IPSJ Transactions on Computer Vision and Applications Vol.5 35–39 (July 2013)

[DOI: 10.2197/ipsjtcva.5.35]

Express Paper

Gait Recognition by Fusing Direct Cross-view Matching Scores for Criminal Investigation

Daigo Muramatsu[1,a)] Yasushi Makihara[1,b)] Yasushi Yagi[1,c)]

Received: March 11, 2013, Accepted: April 24, 2013, Released: July 29, 2013

Abstract: We focus on gait recognition for criminal investigation. In criminal investigation, person authentication is performed by comparing target data at the crime scene and multiple gait data with slightly different views from that of the target data. For this task, we propose fusion of direct cross-view matching. Cross-view matching generally produces worse result than those of same-view matching when view-variant features are used. However, the correlation between cross-view matching with different view pairs is low and it provides improved accuracy. Experimental results performed utilizing large-scale dataset under settings resembling actual criminal investigation cases, show that the proposed approach works well.

Keywords: gait recognition, criminal investication, cross-view, fusion, biometrics

1. Introduction

Recently, many closed-circuit television (CCTV) surveillance cameras have been set up around the world, and are capturing video continuously. These CCTV surveillance cameras could by chance record incidents at a crime scene or at a place near the crime scene. Therefore, such video footage can provide clues in identifying the perpetrators of the crime. Perpetrators can be identified based on their faces if the quality of the footage including resolution and contrast is adequate. However, the resolution of facial images in CCTV footage tends to be low in many cases, while the image contrast may also be low owing to poor illumination conditions. Moreover, the perpetrator may attempt to hide his/her face by wearing a mask for example, to avoid being identified. In these cases face recognition does not work even if the perpetrator appears in the footage, and thus gait recognition is used instead [2], [8], [9].

Gait recognition is a method for biometric person authentication using the shape and motion of the person walking acquired from the footage [12]. Different from many biometrics, gait recognition is possible at a considerable distance from the camera because it works well even if the resolution of the target image sequence is relatively low (e.g., a height of 30 pixels). The accuracy of gait recognition is however, often degraded by several covariates (e.g., views, clothes, and belongings) [13]. Of the covariates, view differences are the most problematic and hence in this paper, we focus on the view issue in gait recognition.

At least two image sequences are necessary for person authentication. These two images sequences are the gallery and the probe. Both the gallery and probe can be captured in controlled [*1] or limited environments for access control application, but they are generally captured in uncontrolled environments for surveillance applications. However, situations for criminal investigation applications are different from those for access control/surveillance applications. In the case of a criminal investigation, once a gait image sequence of the perpetrator associated with the criminal scene is given as a gallery and suspects have been identified, the criminal investigators set out to collect multiple gait image sequences of the suspects in a confidential fashion as probes. Considering the nature of the crime scene, the gallery cannot be captured in a controlled environment, and in the worst case only a single image sequence associated with a single gait cycle from a single view may have been recorded.

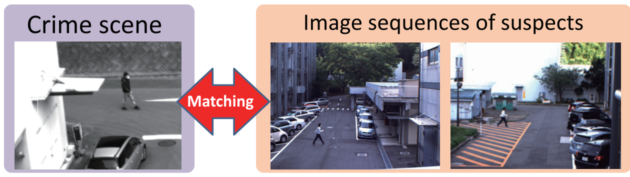


Fig. 1 Typical setting of a criminal investigation.

Probes can be collected in partially controlled environments in the case of a criminal investigation. A probe is acquired after the crime by the criminal investigators so that the view of the probe is similar to that of the gallery. Although criminal investigators apply their best efforts to achieve that in multiple trials, there are some real limitations because the data must be acquired in a confidential fashion. Consequently, multiple gait image sequences from close but different (not exactly the same) views may be collected. **Figure 1** shows a schematic of a typical setting of a

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

*1 In gait recognition, the term *control* means we can set a view that is the relationship between the walking direction and camera position/pose and/or data acquisition settings, which include the type of camera used, and resolution of images.

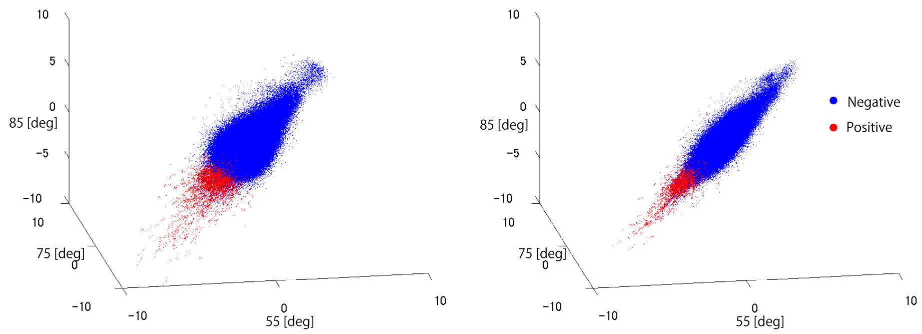


Fig. 2 Scatter plots of positive and negative scores: (left) direct cross-view matching (without view transformation), (right) cross-view matching with view transformation.

criminal investigation.

In this setting, several approaches are possible. A simple approach is to use only the gait feature pair of gallery and probe with the closest (or, if possible, the same) view [4]. This approach works well if the associated views are the same or relatively close. However, a limitation of this approach is that the accuracy degrades if the view difference becomes large. Moreover, this approach discards data from the other close views.

Another approach is to use not only the gait feature pair with the closest view, but also gait feature pairs with different views after transforming the different views to the same view using a view transformation technique, such as the discrete view transformation model (VTM) [6] or arbitrary VTM [11], and to compare the gait feature pairs with the same views. Thereafter, several matching scores are fused for improvement [14]. Although this approach is reasonable and makes use of all the available data, it has certain limitations associated with the applicable view or camera calibration. In the case of applying a discrete view transformation, the applicable views are limited to only several discrete views for which training gait data for the VTM has been collected; this severely limits the applicable criminal scenes. However, although arbitrary VTM is free of view limitations, troublesome camera calibrations are required in all the scenes associated with the gallery and probes.

To solve these problems, we propose a different approach in this paper. In the proposed method, gait features from different views are matched without using a VTM, and the matching scores are fused for authentication. Cross-view matching using view-dependent gait features [3], [7], [10] generally results in worse accuracy than same-view matching, we observe, however, that direct cross-view matching scores are relatively uncorrelated owing to the view-dependent nature of gait features. Since score-level fusion generally works more efficiently with lower score correlations, we make full use of these relatively uncorrelated scores to improve the accuracy.

The contributions of this paper are summarized in the following two points.

1. Accuracy improvement in a criminal investigation setting
Our method works well in settings resembling criminal investigation cases, where views of the gallery and those of the probe sets are close, but slightly different. Owing to the relatively uncorrelated scores obtained from different pairs of cross-view matching, the fusion works efficiently.

2. High usability
Our method does not require troublesome camera calibrations for every scene, because data with different views are directly matched without using an arbitrary VTM.

2. Fusion of Direct Cross-view Matching Scores

2.1 Correlations between Scores

Since the key to success in the proposed method depends on how uncorrelated the scores are for cross-view matching, we first observe score distributions for cross-view matching [*2].

Figure 2 shows the scatter plots of positive (genuine-genuine matching) and negative (genuine-imposter matching) scores associated with cross-view matching of three different view pairs with and without (w/wo) view transformation. From these figures, we observe that the correlations of matching scores with view transformation are higher than those without view transformation. This observation shows that the improvement in fusing scores without view transformation is likely to be larger than that of fusing scores with view transformation.

2.2 Algorithm

After preprocessing and feature extraction, view-dependent gait features are extracted and the extracted probe gait features are directly matched with that of the gallery. The calculated multiple scores are then converted into a single final score using the fusion scheme.

Let x^G and $\mathbf{x}^P = (x_1^P, ..., x_N^P)$ be the gallery gait feature and probe gait feature set from N views, respectively. First, we calculate a distance vector $\mathbf{d}(x^G, \mathbf{x}^P)$ by

$$\mathbf{d}(x^G, \mathbf{x}^P) = (d(x^G, x_1^P), ..., d(x^G, x_N^P)), \quad d(x^G, x_n^P) = \|x^G - x_n^P\|_2, n = 1, 2, ..., N. \tag{1}$$

We then calculate the final score $S(x^G, \mathbf{x}^P)$ by fusing the elements of the distance vector by

$$S(x^G, \mathbf{x}^P) = F_{fusion}(\mathbf{d}(x^G, \mathbf{x}^P); \Theta), \tag{2}$$

where $F_{fusion}(\cdot; \Theta)$ is a fusion function with parameter Θ. Using the final score, a decision is made.

[*2] Explanations of the matching score calculation and database used are given respectively in Sections 3 and 4.

IPSJ Transactions on Computer Vision and Applications Vol.5 35–39 (July 2013)

3. Implementation

3.1 Gait Features

We selected frequency-based features (FDFs) [10] as gait features. The FDFs are generated from a silhouette image sequence in a gait cycle by applying a discrete Fourier transform of the temporal axis. In this paper, we use 0-, 1- and 2-times frequency elements.

3.2 Fusion Function

As the final score, we consider the posterior probability that the suspect is indeed the perpetrator.

Let $\zeta \in \{S(same), D(different)\}$ be a label, and $\bar{\mathbf{d}}(x^G, \mathbf{x}^P)$ be the normalized scores. We set the final score by

$$S(x^G, \mathbf{x}^P) = P(\zeta = S|\bar{\mathbf{d}}(x^G, \mathbf{x}^P)). \qquad (3)$$

We consider the linear logistic regression (LLR) of the posterior probability

$$\log\left(\frac{P(\zeta = S|\bar{\mathbf{d}}(x^G, \mathbf{x}^P))}{1 - P(\zeta = S|\bar{\mathbf{d}}(x^G, \mathbf{x}^P))}\right) = \theta_0 + \sum_{n=1}^{N} \theta_n \bar{d}(x^G, x_n^P), \qquad (4)$$

and calculate the posterior probability as

$$P(\zeta = S|\bar{\mathbf{d}}(x^G, \mathbf{x}^P)) = \frac{1}{1 + e^{-(\theta_0 + \sum_{n=1}^{N} \theta_n \bar{d}(x^G, x_n^P))}}. \qquad (5)$$

Parameter $\Theta = (\theta_0, \theta_1, \ldots, \theta_N)$ is set through training to minimize the objective function proposed in Ref. [1].

4. Experiment

4.1 Database

We used a subset of the OU-ISIR database [5] for accuracy evaluation because it includes gait image sequences from different views. Both the gallery and probes in the database consist of image sequences collected from the same 1,912 subjects, with each image sequence divided into four subsets based on the observation angles. From these sequences, we used the subset with observation angle 65 [deg] as the gallery, and subsets with observation angles 55, 65, 75, and 85 [deg] as probes.

4.2 Experimental Settings

We performed two experiments, a preliminary and a main experiment. The purpose of the preliminary experiment was to confirm that utilizing the VTM as a benchmark improves the accuracy of individual cross-view matching against the database, while that of the main experiment was to evaluate the accuracy of the proposed method compared with that of other methods.

In the main experiment, we considered the four settings associated with criminal investigations as given in Table 2.

In settings A and B, gait features with exactly the same gallery view were not available; instead, gait features with similar, but different views associated with both sides of the gallery view were used. In settings C and D, gait features with exactly the same gallery view were included in the probe gait features.

In all the experiments, we randomly divided the data into two sub-groups, and performed two-fold cross-validation. To eliminate the impact of the grouping, we repeated the cross-validation five times.

Table 1 EERs and Rank-1 values for the same-view and cross-view matching (preliminary experiment).

Matched view (Gallery, Probe)	EER [%] Without VTM	EER [%] With VTM	Rank-1 [%] Without VTM	Rank-1 [%] With VTM
Same (65, 65)	2.35	-	89.7	-
Cross (65, 55)	5.42	3.37	60.9	81.7
Cross (65, 75)	3.55	3.51	73.4	79.6
Cross (65, 85)	7.08	4.43	40.1	68.5

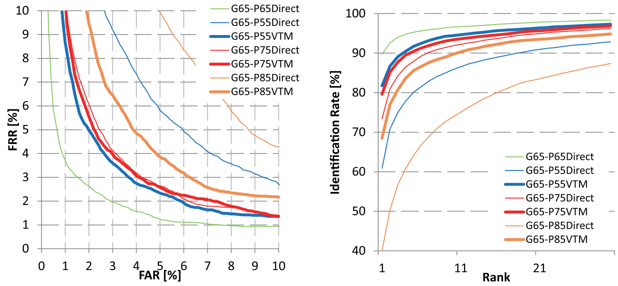


Fig. 3 ROC (upper) and CMC (bottom) curves for the same-view and cross-view matching w/wo VTM.

4.3 Preliminary Experiment

Table 1 gives the equal error rates (EER) and rank-1 identification rates (referred to as Rank-1 in this paper), while **Fig. 3** shows the receiver operating characteristic (ROC) curve depicting the tradeoff between the false acceptance rate (FAR) and false rejection rate (FRR), and the cumulative matching characteristic (CMC) curve showing the relationship between the rank and cumulative matching rates. From these results, we observe that: 1) the EER and Rank-1 improve when applying the VTM, and 2) the EER and Rank-1 for the same view are better than those for the cross-view despite application of the VTM.

4.4 Main Experiment

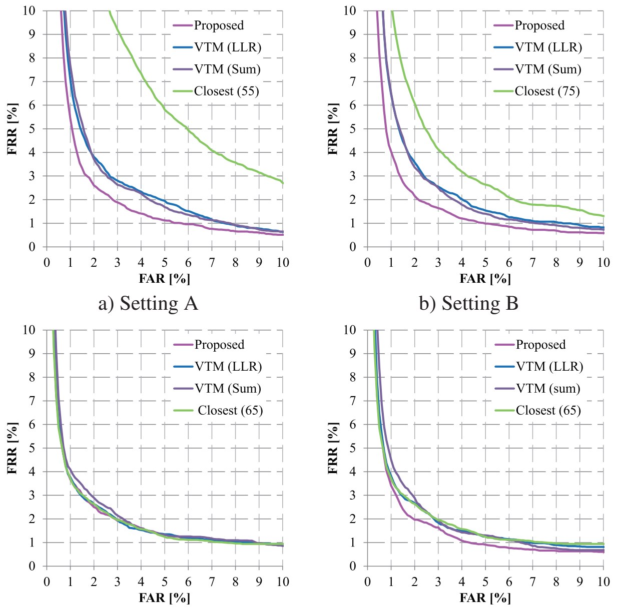
In **Table 2**, **Fig. 4**, and **Fig. 5**, we show the EERs and Rank-1 values, ROC curves, and CMC curves, respectively, for the proposed method in several settings together with those of the closest view and fusions of a cross-view with the VTM using sum rule [14] and LLR. As for the closest view associated with settings A and B, we select the view with the better accuracy.

According to these results, the proposed method achieved the best results in almost all the settings. Experimental results from settings A and B show that the proposed method achieves comparable or better accuracy than that of the same-view matching (see Table 1) even though a probe with exactly the same view as the gallery is not available. Moreover, the proposed method improves the accuracy obtained in setting D, which includes probe data with exactly the same view as the gallery. This result shows that the direct cross-view matching scores are relatively uncorrelated with the same-view scores, and that fusion of these scores works efficiently.

The accuracy of fusion with cross-view matching using the VTM is better than that of the closest view in settings A and B where exactly the same view as the gallery is not available but accuracy is worse than that of the proposed method. In settings C and D, the accuracy of fusion with cross-view matching using the VTM is almost the same as that of the closest view, and worse than that of the proposed method in setting D, despite the

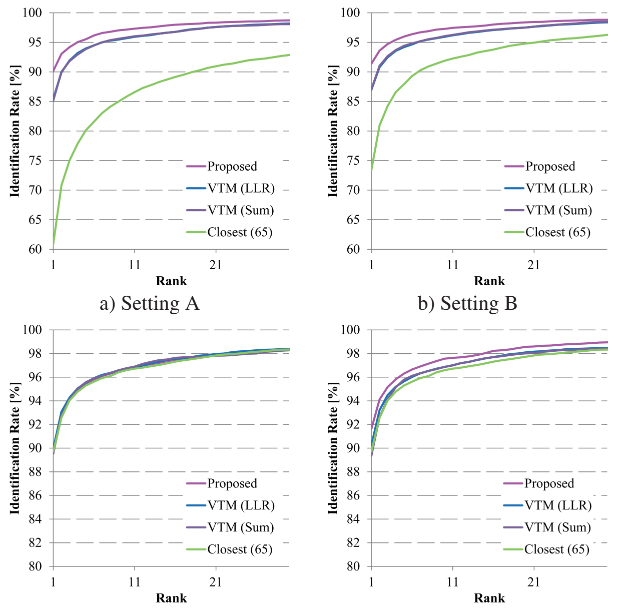
Table 2 EERs and Rank-1 values.

Setting	View		EER [%]				Rank-1 [%]			
	Gallery	Probe	Closest view	VTM (Sum) [14]	VTM (LLR)	Proposed (LLR)	Closest view	VTM (Sum) [14]	VTM (LLR)	Proposed (LLR)
A	65	(55, 85)	5.42	*2.80*	2.87	**2.35**	60.9	85.1	*85.5*	**90.1**
B	65	(55, 75)	3.55	*2.69*	2.73	**2.07**	73.4	*87.1*	87.0	**91.4**
C	65	(55, 65)	*2.35*	2.51	2.40	**2.28**	89.7	89.5	**89.9**	*89.9*
D	65	(55, 65, 85)	*2.35*	2.40	2.36	**1.98**	89.7	89.4	*90.3*	**91.6**



a) Setting A
b) Setting B
c) Setting C
d) Setting D

Fig. 4 ROC curves.



a) Setting A
b) Setting B
c) Setting C
d) Setting D

Fig. 5 CMC curves.

fact that the VTM improves the accuracy of individual cross-view matching as shown in the preliminary experiment (see Fig. 3 and Table 1). These results imply that even though the accuracy of individual cross-view matching is improved by the VTM, fusion of the same-view matching scores with the cross-view matching score by the VTM does not always improve the accuracy, since these scores are highly correlated.

5. Conclusions and Future Work

In this paper, we focus on gait recognition for criminal investigation and propose a method to fuse direct cross-view matching scores.

Experimental results show that individual direct cross-view matching cannot achieve higher accuracy than cross-view matching by the VTM, whereas multiple direct cross-view matching does contribute to improving the accuracy by fusion owing to the relatively uncorrelated scores. Consequently, the proposed method achieved either the best or comparable accuracy. Moreover, the proposed method has another advantage in that it can be applied to any criminal scene without troublesome camera calibrations, which is also important when considering usage in criminal investigations.

Our method assumes that views of the probe are different, but close to that of the gallery. However, if the view difference becomes large, the proposed method does not work well because individual direct cross-view matching does not achieve acceptable accuracy. This is a limitation on the proposed method caused by the tradeoff between accuracy degradation of cross-view matching and uncorrelation among the cross-view matching scores. We need to further analyze which view differences are acceptable for applying the proposed method to improve accuracy.

In this paper, we only focused on view difference between the gallery and probes, and evaluated cases where azimuth angles are different in the experiments. We also need to consider other covariates that influence the accuracy of gait recognition such as tilt angle and speed changes. We believe the proposed approach has the potential to be applied to these covariates. This will be evaluated in a future work.

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] Bouchrika, I., Goffredo, M., Carter, J. and Nixon, M.: On Using Gait in Forensic Biometrics, *Journal of Forensic Science*, Vol.56, No.4, pp.882–889 (2011).

[3] 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).
[4] Han, J., Bhanu, B. and Roy-Chowdhury, A.: A Study On View-Insensitive Gait Recognition, *Proc. IEEE Int. Conf. Image Processing*, Vol.3, pp.297–300 (2005).
[5] 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).
[6] Kusakunniran, W., Wu, Q., Zhang, J. and Li, H.: Gait Recognition Under Various Viewing Angles Based on Correlated Motion Regression, *IEEE Trans. Circuits and Systems for Video Technology*, Vol.22, No.6, pp.966–980 (2012).
[7] Lam, T.H.W., Cheung, K.H. and Liu, J.N.K.: Gait Flow Image: A Silhouette-based Gait Representation for Human Identification, *Pattern Recogn.*, Vol.44, pp.973–987 (2011).
[8] Larsen, P.K., Simonsen, E.B. and Lynnerup, N.: Gait Analysis in Forensic Medicine, *Journal of Forensic Sciences*, Vol.53, No.5, pp.1149–1153 (2008).
[9] Lynnerup, N. and Vedel, J.: Person Identification by Gait Analysis and Photogrammetry, *Journal of Forensic Science*, Vol.50, No.1, pp.112–118 (2005).
[10] Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T. and Yagi, Y.: Gait Recognition Using a View Transformation Model in the Frequency Domain, *Proc. 9th European Conf. Computer Vision*, Graz, Austria, pp.151–163 (2006).
[11] Muramatsu, D., Shiraishi, A., Makihara, Y. and Yagi, Y.: Arbitrary View Transformation Model for Gait Person Authentication, *Proc. 5th IEEE Int. Conf. Biometrics: Theory, Applications and Systems*, No.Paper ID 39, Washington D.C., USA, pp.1–6 (2012).
[12] Nixon, M.S., Tan, T.N. and Chellappa, R.: *Human Identification Based on Gait*, Int. Series Biometrics, Springer-Verlag (2005).
[13] Sarkar, S., Phillips, J., Liu, Z., Vega, I., ther, P.G. and Bowyer, K.: The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.27, No.2, pp.162–177 (2005).
[14] Sugiura, K., Makihara, Y. and Yagi, Y.: Omnidirectional Gait Identification by Tilt Normalization and Azimuth View Transformation, *Proc. IEEE Workshop on OMNIVIS*, Marseille, France (2008).

(Communicated by *Hidekata Hontani*)