Multi-view Multi-modal Person Authentication from a Single Walking Image Sequence

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

Abstract

This paper describes a method for multi-view multi-modal biometrics from a single walking image sequence. As multi-modal cues, we adopt not only face and gait but also the actual height of a person, all of which are simultaneously captured by a single camera. As multi-view cues, we use the variation in the observation views included in a single image sequence captured by a camera with a relatively wide field of view. This enables us to improve the authentication of a person based on multiple modalities and views, while retaining the potential for real applications such as surveillance and forensics using only a single image sequence (a single session with a single camera). In the experiments, we constructed a large-scale multi-view multi-modal score data set with 1,912 subjects, and evaluated the proposed method using the data set in a statistically reliable way. We achieved 0.37% equal error rates for the false acceptance and rejection rates in the verification scenarios, and 99.15% rank-1 identification rate in the identification scenarios.

1. Introduction

Image-based pedestrian authentication has been in high demand for many applications including but not limited to access control, security, surveillance, and forensics, because surveillance cameras have been increasingly installed in both public and private spaces. Image-based biometric cues such as iris, ear, face, and gait, have potential for person authentication, of which face and gait are particularly promising modalities because they can be effective even if target images are captured at relatively large distances.

Face recognition has been extensively studied in recent times and we refer the reader to the excellent surveys of [39, 1, 17] for details. Although face recognition achieves high performance in a controlled situation, it encounters difficulties under the so-called PIE (pose, illumination, and expression) variations [11], and also in the case of significantly low image resolution, occlusion (e.g., mask, sunglass, and full-face helmet), and observation from the rear.

Gait recognition is a relatively new family of biometrics compared with face recognition and we refer the reader to [30] for details. An advantage of gait recognition is its ability to ascertain identity from a distance and hence it works well even if the image resolution is significantly low (e.g., just 30-pixels in height). The performance of gait recognition is, however, often degraded by many of the covariates: views [18, 25, 22], walking speeds [27], clothing [15].

In both face and gait recognition, while the pose or view difference between a probe and a gallery is a troublesome covariate, a common view variation in a probe and a gallery is a useful cue for improving person authentication because the different views contain different types of cues (e.g., forward-backward arm swing in the side-view gait and body width in the frontal-view gait). Therefore, several researchers focus their efforts on multi-view face recognition [8, 29, 20, 37, 28] or multi-view gait recognition [5, 35, 26, 21, 13]. In particular, approaches using multi-view observation from a single image sequence (a single session from a single camera) [13, 34] have good potential for application to real situations, because we cannot always expect view variations from multiple cameras and/or multiple sessions in surveillance and forensic scenarios.

Moreover, fusion of face and gait recognitions [32, 19, 23, 38, 42, 14] is also a promising approach for improving person authentication, because face and gait biometrics can be captured by a single device, namely, a camera, and can also be used as complementary cues to each other.

Therefore, we propose a method of person authentication based on multi-view multi-modal biometrics from a single walking image sequence. We extract multiple cues derived from multiple views as well as multiple modalities including not only face and gait but also the height of a person and we fuse them into a score level for better performance. The contributions of this paper are summarized by the following

three points.

1. Multi-view face and gait recognition from a single image sequence

We simultaneously combine both multiple modalities (face and gait) and multiple views observed from a single walking image sequence in score-level fusion frameworks. This enables us to improve the person authentication performance based on multiple modalities and views, while retaining the potential for real applications such as surveillance and forensics, because we use only a single image sequence (a single session with a single camera).

2. Fusion with the height

We employ the height of a person as another modality in addition to face and gait, while existing methods of gait recognition often overlook the useful height information. Note that we use the actual height, which is computed based on camera calibration and ground plane constraints and is therefore independent of person-to-camera distance.

3. Construction of a large-scale score data set

We have constructed a multi-view multi-modal (face and gait) and height score database using the data set [16] that contains image sequences of 1,912 walking subjects captured by a camera with a relatively wide field of view [1]. Since previous studies on the fusion of face and gait recognition employed at most the order of a hundred subjects [32, 19, 23, 38, 40, 41, 43, 42, 10, 9, 14], we significantly improve the statistical reliability of the performance evaluation with this data set.

2. Related Work

2.1. Multi-view face recognition

It is well known among those using face recognition that video-based recognition achieves a higher performance than single-image-based recognition because the variations in the video such as pose or expression enhance the individual characteristics. A typical method for video-based recognition is a subspace-based method. Fukui and Yamaguchi [8] proposed a constrained mutual subspace method (CMSM), and Nishiyama et al. [29] extended it into multiple CMSM. Kim et al. [20] formulated the multi-view face recognition in a discriminative canonical correlation framework.

Another way of using multiple views is by incorporating a 3D face model. Chen and Hauptmann [37] exploit two orthogonal views to reconstruct a 3D face model for robust face recognition. Mayo and Zhang [28] proposed an algorithm for 3D face recognition based on point cloud rotation, multiple projections, and voted keypoint matching.

2.2. Multi-view gait recognition

Multi-view observations are also effectively used in gait recognition. Cuntoor et al. [5] combine scores provided by width vectors from side and frontal views using a summation rule, while Yu et al. [35] fuse scores provided by a key Fourier descriptor from multiple views using several fusion rules.Multi-view observations are also used to enhance view-invariant gait recognition [33, 25, 22].

Whereas the above approaches assume multiple cameras and/or sessions, several approaches extract multi-view gait features from a single image sequence. Han et al. [13] find and match the optimal pair of subsequences with similar views between two image sequences containing view transitions in the context of the view-invariant gait recognition, while Sugiura et al. [34] extract multi-view gait features captured by a camera with a wide field of view (e.g., omnidirectional camera [36]) and fuse them at a score level.

2.3. Fusion of face and gait

Because we can simultaneously capture face and gait biometrics using a single camera, the fusion of face and gait is a promising fusion pair for real applications.

Kale et al. [19] fuse view-invariant gait recognition and face recognition based on sequential importance sampling with hierarchical (sequential) or score-level fusions. Zhang et al. [38] introduced geometry preserving projections approaches for selecting a subspace of multi-modal biometrics and conducted experiments using chimera data of faces, palm prints, and gaits. Lee et al. [23] brought out a chaotic measure for fusing frontal face and gait. Hofmann et al. [14] used alpha matting preprocessing to accurately separate foreground and background layers and also to obtain alpha GEI for better fusion.

Zhou et al. [40] fused side-view gaits and super-resolution face images using hierarchical or score-level fusions, and further introduced dimension reduction and discriminant analysis for better performance in [41]. Zhou et al. [43] also exploited fusion with face profiles and Zhou and Bhanu [42] proposed a feature-level fusion scheme of side view faces and gaits in conjunction with discriminant analysis.

Whereas the above existing methods make use of single-view face and gait features, Shakhnarovich and Darrell [32] constructed a texture-mapped visual hull using multiple cameras and synthesize a frontal-view face and a side-view gait. Liu and Sarkar focus on outdoor case, and fuse frontal-view face and side-view gait using several fusion scheme [24]. The multi-camera setting is, however, not always available in real scenarios such as surveillance and forensics.

Geng et al. [10] proposed an adaptive fusion scheme of the face and gait to cope with the view and subject-to-camera distance changes and also introduced four different types of fusion weights and determined the weights by prior knowledge or machine learning techniques in [9]. Whereas the change of subject-to-camera distances within a single

[1] http://www.am.sanken.osaka-u.ac.jp/BiometricDB/BioScore.html

image sequence is considered, the change of views are not considered in [10, 9], and hence multi-view fusion is out of scope.

In summary, to the best of our knowledge, up to now there have been no studies on the fusion of multi-view face and gait from a single image sequence.

3. Multi-view multi-modal walking person authentication

3.1. Overview

We consider authenticating a walking person using a single image sequence captured by a single camera at a single session. In this case, the face and gait biometrics are candidates for person authentication, because both face and gait are observed from the camera. In this setting, the image sequences are composed of several image subsequences observed with different views, even if the walking direction is unchanged as long as the camera has a relatively wide field of view [34], and hence we can extract multi-view face and gait features. Moreover, we can extract the height of the subject from the image sequence if the camera used is calibrated in advance and the ground plane constraint is available.

Therefore, our approach employs multi-view multi-modal biometrics, namely, multi-view face features, multi-view gait features, and the height of a subject together as multiple cues for person authentication.

3.2. Multi-view multi-modal fusion

Many fusion approaches with different levels (e.g. sensor-level, feature-level, score-level, and decision level) have been proposed [31]. We focus on score-level fusion in this study.

We extract multi-view face and gait features from the image sequence of the target subject, because the face and gait features from different views include different information (e.g., face texture in the frontal-view and face profile in the side view, body width in the frontal-view and forward-backward arm swing in the side view). In contrast, we extract a single height from the image sequence, because the height of the body is view-independent. Based on the extracted features, we calculate face and gait dissimilarity scores with multiple views, and a height dissimilarity score between two image sequences.

In the case of multi-view multi-modal biometrics, three fusion approaches are possible: a) multi-modal fusion in single view; b) multi-view fusion of unimodal; c) multi-view fusion of multi-modal.

Let $\mathbf{Sf} = [sf_1, sf_2, ..., sf_{N_v}]$, $\mathbf{Sg} = [sg_1, sg_2, ..., sg_{N_v}]$ be N_v dimensional score vectors associated with the face and gait from N_v different views, and let Sh be a score of the height. Our method calculates a fusion score S_{fusion} for authentication by:

$$S_{fusion} = f(\mathbf{Sf}, \mathbf{Sg}, Sh), \quad (1)$$

where $f(\cdot)$ is a fusion function (rule) that combines the input scores. We perform verification or identification using the fused score S_{fusion}.

4. Implementation

4.1. Face-based matching

In this study, we call the part of a body above the neck a *face*, which is defined based on a silhouette. A face feature is then defined as a color texture masked by the face region as shown in Figure 1. The sizes of faces are dependent on the subject and view, and range from 17×19 [pixels] to 21 × 24 [pixel] in the examples of Figure 1.

Given a pair of face features, we calculate a dissimilarity score between them as follows. Let Fp_i^v be a face feature of a probe at the i-th frame from view v, and let $Fg_{j,k}^v$ be a face feature of a gallery at the j-th frame from view v with k-th spatial displacement within the searching regions for template matching. The dissimilarity score between the probe and the gallery from view v is calculated by correlation-based template matching as:

$$Sf_v(probe, gallery) = \min_{i,j,k}[1 - NCC(Fp_i^v, Fg_{j,k}^v)], \quad (2)$$

where $NCC(Fp_i^v, Fg_{j,k}^v)$ is a normalized cross correlation between Fp_i^v and $Fg_{j,k}^v$.

4.2. Gait-based matching

We use GEI [12] as a gait feature in this study. First, image normalization techniques [16] are applied to correct for lens distortion and camera rotation. A silhouette sequence is then extracted from the normalized image sequence using graph-cut-based segmentation [4] and background subtraction. Next, the silhouette sequence is normalized into a 88 × 128 pixel-sized silhouette sequence and finally the GEIs are computed from it. Examples of the GEIs with multiple views from a single image sequence are shown in Figure 2. We can see that GEIs with multiple views contain different types of information (e.g., body width at 55 [deg] and forward-backward arm swing at 85 [deg]).

Given a pair of GEIs, we calculate a dissimilarity score between them as follows. Let Gp^v and Gg^v be the GEI of a probe and a gallery associated with view v. The dissimilarity score between the probe and gallery from view v is calculated by Euclidean distance as:

$$Sg_v(probe, gallery) = ||Gp^v - Gg^v||_2. \quad (3)$$

4.3. Height-based matching

We assume that the camera is calibrated and hence the ground plane constraint on the position of the bottom point

of a foot is available. We first extract the position of the bottom point of the foot $(X_i^f, Y_i^f, 0)$ at the i-th frame in a world coordinate system, and then calculate the position of the head top point (X_i^h, Y_i^h, Z_i^h) at the i-th frame in a world coordinate system based on the position of the head point in the image plane and the assumption that a subject stands perpendicular to the ground plane, namely, $X_i^h = X_i^f$ and $Y_i^h = Y_i^f$. Finally, we calculate the height of the subject by averaging over the image sequence as:

$$h = \frac{1}{N_f} \sum_{i=1}^{N_f} Z_i^h, \tag{4}$$

where N_f is the number of frames in the image sequence.

We then calculate a dissimilarity score from the absolute difference as:

$$Sh = |hp - hg|, \tag{5}$$

where hp and hg are the heights of subjects in the probe and gallery.

4.4. Fusion rule

To eliminate the subject dependency, we normalize the scores before fusion. Let $Sx_v(i,j)$ be the scores of features $x \in \{\text{gait } (g), \text{face } (f), \text{height } (h)\}$ between the i-th probe and the j-th gallery. In this study, we adopt a target-score normalization technique [7] and calculate a normalized score by:

$$\bar{S}x_v(i,j) = \frac{Sx_v(i,j) - \mu x_v(i)}{\sigma x_v(i)}, \tag{6}$$

where $\mu x_v(i)$ and $\sigma x_v(i)$ are the mean and standard deviation associated with feature x from view v of the i-th probe, which are computed based on the dissimilarity scores between the i-th probe and all the subjects in the training data.

In this study, we consider five fusion rules (1) sum rule (denoted as Sum) (2) SVM with a linear kernel[2], (3) linear logistic regression (LLR) [2], (4) a method using kernel density estimation (KDE) [6], and a minimum rule (denoted as Min).

5. Experiment

5.1. Database

We used the subset of OU-ISIR database [16] for performance evaluation. In this data set, individual subjects walked along a course, and data were captured using a single camera placed at a 5-meter distance from the course, as shown in Figure 3. The data were collected from 1,912 subjects, and two image sequences for each subject were available.

To obtain multi-view cues, we considered four views associated with four observation azimuth angles 55, 65, 75, and 85 [deg]. From each image sequence, we determined specific frames whose observation azimuth angles of the subjects coincided with the four angles, and extracted image subsequences associated with one gait cycle around the specific frames as the data associated with the four angles. Consequently, we obtained four image subsequences with different views from a single image sequence and hence the number of views is $N_v = 4$.

[2]Hyper-parameters are automatically selected through cross-validation

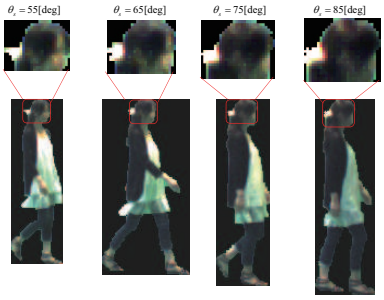
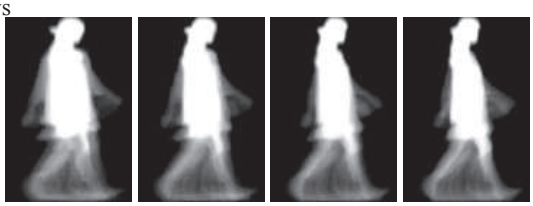


Figure 1. Captured images and zoomed face images of different views



(a) 55 [deg] (b) 65 [deg] (c) 75 [deg] (d) 85 [deg]

Figure 2. Gait features (GEI) of different views

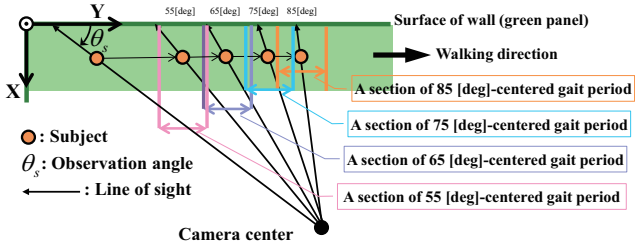


Figure 3. Target situation

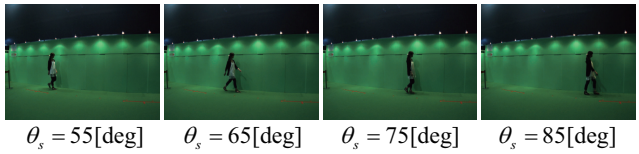


Figure 4. Walking images of different views

5.2. Experimental setting

To generate a fusion model, training data are necessary. In this study, we randomly divided the data into two groups, and performed 2-fold cross-validation. Because the accuracies may be influenced by the random grouping, we repeated the two-fold cross-validation five times so that we

could reduce the impact of the random grouping.

5.3. Experimental results

We evaluated the accuracies in terms of two different scenarios: verification and identification. For the verification scenario, we calculated the false acceptance rates (FARs) and false rejection rates (FRRs), and drew detection error tradeoff (DET) curves to see the tradeoff between FAR and FRR. For the identification scenario, we calculated the cumulative matching rates against ranks, and drew cumulative matching characteristic (CMC) curves. For comparison purposes, we also evaluated the multi-view fusion of face and gait, bi-modal fusion (face and gait) and the multi-modal fusion (face, gait, and height) of each view in this experiment.

Verification scenario

Figures 5, 6, and 7 show the DET curves of bi-modal, multi-modal, multi-view, and multi-view multi-modal fusion. We also summarized the associated equal error rates (EERs) in tables 1 and 3.

From these results, we could observe that the proposed multi-view multi-modal fusions outperform bi-modal and multi-modals for any view, multi-view gait, and multi-view face. The minimum EER (the best result) 0.37% was achieved by the proposed multi-view multi-modal fusion with Sum. From the DET curves, we could observe that the proposed fusions outperformed the comparison fusion schemes for most operating points.

Identification scenario

Figures 8, 9, and 10 shows the CMC curves of bi-modal, multi-modal, multi-view, and multi-view multi-modal fusion. We also summarized the associated rank-1 identification rate in Tables 2 and 4. The best rank-1 identification rate 99.15% was achieved by the proposed fusion with SVM, and the proposed fusion outperformed the comparison fusion schemes.

6. Discussion

Verification scenario

We now review the effects of multi-view and multi-modal fusion on accuracy. As shown in Tables 1 and 3, the accuracies are improved using multi-view and/or multi-modal fusions.

For multi-view fusion, by comparing the best unimodal results (EER=3.86% for face, EER=$2.17\%$for gait), EER was reduced by relatively 59.8% for face, and 50.2% for gait by multi-view fusion. For multi-modal fusion, EER was largely reduced by relatively 72.8% and 84.7%, compared with the best unimodal result. By comparaing the best results of multi-modal (EER=0.59%) and bi-modal (EER=0.81%), EER of multi-modal fusion was reduced by relatively 27.2%. This result shows that the usefulness of the height feature.

Table 3. EERs of multi-view fusion and multi-view multi-modal fusion

	Multi-view fusion		Multi-view
	Face	Gait	multi-modal fusion
Sum	2.12	1.43	0.37
SVM	1.55	1.32	0.41
LLR	2.03	1.32	0.44
KDE	2.23	1.51	0.40
Min	2.13	1.08	0.85

Table 4. Rank-1 identification rates of multi-view fusion and multi-view multi-modal fusion

	Multi-view fusion		Multi-view
	Face	Gait	multi-modal fusion
Sum	96.40	94.87	98.96
SVM	95.33	93.90	99.15
LLR	96.24	94.63	99.09
KDE	90.21	83.34	95.78
Min	94.30	95.43	96.19

By comparing the effects of multi-view and multi-modal, the improvements using multi-modal are greater than those using multi-views. The result is supported by the fact that the strength of the correlation between views within the same modality is stronger than that among modalities within the same view as shown in the scatter plots of the scores (see Figure 11 (a), (b), and (c)). This result is interesting because the observation is slightly different from the research associated with the face and iris case [3], which reports that multi-sample fusion outperforms multi-modal fusion.

Moreover, because the multi-view fusion within the same modality and multi-modal fusion within the same view improve the accuracies at one level or another, we can further improve the accuracies by considering multi-view and multi-modal fusion simultaneously. As a result, the proposed multi-view multi-modal fusion significantly reduces EER by relatively 65.7%, compared with those for the best multi-view fusion within the same modality, and by relatively 37.3%, compared with the best multi-modal fusion within the same view. This is also supported by the fact that the score distribution in the score spaces for multi-view face, multi-view gait, and height show better separability of positive and negative samples in Figure 11 (d) than those in 11 (a), (b) and (c). This is why the multi-view multi-modal fusion achieves better accuracies than individual multi-view and multi-modal fusions.

Consequently, the proposed approach achieves an EER of 0.37% against data composed of 1,912 subjects. These results show that the multi-view multi-modal fusion is a useful approach to person verification tasks.

Identification scenario

Table 1. EERs of Unimodal, bi-modal, and multi-modal fusion in single view

	Uni-modal			Bi-modal (face and gait) fusion					Multi-modal (face, gait, and height) fusion				
view	Face	Gait	Height	Sum	SVM	LLR	KDE	Min	Sum	SVM	LLR	KDE	Min
55	4.23	2.47		1.18	1.72	1.23	1.29	1.81	0.69	0.90	0.76	1.09	1.77
65	3.86	2.27	15.48	0.89	0.97	0.90	1.02	1.66	0.59	0.65	0.61	0.91	1.58
75	4.37	2.17		1.09	1.33	1.09	1.10	1.51	0.84	0.90	0.86	0.91	1.52
85	4.70	2.27		0.86	1.44	0.81	1.06	1.35	0.67	0.74	0.69	0.83	1.46

Table 2. Rank-1 identification rate of unimodal, bi-modal, and multi-modal fusion in single view

	Uni-modal			Bi-modal (face and gait) fusion					Multi-modal (face, gait, and height) fusion				
view	Face	Gait	Height	Sum	SVM	LLR	KDE	Min	Sum	SVM	LLR	KDE	Min
55	88.40	88.89		96.32	96.36	96.19	90.40	90.70	97.10	97.24	97.08	91.87	90.61
65	90.37	90.65	1.92	96.82	96.99	97.17	91.42	91.90	97.76	97.80	97.81	93.21	91.67
75	89.30	90.62		96.78	97.01	96.82	91.43	92.01	97.64	97.54	97.63	93.06	91.83
85	90.39	90.09		97.13	97.17	97.29	91.40	92.26	97.87	97.67	97.92	92.81	92.24

Tables 2 and 4 show the rank-1 identification rates of single-view unimodals, single-view bi-modal, single-view multi-modal, multi-view unimodal, and multi-view multi-modal fusion. Compared with the best single-view unimodal results, rank-1 identification rates for multi-view fusions of individual face and gait are improved to 96.40% (+6.01%) and 95.43% (+4.78%). Comparing the best single-view unimodal results in individual views, rank-1 identification rates for single-view multi-modal fusion are improved to 97.24% (+8.35%), 97.81% (+7.16%), 97.64% (+7.02%), and 97.92% (+7.53%) for 55, 65, 75, and 85-deg views, respectively. And in almost all cases, multi-modal fusion outperforms bi-modal fusion (excluding Min rule). We further improve the rank-1 identification rates to 99.15% by considering multi-modal and multi-view fusion simultaneously.

From these results, we conclude that the multi-view multi-modal fusion is also a useful approach for the person identification task.

7. Conclusions

This paper describes a method employing multi-view multi-modal biometrics from a single walking image sequence. While we adopted not only face and gait but also the actual height of a person which were all simultaneously captured by a single camera as multi-modal cues, we used the variation in observation views included in a single image sequence as multi-view cues. In our experiments, we constructed a multi-view multi-modal (face, gait, and height) score data set consisting of 1,912 subjects. We then evaluated the proposed method with the data set and achieved 0.37% equal error rate and 99.15% rank-1 identification rate.

The data in OU-ISIR used for evaluation was captured in similar conditions (e.g., the same days, the same attire, and natural facial expressions), the obtained results can be thought the upper-bound accuracy of the proposed method. We therefore plan to evaluate the proposed method against more realistic databases in future.

Future work will include the incorporation of spatial and temporal resolution of the image sequence as quality measures, and the treatment of missing data (e.g., face modality when observed from the rear and gait modality when a person stands still).

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] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2d and 3d face recognition: A survey. *Pattern Recogn. Lett.*, 28(14):1885–1906, Oct. 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. of SPIE 6994, Biometric Technologies for Human Identification IV*, Orlando, FL, USA, Mar. 2008.

[3] C. Boehnen, D. Barstow, D. Patlolla, and C. Mann. A multi-sample standoff multimodal biometric system. In *Proc. of the 5th IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, number Paper ID 39.

[4] Y. Boykov and M. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images.

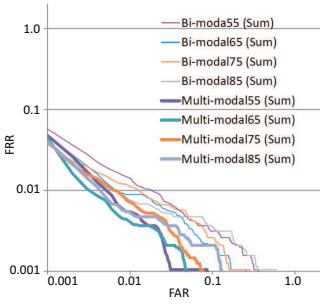


Figure 5. DET curves of bi-modal and multi-modal fusion in single view

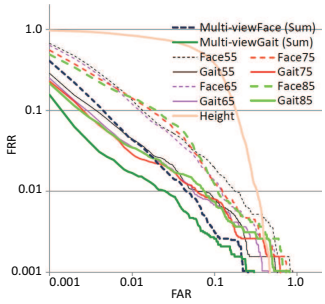


Figure 6. DET curves of each modality and multi-view fusion

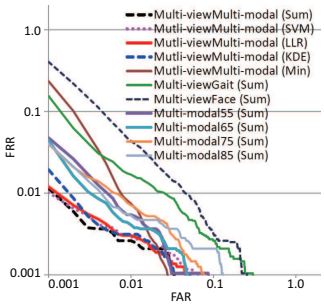


Figure 7. DET curves of multi-view multi-modal fusion

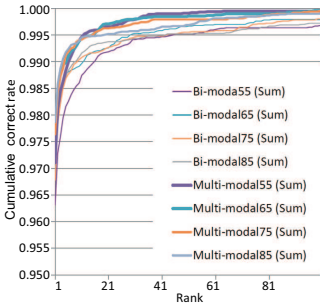


Figure 8. CMC curves of bi-modal and multi-modal fusion in single view

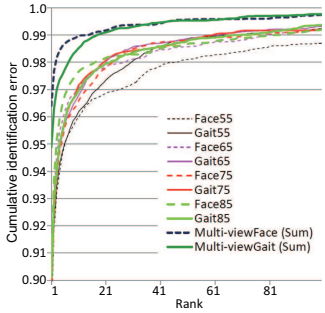


Figure 9. CMC curves of each modality and multi-view fusion

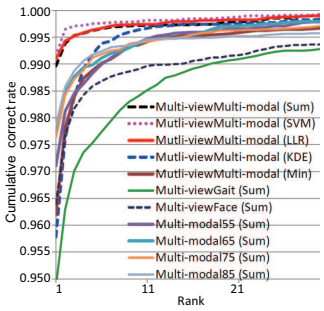
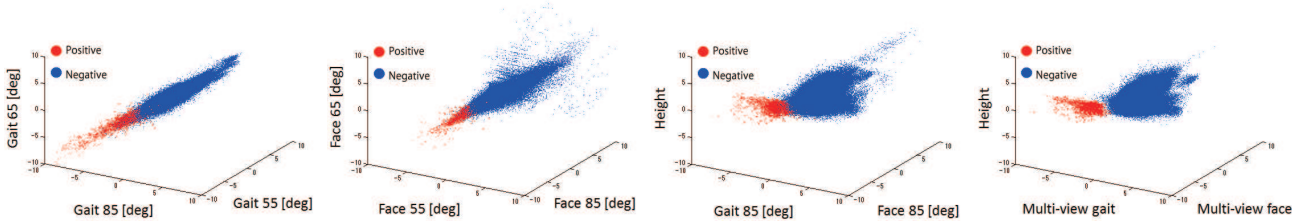


Figure 10. CMC curves of multi-view multi-modal fusion



(a) (b) (c) (b)

Figure 11. Scatter-plots: (a) gait scores of multi-views, (b) face scores of multi-view, (c) gait, face, and height scores, (d) multi-modal scores of multi-view.

In *Proc. of Int. Conf. on Computer Vision*, volume 1, pages 105–112, 2001.

[5] N. Cuntoor, A. Kale, and R. Chellappa. Combining multiple evidences for gait recognition. In *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, volume 3, pages 33–36, 2003.

[6] S. C. Dass, K. Nandakumar, and A. K. Jain. A principled approach to score level fusion in multimodal biometric systems. In *Proc. of the 5th Int. Conf. on Audio- and Video-Based Biometric Person Authentication*, pages 1049–1058, Ny, USA, July 2005.

[7] J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Target dependent score normalization techniques and their application to signature verification. *IEEE Trans. Systems, Man, and Cybernetics-part C: Applications and Reviews*, 35(3):418–425, 2005.

[8] K. Fukui and O. Yamaguchi. Face recognition using multi-viewpoint patterns for robot vision. In *11th Int. Symposium of Robotics Research*, pages 192–201, 2003.

[9] X. Geng, K. Smith-Miles, L. Wang, M. Li, and Q. Wu. Context-aware fusion: A case study on fusion of gait and face for human identification in video. *Pattern Recogn.*, 43(10):3660–3673, Oct. 2010.

[10] X. Geng, L. Wang, M. Li, Q. Wu, and K. Smith-Miles. Adaptive fusion of gait and face for human identification in video. In *Applications of Computer Vision, 2008. WACV 2008. IEEE Workshop on*, pages 1 –6, jan. 2008.

[11] R. Gross, I. Matthews, J. F. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image Vision Comput.*, 28(5):807–813, 2010.

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

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

[14] M. Hofmann, S. M. Schmidt, A. Rajagopalan, and G. Rigoll. Combined face and gait recognition using alpha matte preprocessing. In *Proc. of the 5th IAPR Int. Conf. on Biometrics*, pages 1–8, New Delhi, India, Mar. 2012.

[15] M. A. Hossain, Y. Makihara, J. Wang, and Y. Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, 43(6):2281–2291, Jun. 2010.

[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. *IEEE Transactions on Information Forensics and Security*, 7(5):1511–1521, oct. 2012.

[17] R. Jafri and H. R. Arabni. A survey of face recognition techniques. *Journal of Information Processing*, 5(2):41–68, June 2009.

[18] 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.

[19] A. Kale, A. Roy-Chowdhury, and R. Chellappa. Fusion of gait and face for human identification. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing 2004 (ICASSP'04)*, volume 5, pages 901–904, 2004.

[20] T.-K. Kim, J. Kittler, and R. Cipolla. Discriminative learning and recognition of image set classes using canonical correlations. *IEEE Trans. on Patter Analysis and Machine Intelligence*, 29(6):1005–1018, June 2007.

[21] 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. of of IEEE computer society conferene on Computer Vision and Pattern Recognition 2010*, pages 1–8, San Francisco, CA, USA, Jun. 2010.

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

[23] T. Lee, S. Ranganath, and S. Sanei. Fusion of chaotic measure into a new hybrid face-gait system for human recognition. In *Proc. of the 18th Int. Conf. on Pattern Recognition*, volume 4, pages 541–544, Hong Kong, China, Aug. 2006.

[24] Z. Liu and S. Sarkar. Outdoor recognition at a distance by fusing gait and face. *Image and Vision Computing*, 25:817–832, 2007.

[25] 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.

[26] 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.

[27] Y. Makihara, A. Tsuji, and Y. Yagi. Silhouette transformation based on walking speed for gait identification. In *Proc. of the 23rd IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, Jun 2010.

[28] M. Mayo and E. Zhang. 3d face recognition using multiview keypoint matching. In *Advanced Video and Signal Based Surveillance, 2009. AVSS '09. Sixth IEEE International Conference on*, pages 290 –295, sept. 2009.

[29] M. Nishiyama, O. Yamaguchi, and K. Fukui. Face recognitionwith the multiple constrained mutual subspace method. In *AVBPA05*, pages 71–80, 2005.

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

[31] A. A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics*. Springer Science+Business Media, LLC., 2006.

[32] G. Shakhnarovich and T. Darrell. On probabilistic combination of face and gait cues for identification. In *Proc. Automatic Face and Gesture Recognition 2002*, volume 5, pages 169–174, 2002.

[33] 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.

[34] K. Sugiura, Y. Makihara, and Y. Yagi. Gait identification based on multi-view observations using omnidirectional camera. In *Proc. 8th Asian Conf. on Computer Vision*, pages 452–461, Tokyo, Japan, Nov. 18-22 2007. LNCS 4843.

[35] Y. Wang, S. Yu, Y. Wang, and T. Tan. Gait recognition based on fusion of multi-view gait sequences. In *Proc. of the IAPR Int. Conf. on Biometrics 2006*, pages 605–611, Jan. 2006.

[36] K. Yamazawa, Y. Yagi, and M. Yachida. Hyperomni vision: Visual navigation with an omnidirectional image sensor. *Systems and Computers in Japan*, 28(4):36–47, 1997.

[37] M. yu Chen and A. Hauptmann. Towards robust face recognition from multiple views. In *2004 IEEE Int. Conf. on Multimedia and Expo (ICME 2004)*, volume 2, pages 1191–1194, June 2004.

[38] T. Zhang, X. Li, D. Tao, and J. Yang. Multimodal biometrics using geometry preserving projections. *Pattern Recognition*, 41(3):805–813, 2008.

[39] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, Dec. 2003.

[40] X. Zhou and B. Bhanu. Feature fusion of face and gait for human recognition at a distance in video. In *Proc. of the 18th Int. Conf. on Pattern Recognition*, volume 4, pages 529–532, Hong Kong, China, Aug. 2006.

[41] X. Zhou and B. Bhanu. Integrating face and gait for human recognition at a distance in video. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 37(5):1119 –1137, oct. 2007.

[42] X. Zhou and B. Bhanu. Feature fusion of side face and gait for video-based human identification. *Pattern Recognition*, 41(3):778–795, 2008.

[43] X. Zhou, B. Bhanu, and J. Han. Human recognition at a distance in video by integrating face profile and gait. In *Proceedings of the 5th international conference on Audio- and Video-Based Biometric Person Authentication*, AVBPA'05, pages 533–543, Berlin, Heidelberg, 2005. Springer-Verlag.