RESEARCH ARTICLE OPEN ACCESS

# Cross-view Gait Recognition Based on Spatio-temporal Feature Fusion

Zhongchao Xia[1],Tianqi Yang[2], Xiaofang Wu[3]

1* (College of Information Science and Technology,Jinan University, Guangzhou 510632, China
Email: mmix@stu2017.jnu.edu.cn)

2,3 (College of Information Science and Technology,Jinan University, Guangzhou 510632, China
Email: tytq@jnu.edu.cn, wuxf@stu2017.jnu.edu.cn)

----------------------------------------************************----------------------------------

## Abstract:

A cross-view gait recognition method based on spatio-temporal feature fusion is proposed. This method introduces temporal information of the gait by constructing Gait Temporal Image (GTI), and uses a two-stream convolution neural network STGNet (Spatio-temporal Gait Net) to extract spatio-temporal features of the gait, and finally obtains the recognition result of the identity through the nearest neighbor classifier. The feature fusion deep learning algorithm was tested on the large data set OU-ISIR Large Population and compared with other mainstream cross-view gait recognition methods. The results show that the proposed method has a highest recognition rate and achieves an average recognition rate of 99.5% in cross-view scenario, and shows certain robustness under different angle changes.

*Keywords* **—gait recognition, gait temporal image, two-stream convolutional neural network,spatio-temporal feature fusion, deep learning.**

----------------------------------------************************----------------------------------

## I. INTRODUCTION

With the development of society, biometrics technology is becoming more and more important in people's daily life. In recent years, such technologies have been widely applied to people's daily life, such as access control, intelligent monitor system; gait recognition is an emerging biometric technology that uses people's walking posture to identify people. Compared with other biometric technologies (such as faces, fingerprints, etc.), gait recognition has the advantages of long distance, non-contact, no perception and so on [1][2] , For example, gait recognition can be performed well at a distance (50 meters) where other recognition algorithms have failed. Gait recognition has a wide application prospect and commercial value in the fields of security and intelligent monitoring.

### A. *Gait Templates*

Generally, gait recognition can be classified into two types: model-based and model-free, model-based methods are focus on the characteristics of people's movements, using human body models to track and model different parts of the body, through the model's description of the motion, the identify the human is obtain.

The model-free methods focus on the shape of the silhouette, or the motion state of the entire human body, this type of method is less sensitive to the quality of the obtain image than the model-based method and much faster in calculations. [2], because of its robustness and ease to use, it has been widely studied by scholars [3-6].

According to the different characteristics of gait, different scholars have proposed different gait templates: Han et al[3]. proposed to obtain a gait energy image (GEI) by averaging the silhouette of a period, which is the most widely used template method, and it is simple to use and has strong robustness. Wang et al[4]. proposed to extract the contour of a periodic gait image first, and then use a color mapping function to encode each of gait contour images in the same gait sequence and compositing them to a single Chrono-Gait Image (CGI); Lam et al[5]. proposed a novel gait representation that generated by using an optical flow field without constructing any model; Makihara et al[6]-[8]. propose a method of gait recognition from various view directions using Fourier analysis to extract the frequency-domain features of the volume, and using a view transformation model to model the relationship between two different view; Liu et al[9]. proposed an improved temporal template called Gait History Image (GHI), the time duration of the GHI template is controlled by a finer period resolution (1/4 of a gait cycle).

### *B. Deep Learning based Gait Recognition*

Inspired by the breakthrough of deep learning in otherfields, gait recognition has begun to make use of such technologies. Deep learning have succeed in extracting features (image and video), and much better than handcraft features by human in representation.

Shiraga et al [10]. proposed a simple but effective deep learning based gait recognition method name GEINet which consist of two sequential network layers (composed of a convolutional layer, a pooled layer and a normalized layer) and two fully connected network layers, using the gait energy image (GEI) as the input to the network. By minimized the output of different views and the corresponding ID probability to obtain an angle-invariance gait feature extraction network. Then in the validation process, using the output vector of the neural network as the match feature in the matching process. Such simple method achieves high performance. The effectiveness of the method is also verified in the large data set OU-ISIR Multi-View Large Population Dataset [11].

Li, Min et al [12]. proposed a transfer learning based method, using the large-scale dataset ImageNet pre-training neural network VGG-D network to extract gait features, and using the joint Bayesian method and PCA in the recognition phase to achieve a cross view match and higher recognition rate.

Chao et al [13]. analyzed the characteristics of the silhouette and found that the silhouette in the gait is visually identifiable and recognizable, and the deep learning based method like 3D-CNN convolutional neural network training of silhouettes is very difficult compared to the simple gait template like GEI. Therefore, the authors proposed to use the gait input silhouette as a set, without modeling the timing relationship of the silhouette in the gait, and utilize the deep neural network optimize itself to extract those relationship.

The problem of gait recognition is not a classification problem, but a matching problem. Many scholars have proposed method based on metrics learning for gait recognition[12][14]. Zhang et al [14]. proposed using the two-stream Siamese convolutional neural network with the paired of gait energy image to train the network, and optimized the Euclidean distance of the positive and negative samples to achieve the purpose of metrics learning. Finally, the Euclidean distance of the output vector was used in the validation process.

Takemura et al [11]. discuss the impact of input/output architecture (such as Siamese Network, GEINet etc.) and loss function for different CNN based cross-view gait recognition method. Where the metric learning method 3in+2diff and GEINet methods achieved higher performance compare with other cross-view gait recognitions methods. The experiment was carried out to analyze on the Rank-1 identification rate and equal error rate;

### *C. Spatio-temporal Feature Fusions*

Simonyan et al [15]. proposed a two-stream convolutional network architecture consisting of a temporal network and a spatial network, where the spatial network uses a single-frame input, and the temporal network use dense optical stream as input, and averaged the output score in the sequential stream network, finally get the final probabilities of each category.

Wu et al[16]. referenced the method in the field of face recognition[17], and proposed a weight-sharing convolutional neural network MT and LB, which is a shallow layer networks to learn the similarity between different GEI, and test the proposed method on the CASIA-B and OULP dataset, Achieving a highest recognition rate in both dataset, the paper analyzes the influence of deep convolutional neural network input, resolution and other factors, And try spatio-temporal feature fusion of combining GEI and CGI inputs.

Three-dimensional convolutional neural networks are widely used in the field of video understanding, such as action recognition[18][19], which is mainly used to extract spatio-temporal features. In recent years, it has also been favored by many scholars in the field of gait recognition.

Wolf, Thomas et al[20]. proposed using 3D convolutional neural network for spatio-temporal feature extraction, which used a special input format consisting of a special grayscale image and optical flow, show a comparable to better performance in comparison with other approaches, especially for large view differences.

Thapar et al [21]. propose a two-stage recognition framework, which use 3D convolution deep neural network for gait recognition under multiple view. The first stage is use classification network to identify the viewing point angle, second stage is to use the trained network to identify the person under a particular viewing angle. The experiment shows the propose method have achieved decent result.

### *D. Overview*

Our contributions include: (1) introducing a spatio-temporal feature fusion gait recognition method(STGNet) with a new gait template(Gait Temporal Image) which outperforms traditional gait representations when the gallery and probe gait sequences are from both same view and cross view (2) improved the recognition performances on the Large Population Dataset (OULP) for both identical-view and cross-view settings.

## II. MATRERIALSANDMETHODS

### *A. Datasets*

In the currently disclosed gait database, in order to remove the influence of color, texture, the gait data is a grayscale image of walking human, such as OU-ISIR Large Population, Publish by Osaka University, and CASIA Gait Database, published by the Institute of Automation, Chinese Academy of Sciences. The above dataset provided a grayscale image of the human.

The dataset used in the experiments in this paper is a large gait data set published by Osaka University in Japan in 2009. Its full name is The OU-ISIR Gait Database, Large Population Dataset (OULP). The data set contains 4007 people (2135 women and 1872 men) and is aged from 1 to 94. The gait sequence at angles of four viewing (55°, 65°, 75°, 85°), is one of the largest gait databases [22], which provides two data formats. One is the gait energy image (GEI), and the other is the gait contour image (Silhouette).

Fig. 1OU-ISIR Gait Database samples, same person, from left to right are corresponding to the time order. The size of image is 128 height and 88 width respectively.

### *B. Proposal of Gait Temporal Image*

Because the walking speed and camera acquisition rate are different, the number of images in a walking sequence is inconsistent. In order to ensure the dimension of the extracted gait temporal image, we first propose a Gait Volume. The concept is, by limiting the amount of image in one sequence, we adopted a strategy of repeated sampling to keep the time dimension corresponding to the gait sequence consistent. When the number of images is insufficient, we use a repeat strategy. When the number of images exceeds the preset value, we use the random sampling strategy.

A step for construct gait volume is by stacking gait silhouette according with time dimension order to obtain a pixel volume with a dimension size of [n, C, H, W], (where n is the chronological length of the gait, C For the number of channels of the image, H is the height of the image, W is the width of the image), transpose the coordinate axis to obtain the volume size of the dimension [H, C, n, W], and the gait can be obtained after first axis expansion. In this paper, we call it a Gait Temporal Image (GTI).

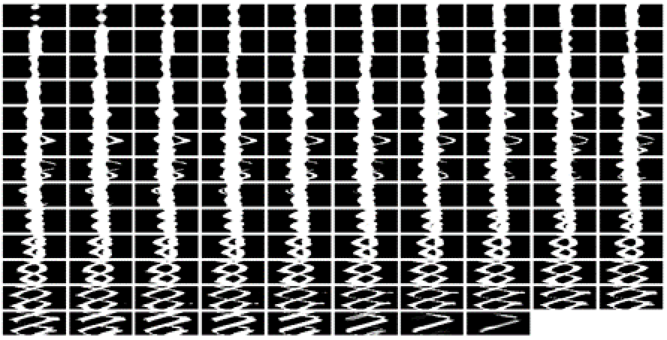

Fig. 2 Generated Gait Temporal Image (GTI) samples. From left to right are corresponding to the pixel height order in silhouette. The height of generated image is 32 pixel, and the width is 88 pixel.

It can be observed from Fig. 2that there is a specific gait pattern in the image according to the GTI on the different segmentation points H, where in the part in the middle of the position can see the trace of the hand swing, and the position in the lower part can be seen a trace of walking, GTI can be used to extract the temporal information of the gait and the dynamic characteristics of the gait.

### *C. Proposal of Network Architecture*

Gait is a kind of time series feature, but most of the methods are based on gait template [3][4][9]. Such a template lose dynamic information of the gait in the synthesis step, to a certain extent, the recognition performance problem is caused; if such a gait template is not used, it is difficult to utilize the silhouette sequence directly [13], and the time dimension lengths in different gait data are different.

To handle the above problems, we have improved on the basis of GEINet, and solved the problem of different lengths of time dimension by constructing Gait Volume which align the shape of the gait sequence. Based on this, we constructed the projection of gait volume to the temporal axis to obtain the Gait Temporal Image, designed and trained a two-stream convolutional neural network STGNet, then independent extract the special and temporal feature and fusion of those features. The gait data of different angles are used in the training phases, and the gait spatio-temporal view-invariance features are obtained.

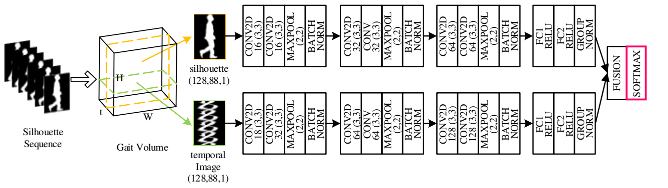


Fig. 3 An illustration of the proposed gait recognition method STGNet (Spatio-temporal Gait Net), The Neural Network is consistent of two different part, one is for extracting the space information of the gait, and the other is for extract the temporal information of the gait. And finally perform a feature-level fusion on the output layer.

The proposed STGNet is consists of two independent two-dimensional convolutional neural networks, which is SGNet and TGNet, the first network is used to extract spatial information of the gait, and the second network is used to extract temporal information of the gait, And finally, the spatio-temporal feature fusion is performed, The architecture of STGNet can be show in Fig. 3. In this paper, the cross-entropy loss function is used to compute the objection function of the neural network, when training the network we use stochastic gradient descent[23] algorithm to optimizer network parameters, Duo to the aim of compare the influences of GTI on the final output of the network, we use a joint-train strategy to train the whole neural network and perform the jointtraining of the SGNet and TGNet.

In the validation step, we fuse the join embedded gait representation of the two networks outputs, and make an average of those output then compute the Euclidean distance of the gallery/probe gait, and finally get the predict result.

In the proposed experiments, stacking small convolution kernels (such as kernels size 3) are widely used, which can achieve more efficient calculation while increasing the size of the receptive field [24], which can help the network capture more global information.

In the neural network architecture design, the widely used Batch Normalization is applied to the training process of STGNet convolutional neural networks, which can increased the generalization and training speed of convolutional neural networks [25]; Group Normalization method is a normalization method in which each formed group can be considered as a sub-vector under each cluster. According to the weights learned by the convolutional neural network, certain features are specific to the specific the distribution, grouping can enable the neural network to learn better features, thus achieving better recognition [26].

### D. Training and Testing Phase

For training, we define a loss function , $L_{loss}$, which is composed of a softmax function and cross entropy of the logit vector $V_{logit}$, then we use the silhouette sequence, constructed GTI and encoded label to train the proposed STGNet network, The output probability of given input $I$ can be formula as:

$$\text{softmax}(I) = \frac{\exp(V_{logit}(I))}{\sum_c^C \exp(V_{logit}(I))}$$

where the $I$ is the input image of the silhouette or gait temporal image, the $V_{logit}$ is the logit output of the neural network, C is the denote class label. The loss function of the input would be:

$$L_{loss}(I) = -\sum_n^N \sum_c^C L_{nc} \log[softmax(I)_{nc} \quad ]$$

where L, N, and C is corresponding to the class labels, the number of training samples inI, and the number of classes.

In the training process of conducted experiment, we use the cross-entropy loss as the minimize objective function and use the stochastic gradient descent (SGD)[23] to optimize the proposed STGNet parameters, which is consistent with GEINet [10], with multi-view training data, the neural network can learn to extract view-invariance features.

To updated network weight, the learning rate of the SGD is set to 0.01 in the below experiment, and the batch size of the update iteration is set to 128. The iteration process of minimizing the loss function is repeated until the neural network converges, in general, the number of iterations is 50k.

When fusion the input of silhouette and gait temporal image in the proposed network, the fusion layers of the proposed STGNet architecture, which can be seen in the fig. 3,the fusion result can be represent as $V_f$,

$$V_f = \frac{1}{2N_s}\sum_{i=0}^{N_s} SNet(SI_i) + \frac{1}{2N_t}\sum_{i=0}^{N_t} TNet(TI_i)$$

where the $V_f$ is the fusion result of the STGNet, and the $N_s$ is the total number of the silhouette of a single gait walking sequence,and $SI_i$is numerically corresponding to the silhouette sequence. And $N_t$ is the total number of the gait temporal image of a single gait walking sequence,and $TI_i$is numerically corresponding to the gait temporal image. The spatial weights and temporal weight are the same, as a result, the weights of the temporal feature and spatial feature is 1/2.

In the validation process, we use the gait identification to demonstrate the performance of the proposed method. And use the nearest neighbor algorithm to get the identity information of the probe subject. The distance function in the nearest neighbor algorithm is defined as:

$$Dist(V_{fg}, V_{fp}) = \|V_{fg} - V_{fp}\|_2$$

where the $V_{fg}$ is the representation of the spatio-temporal fusion feature. The distance is calculated via a l2 norm of the fusion vector $V_{fg}$and $V_{fp}$.

For each of the subjects in the probe set, the distance between the gallery set is compared, let N be number of the subjects in probe set of and M be the total number of subjects in gallery set, for each of the identification score, the number of comparisons would be N*M. the prediction of specify subject in probe set would be:

$$i_{min} = argmin_{i=[0,n]}\left(Dist(V_{fp}, V_{fg=i})\right)$$

where the $i_{min}$ is the prediction of the probe subject, and the finally the accuracy score is calculated,which is based on the prediction result.

## III. EXPERIMENT

Referring to the existed published papers, we conducted experiment I under the identical view setting and experiment II under across view setting, and experiment III self-contrast experiments, and compared it with the mainstream gait recognition method. The identical/cross view indicates whether the gallery set and probe set is at the same perspective or a different perspective.

### A. Experiment I - Identical View

Experiment I is an identical-view experiment, which means that the gallery view of the gait is the same as the view of probe set We divided the

OULP dataset according to different angles, and then trained and tested STGNet using the number of subjects, and compare with DeepGait [12], Siamese Net [14], GEI [3], FDF [6] method.

TABLE 1. COMPARISON OF RANK-1(%) AND RANK-5(%) IDENTIFICATION RATES WITH DIFFERENT GAIT RECOGNITION METHODS UNDER IDENTICAL VIEW SCENARIO. GEI: GAIT ENERGY IMAGE; FDF: FREQUENCY-DOMAIN FEATURE.

| Rank-1/Rank-5 | Dataset | Subject | STGNet (paper) | DeepGait | Siamese Net | GEI | FDF |
|---|---|---|---|---|---|---|---|
| Rank-1 | View-55 | 3,706 | **93.7** | 90.6 | 90.1 | 85.3 | 83.1 |
| | View-65 | 3,770 | **93.2** | 91.2 | 91.1 | 85.6 | 84.7 |
| | View-75 | 3,751 | **93.6** | 91.2 | 91.1 | 86.1 | 86.0 |
| | View-85 | 3,249 | **94.9** | 92.0 | 90.4 | 85.3 | 85.6 |
| | Mean | | **93.9** | 92.3 | 90.6 | 85.6 | 84.9 |
| Rank-5 | View-55 | 3,706 | **97.3** | 96.0 | 94.9 | 91.8 | 91.0 |
| | View-65 | 3,770 | **97.0** | 96.0 | 95.9 | 92.3 | 92.3 |
| | View-75 | 3,751 | **96.9** | 96.1 | 95.9 | 92.2 | 92.5 |
| | View-85 | 3,249 | **97.9** | 96.5 | 95.9 | 92.6 | 92.3 |
| | Mean | | **97.2** | 96.2 | 95.6 | 92.2 | 92.0 |

It can be seen from Table 1 that the proposed method and the comparison methods of DeepGait and GEI have higher recognition rate, which indicates that the use of spatio-temporal feature fusion can improve the recognition rate of gait recognition, and the results are consistent with the actual situation. The main reason for the improvement of the recognition rate is that not only the spatial characteristics of the gait are fully utilized, but also the timing characteristics of the gait are effectively utilized. In the experiments of Rank-1 the highest score was reached, among which the recognition rate of the mean Rank-1 reached 93.9%, which was 1.6% higher than that of DeepGait method, 8.3% higher than the GEI method. And the mean of Rank-5 reached 97.2% which was 1.0% higher than the DeepGait method, 5.0% higher than the GEI method.

### *B. Experiment II - Cross View*

Experiment II is a cross-view experiment, which means that the gallery view of the gait is different from the view of probe set. We used the proposed SGNet, TGNet, STGNetfor this experiment.

The experimental settings are the same as the cross-view experiment in the paper [10] [12], which used the OULP dataset 1912 subjects of 4 views (55,65,75,85) in the OULP data set, each subject has two gait sequence. Half of the 956 subjects are used for training process, and the rest of 956 subjects are for validation process. Multiple views gait data is used in the in the training process in order to get the view-invariance feature extraction network. According to different subjects division, we performed a five two-fold cross-validation. The specific subjects division can be found in the database's official Benchmark. http://www.am.sanken.osaka-u.ac.jp/BiometricDB/dataset/GaitLP/Benchmarks.html.

During the training phase, the gait data from multiple views was used to train the network, and during the verification process, the test data was tested for Rank1 recognition rate, and the report score of tests were averaged, and compared with the paper[7][8][10][12][14]. It can be seen from Table 2 that STGNet outperform the methods of DeepCNN[16], DeepGait [12], GEINet[10], And when the perspective of the gallery peobe is large, it still gives better performance.

The proposed STGNethas surpassed the current mainstream methods and achieved the best performance in cross-view recognition in OULP dataset.

Table 2. Comparison of rank-1(%) identification rates with different gait recognition methods under cross view scenario.

| Gallery View | Method | Rank – 1 Accuracy [%] | | | |
|---|---|---|---|---|---|
| | | 55° | 65° | 75° | 85° |
| 55° | GEINet | (94.7) | 93.2 | 89.1 | 79.9 |
| | w/FDF | (92.7) | 91.4 | 87.2 | 80.0 |
| | TCM+ | | 79.9 | 70.8 | 54.5 |
| | wQVTM | | 78.3 | 64.0 | 48.6 |
| | DeepGait + JB | (97.4) | 96.1 | 93.4 | 88.7 |
| | DeepCNN | (98.8) | 98.3 | 96.0 | 80.5 |
| | **STGNet** | **(99.6)** | **99.5** | **99.6** | **99.3** |
| 65° | GEINet | 93.7 | (95.1) | 93.8 | 90.6 |
| | w/FDF | 92.3 | (93.9) | 92.2 | 88.6 |
| | TCM+ | 81.7 | | 79.5 | 70.2 |
| | wQVTM | 81.5 | | 79.2 | 67.5 |
| | DeepGait + JB | 97.3 | (97.6) | 97.2 | 95.4 |
| | DeepCNN | 96.3 | (98.9) | 97.3 | 83.3 |
| | **STGNet** | **99.7** | **(99.3)** | **99.5** | **99.4** |
| 75° | GEINet | 91.1 | 94.1 | (95.2) | 93.8 |
| | w/FDF | 88.8 | 92.6 | (93.4) | 91.9 |
| | TCM+ | 71.9 | 80.0 | | 79.0 |
| | wQVTM | 70.2 | 80.0 | | 78.2 |
| | DeepGait + JB | 93.3 | 97.5 | (97.7) | 97.6 |
| | DeepCNN | 94.2 | 97.8 | (98.9) | 85.1 |
| | **STGNet** | **99.7** | **99.5** | **(99.5)** | **99.5** |
| 85° | GEINet | 81.4 | 91.2 | 94.6 | (94.7) |
| | w/FDF | 80.9 | 88.4 | 92.2 | (93.2) |
| | TCM+ | 53.7 | 73.0 | 79.4 | |
| | wQVTM | 51.1 | 68.5 | 79.0 | |
| | DeepGait + JB | 89.3 | 96.4 | 98.3 | (98.3) |
| | DeepCNN | 90.0 | 96.0 | 98.4 | (98.9) |
| | **STGNet** | **99.5** | **99.5** | **99.5** | **(99.7)** |

### C. *Experiment III - Self Contrast*

In addition, we compare the Rank-1 recognition rate of STGNet networks in cross view scenario, together with TGNET temporal net and SGNET spatial net. TGNET is the temporal feature single-stream convolutional neural network which only use the Gait Temporal Image as input, while SNET using spatial grayscale gait sequences as input.

The experimental results show that the fusion network has the advantage of high recognition rate. The network that integrates time and space is more advantageous than the input of only one side because the fusion feature can better capture the gait information. As can be seen from Fig. 5, the fusion network STGNet achieved a 99.5% recognition rate in cross-view setting on average, which is 0.8% higher than the unfused spatial network SGNet, and 78.9% higher than the unfused temporal network TGNet.

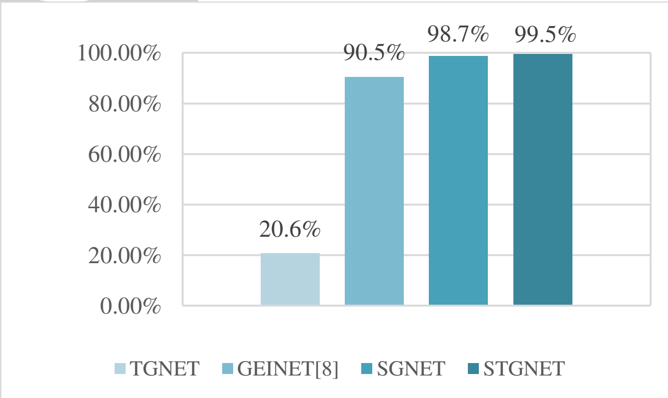


Fig. 5 Rank-1 identification rates of different Gait Recognition method under Cross view Experiment setting. The score is averaged according to the angle combinations

SGNet has achieved good performance before the temporal fusion, and the STGNet has only a small increase after incorporates the temporal features, which can show that the temporal information of the gait is more difficult to utilize than the spatial information.

## IV. CONCLUSIONS

This paper proposes a cross-view gait recognition method based on spatio-temporal feature fusion. This method introduces time information by constructing gait temporal image, and uses SGTNET convolution network to perform gait feature fusion. Compared with other mainstream methods, the effectiveness of the method is verified. The experiment has obtained the highest recognition rate on the OU-ISIR Large Population dataset: the STGNet fusion network reaches 93.9% rank-1 recognition rate under the identical view setting; STGNet achieved the Rank1 recognition rate of 99.5% under cross view setting; through the self-contrast experiments, the enhancement of the fusion for gait recognition was verified: the temporal feature and spatial feature were complementary, and their fusion can improve each other, the network after convergence increased by 78.9% when merge with the temporal network, and increased by 0.8% with the spatial network.

## ACKNOWLEDGMENT

This study is supported by “The Science and Technology Project of Guangdong, China (No.2017A010101036) and “the Fundamental Research Funds for the Central Universities (No.21619413).

## REFERENCES

[1] Gafurov, D. In A survey of biometric gait recognition: Approaches, security and challenges, Annual Norwegian computer science conference, Annual Norwegian Computer Science Conference Norway: 2007; pp 19-21.

[2] Wang, J.; She, M.; Nahavandi, S.; Kouzani, A. In A review of vision-based gait recognition methods for human identification, 2010 international conference on digital image computing: techniques and applications, IEEE: 2010; pp 320-327.

[3] Han, J.; Bhanu, B., Individual recognition using gait energy image. IEEE transactions on pattern analysis and machine intelligence 2005, 28 (2), 316-322.

[4] Wang, C.; Zhang, J.; Pu, J.; Yuan, X.; Wang, L. In Chrono-gait image: A novel temporal template for gait recognition, European Conference on Computer Vision, Springer: 2010; pp 257-270.

[5] Lam, T. H.; Cheung, K. H.; Liu, J. N., Gait flow image: A silhouette-based gait representation for human identification. Pattern recognition 2011, 44 (4), 973-987.

[6] Makihara, Y.; Sagawa, R.; Mukaigawa, Y.; Echigo, T.; Yagi, Y. In Gait recognition using a view transformation model in the frequency domain, european conference on computer vision, Springer: 2006; pp 151-163.

[7] Muramatsu, D.; Makihara, Y.; Yagi, Y., Cross-view gait recognition by fusion of multiple transformation consistency measures. IET Biometrics 2015, 4 (2), 62-73.

[8] Muramatsu, D.; Makihara, Y.; Yagi, Y., View transformation model incorporating quality measures for cross-view gait recognition. IEEE transactions on cybernetics 2015, 46 (7), 1602-1615.

[9] Liu, J.; Zheng, N. In Gait history image: a novel temporal template for gait recognition, 2007 IEEE International Conference on Multimedia and Expo, IEEE: 2007; pp 663-666.

[10] Shiraga, K.; Makihara, Y.; Muramatsu, D.; Echigo, T.; Yagi, Y. In Geinet: View-invariant gait recognition using a convolutional neural network, 2016 international conference on biometrics (ICB), IEEE: 2016; pp 1-8.

[11] Takemura, N.; Makihara, Y.; Muramatsu, D.; Echigo, T.; Yagi, Y., Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. IPSJ Transactions on Computer Vision and Applications 2018, 10 (1), 4.

[12] Li, C.; Min, X.; Sun, S.; Lin, W.; Tang, Z., Deepgait: a learning deep convolutional representation for view-invariant gait recognition using joint bayesian. Applied Sciences 2017, 7 (3), 210.

[13] Chao, H.; He, Y.; Zhang, J.; Feng, J. In Gaitset: Regarding gait as a set for cross-view gait recognition, Proceedings of the AAAI Conference on Artificial Intelligence, 2019; pp 8126-8133.

[14] Zhang, C.; Liu, W.; Ma, H.; Fu, H. In Siamese neural network based gait recognition for human identification, 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE: 2016; pp 2832-2836.

[15] Simonyan, K.; Zisserman, A. In Two-stream convolutional networks for action recognition in videos, Advances in neural information processing systems, 2014; pp 568-576.

[16] Wu, Z.; Huang, Y.; Wang, L.; Wang, X.; Tan, T., A comprehensive study on cross-view gait based human identification with deep cnns. IEEE transactions on pattern analysis and machine intelligence 2016, 39 (2), 209-226.

[17] Taigman, Y.; Yang, M.; Ranzato, M. A.; Wolf, L. In Deepface: Closing the gap to human-level performance in face verification, Proceedings of the IEEE conference on computer vision and pattern recognition, 2014; pp 1701-1708.

[18] Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. In Learning spatiotemporal features with 3d convolutional networks, Proceedings of the IEEE international conference on computer vision, 2015; pp 4489-4497.

[19] Diba, A.; Fayyaz, M.; Sharma, V.; Karami, A. H.; Arzani, M. M.; Yousefzadeh, R.; Van Gool, L., Temporal 3d convnets: New architecture and transfer learning for video classification. arXiv preprint arXiv:1711.08200 2017.

[20] Wolf, T.; Babaee, M.; Rigoll, G. In Multi-view gait recognition using 3D convolutional neural networks, 2016 IEEE International Conference on Image Processing (ICIP), IEEE: 2016; pp 4165-4169.

[21] Thapar, D.; Nigam, A.; Aggarwal, D.; Agarwal, P. In VGR-net: A view invariant gait recognition network, 2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA), IEEE: 2018; pp 1-8.

[22] Iwama, H.; Okumura, M.; Makihara, Y.; Yagi, Y., The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Transactions on Information Forensics and Security 2012, 7 (5), 1511-1521.

[23] Bottou, L.; Bousquet, O. In The tradeoffs of large scale learning, Advances in neural information processing systems, 2008; pp 161-168.

[24] Le, H.; Borji, A., What are the receptive, effective receptive, and projective fields of neurons in convolutional neural networks? arXiv preprint arXiv:1705.07049 2017.

[25] Ioffe, S.; Szegedy, C., Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 2015.

[26] Wu, Y.; He, K. In Group normalization, Proceedings of the European Conference on Computer Vision (ECCV), 2018; pp 3-19.