Anonymization of Gait Silhouette Video by Perturbing Its Phase and Shape Components

Yuki Hirose*, Kazuaki Nakamura*, Naoko Nitta*, and Noboru Babaguchi*
* Graduate School of Engineering, Osaka University, Suita, Osaka, Japan
E-mail: hirose@nanase.comm.eng.osaka-u.ac.jp Tel/Fax: +81-6-6879-7746

***Abstract*—Nowadays there are a lot of videos containing walking people on the web (e.g. YouTube). These videos can cause a privacy issue because the walking people can be identified by silhouette-based gait recognition systems which have been rapidly advanced in recent years. To solve the issue, in this paper, we propose a method for anonymizing human gait silhouettes. A gait silhouette consists of a static component including the body shape and a dynamic component including postures. We refer to the former and the latter as a shape component and a phase component, respectively. The proposed method anonymizes given gait silhouettes as follows: First, each of the given silhouettes is decomposed into its shape and phase components. Next, both components are separately perturbed. Finally, a new gait silhouette is generated from the perturbed components. Owing to the perturbation, the original silhouettes become less informative in the static aspect as well as the dynamic aspect, by which the gait recognition performance is seriously degraded. In our experimental results, the accuracy was actually degraded from 100% to 30% or less, without yielding any unnatural appearance in the output anonymized gait silhouettes.**

I. Introduction

Nowadays, a lot of videos are uploaded and published on video sharing services (VSS) such as YouTube. Although these videos can be freely accessed from all over the world, they often contain the appearance of private citizens, which can cause a privacy issue. To solve the issue, it is desirable that the provider of VSS should anonymize human regions in the uploaded videos before publishing them.

A typical example of the privacy-sensitive human regions is face. Thus, methods for anonymizing face regions in a video have been widely studied. Moreover, in recent years, human gait has also become privacy-sensitive, with the rapid growth of the performance of gait recognition systems [1], especially silhouette-based ones [2]. If an attacker has a sophisticated gait recognition system, he can analyze any video on VSS and identify people in the video. Nevertheless, there are few studies attempting to anonymize gait information.

From the above backgrounds, we focus on the task of anonymizing human gait in a given video. A naïve way to achieve this is to visually abstract whole human region by blocking out, blurring, pixelization, and so on [3]. However, visual abstraction makes the input video unnatural, which can frustrate its viewers and prevent their comfortable viewing. Hence, we consider another strategy as below:

(1) Detect and crop every human region in each frame of the input video.
(2) Binarize the cropped region to extract its silhouette.
(3) Slightly deform the silhouette so that gait recognition systems cannot correctly identify the human.
(4) Map the texture of the original human region onto the deformed silhouette and fill it back to the input video.

We refer to the deformation process in the third step as *human gait anonymization*, which plays a key role in the above strategy. Hence, we only focus on the third step and propose a method for achieving it in this paper. Note that it is not good to just replace each silhouette with some pre-constructed one, because it makes the fourth step quite difficult and easily yields unnatural appearance.

In a video of a person's gait silhouettes, his/her body shape is static and does not change. We call it a *shape component*. In contrast, the person's posture dynamically changes with phase of walking motion. We call it a *phase component*. In the proposed method, we anonymize a given gait silhouette video by perturbing its shape and phase components. The contribution of this paper is summarized as follows: First, this is the first work focusing on the problem of human gait anonymization from the aspects of both shape and phase. Second, we establish a general encoder-decoder framework for human gait anonymization that does not restrict the network structure of the encoder and that of the decoder.

In the remainder of this paper, we first review the related work in Section 2. Next, in Section 3, we describe the proposed method in detail, whose performance is experimentally evaluated in Section 4. Finally, we conclude this paper in Section 5.

II. Related Work

A. Anonymization of Visual Information

People's appearance in images and video includes various kinds of privacy sensitive information. To protect the information, methods for anonymizing the people's appearance, especially their face, have been actively studied in the past decades. Early studies proposed to apply visual abstraction techniques such as blurring and pixelization to a given face region [3], [4], which can successfully prevent human observers from identifying people from their face. However, visual abstraction is not necessarily effective for preventing automated face recognition systems [5]. In addition, visual abstraction makes the given face region quite unnatural. For these reasons,

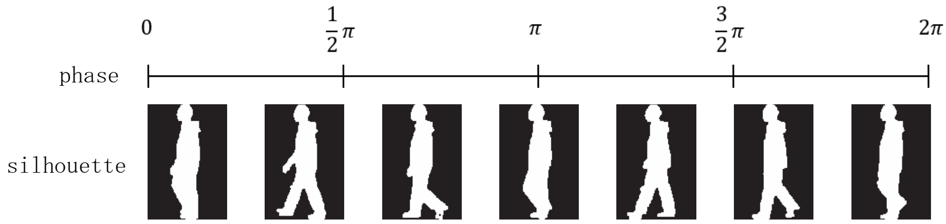

Fig. 1. Relationship between silhouette and its phase.

another approach have been studied in recent years: replacing given face regions with some other face images. A typical example is the work of Bitouk and his colleagues [6]. For a given face region, they proposed to select a face image similar to the given one from a pre-constructed face library and replace the given region with the selected face image by seamlessly blending their colors. Gross et al. proposed the method named k-Same and its extension named k-Same-Select [7], [8]. In these methods, k images that are closest to a given face region are first searched from a pre-constructed face library, and then the given region is replaced with an average of the k-closest images. More recently, Nakashima et al. proposed a patch-based replacing approach, which can achieve face anonymization without missing facial expressions [9].

Compared to face anonymization, there are only a few methods of human gait anonymization. Agrawal et al. proposed a visual abstraction-based approach, which applies a blurring filter to whole human region [10]. This method is not suitable to the videos on VSS because of unnatural appearance in the anonymization result. Unlike this, Tieu et al. proposed to slightly deform human gait silhouettes [11] by simply mixing an input silhouette with another one called a noise silhouette. However, their method only focuses on the static aspect of human gait although the dynamic aspect is also important and privacy sensitive. In contrast, we focus on both aspects in this paper.

B. Silhouette-based Gait Recognition

Silhouette-based gait recognition is another research field deeply related to our work. Existing methods for this task can be divided into two types. One is the methods directly processing a sequence of gait silhouettes, and the other is the methods compressing the gait information contained in a given sequence into a single image before recognizing a person. As an example of the former, Kale et al. proposed to employ a hidden Markov model for directly modeling each person's gait silhouette sequences [12]. Since this type of methods is disadvantageous in computational efficiency, the latter has been more actively studied, whose typical example is gait energy image (GEI) proposed by Man [13]. GEI is obtained by averaging one cycle of gait silhouette images and can be directly used as a feature for gait recognition. Due to the averaging process, GEI loses dynamic information. To cope with this drawback, Bashir et al. proposed gait entropy image (GEnI) [14], which is obtained by computing the Shannon entropy of the gray level of each pixel. Since the pixels corresponding to dynamically moving body parts such as arms and legs tend to have large entropy, GEnI modestly includes dynamic information as well as static information. Frequency domain feature (FDF) [15], which is obtained as the Fourier transform of the given gait silhouette sequence, is another example having dynamic information. These features are input to a deep neural network (DNN) that outputs person ID in modern gait recognition methods [2].

III. Anonymizing Gait Silhouette Video

A. Overview

As mentioned in Section 1, a gait silhouette consists of a static shape component, i.e., body shape, and a dynamic phase component, i.e., posture. Based on this consideration, we let $I_a(\theta)$ denote the gait silhouette of a person a with phase θ. Note that θ can be defined as a real value in the range of $[0, 2\pi]$ because human walking is a periodic motion (see Fig. 1). Using this notation, an input gait silhouette video of a certain person a is represented as $\{I_a(\theta_i) | i = 1, 2, \cdots\}$, where $I_a(\theta_i)$ is the i-th frame of the input video and θ_i is its phase. The purpose of our gait anonymization method is to transform each frame $I_a(\theta_i)$ into $I_{a'}(\theta'_i)$, which means the silhouette of a non-existing person a' with phase θ'_i. We achieve this process as follows, where $\boldsymbol{x}_a$ and $\boldsymbol{y}_\theta$ are feature vectors representing the shape component and the phase component, respectively (see also Fig. 2). In the remainder of this paper, we refer to $\boldsymbol{x}_a$ and $\boldsymbol{y}_\theta$ as *shape code* and *phase code*, respectively.

(1) For each frame $I_a(\theta_i)$, estimate its phase θ_i.
(2) Extract a shape code $\boldsymbol{x}_a$ from $I_a(\theta_i)$ for each i. Ideally, the same shape code should be extracted for any i.
(3) Perturb the phase θ_i and the shape code $\boldsymbol{x}_a$. Let Δ_θ and $\boldsymbol{\Delta}_x$ be the perturbation for the phase and that for the shape code, respectively. Using these perturbations, θ_i is transformed into $\theta'_i = \theta_i + \Delta_\theta$. Similarly, $\boldsymbol{x}_a$ is transformed to $\boldsymbol{x}_{a'} = \boldsymbol{x}_a + \boldsymbol{\Delta}_x$.
(4) Calculate a phase code $\boldsymbol{y}_{\theta'_i}$ from θ'_i, and generate a new silhouette image $I_{a'}(\theta'_i)$ using $\boldsymbol{y}_{\theta'_i}$ and $\boldsymbol{x}_{a'}$.

Thanks to the perturbation $\boldsymbol{\Delta}_x$, we can anonymize the static shape component of the input video. At the same time, its dynamic phase component can also be anonymized due to the perturbation Δ_θ. In the step (3), the same $\boldsymbol{\Delta}_x$ is used

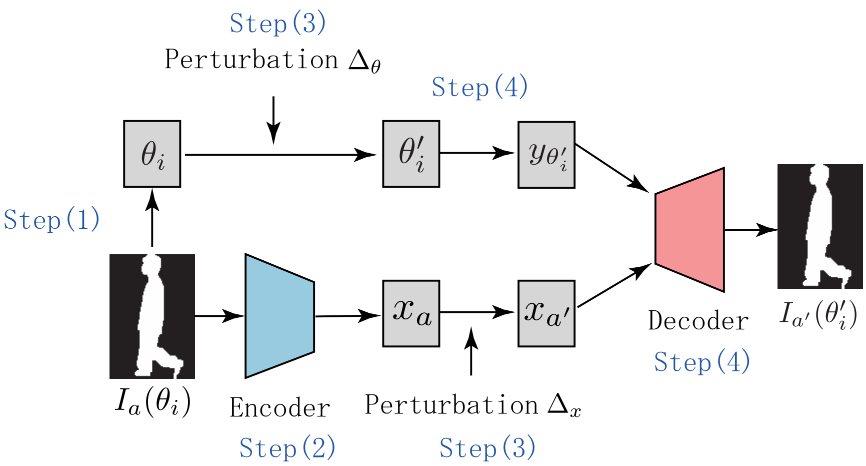

Fig. 2. Overview of the proposed method.

for all i so that the body shape does not unnaturally change in the resultant gait silhouette video. In contrast, different Δ_θ is used for different i. In the steps (2) and (4), we use DNNs for extracting the shape codes as well as generating the new silhouettes. Hereafter, we describe each of the above four steps in detail. Note that, for all gait silhouette videos used in the subsequent subsections, we extract their one cycle using autocorrelation function as a pre-process.

B. *Phase Estimation*

Before describing a phase estimation method in detail, we first define the phase itself clearly. The definition of the phase should satisfy the following two conditions: First, all the silhouettes with the same posture should have the same phase regardless of individual people. Second, silhouette appearance should continuously change with respect to phase. To satisfy these conditions, we use a certain reference video of a cycle of gait silhouettes. Let $R = (r_0,\ r_1,\ \cdots,\ r_{N-1})$ be the reference video, where r_i is its i-th frame and N is the number of the frames in R. For each r_i, we define its phase as $\frac{2\pi i}{N}$. Based on this definition, we estimate the phase of every other gait silhouette video $V = (v_0,\ v_1,\ \cdots,\ v_{M-1})$ by making a correspondence between R and V with DP matching, where v_j is the j-th frame in V and M is the number of the frames in V.

If we directly apply DP matching to R and V, v_0 is always matched with r_0 although a person's posture in v_0 is not necessarily same with that in r_0. To cope with this problem, we circularly shift the frames in V. Let V_l be the l-shifted version of V, that is, $V_l = (v_l,\ v_{l+1},\ \cdots,\ v_{M-1},\ v_0,\ v_1,\ \cdots,\ v_{l-1})$. We apply DP matching to R and V_l. Let $C(R, V_l)$ be the matching cost. We perform the above process for all $l \in \{0,\ 1,\ \cdots,\ M-1\}$ and find the best $\hat{l}$ that minimizes the cost $C(R, V_l)$, that is, $\hat{l} = \mathrm{argmin}_l\ C(R, V_l)$. Based on the matching result of R and $V_{\hat{l}}$, we estimate the phase of v_j as $\frac{2\pi i}{N}$ if v_j was matched to r_i.

C. *Code Extraction and Silhouette Generation*

Using the phase estimated in the previous section, we define the phase code $\boldsymbol{y}_\theta$ of each silhouette $I_a(\theta)$ as

$$\boldsymbol{y}_\theta = (\sin\theta \quad \cos\theta) \ \in \mathbb{R}^2 \tag{1}$$

in order to avoid the discontinuity between 0 and 2π.

Next, we train a DNN that extracts a shape code from a given gait silhouette. Note that this DNN should extract the same code $\boldsymbol{x}_a$ from all $I_a(\theta_i)$ regardless of θ_i. To train the DNN, some ground-truth data of the shape codes are required, which is difficult to be directly collected. Hence, we employ an indirect approach. First, using a certain training dataset of gait silhouette videos, we train a variational autoencoder (VAE) that can compress an input silhouette into a low-dimensional feature vector and reconstruct the same silhouette from the feature vector. Let E and D be the encoder and decoder parts of the trained VAE. By E, each silhouette $I_a(\theta_i)$ in the training dataset is transformed to a feature vector $\boldsymbol{z}_a(\theta_i) = E[I_a(\theta_i)]$. Using the transformed vectors, we calculate their average as

$$\hat{\boldsymbol{x}}_a = \frac{1}{M_a} \sum_{i=1}^{M_a} \boldsymbol{z}_a(\theta_i) \ , \tag{2}$$

where M_a is the length of the gait silhouette video of the person a. The calculated $\hat{\boldsymbol{x}}_a$ is used as the ground truth of the shape code of a person a to train another encoder $E_{\boldsymbol{x}}$ that can extract a shape code from any unknown gait silhouettes. Using the trained $E_{\boldsymbol{x}}$, we can extract the almost same code $\boldsymbol{x}_a = E_x[I_a(\theta)]$ from all $I_a(\theta)$ regardless of θ.

In addition, we train one more DNN that generates a new gait silhouette from an input phase code and a shape code. To this end, we first train a code combiner F that combines $\boldsymbol{y}_{\theta_i}$ and $\boldsymbol{x}_a$ into $\boldsymbol{z}_a(\theta_i)$, using the same training dataset. After successfully training F, we serially concatenate F and D, which is finally used as the silhouette generator. Fig. 3 shows the relationship between E, D, $E_{\boldsymbol{x}}$, and F. As shown in Fig. 3, D, $E_{\boldsymbol{x}}$, and F form a single large network. Hence, we simultaneously train $E_{\boldsymbol{x}}$ and F in practice by minimizing the

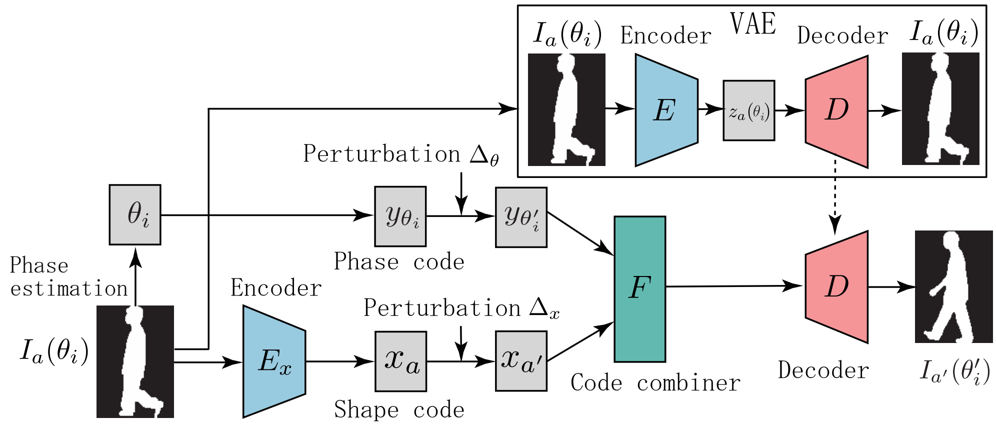

Fig. 3. Detail of the proposed encoder-decoder framework for human gait anonymization.

following loss function, i.e.,

$$L(E_x,\ F) = \sum_a \sum_{i=1}^{M_a} ||E_x[I_a(\theta_i)] - \hat{\boldsymbol{x}}_a||^2 + \lambda\,||F[\boldsymbol{y}_{\theta_i},\ E_x[I_a(\theta_i)]] - \boldsymbol{z}_a(\theta_i)||^2\ , \quad (3)$$

where λ is a weighting constant to control the balance between the performance of $E_{\boldsymbol{x}}$ and that of F.

D. Perturbing Phase and Shape Codes

In this subsection, we describe how to perturb a phase and a shape code in detail.

First, to perturb a phase, we directly change a phase value θ to $\theta' = f(\theta)$ using a certain function f. To avoid the unnatural posture change in the resultant sequence of anonymous gait silhouettes, the function f should be continuous, differentiable, and monotonically increasing. To satisfy this condition, we use

$$f(\varphi) = \begin{cases} \left\{\alpha^\beta - (\alpha - \varphi)^\beta\right\}^{\frac{1}{\beta}} & (0 \le \varphi < \alpha) \\ \alpha + \left\{(1-\alpha)^{\frac{1}{\beta}} - (1-\varphi)^{\frac{1}{\beta}}\right\}^\beta & (\alpha \le \varphi \le 1) \end{cases}, \quad (4)$$

where α and β are constants satisfying $0 \le \alpha \le 1$ and $0 < \beta$. Fig. 4 shows a sketch of this function. Because the domain of the above f is $0 \le \phi \le 1$, we actually calculate θ' as $\theta' = 2\pi f(\frac{\theta}{2\pi})$. This is equivalent to setting $\Delta_\theta = \theta' - \theta = 2\pi f(\frac{\theta}{2\pi}) - \theta$. The phase code after the perturbation is computed as $\boldsymbol{y}_{\theta'} = (\sin\theta' \quad \cos\theta')$.

For shape perturbation, we directly change a shape code by the following strategy: For a given shape code $\boldsymbol{x}_a$, we first find its K-nearest neighbors from a pre-constructed library of shape codes. In practice, the library is constructed by reusing the training dataset that was used to train E, D, E_x, and F. Let $\boldsymbol{x}_a^{(j)}$ $(j = 1, \cdots, K)$ be the j-the nearest neighbor of $\boldsymbol{x}_a$. Their average is next calculated and then used as the perturbed shape code $\boldsymbol{x}_{a'}$, i.e.,

$$\boldsymbol{x}_{a'} = \frac{1}{K}\sum_{j=1}^{K} \boldsymbol{x}_a^{(j)}\ . \quad (5)$$

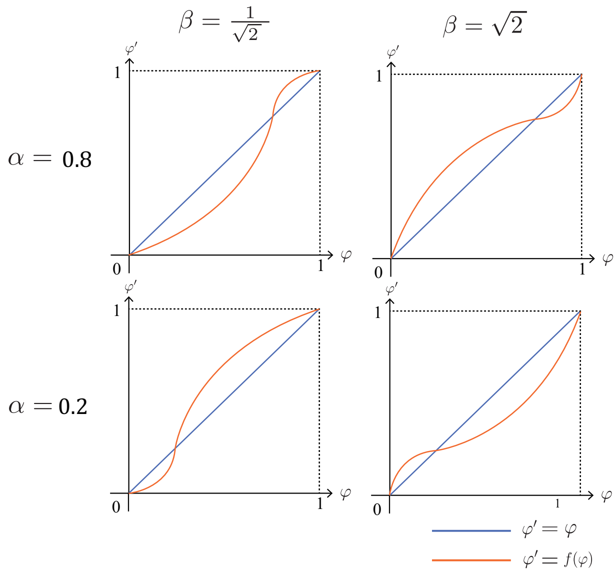

Fig. 4. Sketch of the graph of the function f with several combinations of α and β.

This is equivalent to setting $\boldsymbol{\Delta_x} = \boldsymbol{x}_{a'} - \boldsymbol{x}_a$. The above strategy is inspired by k-Same [7] described in Section 2-A, which has two advantages. First, it can successfully change the appearance of the input silhouette shape. Since the shape codes in the library are not uniformly but biasedly distributed in a feature space in general, the averaged code $\boldsymbol{x}_{a'}$ becomes different from the original code $\boldsymbol{x}_a$. Second, the averaged code can keep a certain level of visual naturalness unless K is too large.

IV. Experiments

A. Experimental Setup

To evaluate the effectiveness of the proposed method, we conducted an experiment, in which we employed *treadmill dataset A* and *treadmill dataset B* from the OU-ISIR Gait Database [1] as the datasets for training and evaluation. The

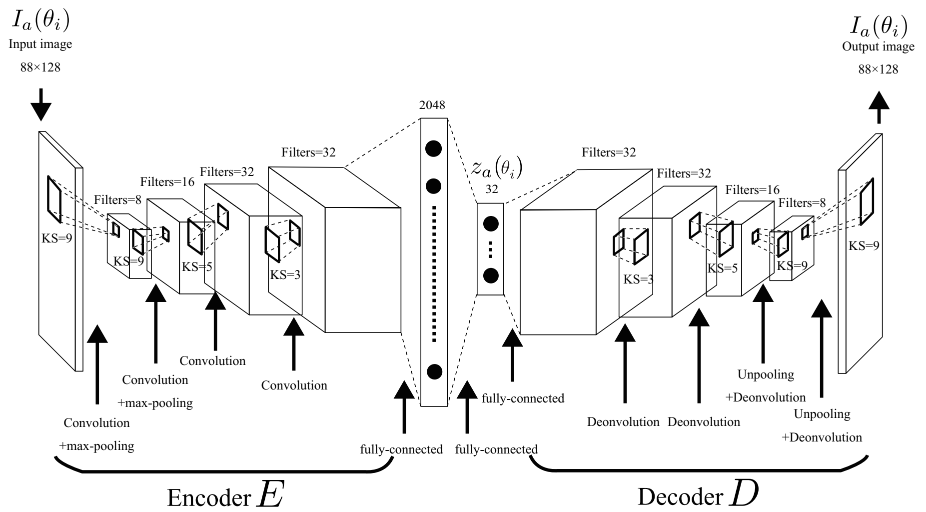

Fig. 5. Structure of variational autoencoder (i.e., E and D) used in the experiment.

treadmill dataset A includes gait silhouette videos of 34 people, whose walking speed is ranged from 2 [km/h] to 10 [km/h]. This dataset has less variety of clothes. On the other hand, the *treadmill dataset B* includes gait silhouette videos of 68 people who are walking with 32 kinds of clothes. The total number of the videos in the dataset is $68 \times 32 = 2176$. In our experiment, we first trained E, D, $E_{\boldsymbol{x}}$, and F using the *treadmill dataset B*, and then anonymized the videos in the *treadmill dataset A* using the proposed method, only targeting 204 videos in which a person is walking at the speed of 4, 5, and 6 [km/h]. This is because only these speeds are natural as a waking speed of ordinary citizens.

For parameter setting, we set the dimension of the shape codes as 32. K was empirically set as 20. α and β in Formula (4) were also empirically set as follows: α was uniformly sampled from the range $[0, 1]$ and β was randomly set as either $\sqrt{2}$ or $\frac{1}{\sqrt{2}}$. For network structures, E and $E_{\boldsymbol{x}}$ were simply designed, consisting of four convolutional layers, two max-pooling layers, and two fully-connected layers with ReLU gates. D was designed in a similar way, employing up-pooling layers instead of pooling layers. F consisted only of three fully-connected layers, whose input layer receives $\boldsymbol{x}_{a'}$ and $\boldsymbol{y}_{\theta'}$. Fig. 5 shows the network structure of E and D, and Fig. 6 shows the structure of F. The structure of $E_{\boldsymbol{x}}$ was totally same with that of E. Note that this design is just an example; any other structures can be used in our proposed method.

The performance of the proposed method was evaluated from the following two aspects: gait recognition accuracy and visual naturalness. For the former, we performed leave-one-out cross validation using the above 204 videos, employing GEI, GEnI, and FDF. Although they are typically input to a DNN in the state-of-the-art methods of gait recognition, we did not have sufficient number of videos for successfully training DNN in this experiment. Therefore, we employed a multi-layer perceptron having three layers as a gait recognizer, whose input is a feature vector extracted from GEI, GEnI, and FDF by compressing them into 34 dimensional vectors by principal component analysis and linear discriminant analysis. For the latter, i.e., visual naturalness, we conducted two kinds of questionnaire surveys, whose detail is described later. To separately evaluate the performance of phase perturbation and that of shape perturbation, we compared the following three cases: only perturbing phase (named *phase-only*), only perturbing shape (*shape-only*), and using both perturbations (*both*).

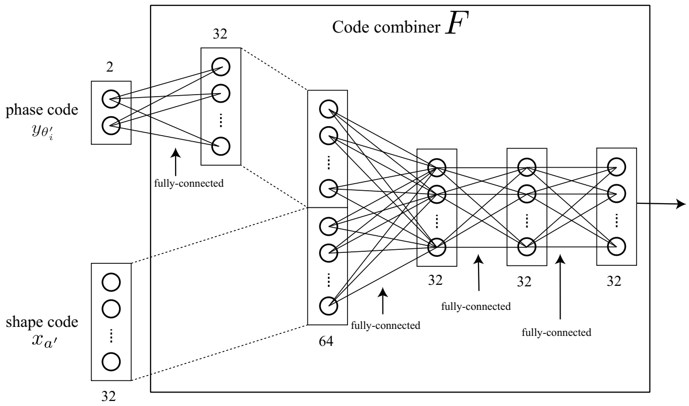

Fig. 6. Structure of code combiner F.

B. Results and Discussion

1) Examples of anonymized gait silhouettes: Fig. 7 shows an example of the anonymization results. We can see that the anonymized silhouettes keep human-like appearance, which are not so different from the original ones. This result demonstrates that the proposed method does not yield unnatural silhouettes.

2) Evaluation of anonymization performance: Table I shows the results of the gait recognition based on GEI, GEnI, and FDF. In Table I, the recognition accuracy is near 100% for

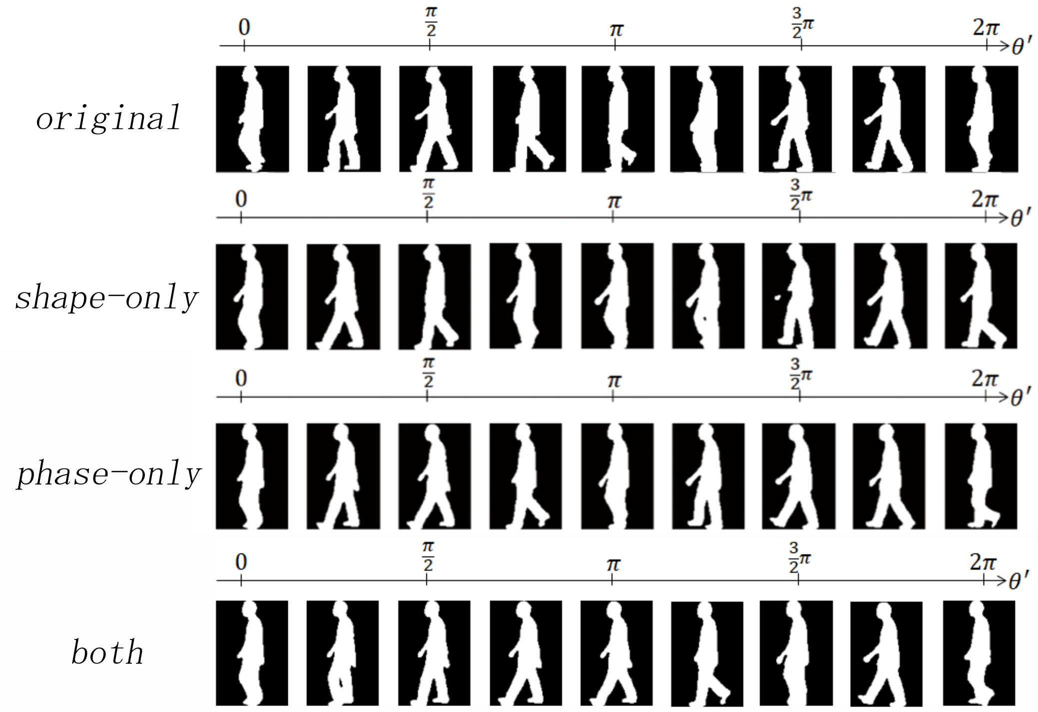

Fig. 7. Anonymization results of gait silhouette videos.

all the methods in the case of using original gait silhouettes. Compared to this, the accuracy is drastically degraded by the proposed method, especially with *shape-only* and *both*. This result demonstrates the effectiveness of the proposed method.

In the case of *phase-only*, a recognition accuracy of around 60% is still achieved by GEI and GEnI. This is because these methods do not reflect the dynamic component of human gait. Since a sequence of phase codes represents the dynamic component, the phase perturbation is expected to be helpful for anonymizing the dynamic aspect of human gait. However, GEI-based gait recognizer originally does not use the dynamic component. Hence, the phase perturbation cannot give a serious effect on the accuracy of GEI. This is also the case with GEnI. Indeed, GEnI reflects the dynamic component a little, but it lacks the information about changes in phase. Unlike GEI and GEnI, FDF reflects the dynamic component as well as the static component, because it is calculated by Fourier transform. Hence, both the phase perturbation and the shape perturbation work well. In fact, the recognition accuracy is degraded to 20% or less with *phase-only* in the case of FDF. This result indicates that the phase perturbation is useful when a gait recognizer considers the dynamic aspect of human gait.

3) Evaluation of visual naturalness: To evaluate the proposed method from the aspect of visual naturalness, we conducted two kinds of questionnaire surveys. For the first survey, we recruited 20 participants and provided them five videos of gait silhouettes, one of which is anonymized by the proposed method and the others are original. Then we asked the participants to identify which is the anonymized one. If the anonymized silhouettes are visually natural as much as the original ones, the accuracy of the participants' answers will be around the chance rate, namely 20%. For the second survey, we recruited other 17 participants and provided them either a video of original silhouettes or that of anonymized ones. Then we asked the participants to evaluate how much the provided video looks visually natural on a five-point scale (1: worst, 5: best). We separately performed the above trials for the three cases, i.e., *phase-only*, *shape-only*, and *both*. Table II shows the result of the two surveys.

As seen in the result of the first survey, the participants cannot correctly identify the anonymized gait silhouettes. This indicates that the proposed method does not yield seriously unnatural appearance in the anonymization results. Only in the

TABLE I
GAIT RECOGNITION ACCURACY USING GEI, GENI, AND FDF AS RECOGNITION METHODS.

	original	*phase-only*	*shape-only*	*both*
GEI	99.0%	59.3%	25.9%	28.9%
GEnI	100%	57.8%	24.5%	21.6%
FDF	100%	23.5%	6.8%	3.9%

TABLE II
RESULTS OF QUESTIONNAIRE SURVEYS ABOUT VISUAL NATURALNESS. WE CALCULATED AVERAGE SCORE IN THE SECOND SURVEY.

	original	*phase-only*	*shape-only*	*both*
1st survey	-	35.0%	10.0%	25.0%
2nd survey	3.75	1.54	2.36	2.25

case of *phase-only*, the accuracy is significantly higher than 20%. In addition, in the result of the second survey, we can see that the score of *phase-only* considerably decreases from that of *original*. This is because we used the original shape code in the case of *phase-only* when generating an anonymous gait silhouette for each frame of a given video. Ideally, it is desirable that the encoder $E_{\boldsymbol{x}}$ provides the same shape code $\boldsymbol{x}_a$ for all frames $I_a(\theta_i)$ regardless of i. However, since it is difficult to perfectly train $E_{\boldsymbol{x}}$ in practice, the shape codes obtained from different $I_a(\theta_i)$, i.e., $E_{\boldsymbol{x}}[I_a(\theta_i)]$, are slightly different with each other. Hence, anonymous gait silhouettes generated from the shape codes are also slightly different with each other in terms of a person's body shape. Due to the difference, the silhouette in the resultant sequence does not change smoothly with time, which might cause the degradation of the evaluation score. In contrast, in the cases of *shape-only* and *both*, we always use the same shape code, i.e., $\boldsymbol{x}_{a'}$ for all frames $I_a(\theta_i)$ regardless of i. Hence, the evaluation score is not so seriously degraded. However, arms sometimes partly disappear in the resultant silhouettes. This is why the evaluation score is somewhat degraded even in the cases of *shape-only* and *both*. This problem can be improved by employing a more sophisticated network for the decoder D.

V. Conclusions

In this paper, we proposed a method for anonymizing gait silhouette video. Our method perturbs the shape and phase codes extracted from each frame of a given video and generates a new silhouette from the perturbed codes for each frame using a DNN. Our experimental results showed that the proposed method degrades the accuracy of silhouette-based gait recognition from 100% to 30% or less without seriously yielding unnatural appearance in the resultant anonymized gait silhouettes.

As a next step, we will try to propose a method for mapping the texture of a human region in the original video onto the corresponding anonymized silhouette in a future work.

This work was supported by JSPS KAKENHI Grant Numbers JP16H06302, JP15H01686, and JP18H04120.

References

[1] Y. Makihara, D. S. Matovski, M. S. Nixon, J. N. Carter, and Y. Yagi: “Gait Recognition: Databases, Representations, and Applications,” John Wiley & Sons, Inc., pp.1-15, 2015.

[2] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, Y. Yagi: “GEINet: View-Invariant Gait Recognition Using a Convolutional Neural Network,” In: Proc. Int'l Conf. Biometrics, pp.1–8, 2016.

[3] M.Boyle, C. Edwards, and Saul Greenberg: “The Effects of Filtered Video on Awareness and Privacy,” In: Proc. Int'l Conf. Computer Supported Cooperative Work, pp.1–10, 2000.

[4] C. Neustaedter, S. Greenberg, and M. Boyle: “Blur Filtration Fails to Preserve Privacy for Home-Based Video Conferencing,” In: Proc. Int'l Conf. Transactions on Computer-Human Interaction, Vol.13, No.1, pp. 1–36, 2006.

[5] E. Newton, L. Sweeney, and B. Malin: “Preserving Privacy by De-Identifying Face Images,” IEEE Trans. Knowledge and Data Engineering, Vol.17, No.2, pp.232–243, 2005.

[6] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, S. K. Nayar: “Face Swapping: Automatically Replacing Faces in Photographs,” ACM Trans. Graphics, Vol.27, No.3, 2008.

[7] R. Gross, E. Airoldi, B. Malin, and L. Sweeney: “Integrating Utility into Face De-identification,” In: Proc. Int'l Conf. Privacy Enhancing Technologies, pp.227–242, 2005.

[8] R. Gross, L. Sweeney, F. de la Torre, S. Baker: “Model-Based Face De-identification,” In: Proc. Int'l Conf. Computer Vision and Pattern Recognition Workshop, 2006.

[9] Y. Nakashima, T. Koyama, N. Yokoya, and N. Babaguchi: “Facial Expression Preserving Privacy Protection using Image Melding,” in Proceedings of the 2015 IEEE International Conference on Multimedia and Expo, pp. 1–6, 2015.

[10] P. Agrawal and P. J. Narayanan: “Person De-Identification in Videos,” IEEE Trans. Circuits and Systems for Video Technology, Vol.21, No.3, pp. 299–310, 2011.

[11] N. T. Tieu, H. H. Nguyen, H. Nguyen-Son, J. Yamagishi, and I. Echizen: “An Approach for Gait Anonymization Using Deep Learning,” In: Proc. IEEE Workshop on Information Forensics and Security (WIFS), 2017.

[12] A. Kale, A. N. Rajagopalan, N. Cuntoor, and V. Kruger: “Gait-Based Recognition of Humans Using Continuous HMMs,” In: Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp.336–341, 2002.

[13] J. Man, B. Bhanu: “Individual Recognition Using Gait Energy Image,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.28, No.2, pp.316–322, 2006.

[14] K. Bashir, T. Xiang, S. Gong: “Gait Recognition using Gait Entropy Image,” In: Proc. Int'l Conf. Crime Detection and Prevention, pp.1–6, 2009.

[15] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi: “Gait Recognition Using a View Transformation Model in the Frequency Domain,” In: Proc. Int'l Conf. Computer Vision, pp.151–163, 2006.