3D Gait Recognition Based on a CNN-LSTM Network with the Fusion of SkeGEI and DA Features

Yu Liu Xinghao Jiang* Tanfeng Sun Ke Xu
School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University
800 Dongchuan RD. Minhang District, Shanghai, China
{moyeliuyu, xhjiang, tfsun, 113025816}@sjtu.edu.cn

Abstract

Gait recognition is a promising technology in biometrics in video surveillance applications for its characteristics of non-contact and uniqueness. With the popularization of the Kinect sensor, human gait can be recognized based on the 3D skeletal information. For exploiting raw depth data captured by Kinect device effectively, a novel gait recognition approach based on Skeleton Gait Energy Image (SkeGEI) and Relative Distance and Angle (DA) features fusion is proposed. They are fused in backward to complement each other for gait recognition. In order to maintain as much gait information as possible, a CNN-LSTM network is designed to extract the temporal-spatial deep feature information from SkeGEI and DA features. The experiments evaluated on three datasets show that our approach performs superior to most gait recognition approaches with multi-directional and abnormal patterns.

1. Introduction

Gait recognition is a kind of emerging technology, which can recognize people with different walking patterns. Compared with other biometric technologies, it has the unique advantages of non-contact and non-invasive. Gait can be identified from a long distance without the detected person knowing. As a behavioural feature, gait is obtained through human learning, including physical conditions and exercise habits, which is difficult to change and difficult to be imitated by others. It is very suitable for anti-terrorism and intelligent security. Therefore, the study of gait recognition is of great significance.

Gait recognition technology can be classified into two types: model-free methods and model-based methods [1]. The model-free based methods employ the silhouettes or partial silhouettes of bodies captured from human walking sequences, in which the gait energy map [2, 6] is the most widely used. Wu et al. [3] proposed a CNN-based method that integrates multiple angles of GEI into a unified network. However, GEI is sensitive to the variations of view angles, wardrobe and the carry-on baggage. Li et at. [4] propose the Gait Energy Response Function (GERF) to make the gait energy image (GEI) more suitable for handling conditions with clothing and carrying variations. The model-based method acquires a series of features by analyzing the posture of human limbs. Due to the emergence of depth cameras such as Kinect and wearable devices, the skeletal-based gait recognition method has gradually increased [5, 6, 11]. Most of the research focuses on the manual extraction of high-dimensional skeleton features for recognition [7, 8, 10], but many of them are identified on the self-built database, and the generalization performance is not good enough. Recently, neural networks (CNNs and LSTMs) achieved exciting success in other recognition tasks, such as face recognition and action recognition. Li et al. [14] proposed using a dynamic LSTM network to learn more descriptive sequence characteristics of the gait cycle directly from the raw motion data. However, 3D skeleton estimation could be hindered when human joints are occluded by clothes, belongings, or other parts of the body. These positional errors in human joint make gait recognition challenging.

In this paper, a novel approach is proposed based on the backward fusion strategy of SkeGEI and DA features. SkeGEI is proposed by the fusion of the advantages of GEI and skeleton. It retains not only the static information of skeleton but also postures change information. Furthermore, it is robust to single-frame skeleton joint detection error and direction change. Meanwhile, considering that SkeGEI loses part motion information such as speed and frequency, we propose DA feature to describe the variance in the relative distance and angle between pairs of joint points in the walking cycle. Finally, they are merged in backward in this paper by a CNN-LSTM network, which is designed for extracting the temporal-spatial deep information. In this CNN-LSTM network, spatial information is acquired from SkeGEI by CNN and temporal information is derived from DA feature by LSTM. Since SkeGEI and DA features contribute differently to gait recognition, the backward fusion of two kinds of features via full convolution layer in the CNN-LSTM network helps in raising recognition precision.

The main contributions of this paper are the proposal of SkeGEI feature and the backward fusion strategy of SkeGEI and DA features based on a CNN-LSTM feature exaction network. The proposed approach is comparable with multi-directional and abnormal patterns.

978-1-5386-9294-3/18/$31.00 ©2019 IEEE

2. Feature Modeling

In order to obtain spatial information and temporal information, two features are proposed in this paper as define in follows:

$$F = \{F_{SkeGEI}, F_{DA}\} \tag{1}$$

which are input into the network as two-channel. The skeleton spatial information and part of the temporal information would be extracted by SkeGEI, and DA feature would provide the rest of temporal information.

2.1. Extraction of skeletal gait energy image

GEI [9] is a kind of average summation algorithm based on the grayscale image of the gait frame in a complete walking cycle, and it shows the energy changes of the human body in this cycle. On the basis of GEI, we obtain skeletal gait energy image (SkeGEI) from skeletal images of a gait cycle by converting the human walking skeleton coordinates into a skeletal grayscale image.

In order to use depth information effectively, we map the 3D skeleton coordinate extracted by the Kinect camera to three perspectives. They are front view (XY plane), top view (ZY plane) and left view (XZ plane) respectively. Then we extract SkeGEI (Figure. 1) from each perspective.

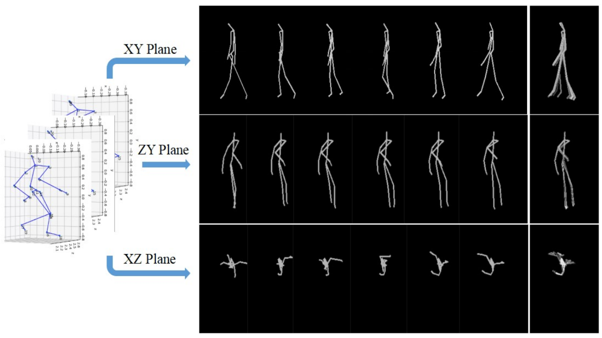

Figure 1: The SkeGEI of three views

Take the XY plane as an example. Suppose $B_t(x, y)$ is the single frame binary skeleton gait image at time t in a sequence. The specific steps are described here:

(1) Get binary skeleton gait image $B_t(x, y)$: Make sure all the skeletal are in the same height range by scaling the skeleton coordinates of the head and foot. Considering the neck is most stable joints during walking, the position of the neck is fixed at a stable point on the image and regarding it as the origin. Then other joints are converted to relative coordinates. Finally, draw skeleton joints and limbs part on the canvas to get binary skeleton gait image $B_t(x, y)$.

(2) Obtain the gait cycle: It's essential to obtain the gait cycle to filter the noise frames and ensure that SkeGEI contains a complete walking process for human. GEI [9] estimate the gait frequency and phase by maximum entropy spectrum estimation from the temporal series signal. This paper extracts the gait cycle by calculating the Euclidean distance between two feet in the vertical view. Calculating the walking distance for each frame in a gait cycle:

$$l_{wd} = \sqrt{(x_r - x_l)^2 + (z_r - z_l)^2} \tag{2}$$

Where (x_r, y_r) is the coordinates of the right foot joint and (x_l, y_l) is the coordinates of the left foot joints in the vertical view.

The walking distance has obviously periodic because each maximum value appears when the distance between the legs is greatest and each minimum value appears when the legs overlap in the vertical position. The differences of frame number between every two adjacent odd maximum frames are calculated. And obtain average differences as the frame number N of gait cycles in the video sequence.

(3) Extracting SkeGEI: Reference [9], given the preprocessed binary gait skeleton images, the SkeGEI is defined as follows:

$$G(x,y) = \frac{1}{N}\sum_{t=1}^{N} B_t(x,y) \tag{3}$$

where x and y are the pixel's coordinates at the front view, N is the number of frames in the complete cycle(s), t is the frame number in the sequence (moment of time).

The pixel values in SkeGEI can indicate the occurrence probability of the pixel in one gait cycle, including both the temporal and spatial information of the gait. In addition, SkeGEI is robust to single frame skeleton detection error.

2.2. Extraction of DA feature

SkeGEI has lost some information about speed and frequency, so we extract dynamic gait features to complement SkeGEI additionally. Preis *et al*. [11] selected 11 skeleton features captured by Kinect as the static feature, used the step length and speed as dynamic features, and integrated both static and dynamic features for recognition. Raviteja *et al*. [12] modeled the 3D geometric relationships between various body parts using special Euclidean group SE(3). Inspired by [11, 12], this paper constructs dynamic skeleton features by relative joint distance and joint angle, which is as described below.

We propose DA features which include distance features(D) and angle features(A) when people were walking, as illustrate in Figure. 2. The DA feature can be denoted as follows:

$$F_{DA} = \{F_{distance}, F_{angle}\} \tag{4}$$

Distance features: The relative distance between hands and foots changes periodically when people were walking, so we extract the relative distances of five pairs of joints which swings a lot when walking as relative distance characteristics(D):

$$F_{distance} = [d_1, d_2, d_3, d_4, d_5] \tag{5}$$

Where d_1 indicates the distance between elbows, and the others indicate the distance between wrists, knees, ankles and feet by ordinal (Figure.3a). Take the d_1 as an example:

$$d_1 = \overrightarrow{j_r j_l} = \sqrt{(x_r - x_l)^2 + (y_r - y_l)^2 + (z_r - z_l)^2} \quad (6)$$

Where (x_r, y_r, z_r) is the 3D coordinate position of right elbow, and the other is the 3D coordinate position of the left elbow.

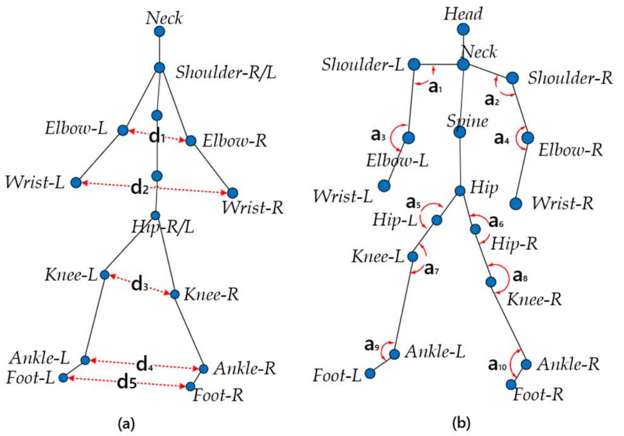

Figure 2: DA feature in human skeleton provided by Kinect V1
(a) distance feature; (b)angle feature

Angle features: The variety of the rotation angle between the limb joints is another significant gait feature. We selected the rotation angle of two sets of arm joints and three sets of leg joints as angle features in this paper, shown in Figure.2b. It can be denoted as follows:

$$F_{angle} = [a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}] \quad (7)$$

Some previous works [5, 8, 10] explored gait features by calculating joint angles in a specific view. To further utilize the advantages of 3D coordinate and achieve high-dimensional angle information, the special orthogonal group SO(3) is used to describe the rotation of the 3D space. The SO(3) is expressed as so(3) which is a collection of 3x3 oblique symmetric matrices in Lie algebra, shown in Eq. (8). The so(3) is a sparse matrix that can be mapped to a 1x3 3D vector, expressed as Eq. (9).

$$R = \begin{bmatrix} 0 & -r_3 & r_2 \\ r_3 & 0 & -r_1 \\ -r_2 & r_1 & 0 \end{bmatrix} \in so(3) \quad (8)$$

$$vec(R) = [r_1, r_2, r_3] \quad (9)$$

Take a_i as an example, it can be computed as follows:

$$a_i = vec(R_i) \quad (10)$$

Where R_i is a special orthogonal group calculated by this approach [8].

3. Proposed Framework of Approach

In this section, we describe the overall framework and the CNN-LSTM network architecture used

3.1. Framework of The Proposed Approach

In this paper, the proposed approach using the backward fusion strategy of SkeGEI and DA features for gait recognition, which makes them complementary. The work-flow of the proposed approach is illustrated in Figure 3.

The first step is deriving the SkeGEI from skeleton coordinate sequences captured from Kinect, which contains skeleton coordinate preprocessing, calculating the number of frames in a complete gait cycle, and extracting SkeGEI feature. Then, acquire the DA feature of each frame in the skeleton sequence and concatenate by the maximum length of the gait cycle in the data. Finally, input SkeGEI and DA feature into the CNN-LSTM networks separately to extract high-dimensional temporal and spatial information, and backward fuse in the network.

3.2. The CNN-LSTM Network Architecture

The CNN-LSTM network is proposed as illustrated in Figure 4. We use CNN and LSTM deep network structures to extract the high-dimensional temporal-spatial features of SkeGEI and DA features respectively. The multi-stream skeleton features are merged via two full connected layers in the CNN-LSTM network.

CNN for Spatial Features: With the development of deep learning, it has become a consensus that neural networks can extract high-dimensional features from data. Shiraga *et al*. [13] present a CNN-based approach to automatically learn gait features or representations directly from low-level input raw data (i.e. Gait Energy Image (GEI)). Unlike [13], an SVM classifier is used to replace the softmax layer at last. Experiments have shown that SVM classifier performs better in the field of multi-class. And like RGB images, SkeGEI in three views are inputted into the network as three channels for learning 3D information at the same time

LSTM for Temporal Features: CNN-based network is good at extracting spatial features, but temporal features are not sufficiently extracted. Therefore, recurrent network-based approaches have been proposed [14,15]. Inspired by these studies, we employ the LSTM to extract the temporal features from DA features sequences.

Backward Fusion: Since SkeGEI and DA features represent different information in the gait cycle and hard to be unified normalized, pre-experiment has verified the effect of forwarding fusion poorly performed. Therefore,

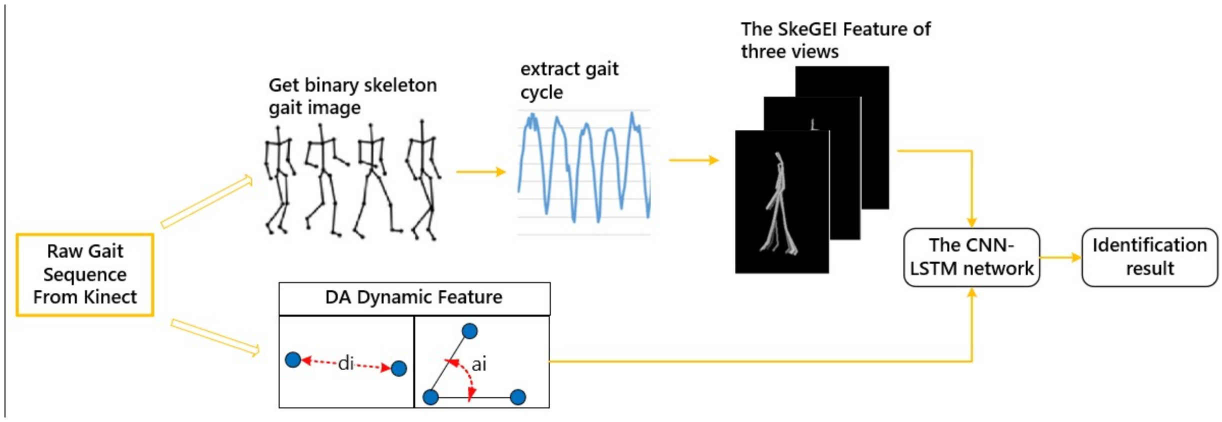

Figure 3: Framework of proposed approach

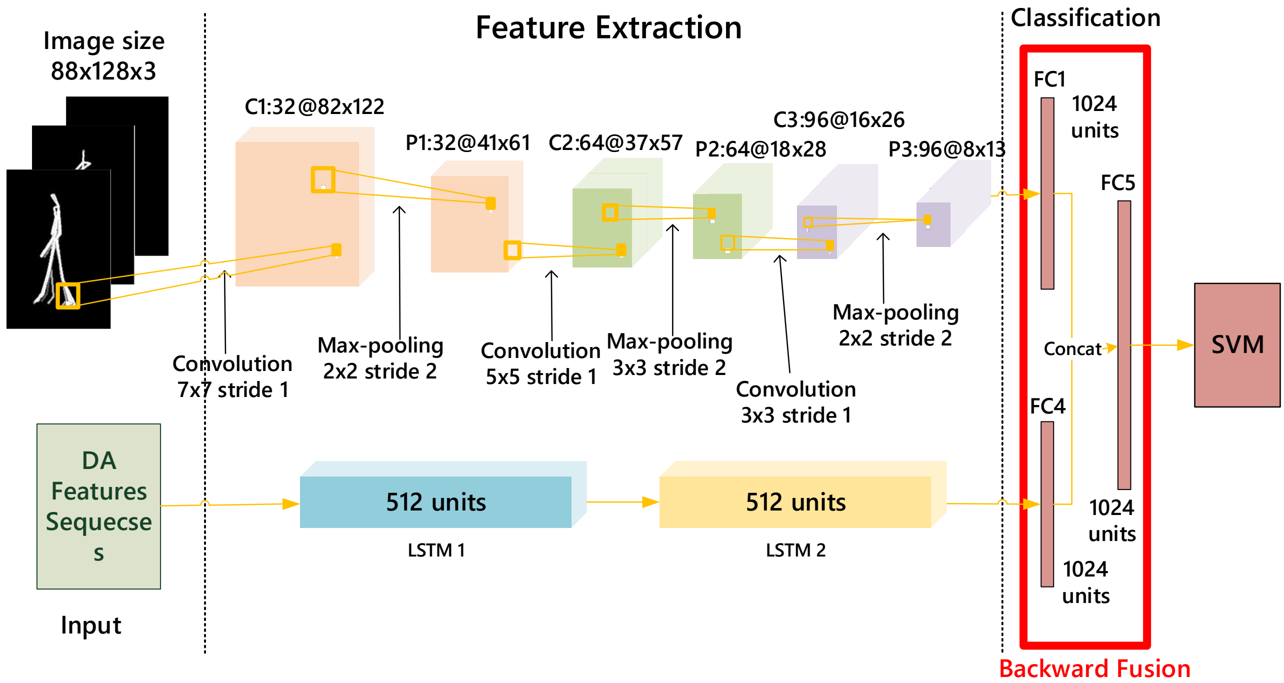

Figure 4: The proposed CNN-LSTM network for gait recognition.

the backward fusion via full connected layers is applied in this paper. In contrast with traditional approach that determine the fusion coefficient manually, the full connected layer is able to efficiently fuse features together using learnable weights.

4. Experiments and Analysis

4.1. Datasets

To evaluate the capability of the proposed approach, three Kinect-based gait datasets are used. These datasets are captured with different procedures and settings. The descriptions of the datasets are shown below.

Kinect Gait Biometry Dataset [17]. It is a unidirectional gait dataset which contains walking sequences from 160 subjects captured by Kinect V1 sensor. Each subject was asked to walk five round trips in front of the Kinect in a semi-circular trajectory (Figure.5a). And the Kinect can rotate to follow the subject, to guarantee the subject is always in the centre of the view..

SDU gait dataset [18]. This dataset contains a total of 1040 walking sequences of 52 subjects captured by two Kinect V2 sensors. Each subject has 20 walking sequences acquired from 6 fixed (0, 90, 135, 180, 225, 270 degrees) and 2 arbitrary directions (Figure.5b). It is a challenging dataset due to directional changes and fewer gait data of each person.

CILgait dataset [19]. This dataset obtained walking sequences from 12 subjects by Kinect V2 sensor. Each subject was recorded 22 sequences. And 16 of them are directional sequences (0, 45, 90, 135, 180, 225, 270, and 315 degrees), the others are abnormal patterns. It is a challenging dataset due to a variety of situations that may exist in a realistic environment.

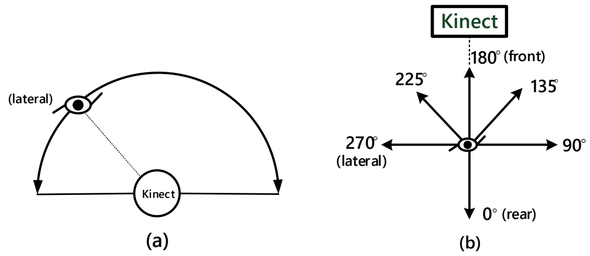

Figure 5: (a) semi-circular trajectory,
(b) six walking directions

4.2. Experimental Settings

The vast majority of gait databases are self-built without uniform standards, which leading different experiment settings between databases.

On Kinect Gait Biometry Dataset, the protocol of [14] is followed, in which the experiments are conducted using the 10-fold cross validation technique.

On SDU gait dataset, three protocols are applied according to [19]. They are SDU-front, SDU-side and SDU-arb respectively. The SDU-front protocol consists of only four walking sequences from 180 degree. Two sequences are employed for training and the others are used for testing. The SDU-front protocol includes six walking sequences in the side direction (90 and 270 degrees). In the SDU-front protocol, three sequences are adopted as training, and the remaining three are used for testing. Define the SDU-arb protocol comprising of a total of 16 walking sequences of six directions for training and four sequences with arbitrary walking directions for testing. In these protocols, statistical evaluation is not required since the sequences for training and testing are separated.

On CILgait dataset, the protocol of [19] is followed. Each subject was recorded 22 sequences, 16 of which are used for training and 6 for testing. In the training sequences, eight directional sequences (0, 45, 90, 135, 180, 225, 270, and 315 degrees) are included and each direction contains two sequences with a linear walking. On the other hand, divide the testing sequences, which are abnormal walking patterns, into three categories (CILgait-S, CILgait-SC, and CILgait-SV) with two sequences each. CILgait-S meanings walking freely and even stop walking. CILgait-SC denotes calling while walking(C) and CILgait-SV represents subject are looking at the phone (V) while walking.

Unlike the other work, our approach processed each gait cycle separately. It means that for every sequence, we obtained as many evaluations as there were in gait cycles. Comparing with approaches analyzing the whole sequence, analyzing a single cycle effectively filter out noise frames in gait sequence. Since some related work evaluates the entire sequence, in order to compare with their recognition results, we achieve the final performance by evaluating each gait cycle in the sequence. We also employed 20 joints which can be extracted from the Kinect V1 even if the two datasets are captured from the Kinect V2 for a fair comparison.

4.3. Features Analysis Experiment

In order to compare the recognition accuracy of SkeGEI and DA feature separately, we removed the concatenation before the full connection layer (Figure 4.). Considering the characteristics of these two features, the CNN branch is arranged to SkeGEI and the LSTM branch is provided to DA feature. However, note that the dimensionality of DA feature depends on the frame amount of gait cycle. Therefore, DA feature can't input to the CNN branch because it is a variable sequence. The recognition results for SkeGEI and DA feature are shown in Table 1, Table 2 and Table 3. Experimental settings are the same for the two features.

Table 1 reports the performance on Kinect Gait Biometry Dataset using the protocol of [14]. From Table 1, we can find that the SkeGEI has achieved better accuracy than DA feature under the unidirectional condition. What's more, the backward fusion based the CNN-LSTM network increased respectively by 4.22% and 10.17% compared with SkeGEI and DA single feature. This accuracy rate indicates that SkeGEI and DA features complement each other well.

Table 2 reports the recognition rates on SDU gait dataset using the protocol of [19] where results in the last row are the average of results for the three subsets SDU-front, SDU-side and SDU-arb. On the three subsets, the accuracy of SkeGEI is respectively 6.5%, 12.38% and 11.05% better than DA feature. However, after SkeGEI and DA features are fused in backwardly, the DA feature plays its part. With the help of DA, the average accuracy of three subsets reach 88.11%, increased by 3.06% and 13.03%, respectively. The experimental results show that SkeGEI is relatively less sensitive to directional variations, and the fusion of it with DA feature can improve the results to be better.

Table 3 reports the recognition rates on CIL gait dataset using the protocol of [19] where results in the last row are the average of results for the three subsets CIL-S, CIL-SC and CIL-SV. On the CIL-S dataset, the accuracy of SkeGEI is respectively 13.82% better than DA feature. Nevertheless, the accuracy of DA feature is respectively

6.37% and 8.78% better than SkeGEI on the CIL-SC and CIL-SV datasets, which means DA features is more robust than SkeGEI when people walk in abnormal patterns (i.e. looking at phone or calling). On the CIL-SV dataset, the accuracy of backward fusion performs worse than single DA feature, which indicates SkeGEI play a negative effect. The reason may be that the SkeGEI bases on skeleton shape and the shape changes a lot between training and testing on the CIL-SC and CIL-SV datasets. It should be noted that the average accuracy of our approach is respectively 5.49% and 5.04% higher than the average accuracy of SkeGEI and DA. Better results indicate that our approach is comparable in complex realistic environment.

Table 1 Recognition rates on Kinect Gait Biometry Dataset using the protocol of [10]

Approach	Accuracy(%)
DA + LSTM	86.67
SkeGEI + CNN	93.17
Backward fusion	**97.39**

Table 2 Recognition rates on SDU datasets using the protocol of [19]

Approach	Accuracy(%)			
	front	Side	arb	Average
DA + LSTM	78.84	76.92	69.47	75.08
SkeGEI + LSTM	85.34	89.3	80.52	85.05
Backward fusion	**90.14**	**91.67**	**82.51**	**88.11**

Table 3 Recognition rates on CIL datasets using the protocol of [19]

Approach	Accuracy(%)			
	S	SC	SV	Average
DA + LSTM	72.71	76.56	**76.2**	75.16
SkeGEI + LSTM	86.53	70.19	67.42	74.71
Backward fusion	**88.83**	**76.7**	75.08	**80.20**

4.4. Comparison with Prior Work in Human Recognition

In this part, the recognition accuracy of our approach is compared with the accuracy of state-of-the-art approaches including skeleton-based and other types. Our approach achieves the same or better performance in three datasets.

(1) **Kinect Gait Biometry Dataset**: Table 4 reports the accuracy of our approach along with existing approaches under experimental settings [14] on Kinect Gait Biometry Dataset.

The approached of [17, 20, 23] extract anthropometric and gait features from skeleton data. Li et al. [14] proposed using a dynamic LSTM network to learn more descriptive sequence characteristics of the gait cycle directly from the raw motion data. We compare the results of the proposed approach with three existing techniques as shown in Table4.In the protocol of [14], the accuracy of our approach is respectively 0.39%, 1.99% and 9.69% higher than the accuracy of the state-of-the-art approaches based manual feature [23, 20, 17]. Compared with [14] which use a dynamic LSTM network to learn sequence characteristics from the raw gait data, our approach also achieves an increase of 0.83% performance. While [17, 20, 23] are based on spatial information and [14] focus on temporal information, our approach fuses SkeGEI and DA backwards, which captures not only skeleton spatial information but also the variation of amplitude and angles during the temporal window of gait cycle.

One of the assessments of human identification techniques is how they perform with different gallery sizes. Figure6. Shows the accuracy rates of the proposed technique with different gallery sizes, in comparison to techniques from [14], [17], [20] and [1]. For each size, individuals were randomly selected from the same 140 individuals [14,20]. The approach of [17] produces much lower accuracy than other technologies even if the gallery size is small and get worse accuracy with bigger gallery size. The performance of the proposed approach and [23], [14] are comparable when gallery size is less than 50. However, performance of [14] decrease, with a higher rate than the proposed approach. In particular, the proposed approach acquired good performance of higher than 97% at different gallery size.

Table 4: Performance comparison of existing approaches using the protocol [14] (%)

Appaorch	Kinect Gait Biometry Dataset [14]
Khamsemanan *et al.* [23]	97.00
Andersson and Araujo [17]	87.7
Yang *et al.* [20]	95.4
Li *et al.* [14]	96.56
Ours	**97.39**

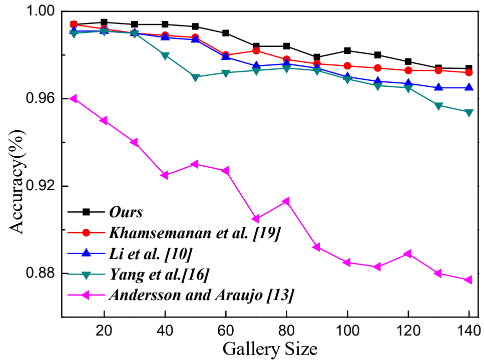

Figure 6: Average recognition accuracy under different gallery sizes

(2) **SDU gait dataset**: Choi et al. [19] evaluate a robust skeleton-based frame-level matching approach on SDU and CIL gait datasets. Figure 7. reports the performance on SDU gait dataset along with existing approaches using the protocol of [19]. On the SDU-front dataset, our approach fails to achieve the first place. Compared with other unidirectional gait datasets, the training set in SDU-front dataset is small and lead to the CNN-LSTM network of our approach lack of enough training. On the SDU-side and SDU-arb dataset, the accuracy of our approach is

respectively 3.85% and 9.91% better than the highest accuracy of the skeleton-based approach. Considering that the walking direction changes in SDU-arb dataset, the contribution of our approach is acceptable. Specifically, the average accuracy of our approach is higher than the average accuracy of other skeleton-based approaches.

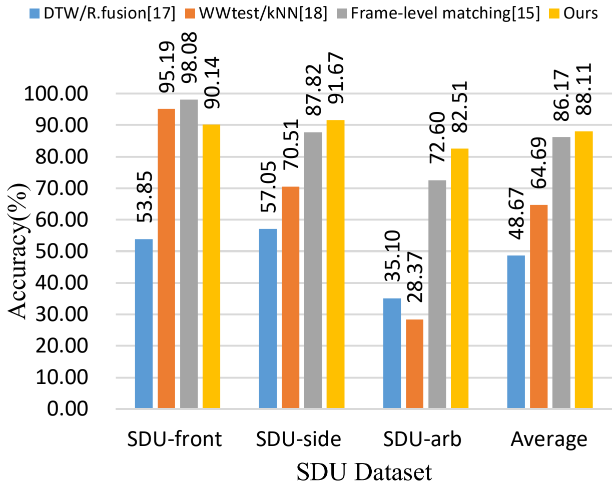

Figure 7. Performance comparison of existing approaches on SDU dataset using the protocol [19] (%)

(3) **CIL gait dataset:** Figure 8 reports the recognition rates on CIL gait dataset along with existing approaches using the protocol of [19]. On the CIL-S and CIL-SC datasets, the accuracy of our approach is respectively 1.33%, 1.70% better than the highest accuracy of frame-level matching [19]. While our approach doesn't perform as well as frame-level matching [19] on CIL-SV with respectively lower 8.25%. Whereas the difference between training and testing set is wide, the CNN-LSTM network of our approach couldn't learn features that do not appear in the training set. The average accuracy of our approach is comparable with the frame-level matching approach [19].

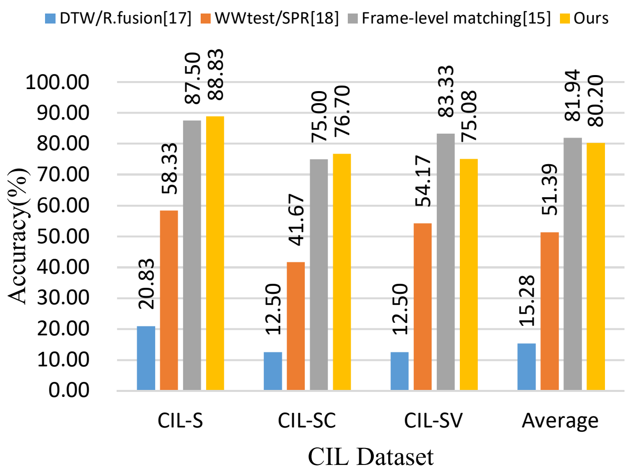

Figure 8. Performance comparison of existing approaches on CIL dataset using the protocol [19] (%)

5. Conclusions

In this paper, a novel approach is proposed based on backward fusion strategy of SkeGEI and DA features for the gait recognition with multi-directional and abnormal patterns. SkeGEI and DA features is first proposed in our work and the CNN-LSTM network effectively extract the temporal-spatial deep feature information from handcraft features. The simulation results show that the proposed algorithm has some advantages over the existing classical algorithms, and substantially improved the gait recognition accuracy with direction variations. In future work, we are planning to explore other types of network structure such as 3DCNN for extracting feature from raw gait sequences.

Acknowledgements: This work is funded by National Natural Science Foundation of China (Grant No.61572321 & 61572320). It is supported by the National Key Research and Development Projects of China （2018YFC0830703）.It is also supported by the Foundation of Shanghai Fusion and Innovative Research Laboratory for Procuratorial Big Data . The Corresponding Author is Dr. Xinghao Jiang.

References

[1] D. S. Matovski, M. S. Nixon, S. Mahmoodi and J. N. Carter, The Effect of Time on Gait Recognition Performance, IEEE Transactions on Information Forensics and Security, 7(2):543-552, 2012.

[2] J. Han and B. Bhanu, Individual recognition using gait energy image, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(2):316-322, 2006.

[3] Z. Wu, Y. Huang, L. Wang, X. Wang and T. Tan, A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(2): 209-226, 2017.

[4] X. Li, Y. Makihara, C. Xu, D. Muramatsu, Y. Yagi, M. Ren, Gait energy response functions for gait recognition against various clothing and carrying status, Applied Sciences, 8(8):1380, 2018.

[5] Z. Kang, M. Deng and C. Wang, Frontal–view human gait recognition based on Kinect features and deterministic learning, 2017 36th Chinese Control Conference (CCC), Dalian, 2017, pp. 10834-10839.

[6] B. Tekin, A. Rozantsev, V. Lepetit and P. Fua, Direct Prediction of 3D Body Poses from Motion Compensated Sequences, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 991-1000

[7] J. Preis, M. Kessel, M. Werner, and C. Linnhoff-Popien, Gait recognition with Kinect, Proc. 1st Int. Workshop Kinect Pervasive Comput, New Castle, UK, 2012, pp. 1-4.

[8] H. Zhen, M. Deng, P. Lin and C. Wang, Human gait recognition based on deterministic learning and Kinect sensor, 2018 Chinese Control And Decision Conference (CCDC), Shenyang, 2018, pp. 1842-1847.

[9] V. G. M. Guru, V. N. Kamalesh and R. Dinesh, Human gait recognition using four directional variations of gradient gait energy image, International Conference on Computing, Communication and Automation (ICCCA), Noida, 2016, pp. 1367-1371.
[10] M. Deng and C. Wang, Human gait recognition based on deterministic learning and data stream of Microsoft Kinect, in IEEE Transactions on Circuits and Systems for Video Technology. doi: 10.1109/TCSVT.2018.2883449
[11] M. W. Rahman and M. L. Gavrilova, Kinect gait skeletal joint feature-based person identification, IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), Oxford, 2017, pp. 423-430.
[12] R. Vemulapalli, F. Arrate and R. Chellappa, Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group, IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 588-595.
[13] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo and Y. Yagi, GEINet: View-invariant gait recognition using a convolutional neural network, International Conference on Biometrics (ICB), Halmstad, 2016, pp. 1-8.
[14] J. Li, L. Qi, A. Zhao, X. Chen and J. Dong, Dynamic long short-term memory network for skeleton–based gait recognition, IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), San Francisco, CA, 2017, pp. 1-6.
[15] R. Liao, C. Cao, E.B. Garcia, S. Yu, Y. Huang, Pose–based temporal-spatial network (ptsn) for gait recognition with carrying and clothing variations, Chinese Conference on Biometric Recognition(CCBR), Beijing, 2017, pp.474-483.
[16] W. An, R. Liao, S. Yu, Y. Huang, P. Yuen. Improving Gait Recognition with 3D Pose Estimation, Chinese Conference on Biometric Recognition(CCBR), Urumchi, 2018, pp. 137-147.
[17] V. O. Andersson and R. M. de Ara′ujo, Person identification using anthropometric and gait data from kinect sensor. in AAAI, Texas, 2015, pp. 425-431.
[18] J. Sun, Y. Wang, J. Li, W. Wan, D. Cheng, and H. Zhang, View-invariant gait recognition based on kinect skeleton feature, Multimed. Tools Appl, 77:24909-24935, 2018.
[19] S. Choi, J. Kim, W. Kim and C. Kim, Skeleton-based Gait Recognition via Robust Frame-level Matching, in IEEE Transactions on Information Forensics and Security. doi: 10.1109/TIFS.2019.2901823
[20] K. Yang, Y. Dou, S. Lv, F. Zhang, and Q. Lv, Relative distance features for gait recognition with kinect, Journal of Visual Communication and Image Representation, 39: 209-217, 2016.
[21] F. Ahmed, P. P. Paul, and M. L. Gavrilova, Dtw–based kernel and ranklevel fusion for 3d gait recognition using kinect, The Visual Computer, 31(6-8):915-924, 2015.
[22] D. Kastaniotis, I. Theodorakopoulos, C. Theoharatos, G. Economou, and S. Fotopoulos, A framework for gait–based recognition using kinect, Pattern Recognit. Letters, 68(2): 327-335, 2015.
[23] N.Khamsemanan, C. Nattee and N. Jianwattanapaisarn, Hu man Identification From Freestyle Walks Using Posture-Bas ed Gait Feature, IEEE Transactions on Information Forensic s and Security, 13(1): 119-128, 2018.