Edinburgh Research Explorer

An Approach for Gait Anonymization Using Deep Learning

An Approach for Gait Anonymization Using Deep Learning

Ngoc-Dung T. Tieu[1], Huy H. Nguyen[1], Hoang-Quoc Nguyen-Son[2], Junichi Yamagishi[1,2,3], Isao Echizen[1,2]

[1] *SOKENDAI (The Graduate University for Advanced Studies), Kanagawa, Japan*
[2]*National Institute of Informatics, Tokyo, Japan*
[3]*The University of Edinburgh, Edinburgh, UK*
{dungtieu,nhhuy,nshquoc,jyamagis,iechizen@nii.ac.jp}

Abstract—The human gait has become another biometric trait used in security systems because it is unique to each person and can be recognized at a distance. However, a bad actor could use a gait recognition system to identify a person on the basis of his or her gait. We have developed a gait anonymization method that prevents unauthorized gait recognition. It modifies the gait so that the person cannot be identified while maintaining the naturalness of the gait. The modification is done by adding another gait, called "noise gait". A convolutional neural network makes this modification by taking two gaits as input, the original gait and the noise gait, and outputting an anonymized gait. The proposed method was evaluated using the success rate and mean opinion score (MOS). The success rate is the rate of failed gait recognition, and the MOS is a measure of the naturalness of the anonymized gait. In our experiments, the success rate achieved 98.86% at most while the highest naturalness score is 3.73 in the MOS scale. These findings should open new research directions regarding privacy protection related to gait recognition.

Index Terms—gait; biometric trait; security; gait anonymization; deep learning

I. Introduction

The human gait, which represents the manner and pattern of walking, has become an important biometric trait because it is unique to each person [1] and can be recognized at a distance without physical contact or the persons cooperation. It is thus particularly advantageous considering that most other biometrics (face, fingerprint, iris, etc.) can be recognized only at close range or with physical contact. Therefore, gait has become a biometric trait that can be used to identify people at a distance [2]–[6].

Nowadays, due to the utility of the social networks such as Facebook, enormous numbers of internet users can easily upload their photos or videos to the social networks with or without permissions of people captured in the video and share them with anybody else instantly. A serious privacy problem may happen if someone captured in those videos is identified unintentionally by gait recognition systems and if their personal information is revealed eventually. Privacy concerns related to the sensitive information (e.g., ethnicity, gender, age) contained in faces or body shapes have led to the development of methods for anonymizing personal characteristics. To give a few examples, Yamada et al. developed a wearable device that prevents detection by face detection systems [7]. Othman and Ross [8] developed a method for suppressing gender but retaining identity. Ruchaud et al. [9] proposed an approach

for degenderizing while preserving enough information for recognizing body shape and motion.

We have developed a method for anonymizing gaits that prevents a person's gait from being identified with a gait recognition system. Our aim is to enable internet users to upload and share videos safely while maintaining the original appearance of the videos as much as possible. It could be implemented in a social network, for example, as a utility that enables the sharing of videos while preventing the revelation of personal information obtained from gait biometrics. Another potential application is police video redaction. Often times when a video of a suspect is shown on television, the face is blurred, but not the rest of the body. If the suspect is walking, it may be possible to identify the person from his or her gait. Anonymization of the gait would prevent this. The scenario of the gait anonymization is shown in Fig.1.
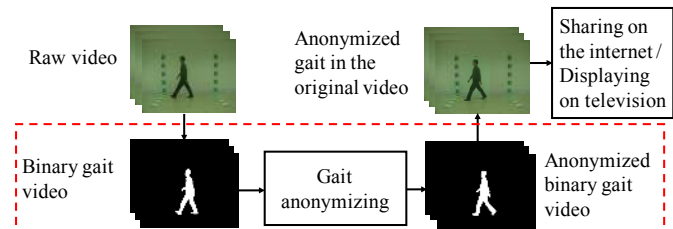
Fig. 1: Scenario of gait anonymization.

In this research, we focused on anonymizing gaits in silhouette sequences, as indicated by the dashed rectangular box in Fig.1. The main idea of our method is to change the shape of a gait by adding another gait, a "noise gait". This is done using the contour coordinates of the gait in each frame instead of the gait image; these coordinates are converted into a vector. We designed a convolutional neural network (CNN) that takes two inputs, the contour vector of the original gait and that of the noise gait, and outputs a modified contour vector. An anonymized gait is generated from this modified vector.

We evaluate the proposed method on the CASIA-B gait dataset [10] using two metrics: the success rate and the mean opinion score (MOS). The success rate is the rate of failed gait recognition, and the MOS is a measure of the naturalness of the anonymized gait. To evaluate the success rate, we used the gait recognition system developed by Zheng et al. [5],

which finds the gait in the database that is most similar to the query gait. It partially overcomes the cross-view problem, so even if the gait view angle differs from that in the database image, it may still recognize the gait. In our experiment, the highest identification ratio based on Zheng's method was 89%. For the later metric, we asked 30 subjects to do the MOS test. To obtain the MOS, we asked 30 people to evaluate the anonymized gaits.

The contributions of our work include

- The introduction of a new study direction, gait anonymization, for preventing personal information from being revealed through unauthorized use of gait recognition.
- The introduction of a CNN that anonymizes a person's gait by adding another gait, the noise gait.
- The suggestion of using the proposed method is not only for gait anonymization but also for other object image anonymization, which will be investigated in future work.

II. RELATED WORK

A. Gait Recognition Systems

Gait recognition systems aim to recognize a person on the basis of their manner and pattern of walking. They predict the identity of a person from a probe sample by comparing it to the gaits in a gallery composed of registered gait samples. In the current state-of-the-art, there are two mainstream approaches to gait recognition: model-free and model-based. The model-based approaches use a series of dynamic or static parameters of body parts, such as arms, legs, limbs, and thighs, while the model-free approaches use either silhouettes or the average silhouette. Generally, the model-free approaches tend to be less sensitive to the quality of the gait sequence and have lower computational cost [2]. One of the most challenging tasks of gait recognition is handling the multi-view angle problem, i.e., when the view angle of the probe gait is not the same as that of the gallery gait. Zheng et al. [5] proposed a robust, easy-to-implement, and rapid method that transforms the feature of the gait in the probe view into that of the gait in the gallery view. The rest of this section briefly summarizes this method. In their method, the gait energy image (GEI) obtained from a silhouette sequence is used as the feature of the gait. The GEI was defined by Han and Bhanu [4] as the average silhouette in a sequence of silhouettes:

$$g(x,y) = \frac{1}{T} \sum_{t=1}^{T} I_t(x,y), \qquad (1)$$

where T is the number of frames in the silhouette sequence, I_t is the silhouette image at frame t, and x and y are values in the $2D$ image coordinate. An example silhouette sequence and its GEI are shown in Fig.2.

There are two separate processes in their method, gait registration and gait recognition. In each process, they apply a supervised dimension reduction approach named Partition Least Square as a feature extraction method on the original gait feature, i.e., the GEI. After the features are extracted in

Fig. 2: Example silhouette sequence (left side), and corresponding GEI (right side).

the registration process, singular value decomposition is used to construct a vector transform model (VTM). The VTM is used in the recognition process to transform the view angle of the probe gait to that of the gallery gait. Finally, gait similarity is measured by using the L1-norm distance between the gait features of the two gaits which are in the same view angle. The smaller the distance value, the greater the similarity.

B. Gait Spoofing and Anti-spoofing

Another line of research has been developing methods for defending against attacks on these systems. Bustard et al. [11] examined the effects of two types of spoofing attacks against two gait recognition systems. One type is a clothing impersonation attack in which an imposter replicates the clothing of an enrolled individual. The other type is a targeted attack in which an imposter selects the enrolled individual whose gait signature is closest to that of the attacker. Testing showed that both systems could be spoofed, especially when the types of attack were combined. Bustard et al. also proposed a countermeasure using body-part-based gait analysis. Testing showed that it reduced the false acceptance rate for both gait recognition systems.

C. Generative Models

A recent research direction is the use of deep learning for feature representation, classification, and object synthesis. Deep learning can also be used to generate various types of objects. Motion and image generation are particularly relevant to our research in a broader sense. Holden et al. [12] proposed a deep learning framework that enables animators to synthesize and edit motion animation using a skeleton structure dataset. Dosovitskiy et al. [13] proposed a CNN that generates images of objects given their style, viewpoint, and color. Generative adversarial networks (GANs), which were first introduced by Goodfellow et al. [14], have been actively investigated for image generation. A GAN includes two deep neural networks working against each other - one generates data fitting the data distribution of interest, and the other tries to discriminate the generated data from the true data. The GAN-based approach has been shown to generate very natural images (e.g., [15], [16]).

Among the various GAN-based image synthesis methods, a variant of GAN called "auxiliary classifier GAN" [17] is most relevant to our approach because its discriminator needs to identify the class label of the image and then determine whether the image is natural or fake. While our CNN-based approach is related to these other approaches, it clearly differs as we want to generate a natural motion image that prevents gait recognition systems from identifying the person. Although

it might be possible to modify a criterion of the auxiliary classifier GAN, it is relatively difficult to train GAN models [18]. We therefore adopted a relatively simple CNN architecture that modifies the original gait by adding a noise gait. The architecture of the proposed model includes two independent networks that are merged into one to generate a new gait.

III. METHOD

A. Overview

An anonymized gait is generated by using a CNN to mix a noise gait into the original gait. Since we want to change the shape of the gait, we use the contour coordinate of the silhouette instead of the gait image. The three steps are illustrated in Fig.3.

Step 1 - Pre-processing: The two inputs, the original gait and the noise gait, are first pre-processed to extract their silhouette contours. These contours are then transformed into vectors, an original contour vector and a noise contour vector.

Step 2 - CNN: The two contour vectors are input to the CNN, which changes the original gait and outputs a modified contour vector.

Step 3 - Post-processing: The modified contour vector is post-processed to obtain the anonymized gait. This gait is then placed in the original scene.
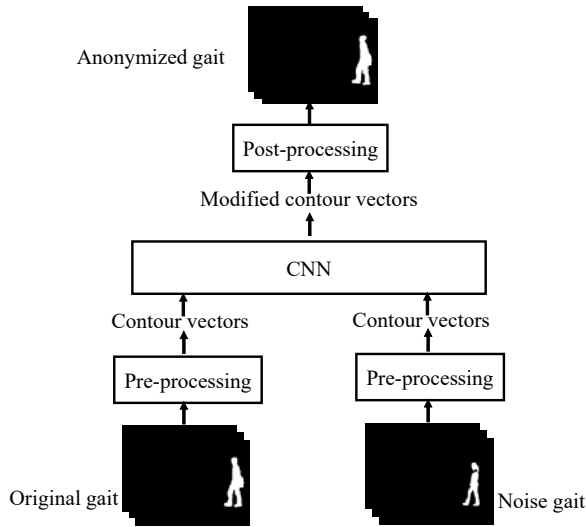

Fig. 3: Overview of proposed method.

B. Pre-processing

The key idea of our proposed gait anonymization is to use the shape of a noise gait to alter the shape of the original gait. Therefore, as illustrated in Fig. 4, there are three steps in pre-processing. First, the regions containing silhouettes are extracted and resized to the same size (240×240). Then, the coordinates of the pixels on the contours of the silhouettes are extracted. Finally, the coordinates for each frame are transformed into a vector, which is input to the CNN in the modification phase. The lengths of the contour vectors are fixed to 4000, which is equivalent to 2000 pixels on

the contour. This is because, in our database, there were no contours with more than 2000 pixels. Each modified silhouette is stored in a vector of length 4000, consisting of two parts: the first part is the row and column coordinates of pixels on the contour, and the second part is a zero-padded area in case there are fewer than 2000 pixels on the contour.
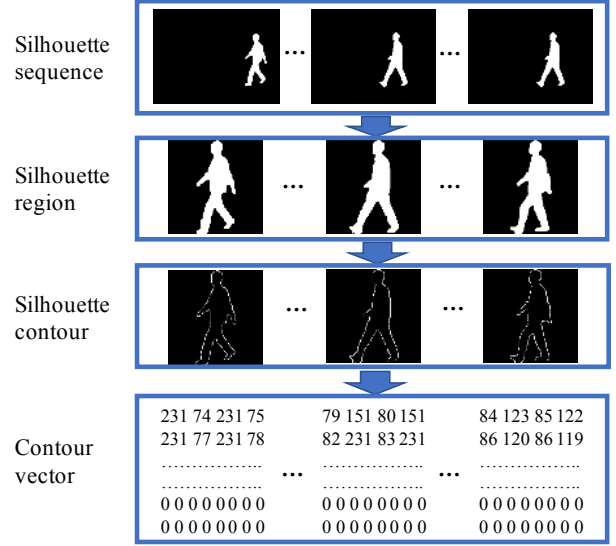

Fig. 4: Pre-processing.

C. Contour Vector Modification

The original gait is modified using one-dimension convolutional networks, as illustrated in Fig. 5. The original contour vector and a noise contour vector are input, and a modified contour vector is output. The two input vectors are abstracted by passing each one through a shared weights network. These networks are concisely formulated as

$$\Phi_1(X_1) = ReLU(W_1 * X_1 + b_1) \tag{2}$$
$$\Phi_1(X_2) = ReLU(W_1 * X_2 + b_1), \tag{3}$$

consisting of convolutional functions (denoted *) of weights matrix W_1 (or filters for convolution layers) and input vectors X_1, X_2, addition of the bias b_1, where X_1, X_2 are respectively contour vector of original and noise gait. These functions are followed by nonlinear operation $ReLU(x) = \max(x, 0)$. Since the aim is to generate a modified the gait, these two independent networks are then merged into one network:

$$\Phi_2(X_1, X_2) = ReLU(W_2 * (\Phi_1(X_1) + \Phi_1(X_2)) + b_2), \tag{4}$$

where W_2 and b_2 are the weight matrix and bias of the merged network. The weights and biases are learned by minimizing the following cost function:

$$\frac{1}{D_{X_1}} \Big(||\Phi_2(X_1, X_2) - X_1||^2 + \alpha ||\Phi_2(X_1, X_2) - X_2||^2 \Big), \tag{5}$$

where the first term is used to preserve the silhouette of the original gait so that the naturalness of the gait images is maintained. The second term is used to modify the original gait

by adding the noise gait, and α is a parameter that specifies how much of the noise gait should be added to the original gait. We set $\alpha = 0.3$ in our experiments. D_{X_1} is the length of a contour vector (4000 in our case). This function is minimized using the stochastic gradient descent algorithm. Training is performed for 100 epochs on a NVIDIA GeForce GTX 1070 GPU.
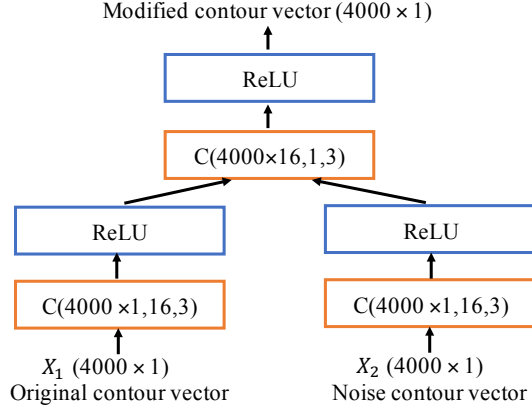

Fig. 5: Architecture of the CNN: C represents one-dimension convolution; string following each C shows number of dimensions of feature maps, number of filters, and size of filters.

D. Post-processing

The purpose of post-processing is to create a video in which the original gait is replaced with the modified one. As illustrated in Fig. 6, a contour image is first created from the modified contour vector. The inside region of this contour is then filled to create a modified silhouette image. Finally, the original silhouette is replaced with the modified one at the same position in the original video.
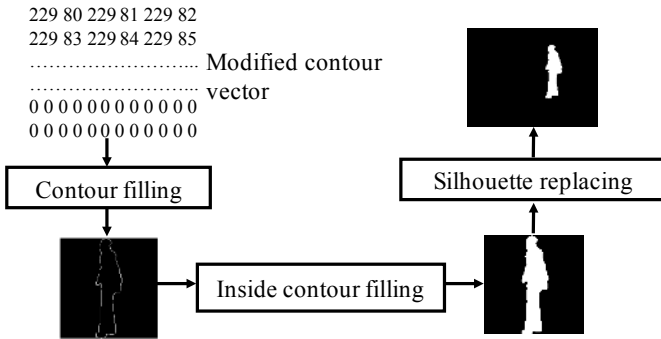

Fig. 6: Post-processing.

E. Noise Gait Selection

An important aspect of our method is selecting the noise gait. The noise gait should anonymize the original gait but should not significantly reduce the naturalness of the original gait. Two cases can be considered: (1) the views of the original and noise gait differ; (2) the views are the same, as illustrated in Figs. 7 and 8, respectively. The first row in each figure is

Fig. 7: Views of original and noise gait differ.

Fig. 8: Views of original and noise gait are the same.

Comparing original and noise gaits shown in the two figures, we can see that the gait looks more natural when the views are the same. Under each noise gait, we show anonymized gaits that differ from the original gaits but are have the same view angle.

III. EVALUATION

To evaluate our method, we used the CASIA-B gait dataset [10]. This dataset contains 124 subjects in total, with 110 sequences (10 sequences for each of 11 viewing angles $(0^0, 18^0, \ldots, 180^0)$) for each subject. In general, our method does not require sequences to have a fixed length, but having the same length for all sequences not only facilitates implementation but also reduces the training time. We thus created a dataset in which the sequences all had a length of 50 frames as we had determined that this was sufficient for evaluating gait naturalness. Starting with the CASIA-B dataset, we removed the sequences shorter than 50; for the sequences longer than 50 frames, we extracted the last 50 frames because we had observed that the quality of the later frames was usually better than that of the earlier frames. This process reduced the number of sequences from 13,640 to 12,989. We used this new dataset to train and test our proposed model against the gait recognition system developed by Zheng et al. [5].

From the 124 subjects, we used the last 24 (2541 sequences) to train the recognition system, the first 10 (1007 sequences) to train our CNN, and the remaining 90 for testing the model (we use D_1, D_2, and D_3 to denote these sets, respectively). After training the CNN, we fed the gaits in D_3 into the CNN to obtain anonymized gaits. We then passed these anonymized gaits through the pre-trained gait recognition system.

From D_2 and D_3, we randomly chose one subject as the noise gait for that set. Fig. 9 shows the results for one person for various view angles, where the first row for each view angle shows the silhouettes of the original gait, and the second row shows the anonymized gait.

Performance was evaluated using the *naturalness* and *success rate* metrics, which are defined below.

A. Naturalness

The mean opinion score (MOS) has been used for decades to measure the quality of media from the users perspective. It was used, for example, by van den Oord et al. to assess user preferences for audio waveforms generated by the WaveNet deep generative model [19] and by Ledig et al. to evaluate the quality of images created using the SRGAN generative adversarial network [15].

We asked 30 people to evaluate the naturalness of the anonymized gaits. We gave each person 30 random pairs of original and anonymized gait videos; the length of each video was 10 seconds. After watching each pair, they rated the naturalness of the anonymized gaits on a five-point scale (1: Bad, 2: Poor, 3: Fair, 4: Good, 5: Excellent). Fig. 10a shows the average MOS values by view angle. The naturalness was lowest (2.94) for a view angle of 180^0 and highest (3.73) for a view angle of 54^0.

B. Success Rate

As mentioned, we defined the success rate as the rate at which the gait recognition system fails to identify the gait. This definition is similar to that of Sharif [20]. We used the gait recognition system presented by Zheng et al. [5].

We use S to denote the set of gaits in D_3 that the gait recognition system recognized and S' to denote the set of gaits obtained by anonymizing the gaits in S.

$$success\_rate(\%) = \frac{M}{|S|} \times 100\%, \qquad (6)$$

where M is the number of gaits in S' that the gait recognition system failed to identify. The success rates for each probe view and gallery view are shown in Table I. The average success rate for each view angle is summarized in Fig. 10b.

From our experimental results, we would like to highlight two points:

(1) As shown in Table I, the success rate was high in some cases but varied depending on the cross-view angle or the difference between the probe and gallery view. This is because the performance of gait recognition is degraded when the gait appearance changes drastically.

(2) A comparison of Figs. 10a and 10b shows that, in general, the success rate was somehow inversely proportional to naturalness. This does not hold in some cases, such as for a view angle of 0^0, because of the low gait recognition performance at those view angles.

V. CONCLUSION

Our proposed gait anonymization method prevents gait recognition systems from recognizing a person's gait and therefore prevents unauthorized gait recognition. It uses a newly developed convolutional neural network to modify the original gait in a video so that it cannot be identified by a gait recognition system, while still maintaining the gaits
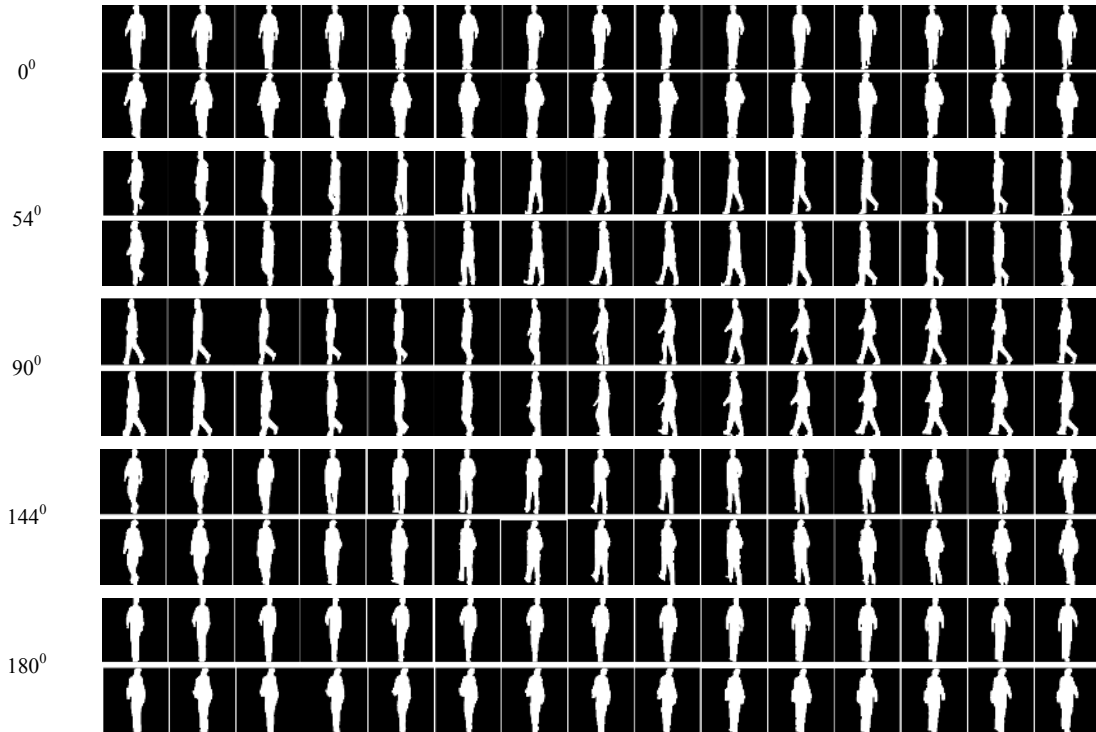

Fig. 9: Silhouettes of original and anonymized gaits of one person for various view angles.

TABLE I: Success rate (%) by view angle.

| Gallery	Probe view										
	0^0	18^0	36^0	54^0	72^0	90^0	108^0	126^0	144^0	162^0	180^0
0^0	52.05	45.41	82.85	45.74	68.84	58.33	85.00	79.00	**88.33**	72.50	55.00
18^0	94.11	33.61	65.94	50.78	68.33	60.11	43.46	73.66	76.29	81.42	63.82
36^0	**98.86**	55.74	29.36	23.74	88.33	73.57	**89.37**	76.53	78.63	71.29	69.28
54^0	92.85	66.57	65.92	23.54	73.97	75.57	63.06	**90.37**	70.00	76.67	77.67
72^0	98.27	81.00	30.06	28.73	28.27	60.14	61.86	52.50	50.29	**87.78**	51.67
90^0	98.50	**95.57**	81.52	**80.00**	46.42	31.51	38.91	59.91	84.68	62.94	**95.74**
108^0	95.07	91.57	51.36	77.17	49.61	50.85	24.53	45.76	33.18	57.50	95.00
126^0	98.07	71.11	39.81	54.69	87.14	65.79	57.26	39.63	23.76	61.21	59.28
144^0	83.33	54.02	48.33	55.00	85.95	63.00	63.82	55.26	24.67	65.56	77.33
162^0	73.33	44.74	73.97	49.00	**89.19**	65.71	88.33	58.89	51.10	30.37	68.79
180^0	64.28	46.66	**93.57**	45.90	50.71	**90.33**	70.22	87.69	67.38	46.20	46.73

(a) Naturalness (b) Success rate
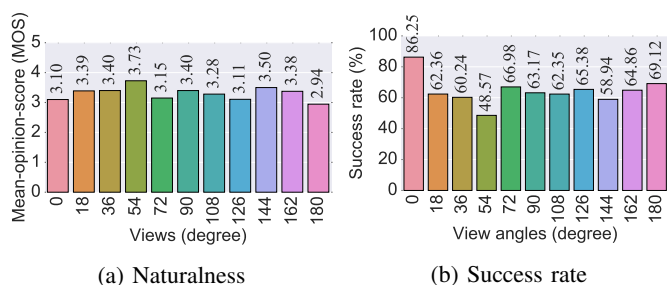
Fig. 10: Naturalness and success rates for anonymized gaits

naturalness. In our experiments, the average success rate ranged from 48.57% to 86.25% depending on the view angle while the average naturalness score (MOS) ranged from 2.94 to 3.73.

These findings should open new research directions regarding privacy protection related to gait recognition. Our research simply used silhouette sequences as input. Future work will address the use of color videos as input in order to make our findings more useful. It also includes the application of this method to object image anonymization.

Acknowledgments

References

[1] M. P. Murray, "Gait as a total pattern of movement," *Am. J. Phys. Med.*, vol. 46, no. 1, pp. 290–333, 1967.

[2] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "A Review of Vision-Based Gait Recognition Methods for Human Identification," in *Digital Image Computing: Techniques and Applications (DICTA)*, 2010, pp. 320–327.

[3] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, pp. 1501–1518, 2003.

[4] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, pp. 316–322, 2006.

[5] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, "Robust view transformation model for gait recognition," in *2011 18th IEEE International Conference on Image Processing (ICIP)*, Sept 2011, pp. 2073–2076.

[6] 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 Trans. on Pattern Anal. and Mach. Intell.*, pp. 209–226, 2017.

[7] T. Yamada, S. Gohshi, and I. Echizen, "Privacy visor: Method based on light absorbing and reflecting properties for preventing face image detection," in *IEEE SMC*, 2013, pp. 1572–1577.

[8] A. Othman and A. Ross, "Privacy of facial soft biometrics: Suppressing gender but retaining identity," in *Proceedings, Part II, Computer Vision - ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014.* Springer International Publishing, 2015, pp. 682–696.

[9] N. Ruchaud and J. L. Dugelay, "De-genderization by body contours reshaping," in *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, 2017, pp. 1–6.

[10] 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*, 2006, pp. 441–444.

[11] J. D. Bustard, M. Ghahramani, J. N. Carter, A. Hadid, and M. S. Nixon, "Gait anti-spoofing," in *Handbook of Biometric Anti-Spoofing: Trusted Biometrics under Spoofing Attacks.* Springer London, 2014, pp. 147–163.

[12] D. Holden, J. Saito, and T. Komura, "A deep learning framework for character motion synthesis and editing," *ACM Trans. Graph.*, pp. 138:1–138:11, 2016.

[13] A. Dosovitskiy, J. T. Springenberg, M. Tatarchenko, and T. Brox, "Learning to Generate Chairs, Tables and Cars with Convolutional Networks," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, pp. 692–705, 2017.

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *NIPS*, pp. 1–9, 2014.

[15] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," 2016. [Online]. Available: http://arxiv.org/abs/1609.04802

[16] J. Y. Zhu, P. Krahenbuhl, E. Shechtman, and A. A. Efros, "Generative visual manipulation on the natural image manifold," in *ECCV*, 2016, pp. 597–613.

[17] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *ICML*, 2017, pp. 2642–2651.

[18] X. Yan, J. Yang, K. Sohn, and H. Lee, "Attribute2Image: Conditional image generation from visual attributes," in *ECCV*, 2016, pp. 776–791.

[19] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," 2016. [Online]. Available: http://arxiv.org/abs/1609.03499

[20] M. Sharif, S. Bhagavatula, L. Bauer, and M. K. Reiter, "Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition," in *ACM SIGSAC*, 2016, pp. 1528–1540.