Distinguishing Pedestrians Facing to the Front and the Side by Gait Observation

Mitsuru Nakazawa*, Ikuhisa Mitsugami*, Hirotake Yamazoe† and Yasushi Yagi*
*ISIR, Osaka Univ.
8-1 Mihogaoka, Ibaraki, Osaka, Japan 567–0047
E-mail: {nakazawa, mitsugami, yagi}@am.sanken.osaka-u.ac.jp
†OSIPP, Osaka Univ.
1-6 Machikaneyama-cho, Toyonaka, Osaka, Japan 560–0043
Email: yamazoe@osipp.osaka-u.ac.jp

***Abstract*—In this paper, we propose a method to distinguish pedestrians facing to the front and the side by using a low resolution and quality surveillance image sequence. In the past, there have been many methods to estimate the head orientation of a pedestrian. However, because all these methods use facial texture information to achieve the goal, it is difficult to apply the methods to a low resolution and quality image sequence that does not include enough information. Therefore, we focus on the gait change of a pedestrian, which can be acquired even from a low-resolution silhouette sequence. Experiments confirm the effectiveness of the proposed method by using low-resolution image sequences of over one hundred subjects.**

***Keywords*—head orientation estimation; gait; low resolution images;**

I. Introduction

In both indoor and outdoor environments, many surveillance cameras are located for safety, security and traffic measurement. If we can estimate the head orientation of pedestrians by using these surveillance cameras, it would be possible to estimate their intentions and behaviors. For example, in marketing applications, it would be possible to understand whether or not people become interested in an advertisement based on their head orientation [1]. In safety applications, it would be possible to alert careless people whose head orientation is not front near an area where they must watch their step.

In the past, there have been many methods to estimate his/her head orientation from a surveillance camera [2]. Benfold et al. [3] proposed a classification approach using randomized ferns with decision branches based on HOG and color-based features. Zhang et al. [4] proposed an approach using multiple orientation-specific face detectors based on a boosting algorithm. Tosato et al. [5] proposed an approach using non-linear regression based on the continuity of a head orientation space. Jung et al. [6] proposed an approach using head pose tracking with a 3D ellipsoidal model. To keep the tracking accuracy even in bad capturing conditions, they also utilize a walking trajectory to approximately estimate a head region. In common with all these methods, they use facial texture information for the head orientation estimation. Thus, they cannot be applied to a low resolution and quality image sequence that does not include enough information.

We therefore propose a method to estimate the head orientation of a pedestrian without using facial texture information. As information except facial texture one, we focus on a way of walking (i.e., gait) [7]. Gait has already used in many applications using a silhouette image sequence such as human identification [8]–[10], age estimation [11] and gender estimation [12]. When distinguishing the head orientations, we can utilize the gait change of not only the head region but also the other body regions. They must be useful when the pixel size of the head region is quite small. In this paper, to confirm the effectiveness of our approach, we set the goal as distinguishing pedestrians who face to the front and the side.

The remaining of the paper is organized as follows. In Section II, we explain how to distinguish pedestrians who face to the front and the side using a gait change. In Section III, we show some experiments with more than one hundred subjects to prove the effectiveness of using the gait change. In Section IV, the conclusion of this paper and future works are described.

II. Method to distinguish the head orientations

The overview of the distinction method is shown in Fig. 1. First, silhouette images of a pedestrian are extracted from

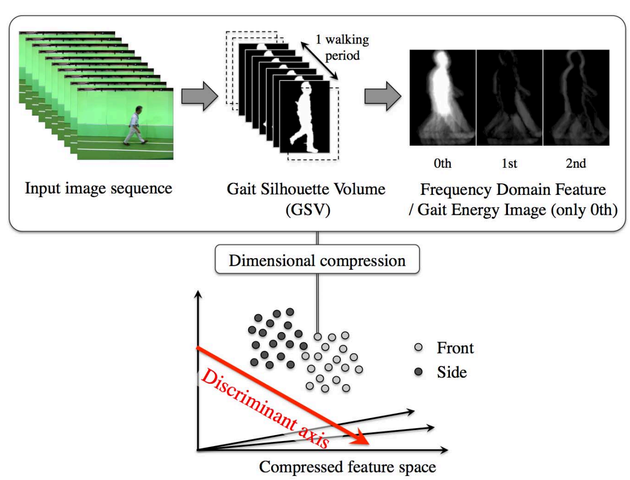


Fig. 1: Overview of the distinction method

978-1-4799-2190-4/13 $26.00 © 2013 IEEE
DOI 10.1109/ACPR.2013.136

an input image sequence. Next, the position alignment and the size normalization are performed on the extracted silhouette images. We define the silhouette images stacked in the direction of the time axis as the Gait Silhouette Volume (GSV). Then, a walking period is determined as the number of frames that gives the highest value of the normalized autocorrelation of the GSV in the time axis. Finally, we obtain the Frequency Domain Features (FDF) [8], by performing Discrete Fourier Transformation (DFT) at each pixel of the GSV as follows:

$$G(x, y, k) = \sum_{n=1}^{N_{\mathrm{gait}}} g(x, y, n) e^{-j\omega_0 k n}, \quad (1)$$

where N_{gait} is the gait period and $g(x, y, n)$ is a silhouette value of the GSV at a pixel (x, y) in the n-th frame. $\omega_0 = 2\pi/N_{\mathrm{gait}}$ is the base angular frequency for the gait period N_{gait}. $G(x, y, k)$ is the DFT of the GSV for k-times the base frequency. Then its amplitude spectrum $A(x, y, k)$ is calculated as follows:

$$A(x, y, k) = \frac{1}{N_{\mathrm{gait}}} |G(x, y, k)|. \quad (2)$$

As with [8], the direct component ($k = 0$, i.e. averaged silhouette) and low frequency ones ($k = 1, 2$) are used as the FDF in our method. Note that the only direct component of the FDF is equivalent to the Gait Energy Image (GEI) [7]. In this paper, the FDF and the GEI are used as gait features.

To classify the head orientation of pedestrians into two classes ("Front" and "Side"), we generate a discriminant model by using gait features whose class label is given. In our method, first each gait feature is compressed by the Principal Component Analysis. Then, we obtain the discriminant model that projects the compressed gait features into a space where the ratio of the variance between the classes to the variance within the classes is maximized by the Linear Discriminant Analysis (LDA).

III. Experiments

A. Experimental setting

To capture many pedestrians who turned his/her head to the front or the side, we prepared a game where a subject sought and followed a moving character shown in an 18m width screen while walking in a lane (Fig. 2). At the top of the screen, range sensors (Microsoft Kinects) were located so that the subjects were captured from their side. Using range data captured from the Kinects, the silhouette images of subjects could be easily extracted by background subtraction. In the size-normalization of the silhouette images, we set the resolution of GSVs as 128×88 pixel, which was adopted in [13].

To give the ground truth, we acquired the head orientation of each subject by visual judgment from a color image sequence that was also captured from the Kinects. From captured data of 113 subjects, we labeled a total of 230 and 79 subjects as "Front" and "Side", respectively. In the evaluation of our discrimination, since the number of subjects of the two classes was uneven, we first randomly sampled subjects so that their numbers should be equal. Then we calculated the accuracy rates of the discrimination by Leave-one-out Cross Validation. After fifty loops of the random subject sampling and the accuracy rate calculation, the total performance was obtained by the average of the accuracy rates.

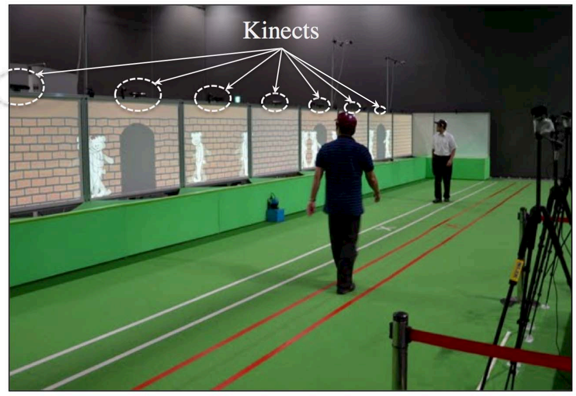


Fig. 2: Scene of a walking game

TABLE I: Confusion matrix of the head orientation discrimination

	Front	Side
Front	93.0%	7.0%
Side	8.7%	91.3%

B. Results of head orientation discrimination

TABLE I shows the confusion matrix of the head orientation discrimination using the FDF. Our method achieved over 90% accuracy rates without using facial texture information. Additionally, as can be seen in Fig. 3, the distributions of two classes were well discriminated in the projected feature space. To see the representation of the discriminant axis, (A) and (B) of Fig. 3 were chosen as the representative sample point of two classes and reprojected to their FDFs (Figures 4 (a) and (b)). From this result, it is found that although the appearance of the head shape, the back shape and the length of stride varied by changing the head orientation, the most markedly-changed region is the right side of the head. It was caused by the brim of a cap. In fact, we asked all subjects to wear the cap because it is difficult for Kinect to obtain range data of a hair region. Therefore, to eliminate the effect of the brim in the estimation, we masked the brim region in all GSVs. TABLE II shows the confusion matrix in the case that the brim region was masked. Although the accuracy rates fell a little, we could keep high accuracy rates as about 88%. In the remaining experiments, the brim region of the cap was masked on GSVs beforehand for the above reason.

C. Results comparison between the FDF and the GEI

We compared the result of the head orientation discrimination between the FDF [8] and the GEI [7]. Additionally, we also show the result using only low frequency components

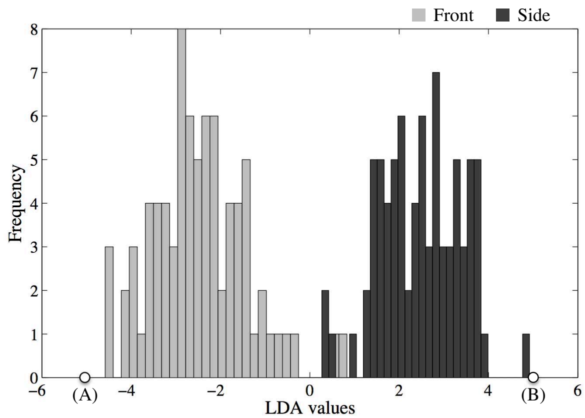
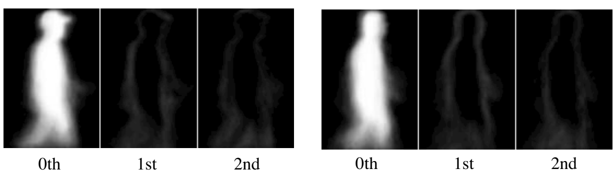


Fig. 3: LDA value histogram of each head orientation class



(a) Point (A) (b) Point (B)

Fig. 4: FDF reprojected from two points of Fig. 3

(1st+2nd) to see whether they have an effect on the head orientation discrimination.

Figure 5 presents their accuracy rates. As a result, the accuracy rates of the FDF and the GEI were almost same, and the ones of low frequency components were lower than the others. Therefore, it can be said that the direct component is the most crucial, and the lower frequency ones have less information for the head orientation discrimination. Here, let us see the reprojected FDFs of Fig. 4. It can be said that the low frequency components appeared near the contour of the direct component. Because the lower frequency ones were redundant information of the GEI, they were not useful for the head orientation discrimination. In the remaining experiments, we used the GEI for the above reason.

D. Results comparison in GEIs masked on body parts while changing their resolution

To confirm the effectiveness of using the gait of the whole body, we prepared some masked GEIs on body regions and obtained the accuracy rates of GEIs that included only head, only torso and only leg regions. Additionally, we downsampled these GEIs and obtained their accuracy rates not only to confirm whether our method could be applied to a lower resolution image sequence but also to compare the change of the accuracy rates in the body regions.

Figures 6 (a), (b) and (c) show the accuracy rates of the GEIs of each body region, respectively. As we had expected, the head region was crucial for the head orientation discrimination. However, its accuracy rates were decreased by downsampling the GEIs because it became difficult to express the difference of the head shape in the head orientation. On the other hands, although the accuracy rates of the torso and the leg region were lower than the ones of the head regions, their accuracy rates were more stable. This is because these regions that are larger than the head region could express the difference in the head orientation more stably. When comparing between Figures 6 (a) and (d), it can be said that the result of the whole body was better than the one of the only head region, especially in lower resolution GEIs. From this result, we could confirm the effectiveness of using not only the head but also the other body regions in the head orientation discrimination.

TABLE II: Confusion matrix of the head orientation discrimination in the case that the brim region of the cap was masked

	Front	Side
Front	87.8%	12.2%
Side	12.5%	87.5%

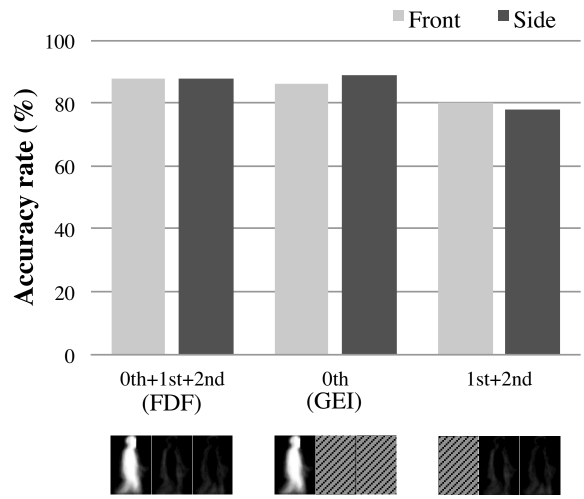
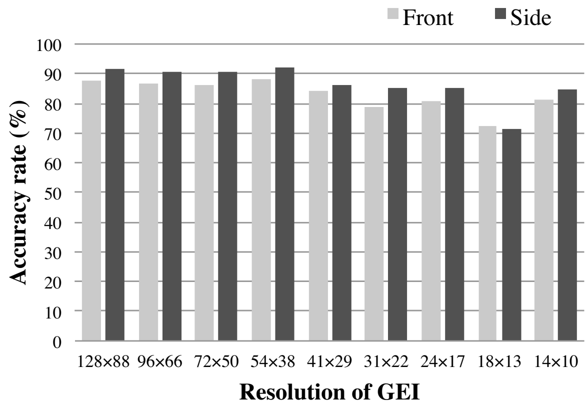


Fig. 5: Accuracy rates of the combinations of FDF components

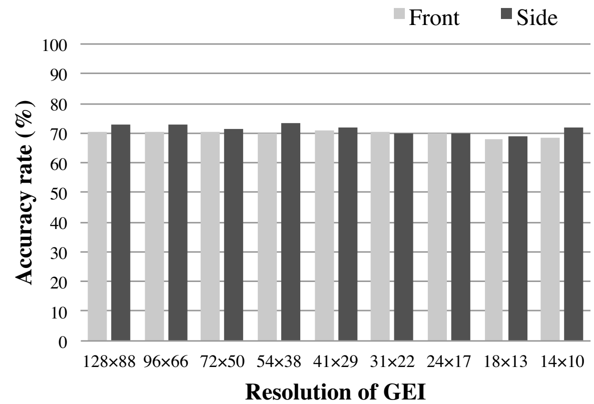
IV. Conclusion

In this paper, we proposed a method to distinguish pedestrians who face to the front and the side by using a gait change that can be acquired from a low resolution and quality surveillance image sequence. In the experiments, it was confirmed that the gait change of not only the head region but also the other body regions is effective for the head orientation discrimination in a low resolution image sequence.

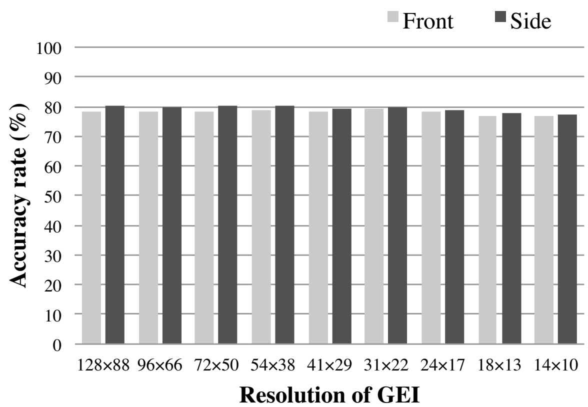
In future work, based on our proposed method, we will try to distinguish pedestrians facing to the front and the side in a shopping mall as one of real environments. It will be possible



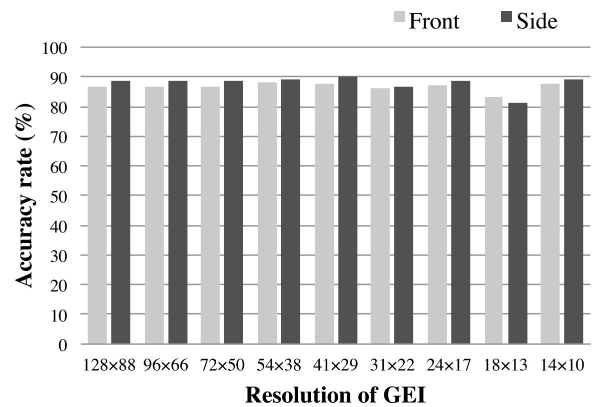
(a) Only head region



(b) Only torso region



(c) Only leg region



(d) Whole body region

Fig. 6: Accuracy rates of low resolution GEIs

to understand whether walking shoppers are interested in a shop and its products that are located laterally on a walkway in the mall. For the achievement of the distinction in a real environment, we will extend our proposed method so that a gait feature can be obtained even when a pedestrian walks not only straightly but also freely. Furthermore, to model the relation between a gait change and the head orientation, we will try to compute the regression analysis between a gait feature and its head orientation angle.

Acknowledgment

This work was supported by Core Research for Evaluation Science and Technology (CREST) of the Japan Science and Technology Agency (JST).

References

[1] K. Smith, S. O. Ba, J.-M. Odobez, and D. Gatica-Perez, "Tracking the visual focus of attention for a varying number of wandering people," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 30, no. 7, pp. 1212–1229, 2008.

[2] T. Siriteerakul, "Advance in head pose estimation from low resolution images: A review," *International Journal of Computer Science Issues (IJCSI)*, vol. 9-2, no. 3, pp. 442–449, March 2012.

[3] B. Benfold and I. Reid, "Guiding visual surveillance by tracking human attention," in *Proceedings of British Machine Vision Conference (BMVC) 2009*, London, UK, September 2009.

[4] M. L. Zhenqiu Zhang, Yuxiao Hu and T. Huang, "Head pose estimation in seminar room using multi view face detectors," in *Proceedings of the 1st international evaluation conference on Classification of events, activities and relationships (CLEAR2006)*, Southampton, UK, 2006, pp. 299–304.

[5] D. Tosato, M. Farenzena, M. Spera, V. Murino, and M. Cristani, "Multi-class classification on riemannian manifolds for video surveillance," in *Proceedings of the 11th European Conference on Computer Vision (ECCV2010)*, Crete, Greece, September 2010, pp. 378–391.

[6] S.-U. Jung and M. Nixon, "On using gait biometrics to enhance face pose estimation," in *Proceedings of The IEEE Fourth International Conference on Biometrics: Theory, Applications and Systems (BTAS 10)*, Washington D.C., USA, September 2010, pp. 1–6.

[7] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 28, no. 2, February 2006.

[8] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi,

"Gait recognition using a view transformation model in the frequency domain," in *Proceedings of the 9th European conference on Computer Vision (ECCV2006)*, vol. 3, Graz, Austria, May 2006, pp. 151–163.

[9] H. Ali, J. Dargham, C. Ali, and E. G. Moung, "Gait recognition using gait energy image," *International Journal of Signal Processing*, vol. 4, no. 3, pp. 141–512, September 2011.

[10] 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*, vol. 7, no. 5, pp. 1511–1521, 2012.

[11] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi, "Gait-based age estimation using a whole-generation gait database," in *Proceedings of the International Joint Conference on Biometrics (IJCB2011)*, no. 195, Washington D.C., USA, October 2011, pp. 1–6.

[12] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu, "A study on gait-based gender classification," *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1905–1910, 2009.

[13] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanid gait challenge problem: Data sets, performance, and analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 27, pp. 162–177, 2005.