PAPER

Temporal and spatial analysis of local body sway movements for the identification of people

Takuya KAMITANI[†], **Hiroki YOSHIMURA**[†], **Masashi NISHIYAMA**[†,††a)], ***and*** **Yoshio IWAI**[†,††],

SUMMARY We propose a method for accurately identifying people using temporal and spatial changes in local movements measured from video sequences of body sway. Existing methods identify people using gait features that mainly represent the large swinging of the limbs. The use of gait features introduces a problem in that the identification performance decreases when people stop walking and maintain an upright posture. To extract informative features, our method measures small swings of the body, referred to as body sway. We extract the power spectral density as a feature from local body sway movements by dividing the body into regions. To evaluate the identification performance using our method, we collected three original video datasets of body sway sequences. The first dataset contained a large number of participants in an upright posture. The second dataset included variation over the long term. The third dataset represented body sway in different postures. The results on the datasets confirmed that our method using local movements measured from body sway can extract informative features for identification.

key words: *Body Sway, Identification, Local Movements, Soft Biometrics*

1. Introduction

The progressively widespread use of video surveillance cameras has led to the development of convenient and non-intrusive biometric authentication systems [1]–[3]. Identification using video from surveillance cameras is a key technology for developing various authentication systems. To obtain high identification performance, it is important to design methods of extracting informative features from video sequences. Recently, soft biometrics [4]–[6] that represent human attributes have been an active topic in pattern recognition research in terms of extracting informative features for identification. Human attributes can be split intuitively into three types: physical characteristics [7], [8] (e.g., gender and age), adhered human characteristics [9], [10] (e.g., clothing and belongings), and behavioral characteristics [11], [12] (e.g., gestures and gait). With respect to person re-identification, researchers [13]–[15] have reported that combining these characteristics increases the identification performance. In particular, behavioral characteristics have the advantage that individuals' movements differ and can thus be used to identify people even when attributes such as their gender, age, and clothing are the same. For instance, there are situations where many of the office workers in a building wear a suit or workers in a factory wear a uniform.

Manuscript received May 21, 2018.
Manuscript revised August 20, 2018.

[†]Graduate School of Engineering Department of Information and Electronics, Tottori University, Tottori 680-8550, Japan
[††]Cross-informatics Research Center, Tottori University
a) E-mail: nishiyama@tottori-u.ac.jp

DOI: 10.1587/transinf.E102.D.1

In this paper, we focus on how to design a method of extracting and using behavioral characteristics. Existing methods [11], [12] that use behavioral characteristics generally exploit gait features acquired from video sequences. Gait features represent periodic movements of body parts such as limbs. However, people frequently stop walking and maintain the same posture, for example, while waiting for a security gate to open. Thus, gait features do not sufficiently represent behavioral characteristics when people have stopped walking because periodic movements of the body parts do not always occur and thus are not informative for identification. Indeed, gait features sometimes cause a decrease in identification performance when people stop walking.

We consider an authentication system that requires people to maintain an upright posture for several tens of seconds. When people maintain their posture, their bodies do not remain completely still but slightly and continuously move in all directions. This body movement occurs naturally to maintain a person's posture and is called body sway. Note that we consider an upright posture as a specific example of the posture of a person who has stopped walking. In the field of medical science, many researchers [16]–[20] have attempted to measure the center of gravity of body sway using force plates embedded in the floor. These methods are not intended for identification purposes, but can be used to classify gender and age [16], [17], people with lower-back pain [18], women with morning sickness [19] and patients with neuropathy [20]. We thus assume that body sway contains information about the identity of people, and is a behavioral characteristic that can be used as a human attribute in soft biometrics. In this study, we tackle the challenging task of extracting an informative feature for identification from video sequences of body sway. In addition, body sway has the advantage that an overhead camera can be used to observe people passively because body sway movement can be measured from the upper half of the body. Hence, it is not necessary to locate a camera to the side of the person, which is commonly required for extracting gait features [11], [12]. The use of an overhead camera avoids the occlusion of people when the number of people increases.

To this end, we propose a method of identifying people using video sequences of body sway acquired from people in an upright posture. Our method computes the center of body sway from a video sequence, and spatially divides the body into small local regions using the center of body sway. It then measures the temporal and spatial changes in local

movements in these regions and conducts a frequency analysis for feature extraction. The main purpose of this study is to confirm whether the body sway observed when people maintain an upright posture can be used for identification. We collected a novel dataset of body sway for 118 participants. The experimental results showed that identification performance improved from 63.5% using the gait feature of an existing method to 88.8% using our local movement feature. Furthermore, we demonstrated that our method can extract informative features even when there is variation over time or variation in posture.

This paper is an extended version of our previous research [21]. Here, we provide improved results through further investigation and an extended experimental evaluation of our method. The remainder of the paper is organized as follows. Section 2 describes our method of extracting features for body sway, Section 3 presents the identification performance when using body sway with upright postures, Section 4 shows the influence of the variation of different postures, and Section 5 presents our concluding remarks.

2. Design of features in terms of body sway

2.1 Overview

We consider the informative features extracted from a video sequence of body sway acquired from people in an upright posture. The first feature is the temporal and spatial swinging movement of the body. The movement contains information about identity differences, including gender, age, chronic disease, how muscles are attached, and the sense of balance. The second feature represents the body shape, such as whether the person is obese or thin. The third feature represents the body posture, such as a stooping or slouching posture.

The features of existing gait recognition methods [11], [12] mainly represent the body shape together with the temporal swinging movements, and action recognition methods [11], [22] are similar. Researchers [23], [24] exploited temporal movements as features for gaze authentication. The existing methods have been designed to represent the features of gait, action or gaze. In preliminary experiments, we used the existing methods to extract features from video sequences of body sway. However, we could not obtain high identification performance using these methods.

We focus on how to represent identity using temporal and spatial changes in movements due to body sway. Figure 1 provides an overview of our method. We divide the body into small local regions to represent the spatial movements of body sway. We measure temporal changes in local movements for each local region and compute a query feature from measured temporal changes. We store target features in an authentication system in advance. Our method computes the distance between query and target features using a metric learning technique [25]. Finally, the distance is transformed into the likelihood of the query and target features belonging to the same person using the technique described in [26].

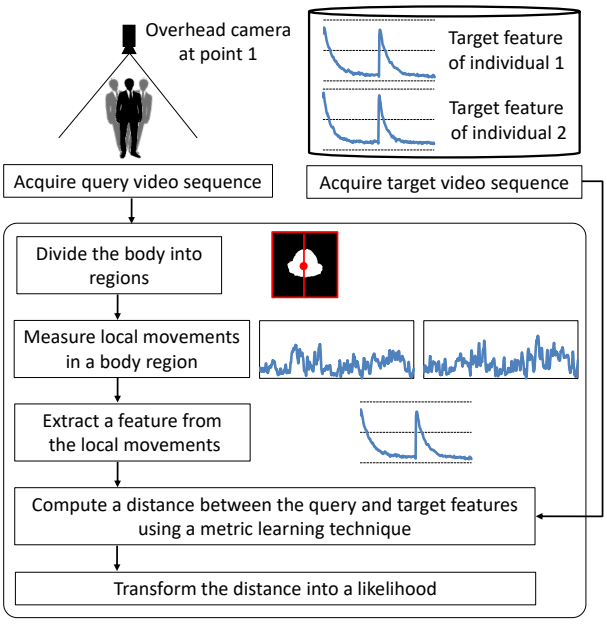


Fig. 1 Overview of our method.

The details of our method are described below.

2.2 Measuring temporal and spatial changes in local movements

We describe a method of measuring temporal and spatial changes in local movements from a video sequence of body sway. Movements of whole body occur around a central position. An existing method [27] measures the movements using the body regions, under the assumption that all body parts move synchronously in the same direction. Although the method considers the temporal changes in movements, it ignores the spatial changes. We thus extend the method to represent spatial changes in movement and extract informative features measured from body sway.

From a frame of the video sequence at time $t \in 1, ..., T$, we compute a mask image $\boldsymbol{m}_t$ in which a pixel takes a value of 1 if it is within a body region and 0 otherwise. The original method [27] uses algorithm 1 to infer the reference time r, which represents the temporal center of swings, from the whole body region in the mask image.

Algorithm 1 Determining reference time

Input: Mask images $\{\boldsymbol{m}_t | t \in 1, ..., T\}$
Output: Reference time r
1: **for** $\tilde{r} = 1$ to T **do**
2: Initialize $D_{\tilde{r}} \leftarrow 0$
3: **for** $t = 1$ to T **do**
4: compute $\tilde{d} = \|\boldsymbol{m}_{\tilde{r}} - \boldsymbol{m}_t\|_1$
5: $D_{\tilde{r}} \leftarrow D_{\tilde{r}} + \tilde{d}$
6: **end for**
7: **end for**
8: $r \leftarrow \arg\min D_{\tilde{r}}$

To consider the spatial change in movement, our method

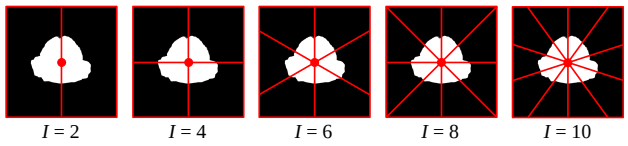


Fig. 2 Examples of local regions radially divided from the body region. I is the number of local regions.

divides the body region into numerous local regions and computes the local movement in each region. The simplest method is to divide the body region into a lattice. However, we cannot stably measure movements around the center of the body region when the lattice cells are small. Therefore, we divide the body radially into local regions using the position of the center of the body, as illustrated in Figure 2. We compute the center position using the body region of the mask image $\boldsymbol{m}_r$ acquired at reference time r. Note that we assume that the center is in the same position at all times $t \in 1, ..., T$. We measure the temporal changes in local movements from the spatially divided local regions using Algorithm 2. We aim to represent the spatial changes in movement in more detail by increasing the number of divisions. The local movement $d_{i,t}$ in a local region $i \in 1, ..., I$ is computed as

$$d_{i,t} = \sum_{\boldsymbol{x} \in \text{region}(i)} \|\boldsymbol{m}_r(\boldsymbol{x}) - \boldsymbol{m}_t(\boldsymbol{x})\|_1 \quad (1)$$

where $\boldsymbol{m}_r(\boldsymbol{x})$ and $\boldsymbol{m}_t(\boldsymbol{x})$ are the binary pixel values at point $\boldsymbol{x}$ in each mask image, and region(i) consists of the pixels in the i-th local region. We can use the $\boldsymbol{L}_1$-norm to count the number of positive pixels because the mask images are binary.

Algorithm 2 Computing local movement

Input: Reference time r, mask images $\{\boldsymbol{m}_t | t \in 1, ..., T\}$, the length of the video sequence T, the number of local regions I
Output: The local movement $\{d_{i,t} | t \in 1, ..., T, i \in 1, ..., I\}$
1: compute the position of the center of body region in $\boldsymbol{m}_r$
2: **for** $i = 1$ to I **do**
3: set the i-th region using the computed center
4: **for** $t = 1$ to T **do**
5: compute $d_{i,t}$ using Equation (1)
6: **end for**
7: **end for**

2.3 Extracting the feature for identification

We describe a method of extracting the feature for identification from the temporal and spatial changes in local movements. The identification performance decreases when directly using the changes in local movements because the direction of body sway varies randomly. We thus need to consider a feature that is invariant to the randomness of movements.

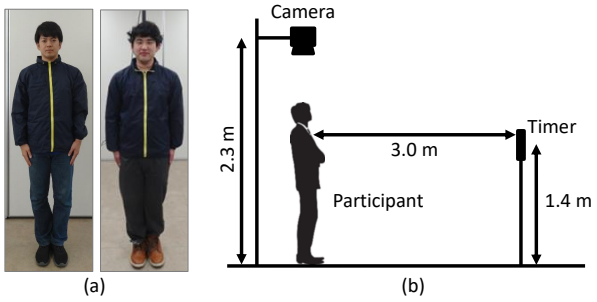


Fig. 3 Setup for acquiring video sequences of body sway.

In the signal processing field, frequency analysis techniques are widely used to extract informative features from time series signals. Because the changes in local movements are also time series signals, we assume that the frequency analysis techniques are adequate for achieving high performance. We assume that the phase components are shifted each time when we measure the local movements. To alleviate the randomness of swings, we do not use the phase components.

Our method estimates the power spectral density (PSD) from the local movements $d_{i,t}$ using Welch's method [28], and extract the feature $\boldsymbol{f}$ for identification using Algorithm 3. To compute the PSD, we divide the local movements into small segments by convoluting a Hann window. We denote the length of each segment L. If L takes a large value, the frequency resolution increases. We believe that features can capture the details of movements when using a large L when there is no influence of noise. However, the appropriate value of L needs to be chosen experimentally because we cannot ignore noise. We use all of the values from the DC component to the $L/2$-th component computed by the PSD for identification. The dimension of $\boldsymbol{f}_i$ is $L/2$. The feature for identification is represented as $\boldsymbol{f} = [\boldsymbol{f}_1^{\mathrm{T}}, ..., \boldsymbol{f}_I^{\mathrm{T}}]^{\mathrm{T}}$. The dimension of $\boldsymbol{f}$ is $IL/2$. We expect $\boldsymbol{f}$ to represent the identify of a person while alleviating the random swings in the temporal and spatial changes of local movements.

Algorithm 3 Extracting the feature using local movements

Input: The local movement $\{d_{i,t} | t \in 1, ..., T, i \in 1, ..., I\}$, the length of the video sequence T, the number of local regions I
Output: The feature $\boldsymbol{f}$
1: **for** $i = 1$ to I **do**
2: compute the PSD with L from $\{d_{i,t} | t \in 1, ..., T\}$
3: compute a value by taking the logarithm of the PSD for each frequency
4: set $\boldsymbol{f}_i$ using the values of all frequencies
5: **end for**
6: concatenate $\{\boldsymbol{f}_i | i \in 1, ..., I\}$ to $\boldsymbol{f}$

3. Experiments with upright postures

3.1 Dataset of video sequences of body sway

We evaluated whether the features extracted from the video

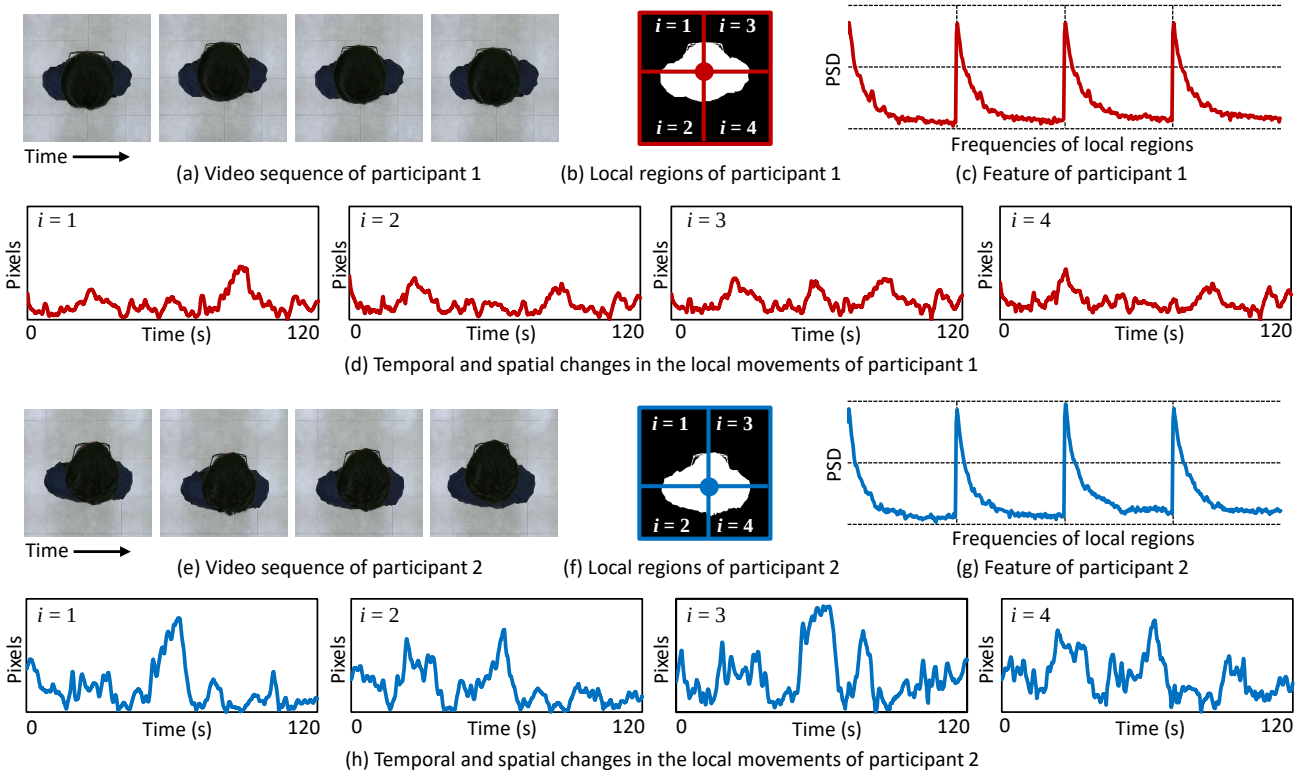


Fig. 4 Examples of temporal and spatial changes in local movements measured from the video sequences of two participants.

sequences of body sway contained identities. We collected video sequences of body sway from 118 participants (average age 22.1 ± 4.3 years; 83 males and 35 females). Each participant maintained an upright posture while standing with their heels aligned as shown in Figure 3 (a). We asked all participants to wear the same dark-blue nylon outerwear, similar to a uniform worn by factory workers. We set an overhead camera at a height of 2.3 m. We applied a camera calibration technique such that the optical axis coincided with the normal direction of the floor. Each participant stood under the camera as shown in Figure 3 (b). A marker was set to indicate the position of the participant's heel in the standing position. We asked each participant to look at a timer placed 3 m away. We displayed the time lapse on the timer. We used video sequences comprising images of 1920 × 1080 pixels captured at 30 fps by a Microsoft Kinect V2. The highest PSD frequency was 15 Hz. The time length of a video sequence was 120 s, and the number of sampled movements was $T = 120 \times 30 = 3600$ for each local region. We used a fixed 1000×1000 bounding box to measure local movements. We observed each participant three times. The participant sat and rested between each sequence. To generate the mask images of body regions, we applied a background subtraction technique using images without participants.

3.2 Evaluation of the parameters of our method

Figure 4 shows examples of our local movement features for two participants. The acquired video sequences are shown in (a) and (e), and the four local regions are presented in (b) and (f). The features in (c) and (g) were extracted from the temporal and spatial changes of the local movements in (d) and (h). The features differed between participants while they maintained an upright posture even though the video sequences appeared to be almost the same.

We evaluated the identification performance while changing the number of local regions I, the length of the video sequence T and the length of each segment L, separately. Our method used a metric learning technique, the large margin nearest neighbor (LMNN) method [25]. We randomly selected 59 participants from the dataset described in Section 3.1. The remaining 59 participants were used to train a metric matrix for LMNN. We repeated the random selection five times to generate different test sample sets. In each set, we tested ${}_3P_2 = 6$ times for each participant by selecting a single video sequence as a target and a single video sequence as a query. We used the first matching rate for the identification performance. We used a nearest-neighbor algorithm for identification.

Figure 5 shows the mean and standard deviation of the correct matching rate when a certain parameter was fixed and other parameters were changed. The best identification performance was 88.8 ± 3.8% using $I = 25$, $L = 256$ and $T = 3600$. Figure 5 (a) shows that the identification performance was improved by increasing the number of lo-

cal regions I. When the number of local regions exceeded 15, the identification performance was almost constant. Figure 5 (b) shows that the identification performance was stable when the length of each segment L was set between 64 and 512. Figure 5 (c) shows the identification performance when the time length of the video sequences was less than 120 s ($T = 3600$). When the length of the video was reduced to $3/4$ or $1/2$, the performance was reduced by 3.6 and 10.7 points, respectively. We believe that the degradation in performance caused by using short video is related to the noise in the body sway during observation. Noise was reduced by observing body sway over a longer duration and identification performance was improved. A short time length is preferable for the development of practical applications. We will conduct further experiments to reduce the time length in future work.

3.3 Comparison with features extracted using existing methods

We compared the identification performance between the local movement features obtained using our method with those obtained using existing methods.

- **LM** (Local Movements): We computed a feature using our method with the parameters set as $I = 25$, $L = 256$, $T = 3600$.
- **GEI** (Gait Energy Image) [11]: We assumed a walking cycle T. We computed a feature by averaging the mask images as $\Sigma_{t=1}^{T} \boldsymbol{m}_t / T$. Figure 6 (a) shows an example of a GEI.
- **MHI** (Motion History Image) [22]: We assigned a weight $\tau = t/T$ for each time at a position where movement was generated. We added the temporal weights for each position. Figure 6 (b) shows an example of an MHI.
- **MEI** (Motion Energy Image) [22]: We set the positions where movements were generated as $\cup_{t=2}^{T} |\boldsymbol{m}_t - \boldsymbol{m}_{t-1}|$. Figure 6 (c) shows an example of an MEI.
- **C** (Cepstrum) [23]: We applied cepstrum analysis to the temporal change in local movements. We used frequencies from the DC component to the 1100-th component for a feature.
- **MFCC** (Mel-frequency Cepstrum Coefficients) [24]: We computed a feature using 40 coefficients.

The same experimental conditions were used for the query and target sequences as described in Section 3.2. Note that C and MFCC were computed from 25 local regions in our method. We set the walking cycle for GEI to 3600 because we obtained the best identification performance with this value, as shown in Figure 7.

Figure 8 compares the identification performance achieved using features extracted using our method and existing methods; the figure shows that LM outperformed GEI, MHI, and MEI. We believe that the performances of the other methods were lower because they cannot represent small movements of the body: GEI was designed for gait recognition, which involves large limb movements, while MHI and MEI were designed for action recognition with dynamic movement of the body. The performances of MHI and MEI

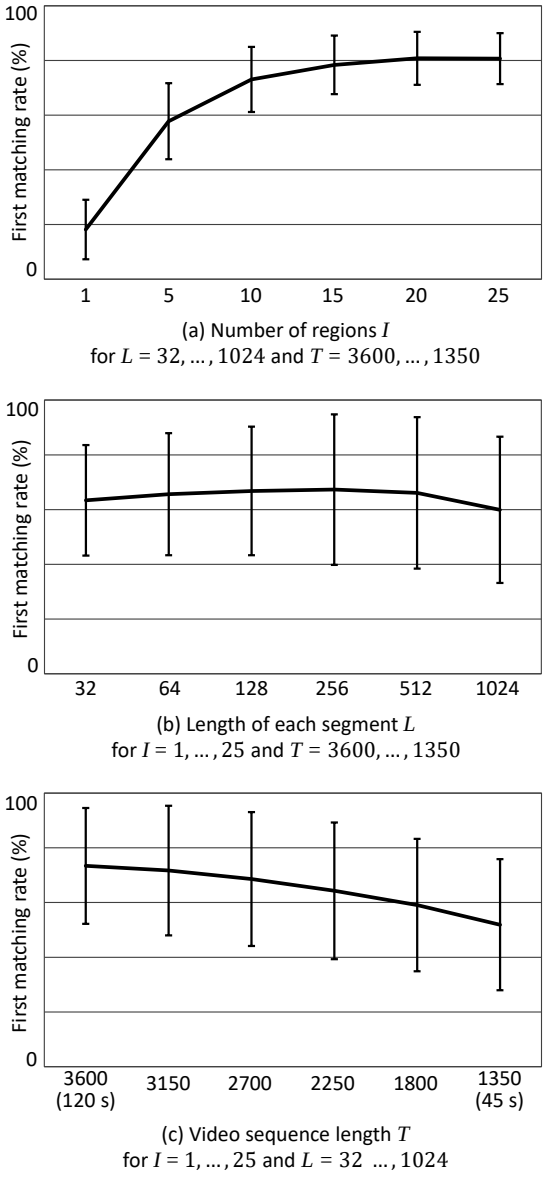


Fig. 5 Identification performance when changing the parameters of our method: (a) number of local regions, (b) length of each segment, (c) video sequence length.

Fig. 6 Examples of GEI, MHI and MEI computed from the video sequence of the same participant.

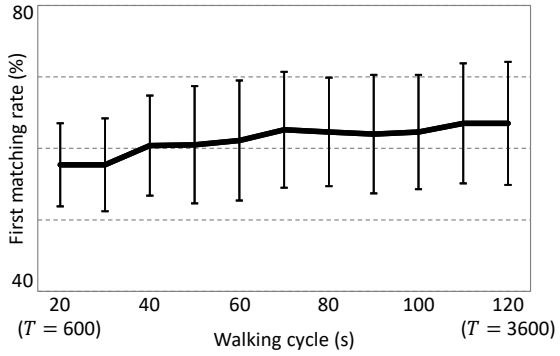


Fig. 7 Identification performance of GEI while changing the walking cycle parameter.

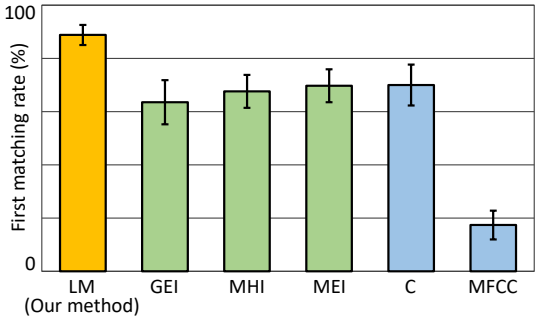


Fig. 8 Comparison of the first matching rate achieved with the local movement feature obtained using our method and those obtained using existing methods.

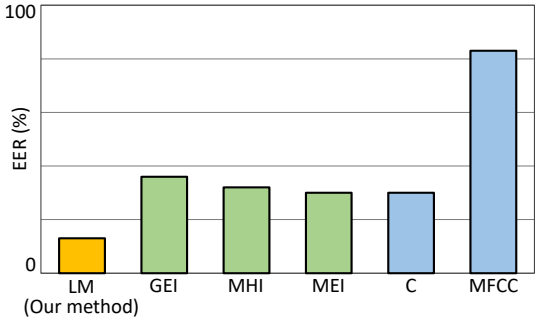


Fig. 9 Comparison of the EER achieved with the local movement feature obtained using our method and those obtained using existing methods.

were almost equivalent, while the performance of GEI was lower. We see that LM outperformed C and MFCC. Furthermore, we evaluated equal error rate (EER), which indicates the point at which the false acceptance rate is equal to the false rejection rate. Figure 9 shows the EER of LM, GEI, MHI, MEI, C, and MFCC. We again see that our method significantly improved identification performance compared with all other methods. We believe that the lower performances of C and MFCC were because these methods were designed for gaze authentication with abrupt and rapid movements of the eyes, whereas body sway is characterized

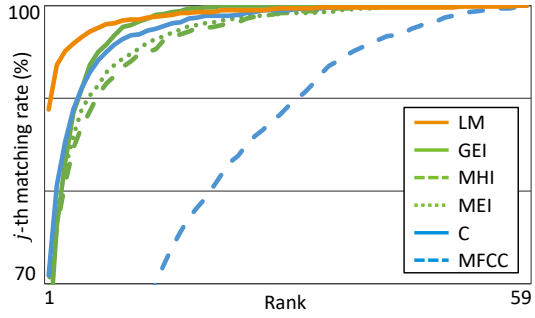


Fig. 10 Comparison of CMC curves achieved with local movement features obtained using our method and those obtained using existing methods.

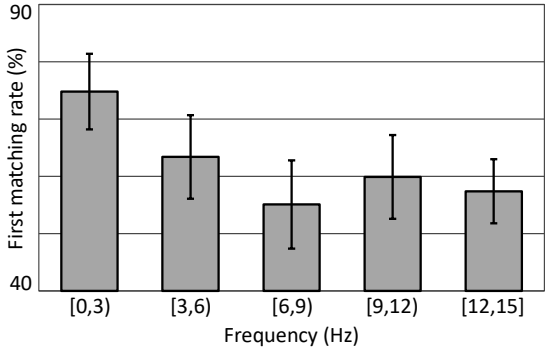


Fig. 11 Comparison of the identification performance achieved using each frequency band at intervals of 3 Hz.

by low-frequency components over a longer time. We also evaluated the identification performance using the CMC (Cumulative Match Characteristic) curve, which represents the j-th matching rate. The results shown in Figure 10 confirm that our method outperforms existing methods in identifying people using body sway.

3.4 Frequency analysis of temporal changes in local movements

We evaluated the identification performance using each frequency band obtained from temporal changes in local movements. We used the frequency band at intervals of 3 Hz. We set the same parameters for LM as described in Section 3.3.

Figure 11 shows the identification performance using each frequency band. We can see that the identification performance using frequency band $[0, 3)$ Hz was higher than that using the different frequency bands. We believe that the low frequency components of temporal changes in local movements contain more identity information than the high frequency components.

3.5 Improvement of the identification performance by combining likelihoods.

To improve the identification performance, we combined two different characteristics: a likelihood obtained by LM

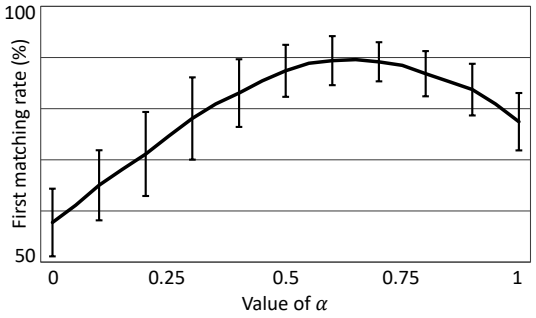


Fig. 12 Identification performance using the combination of the likelihood of LM and the likelihood of GEI while changing α.

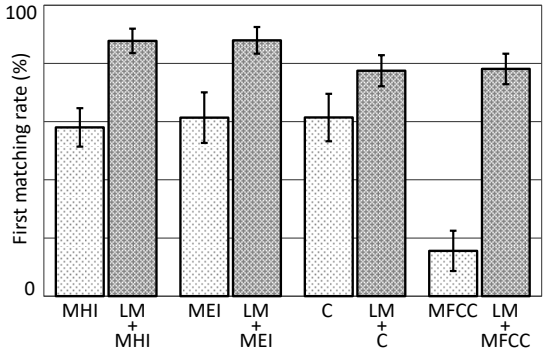


Fig. 13 Comparison of the identification performance using a combination of likelihoods and using each likelihood. We set $T = 1350$ (45 s).

and a likelihood obtained by GEI, MHI, MEI, C or MFCC. We applied the weighted linear sum for a combination of likelihoods. We used the weight α for the likelihood of LM and the weight $1-\alpha$ for the likelihood of another feature. We set the same feature parameters as described in Section 3.3 and $T = 1350$ (45 s) instead of $T = 3600$ (120 s).

Figure 12 shows the identification performance using a combination of a likelihood of LM and a likelihood of GEI while changing α. When α was 0 or 1, a single likelihood of GEI or LM was used. The highest identification performance was 89.6 ± 4.2% using $\alpha = 0.65$. The performance using a combination of likelihoods was higher than that using each likelihood. Interestingly, a combination of likelihoods using $T = 1350$ obtained almost the same performance as a single likelihood of LM using $T = 3600$.

Figure 13 shows the identification performance using a combination of a likelihood of LM and a likelihood of MHI, MEI, C or MFCC. We set $\alpha = 0.65$ for HEI and MEI and $\alpha = 0.9$ for C and MFCC. The identification performance using a combination with a likelihood of LM was higher than that using any other likelihood. We confirmed that LM improved the identification performance by combining different characteristics.

3.6 Evaluation of the variation in identification performance over the long term

We checked the variation in the identification performance over the long term. We collected video sequences for 10 participants (average age 22.6 ± 1.3 years; 9 males and 1 female) using the camera setup in Figure 3. We acquired three target video sequences for each participant. After 128 days, we acquired three query video sequences for each participant. We used a single video sequence as a target and a single video sequence as a query for each participant. We generated a metric matrix using 108 participants by removing the 10 test participants from the dataset described in Section 3.1. We compared the identification performance of our method with those of GEI, HMI, MEI and C, which obtained better performance in Section 3.3.

The identification performances using LM, GEI, MHI, MEI, and C were 66.7 ± 9.4%, 54.4 ± 6.8%, 41.1 ± 12.9%, 47.8 ± 13.1%, and 51.1 ± 12.0%, respectively. The chance rate was 10.0%. Although this result confirmed that our method performed better than the existing methods, the variation over the long term was still too high. We need to improve the performance to construct a practical application in future work.

4. Experiments with different postures

4.1 Datasets

We evaluated the identification performance under the condition that the postures of each participant were different between the query and target video sequences. We collected video sequences of body sway from 31 participants (average age 22.2 ± 1.2 years; 22 males and 9 females) using the camera setup in Figure 3 (b). Each participant maintained a feet-closed posture as in Figure 14 (a) and a feet-open posture as in Figure 14 (b), respectively. We observed each participant three times in each posture.

Figure 15 shows the distributions of the positions of the centers computed from the body regions of the mask images $\boldsymbol{m}_t$ ($t = 1,..., 3600$). To determine the origin of the distribution, we used the position of the center of the body region in the reference mask image $\boldsymbol{m}_r$. Each distribution in Figure 15 followed a normal distribution. We regarded the distribution in (a) as isotropic and that in (b) as anisotropic. The movements became small in the direction of the straight line between the feet. There seemed to be variation in body sway between the feet-closed and feet-open postures.

4.2 Comparison of the identification performance between feet-closed and feet-open postures

We used a video sequence of the feet-closed posture as the query and a video sequence of the feet-open posture as the target, and vice versa. We removed the 31 test participants from the dataset described in Section 3.1 to generate a metric

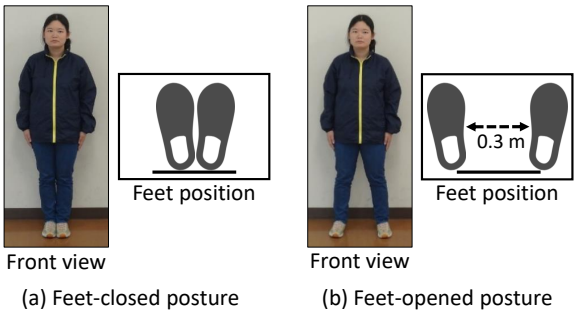


Fig. 14 Different postures of each participant.

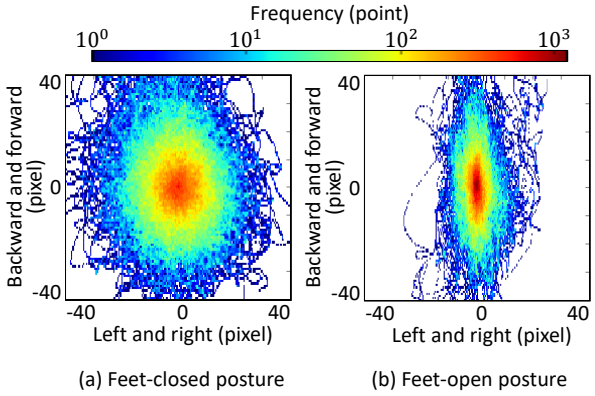


Fig. 15 Distributions of the positions of the center of the body. The vertical axis shows backward and forward movements. The horizontal axis shows left and right movements. The color bar shows the frequency of appearance.

Table 1 Comparison of the first matching rate (%) of LM, GEI and a combination of LM and GEI when the participants' postures differed between the query and target video sequence.

LM	GEI	LM+GEI
55.0 ± 11.8	62.5 ± 8.1	64.9 ± 11.3

matrix of 87 participants. We used LM and GEI to evaluate the performance.

Table 1 shows the identification performance between feet-closed and feet-open postures. The performance of LM was 7.5 points lower than that of GEI. The difference in postures clearly influenced the identification performance of LM. However, because the highest performance was achieved using a combination of LM and GEI, we believe that LM helped to improve the identification performance. However, the performances were not high because of the variation between feet-closed and feet-open postures. We will work on improving the performance to construct a practical application in future work.

5. Conclusions

We proposed a method of identifying people using video sequences of body sway. We designed a feature extraction method for identification by measuring temporal and spatial changes in local movements. To evaluate our method, we originally collected three novel datasets containing video sequences of body sway. The first dataset included 118 participants in an upright posture, the second included the variation over 128 days and the third included variation in feet-closed and feet-open postures. We confirmed that our method can extract informative features from video sequences of body sway for the identification of people.

Spatio-Temporal Histograms of Oriented Gradient (STHOGs) [29] have been previously proposed as a spatio-temporal feature. For instance, El-Alfy et al. [30] reported improved person re-identification performance using STHOG features. This method counts spatial changes and temporal changes simultaneously by combining the distributions of gradients and treating them equivalently. In contrast, our method extracts a feature that represents the temporal changes after roughly gathering the spatial changes because our aim is to emphasize the movements of body sway. We plan to compare the performance of STHOG features in future work.

As described Section 1, an overhead camera has the advantage of lowering the level of mutual occlusion among people. However, this camera has the disadvantage of narrow field angle such that it only watches one spot an area. We believe that the observation of body sway within a single spot could lead to some desirable applications, e.g., identification when people are riding an elevator, waiting for a traffic light to change, or waiting in line to use a cash machine.

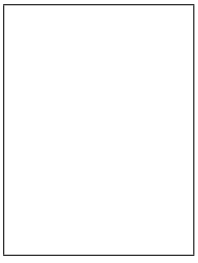
As part of our future work, we intend to increase the tolerance of our method when there is variation over the long term and variation in postures. Furthermore, we plan to develop a practical application by reducing the time required to measure body sway. We also intend to investigate identification performance while participants change their standing positions. We will expand our evaluation using a convolutional neural network-based method after collecting a large dataset of body sway.

Acknowledgment This work was partially supported by JSPS KAKENHI under grant number JP17K00238 and MIC SCOPE under grant number 172308003.

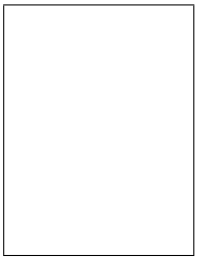
References

[1] A.K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," IEEE Transactions on circuits and systems for video technology, vol.14, no.1, pp.4–20, 2004.
[2] X. Wang, "Intelligent multi-camera video surveillance: A review," Pattern Recognition Letters, vol.34, no.1, pp.3 – 19, 2013.
[3] J. Ashbourn, Biometrics: Advanced Identity Verification, Springer, 2000.
[4] A. Dantcheva, C. Velardo, A. D'Angelo, and J.L. Dugelay, "Bag of soft biometrics for person identification," Multimedia Tools and Applications, vol.51, no.2, pp.739–777, 2011.
[5] M.S. Nixon, P.L. Correia, K. Nasrollahi, T.B. Moeslund, A. Hadid, and M. Tistarelli, "On soft biometrics," Pattern Recognition Letters, vol.68, pp.218 – 230, 2015.
[6] P. Tome, J. Fierrez, R. Vera-Rodriguez, and M.S. Nixon, "Soft biometrics and their application in person recognition at a distance,"

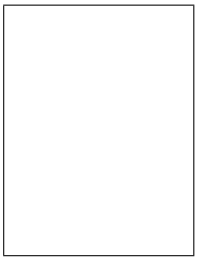
IEEE Transactions on Information Forensics and Security, vol.9, no.3, pp.464–475, 2014.
[7] G. Antipov, S. Berrani, N. Ruchaud, and J. Dugelay, "Learned vs. hand-crafted features for pedestrian gender recognition," Proceedings of the 23rd ACM International Conference on Multimedia, pp.1263–1266, 2015.
[8] H. Tang, H. Liu, and W. Xiao, "Gender classification using pyramid segmentation for unconstrained back-facing video sequences," Proceedings of the 23rd ACM International Conference on Multimedia, pp.1183–1186, 2015.
[9] A. Li, L. Liu, K. Wang, S. Liu, and S. Yan, "Clothing attributes assisted person reidentification," IEEE Transactions on Circuits and Systems for Video Technology, pp.134–146, 2015.
[10] C.H. Kuo, S. Khamis, and V. Shet, "Person re-identification using semantic color names and rankboost," Proceedings of IEEE Workshop on Applications of Computer Vision, pp.281–287, 2013.
[11] J. Han and B. Bhanu, "Individual recognition using gait energy image," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.28, no.2, pp.316–322, Feb. 2006.
[12] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, "Gait recognition using a view transformation model in the frequency domain," Proceedings of 9th European Conference on Computer Vision, pp.151–163, 2006.
[13] Z. Shi, T. Hospedales, and T. Xiang, "Transferring a semantic representation for person re-identification and search," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.4184–4193, 2015.
[14] R. Layne, T. Hospedales, and S. Gong, "Person re-identification," chapter Attributes-Based Re-identification, pp.93–117, 2014, Springer.
[15] J. Roth and X. Liu, "On the exploration of joint attribute learning for person re-identification," Proceedings of 12th Asian Conference on Computer Vision, pp.673–688, 2014.
[16] H. Kollegger, C. Baumgartner, C. Wöber, W. Oder, and L. Deecke, "Spontaneous body sway as a function of sex, age, and vision: posturographic study in 30 healthy adults," European Neurology, vol.32, no.5, pp.253–259, 1992.
[17] G. Cavalheiro, M. Almeida, A. Pereira, and A. Andrade, "Study of age-related changes in postural control during quiet standing through linear discriminant analysis," BioMedical Engineering On-Line, vol.8, no.35, 2009.
[18] N. Nies and P. Sinnott, "Variations in balance and body sway in middle-aged adults. subjects with healthy backs compared with subjects with low-back dysfunction," Spine, vol.16, no.3, pp.325–330, 1991.
[19] Y. Yu, H.C. Chung, L. Hemingway, and T.A. Stoffregen, "Standing body sway in women with and without morning sickness in pregnancy," Gait & posture, vol.37, no.1, pp.103–107, 2013.
[20] P. Bergin, A. Bronstein, N. Murray, S. Sancovic, and D. Zeppenfeld, "Body sway and vibration perception thresholds in normal aging and in patients with polyneuropathy," Neurol Neurosurg Psychiatry, vol.58, no.3, pp.335–340, 1995.
[21] T. Kamitani, H. Yoshimura, M. Nishiyama, and Y. Iwai, "Identifying people using temporal and spatial changes in local movements measured from body sway," Proceedings of the 4th Asian Conference on Pattern Recognition, pp.828–833, 2017.
[22] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.23, no.3, pp.257–267, 2001.
[23] P. Kasprowski and J. Ober, "Eye movements in biometrics," Proceedings of International Workshop on Biometric Authentication, pp.248–258, 2004.
[24] N. Cuong, V. Dinh, and L. Ho, "Mel-frequency cepstral coefficients for eye movement identification," Proceedings of 24th International Conference on Tools with Artificial Intelligence, pp.253–260, 2012.
[25] K. Weinberger, J. Blitzer, and L. Saul, "Distance metric learning for large margin nearest neighbor classification," Journal of Machine Learning, pp.207–244, 2009.
[26] H.T. Lin, C.J. Lin, and R.C. Weng, "A note on platt ' s probabilistic outputs for support vector machines," Machine learning, vol.68, no.3, pp.267–276, 2007.
[27] M. Nishiyama, T. Miyauchi, H. Yoshimura, and Y. Iwai, "Synthesizing realistic image-based avatars by body sway analysis," Proceedings of the Fourth International Conference on Human Agent Interaction, pp.155–162, 2016.
[28] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," IEEE Transactions on Audio and Electroacoustics, vol.15, no.2, pp.70–73, 1967.
[29] C. Hua, Y. Makihara, and Y. Yagi, "Pedestrian detection by using a spatio-temporal histogram of oriented gradients," IEICE Transactions on Information and Systems, vol.96, no.6, pp.1376–1386, 2013.
[30] H. El-Alfy, D. Muramatsu, Y. Teranishi, N. Nishinaga, Y. Makihara, and Y. Yagi, "A visual surveillance system for person re-identification," Proceedings of 13th International Conference on Quality Control by Artificial Vision, no.103380D, pp.1–7, 2017.



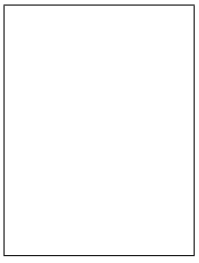
Takuya Kamitani graduated from Tottori University in 2017. He currently attends the master's course at Graduate School of Sustainability Science, Tottori University. He is engaged in studies relating to soft biometrics.



Hiroki Yoshimura graduated from Tottori University in 1993 and completed the M.S. and doctoral programs in 1995 and 1998, respectively. I received my Ph.D. degree from Totori University. I am currently an asisitant professor in the Graduate School of Engineering at Tottori University. I am engaged in studies relating to machine learning.



Masashi Nishiyama is an associate professor in Graduate School of Engineering at Tottori University, Japan. He received his M.S. degrees in Engineering from Graduate School of Natural Science and Technology, Okayama University, Japan, in 2002. He joined in Corporate Research & Development Center, TOSHIBA Corporation during 2002–2015. He received his Ph.D. degrees in Interdisciplinary Information Studies from Graduate School of Interdisciplinary Information Studies, University of Tokyo, Japan, in 2011. His recent research has focused on novel principles for representing identities and behaviors of humans.



Yoshio Iwai graduated from Osaka University in 1992 and completed the M.S. and doctoral programs in 1994 and 1997, respectively. He was then appointed a research associate at the university, subsequently becoming an associate professor. From 2004 to 2005, he was a visiting researcher at the Cambridge University. He is currently a professor in the Graduate School of Engineering at Tottori University. He is engaged in studies relating to computer vision. He is a member of IEEE, the Information Processing Society, He holds a D.Eng. degree.