IPSJ Transactions on Computer Vision and Applications Vol.7 94–98 (July 2015)

[DOI: 10.2197/ipsjtcva.7.94]

Express Paper

Depth-based Gait Authentication for Practical Sensor Settings

Taro Ikeda[1,a) Ikuhisa Mitsugami[1,b)] Yasushi Yagi[1,c)]

Received: March 13, 2015, Accepted: April 20, 2015, Released: July 27, 2015

Abstract: This paper investigates performances of silhouette-based and depth-based gait authentication considering practical sensor settings where sensors are located in an environments afterwards and usually have to be located quite near to people. To realize fair comparison between different sensors and methods, we construct full-body volume of walking people by a multi-camera environment so as to reconstruct virtual silhouette and depth images at arbitrary sensor positions. In addition, we also investigate performances when we have to authenticate between frontal and rear views. Experimental results confirm that the depth-based methods outperform the silhouette-based ones in the realistic situations. We also confirm that by introducing Depth-based Gait Feature, we can authenticate between the frontal and rear views.

Keywords: gait authentication, range sensor, dissimilarity measure

1. Introduction

There are various biometrics, such as fingerprint, vein and iris. These biometrics get attention because they are applicable to criminal investigation or security purposes. Gait, way of walking, is also regarded as one of the biometrics. Although its authentication ability is not high enough to be applied to such serious purposes, gait has an advantage that it can be obtained without any contact to devices; a person can be authenticated only by walking. Considering this great property of gait, we would like to apply gait authentication for automatic security door, visitor logging, and global people tracking, for example. Indeed we would be happy if we could change a door to one that automatically opens when we approach to the door just by putting a sensor around the door. It is commercially useful if we could automatically count the number of visits of a customer just by locating a sensor at an entrance.

Considering this advantage of gait, there are various studies about gait authentication. Among them, the most popular gait feature is Gait Energy Image (GEI) [1], which is an averaged image of the silhouette image sequence corresponding with a walking period. There are also many other silhouette-based studies [2], [3], [4], [5], [6]. In these studies, they report very high authentication accuracy that looks enough for real application, even for the serious applications above mentioned. They, however, have not been introduced in real environments. This is because they require conditions that cannot be fulfilled in the real environments. For example, most silhouette-based methods assume that a walking person is captured from a constant direction. To realize this condition, they have to locate a camera at a distance from the walking way, which is difficult to fulfill in the real environments; we guess in many cases we can put a sensor just at a ceiling or wall near to the walking way. This might have prevented gait authentication techniques from real use, but no study confirms this query objectively and quantitatively.

Currently there are also several studies that use range sensors instead of cameras. Especially after emergence of Microsoft Kinect, the increase of the number of volume-based and depth-based methods is accelerated. We can also find a public dataset [7]. Gait Energy Volume (GEV) [8] is one of range data-based method, which is a simple extension of GEI to 3-D; average volume of the volume sequence of whole body. Sivapalan et al. [9] extracted histogram of oriented gradients (HOG) and local directional patterns (LDP) from the volume sequence to achieve better authentication performance. There are also other depth-based methods [10], [11], [12], [13], [14], [15], [16]. Depth-based Gait Feature (DGF) [17] is also one of the state-of-the-art methods categorized in the depth-based methods, but DGF has a unique advantage that it can separate shape and motion information from gait observation. This property is very effective when we can capture a person from either of his/her front or back, which should often happen in the real situations. However, quantitative evaluation considering such a scenario has not been done.

In this paper, therefore, we evaluate the performances of silhouette-based and depth-based methods considering "realistic" situations. The "realistic" situation is defined, considering the above discussion, as a case where a sensor is located in an environment afterwards and thus has to be near to a walk way, and where the sensor can capture just either of the front or back sides of people. We use GEI and DGF as representatives of the silhouette-based and depth-based methods, respectively. To realize fair comparison between different sensors, we construct full-

1 Osaka University, Ibaraki, Osaka 567–0047, Japan
a) ikeda@am.sanken.osaka-u.ac.jp
b) mitsugami@am.sanken.osaka-u.ac.jp
c) yagi@am.sanken.osaka-u.ac.jp

body volume of walking people by multi-camera environment so as to reconstruct virtual silhouette and depth images at arbitrary sensor positions. In addition, we also investigate performances when we have to authenticate between frontal and rear views. Experimental results show that in the realistic situations the depth-based methods outperform the silhouette-based ones. It is also an important discussion that by introducing DGF, we can authenticate even between the frontal and rear views. These results confirm that DGF is most suitable when we would like to construct a gait authentication system under "realistic" situations.

2. Depth-based Gait Feature

DGF was originally proposed by Nakajima et al. [17]. This section briefly explains how to obtain DGF from observation.

From captured depth images, a walking person is detected by background subtraction. A virtual range sensor is located in his/her front at a certain distance so that we obtain an image sequence where his/her position is normalized. This sequence is called a Gait Depth-map Sequence (GDS). Considering the periodicity of walking, we then estimate a period N and extract frames corresponding with a cycle from GDS by evaluating its autocorrelation. We then apply Discrete Fourier Transformation (DFT) to the extracted frames.

$$G(x, y, k) = \sum_{n=0}^{N-1} g(x, y, n)e^{-j\omega_0 kn}, \quad (1)$$

where $g(x, y, n)$ denotes a depth value at a pixel (x, y) in the n-th frame, ω_0 is a base angular frequency for the gait period N, and $G(x, y, k)$ is the DFT of GDS for k-times the gait period. Note that this operation is performed only for pixels that always correspond with foreground region. Pixels that are always or sometimes correspond with background region are masked so as not to consider them in dissimilarity measure. From $G(x, y, k)$, an amplitude spectrum $A(x, y, k)$ is calculated as follows:

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

In the original DGF paper [17], they picked up only $A(x, y, 0)$ and $A(x, y, 1)$ since $A(x, y, k)$ $(k \geq 2)$ are expected to be noisy and less reliable considering the number of frames included in a cycle and noise of each range sensor. They also calculate the phase information for the basic frequency component $G(x, y, 0)$, but they reported that the phase component is unstable and so not effective for authentication task. Considering this report, we use $A(x, y, 0)$ and $A(x, y, 1)$ in this paper. **Figure 1** shows examples of $A(x, y, 0)$ and $A(x, y, 1)$. $A(x, y, 0)$, which we call the direct component, mainly describes the shape of the person. On the other hand, $A(x, y, 1)$, which is called the amplitude component, corresponds with the fundamental motion of walking such as arm swings and steps that occur only once in a period, so that pixel values around the arms and legs are higher than ones around the trunk.

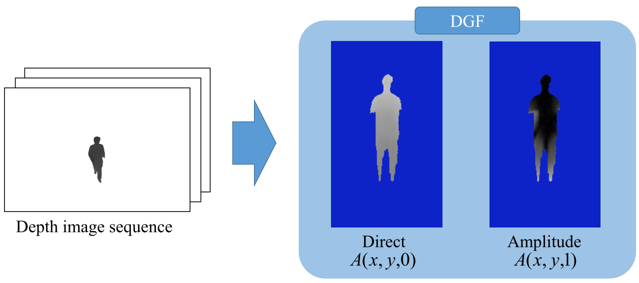

Fig. 1 Depth-based Gait Feature.

3. Scene Generation

To correctly compare between authentication performances of different feature descriptions or different view angles, we need to use the same scene where the same person walks in the same way. We, therefore, adopt a strategy that we firstly construct full-body volume of the person and reconstruct silhouette and depth images from the volume. The full-body volume is captured in the treadmill environment shown in **Fig. 2**. This environment consists of 25 cameras and treadmill. There is a treadmill in the center and it is surrounded by 25 cameras. **Figure 3** shows examples of the full-body volume data and the depth image generated from the volume.

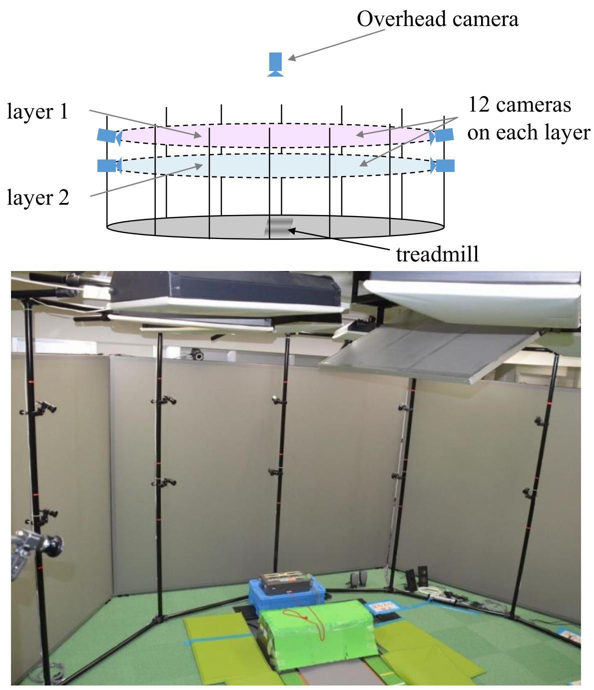

Fig. 2 Treadmill environment. It consists of 25 cameras surrounding the treadmill.

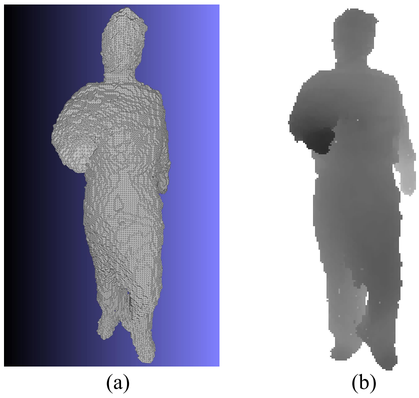

Fig. 3 3-D person data. (a) Full-body volume obtained by the treadmill environment. (b) Depth image generated from the full-body volume.

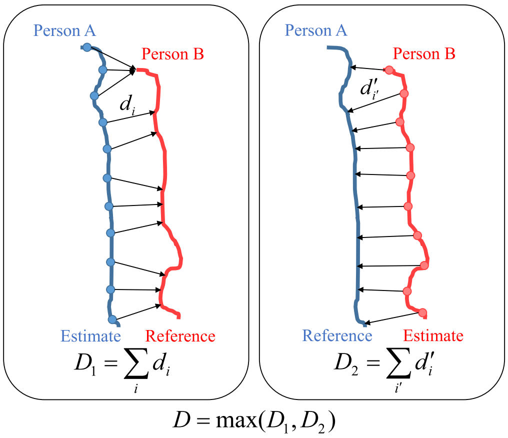

Fig. 4 Dissimilarity measure for masked DGF.

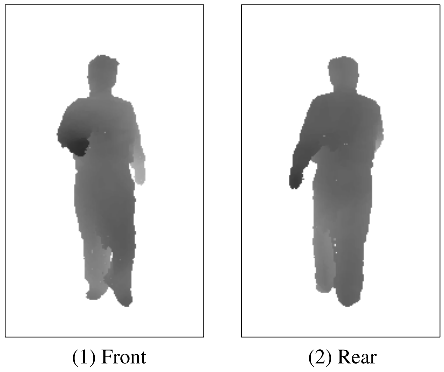

Fig. 5 Front and rear images.

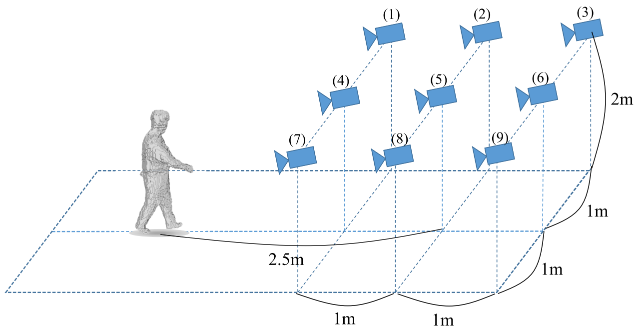

Fig. 6 Virtual range sensor locations.

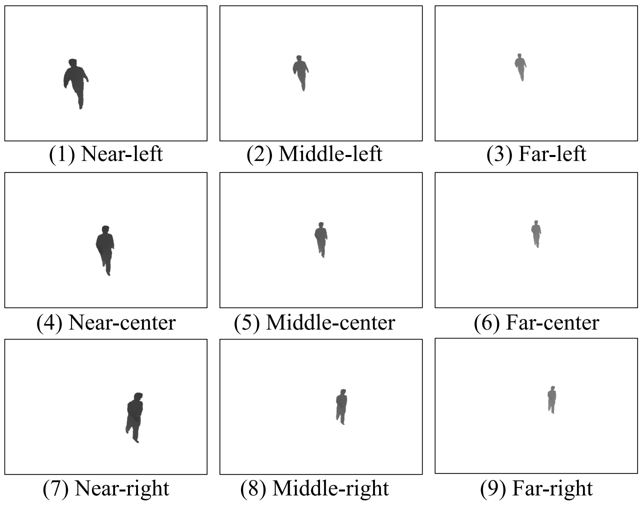

Fig. 7 Virtually captured depth images of a walking person. By binarizing them, we can also obtain virtually captured silhouettes.

4. Dissimilarity Measure

In evaluating performance of authentication, not only the feature description but also choice of dissimilarity (or similarity) measure is important. In most studies related to authentication, they calculate difference of two feature vectors and evaluate its L2/L1/L0 norms as the dissimilarity measure. Nakajima et al. [17] also simply use L2 norm of feature vectors they proposed. When we compare two depth images with mask regions, however, we cannot simply use this way. Let us consider a case that we would like to measure dissimilarity of the direct component of two different people, for example. One way is to simply calculate a difference of two images without considering the masks. In this case, however, the difference of the number of foreground pixels should be dominant and the difference of depth value of each pixel affects very little. Another way is to consider the masks and sum up just the pixel value differences of the overlapped region. This dissimilarity measure, however, tends to be smaller when the overlapped region is smaller, which means the shapes of two people are so different though.

In this paper, we adopts another dissimilarity measure that is used for evaluating accuracy of 3-D shape reconstruction considering that DGF can also be regarded as a kind of 3-D surface of a person. **Figure 4** shows this method. In this method, for a pair of two people's DGF, we first regard DGF of a person as a reference and that of the other person as an estimate. To evaluate the accuracy of the estimate, for each point on the estimate surface its nearest point on the reference is selected, and a distance d_i between the points are summed up among all the estimate surface to obtain a measurement D_1. We then exchange the reference and estimate and sum up the distance to get D_2. By this operation, we get two measurements and choose the larger one as the final dissimilarity measure:

$$D = \max(D_1, D_2). \tag{3}$$

5. Experiment and Discussion

5.1 Experimental Settings

The main purpose of this paper is to evaluate the performances of GEI and DGF in realistic situations, where a camera or range sensor have to be located quire near to a walking person so that relative positions between a sensor and a person cannot be regarded to be constant; locations of people in captured images are different, moreover even for a certain person, his/her appearance gradually changes as he/she walks. In addition, it often happens that a person can be captured only from either of the front and rear, so that we might have to authenticate observations from opposite directions as shown in **Fig. 5**.

Considering this purpose, we virtually locate cameras and range sensors at nine locations as shown in **Fig. 6**. The corresponding captured images are shown in **Fig. 7**. These sensors are fixed not in the person coordinate but in the environment coordinate, so that his/her positions in images gradually moves according to his/her walking. Moreover, we locate virtual sensors at the back side of the person as well as the front side for performance

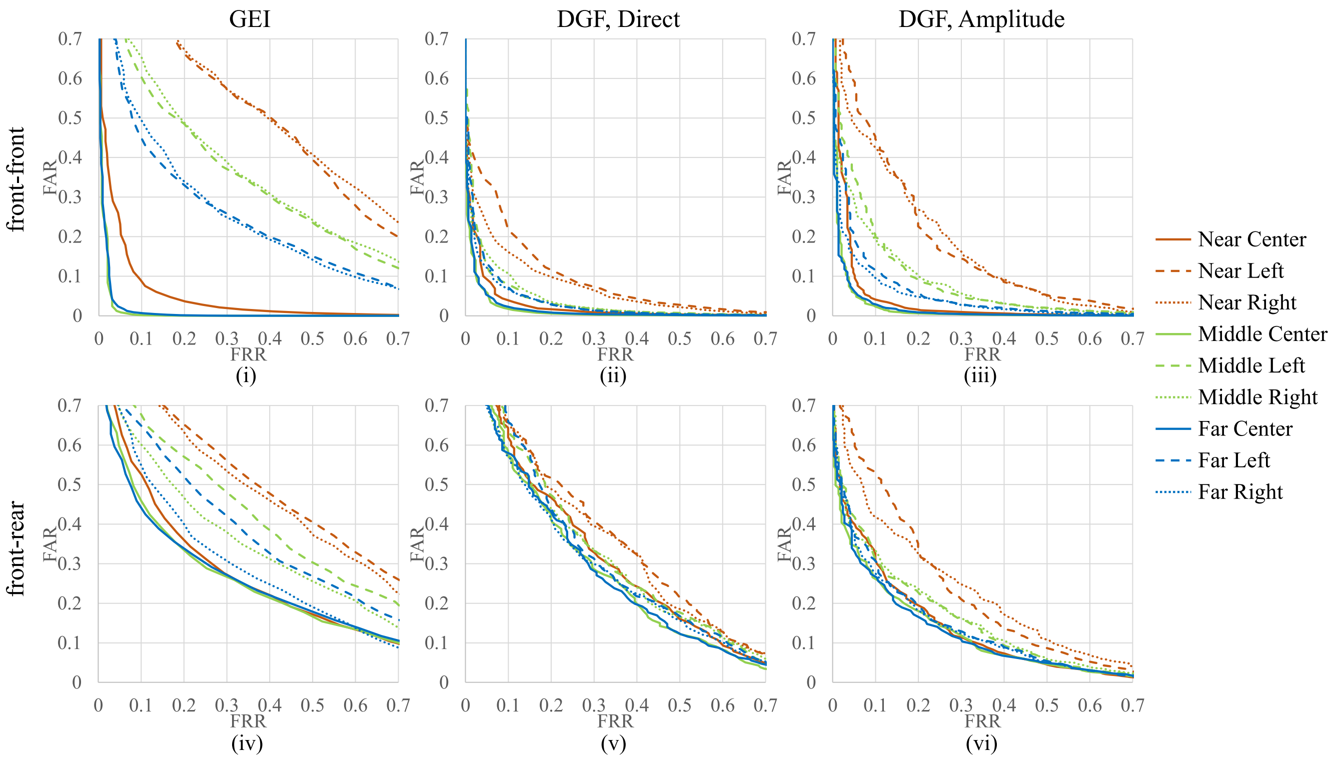

Fig. 8 ROC curves of GEI, direct and amplitude components of DGF.

evaluation in the front-rear cases.

For performance evaluation, we use 97 subjects. There are three sequences for each subjects. Their DGF obtained from the sensor position (5) (called "middle-center") is used as galleries, those from all the positions (1)-(9) (including (5)) are used as probes.

5.2 Experimental Results

For comparative evaluation, we employed the ROC curve which indicates the trade-off between the false rejection rate (FRR) of the genuine and the false acceptance rate (FAR) of the imposter when an acceptance threshold changes. **Figure 8** summarizes all experimental results. Three columns correspond with results of GEI, the direct and amplitude components of DGF, respectively. The results in the first row are of the front-front cases, and those of the second row are of the front-rear cases.

5.2.1 Front-front Authentication

According to (i), these nine curves are so different. When the sensor positions are the same or very similar between the galleries and probes, performances are fine. In the other sensor locations, however, performances are terribly degraded. Especially in "near-right/left" cases, they are almost chance rate. This result indicates that GEI is not effective for authentication task in realistic situations where a walking person has to be captured by a nearly-located sensors. GEI, indeed, has been known as a good gait feature, but it can be said only in a well-controlled environment where people can be observed from a distant and from a constant direction.

On the other hand, the variation of the curves in (ii) is much smaller than in (i). Although the performances in "middle-center" and "far-center" are slightly worse than those in (i), the others are apparently better. In (iii), each curve is slightly worse than that in (i). This fact says that the motion information is less effective than the shape information, which is consistent with discussion of Nakajima et al. [17]. Note that, however, the amplitude component still gives better performance than (i). From these results, DGF works much better than GEI in the realistic situations, and especially the direct component of DGF, which denotes shape of person, is most effective for authentication task.

5.2.2 Front-rear Authentication

We also evaluate their performances in "front-rear" cases where we have to authenticate between frontal and rear views. Considering that GEI is based on silhouettes that should be the same (mirrored, strictly speaking) between the frontal and rear views, performance of GEI was expected to be fine. The result is, however, very bad as shown in (iv). This is because the camera angle is not horizontal but has a certain elevation. Though the result would get better if we could put a camera horizontally, it is often impossible to fulfill this camera angle condition.

In (v), all the results are so bad, but it is just natural. Since the direct component encodes a person's shape, this operation means comparison between frontal and rear surface of a person, which must not match, of course. In (vi), on the other hand, the results are much better than those of GEI and the direct component. It is considered because the dominant motions in walking (arm and leg swings) are all anteroposterior motions, which can be well encoded by the amplitude component of DGF and are expected to be similar between the frontal and rear views.

5.3 Discussion

The results argue that when we would like to construct an authentication system in real environment where we often have locate a sensor quire near to a person we should not choose a camera but a range sensor. In addition, when we use DGF for gait feature description, which can separate the shape and the motion information, we can authenticate even between the frontal and rear views by focusing only on the motion information.

6. Conclusion

This paper investigated performances of silhouette-based and depth-based gait authentication methods considering practical sensor settings where sensors for authentication are located in the environments afterwards so that they usually have to be located quite near to people. To realize a fair comparison between different sensors, we constructed the full-body volume of walking people by a multi-camera environment so as to reconstruct virtual silhouette and depth images at arbitrary sensor positions. In addition, we also investigated performances when we have to authenticate between frontal and rear views. Experimental results confirmed that in the realistic situations the depth-based methods outperform the silhouette-based ones. We also confirmed that by introducing DGF, we can authenticate even between the frontal and rear views.

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

References

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

[2] Lynnerup, N. and Vedel, J.: Person identification by gait analysis and photogrammetry, *Journal of Forensic Science*, Vol.50, No.1, pp.112–118 (2005).

[3] 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*, pp.151–163 (2006).

[4] 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).

[5] Lam, T.H.W., Cheung, K.H. and Liu, J.N.K.: Gait flow image: A silhouette-based gait representation for human identification, *Pattern Recognition*, Vol.44, pp.973–987 (2011).

[6] 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).

[7] Hofmann, M., Geiger, J., Bachmann, S., Schuller, B. and Rigoll, G.: The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits, *Journal of Visual Communication and Image Representation*, Vol.25, No.1, pp.195–206 (2014).

[8] Sivapalan, S., Chen, D., Denman, S., Sridharan, S. and Fookes, C.: Gait Energy Volumes and Frontal Gait Recognition using Depth Images, *Proc. International Joint Conference on Biometrics*, pp.1–6 (2011).

[9] Sivapalan, S., Chen, D., Denman, S., Sridharan, S. and Fookes, C.: Histogram of Weighted Local Directions for Gait Recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.125–130 (2013).

[10] Sivapalan, S., Chen, D., Denman, S., Sridharan, S. and Fookes, C.: The Backfilled GEI — A Cross-Capture Modality Gait Feature for Frontal and Side-View Gait Recognition, *Proc. International Conference on Digital Image Computing Techniques and Applications*, pp.1–8 (2012).

[11] Igual, L., Lapedriza, À. and Borràs, R.: Robust gait-based gender classification using depth cameras, *EURASIP Journal on Image and Video Processing*, Vol.2013:1 (2013).

[12] Chattopadhyay, P., Roy, A., Sural, S. and Mukhopadhyay, J.: Pose Depth Volume extraction from RGB-D streams fro frontal gait recognition, *Journal of Visual Communication and Image Representation*, Vol.25, pp.53–63 (2014).

[13] Chattopadhyay, P., Sural, S. and Mukherjee, J.: Exploiting Pose Information for Gait Recognition from Depth Streams, *Proc. 4th IEEE Workshop on Consumer Depth Cameras for Computer Vision*, pp.341–355 (2014).

[14] Tang, J., Luo, J., Tjahjadi, T. and Gao, Y.: 2.5D Multi-View Gait Recognition Based on Point Cloud Registration, *Sensors*, Vol.14, pp.6124–6143 (2014).

[15] Afendi, T., Kurugollu, F., Crookes, D. and Bouridane, A.: A frontal view gait recognition based on 3D imaging using a time of flight camera, *Proc. 22nd European Signal Processing Conference*, pp.2435–2439 (2014).

[16] Procházka, A., Vyšata, O., Valis, M., Ťupa, O., Schätz, M. and Mařík, V.: The MS kinect image and depth sensors use for gait features detection, *Proc. IEEE International Conference on Image Processing*, pp.2271–2274 (2014).

[17] Nakajima, H., Mitsugami, I. and Yagi, Y.: Depth-Based Gait Feature Representation, *IPSJ Trans. Computer Vision and Applications*, Vol.5, pp.94–98 (2013).

(Communicated by *Takeshi Masuda*)