2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops

GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Networks

Shiqi Yu, Haifeng Chen,
Computer Vision Institute,
College of Computer Science
and Software Engineering,
Shenzhen University, China.
shiqi.yu@szu.edu.cn
chenhaifeng@email.szu.edu.cn

Edel B. García Reyes
Advanced Technologies
Application Center,
7ma A 21406, Playa,
Havana, Cuba.
egarcia@cenatav.co.cu

Norman Poh
Department of Computer Science,
University of Surrey,
Guildford, Surrey,
GU2 7XH, United Kingdom.
n.poh@surrey.ac.uk

Abstract

The performance of gait recognition can be adversely affected by many sources of variation such as view angle, clothing, presence of and type of bag, posture, and occlusion, among others. In order to extract invariant gait features, we proposed a method named as GaitGAN which is based on generative adversarial networks (GAN). In the proposed method, a GAN model is taken as a regressor to generate invariant *gait images that is side view images with normal clothing and without carrying bags. A unique advantage of this approach is that the view angle and other variations are not needed before generating invariant gait images. The most important computational challenge, however, is to address how to retain useful identity information when generating the invariant gait images. To this end, our approach differs from the traditional GAN which has only one discriminator in that GaitGAN contains two discriminators. One is a fake/real discriminator which can make the generated gait images to be realistic. Another one is an identification discriminator which ensures that the the generated gait images contain human identification information. Experimental results show that GaitGAN can achieve state-of-the-art performance. To the best of our knowledge this is the first gait recognition method based on GAN with encouraging results. Nevertheless, we have identified several research directions to further improve GaitGAN.*

1. Introduction

Gait is a behavioural biometric modality with a great potential for person identification because of its unique advantages such as being contactless, hard to fake and passive in nature which requires no explicit cooperation from the subjects. Furthermore, the gait features can be captured at a distance in uncontrolled scenarios. Therefore, gait recognition has potentially wide application in video surveillance. Indeed, many surveillance cameras has been installed in major cities around world. With improved accuracy, the gait recognition technology will certainly be another useful tools for crime prevention and forensic identification. Therefore, gait recognition is and will become an ever more important research topic in the computer vision community.

Unfortunately, automatic gait recognition remains a challenging task because there are many variations that can alter the human appearance drastically, such as view, clothing, variations in objects being carried. These variations can affect the recognition accuracy greatly. Among these variations, view is one of the most common one because we cannot control the walking directions of subjects in real applications, which is the central focus of our work here.

Early literature reported in [9] uses static body parameters measured from gait images as a kind of view invariant feature. Kale *et al.* [10] used the perspective projection model to generated side view features from arbitrary views. Unfortunately, the the relation between two views is hard to be modelled by a simple linear model, such as the perspective projection model.

Some other researchers employed more complex models to handle this problem. The most commonly used model is the view transformation model (VTM) which can transform gait feature from one view to another view. Makihara *et al.* [15] designed a VTM named as FD-VTM, which can work in the frequency-domain. Different from FD-VTM, RSVD-VTM proposed in [11] operates in the spatial domain. It uses reduced SVD to construct a VTM and optimized Gait Energy Image (GEI) feature vectors based on linear discriminant analysis (LDA), and has achieved a relatively good improvement. Motivated by the capability of robust principal component analysis (RPCA) for feature ex-

2160-7516/17 $31.00 © 2017 IEEE
DOI 10.1109/CVPRW.2017.80

traction, Zheng *et al.* [23] established a robust VTM via RPCA for view invariant feature extraction. Kusakunniran *et al.* [12] considered view transformation as a regression problem, and used the sparse regression based on the elastic net as the regression function. Bashir *et al.* [1] formulated a gaussian process classification framework to estimate view angle in probe set, then uses canonical correlation analysis(CCA) to model the correlation of gait sequences from different, arbitrary views. Luo *et al.* [14] proposed a gait recognition method based on partitioning and CCA. They separated GEI image into 5 non-overlapping parts, and for each part they used CCA to model the correlation. In [19], Xing *et al.* also used CCA; but they reformulated the traditional CCA to deal with high-dimensional matrix, and reduced the computational burden in view invariant feature extraction. Lu *et al.* [13] proposed one method which can handle arbitrary walking directions by cluster-based averaged gait images. However, if there is no views with similar walking direction in the gallery set, the recognition rate will decrease.

For most VTM related methods, one view transformation model [2, 3, 11, 12, 15, 23] can only transform one specific view angle to another one. The model heavily depends on the accuracy of view angle estimation. If we want to transform gait images from arbitrary angles to a specific view, a lot of models are needed. Some other researchers also tried to achieve view invariance using only one model, such as Hu *et al.* [8] who proposed a method named as ViDP which extracts view invariant features using a linear transform. Wu *et al.* [18] trained deep convolution neural networks for any view pairs and achieved high accuracies.

Besides view variations, clothing can also change the human body appearance as well as shape greatly. Some clothes, such as long overcoats, can occlude the leg motion. Carrying condition is another factor which can affect feature extraction since it is not easy to segment the carried object from a human body in images. In the literature, there are few methods that can handle clothing invariant in gait recognition unlike their view invariant counterparts. In [7] , clothing invariance is achieved by dividing the human body into 8 parts, each of which is subject to discrimination analysis. In [5], Guan *et al.* proposed a random subspace method (RSM) for clothing-invariant gait recognition by combining multiple inductive biases for classification. One recent method named as SPAE in [21] can extract invariant gait feature using only one model.

In the paper, we propose to use generative adversarial networks (GAN) as a means to solve the problems of variations due to view, clothing and carrying condition simultaneously using only one model. GAN is inspired by two person zero-sum game in Game Theory, developed by Goodfellow *et al.* [4] in 2014, which is composed of one generative model G and one discriminative model D. The generative model captures the distribution of the training data, and the discriminative model is a second classifier that determines whether the input is real or generated. The optimization process of these two models is a problem of minimax two-player game. The generative model produces a realistic image from an input random vector z. As we know the early GAN model is too flexible in generating image. In [16], Mirza *et al.* fed a conditional parameter y into both the discriminator and generator as additional input layer to increase the constraint. Meanwhile, Denton *et al.* proposed a method using a cascade of convolutional networks within a Laplacian pyramid framework to generate images in a coarse-to-fine fashion. As a matter of fact, GANs are hard to train and the generators often produce nonsensical outputs. A method named Deep Convolutional GAN [17] proposed by Radford *et al.*, which contains a series of strategies such as using fractional-strided convolutions and batch normalization to make the GAN more stable in training. Recently, Yoo *et al.* [20] presented an image-conditional generation model which contains a vital component named domain-discriminator. This discriminator ensures that a generated image is relevant to its input image. Furthermore, this method proposes domain transfer using GANs at the pixel level; and is subsequently known as pixel-level domain transfer GAN, or PixelDTGAN in [20].

In the proposed method, GAN is taken as a regressor. The gait data captured with multiple variations can be transformed into the side view without knowing the specific angles, clothing type and the object carried. This method has great potential in real scenes.

The rest of the paper is organized as follows. Section 2 describes the proposed method. Experiments and evaluation are presented in Section 3. The last section, Section 4, gives the conclusions and identifies future work.

2. Proposed method

To reduce the effect of variations, GAN is employed as a regressor to generate *invariant* gait images, that contain side view giat images with normal clothing and without carrying objects. The gait images at arbitrary views can be converted to those at the side view since the side view data contains more dynamic information. While this is intuitively appealing, a key challenge that must be address is to preserve the human identification information in the generated gait images.

The GaitGAN model is trained to generate gait images with normal clothing and without carrying objects at the side view by data from the training set. In the test phase, gait images are sent to the GAN model and invariant gait images contains human identification information are generated. The difference between the proposed method and most other GAN related methods is that the generated image here can help to improve the discriminant capability, not just generate some images which just looks realistic. The most challenging thing in the proposed method is to preserve human identification when generating realistic gait images.

2.1. Gait energy image

The gait energy image [6] is a popular gait feature, which is produced by averaging the silhouettes in one gait cycle in a gait sequence as illustrated in Figure 1. GEI is well known for its robustness to noise and its efficient computation. The pixel values in a GEI can be interpretted as the probability of pixel positions in GEI being occupied by a human body over one gait cycle. According to the success of GEI in gait recognition, we take GEI as the input and target image of our method. The silhouettes and energy images used in the experiments are produced in the same way as those described in [22].

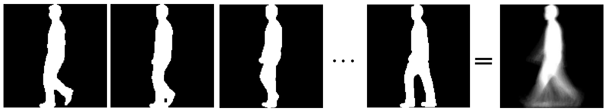


Figure 1: A gait energy image (the right most one) is produced by averaging all the silhouettes (all the remaining images on the left) in one gait cycle .

2.2. Generative adversarial networks for pixel-level domain transfer

Generative adversarial networks (GAN) [4] are a branch of unsupervised machine learning, implemented by a system of two neural networks competing against each other in a zero-sum game framework. A generative model G that captures the data distribution. A discriminative model D then takes either a real data from the training set or a fake image generated from model G and estimates the probability of its input having come from the training data set rather than the generator. In the GAN for image data, the eventual goal of the generator is to map a small dimensional space z to a pixel-level image space with the objective that the generator can produce a realistic image given an input random vector z. Both G and D could be a non-linear mapping function. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation.

The input of the generative model can be an image instead of a noise vector. GAN can realize pixel-level domain transfer between input image and target image such as PixelDTGAN proposed by Yoo *et al.* [20]. PixelDTGAN can transfer a visual input into different forms which can then be visualized through the generated pixel-level image. In this way, it simulates the creation of mental images from visual scenes and objects that are perceived by the human eyes. In that work, the authors defined two domains, a source domain and a target domain. The two domains are connected by a semantic meaning. For instance, the source domain is an image of a dressed person with variations in pose and the target domain is an image of the person's shirt. So PixelDTGAN can transfer an image from the source domain which is a photo of a dressed person to the pixel-level target image of shirts. Meanwhile the transferred image should look realistic yet preserving the semantic meaning. The framework consists of three important parts as illustrated in Figure 2. While the real/fake discriminator ensures that the generated images are realistic, the domain discriminator, on the other hand, ensures that the generated images contain semantic information.

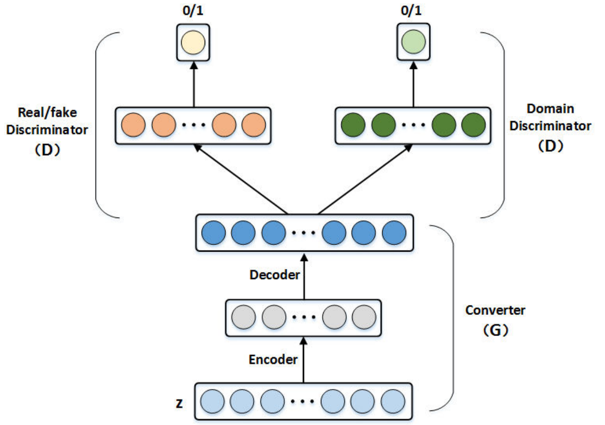


Figure 2: The framework of PixelDTGAN which consists of three important parts.

The first important component is a pixel-level converter which are composed of an encoder for semantic embedding of a source image and a decoder to produce a target image. The encoder and decoder are implemented by convolution neural networks. However, training the converter is not straightforward because the target is not deterministic. Consequently, on the top of converter it needs some strategies like loss function to constrain the target image produced. Therefore, Yoo *et al.* connected a separate network named domain discriminator on top of the converter. The domain discriminator takes a pair of a source image and a target image as input, and is trained to produce a scalar probability of whether the input pair is associated or not. The loss function L_A^D in [20] for the domain discriminator D_A is defined as

$$L_A^D(I_S, I) = -t \cdot log[D_A(I_S, I)] + (t-1) \cdot log[1 - D_A(I_S, I)],$$

$$s.t. \quad t = \begin{cases} 1 & if\ I = I_T \\ 0 & if\ I = \hat{I}_T \\ 0 & if\ I = I_T^- . \end{cases} \qquad (1)$$

where I_S is the source image, I_T is the ground truth target, I_T^- the irrelevant target, and $\hat{I}_T$ is the generated image from converter.

Another component is the real/fake discriminator which similar to traditional GAN in that it is supervised by the labels of real or fake, in order for the entire network to produce realistic images. Here, the discriminator produces a scalar probability to indicate if the image is a real one or not. The discriminator 's loss function L_R^D, according to [20], takes the form of binary cross entropy:

$$L_R^D(I) = -t \cdot log[D_R(I)] + (t-1) \cdot log[1 - D_R(I)],$$
$$s.t. \quad t = \begin{cases} 1 & if\ I \in \{I^i\} \\ 0 & if\ I \in \{\hat{I}^i\}. \end{cases} \tag{2}$$

where $\{I^i\}$ contains real training images and $\{\hat{I}^i\}$ contains fake images produced by the generator.

Labels are given to the two discriminators, and they supervise the converter to produce images that are realistic while keeping the semantic meaning.

2.3. GaitGAN: GAN for gait gecognition

Inspired by the pixel-level domain transfer in PixelDTGAN, we propose GaitGAN to transform the gait data from any view, clothing and carrying conditions to the invariant view that contains side view with normal clothing *and* without carrying objects. Additionally, identification information is preserved.

We set the GEIs at all the viewpoints with clothing and carrying variations as the source and the GEIs of normal walking at 90° (side view) as the target, as shown in Figure 3. The converter contains an encoder and a decoder as shown in Figure 4.

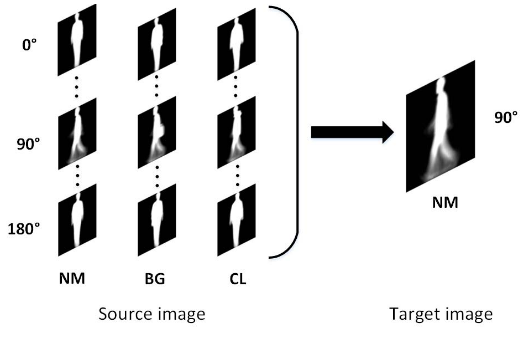


Figure 3: The source images and the target image. 'NM' stands for the normal condition, 'BG' is for carrying a bag and 'CL' is for dressing in a coat as defined in CASIA-B database.

There are two discriminators. The first one is a real/fake discriminator which is trained to predict whether an image is real. If the input GEI is from real gait data at 90° view in normal walking, the discriminator will output 1. Otherwise, it will output 0. The structure of the real/fake discriminator is shown in Figure 5.

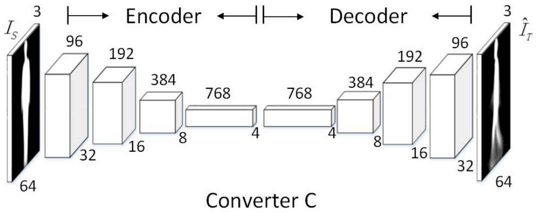


Figure 4: The structure of the converter which transform the source images to a target one as shown in Figure 3.

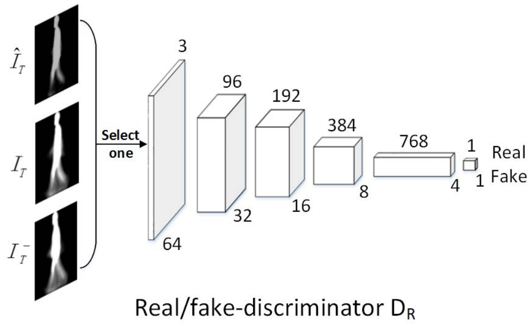


Figure 5: The structure of the Real/fake-discriminator. The input data is target image or generated image.

With the real/fake discriminator, we can only generate side view GEIs which look well. But, the identification information of the subjects may be lost. To preserve the identification information, another discriminator, named as identification discriminator, which is similar to the domain discriminator in [20] is involved. The identification discriminator takes a source image and a target image as input, and is trained to produce a scalar probability of whether the input pair is the same person. If the two inputs source images are from the same subject, the output should be 1. If they are source images belonging to two different subjects, the output should be 0. Likewise, if the input is a source image and the target one is generated by the converter, the discriminator function should output 0. The structure of identification discriminator is shown in Figure 6.

3. Experiments and analysis

3.1. Dataset

CASIA-B gait dataset [22] is one of the largest public gait databases, which was created by the Institute of Automation, Chinese Academy of Sciences in January 2005. It consists of 124 subjects (31 females and 93 males) captured from 11 views. The view range is from 0° to 180°

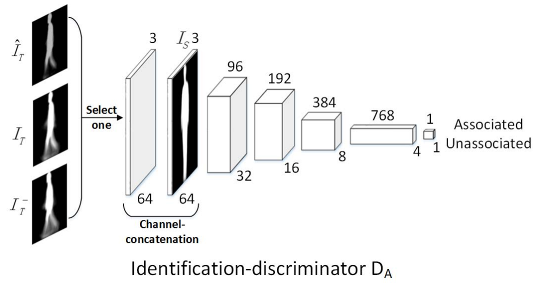


Figure 6: The structure of the identification-discriminator. The input data is two image concatenated by channel.

with 18° interval between two nearest views. There are 11 sequences for each subject. There are 6 sequences for normal walking (”nm”), 2 sequences for walking with a bag (”bg”) and 2 sequences for walking in a coat (”cl”).

3.2. Experimental design

In our experiments, all the three types of gait data including ”nm”, ”bg” and ”cl” are all involved. We put the six normal walking sequences, two sequences with coat and two sequences containing walking with a bag of the first 62 subjects into the training set and the remaining 62 subjects into the test set. In the test set, the first 4 normal walking sequences of each subjects are put into the gallery set and the others into the probe set as it is shown in Table 1. There are four probe sets to evaluate different kind of variations.

Table 1: The experimental design.

Training set		ID: 001-062 nm01-nm06, bg01, bg02, cl01, cl02
Gallery set		ID: 063-124 nm01-nm04
Probe set	ProbeNM	ID: 063-124 nm05, nm06
	ProbeBG	ID: 063-124 bg01, bg02
	ProbeCL	ID: 063-124 cl01, cl02
	ProbeALL	ID: 063-124 nm05, nm06 bg01, bg02, cl01, cl02

3.3. Model parameters

In the experiments, we used a similar setup to that of [20], which is shown in Figure 4. The converter is a unified network that is end-to-end trainable but we can divide it into two parts, an encoder and a decoder. The encoder part is composed of four convolutional layers to abstract the source into another space which should capture the personal attributes of the source as well as possible. Then the resultant feature z is then fed into the decoder in order to construct a relevant target through the four decoding layers. Each decoding layer conducts fractional stride convolutions, where the convolution operates in the opposite direction. The details of the encoder and decoder structures are shown in Table 2 and 3. The structure of the real/fake discriminator and the identification discriminator are similar to the encoder's first four convolution layers. The layers of the discriminators are all convolution layers.

Table 2: Details of the encoder. The first four layers of encoder is the same as real/fake discriminator and domain discriminator. After Conv.4, the real/fake and identification discriminator connect Conv.5 to output binary value

Layers	Number of filters	Filter size	Stride	Batch norm	Activation function
Conv.1	96	4× 4× {1,1,2}	2	N	L-ReLU
Conv.2	192	4× 4× 96	2	Y	L-ReLU
Conv.3	384	4× 4× 192	2	Y	L-ReLU
Conv.4	768	4× 4× 384	2	Y	L-ReLU

Table 3: Details of the decoder. F denotes fractional-stride.

Layers	Number of filters	Filter size	Stride	Batch norm	Activation function
F-Conv.1	768	4× 4× 384	1/2	Y	L-ReLU
F-Conv.2	384	4× 4× 192	1/2	Y	L-ReLU
F-Conv.3	192	4× 4× 96	1/2	Y	L-ReLU
F-Conv.4	96	4× 4× 1	1/2	N	Tanh

Normally to achieve a good performance using deep learning related methods, a large number of iterations in training are needed. From Figure 7, we can find that more iterations can indeed result in a higher recognition rate, but the rate peaks at around 450 epoches. So in our experiments, the training was stopped after 450 epoches.

3.4. Experimental results on CASIA-B dataset

To evaluate the robustness of the proposed GaitGAN, three sources of variations have been evaluated, covering view, clothing, and carrying objects. The performance of the model under these conditions are shown in Tables 4-6.

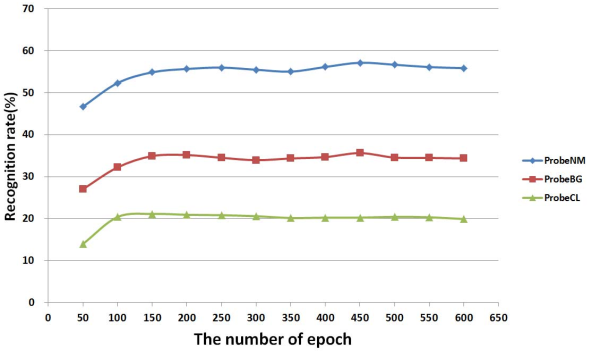


Figure 7: The recognition rate as a function of the number of iterations. The blue, red and green respectively corresponds to normal walking (ProbeNM), walking with a bag (ProbeBG) and walking in a coat (ProbeCL) sequences respectively in the probe set.

Table 4: Recognition Rate of ProbeNM in experiment 2, training with sequences containing three conditions.

		Probe set view(Normal walking, nm05,nm06)										
		0	18	36	54	72	90	108	126	144	162	180
Gallery set view	0	100.0	79.03	45.97	33.87	28.23	25.81	26.61	25.81	31.45	54.84	72.58
	18	78.23	99.19	91.94	63.71	46.77	38.71	37.90	44.35	43.55	65.32	58.06
	36	56.45	88.71	97.58	95.97	75.00	57.26	59.68	72.58	70.16	60.48	35.48
	54	33.87	53.23	85.48	95.97	87.10	75.00	75.00	77.42	63.71	37.10	22.58
	72	27.42	41.13	69.35	83.06	100.0	96.77	89.52	73.39	62.10	37.10	17.74
	90	22.58	37.10	54.84	74.19	98.39	98.39	96.77	75.81	57.26	35.48	21.77
	108	20.16	32.26	58.06	76.61	90.32	95.97	97.58	95.97	74.19	38.71	22.58
	126	29.84	37.90	66.94	75.00	81.45	79.03	91.13	99.19	97.58	59.68	37.10
	144	28.23	45.97	60.48	66.94	61.29	59.68	75.00	95.16	99.19	79.84	45.97
	162	48.39	63.71	60.48	50.00	41.13	33.87	45.16	65.32	83.06	99.19	71.77
	180	73.39	56.45	36.29	25.81	21.77	19.35	20.16	31.45	44.35	72.58	100.0

Table 5: Recognition Rate of ProbeBG in the experiment 2 training with sequences containing three conditions.

		Probe set view(walking with a bag, bg01,bg02)										
		0	18	36	54	72	90	108	126	144	162	180
Gallery set view	0	79.03	45.97	33.06	14.52	16.13	14.52	11.29	15.32	22.58	33.87	41.13
	18	54.84	76.61	58.87	31.45	26.61	16.13	24.19	29.84	32.26	41.94	32.26
	36	36.29	58.87	75.81	53.23	44.35	30.65	34.68	46.77	42.74	34.68	20.16
	54	25.00	45.16	66.13	68.55	57.26	42.74	41.13	45.97	40.32	20.16	13.71
	72	20.16	24.19	38.71	41.94	65.32	56.45	57.26	51.61	39.52	16.94	8.87
	90	15.32	27.42	37.90	38.71	62.10	64.52	62.10	61.29	38.71	20.97	12.10
	108	16.13	25.00	41.13	42.74	58.87	58.06	69.35	70.16	53.23	24.19	11.29
	126	19.35	29.84	41.94	45.16	46.77	52.42	58.06	73.39	66.13	41.13	22.58
	144	26.61	32.26	48.39	37.90	37.10	36.29	38.71	67.74	73.39	50.00	32.26
	162	29.03	34.68	36.29	25.00	19.35	16.13	20.16	37.90	51.61	76.61	41.94
	180	42.74	28.23	24.19	12.90	11.29	11.29	14.52	21.77	30.65	49.19	77.42

Table 6: Recognition Rate of ProbeCL in the experiment 2 training with sequences containing three conditions.

		Probe set view(walking wearing a coat, cl01,cl02)										
		0	18	36	54	72	90	108	126	144	162	180
Gallery set view	0	25.81	16.13	15.32	12.10	6.45	6.45	9.68	7.26	12.10	11.29	15.32
	18	17.74	37.90	34.68	20.97	13.71	8.87	12.10	19.35	16.94	24.19	19.35
	36	13.71	24.19	45.16	43.55	30.65	19.35	16.94	22.58	28.23	20.16	10.48
	54	2.42	19.35	37.10	55.65	39.52	22.58	29.03	29.84	29.84	16.94	8.06
	72	4.84	12.10	29.03	40.32	43.55	34.68	32.26	28.23	33.87	12.90	8.06
	90	4.03	10.48	22.58	31.45	50.00	48.39	43.55	36.29	31.45	13.71	8.06
	108	4.03	12.90	27.42	27.42	38.71	44.35	47.58	38.71	32.26	15.32	4.84
	126	10.48	10.48	23.39	27.42	26.61	25.81	37.10	45.97	41.13	15.32	10.48
	144	8.87	13.71	26.61	22.58	18.55	19.35	21.77	35.48	43.55	20.97	12.90
	162	14.52	18.55	20.97	17.74	12.10	12.10	17.74	21.77	37.10	35.48	21.77
	180	17.74	13.71	11.29	6.45	10.48	5.65	6.45	5.65	14.52	29.03	27.42

For Table 4, the first four normal sequences at a specific view are put into the gallery set, and the last two normal sequences at another view are put into the probe set. Since there are 11 views in the database, there are 121 pairs of combinations. In each table, each row corresponds to a view angle of the gallery set, whereas each column corresponds to the view angle of the probe set. The recognition rates of these combinations are listed in Table 4. For the results in Table 5, the main difference with those in Table 4 are the probe sets. The probe data contains images of people carrying bags, and the carrying conditions are different from that of the gallery set. The probe sets for Table 6 contain gait data with coats.

3.5. Comparisons with GEI+PCA and SPAE

Since GEIs are used as input trying to extract invariant features, we first compare our method with GEI+PCA [6] and SPAE [21]. The experiment protocols in terms of the gallery and probe sets for GEI+PCA and SPAE are exactly the same as those presented in Table. 1. Due to limited space, we only list 4 probe angles with a 36° interval. Each row in this figure represents a probe angle. The compared angles are 36°, 72°, 108° and 144°. The first column of Figure 8 compares the recognition rates of the proposed GaitGAN with GEI+PCA and SPAE at different probe angles in normal walking sequences. The second column shows the comparison with different carrying conditions, and the third shows the comparison with different clothings. As illustrated in Figure 8, the proposed method outperforms GEI+PCA at all probe angle and gallery angle pairs. Meantime, its performance is comparable to that of SPAE and better than SPAE at many points. The results show that the proposed method can produce features that as robust as the state of the art methods in the presence of view, clothing and carrying condition variations.

We also compared the recognition rates without view variation. This can be done by taking the average of the rates on the diagonal of Table 4, Table 5 and Table 6. The corresponding average rates of GEI+PCA and SPAE are also obtained in the same manner. The result are shown in Figure 9. When there is no variation, the proposed method achieve a high recognition rate which is better than GEI+PCA and SPAE. But when variation exists, the proposed method outperforms GEI+PCA greatly.

3.6. Comparisons with state-of-the-art

In order to better analyse the performance of the proposed method, we further compare the proposed GaitGAN with additional state-of-the-art methods including FD-VTM [15], RSVD-VTM [11], RPCA-VTM [23], R-VTM [12], GP+CCA [1] , C3A [19] and SPAE [21].

The probe angles selected are 54°, 90° and 126° as in experiments of those methods. The experimental results are listed in Figure 10. From the results we can find that the proposed method outperforms others when the angle difference between the gallery and the probe is large. This shows that

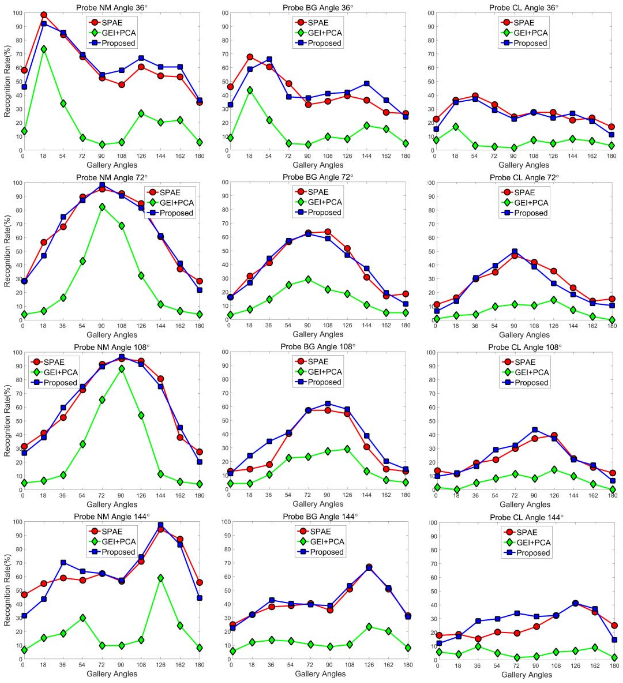


Figure 8: Comparison with GEI+PCA and SPAE at different probe angle. Each row represents a probe angle and each column represents different probe sequences. The blue lines are achieved by proposed method.

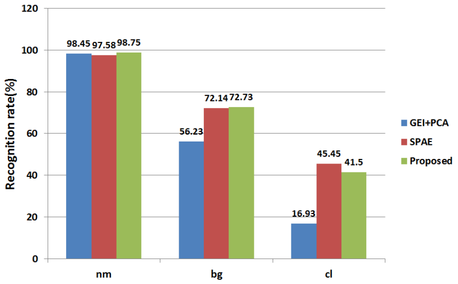
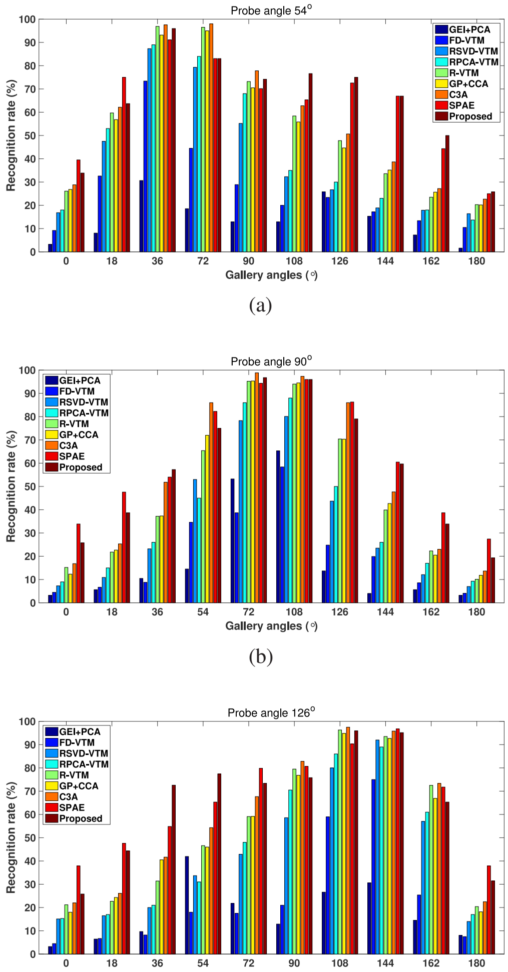


Figure 9: The average recognition rate on the diagonal compare with GEI+PCA and SPAE at three condition. The green bars are achieved by the proposed method.



(a)

(b)

(c)

Figure 10: Comparing with existing methods at probe angles (a)54°, (b)90° and (c)126°. The gallery angles are the rest 10 angles except the corresponding probe angle.

Table 7: Average recognition rates at probe angles 54°, 90° and 126°. The gallery angles are the rest 10 angles except the corresponding probe angle. The values in the right most column are the averages rate at the three probe angles 54°, 90° and 126°.

Method	Probe angle			
	54°	90°	126°	Average
C3A [19]	56.64%	54.65%	58.38%	56.56%
ViDP [8]	64.2%	60.4%	65.0%	63.2%
CNN [18]	77.8%	64.9%	76.1%	72.9%
SPAE [21]	63.31%	62.1%	66.29%	63.9%
Proposed	64.52%	58.15%	65.73%	62.8%

the model can handle large viewpoint variation well. When the viewpoint variation is not large enough, the proposed method can also improve the recognition rate obviously.

In Table 7, the experimental results of C3A [19], ViD-P [8], CNN [18], SPAE [21] and the proposed method are listed. Here we want to emphasis that the proposed method obtains comparable results using only one generative model for any views, and for clothing or carrying condition variations, simultaneously. Meanwhile, this method is the first use of GAN for gait recognition and the experimental results show that GAN is feasible for gait recognition under significant variations.

4. Conclusions and future work

In this paper, we proposed to apply PixelDTGAN which is a variant of generative adversarial networks to deal with variations in viewpoint, clothing and carrying conditions simultaneously in gait recognition. The resultant model is known as GaitGAN. Extensive experiments based on the CASIA-B database shows that the GaitGAN can transform gait images obtained from any viewpoint to the side view, from abnormal walking sequences to normal walking sequences without the need to estimate the subject's view angle, clothing type and carrying condition beforehand. Experimental results show that the recognition rate of proposed model is comparable to that of the state-of-the-art methods. Indeed, GaitGAN is shown to be promising for practical applications in video surveillance.

There are however, a number of limitations which need to be addressed in future work. The number of database samples currently used is relatively small. In the future, we will use a larger database for training. In addition, we can use more complex and powerful networks. In this work, we have merely utilized the generated side view image as the matching feature without exploring intermediate layer features neither others effective feature extraction methods. We believe that better GAN technologies will further improve gait recognition in future. Besides gait recognition, other recognition and classification problems could also benefit from the development of GAN.

Acknowledgment

The authors would like to thank Dr. Xianglei Xing for his support on experimental results of some methods. The work is supported by the Science Foundation of Shenzhen (Grant No. JCYJ20150324141711699).

References

[1] K. Bashir, T. Xiang, and S. Gong. Cross-view gait recognition using correlation strength. In *BMVC*, 2010.

[2] X. Ben, W. Meng, R. Yan, and K. Wang. An improved biometrics technique based on metric learning approach. *Neurocomputing*, 97:44 – 51, 2012.

[3] X. Ben, P. Zhang, W. Meng, R. Yan, M. Yang, W. Liu, and H. Zhang. On the distance metric learning between cross-domain gaits. *Neurocomputing*, 208:153 – 164, 2016. SI: BridgingSemantic.

[4] I. J. Goodfellow, J. Pougetabadie, M. Mirza, B. Xu, D. Wardefarley, S. Ozair, A. Courville, Y. Bengio, Z. Ghahramani, and M. Welling. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 3:2672–2680, 2014.

[5] Y. Guan, C. T. Li, and Y. Hu. Robust clothing-invariant gait recognition. In *2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 321–324, July 2012.

[6] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE TPAMI*, 28(2):316–322, 2006.

[7] M. A. Hossain, Y. Makihara, J. Wang, and Y. Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, 43(6):2281 – 2291, 2010.

[8] M. Hu, Y. Wang, Z. Zhang, J. J. Little, and D. Huang. View-invariant discriminative projection for multi-view gait-based human identification. *IEEE TIFS*, 8(12):2034–2045, Dec 2013.

[9] A. Y. Johnson and A. F. Bobick. A multi-view method for gait recognition using static body parameters. In *Proc. of 3rd International Conference on Audio and Video Based Biometric Person Authentication*, pages 301–311, Jun. 2001.

[10] A. Kale, A. K. R. Chowdhury, and R. Chellappa. Towards a view invariant gait recognition algorithm. In *IEEE Conference on Advanced Video and Signal Based Surveillance*, pages 143–150, July 2003.

[11] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang. Multiple views gait recognition using view transformation model based on optimized gait energy image. In *ICCV Workshops*, pages 1058–1064, 2009.

[12] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Gait recognition under various viewing angles based on correlated motion regression. *IEEE TCSVT*, 22(6):966–980, 2012.

[13] J. Lu, G. Wang, and P. Moulin. Human identity and gender recognition from gait sequences with arbitrary walking directions. *IEEE TIFS*, 9(1):51–61, Jan 2014.

[14] C. Luo, W. Xu, and C. Zhu. Robust gait recognition based on partitioning and canonical correlation analysis. In *IEEE International Conference on Imaging Systems and Techniques*, 2015.

[15] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *ECCV*, pages 151–163, 2006.

[16] M. Mirza and S. Osindero. Conditional generative adversarial nets. *Computer Science*, pages 2672–2680, 2014.

[17] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *Computer Science*, 2015.

[18] 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 TPAMI*, 39(2):209–226, Feb 2017.

[19] X. Xing, K. Wang, T. Yan, and Z. Lv. Complete canonical correlation analysis with application to multi-view gait recognition. *Pattern Recognition*, 50:107–117, 2016.

[20] D. Yoo, N. Kim, S. Park, A. S. Paek, and I. S. Kweon. Pixel-level domain transfer. *arXiv:CoRR*, 1603.07442, 2016.

[21] S. Yu, H. Chen, Q. Wang, L. Shen, and Y. Huang. Invariant feature extraction for gait recognition using only one uniform model. *Neurocomputing*, 239:81–93, 2017.

[22] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *ICPR*, pages 441–444, August 2006.

[23] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan. Robust view transformation model for gait recognition. In *ICIP*, pages 2073–2076, 2011.