A Geometric View Transformation Model using Free-form Deformation for Cross-view Gait Recognition

Hazem El-Alfy*, Chi Xu‡*, Yasushi Makihara*, Daigo Muramatsu* and Yasushi Yagi*
*Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan
‡School of Computer Science and Engineering,
Nanjing University of Science and Technology, Nanjing, China
{hazem, xu, makihara, muramatsu, yagi}@am.sanken.osaka-u.ac.jp

***Abstract*—Gait is a commonly used behavioral biometric that proved to be successful for authenticating people at a distance, however, its recognition accuracy severely deteriorates due to view changes. We therefore propose a Geometric View Transformation Model (GVTM) to enhance the robustness of gait recognition under cross-view conditions. Specifically, we train a subject-independent warping field with a free-form deformation framework which geometrically transforms gait features from two different views into those from an intermediate view. We then apply it to gait features of a test subject to register them and subsequently match them under the same intermediate view. Unlike existing appearance-based view transformation models that may corrupt the gait features, the proposed GVTM does not corrupt them because it preserves their spatial proximity. In addition, the GVTM can transform features more flexibly than simple weak perspective projection-based geometric approaches and more efficiently than 3D model-based approaches. We conduct experiments on the OU-ISIR Large Population gait dataset, the largest such database, and show that the proposed method outperforms state-of-the-art accuracy of generative and discriminative approaches under both identification and verification scenarios.**

***Keywords*-gait recognition; view-invariance; free-form deformation**

I. Introduction

The increased need for person authentication methods that do not require the cooperation of subjects has motivated many researchers to use gait as a biometric. In that regard, gait has established itself as a promising tool in forensic applications [1], in particular when subjects are captured at a distance, where conventional biometrics such as DNA, fingerprinting, iris, or face, are unavailable without subject collaboration. Gait recognition, however, often suffers from large intrasubject variations due to covariates such as view [2], clothing [3], walking speed [4], etc... Among these, view variation is particularly important since it is typical in CCTV videos. It affects the appearance 1 and hence, significantly reduces the recognition accuracy.

Previous approaches to cross-view gait recognition mainly fall into two families: geometric approaches and training-based approaches. 3D model-based approaches are typical instances of the geometric approaches. Those approaches build 3D models of gallery subjects by fitting a standard human body model to a single-view image [5] or by a visual intersection method from multi-view images [6]. Because 2D silhouettes/images from arbitrary views can be generated once we build the 3D models, the gallery 3D models are projected to 2D silhouettes/images from the same view as the probe subjects to do matching under the same conditions. This makes the 3D model-based approaches naturally view-invariant. However, it is generally difficult to get highly accurate 3D models from a single view and it is unlikely that multi-view images are available for uncooperative subjects such as a perpetrator or a suspect. In addition, the 3D model-based approaches require a relatively high image resolution to get reasonable human model fitting results and incur high computational costs.

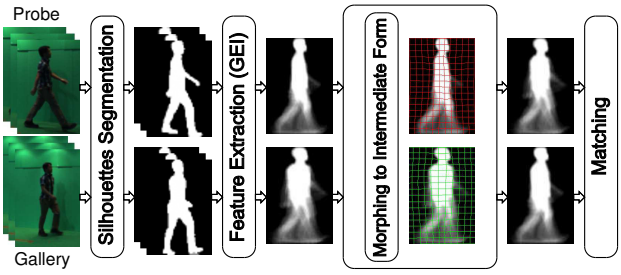

Figure 1. General overview of our approach.

Other geometric approaches exploit assumptions via the sagittal plane, which separate a human body into left and right parts. For example, Kale et al. [7] assume perspective projection where a human body is well approximated on the sagittal plane with no thickness. They then project silhouettes from front-oblique view into those from a canonical view, i.e., side view, based on the above mentioned assumption and then match a pair of gallery and a probe under the same canonical view. Jean et al. [8] approximate a walking trajectory by piecewise linear sub-trajectories, then, organize piecewise canonical vertical planes which intersect with the sub-trajectories on the ground plane and which also coincide with the sagittal plane. They then match a pair of gallery and probe trajectories of the top of head and bottom of feet

2327-0985/17 $31.00 © 2017 IEEE
DOI 10.1109/ACPR.2017.153

points, based on the piecewise canonical vertical planes for view invariance. These frameworks are acceptable when a forming angle between the sagittal plane and an image plane, i.e., view angle is close to the side view. On the other hand, when the view angle is close to the frontal view, they do not work well because the above mentioned assumptions do not hold.

The second family of approaches, i.e. the training-based approaches, further fall into two categories: generative approaches and discriminative approaches. As for generative approaches, view transformation models (VTMs) are trained from pairs of gait features of cooperative training subjects (e.g., students and staff in a lab.) from different views, using singular value decomposition [2] or support vector regression [9]. The trained VTMs are then used to transform the gait features of an uncooperative test subject (e.g., a perpetrator or a suspect) from a source view (e.g. gallery) to those of a destination view (e.g. probe). The VTM approaches usually exploit appearance-based gait features such as gait energy image (GEI) [10] and then unfold the image into a vector in a raster-scan way, where spatial proximity is disregarded. Hence, the transformed features may be corrupted (e.g., non-human-like gait features may be generated) when a test subject is not well represented by the training subjects [11].

The discriminative approaches aim at learning the view-invariant subspace or metrics without the above mentioned transformation. For example, Lu et al. [12] employ uncorrelated discriminant simplex analysis for subspace learning, while Martín-Félez et al. [13] employ rank support vector machine (Rank SVM) for metric learning. In addition, due to the recent success of deep learning-based approaches in many computer vision and pattern recognition fields, approaches to convolutional neural network-based cross-view gait recognition have also been proposed in [14], [15]. The discriminative approaches, however, work poorly when a large misalignment of the gait features (such as GEI) occurs due to view differences, because it is generally difficult to find a common subspace or metric robust against such non-registered gait features.

Taking a glance at the face recognition community, we find that geometric registration is commonly used as a preprocessing step to get pose invariance. The nice thing about geometric registration is its flexible deformation (considering prior knowledge of the face), in addition to its high generalization capability, i.e., the geometric deformation of a test subject is relatively well represented even with a limited number of training subjects. After surveying the literature on cross-view gait recognition, we notice that this idea of geometric registration has, however, not yet been brought into the gait recognition community despite its effectiveness.

We therefore propose a training-based and a geometric deformation-based approach to cross-view gait recognition, called a geometric view transformation model (GVTM). Specifically, we train a common geometric warping field based on free-form deformation (FFD) using gait features of training subjects captured from pairs of different views, and then apply the GVTM to register a pair of gait features of a test subject from different views (see Fig. 1). Our contributions in this paper are thus summarized as follows.

1. The GVTM for cross-view gait recognition

We introduce the GVTM as a geometric registration-based approach to cross-view gait recognition, which is the first time to be used in the gait recognition community to the best of our knowledge. Unlike the existing generative approaches which may corrupt appearance-based gait features [2], [9], [11], the proposed geometric deformation keeps the spatial proximity relationship and hence, never corrupts the appearance-based gait features. In addition, the proposed method realizes a good tradeoff between simple geometric approaches based on a too strong weak perspective assumption [7], [8] and 3D model-based approaches [5], [6]. In other words, it is more flexible than the simple geometric transformations but also, more efficient and more easily applicable than the 3D model-based approaches.

2. State-of-the-art accuracy on robust gait recognition

We achieve state-of-the-art accuracy on cross-view gait recognition with the OU-ISIR large population dataset [16], the largest publicly available gait database, compared with the existing generative approaches. In addition, because the proposed method does not include the discriminative approaches as a post-process, a combination of the proposed GVTM and a standard discriminative approach outperform the state-of-the-art accuracy of the existing discriminative approaches as well.

II. Cross-view Gait Recognition using a GVTM

Given a pair of input image sequences (probe and gallery), captured from different views, we extract gait features, register them into an intermediate view, then feed them into a matching module as detailed below and illustrated in Fig. 1.

A. GEI Templates

We employ GEI as an appearance-based gait feature due to its simplicity yet effectiveness. As an initial step, input images undergo a background subtraction process using graph-cut segmentation [17]. The heights of the segmented silhouettes are normalized and the locations of their centroids smoothed and aligned at the center of the images [16]. Gait periods are then determined so as to maximize the normalized auto-correlation along with the temporal axis of the size-normalized silhouette sequences [2]. Finally, the GEI templates are computed as the average of the silhouettes over one gait cycle. Specifically, given the size-normalized silhouette sequence as $I(x, y, t)$, where $I(x, y, t)$ indicates silhouette values at the position (x, y) at the t^{th} frame (i.e., 0

and 255 for background and foreground, respectively), and the detected gait cycle n [frame], the GEI is computed as

$$G(x,y) = \frac{1}{n}\sum_{t=1}^{n} I(x,y,t), \quad (1)$$

where $G(x,y)$ indicates an energy value at the position (x,y).

B. GVTM

Next, we introduce the GVTM to transform gait features captured from different views into those under the same view for better matching. More specifically, we employ an FFD framework [18] because it is highly flexible and preserves the proximity of adjacent regions. It is thus suited to the transformation of human body silhouettes.

Given gait features from two different views, we usually transform a gait feature from a view (let it be a source view) into that from the other view (let it be a target view) and then match the transformed gait feature and the original gait feature under the same target view. This however raises asymmetry between two views, and hence we often consider a bi-directional transformation, i.e., from the source view to the target view as well as from the target view to the source view [19]. Moreover, in the cross-view gait recognition community, the effectiveness of the transformation of gait features from two different views into those of an intermediate view is well known [20]. We therefore also consider transformation between the source/target view and the intermediate view because of its effectiveness as well as symmetric property.

To achieve this task, we assign a set of control points on a lattice on the GEI templates. The lattice size is experimentally determined. We then define displacement vectors $\boldsymbol{u}$ on those points to represent the transformation. The displacements are then linearly interpolated to construct a full warping field throughout the entire GEI areas. We denote the full warping field from an intermediate-view GEI to a source-view GEI as $F(\boldsymbol{u})$ and that to the reverse direction, i.e., from an intermediate-view GEI to a target-view GEI by $F(-\boldsymbol{u})$. Let a pair of source and target GEIs, corresponding to the j^{th} subject, be G_j^S and $G_j^T, (j = 1, \ldots, N)$, where N is the number of training subjects.

We then estimate the optimal displacement $\boldsymbol{u}^*$ by minimizing the sum of the differences between the intermediate-view GEIs transformed from the source and target views as

$$\boldsymbol{u}^* = \min_{\boldsymbol{u}} E(\boldsymbol{u}) \quad (2)$$

$$E(\boldsymbol{u}) = \sum_{j=1}^{N} \|G_j^S \circ F^{-1}(\boldsymbol{u}) - G_j^T \circ F^{-1}(-\boldsymbol{u})\|_2 + \lambda R(\boldsymbol{u}), \quad (3)$$

where $\circ$ represents the transformation operator, $R(\boldsymbol{u})$ is a regularization term to enforce smoothness of displacement between neighboring control points based on linear elasticity, and λ is a coefficient for the regularizer. Iterative gradient descent methods are employed to solve the above equations. Details of the FFD approach are presented in [19].

C. Matching

Once the probe and gallery gait features have been transformed (later, we use terms "transform" and "morph" interchangeably) into intermediate ones, we perform matching between pairs of those morphed features. We use the Euclidean distance (Eq. 4) as a simple yet efficient and broadly accepted metric. Naturally, smaller distances correspond to more similar pairs of subjects, and vice versa. Given a pair of gait features, i.e., G^S from the source view and G^T from the target view, the distance between the morphed features is computed as

$$D(G^S, G^T) = \|G^S \circ F^{-1}(\boldsymbol{u}^*) - G^T \circ F^{-1}(-\boldsymbol{u}^*)\|_2. \quad (4)$$

III. Experiments

A. Data sets

To validate our approach, we use a subset of the OU-ISIR Large Population dataset (OULP, Version 2) [16]. To compare our method to recently published approaches, we use the same subset of 1,912 subjects that they used. The subjects are captured, using calibrated cameras, from four different view angles: 55°, 65°, 75° and 85°, and that enables us to evaluate the proposed method under various levels of difficulty. The dataset provides two image sequences for each viewpoint: one for probe and one for gallery. Image sizes are normalized to 128×88 pixels.

B. Qualitative Evaluation of the GVTM

To better understand the effectiveness of the proposed GVTM, we illustrate the morphing process in Fig. 2 for two situations with differing levels of difficulty: a gallery subject viewed at 55° against probes at 65° (easier) and 85° (harder).

The original GEIs from two different views (Figs. 2(a) and (d)) are transformed into intermediate views (Figs. 2(c) and (f)) using the GVTM which is shown by the warping fields from the source view (Fig. 2(b)) and the target view (Fig. 2(e)). The difference images of two original GEIs (Fig. 2(g)) show that they have relatively large difference due to the view variations, especially for the difficult case (i.e., 55° vs. 85°). After transformation into intermediate views, the differences are significantly reduced in both situations (Fig. 2(h)), which demonstrates the effectiveness of the proposed GVTM.

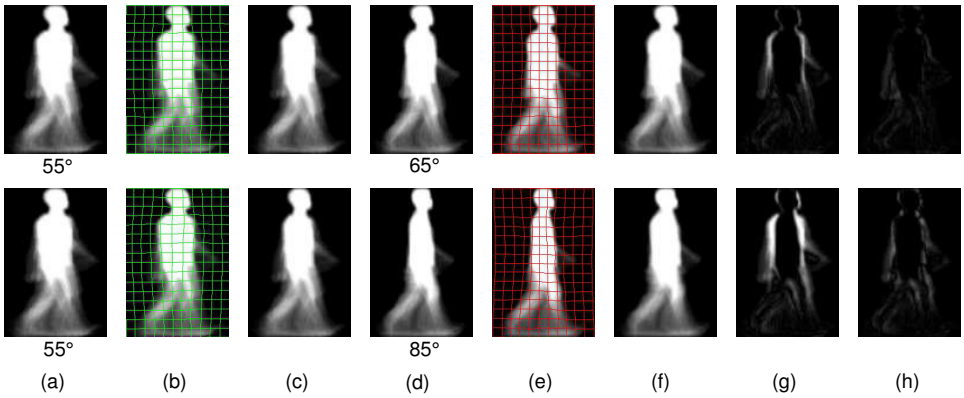

Figure 2. Morphing improves the matching quality. First row matches a 55°-view gallery to a 65°-view probe while second row is for 55° vs. 85° matching. (a) Original gallery (55°). (b) Gallery warping field. (c) Morphed gallery. (d) Original probe. (e) Probe warping field. (f) Morphed probe. (g) Difference of original images. (h) Difference of morphed images.

C. Comparison with Generative Approaches

We compare the recognition accuracy of our proposed GVTM against recent generative view transformation models. Specifically, we consider the published results of quality-dependent arbitrary VTM (QVTM) [11] and those of transformation consistency measures (TCM) [21]. We also report direct matching results using GEI [10] as a baseline. We evaluate using two-fold cross-validation. The validation subset of the OULP database is randomly partitioned into two disjoint subsets, one for training and one for testing, with 956 subjects in each. To limit any influence that random splitting may have on performance, the two-fold cross-validation is repeated five times, each with a different random partition. For a fair comparison of results, we use the exact same partitions used in [21], [11], which are also publicly available on the "protocols and benchmarks" section of the dataset's homepage.

We evaluate the recognition accuracy of our approach under both verification and identification scenarios. We carry out exhaustive experiments for all the view combinations. For verification, we compute the false acceptance / rejection rates (FAR / FRR) and report the equal error rates (ERR) as a measure of the verification ability. The EER is obtained by combining all the results obtained from the five random partitions mentioned earlier. As for the identification task, we compute the rank-1 identification rate for each partition, then average all of them into an overall rate that measures the identification capability of our method.

We summarize the results of the generative approaches in Tables I and II. We also display CMC and ROC curves in Fig. 3 for the most challenging view pair of 55°-gallery vs. 85°-probe. It is clear that the accuracy of our GVTM approach is considerably higher than the state-of-the-art generative approaches in both identification and verification scenarios. On a side note, we observe that direct matching

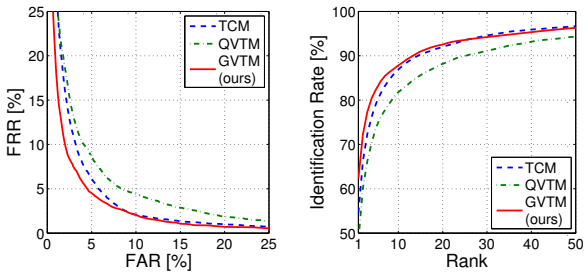

Figure 3. Comparison with generative approaches for the view pair 55°-gallery vs. 85°-probe. (a) ROC curves. (b) CMC curves.

with GEI performs reasonably well at near fronto-parallel views. That is the case since GEI has long been established to succeed in capturing the dynamics of limbs motion when captured from the side.

D. Comparison with Discriminative Approaches

We also evaluate our GVTM approach against the latest published results of discriminative methods. In this respect, we compare our approach to the basic linear discriminant analysis (LDA) [22], the generalized multi-view linear discriminant analysis (GMLDA) [23], the multi-view discriminant analysis (MvDA) [24] and deep learning (GEINet) [14]. To contrast our results with the published ones, we use the same protocol used in [24]. Here also, a subset of the OULP dataset, consisting of 1912 subjects, is randomly partitioned into a training and a testing sets of equal sizes. However, the experiment is performed only once on that particular partition.

For a fair comparison with discriminative approaches, we first use PCA on the morphed images to reduce the feature dimension while retaining 99% of the variance. LDA is then

Table I
COMPARISON OF THE EER [%] OF OUR PROPOSED GVTM APPROACH TO THOSE OF BENCHMARK GENERATIVE APPROACHES FOR ALL VIEW GALLERY/PROBE VIEW COMBINATIONS. BOLDFACE FONT INDICATES THE BEST PERFORMANCE AND BOLD ITALICS ARE USED FOR THE SECOND BEST.

Gallery View		Method	Probe View 55°	65°	75°	85°
	55°	GEI [10]		12.3	25.8	30.8
		TCM [21]		***3.2***	***4.0***	***5.7***
		QVTM [11]		3.6	4.8	6.5
		GVTM [ours]		**2.4**	**3.2**	**4.7**
	65°	GEI [10]	9.3		5.9	14.5
		TCM [21]	***3.0***		***3.4***	***4.2***
		QVTM [11]	3.5		3.4	5.1
		GVTM [ours]	**2.8**		**2.4**	**3.0**
	75°	GEI [10]	21.3	5.4		***3.1***
		TCM [21]	***4.0***	***3.4***		3.8
		QVTM [11]	4.7	3.7		3.8
		GVTM [ours]	**3.4**	**2.3**		**2.4**
	85°	GEI [10]	25.5	11.8	***3.0***	
		TCM [21]	***5.5***	***4.4***	3.7	
		QVTM [11]	6.5	4.9	3.7	
		GVTM [ours]	**4.3**	**3.1**	**2.3**	

Table II
COMPARISON OF THE RANK-1 IDENTIFICATION RATES [%] OF OUR PROPOSED GVTM APPROACH TO THOSE OF BENCHMARK GENERATIVE APPROACHES FOR ALL VIEW GALLERY/PROBE VIEW COMBINATIONS.

Gallery View		Method	Probe View 55°	65°	75°	85°
	55°	GEI [10]		19.7	4.2	2.3
		TCM [21]		***79.9***	***70.8***	***54.5***
		QVTM [11]		78.3	64.0	48.6
		GVTM [ours]		**87.0**	**77.7**	**62.3**
	65°	GEI [10]	21.4		46.9	16.7
		TCM [21]	***81.7***		***79.5***	***70.2***
		QVTM [11]	81.5		79.2	67.5
		GVTM [ours]	**87.5**		**87.7**	**83.0**
	75°	GEI [10]	3.3	49.7		77.1
		TCM [21]	***71.9***	***80.0***		***79.0***
		QVTM [11]	70.2	80.0		78.2
		GVTM [ours]	**78.8**	**87.5**		**88.1**
	85°	GEI [10]	1.8	15.0	79.2	
		TCM [21]	***53.7***	***73.0***	***79.4***	
		QVTM [11]	51.1	68.5	79.0	
		GVTM [ours]	**64.9**	**83.2**	**88.4**	

employed while choosing its dimension within the range [1, 20, 40, ..., 300] to maximize the accuracy of the morphed training set in both verification and identification scenarios. In addition, when multiple LDA dimensions achieve the best accuracy for the training set, we choose the largest LDA dimension for the testing set to preserve more information. For example, in the identification scenario, the dimension of LDA are set to 160, 180, and 240 for the 55°, 65° and 75° probes, respectively.

Table III
COMPARISON OF THE EER [%] AND RANK-1 IDENTIFICATION RATES [%] OF OUR PROPOSED GVTM APPROACH TO THOSE OF BENCHMARK DISCRIMINATIVE APPROACHES. GALLERY VIEW FIXED TO 85° AGAINST VARIABLE PROBE VIEW. BOLDFACE FONT INDICATES THE BEST PERFORMANCE AND BOLD ITALICS ARE USED FOR THE SECOND BEST.

Method	EER [%] 55°	65°	75°	Rank-1 [%] 55°	65°	75°
LDA [22]	8	5	4	56	91	96
GMLDA [23]	12	9	5	68	82	95
MvDA [24]	7	5	4	***88***	**96**	***97***
GEINet [14]	**3**	**2**	**1**	80	92	95
GVTM [ours]	***4***	***3***	***2***	**92**	**96**	**98**

Table III shows the evaluation results of the discriminative approaches. As with generative approaches, we report EER and rank-1 rates for verification and identification, respectively. In line with [24], we evaluate cross-view gait recognition for a fixed gallery view of 85° and different probe views of 55°, 65° and 75°.

We notice that our GVTM approach clearly outperforms the benchmarks with regards to identification. In particular, we significantly outperform the deep learning approach, GEINet. Our identification results also supersede the best performing benchmark, MvDA, specially when the difference of viewing angles is large. On the other hand, although the results of smaller view angle differences are only slightly better than the MvDA, we still perform quite well compared with other benchmarks. As for verification, the proposed approach performs the second best, right after GEINet, with a big gap to the other approaches. The inconsistency between the verification and identification scenarios is common in gait recognition, since the identification performance is evaluated by probe-dependent rank statistics, while the verification results are computed by combined score distributions [14].

IV. CONCLUSION

In this paper, we introduce the GVTM based on an FFD framework for cross-view gait recognition. More precisely, we extract GEIs as appearance-based features from input image sequences captured from two different views. For every pair of views, we train a warping field from the two views to an intermediate view using gait features of training subjects. The trained GVTM is thereafter applied to morph the gait features of a test subject from the two different views to the intermediary view for a better matching. The effectiveness of the proposed method is validated through experiments using the large population gait database of OU-ISIR, where we outperform the latest published approaches.

While we employ a common (i.e., subject-independent) GVTM so far, a future research avenue is to extend it to a subject-dependent one for better transformation accuracy as the previous appearance-based VTM do so. Moreover, we

plan to further validate the proposed method against a larger variability of views, because view variations in the data sets used are still limited.

ACKNOWLEDGEMENT

This work was supported by JSPS Grant-in-Aid for Scientific Research (A) JP15H01693.

REFERENCES

[1] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon, "On using gait in forensic biometrics," *Journal of Forensic Sciences*, vol. 56, no. 4, pp. 882–889, 2011. 1

[2] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, "Gait recognition using a view transformation model in the frequency domain," in *Computer Vision–ECCV 2006*. Springer, 2006, pp. 151–163. 1, 2

[3] K. Bashir, T. Xiang, and S. Gong, "Gait recognition without subject cooperation," *Pattern Recognition Letters*, vol. 31, no. 13, pp. 2052–2060, Oct. 2010. 1

[4] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Gait recognition across various walking speeds using higher order shape configuration based on a differential composition model," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 6, pp. 1654–1668, Dec. 2012. 1

[5] H.-D. Yang and S.-W. Lee, "Reconstruction of 3D human body pose for gait recognition," in *International Conference on Biometrics*. Springer, 2006, pp. 619–625. 1, 2

[6] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated face and gait recognition from multiple views," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–I. 1, 2

[7] A. Kale, A. R. Chowdhury, and R. Chellappa, "Towards a view invariant gait recognition algorithm," in *Advanced Video and Signal Based Surveillance, 2003. Proceedings. IEEE Conference on*. IEEE, 2003, pp. 143–150. 1, 2

[8] F. Jean, R. Bergevin, and A. B. Albu, "Computing and evaluating view-normalized body part trajectories," *Image Vision Comput.*, vol. 27, no. 9, pp. 1272–1284, Aug. 2009. 1, 2

[9] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Support vector regression for multi-view gait recognition based on local motion feature selection," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 974–981. 2

[10] J. Han and B. Bhanu, "Individual recognition using gait energy image," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 2, pp. 316–322, 2006. 2, 4, 5

[11] D. Muramatsu, Y. Makihara, and Y. Yagi, "View transformation model incorporating quality measures for cross-view gait recognition," *IEEE transactions on cybernetics*, vol. 46, no. 7, pp. 1602–1615, 2016. 2, 4, 5

[12] J. Lu and Y.-P. Tan, "Uncorrelated discriminant simplex analysis for view-invariant gait signal computing," *Pattern Recognition Letters*, vol. 31, no. 5, pp. 382–393, 2010. 2

[13] R. Martin-Felez and T. Xiang, "Uncooperative gait recognition by learning to rank," *Pattern Recognition*, vol. 47, no. 12, pp. 3793 – 3806, 2014. 2

[14] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "Geinet: View-invariant gait recognition using a convolutional neural network," in *Biometrics (ICB), 2016 International Conference on*. IEEE, 2016, pp. 1–8. 2, 4, 5

[15] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A comprehensive study on cross-view gait based human identification with deep cnns," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 2, pp. 209–226, 2017. 2

[16] 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," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 5, pp. 1511–1521, 2012. 2, 3

[17] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in nd images," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1. IEEE, 2001, pp. 105–112. 2

[18] T. W. Sederberg and S. R. Parry, "Free-form deformation of solid geometric models," *ACM SIGGRAPH computer graphics*, vol. 20, no. 4, pp. 151–160, 1986. 3

[19] A. Leow, S.-C. Huang, A. Geng, J. Becker, S. Davis, A. Toga, and P. Thompson, "Inverse consistent mapping in 3D deformable image registration: its construction and statistical properties," in *Biennial International Conference on Information Processing in Medical Imaging*. Springer, 2005, pp. 493–503. 3

[20] D. Muramatsu, Y. Makihara, and Y. Yagi, "Are intermediate views beneficial for gait recognition using a view transformation model?" in *Proc. of the 20th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV 2014)*, Nago, Japan, Feb. 2014, pp. 222–227. 3

[21] ——, "Cross-view gait recognition by fusion of multiple transformation consistency measures," *IET Biometrics*, vol. 4, no. 2, pp. 62–73, 2015. 4, 5

[22] N. Otsu, "Optimal linear and nonlinear solutions for least-square discriminant feature extraction," in *Proceedings of the 6th International Conference on Pattern Recognition, 1982*, 1982, pp. 557–560. 4, 5

[23] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2160–2167. 4, 5

[24] A. Mansur, Y. Makihara, D. Muramatsu, and Y. Yagi, "Cross-view gait recognition using view-dependent discriminative analysis," in *Biometrics (IJCB), 2014 IEEE International Joint Conference on*. IEEE, 2014, pp. 1–8. 4, 5