IPSJ Transactions on Computer Vision and Applications Vol.6 63–67 (July 2014)

[DOI: 10.2197/ipsjtcva.6.63]

Express Paper

Head Orientation Estimation using Gait Observation

Mitsuru Nakazawa[1,a)] Ikuhisa Mitsugami[1,b)] Hirotake Yamazoe[2,c)]
Yasushi Yagi[1,d)]

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

Abstract: We propose a novel method to estimate the head orientation of a pedestrian. There have been many methods for head orientation estimation based on facial textures of pedestrians. It is, however, impossible to apply these methods to low-resolution images which are captured by a surveillance camera at a distance. To deal with the problem, we construct a method that is not based on facial textures but on gait features, which are robustly obtained even from low-resolution images. In our method, first, size-normalized silhouette images of pedestrians are generated from captured images. We then obtain the Gait Energy Image (GEI) from the silhouette images as a gait feature. Finally, we generate a discriminant model to classify their head orientation. For this training step, we build a dataset consisting of gait images of over 100 pedestrians and their head orientations. In evaluation experiments using the dataset, we classified their head orientation by the proposed method. We confirmed that gait changes of the whole body were efficient for the estimation in quite low-resolution images which existing methods cannot deal with due to the lack of facial textures.

Keywords: head orientation, Gait, low-resolution images

1. Introduction

In both indoor and outdoor environments, many surveillance cameras are located for safety, security and traffic measurement. If we can estimate head orientations of pedestrians by using these surveillance cameras, it would be possible to estimate their intentions and behaviors. For example, in a marketing application, it would be able to understand whether or not people are interested in signage based on their head orientation [17]. As an application of intelligent driver support systems, head orientation would help to predict the walking direction of a pedestrian for their protection [6]. In social psychology studies, head orientation would be a proxy for determining the social attention direction in group interaction [18].

There have been many methods for estimating head orientation from images [14], and some of them estimate the head orientation even from low-resolution images as are captured by a surveillance camera [4], [9], [19], [23]. These methods are based on the facial texture information of a pedestrian to achieve the goal. Thus, it is impossible to apply them to low-resolution images where facial textures are not rich enough.

To deal with the problem, we propose a novel method to estimate the head orientation of a pedestrian even from low-resolution images without using facial textures. In this method, we utilize gait [7], i.e., way of walking.

[1] The Institute of Science and Industrial Research, Osaka University, Ibaraki, Osaka 567–0047, Japan
[2] Osaka School of International Public Policy, Osaka University, Toyonaka, Osaka 560–0043, Japan
[a)] nakazawa@am.sanken.osaka-u.ac.jp
[b)] mitsugami@am.sanken.osaka-u.ac.jp
[c)] yamazoe@osipp.osaka-u.ac.jp
[d)] yagi@am.sanken.osaka-u.ac.jp

Since gait is obtained from the whole body region, it consists of more pixels than only the head region, which means it is applicable for even lower-resolution images. Gait has been used for several applications such as human identification [1], [8], [12], age estimation [11] and gender estimation [22]. This fact means that gait is actually different among people and is affected by those properties of people. Moreover, it is reported that gait is affected by the head orientation of pedestrians with respect to their walking direction [15]. Motivated by this report, our method is designed to estimate the head orientation from gait observation.

Our proposed method comprises the following steps: first, size-normalized silhouette images of pedestrians are generated from input images for the preparation of gait feature acquisition. Then, from the silhouette images, we obtain the Gait Energy Image (GEI) [7], which has the advantage that it is less affected by segmental errors of gait silhouette images [1]. Finally, to classify the head orientation of a pedestrian, we generate a discriminant model.

The proposed method was experimentally evaluated. For experiments, we built a dataset consisting of gait images of over 100 pedestrians and their three head orientations, that is, the front, the diagonal front and the side. Our dataset is novel in that it involves whole body images, which have not been included in existing datasets [5], [13], [20]. Using this dataset as training and testing data, the classification performance was evaluated in various resolutions to confirm its robustness to the decrease of resolution. From the results, we confirmed that the proposed method can estimate the head orientation with high accuracy from low-resolution images that facial-texture-based methods cannot cope with. Note that although the head orientations were classified into just three classes in the experiments, it is still effective; in a marketing ap-

© 2014 Information Processing Society of Japan

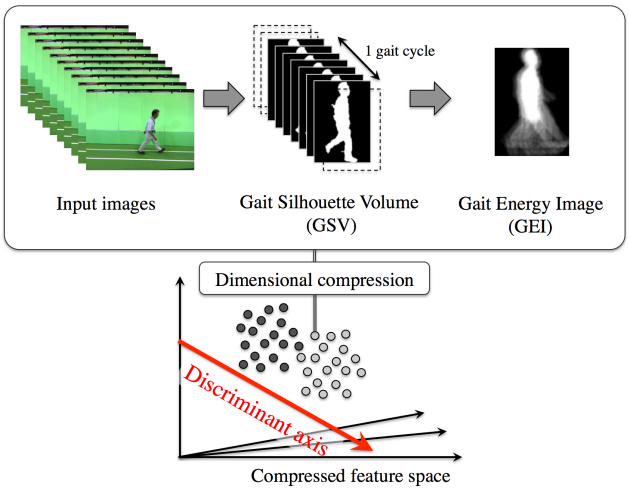


Fig. 1 Overview of our estimation method.

plication, for example, it will be helpful to understand whether shoppers are less or more interested in a shop that is located laterally on a walkway of a shopping mall.

2. Head Orientation Estimation

An overview of our estimation method is shown in **Fig. 1**. First, pedestrian silhouettes are extracted from input images. Next, position alignment and size normalization are performed on the extracted silhouette images. The silhouette images stacked in the direction of the temporal axis is defined as the Gait Silhouette Volume (GSV) [12]. Then, to determine a gait cycle, we acquire the number of frames that gives the highest value of the normalized autocorrelation of the GSV in the temporal axis. Finally, we obtain a gait feature from the GSV. In our method, GEI is employed because it is less affected by segmental errors of gait silhouette images [1]. GEI $G(x, y)$ is computed as follows:

$$G(x, y) = \frac{1}{N} \sum_{n=1}^{N} s(x, y, n), \tag{1}$$

where N is the number of frames in a gait cycle and $s(x, y, n)$ is a silhouette value of the GSV at a pixel (x, y) in the n-th frame.

To estimate the head orientation of a pedestrian, we generate a discriminant model by using training data. In our method, first the principal component analysis is performed on the gait features to compress the feature dimension. 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 Linear Discriminant Analysis (LDA).

3. Gait Dataset for Head Orientation Estimation

To realize our estimation method, it is necessary to prepare training data that consist of gait features of pedestrians and their head orientations. Therefore, we created a gait dataset that varies head orientation.

Each subject walked in an environment (**Fig. 2**) while paying attention to the walking direction or other directions through a target shown beside them. The subjects were captured from their

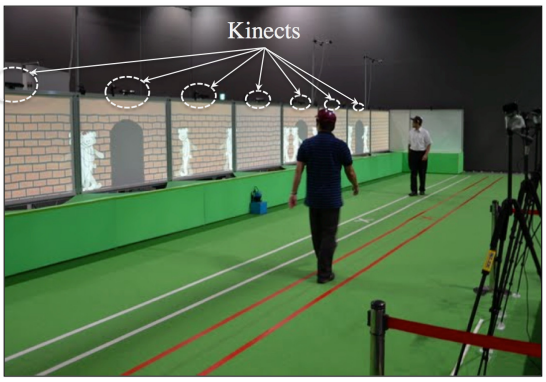


Fig. 2 Environment for dataset construction.

(a) Front

(b) Diagonal front

(c) Side

Fig. 3 Example silhouette images composing GSVs of the dataset.

side by using range sensors (Microsoft Kinects) located within the environment. Using captured range data, the silhouettes 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 pixels, which was adopted in Ref. [16].

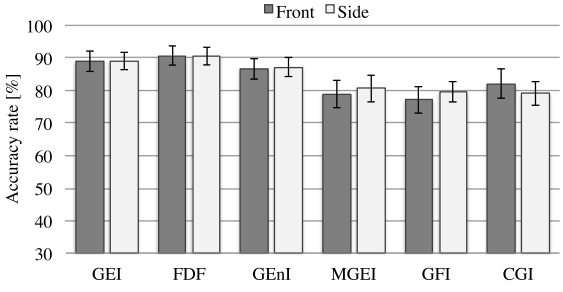
To acquire the ground truth of their head orientation, we visually annotated the head orientation of each subject by using color images captured from the Kinects. From captured data of 113 subjects, we labeled the head orientation of a total of 224, 162 and 77 subjects as "Front," "Diagonal front" and "Side" relative to the walking direction, respectively. **Figure 3** shows some examples of the silhouette image sequences.

4. Experiments

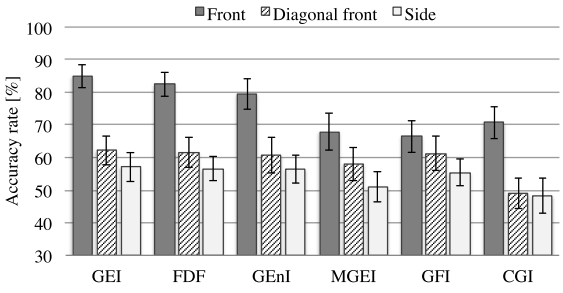
In the evaluation of our proposed method, we conducted (1) classification of "Front" and "Side" subjects and (2) classification of "Front," "Diagonal front" and "Side" subjects as an easy and a challenging task. Henceforth, we call them as two-class and three-class estimation, respectively.

In gait feature acquisition, we masked the right of the head region, which corresponds to the brim of a cap. The reason why this region appeared is that we asked all subjects to wear a cap so that the Kinects could obtain range data of the hair region. The brim region can be seen in Fig. 3 (a). To prevent the region from

© 2014 Information Processing Society of Japan



(a) Two-class estimation



(b) Three-class estimation

Fig. 4 Accuracy rates of each gait feature. The standard deviations are depicted by error bars.

having an effect on the estimation, we eliminated the region by the masking operation.

In head orientation estimation, since the number of subjects of the classes was uneven, we first randomly sampled subjects so that their numbers became equal. Then, we calculated the accuracy rates by leave-one-out cross validation. After 50 iterations of the random subject sampling and the accuracy rate calculation, the total performance was obtained by the average of the accuracy rates.

4.1 Result Comparison with Other Gait Features

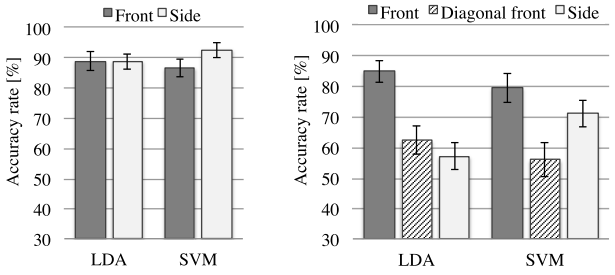
To verify the validity of our gait feature acquisition, we compared the result of head orientation estimation between GEI and other gait features, which are state-of-the-art ones representing dynamic gait information in more detail; frequency domain feature (FDF) [12], gait entropy image (GEnI) [2], masked GEI based on GEnI (MGEI) [*1] [3], gait flow image (GFI) [10] and chrono-gait image (CGI) [21].

Figure 4 shows the accuracy rates of each gait feature when two-class and three-class estimation were conducted. In their accuracy rates, GEI and FDF were higher than the other gait features. From the results, it can be said that dynamic gait information of the other features were less effective for the estimation. In comparison with GEI and FDF, there was little difference in their accuracy. Considering that FDF consists of both the same component as GEI and different ones that represent more dynamic gait information, it is obvious that the GEI component was the most crucial and the other components were redundant information for the head orientation estimation. From the above results, we confirmed that GEI captured sufficient features for the estimation.

[*1] Originally, MGEI was proposed for identification. In Ref. [3], the authors generated MGEI by masking GEI with the regions where both the probe and the gallery GEnI values were over a threshold. By contrast, we generated it by masking GEI with the regions where the GEnI values were over a threshold.

Table 1 Confusion matrix of three-class estimation using GEI.

	Front	Diagonal front	Side
Front	**84.9%**	11.7%	3.5%
Diagonal front	8.4%	**62.3%**	29.3%
Side	6.1%	36.8%	**57.1%**



(a) Two-class estimation (b) Three-class estimation

Fig. 5 Accuracy rates of each classifier. The standard deviations are depicted by error bars.

In comparison between Fig. 4 (a) and Fig. 4 (b), it can be seen that the accuracy rate of "Side" became dramatically worse in three-class estimation. For detailed discussion, the confusion matrix of three-class estimation using GEI is shown in **Table 1**. From this table, it turns out that "Side" was easily misclassified as "Diagonal front" and vice versa. It means that although it was easy to observe gait differences when a pedestrian faced to the side or the front, it was difficult to figure out how a pedestrian faced to the side because of unclear differences.

4.2 Result Comparison with a Non-linear Classifier

In our method, there is a possibility that the accuracy rates could be improved by using a non-linear classifier instead of LDA. Therefore, we employed a Support Vector Machine (SVM) with a radial basis function kernel as a typical non-linear classifier and compared the accuracy rates between LDA and SVM. Because SVM is a two-class classifier, it is impossible to apply SVM to three-class estimation straightforwardly. Our method dealt with it by constructing a one-versus-one classifier.

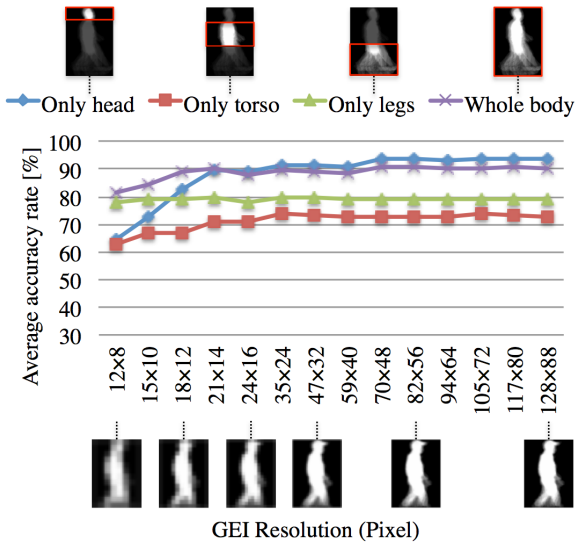
Figure 5 shows the accuracy rates of each classifier when two-class and three-class estimation were conducted. Although the trend of accuracy rates differed between LDA and SVM, there was little difference in terms of the average of their accuracy rates. Considering that the non-linear classifier could not achieve the accuracy improvement, classifier selection is less important for head orientation estimation.

4.3 Contribution of Body Regions in Lower-resolution Cases

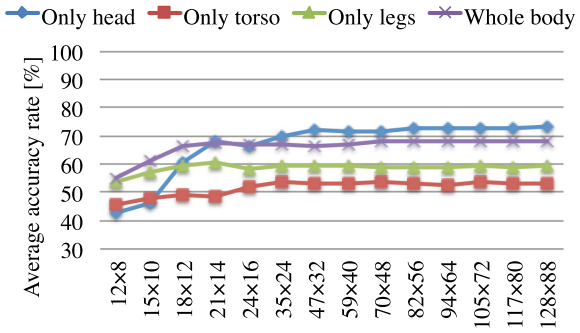
To figure out where body regions become crucial in the estimation while decreasing the GEI resolution, we compared the accuracy rates when some downsampled GEIs of only head, only torso, only leg region were used. Moreover, we verified the effectiveness of using the whole body GEI by comparing its result with that of the above specified regions.

Figure 6 shows the average accuracy rate when the GEI of

© 2014 Information Processing Society of Japan



(a) Two-class estimation



(b) Three-class estimation

Fig. 6 Average accuracy rate of each body region of GEI.

each body region was used. Note that in Fig. 6, the resolution is described as not that of the head region but that of the whole body region. Considering that existing texture-based methods assume the size of a head region as at least 10×10 pixels [4], it would be impossible to apply existing methods to whole body images whose resolution is lower than about 60×40 pixels due to the lack of facial texture information.

When the resolution was relatively large, the head region was the most crucial for the head orientation estimation, as expected. However, when it became smaller than 21×14, its accuracy rates were dramatically decreased because it would become difficult to express the difference of the head shape in the head orientation. In contrast, although the accuracy rate of the leg region was lower than that of the head region, it was more stable. This was because the leg region was larger than the head one and was thus able to express the difference between the head orientation even at quite low-resolutions. When the GEI of the whole body region was used, it could achieve the highest accuracy rate in quite low-resolution images. From the results, we could confirm the effectiveness of using not only the head but also the other body regions in estimating from quite low-resolution images.

5. Conclusions

In this paper, we proposed a novel method to estimate the head orientation of a pedestrian based on his/her gait that can be acquired from low-resolution silhouette images. In the experiments, it was confirmed that the gait change of not only the head region but also the other body regions was effective for head orientation estimation in quite low-resolution images.

In future work, to improve the estimation accuracy, we will consider the fusion of gait features and facial textures, which must be helpful for our estimation method although it is quite difficult to achieve the goal alone. Moreover, we will try to estimate head orientations of pedestrians in a shopping mall as a real environment. It would be possible to understand whether walking shoppers are interested in a shop that is located laterally on a walkway in the mall. Toward achieving head orientation estimation in a real environment, we will extend our proposed method so that it can estimate the head orientation of a pedestrian who has luggage or walks freely.

Acknowledgments This work was partly supported by the JST CREST "Behavior Understanding based on Intention-Gait Model" project.

References

[1] Ali, H., Dargham, J., Ali, C. and Moung, E.G.: Gait Recognition using Gait Energy Image, *International Journal of Signal Processing*, Vol.4, No.3, pp.141–512 (2011).
[2] Bashir, K., Xiang, T. and Gong, S.: Gait Recognition Using Gait Entropy Image, *Proc. 3rd International Conference on Crime Detection and Prevention* (*ICDP 2009*), London, UK, pp.1–6 (2009).
[3] Bashir, K., Xiang, T. and Gong, S.: Gait recognition without subject cooperation, *Pattern Recognition Letters*, Vol.31, No.13, pp.2052–2060 (2010).
[4] Benfold, B. and Reid, I.: Colour Invariant Head Pose Classification in Low Resolution Video, *Proc. British Machine Vision Conference* (*BMVC*) *2008*, pp.49.1–49.10, London, UK (2008).
[5] Gourier, N., Hall, D. and Crowley, J.L.: Estimating Face orientation from Robust Detection of Salient Facial Structures, *Proc. Pointing 2004, ICPR, International Workshop on Visual Observation of Deictic Gestures*, Cambridge, UK (2004).
[6] Guangzhe, Z., Mrutani, T., Kajita, S. and Mase, K.: Video based estimation of pedestrian walking direction for pedestrian protection system, *Journal of Electronics* (*China*), Vol.29, No.1–2, pp.72–81 (2012).
[7] Han, J. and Bhanu, B.: Individual Recognition Using Gait Energy Image, *IEEE Trans. Pattern Analysis and Machine Intelligence* (*PAMI*), Vol.28, No.2, pp.316–322 (2006).
[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] Jung, S.-U. and Nixon, M.: On using gait biometrics to enhance face pose estimation, *Proc. IEEE 4th International Conference on Biometrics: Theory, Applications and Systems* (*BTAS 10*), Washington D.C., USA, pp.1–6 (2010).
[10] Lam, T.H., Cheung, K. and Liu, J.N.: Gait flow image: A silhouette-based gait representation for human identification, *Pattern Recognition*, Vol.44, No.4, pp.973–987 (2011).
[11] Lu, J. and Tan, Y.-P.: Gait-based human age estimation, *IEEE Trans. Information Forensics and Security*, Vol.5, No.4, pp.761–770 (2010).
[12] 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 Conference on Computer Vision* (*ECCV2006*), Vol.3, Graz, Austria, pp.151–163 (2006).
[13] Muñoz-Salinas, R., Yeguas-Bolivar, E., Saffiotti, A. and Medina-Carnicer, R.: Multi-camera head pose estimation, *Machine Vision and Applications*, Vol.23, No.3, pp.479–490 (2012).
[14] Murphy-Chutorian, E. and Trivedi, M.M.: Head Pose Estimation in Computer Vision: A Survey, *IEEE Trans. Pattern Analysis and Machine Intelligence* (*PAMI*), Vol.31, No.4, pp.607–626 (2009).
[15] Okada, T., Yamazoe, H., Mitsugami, I. and Yagi, Y.: Preliminary

© 2014 Information Processing Society of Japan

Analysis of Gait Changes that Correspond to Gaze Directions, *Proc. 2nd IAPR Asian Conference on Pattern Recognition* (*ACPR2013*), Naha, Japan, pp.788–792 (2013).

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

[17] Smith, K., Ba, S.O., Odobez, J.-M. and Gatica-Perez, D.: Tracking the Visual Focus of Attention for a Varying Number of Wandering People, *IEEE Trans. Pattern Analysis and Machine Intelligence* (*PAMI*), Vol.30, No.7, pp.1212–1229 (2008).

[18] Subramanian, R., Yan, Y., Staiano, J., Lanz, O. and Sebe, N.: On the relationship between head pose, social attention and personality prediction for unstructured and dynamic group interactions, *Proc. 15th ACM on International Conference on Multimodal Interaction* (*ICMI2013*), Sydney, Australia, pp.3–10 (2013).

[19] Tosato, D., Farenzena, M., Spera, M., Murino, V. and Cristani, M.: Multi-class Classification on Riemannian Manifolds for Video Surveillance, *Proc. 11th European Conference on Computer Vision* (*ECCV2010*), Crete, Greece, pp.378–391 (2010).

[20] Tosato, D., Spera, M., Cristani, M. and Murino, V.: Characterizing humans on Riemannian manifolds, *IEEE Trans. Pattern Analysis and Machine Intelligence* (*PAMI*), Vol.35, No.8, pp.1972–1984 (2013).

[21] Wang, C., Zhang, J., Pu, J., Yuan, X. and Wang, L.: Chrono-Gait Image: A Novel Temporal Template for Gait Recognition, *Proc. 11th European Conference on Computer Vision* (*ECCV2010*), Crete, Greece, pp.257–270 (2010).

[22] Yu, S., Tan, T., Huang, K., Jia, K. and Wu, X.: A Study on Gait-Based Gender Classification, *IEEE Trans. Image Processing*, Vol.18, No.8, pp.1905–1910 (2009).

[23] Zhang, Z., Hu, Y., Liu, M. and Huang, T.: Head Pose Estimation in Seminar Room using Multi View Face Detectors, *Proc. 1st International Evaluation Conference on Classification of Events, Activities and Relationships* (*CLEAR2006*), Southampton, UK, pp.299–304 (2006).

(Communicated by *Takeshi Oishi*)

© 2014 Information Processing Society of Japan