets. Aside from that, combining VGG-16 and MLP can reduce the effect of noise, improving the performance of proposed approach.

ACKNOWLEDGMENT

This work was supported by Science and Engineering Research Board (grant no. CRG/2019/001414).

REFERENCES

[1] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun., "Deep residual learning for image recognition." *IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.

[2] Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. "Densely connected convolutional networks." *IEEE conference on computer vision and pattern recognition*, pp. 4700-4708, 2017.

[3] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Communications of the ACM*, vol. 60, no. 2, pp. 84-90, 2017

[4] Vedaldi, Andrea, and Andrew Zisserman. "Vgg convolutional neural networks practical." *Department of Engineering Science, University of Oxford* vol. 66, 2016.

[5] Kastaniotis, Dimitris, Ilias Theodorakopoulos, Christos Theoharatos, George Economou, and Spiros Fotopoulos. "A framework for gait-based recognition using Kinect." *Pattern Recognition Letters*,vol. 68, pp. 327-335, 2015.

[6] Lopez-Fernandez, David, Francisco Jose Madrid-Cuevas, A. Carmona-Poyato, Rafael Munoz-Salinas, and R. Medina-Carnicer. "A new approach for multi-view gait recognition on unconstrained paths." *Journal of Visual Communication and Image Representation*,vol. 38, pp. 396-406, 2016.

[7] Chen, Xin, Jiaming Xu, and Jian Weng. "Multi-gait recognition using hypergraph partition." *Machine Vision and Applications*, vol. 28, pp. 117-127, 2017.

[8] Rida, Imad, Xudong Jiang, and Gian Luca Marcialis. "Human body part selection by group lasso of motion for model-free gait recognition." *IEEE Signal Processing*,vol. 23, no. 1, pp. 154-158, 2015.

[9] Nandy, Anup, Rupak Chakraborty, and Pavan Chakraborty. "Cloth invariant gait recognition using pooled segmented statistical features." *Neurocomputing*, vol. 191, pp. 117-140, 2016.

[10] Wu, Zifeng, Yongzhen Huang, Liang Wang, Xiaogang Wang, and Tieniu Tan. "A comprehensive study on cross-view gait based human identification with deep cnns." *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 2, pp. 209-226, 2016.

[11] Liu, Yu, Xinghao, Tanfeng, and Ke "3D gait recognition based on a CNN-LSTM network with the fusion of SkeGEI and DA features." *IEEE International Conference* , pp. 1-8, 2019.

[12] Guru, VG Manjunatha, V. N. Kamalesh, and R. Dinesh. "Human gait recognition using four directional variations of gradient gait energy image." *Computing, Communication, and Automation International Conference*, pp. 1367-1371, 2016.

[13] Bari, ASM Hossain, and Marina L. Gavrilova. "KinectGaitNet: Kinect-based gait recognition using deep convolutional neural network." Sensors,vol. 22, no. 7, pp. 26-31, 2022.

[14] Sun, Jiande, Yufei Wang, Jing Li, Wenbo Wan, De Cheng, and Huaxiang Zhang. "View-invariant gait recognition based on kinect skeleton feature." *Multimedia Tools and Applications*,vol. 77,pp. 24909-24935, 2018.

[15] Choi, Seokeon, Jonghee Kim, Wonjun Kim, and Changick Kim. "Skeleton-based gait recognition via robust frame-level matching." *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2577-2592, 2019.

[16] Li, Jie, Lin Qi, Aite Zhao, Xingnan Chen, and Junyu Dong. "A skeleton based gait recognition network with a dynamic long short term memory network." *IEEE, SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI*, pp. 1-6, 2017.

[17] Khamsemanan, Nirattaya, Cholwich Nattee, and Nitchan Jianwattanapaisarn. "Human identification from freestyle walks using posture based gait feature." *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 1, pp. 119-128, 2017.

[18] Yang, Ke, Yong, Shaohe, Fei Zhang, and Qi. "Kinect's relative distance features for gait recognition, " *Image Representation and Visual Communication Journal*, vol. 39, pp. 209-217, 2016.

[19] Liao, Rijun, Chunshui Cao, Edel B. Garcia, Shiqi Yu, and Yongzhen, "A time-space network based on poses for gait recognition with carrying and clothing variability." *Biometric Recognition: CCBR*, vol. 12, pp. 474-483, 2017.