IPSJ Transactions on Computer Vision and Applications Vol.6 53–57 (July 2014)

[DOI: 10.2197/ipsjtcva.6.53]

Express Paper

Quality-dependent Score-level Fusion of Face, Gait, and the Height Biometrics

Takuhiro Kimura[1,a) Yasushi Makihara[1,b) Daigo Muramatsu[1,c)
Yasushi Yagi[1,d)

Received: March 14, 2014, Accepted: April 24, 2014, Released: July 25, 2014

Abstract: This paper describes a quality-dependent score-level fusion framework of face, gait, and the height biometrics from a single walking image sequence. Individual person authentication accuracies by face, gait, and the height biometrics, are in general degraded when spatial resolution (image size) and temporal resolution (frame-rate) of the input image sequence decrease and the degree of such accuracy degradation differs among the individual modalities. We therefore set the optimal weights of the individual modalities based on linear logistic regression framework depending on a pair of the spatial and temporal resolutions, which are called qualities in this paper. On the other hand, it is not a realistic solution to compute and store the optimal weights for all the possible qualities in advance, and also the optimal weights change across the qualities in a nonlinear way. We thus propose a method to estimate the optimal weights for arbitrary qualities from a limited training pairs of the optimal weights and the qualities, based on Gaussian process regression with a nonlinear kernel function. Experiments using a publicly available large population gait database with 1,935 subjects under various qualities, showed that the person authentication accuracy improved by successfully estimating the weights depending on the qualities.

Keywords: multi-modal person authentication, face, gait, height, spatial and temporal resolution, quality-dependent

1. Introduction

Forensic technologies are essential for recent criminal investigation and also realization of a safe and secure society, since they contribute to prompt solution of criminal cases as well as to act as deterrents to crimes. Biometrics [10] such as DNA, fingerprint, voice, and signature, are now regarded as powerful tools to verify a person (e.g., a perpetrator and a suspect) in the forensics community and have been exploited for criminal investigation as well as expert evidences in the court. In particular, due to the exponential increase of CCTVs in the public space, the CCTV footage-based forensics are indispensable for modern criminal investigation.

The CCTV footage provides a variety of biometric traits including but not limited to the face [19], the gait [15], and soft biometrics [9] (e.g., the height, shoulder width, arm length, hair color and type, and tattoo). While these individual biometric traits have their own advantages (e.g., identification at a distance from a camera for gait), their accuracies are in general inferior to those by hard biometrics such as DNA and fingerprint.

One of reasonable solutions to enhance the accuracy and reliability for forensic use, is multi-modal biometrics [16]. In fact, fusion of face and gait have been extensively studied in biometrics field [4], [6], [11], [20], [21] because both the face and gait are often simultaneously captured by a single CCTV. Moreover, the height information is also available as an additional soft biometric modality when a camera calibration is done in advance [13].

Whatever the modality combination is, the most prevailing way to fuse the multiple biometric traits is the score-level fusion, and a key to success in the score-level fusion framework is to assign the optimal weights to the individual scores considering accuracies of the individual modalities. Furthermore, the optimal weights may depend on a kind of data quality, i.e., *quality measure* [2], which are utilized to improve the accuracy of multi-modal biometrics [2], [14], [18].

Considering a case of the CCTV footage-based forensics (e.g., fusion of the face, the gait, and the height biometrics [13]), two of the most significant quality measures are (1) the spatial resolution (or image size) and (2) the temporal resolution (or frame-rate). Since the spatial resolution (SR) of a target person highly depends on an observation distance from the camera as well as the CCTV footage size itself, it may considerably differ among situations. In addition, the SR obviously makes a large impact on the accuracies of the face trait as well as the height trait.

Moreover, the temporal resolution (TR) of the CCTV footage may largely differ among situations (e.g., a low frame-rate due to limitation of the communication band width and storage size, while a high frame-rate in places on high alert). Although the TR may make a little impact on face and height traits since they are static features in essence, it severely affects the gait trait since it exploits the temporal aspect.

We therefore propose a framework of quality-dependent score-level fusion of the face, the gait, and the height biometrics, where

[1] The Institute of Scientific and Industrial Research, Osaka University, Ibaraki, Osaka 567–0047, Japan
a) kimura@am.sanken.osaka-u.ac.jp
b) makihara@am.sanken.osaka-u.ac.jp
c) muramatsu@am.sanken.osaka-u.ac.jp
d) yagi@am.sanken.osaka-u.ac.jp

the weights for face, gait, and the height traits are appropriately set by considering the SRs and TRs as a quality. In this context, contributions of this paper are summarized as the follows two points.

Quality-dependent score-level fusion of face, gait, and the height biometrics: While the previous study [13] sets fixed weights to individual biometric traits, we set flexible weights so as to improve the accuracy as much as possible under a given quality. More specifically, we employ a linear logistic regression (LLR) framework [1], since the LLR provides a probabilistic value rather than just a fused score, which is essential for forensic use [17].

The optimal weights estimation from finite training sets: Although we need to cope with a test subject observed with an arbitrary quality, it is an unrealistic solution to compute and store the optimal weights for all the possible qualities in advance. We therefore estimate the optimal weights for such an arbitrary quality from a finite training pairs of the weights and the qualities. In addition, because the optimal weights may change in a non-linear way depending on the qualities, we introduce a Gaussian process regression (GPR) with a non-linear kernel function as a weight estimator.

2. Matching of Individual Modalities

In this section, we explain how to calculate the dissimilarity scores of individual face, gait, and height biometrics from an original image and silhouette as shown in **Fig. 1** (see [13] for more details).

In this paper, we refer *face* to a peripheral region of face including hair and face contour parts in addition to face region itself, namely, a head region (see **Fig. 2**). We calculate a face dissimilarity score s_{face} as a cosine distance between a gallery template face image and a probe face image after face alignment by template matching.

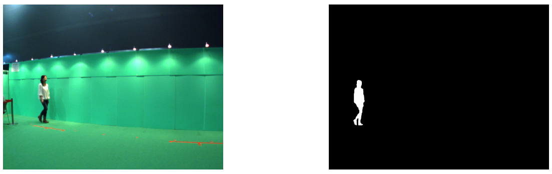

(a) Original image (b) Silhouette

Fig. 1 An original walking image and an extracted silhouette. The average size of the silhouette region over subjects is approximately 90 × 180 pixels.

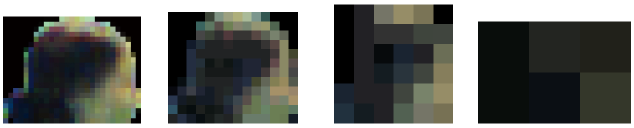

(a) 640 × 480 (b) 320 × 240 (c) 160 × 120 (d) 80 × 60

Fig. 2 Examples of face templates for various sizes [pixels] of original images. As a reference, the sizes [pixels] of face regions of this specific subject is 27 × 21, 14 × 12, 6 × 6, and 3 × 2 for (a)–(d), respectively.

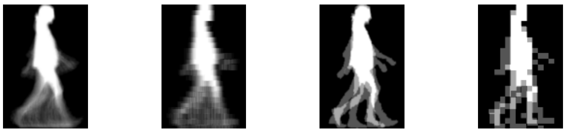

(a) 30, 640×480 (b) 30, 80×60 (c) 3, 640×480 (d) 3, 80×60

Fig. 3 Examples of GEIs (denoted as a pair of frame-rate [fps], the original image size [pixels]).

As for the gait, we first construct a size-normalized silhouette sequence with 88×128 pixel-size and then average them over one gait period to compute gait energy image (GEI) [5] (see **Fig. 3**). We calculate a gait dissimilarity score s_{gait} as the Euclidean distance between probe and gallery GEIs.

We compute the apparent height of the subject based on the bounding box of the silhouette. Since we assume that a camera is calibrated and that the ground plane constraint on the bottom of the foot is available, we convert the height in the image coordinate into that in the world coordinate, namely, the actual height. We then compute the height dissimilarity score s_{height} as an absolute difference between the gallery and probe heights.

3. Proposed Method

3.1 Quality-dependent Score-level Fusion Framework

We first normalize the dissimilarity scores to eliminate the subject dependency before fusion. Let $s_m(i, j)$ be a dissimilarity score of modality $m \in \{face, gait, height\}$ between the i-th probe and the j-th gallery, and let $\bar{s}_m(i, j)$ be the associated normalized score. We compute the normalized score $\bar{s}_m(i, j)$ as

$$\bar{s}_m(i, j) = \frac{s_m(i, j) - \mu_m(i)}{\sigma_m(i)}, \tag{1}$$

where $\mu_m(i)$ and $\sigma_m(i)$ is the mean and the standard deviation associated with the i-th probe and the modality m, respectively, and they are calculated using the i-th probe and available independent training data.

We then calculate a posterior probability of an event that the two gait image sequences originate from the same subject (the event is denoted as $X = 1$, while the complementary event, i.e., originating from different subjects, is denoted as $X = 0$) conditional on a set of the normalized dissimilarity scores $\bar{\boldsymbol{s}} = [\bar{s}_{face}, \bar{s}_{gait}, \bar{s}_{height}]^T$. Note that this kind of probabilistic score is meaningful for the purpose of forensics and expert evidence in the court, unlike a non-probabilistic score such as a sum of scores [12].

Moreover, because the SR q_S and TR q_T impact the scores and their discrimination capabilities, the quality $\boldsymbol{q} = [q_S, q_T]^T$ should be taken into account for the posterior calculation. We therefore consider the quality-dependent posterior probability $P(X = 1|\bar{\boldsymbol{s}}; \boldsymbol{q})$ and compute it by the LLR framework [1], where a logit function of the posterior is approximated by a weighted sum of the dissimilarity scores as

$$\log\left(\frac{P(X = 1|\bar{\boldsymbol{s}}; \boldsymbol{q})}{1 - P(X = 1|\bar{\boldsymbol{s}}; \boldsymbol{q})}\right) = \sum_{\substack{m\in\{face,\\ gait,height\}}} \alpha_m(\boldsymbol{q})\bar{s}_m + \alpha_c(\boldsymbol{q}), \tag{2}$$

where $\alpha_m(\boldsymbol{q})$ is the weight for the modality m, and $\alpha_c(\boldsymbol{q})$ is a constant. These LLR weights α_{face}, α_{gait}, α_{height}, and α_c are optimized using a training set of the dissimilarity scores $\boldsymbol{s}$ and the labels $X \in \{0, 1\}$, depending on the quality $\boldsymbol{q}$ of the given gait image sequences.

3.2 Weight Estimation

As also described in Section 1, it is an unrealistic solution to compute and store the optimal weights for all the possible qualities in advance. We therefore estimate the optimal weight α_* with a quality $\boldsymbol{q}_*$ given in a test case, and also with a finite training set $D = [Q, \boldsymbol{\alpha}]^T$, which are N pairs of the qualities $Q = \{\boldsymbol{q}_i\} (i = 1, \ldots, N)$ and the weights $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]^T$ which is a set using a training set of the dissimilarity scores associated with the training quality Q. Since we independently estimate the weights for individual modalities and constant term, we omit the subscripts for the modalities and constant instead of adding a training sample index i as a subscript for the weight α_i.

In addition, since it is known that the accuracies of individual modalities vary in a non-linear way (e.g., the accuracy of face biometrics rapidly drops as the SR is less than a certain value, while that of gait biometrics does so as the TR is less than a certain value), the weights should be represented as a non-linear function of the qualities. We therefore introduce GPR with a non-linear kernel function as a weight estimator. More specifically, we estimate the weight α_* as a posterior distribution given a quality $\boldsymbol{q}_*$ and the training set D, i.e., $P(\alpha_* | \boldsymbol{q}_*, D)$.

Assuming that a regression parameter is drawn from a Gaussian distribution and that each training weight α_i is drawn from each Gaussian distribution $\mathcal{N}(\alpha_i; \mu_i, \sigma_{o,i}^2)$, where μ_i and $\sigma_{o,i}^2$ are an expectation and a variance (i.e., observation noise level) of the weight α_i, it is well known that the posterior distribution $P(\alpha_* | \boldsymbol{q}_*, D)$ is also derived as a Gaussian distribution $\mathcal{N}(\alpha_*; \mu_*, \sigma_*^2)$ [3].

For this derivation, we first introduce a radial basis function (RBF) kernel $k(\boldsymbol{q}_i, \boldsymbol{q}_j; \boldsymbol{\theta})$ which indicates an affinity between two qualities $\boldsymbol{q}_i$ and $\boldsymbol{q}_j$ as

$$k(\boldsymbol{q}_i, \boldsymbol{q}_j; \boldsymbol{\theta}) = v \exp\left(-\frac{\|\boldsymbol{q}_i - \boldsymbol{q}_j\|^2}{2r^2}\right), \tag{3}$$

where $\boldsymbol{\theta} = [v, r]^T$ is a parameter vector for the RBF kernel. Note that this is originally defined as an inner product in a higher-dimensional space mapped from the quality via a kernel trick. Considering a linear regression of the weight from the quality mapped in the higher-dimensional space, the posterior distribution $\mathcal{N}(\alpha_*; \mu_*, \sigma_*^2)$ is derived as Ref. [3]

$$\mu_* = \boldsymbol{k}_*^T (K + \Sigma)^{-1} \boldsymbol{\alpha} \tag{4}$$

$$\sigma_*^2 = k(\boldsymbol{q}_*, \boldsymbol{q}_*; \boldsymbol{\theta}) - \boldsymbol{k}_*^T (K + \Sigma)^{-1} \boldsymbol{k}_* + \sigma_{o,*}^2, \tag{5}$$

where K is an $N \times N$ square matrix whose (i, j) component is $k(\boldsymbol{q}_i, \boldsymbol{q}_j; \boldsymbol{\theta})$, $\boldsymbol{k}_*$ is an N-dimensional vector whose i-th row is $k(\boldsymbol{q}_i, \boldsymbol{q}_*; \boldsymbol{\theta})$, Σ is an $N \times N$ diagonal matrix whose (i, i) component is σ_i^2, and $\sigma_{o,*}^2$ is an observation noise level for the test weight.

We finally adopt the expectation μ as the weight α_* for the given quality $\boldsymbol{q}_*$, which is represented as a weighted sum of the training weights $\boldsymbol{\alpha}$, considering the affinity $\boldsymbol{k}_*$ between the quality α_* and the training qualities Q calculated via the RBF kernel as well as the affinity K within the training qualities Q and training weight variance Σ.

Table 1 The combinations of SRs [pixels] and TRs [fps].

Data set	SR	TR
Training	640 × 480, 320 × 240, 160 × 120, 106 × 80, 80 × 60, 53 × 40	30, 15, 7.5, 5, 3, 1
Test	480 × 360, 213 × 160, 128 × 96, 91 × 68, 64 × 48	10, 6, 3.75, 2

4. Experiment

4.1 Setup

As a data set used in our experiments, we drew 1,935 subjects from the OU-ISIR Gait Database, the large population data set [8], and picked up two walking image sequences per subject observed from a 85-deg view (approximately side view) as a gallery and a probe, which were originally captured with the image size of 640 × 480 pixels at 30 fps.

We randomly divided the 1,935 subjects into disjoint training and test sets, and then trained the weights by LLR with the training set. We repeated this two-fold cross validation 100 times so as to reduce the influence of the random divisions, and used the expectation and variance of the weights over the 100-time trials for GPR.

Moreover, we downsampled the walking image sequences with respect to both the SR and the TR (see Figs. 2 and 3 for examples of downsampled face templates and GEIs), and defined training and test sets of the qualities (combinations of SRs and TRs) used for GPR as listed in **Table 1**. We set the quality value used in GPR as a log of the ratio of the downsampled SR or TR to the original one (e.g., $q_S = \log(0.5)$ for a half-sized image) and also experimentally set the RBF kernel parameters as $r = 0.2$ and $v = 1$.

We compared the proposed quality-dependent LLR whose parameters are estimated by GPR (denoted as LLR (GPR)) with three benchmarks: (1) sum-rule [12] (denoted as Sum), (2) LLR with fixed parameters regardless of the quality [13] (denoted as LLR (Fixed)), and (3) quality-dependent LLR whose parameters are trained using the same quality of the test set, namely, using a sort of ground truth (GT) data (denoted as LLR (GT)). Note that LLR (GT) is actually unavailable under situations of the finite training sets.

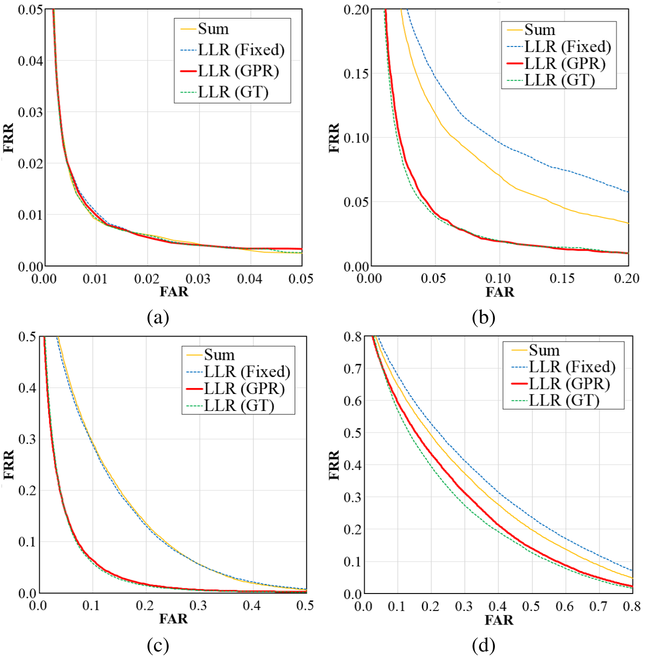
4.2 Results

We evaluate the accuracy in a verification scenario (one-to-one matching) with a receiver operating characteristics (ROC) curve which indicates a tradeoff between the false acceptance rate (FAR) of different subjects and the false rejection rate (FRR) of the same subject. We show the ROC curves for combinations of typical high and low SRs and TRs as shown in **Fig. 4**. As a result, while all the benchmarks are comparative for a high SR and a high TR setting (Fig. 4 (a)), it turns out that the proposed method clearly outperforms Sum and LLR (Fixed) and that it is comparative to LLR (GT) when either the SR or TR, or both of them are low (Fig. 4 (b)–(d)).

In addition, we summarize equal error rates (EERs) of FARs and FRRs as shown in **Table 2**, and also analyze the EER transition along with either SR or TR as shown in **Fig. 5**. Consequently,

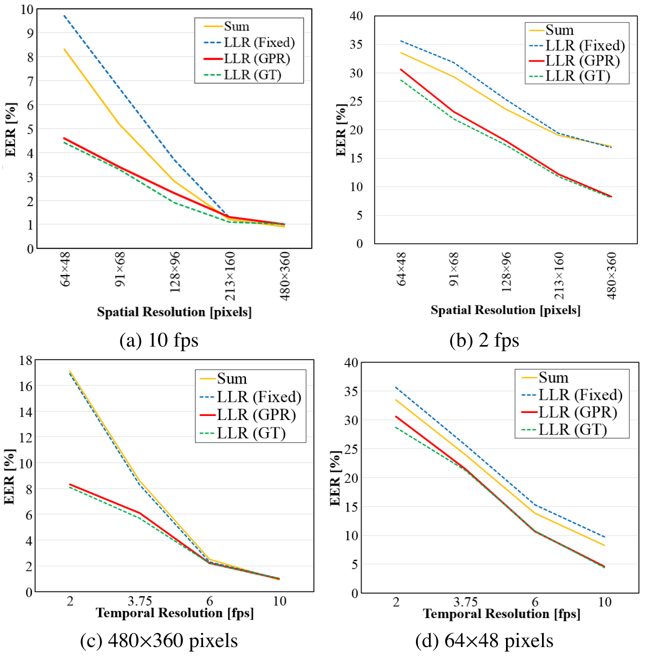
Table 2 EERs [%] for the combinations of SRs [pixels] and TRs [fps] of test data. Bold indicate the best accuracy among Sum, LLR (Fixed), and LLR (GPR).

Fusion	SR	480 × 360				213 × 160				128 × 96				91 × 68				64 × 48			
rule	TR	10	6	3.75	2	10	6	3.75	2	10	6	3.75	2	10	6	3.75	2	10	6	3.75	2
Sum		**0.9**	2.5	8.6	17.1	**1.2**	**3.4**	9.7	19.0	2.8	6.2	14.0	23.6	5.2	9.3	18.9	29.3	8.3	13.8	24.0	33.5
LLR (Fixed)		1.0	2.3	8.3	16.9	1.3	3.7	10.0	19.4	3.7	7.7	15.4	25.3	6.7	11.5	21.3	31.8	9.7	15.3	25.7	35.6
LLR (GPR)		1.0	**2.2**	**6.1**	**8.3**	1.3	**3.4**	**8.7**	**12.2**	**2.3**	**4.9**	**12.7**	**18.0**	**3.4**	**6.8**	**15.9**	**23.2**	**4.6**	**10.7**	**21.5**	**30.6**
LLR (GT)		1.0	2.2	5.7	8.1	1.1	3.4	8.4	11.8	1.9	4.9	12.5	17.3	3.3	6.8	15.7	21.9	4.4	10.7	21.3	28.7



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

Fig. 4 ROC curves on the SRs (left: 480 × 360 pixels, right: 64 × 48 pixels) and the TRs (top: 10 fps, bottom: 2 fps) of test data.



(a) 10 fps (b) 2 fps (c) 480×360 pixels (d) 64×48 pixels

Fig. 5 EER transition along with the SRs under fixed TRs (top) and with the TRs under fixed SRs (bottom).

although EERs get worse as either SR or TR decrease for all the methods, the proposed LLR (GPR) successfully mitigates such accuracy degradations since it appropriately sets the weights of individual modalities depending on the quality, namely, the SR and TR. As a result, we can see that the proposed LLR (GPR) achieves the best EER for almost all the combinations of SR and TR.

5. Discussion

Other approaches to score-level fusion: In addition to LLR used in this paper, several approaches to score-level fusion such as support vector machine (SVM), kernel density estimation (KDE), and minimum-rule could be exploited in multi-modal biometrics. These approaches were evaluated in score-level fusion of face, gait, and the height biometrics and it is reported that some of them are comparable to LLR in Ref. [13] *1. However, note that some of the approaches such as SVM, sum-rule, and minimum-rule just return a score, while LLR returns a posterior probability of the same subject, which is quite important for the purpose of criminal investigation and forensics.

The required number of SRs and TRs of training data: Although we used the fixed number of SRs and TRs as training data to estimate the weights, it is naturally expected that the weight estimation accuracy changes as the number of SRs and TRs of training data changes (e.g., the accuracy would get worse if we decrease the number of SRs and TRs of training data, and vice versa). We therefore plan to confirm how many combinations of SRs and TRs are necessary for a reasonable weight estimation accuracy through experiments of sensitivity analysis in future research.

Effect of quality errors: Although we assume that SRs and TRs as qualities are accurately given in advance based on the camera specification (e.g., image size and frame-rate) and the depth to the target subject, quality errors may be induced by several factors: camera calibration errors, bounding box errors, and frame-rate fluctuations with network cameras. We therefore need to evaluate the sensitivity of such quality errors on the accuracies of the multi-modal biometrics in future research.

6. Conclusion

This paper described a quality-dependent score-level fusion framework of face, gait, and the height biometrics from a single walking image sequence. We considered the SR and TR as a quality and set the optimal weights of the individual modalities based on the LLR framework depending on the qualities. We also estimated the optimal weights for a given test quality from a finite training pairs of the optimal weights and the qualities with a GPR framework. Experiments using a publicly available large popula-

*1 As a reference, EERs by SVM for the typical high and low combinations of SRs [pixels] and TRs [fps] of training data are shown in the supplementary material.

tion gait database showed that the proposed method outperformed the other benchmarks in verification scenarios.

Since we evaluated the proposed method with the database captured under a relatively controlled indoor environment, a future avenue of research involves rigorous experimental validation with more realistic data sets (e.g., CCTV footage in outdoor). In addition, it is also important to analyze sensitivities of the number of SRs and TRs of training data as well as the quality errors on the accuracies of the proposed multi-modal biometrics.

Another future avenue is the implementation of CCTV footage-based person verification system for criminal investigators [7] including the proposed quality-dependent multi-modal biometrics so as to yield practical applications.

Acknowledgments 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] Alonso-Fernandez, F., Fierrez, J., Ramos, D. and Ortega-Garcia, J.: Dealing with sensor interoperability in multi-biometrics: the UPM experience at the Biosecure Multimodal Evaluation 2007, *Proc. SPIE 6994, Biometric Technologies for Human Identification IV*, Orlando, FL, USA (2008).

[2] Bengio, S., Marcel, C., Marcel, S. and Mariethoz, J.: Confidence measures for multimodal identity verification, *Information Fusion*, Vol.3, No.4, pp.267–276 (2002).

[3] Carl Edward Rasmussen, C.K.I.W.: *Gaussian Processes for Machine Learning*, The MIT Press (2006).

[4] Geng, X., Smith-Miles, K., Wang, L., Li, M. and Wu, Q.: Context-aware fusion: A case study on fusion of gait and face for human identification in video, *Pattern Recogn.*, Vol.43, No.10, pp.3660–3673 (2010).

[5] Han, J. and Bhanu, B.: Individual Recognition Using Gait Energy Image, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.28, No.2, pp.316–322 (2006).

[6] Hofmann, M., Schmidt, S.M., Rajagopalan, A. and Rigoll, G.: Combined Face and Gait Recognition using Alpha Matte Preprocessing, *Proc. 5th IAPR Int. Conf. on Biometrics*, New Delhi, India, pp.1–8 (2012).

[7] Iwama, H., Muramatsu, D., Makihara, Y. and Yagi, Y.: Gait Verification System for Criminal Investigation, *IPSJ Trans. Computer Vision and Applications*, Vol.5, pp.163–175 (2013).

[8] Iwama, H., Okumura, M., Makihara, Y. and Yagi, Y.: The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition, *IEEE Trans. Information Forensics and Security*, Vol.7, No.5, pp.1511–1521 (2012).

[9] Jain, A.K., Dass, S.C. and Nandakumar, K.: Soft Biometric Traits for Personal Recognition Systems, pp.731–738 (2004).

[10] Jain, A.K., Flynn, P. and Ross, A.A.: *Handbook of Biometrics*, Springer-Verlag New York, Inc., Secaucus, NJ, USA (2007).

[11] Kale, A., Roy-Chowdhury, A. and Chellappa, R.: Fusion of gait and face for human identification, *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing 2004* (*ICASSP'04*), Vol.5, pp.901–904 (2004).

[12] Kittler, J., Hatef, M., Duin, R.P.W. and Matas, J.: On Combining Classifiers, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.20, No.3, pp.226–239 (1998).

[13] Muramatsu, D., Iwama, H., Makihara, Y. and Yagi, Y.: Multi-view multi-modal person authentication from a single walking image sequence, *2013 International Conference on Biometrics* (*ICB*), pp.1–8 (2013).

[14] Nandakumar, K., Chen, Y., Dass, S. and Jain, A.: Quality-based score level fusion in multibiometric systems, *Proc. 18th Int. Conf. Pattern Recognition*, Vol.4, pp.473–476 (2006).

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

[16] Ross, A.A., Nandakumar, K. and Jain, A.K.: *Handbook of Multibiometrics*, Int. Series on Biometrics, Springer-Verlag New York, Inc., Secaucus, NJ, USA (2006).

[17] Tang, Y. and Srihari, S.N.: Likelihood Ratio Estimation in Forensic Identification using Similarity and Rarity, *Pattern Recognition*, Vol.47, No.3, pp.945–958 (2014).

[18] Toh, K., Yau, W., Lim, E., Chen, L. and Ng, C.: Fusion of auxiliary information for multi-modal biometrics authentication, *Proc. Int. Conf. Biometrics*, Hong Kong, pp.678–685 (2004).

[19] Turk, M. and Pentland, A.: Eigenfaces for Recognition, *J. Cognitive Neuroscience*, Vol.3, No.1, pp.71–86 (1991).

[20] Zhang, T., Li, X., Tao, D. and Yang, J.: Multimodal biometrics using geometry preserving projections, *Pattern Recognition*, Vol.41, No.3, pp.805–813 (2008).

[21] Zhou, X. and Bhanu, B.: Feature fusion of side face and gait for video-based human identification, *Pattern Recognition*, Vol.41, No.3, pp.778–795 (2008).

(Communicated by *Shin'Ichi Satoh*)