Arbitrary View Transformation Model for Gait Person Authentication

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

Abstract

Camera-based gait recognition is a useful method for authenticating a person from a distance, even if the resolution of the acquired images is not high. However, different views of the compared gallery and probe decrease the recognition accuracy. To solve this problem, we propose a gait based authentication method that uses an arbitrary view transformation scheme. The proposed method constructs a transformation matrix associated with the view of the set of gallery and probe using a 3D gait database composed of non-target multiple subjects' visual hulls. This matrix is used to transform the gallery gait features with varying views into features with the same view as the probe. Using this scheme, we can minimize the impact of the view difference. We evaluated the accuracy of the proposed method using a gait database composed of 52 subjects. The experimental results show that the proposed method is promising.

1. Introduction

Biometric person authentication methods are becoming more important in everyday life and are used in many situations such as access control, credit card verification, and surveillance. A wide variety of modalities such as fingerprints, finger or hand veins, voice, irises, faces, handwriting, keystrokes, and gait have been proposed and used [4]. Each modality has both advantages and disadvantages, and therefore, the appropriate modality must be selected depending on the situation or application.

In considering the usage for surveillance, gait is a promising modality because gait data can be acquired from a remotely-positioned camera, and used to authenticate a person even if the resolution of the acquired data is low. However, for the purpose of surveillance, the subject' s walking directions cannot be controlled. This leads to a difference in views, which in turn causes appearance changes in gait, making gait-based authentication more difficult. In fact, BenAbdelkader [1] and Yu et al. [17] reported that view changes cause a deterioration in gait recognition accuracy. Two different types of approach have been proposed for gait person authentication, appearance-based and model-based; and solutions against accuracy deterioration due to the view difference have been discussed in both approaches. In this paper, we focus on the solutions against appearance-based approaches. The solutions for model-based approaches are discussed in [2].

To cope with view direction changes, several approaches have been proposed. Kale et al. proposed a method to synthesize side view gait images from other arbitrary views [5]. In their approach, they assume that a person (3D object) can be represented by a planar object on a sagittal plane. This approach works well if the angle between the sagittal plane of the person and the image plane is small; however, where the angle is large, the accuracy is drastically degraded.

Visual hull based approaches have been proposed by Shakhnarovich et al. [15], Lee [10], and Iwashita et al. [3]. In these methods, the gait image of a virtual view is synthesized using an image-based visual hull of the target subject. From the visual hull, gait images of any arbitrary view can be synthesized. A limitation of the visual hull based approach is that visual hulls of all the target subjects are needed. To generate a 3D visual hull of a target subject, gait image sequences of the subject must be captured in advance using multiple synchronized cameras. As such, visual hull based approaches cannot be used if gait image sequences of the target subjects captured with multiple synchronized cameras are not available.

On the other hand, view transformation model (VTM)-based approaches have also been proposed, and these can be used in the above case. In this type of approach, a VTM that transforms the data of varying views into those of a different view is constructed using a training database. To construct the VTM , some methods use a matrix factorization process, adopting a singular value decomposition (SVD) technique [16, 11, 6], while others use regression [7, 8, 9]. Unlike the visual hull based approach, gait image sequences of the target subject captured with multiple synchronized cameras are not necessary; instead, a pre-existing gait database can be used to construct the VTM. The pre-existing gait database must be composed of gait image sequences of mul-

Table 1. Comparison of the methods with respect to view changes

Approach	Necessary training data	Arbitrary view
Proposed	non-target	applicable
Visual hull based	target	applicable
VTM based	non-target	not applicable

tiple subjects taken with multiple synchronized cameras; however, it does not need to include the gait data associated with the target person. Therefore, we can apply this type of approach to the case where only a single gait image sequence associated with the target person is available. A limitation of these methods is the applicability to an arbitrary view. Possible views for application are limited to several discrete views under which the training gait data are collected. In this paper, we refer to the list of available views as the *training view list*. In those cases where a view of the gallery or probe is not included in the training view list, these methods cannot be applied directly. A possible solution is to approximate the target view to the nearest one in the training view list, although this results in lower accuracy.

To overcome the limitation, we propose a method that uses an arbitrary view transformation scheme. The proposed method constructs view transformation matrices associated with views of the set of gallery and probe using a 3D gait database composed of non-target multiple subject's visual hulls. With the 3D gait database, 2D gait image data of arbitrary views can be generated by re-projections. Therefore, given the target views (views of the set of gallery and probe), 2D gait image data associated with the target views can be generated even if the target views are not included in the training view list. The 2D gait image data then enable the construction of the view transformation matrix associated with the target views. Using the matrix, we can transform the gallery gait features into gallery gait features with the same view as the probe. It should be noted that visual hulls associated with the target subjects are not necessary in the proposed method, as the VTM is constructed from visual hulls of non-target subjects. Therefore, our proposed method can be used in the case that visual hulls associated with the target subject are not available. Moreover, thanks to the visual hulls of non-target subjects and the view transformation matrix, our proposed method is applicable to transformation to/from an arbitrary view. A comparison of the proposed method and visual hull and view transformation based approaches is given in Table 1.

The rest of this paper is organized as follows. The proposed algorithm is explained in Section 2. The experimental setup and results are reported in Section 3, and our conclusions are given in Section 4.

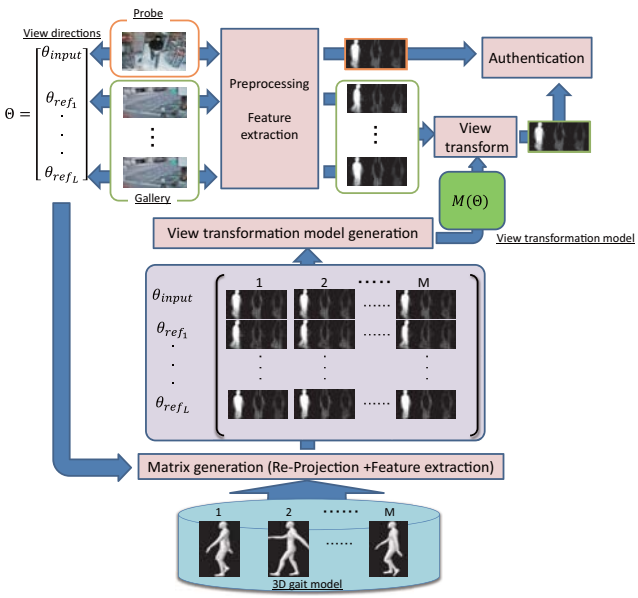

Figure 1. Overview of the algorithm

2. Authentication algorithm

2.1. Overview

Figure 1 shows the whole algorithm for the proposed method. As described in the introduction, the accuracy of gait recognition algorithms degrades if the views of the gallery and probe differ. Therefore, we transform the gallery gait features with varying views into gallery gait features with the same view as the probe. This transformation is performed using the view transformation matrix associated with the views of the set of gallery and probe. This matrix is constructed using the 3D gait database composed of gait visual hulls of multiple non-target subjects.

Assume that there are L views of gait image sequences per subject in the gallery and one gait image sequence as the probe. Let $\mathbf{s}_{G_n} = (s^1_{G_n}, ...s^L_{G_n})$ be the L views of gait image sequences associated with the gallery of the n-th subject, and s_P be a gait image sequence associated with the probe. After preprocessing, gait features are extracted from the preprocessed gait image sequences. Let $a^l_{G_n}$ be a feature vector extracted from $s^l_{G_n}$, and a_P be a feature vector extracted from s_P. Let $\theta_{ref^n_l}$ and θ_{input} be the views associated with the l-th gallery of the n-th subject and the probe, respectively. For simplicity, we omit the subindex n and express $\theta_{ref^n_l}$ as θ_{ref_l}. Let $\mathbf{\Theta}$ be a view vector composed of views associated with the set of gallery and probe. We set this as

$$\mathbf{\Theta} = [\theta_{input}, \theta_{ref_1}, ..., \theta_{ref_L}]^T. \quad (1)$$

Here, T is the matrix transpose operator. Given view vector $\mathbf{\Theta}$, we generate a gallery gait feature with view θ_{input}

from the multiple gallery gait features with views θ_{ref_l} as follows:

$$\hat{a}_{G_n}(\theta_{input}) = M(\mathbf{\Theta}) \begin{bmatrix} a^1_{G_n} \\ \vdots \\ a^L_{G_n} \end{bmatrix}. \quad (2)$$

Here, $M(\mathbf{\Theta})$ is a VTM (matrix) associated with $\mathbf{\Theta}$, and $\hat{a}_{G_n}(\theta_{input})$ is the generated gallery gait feature with view θ_{input}. The way matrix $M(\mathbf{\Theta})$ is constructed is described in Subsection 2.2.

Having obtained the generated gallery gait feature, we compare the probe gait feature with the generated gallery gait feature and calculate a dissimilarity score.

Finally, a decision is made based on the calculated dissimilarity score.

2.2. View transformation

Assume that there are M subjects' gait visual hulls (3D gait sequences) in the training database. Using the 3D gait sequences, we can generate 2D gait image sequences of arbitrary views by re-projection.

Given a view vector $\mathbf{\Theta} = [\Theta_1, ..., \Theta_K]^T$, all the 3D gait sequences are re-projected onto the K views of the 2D gait image sequences with views $\Theta_k, k = 1, 2, ..., K$, respectively, and gait features are extracted from the re-projected 2D gait image sequences. Using the extracted features, we construct the view transformation matrix $M(\mathbf{\Theta})$.

We briefly describe the formulation of the VTM in a similar way to that given in [16]. Note that we apply the VTM to the phase-invariant gait features extracted from the gait image sequences[1].

Let $\mathbf{a}^m_T(\Theta_k)$ be a N_A dimensional feature vector of the m-th subject; this vector is extracted from the 2D gait image sequences of the m-th subject re-projected onto view Θ_k. Here, although the type of gait feature is the same as those of the gallery and probe, to avoid confusion, we use a different symbol $\mathbf{a}^m_T(\theta)$ for the gait feature vectors originating from the 3D gait sequences in the training database, to those of the gallery a^l_G and probe a_P . Using the gait feature vectors associated with the training database, we construct a matrix, where the rows indicate views and the columns each represent a subject in the training database, which is then decomposed by SVD as

$$\begin{bmatrix} \mathbf{a}^1_T(\Theta_1) & \cdots & \mathbf{a}^M_T(\Theta_1) \\ \vdots & \ddots & \vdots \\ \mathbf{a}^1_T(\Theta_K) & \cdots & \mathbf{a}^M_T(\Theta_K) \end{bmatrix}$$
$$= USV^T = \begin{bmatrix} P(\Theta_1) \\ \vdots \\ P(\Theta_K) \end{bmatrix} \begin{bmatrix} \mathbf{v}^1 & \cdots & \mathbf{v}^M \end{bmatrix}, \quad (3)$$

[1]The transformation model can be applied directly to gait images if the phases of the gait image sequences are synchronized. However, for phase-invariant gait features, this model can be applied without phase synchronization.

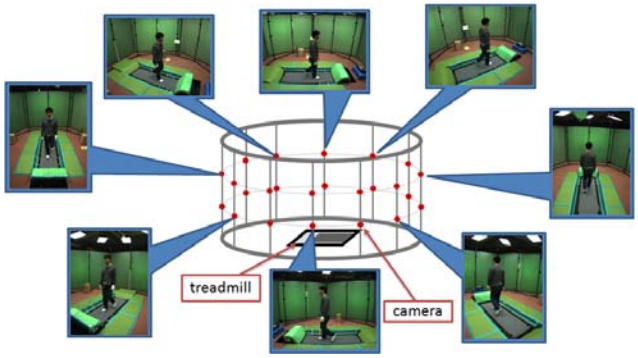

Figure 2. Data acquisition setup. Gait image sequences are taken from 24 synchronized cameras.

where U is the $KN_A \times M$ orthogonal matrix, V is the $M \times M$ orthogonal matrix, S is the $M \times M$ diagonal matrix composed of singular values, $P(\Theta_k), k = 1, 2, ..., K$ is the $N_A \times M$ submatrix of US, and $\mathbf{v}^m$ is the M dimensional column vector.

Vector $\mathbf{v}^m$ is an intrinsic feature vector of the m-th subject and is independent of views. The submatrix $P(\Theta_k)$ is a projection matrix from the intrinsic vector $\mathbf{v}$ to the feature vector for view Θ_k, and is common for all subjects; that is, it is independent of the subject. Thus, feature vector $\mathbf{a}^m_T(\Theta_i)$ for view Θ_i of the m-th subject is expressed as

$$\mathbf{a}^m_T(\Theta_i) = P(\Theta_i)\mathbf{v}^m. \quad (4)$$

Then, the feature vector transformation from view Θ_j to Θ_i is easily obtained by the least squares method as

$$\mathbf{a}^m_T(\Theta_i) = P(\Theta_i)P^+(\Theta_j)\mathbf{a}^m_T(\Theta_j), \quad (5)$$

where $P^+(\Theta_j)$ is the pseudo inverse matrix of $P(\Theta_j)$.

Sometimes, transformation from one view may be insufficient in practice because movements orthogonal to the image plane are degenerated in the silhouette image. In this case, by using feature vectors of more than one view (let these be $\Theta_j, j = 1, 2, ..., L$), we can generate a feature vector for view Θ_i more precisely as

$$\mathbf{a}^m_T(\Theta_i) = P(\Theta_i) \begin{bmatrix} P(\Theta_1) \\ \vdots \\ P(\Theta_L) \end{bmatrix}^+ \begin{bmatrix} \mathbf{a}^m_T(\Theta_1) \\ \vdots \\ \mathbf{a}^m_T(\Theta_L) \end{bmatrix}. \quad (6)$$

By setting $\Theta_i = \theta_{input}$ and $\Theta_j = \theta_{ref_j}$, we can generate a feature vector with view θ_{input} from the L views' feature vectors with views $\theta_{ref_j}, j = 1, 2, ..., L$.

3. Experiments

3.1. 3D gait model

Figure 2 shows the setup for data acquisition. We constructed a circular studio with a treadmill in the center.

Twelve poles were set 30 degrees apart around the circumference, and two cameras were set at a height of 130 cm and 200 cm on each pole. Thus, in total we positioned 24 cameras, which were temporally synchronized. The resolution of each image captured by the cameras was 640 by 480 pixels, with 60 frames captured per second.

Using this studio with the 24 synchronized cameras, we collected gait image sequences of 20 subjects from multiple views (associated with the camera positions). During the collection of gait image sequences, the subjects were asked to walk on the treadmill in a natural manner. Here, projection matrices for the individual cameras are computed by means of camera calibration. The lens distortion is calibrated by a non-parametric calibration method [14]. A silhouette of a gait image in each view in each frame is extracted using graph-cut based segmentation [13], and thereafter, a visual cone intersection technique is used to reconstruct the 3D gait model.

3.2. Evaluation database

We collected gait image sequences from 53 subjects for evaluation. The gait image sequences were taken in the same way as those used for generating the 3D gait model. Each subject was asked to walk on the treadmill in a natural manner while gait image sequences were captured. The same 24 cameras described in Section 3.1 were used to collect the images for this database.

3.3. Gait authentication algorithm

Frequency-based features [11] were considered as the gait features in this experiment.

The Euclidian distance calculated as

$$d(\mathbf{s}_{G_n}, s_P) = ||\hat{a}_{G_n}(\theta_{input}) - a_P|| \tag{7}$$

was used as the dissimilarity score between the gallery and probe. Based on the score, subject X who gave the probe s_P was decided by

$$X \text{ is identified as } I\text{th subject} \quad \text{where} I = \underset{n}{argmin}\, d(\mathbf{s}_{G_n}, s_P), \tag{8}$$

in identification mode, and

$$X = \begin{cases} \text{accepted as } n\text{th subject if } d(\mathbf{s}_{G_n}, s_P) < threshold \\ \text{rejected} \qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{9}$$

in verification mode.

3.4. Experimental setup

The following two methods were evaluated together with the proposed method for the sake of comparison:

- No-transformation method (Notrans)
  Probe gait features are compared directly with the gallery gait features without any transformation.
- Discrete view transformation model-based method (DVTM)
  Gallery gait features are transformed into the nearest view of the probe's views in the training view list using a discrete view transformation matrix [11], and probe gait features are compared with the transformed gallery features.

The following two experimental setups are considered in this paper:

1. Preliminary experiment
   We compared the original gait features (extracted from the gait image sequences of the target view) with the transformed features generated from gait features with different views by visual inspection, as a qualitative analysis.
2. Main experiment
   We considered the situation in which a target view associated with the probe gait image sequences was *not* included in the training view list, and evaluated the accuracy of the proposed method. We also evaluated the accuracies of compared methods under the same circumstances.

In all experiments, we assumed that a gait image sequence of one view was given per subject in the gallery ($L = 1$). In this paper, we define a view θ using two parameters: azimuth ϕ and height ψ associated with the camera position. For example, $\theta = (240, 200)$ means that the images of view θ were captured by a camera set at an azimuth of 240 degrees, and a height of 200 cm. The azimuth of the camera positioned in front of the subject was defined as 0 degrees, and this increased in a counterclockwise direction.

3.5. Evaluation criteria

Receiver operating characteristic (ROC) curves and equal error rates (EERs) were used to evaluate the accuracy of verification, while cumulative matching characteristic (CMC) curves were used to evaluate the accuracy of identification.

3.6. Experimental results

3.6.1 Preliminary experiment

Figure 3 shows examples of the transformed gait features of a subject. The first column shows the gait feature extracted from the original image with different views. The second, third, fourth, and fifth columns show the transformed gait features. In each column, 12 gait features associated with the 12 views (with an interval of 30 degrees) are displayed. In the second, third, fourth, and fifth columns, the reference data (framed by a red line) are transformed into the gait features of the associated views (framed in blue). By comparing the features in each row, we can see that the reference gait features are successfully transformed into those of the associated views.

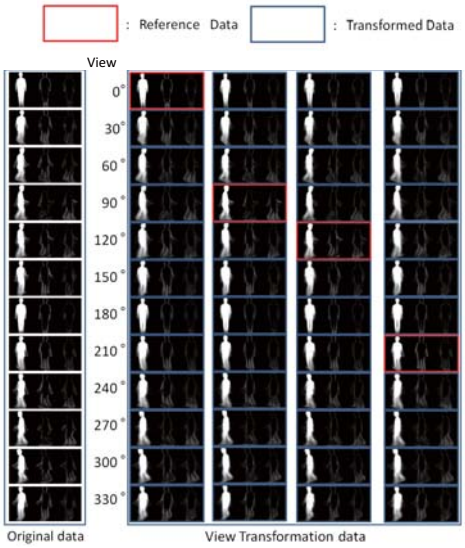

Figure 3. Transformed gait features. The first column shows original gait features. The second, third, fourth, and fifth columns show the transformed gait features; in each column, gait features framed in blue are generated from a gait feature framed in red using view transformation model.

3.6.2 Main experiment

We set the view associated with the probe gait image sequences (input view) as $\theta_{input} = (330, 200)$, and evaluated the accuracy by changing the gallery gait image sequences with different views (reference view). Note that the purpose of this experiment was to evaluate the accuracy under the condition that the target view was not included in the training view list. In this experiment, we assumed that the reference views were included in the training view list, but the input view was *not*. Under these circumstances, the DVTM could not transform the gait features of the reference view into those of the input view, because the transformation matrix for the input view could not be constructed. The DVTM could only transform the gait features of the reference view into those of the nearest view to the input view, and use these for authentication[2]. As for the nearest views, we consider the following three in this experiment: $\theta_{input'_1} = (300, 200), \theta_{input'_2} = (330, 130)$, and $\theta_{input'_3} = (360, 200)$.

Figure 4 shows the CMC curves associated with the two reference views, while Figure 5 shows the ROC curves for the same view settings. The EERs are summarized in Figure 6. From these figures, it is clear that the proposed arbitrary view transformation model based method (AVTM) outperforms the discrete VTM in many cases, while the proposed method outperforms the no-transformation method

[2]In the case that the input view is equal to reference view, DVTM uses original features without transformation.

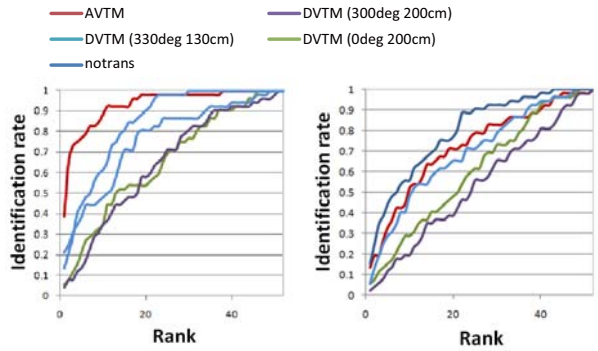

Figure 4. CMC curves. (Left) $\theta_{input} = (330, 200)$ and $\theta_{ref} = (150, 200)$. (Right) $\theta_{input} = (330, 200)$ and $\theta_{ref} = (180, 200)$.

where $\theta_{ref} = (30, 200)$, $(60, 200)$, $(120, 200)$, $(150, 200)$, and $(300, 200)$. On the other hand where $\theta_{ref} = (180, 200)$ and $(210, 200)$, the accuracy of the proposed method is better than that of the discrete view transformation method, but worse than that of the no-transformation method. If we assume that the gait image is a weak perspective projection and that gait has mirror symmetry, the gait features of $\phi = 330$ degrees coincide with those of $\phi = 210$ degrees as reported in literature [12]. This is why the no-transformation method achieves better accuracy than the proposed method with $\theta_{ref} = (210, 200)$. In these cases, the accuracy of the proposed method can be improved by adding a virtual view [12].

4. Conclusion

We proposed a gait based person authentication method that uses an arbitrary view transformation scheme to decrease the accuracy deterioration due to view changes. In the proposed method, arbitrary view transformation matrices can be constructed using a 3D gait model of multiple (non-target) subjects. In contrast to the previously proposed visual hull based approaches, a visual hull of the target subject is not necessary. This method can thus be used, even if only a single gait image sequence of the target subject is available. As such, the proposed method can be used in many real situations. Moreover, differing from the discrete view transformation based approach, the proposed approach can be applied to an arbitrary view. Therefore, the proposed method achieves better accuracy than the discrete view transformation based approach if the target view is not included in the training view list. In fact, the experimental results show that the proposed method works reasonably well compared with the discrete view transformation based method. However, experimental results also show that for some views, the accuracy of the proposed method is worse than that of the no-transformation method. In these views, the accuracy of the proposed method can be improved by adding a virtual reference view based on the geometric

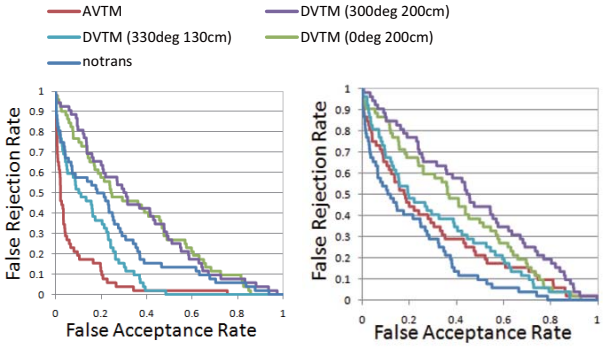

Figure 5. ROC curves. (Left) $\theta_{input} = (330, 200)$ and $\theta_{ref} = (150, 200)$, (Right) $\theta_{input} = (330, 200)$ and $\theta_{ref} = (180, 200)$

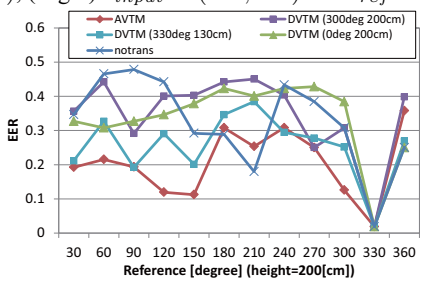

Figure 6. EERs comparison. EERs of proposed and compared methods are evaluated by varying the Azimuth associated with reference data from 30 to 360 degree. The azimuth of input is set to 330 degree, and heights of input and reference are set to 200 cm.

structure [12]. This is our future work.

Acknowledgment

This work was partly supported by JSPS KEKENHI Grant Number 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] C. BenAbdelkader. *Gait as a Biometric For Person Identification in Video*. PhD thesis, Maryland Univ., 2002.

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

[3] 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.

[4] A. K. Jain, P. Flynn, and A. A. Ross. *Handbook of Biometrics*. Springer Science+Business Media, LLC., 2008.

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

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

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

[8] 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. CVPR*, pages 974–981, 2010.

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

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

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

[12] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Which reference view is effective for gait identification using a view transformation model? In *Proc. of the IEEE Computer Society Workshop on Biometrics 2006*, New York, USA, Jun. 2006.

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

[14] R. Sagawa, M. Takatsuji, T. Echigo, and Y. Yagi. Calibration of lens distortion by structured-light scanning. In *Proc. 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1349–1354, Edmonton, Canada, August 2005.

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

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

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