Quality-dependent View Transformation Model for Cross-view Gait Recognition

Daigo Muramatsu, Yasushi Makihara, and Yasushi Yagi
The Institute of Scientific and Industrial Research, Osaka University
8-1 Mihogaoka, Ibaraki, Osaka, 567-0047, Japan
{muramatsu,makihara,yagi}@am.sanken.osaka-u.ac.jp

Abstract

View difference is a factor that degrades the accuracy of gait recognition. A solution to reducing accuracy degradation is applying a view transformation model (VTM) that encodes a joint subspace of multi-view gait features trained from multiple training subjects. In the VTM framework, once an intrinsic vector of a test subject in the joint subspace is estimated from a gait feature with a source view (e.g. probe view), a gait feature with a destination view (e.g. gallery view) is generated for the same-view matching. Although this family of methods can improve the total accuracy, the quality of generated gait features depends on a test gait feature, and may be relevant to the accuracy of gait recognition. We therefore propose a method of incorporating the quality measure of the generated gait feature into the VTM framework. We employ the projection error into the joint subspace as the quality measure. A posterior probability is then computed by incorporating the quality measure. The accuracy evaluation against a subset of a public database collected from 1,912 subjects shows that the proposed method further improves the accuracy.

1. Introduction

Gait recognition is a biometric method used to recognize a person from their walking style [40], which can be acquired from a camera. Unlike many biometric techniques such as fingerprint, iris or face recognition, gait recognition can authenticate a person some distance from the camera, because it has high accuracy even when the resolution of an image sequence is relatively low (e.g., individual with its height of 30 pixels). However, the accuracy of gait recognition is often degraded by several covariates (e.g., views, clothes, and belongings) [44]. Among these, view difference is the most problematic covariate inducing accuracy degradation in gait recognition [4, 49], and hence in this paper we focus on the issue of view difference in gait recognition. The main reason why the change of view causes accuracy degradation is that it changes the appearance of a person as shown in Fig. 1. We can observe that the images are different even though they are captured simultaneously from the same subject with the same pose. Because appearance-based methods use image-based features, their accuracy is highly influenced by view change.

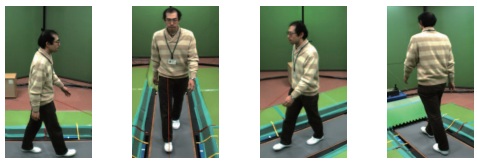

Figure 1. Images of a subject captured from different views

In general, two families of approaches to gait recognition have been proposed: appearance-based [44, 14, 35, 30] and model-based [5, 47, 48, 3]. Appearance-based approaches use captured image sequences directly, while model-based approaches calculate the model parameters from the images. From the viewpoint of appearance change caused by view difference, model-based approaches seem to be preferable because of their view-invariant nature. However, calculating model parameters with high accuracy is often difficult from the images captured in surveillance or criminal scenes; thus, there are inherent difficulties in applying model-based approaches in these cases, and consequently appearance-based approaches generally achieve higher accuracy than model-based ones. Therefore, we focus on the appearance-based approach and discuss possible solutions to the view change problem. We refer the readers to [12] for solutions regarding model-based approaches to this issue.

There are several approaches to the problems associated with matching gaits with different views (cross-view matching): visual-hull based approaches [45, 31, 6, 18], view-invariant-feature based approaches [19, 21, 15, 34, 33], and view transformation model (VTM)-based approaches [35, 26, 27, 28, 29, 38] have all been proposed. In this paper, we focus on the VTM-based approach, the mainstream approach for cross-view matching.

In the VTM framework, the training phase involves construction of the VTM (e.g., a joint subspace of multi-view

gait features in [35] and a regression model from one view to another view in [29]) using multi-view gait features of multiple training subjects. Given a pair of gait features with two different views (let them be "source" and "destination" views, respectively) in the test phase, a gait feature in the destination view is generated from a gait feature in the source view, and a matching score between the original gait feature and generated gait features is calculated. In this approach, only calculated matching score is used for recognition without considering the quality of the generated gait feature (i.e., the transformation accuracy). We believe, however, that the quality should be different from feature to feature, and that it should also be relevant to the matching score. Once we quantify such a quality, it can be used as auxiliary information to improve the accuracy of recognition, because many studies report improvements in accuracy using these kinds of quality measures [16, 32, 37]. Therefore, we estimate the transformation accuracy and use it to improve the recognition accuracy of VTM-based approaches for cross-view matching.

The contributions of this paper are summarized in the following two points to enhance the VTM framework:

Proposal of a quality measure

We focus on a VTM [35] that is based on the joint subspace learning of multi-view gait features. In this VTM, the joint subspace is learned by training multi-view gait features, and then an input gait feature is projected into the subspace for future transformation. We hypothesize that the transformation accuracy must be related to the projection error into the subspace, and dependent on the input gait feature. Therefore, in this paper, we use data captured by multiple temporally synchronized cameras to show that the transformation accuracy is dependent on the input gait feature. Using the same data, we also show the relationship between the projection error of each input gait feature into the learned subspace and the transformation error that is a measure of transformation accuracy. Further, we propose to use the projection error as an estimate of transformation error that is related to the quality of generated gait features.

Quality-dependent score normalization in the posterior domain

The computed quality measure is used together with a matching score to calculate a normalized score, namely a posterior probability that the given pair of gait features originate from the same person. More specifically, the posterior distribution is estimated in the framework of linear logistic regression (LLR) using a training set composed of positive (genuine) and negative (imposter) pairs of the matching score and the quality measure. The experimental results, tested against a subset of a public large population gait database [17], show that the proposed method further improves the recognition accuracy of gait recognition for cross-view matching.

2. Related work

Gait recognition approach for view difference

Appearance-based approaches that can tackle the view difference issues are classified into three groups: visual-hull-based approaches [45, 31, 6, 18], view-invariant feature-based approaches [19, 21, 15, 34, 33], and view transformation-based approaches [35, 26, 27, 28, 29, 38].

Visual-hull-based approaches assume that for all target subjects, 3D gait models or gait image sequences captured by multiple temporally synchronized cameras are available. From the 3D models, images that correspond to a virtual view can be generated by re-projecting the 3D objects to the image plane, and hence gait images from any view can be generated from the 3D gait model. These generated images are then used for recognition [45, 31, 18]. On the other hand, Bodor et al. apply image-based rendering to reconstruct gait features for any required view direction [6]. This type of approach can achieve high accuracy, but one limitation is that data from all target subjects captured by multiple temporally synchronized cameras are necessary for 3D gait model generation or image-based rendering. Thus, this approach cannot be applied to cases where only two gait image sequences (one for the gallery, and another for the probe) are given for verification.

In a different way from the visual-hull-based approach, the view-invariant feature-based approach does not necessitate that the gait data of the target subjects be synchronized, and hence this method can be applied to cases where only two gait image sequences are given for verification. View-invariant feature-based approaches can be further divided into two types: geometrical-based approaches [19, 21, 15] and subspace-based approaches [34, 33]. Geometrical-based approaches extract view-invariant features by considering their geometrical properties, whereas subspace-based approaches learn a joint subspace using training data and compute view-invariant features by projecting the original features onto the learned subspace. Although the subspace-based approaches are similar to those of VTM-based approaches, the subspace-based approaches do not reconstruct gait features unlike VTM-based approaches.

The limitation of the geometrical-based approach is its narrow range of applicable view differences. For example, Kale et al. propose a method to generate side view gait images from any other arbitrary view [21]. In this method, they generate gait images of the side view from any other arbitrary view under the assumption that the person (3D object) can be represented by a planar object in the sagittal plane. This type of approach can work well when the angle between the sagittal plane of the person and the image plane is small; however, when the angle is large, the accuracy is drastically degraded. The limitations of the subspace-based approach are almost the same as those of the VTM-based approach, as we discuss later.

Methods that take the VTM-based approach construct a view transformation model using gait image sequences of multiple subjects taken by multiple temporally synchronized cameras set on the target views, and use the model to generate gait images or gait features with a destination view from those with a different view. For construction of the VTM, either a matrix factorization process by Singular Value Decomposition (SVD) technique [46, 35, 26, 38], or a regression [27, 28, 29] is used. Whereas regression-based VTMs generate a gait feature directly from a source gait feature, SVD-based VTMs calculate an intrinsic vector by first projecting the source gait feature into the learned subspace, and then generating a gait feature from the intrinsic vector.

These types of method require data not from the recognition-target, but from non-recognition-target subjects captured by multiple temporally synchronized cameras, unlike the visual hull-based methods, which require data from recognition-target subjects. Therefore, methods using this approach can verify a target subject only from a pair of gait image sequences, and have a wider range of applicable view differences than those of the geometrical view-invariant feature-based approach. A limitation of this method is that training data associated with the required views are necessary and transformation accuracy depends on these data. In order to solve the former limitation, Muramatsu et al. proposed an arbitrary VTM method that generates training data from a 3D gait model of non-target subjects [38]. However, the latter problem has not yet been addressed to the best of our knowledge.

The evaluation of transformation accuracy is not a straightforward task, because we do not have the ground truth in a cross-view setting, and thus we cannot directly measure the accuracy. An alternative method is to estimate the accuracy. Thanks to the extra processing associated with intrinsic vector calculation inherent in of SVD-based VTMs, the intrinsic vector can be a clue to estimating the transformation accuracy. Therefore, using the SVD-based VTM, transformation accuracy can be incorporated into recognition.

Quality of biometrics

Because the quality of the biometric sample is very important in obtaining high accuracy in biometric systems, quality measurement/assessment algorithms are discussed, and the measured quality is used to improve recognition accuracy [13, 42].

Generally, the term *quality* is considered to be measurable auxiliary information that cannot be used to distinguish the genuine class and the imposter class by itself, but that makes an impact on the matching scores. For example, texture richness [8], ridge frequency [20] and local orientation quality [7] are all considered in fingerprint recognition. Similarly, illumination variation, head pose, and facial expression are considered in face recognition [1], while texture richness [8], the difference between iris diameter and pupil diameter [41], and occlusion and blurring of the iris region [22] are considered in iris recognition; entropy of signature is considered in signature verification [11]; and quality of silhouette [32, 37] and clothing variation [16] are considered in gait recognition. However, no quality measures for cross-view gait recognition have been proposed to the best of our knowledge.

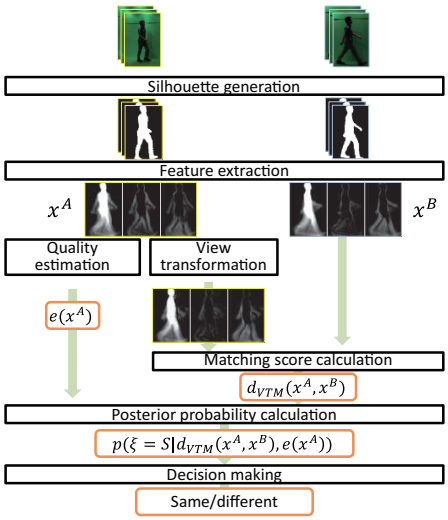

Figure 2. Algorithm of the proposed method

These measured qualities are used in different levels of processing. For example, quality measures are used at the data acquisition level in order to make decisions about whether the input data is used or rejected for recapturing [13]; at the preprocessing level, they are used to set a threshold associated with preprocessing [37]; at the score calculation level, they are used to normalize a uni-modal score, or generate uni- or multi- modal fusion [9, 24, 16, 10, 39, 43, 41]; and they are also used at the decision making level [25]. Among these, usage at the score calculation level is the most popular in biometric fields, and these methods are unified in biometric person authentication frameworks [23, 42].

3. Algorithm

3.1. Overview

Figure 2 shows an algorithm of the proposed method. Given a pair of gait image sequences with different views, we generate silhouette image sequences from each given image sequence by background subtraction-based graphcut segmentation [36]. We then extract a gait feature from each silhouette image sequence. Because the extracted gait features are derived from different views, we generate a gait

feature with the same view from one of the gait features with another view using the VTM. In a different way from the VTM methods reported in [35, 26, 27, 28, 29, 38], the proposed method estimates the transformation error of each generated gait feature as a quality measure, and uses this quality measure to calculate a posterior probability for decision making. In this section, we first briefly explain the method of view transformation between two views, and then explain how to estimate the quality measure associated with the generated gait feature. We finally explain how to incorporate this quality measure into a posterior probability for decision-making.

3.2. VTM

View-dependent transformation matrix generation
We assume that pairs of gait features with two views θ and ϕ from M subjects are available. Let $\boldsymbol{X}_\theta^m$ and $\boldsymbol{X}_\phi^m$ be N-dimensional gait features of the m-th subject with views θ and ϕ, respectively. Using these gait features, we generate a training matrix by arranging the gait features as Eq. (1), and decompose it by applying singular value decomposition (SVD):

$$\begin{aligned}\begin{bmatrix} \boldsymbol{X}_\theta^1 \cdots \boldsymbol{X}_\theta^M \\ \boldsymbol{X}_\phi^1 \cdots \boldsymbol{X}_\phi^M \end{bmatrix} &= \mathbf{U}\mathbf{S}\mathbf{V}^T \\ &= \begin{bmatrix} R(\theta) \\ R(\phi) \end{bmatrix} \begin{bmatrix} v^1 \cdots v^M \end{bmatrix}, \quad (1)\end{aligned}$$

where $\mathbf{U} \in \mathbb{R}^{2N\times M}$ is an orthogonal matrix that contains a set of bases for the joint subspace of concatenated gait features from the two views $[\boldsymbol{X}_\theta^T, \boldsymbol{X}_\phi^T]^T$, $\mathbf{V} \in \mathbb{R}^{M\times M}$ is also an orthogonal matrix, $\mathbf{S} \in \mathbb{R}^{M\times M}$ is a diagonal matrix whose on-diagonal elements are singular values. $R(\theta)$ and $R(\phi)$ are submatrices of $\mathbf{US}$ which back-project the individual vector v into gait features $\boldsymbol{X}_\theta$ and $\boldsymbol{X}_\phi$, respectively. Consequently, the gait feature $\boldsymbol{X}_\psi^m$ of the m-th subject with view $\psi \in \{\theta, \phi\}$ is described by

$$\begin{aligned}\boldsymbol{X}_\psi^m &= R(\psi)v^m, \\ &\quad \psi \in \{\theta, \phi\}, m = 1, 2, ..., M. \quad (2)\end{aligned}$$

Generating gait features by view transformation
From a gait feature with view θ/ϕ, we generate a gait feature with view ϕ/θ. For simplicity, we assume that we generate a gait feature with view ϕ from a gait feature with view θ in this section, and describe the target input gait feature by $\boldsymbol{x}_\theta$ and the generated gait feature by $\hat{\boldsymbol{x}}_\phi$. In order to distinguish test data gait features from training data gait features, we describe the test data gait features with the symbol $\boldsymbol{x}$ and not by $\mathbf{X}$, though the type of both gait features is the same.

The generated gait feature $\hat{\boldsymbol{x}}_\phi$ is derived from the following two steps: Step 1: We define a projection error $e\left(v; \boldsymbol{x}_\theta, R(\theta)\right)$ by

$$e\left(v; \boldsymbol{x}_\theta, R(\theta)\right) = ||\boldsymbol{x}_\theta - R(\theta)v||_2, \quad (3)$$

and estimate the vector $\hat{v}(\theta)$ in the joint subspace that minimizes the squared error in Eq. (3) by

$$\hat{v}(\theta) = \underset{v}{\operatorname{argmin}}\, e\left(v; \boldsymbol{x}_\theta, R(\theta)\right)^2. \quad (4)$$

Step 2: Using $\hat{v}(\theta)$, a gait feature with view ϕ is generated by

$$\hat{\boldsymbol{x}}_\phi(\theta) = R(\phi)\hat{v}(\theta). \quad (5)$$

3.3. Quality estimation

The success of VTM-based cross-view matching relies closely on transformation accuracy, i.e., how similar the generated gait feature is to an original gait feature with the destination view. Therefore, the transformation error between the generated gait feature $\hat{\boldsymbol{x}}_\phi(\theta)$ and the original gait feature $\boldsymbol{x}_\phi$, defined by

$$\text{Transformation error} = ||\hat{\boldsymbol{x}}_\phi(\theta) - \boldsymbol{x}_\phi||_2, \quad (6)$$

can be useful auxiliary information for accuracy improvement if we can measure it. However, we cannot measure it directly in a real situation, because an original gait feature $\boldsymbol{x}_\phi$ with the destination view is unavailable in a cross-view matching problem setting. On the other hand, we note that the success of the view transformation relies on how well the joint subspace of multi-view gait features derived from the non-recognition-target training subjects describes a gait feature of a target subject, namely projection error. This error can be described by

$$\text{Projection error} = ||\boldsymbol{x}_\theta - R(\theta)\hat{v}_\theta(\theta)||_2. \quad (7)$$

Figure 3 shows scatter plots of the relationship between the projection error and the transformation error of gait features from several subjects. In this figure, different point shows error of different subject. These errors are calculated using gait features that are extracted from gait image sequences captured by two temporally synchronized cameras with different views. From this figure, we obtain the following two observations:

- Transformation errors differ from feature to feature, namely, the degree of success of the view transformation varies depending on the feature.
- The correlation between the projection errors of the source-view gait features and the transformation error of the destination-view gait features is quite high. This result provides evidence that the projection errors of the source-view gait feature can be used to infer the transformation error of the destination-view gait features.

From these observations, we use the projection error as a quality measure of the generated gait feature, and incorporate it to calculate a posterior probability as explained later.

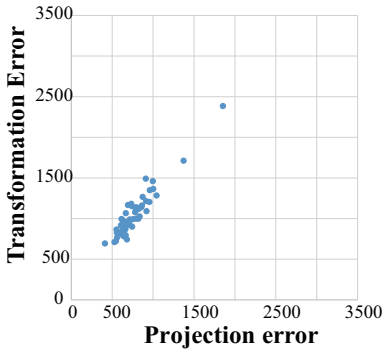

Figure 3. Scatter plots between the projection error and the transformation error

3.4. Matching score calculation

Let $\boldsymbol{x}_\theta^G$ and $\boldsymbol{x}_\phi^P$ be a gallery gait feature with view θ and a probe gait feature with view ϕ, respectively. When we transform the gallery gait feature with view θ to that with probe view ϕ, a matching score is calculated by

$$d_{VTM}(\boldsymbol{x}_\theta^G, \boldsymbol{x}_\phi^P) \quad = \quad \left|\left|\hat{\boldsymbol{x}}_\phi^G(\theta) - \boldsymbol{x}_\phi^P\right|\right|_2 . \qquad (8)$$

On the other hand, when we transform the probe gait feature with view ϕ into that with gallery view θ, the matching score is

$$d_{VTM}(\boldsymbol{x}_\phi^P, \boldsymbol{x}_\theta^G) \quad = \quad \left|\left|\hat{\boldsymbol{x}}_\theta^P(\phi) - \boldsymbol{x}_\theta^G\right|\right|_2 . \qquad (9)$$

Note that the first argument of the dissimilarity function $d_{VTM}(\cdot,\cdot)$ is transformed to a gait feature with another view.

3.5. Posterior probability calculation

We consider the posterior probability that a pair of gait features originates from the same subject given the matching score and the quality measure. Let $\xi \in \{S(same), D(different)\}$ be a label, and let $d_{VTM}(\boldsymbol{x}^A, \boldsymbol{x}^B)$ and $e(\boldsymbol{x}^A)$ be a matching score and a quality measure associated with $\boldsymbol{x}^A$. We then calculate a posterior probability distribution $P\left(\xi = S | d_{VTM}(\boldsymbol{x}^A, \boldsymbol{x}^B), e(\boldsymbol{x}^A)\right)$ over two-dimensional evidence space $(d_{VTM}(\boldsymbol{x}^A, \boldsymbol{x}^B), e(\boldsymbol{x}^A))$ based on the linear logistic regression (LLR) framework. More specifically, the logarithm of the posterior probability ratio is approximated by a weighted linear sum of the matching score and the quality measure as

$$\log\left(\frac{P(\xi = S | d_{VTM}(\boldsymbol{x}^A, \boldsymbol{x}^B), e(\boldsymbol{x}^A))}{1 - P(\xi = S | d_{VTM}(\boldsymbol{x}^A, \boldsymbol{x}^B), e(\boldsymbol{x}^A))}\right) = \Theta_0 + \Theta_1 d_{VTM}(\boldsymbol{x}^A, \boldsymbol{x}^B) + \Theta_2 e(\boldsymbol{x}^A), \qquad (10)$$

where $\boldsymbol{\Theta} = (\Theta_0, \Theta_1, \Theta_2)$ are parameters of the LLR. Once the parameters $\boldsymbol{\Theta}$ are trained by minimizing the objective function of the logistic loss proposed in [2], the posterior probability is calculated as

$$P\left(\xi = S | d_{VTM}(\boldsymbol{x}^A, \boldsymbol{x}^B), e(\boldsymbol{x}^A)\right) = \frac{1}{1 + e^{-(\Theta_0 + \Theta_1 d_{VTM}(\boldsymbol{x}^A, \boldsymbol{x}^B) + \Theta_2 e(\boldsymbol{x}^A))}} . \qquad (11)$$

4. Implementation

We chose an arbitrary view transformation model (AVTM) [38] from the VTM family because of its applicability to any combinations of arbitrary views. We constructed 3D gait volumes of 103 training subjects from multi-view synchronized silhouette sequences by a visual cone intersection technique. We then projected the 3D gait volumes into silhouettes of source and destination views using the associated projection matrices. Thereafter, we extracted frequency-domain features [35] as the gait features from the projected silhouette, and constructed the training matrix in Eq. (1). Note that frequency-domain features are phase-independent and that this fact enables us to skip troublesome phase synchronization processing in both view transformation and matching stages.

5. Experiment

5.1. Database

We used a subset of the publicly available OU-ISIR gait database, a large population data set called OULP-C1V2 (later called OULP for simplicity) [17]. In this database, individual subjects walked along a straight course, and images were captured using a single camera placed approximately 5 meters from the course. We chose a subset of 1,912 subjects captured by a calibrated camera and two image sequences per subject were available. Each image sequence was divided into four sets based on the observation azimuth angles, 55, 65, 75 and 85 deg, where 90 and 0 degs means side and frontal views, respectively. In this paper, we used the 55 deg set from the second image sequence per subject as a gallery, and the 75 and 85 deg sets from the first image sequence per subject as probes. Examples of the gallery and probe images are shown in Fig. 4, and gait features extracted from normalized silhouette sequences with the size of 44×64 pixels are shown in Fig. 5. The OULP is suitable for evaluating the accuracy of the cross-view matching, because the two image sequences per subject in OU-ISIR were captured in similar conditions (e.g., the same day, the same attire, and with natural facial expressions), namely excluding covariates other than view. Note that the OULP evaluation data set is different from the 3D gait volume data set for VTM training explained in Section 4.

5.2. Experimental setting

Because the parameter $\boldsymbol{\Theta}$ for the posterior probability calculation needed to be trained, we divided the OULP ran-

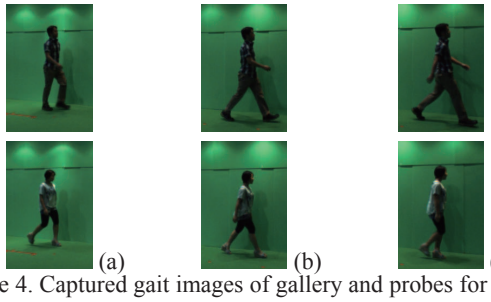

Figure 4. Captured gait images of gallery and probes for two subjects: (a) gallery gait images with 55-deg observation angle, (b) probe gait images with 75-deg observation angle, (c) probe gait images with 85-deg observation angle.

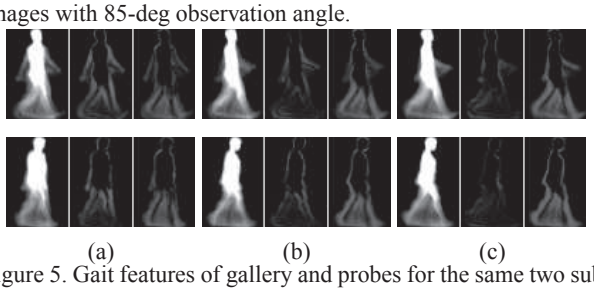

Figure 5. Gait features of gallery and probes for the same two subjects in Fig. 4: (a) gallery gait feature with 55-deg observation angle, (b) probe gait feature with 75-deg observation angle, (c) probe gait feature with 85-deg observation angle.

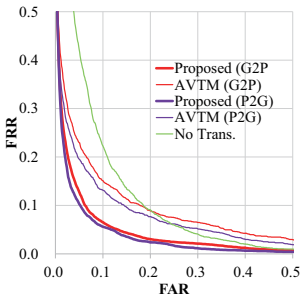

Figure 6. ROC curves: gallery of 55 [deg] and probe of 75 [deg]

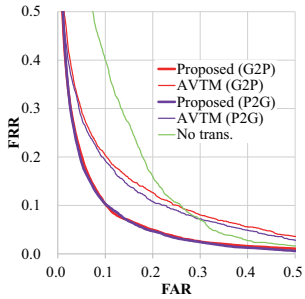

Figure 7. ROC curves: gallery of 55 [deg] and probe of 85 [deg]

Table 1. EERs [%]

Method		Probe [deg] 75	85
Proposed (G2P):	$d_{VTM}(\boldsymbol{x}^G, \boldsymbol{x}^P), e(\boldsymbol{x}^G)$	*8.0*	**10.1**
AVTM (G2P):	$d_{VTM}(\boldsymbol{x}^G, \boldsymbol{x}^P)$	13.1	15.4
Proposed (P2G):	$d_{VTM}(\boldsymbol{x}^P, \boldsymbol{x}^G), e(\boldsymbol{x}^P)$	**7.4**	*10.2*
AVTM (P2G):	$d_{VTM}(\boldsymbol{x}^P, \boldsymbol{x}^G)$	11.7	14.5
No trans.:	$d(\boldsymbol{x}^G, \boldsymbol{x}^P)$	14.3	18.6

domly into two disjoint groups, training and test sets, and performed two-cross validation tests. In order to remove any effects associated with the random grouping, we repeated the two-cross validations five times with different groupings.

The projection matrices associated with the 55, 75 and 85 deg views for the AVTM were calculated using the calibrated camera parameters and estimated subject positions.

For comparison purpose, we evaluated the following two methods:

- AVTM [38]
  The matching score $d_{VTM}(\boldsymbol{x}_\theta^G, \boldsymbol{x}_\phi^P)$ in Eq. (8) or $d_{VTM}(\boldsymbol{x}_\phi^P, \boldsymbol{x}_\theta^G)$ in Eq. (9) was directly used for recognition. In this method, the quality measure $e(\boldsymbol{x}_\theta^G)$ or $e(\boldsymbol{x}_\phi^P)$ was excluded.
- No transformation (No trans.)
  The matching score was directly calculated between original gait features with different views without the view transformation as $||\boldsymbol{x}_\theta^G - \boldsymbol{x}_\phi^P||_2$, and was used for recognition.

5.3. Experimental result

We calculated false acceptance rates (FARs) and false rejection rates (FRRs) in a verification scenario (one-to-one matching), and drew receiver operation characteristic (ROC) curves for recognition accuracy evaluation. Figures 6 and 7 show ROC curves for two settings; probe observation angles were 75 [deg] and 85 [deg], while the gallery observation was 55 [deg] in both settings. In addition, we also summarize equal error rates (EERs) in Table 1. Because recognition accuracy is somewhat dependent on the combination of the source and destination views, namely, dependent on whether the gallery feature or the probe feature is transformed in our experimental setup, we show both results in the figures and the table, and distinguish them by denoting them as "G2P" and "P2G", respectively.

From these results, we can observe that recognition accuracy for the 85-deg probe improves from 18.6 % to 15.4 and 14.5 % for G2P and P2G, respectively, and for the 75-deg probe improve from 14.3 % to 13.1 and 11.7 %, respectively. Moreover, the proposed quality-dependent method produces a further improvement. More specifically, the EERs for the 85-deg probe further improve to 10.1 and 10.2 %, and those for the 75-deg probe improve to 8.0 and 7.4 [%] for G2P and P2G, respectively. We can also confirm that the proposed method outperforms the benchmarks for all the operating points in the ROC curves.

We therefore conclude that the proposed method achieves better accuracy than the AVTM without the quality measure and the no transformation method.

6. Conclusion

We proposed a quality measure for cross-view gait recognition, and also a quality-dependent VTM framework

for better recognition accuracy. Since the success of the VTM relies on how well the joint subspace derived from the multi-view gait features of training subjects describes a gait feature of a test subject, we adopted the projection error of the source-view gait feature to the joint subspace as a quality measure. We then incorporated the quality measure into the LLR framework to calculate the posterior probability of the genuine subject.

The proposed method was evaluated against a subset of the OU-ISIR gait database, a large population data set, and the experimental results showed that the proposed method improved the recognition accuracy.

In this paper, we only use a projection error as a quality measure. However, we think that there are other factors that impact the matching scores (e.g., silhouette qualities of the original gait features). By using these factors, the proposed method can further improve the recognition accuracy.

Moreover, since we used the indoor OU-ISIR gait database with relatively clear silhouette sequences to evaluate the improvement of the proposed method and did not consider other covariates, in a way we investigated a kind of upper bound accuracy as a fundamental study. We therefore plan to evaluate the proposed method against more realistic outdoor gait databases such as the USF gait database [44] in our future work.

Acknowledgment

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (S) 21220003, "R&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society", Strategic Funds for the Promotion of Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government, and the JST CREST "Behavior Understanding based on Intention-Gait Model" project.

References

[1] M. Abdel-Mottaleb and M. Mahoor. Application notes - algorithms for assessing the quality of facial images. *IEEE Computer Intelligence Magazine*, 2(2):10–17, 2007.

[2] F. Alonso-Fernandez, J. Fierrez, D. Ramos, and J. Ortega-Garcia. Dealing with sensor interoperability in multi-biometrics: the upm experience at the biosecure multimodal evaluation 2007. In *Proc. SPIE 6994, Biometric Technologies for Human Identification IV*, Orlando, FL, USA, Mar. 2008.

[3] G. Ariyanto and M. Nixon. Marionette mass-spring model for 3d gait biometrics. In *Proc. of the 5th IAPR Int. Conf. on Biometrics*, pages 354–359, March 2012.

[4] C. BenAbdelkader. *Gait as a Biometric For Person Identification in Video*. PhD thesis, Maryland Univ., 2002.

[5] A. Bobick and A. Johnson. Gait recognition using static activity-specific parameters. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 423–430, 2001.

[6] R. Bodor, A. Drenner, D. Fehr, O. Masoud, and N. Papanikolopoulos. View-independent human motion classification using image-based reconstruction. *Image and Vision Computing*, 27(8):1194 –1206, 2009.

[7] T. Chen, X. Jiang, and W. Yau. Fingerprint image quality analysis. In *International Conference on Image Processing*, volume 2, pages 1253–1256 Vol.2, 2004.

[8] Y. Chen, S. Dass, and A. Jain. Localized iris image quality using 2-d wavelets. In D. Zhang and A. Jain, editors, *Advances in Biometrics*, volume 3832 of *Lecture Notes in Computer Science*, pages 373–381. Springer Berlin Heidelberg, 2005.

[9] J. Fierrez-Aguilan, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun. Kernel-based multimodal biometric verification using quality signals authors. In *Proc. the SPIE, Biometric Technology for Human Identification*, volume 5404, pages 544–554, 2004.

[10] H. Fronthaler, K. Kollreider, J. Bigun, J. Fierrez, F. Alonso-Fernandez, and J. Ortega-Garcia. Fingerprint image quality estimation and its application to multi-algorithm verification. *IEEE Trans. Information Forensic and Security*, 3(2):331–338, 2008.

[11] S. Garcia-Salicetti, N. Houmani, and B. Dorizzi. A novel criterion for writer enrolment based on a time-normalized signature sample entropy measure. *EURASIP Journal on Advances in Signal Processing*, 2009(1):964746, 2009.

[12] M. Goffredo, I. Bouchrika, J. Carter, and M. Nixon. Self-calibrating view-invariant gait biometrics. *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics,*, 40(4):997 –1008, aug. 2010.

[13] P. Grother and E. Tabassi. Performance of biometric quality measures. *IEEE Trans. Pattern Anallysis and Machine Intelligence*, 29(4):531–543.

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

[15] J. Han, B. Bhanu, and A. Roy-Chowdhury. A study on view-insensitive gait recognition. In *Proc. IEEE Int. Conf. on Image Processing*, volume 3, pages 297–300, Sep. 2005.

[16] M. A. Hossain, Y. Makihara, J. Wang, and Y. Yagi. Clothes-invariant gait identification using part-based adaptive weight control. In *Proc. the 19th Int. Conf. on Pattern Recognition*, pages 1–4, Tampa, Florida USA, Dec. 2008.

[17] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi. The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Trans. Information Forensics and Security*, 7(5):1511–1521, oct. 2012.

[18] Y. Iwashita, R. Baba, K. Ogawara, and R. Kurazume. Person identification from spatio-temporal 3D gait. In *Proc. International conference on Emerging Security and Technology*, pages 30–35, 2010.

[19] F. Jean, R. Bergevin, and A. B. Albu. Computing and evaluating view-normalized body part trajectories. *Image Vision Comput.*, 27(9):1272–1284, Aug. 2009.

[20] X. Jiang. Fingerprint image ridge frequency estimation by higher order spectrum. In *Proc. International Conference on Image Processing*, volume 1, pages 462–465 vol.1, 2000.

[21] A. Kale, A. Roy-Chowdhury, and R. Chellappa. Towards a view invariant gait recognition algorithm. In *Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance*, pages 143–150, 2003.

[22] E. Krichen, S. Garcia-Salicetti, and B. Dorizzi. A new probabilistic iris quality measure for comprehensive noise detection. In *Proc. the First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, 2007.

[23] K. Kryszczuk and A. Drygajlo. Improving biometric verification with class-independent quality information. *IET Signal Processing, Special Issue on Biometric Recognition*, 3(4):310–321, 2009.

[24] K. Kryszczuk and A. Drygajlo. Improving classification with class-independent quality measures: Q-stack in face verification. In S.-W. Lee and S. Li, editors, *Advances in Biometrics*, volume 4642 of *Lecture Notes in Computer Science*, pages 1124–1133. Springer Berlin Heidelberg, 2007.

[25] K. Kryszczuk, J.Richiardi, P.Prodanov, and A. Drygajlo. Error handling in multimodal biometric systems using reliability measures. In *Proc. the 13th European Conference on Signal Processing EUSPICO*, 2005.

[26] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang. Multiple views gait recognition using view transformation model based on optimized gait energy image. In *IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1058–1064, 2009.

[27] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Multi-view gait recognition based on motion regression using multilayer perceptron. In *Proc. International Conference on Pattern Recognition*, pages 2186–2189, 2010.

[28] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Support vector regression for multi-view gait recognition based on local motion feature selection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 974–981, 2010.

[29] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Gait recognition under various viewing angles based on correlated motion regression. *IEEE Trans. Circuits and Systems for Video Technology*, 22(6):966 –980, june 2012.

[30] T. H. W. Lam, K. H. Cheung, and J. N. K. Liu. Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognition*, 44:973–987, April 2011.

[31] L. Lee. *Gait Analysis for Classification*. PhD thesis, Massachusetts Institute of Technology, 2002.

[32] J. Liu, N. Zheng, and L. Xiong. Silhouette quality quantification for gait sequence analysis and recognition. *Signal Processing*, 89(7):1417 – 1427, 2009.

[33] N. Liu, J. Lu, and Y.-P. Tan. Joint subspace learning for view-invariant gait recognition. *IEEE Signal Processing Letters*, 18(7):431–434, 2011.

[34] J. Lu and Y.-P. Tan. Uncorrelated discriminant simplex analysis for view-invariant gait signal computing. *Pattern Recognition Letters*, 31(5):382 – 393, 2010.

[35] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *Proc. the 9th European Conf. on Computer Vision*, pages 151–163, Graz, Austria, May 2006.

[36] Y. Makihara and Y. Yagi. Silhouette extraction based on iterative spatio-temporal local color transformation and graph-cut segmentation. In *Proc. the 19th Int. Conf. on Pattern Recognition*, Tampa, Florida USA, Dec. 2008.

[37] D. S. Matovski, M. S. Nixon, S. Mahmoodi, and T. Mansfield. On including quality in applied automatic gait recognition. In *Proc. the 21st International Conference on Pattern Recognition*, pages 3272–3275, 2012.

[38] D. Muramatsu, A. Shiraishi, Y. Makihara, and Y. Yagi. Arbitrary view transformation model for gait person authentication. In *Proc. IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2012.

[39] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain. Likelihood ratio-based biometric score fusion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(2):342–347, 2008.

[40] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait*. International Series on Biometrics. Springer-Verlag, Dec. 2005.

[41] N. Poh, T. Bourlai, and J. Kittler. A multimodal biometric test bed for quality-dependent, cost-sensitive and client-specific score-level fusion algorithms. *Pattern Recognition*, 43(3):1094–1105, 2010.

[42] N. Poh and J. Kittler. A unified framework for biometric expert fusion incorporationg quality measure. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2013. (Accepted) downloaded from http://personal.ee.surrey.ac.uk/Personal/Norman.Poh/publications.

[43] N. Poh, J. Kittler, and T. Bourlai. Quality-based score normalisation with device qualitative information for multimodal biometric fusion. *IEEE Trans. Systems, Man, Cybernatics Part B : Systems and Humans*, 40:539–554, 2010.

[44] S. Sarkar, J. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005.

[45] G. Shakhnarovich, L. Lee, and T. Darrell. Integrated face and gait recognition from multiple views. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 439–446, 2001.

[46] A. Utsumi and N. Tetsutani. Adaptation of appearance model for human tracking using geometrical pixel value distributions. In *Proc. the 6th Asian Conf. on Computer Vision*, volume 2, pages 794–799, 2004.

[47] C. Yam, M. Nixon, and J. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5):1057–1072, 2004.

[48] H.-D. Yang and S.-W. Lee. Reconstruction of 3d human body pose for gait recognition. In *Proc. the IAPR Int. Conf. on Biometrics 2006*, pages 619–625, Jan. 2006.

[49] S. Yu, D. Tan, and T. Tan. Modelling the effect of view angle variation on appearance-based gait recognition. In *Proc. 7th Asian Conf. on Computer Vision*, volume 1, pages 807–816, Jan. 2006.