Identifying People using Temporal and Spatial Changes in Local Movements Measured from Body Sway

Takuya Kamitani, Hiroki Yoshimura, Masashi Nishiyama and Yoshio Iwai
Graduate School of Engineering, Tottori University, Japan
s132017@eecs.tottori-u.ac.jp, nishiyama@eecs.tottori-u.ac.jp

Abstract

We propose a novel method of identifying people using temporal and spatial changes in local movements measured from a video sequence of body sway. Existing methods identify people using a gait feature mainly representing the large swinging of the limbs. The use of the gait feature introduces a problem in that the identification performance decreases when people stop walking. To extract an informative feature from people who have stopped walking, our method measures small swings of the body, which is called body sway. We extract the feature from local movements of body sway by participially dividing the body into regions. Experimental results for a dataset of body sway of 118 participants show that the local movement feature obtained using our method outperforms the gait feature obtained using an existing method.

1. Introduction

Identification using video surveillance cameras allows the development of a convenient and non-intrusive biometric authentication system. Recently, soft biometrics [8] that represent human attributes have been an active topic of pattern recognition research in terms of extracting informative features for identification. Human attributes can be split intuitively into three types: physical characteristics [1, 18] (e.g., gender and age), adhered human characteristics [14, 13] (e.g., clothing and belongings), and behavioral characteristics [10, 15] (e.g., gestures and gait). In particular, behavioral characteristics have the advantage that there are differences in movements among individuals. Behavioral characteristics can thus be used to identify people even if those people have the same attributes of gender and age, or the same attributes of clothing.

Existing methods [10, 15] that use behavioral characteristics generally exploit gait features acquired by cameras. However, we cannot assume that people are always walking. People frequently stop walking when, for example, riding an elevator, waiting for a traffic light to change, or waiting in line to use a cash machine. In these situation, gait features insufficiently represent behavioral characteristics, and they are not informative for identification. The identification performance is therefore sometimes lower when using gait features.

When people stop walking, their bodies do not remain completely still but slightly and continuously move in all directions. This movement of the body naturally occurs to maintain a person's posture, and is called body sway. In the field of medical science, many researchers [12, 4, 16, 20, 2] have attempted to measure the center of gravity of body sway using force plates embedded in a floor. The purposes of existing methods are not identification, and instead it has been reported that the center of gravity can be used to classify gender and age [12, 4], people with lower-back pain [16], women with morning sickness [20], and patients with neuropathy [2]. We thus assume that body sway contains the identity of people, and is a behavioral characteristic that can be used as a human attribute in soft biometrics. In this paper, we tackle the challenging task of extracting an informative feature for identification from a video sequence of body sway. In addition, body sway has the advantage that we can passively observe people with an overhead camera because the movement of body sway can be measured for the upper half of the body. We thus do not need to set a camera to the side of a person, as widely done in extracting gait features [10, 15]. The use of an overhead camera can avoid 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 having stopped walking by measuring local movements in body regions. Our method computes the center of body sway from a video sequence, and spatially divides the body region into small local regions using the center of body sway. Our method measures temporal and spatial changes in local movements in the regions, and conducts frequency analysis for feature extraction. Note that we treat an upright posture as a specific example of the posture of a person having stopped walking. We originally collected a novel dataset of body sway for 118 subjects. Experimental results show that the identification performance improved from 52.0% when using a gait feature of an existing method to 94.6% when using our local movement feature.

The remainder of the paper is organized as follows. Section 2 describes our method of extracting the feature from a video sequence of body sway, Section 3 presents the identification performance when using body sway, and Section 4 gives our concluding remarks.

2. Our method

2.1. Overview

We consider the informative feature extracted from a video sequence of body sway acquired for people having an upright posture. The first feature is movement representing how much the body is temporally and spatially swinging. The movement contains an identity that is the difference in, for example, gender, age, chronic disease, how muscles attach, or sense of balance. The second feature represents the body shape, such as an obese or thin body type. The third feature represents the body posture, such as a stooping or slouching posture.

With respect to gait recognition, existing methods [10, 15] mainly exploit the body shape and slightly add movements of the limbs. Existing methods [10, 3] of action recognition have a basis similar to that of gait recognition. Researchers [11, 7] have exploited temporal movement for gaze authentication. Although existing methods are designed for the different propose of identifying people using body sway, we can simply use the methods for feature extraction. In preliminary experiments, however, we could not obtain high identification performance using existing methods for video sequences of body sway.

We focus on how to represent the identity using temporal and spatial changes in movements due to body sway. Figure 1 is an overview of our method. We divide the body region into small local regions to represent the spatial movements of body sway. We measure temporal changes in local movements from the local regions, and compute a feature for identification. The detail of our method is described below.

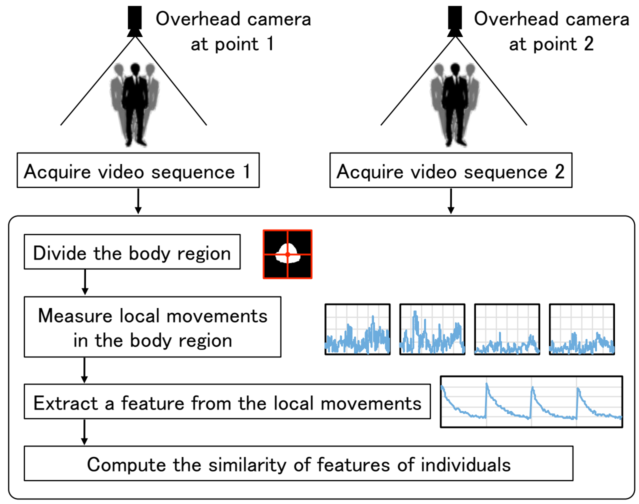

Figure 1. Overview of our method.

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. The movements of body sway occur around a certain position that is the center of the swinging movement. An existing method [17] measures the movements using a whole body region under the assumption that all body parts synchronically move in the same direction. Although the existing method considers temporal changes in the movement, it ignores spatial changes in the movement. We thus extend the existing method to represent spatial changes in movement and extract informative features measured from body sway.

We compute a mask image $\boldsymbol{m}_t$ in which a pixel takes a value 1 if it is within a body region and zero otherwise from a frame of the video sequence at time $t \in 1, ..., T$. The existing method [17] infers the reference time r representing the temporal center of swings using algorithm 1 from whole body regions in mask images.

Algorithm 1 Determining reference time r

for $\tilde{r} = 1$ to T **do**
  $D_{\tilde{r}} \leftarrow 0$
  for $t = 1$ to T **do**
    compute $\tilde{d} = \|\boldsymbol{m}_{\tilde{r}} - \boldsymbol{m}_t\|_1$
    $D_{\tilde{r}} \leftarrow D_{\tilde{r}} + \tilde{d}$
  end for
end for
$r \leftarrow \arg\min D_{\tilde{r}}$

To consider the spatial change in movement, our method divides the body region into a plurality of local regions, and computes the local movement in each region. The simple idea is to divide the body region into a lattice. However, we cannot stably measure movements around the center of gravity of the body region when the lattice cells are made small. We thus radially divide the body region into local regions using the center of gravity of the body, as illustrated in Figure 2. Our method computes the center of gravity of the body region using the mask image $\boldsymbol{m}_r$ acquired at reference time r. Note that we assume that the center of gravity

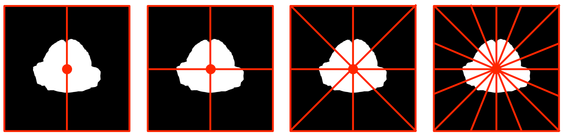

Figure 2. Examples of local regions radially divided from the body region. I is the number of local regions.

is at the same position for all times $t \in 1, ..., T$. Our method measures the temporal changes in local movements from the local regions divided spatially 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 pixel values indicated by $\boldsymbol{x}$, and region(i) is the i-th local region. We regularized with the $\boldsymbol{L}_1$-norm.

Algorithm 2 Computing local movement $d_{i,t}$

```
compute the center of gravity in m_r
for i = 1 to I do
  set the i-th region using the computed center
  for t = 1 to T do
    compute d_{i,t} using Equation (1)
  end for
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 swings of body sway vary in random directions. We thus need to consider a feature invariant to the randomness of the swings.

In the field of signal processing, 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 believe that the frequency analysis techniques are adequate for achieving high performance. We assume that the phase components are shifted each time when measuring local movements. To alleviate the randomness of swings, we do not use the phase components.

Our method estimates the power spectral density (PSD) employing Welch's method [19] from the local movements $d_{i,t}$ to extract the feature $\boldsymbol{f}$ for identification using Algorithm 3. We assume that high-frequency components contain unnecessary noise for identification. Our method exploits only low-frequency components $\boldsymbol{f}_i$ by selecting them from the DC component to the M-th component. The dimension of $\boldsymbol{f}_i$ is M. The feature for identification is represented as $\boldsymbol{f} = [\boldsymbol{f}_1^{\mathrm{T}}, \ldots, \boldsymbol{f}_I^{\mathrm{T}}]^{\mathrm{T}}$. The dimension of $\boldsymbol{f}$ is IM. We expect that $\boldsymbol{f}$ represents the identify of people while alleviating the randomness of the swings in the temporal and spatial changes of local movements.

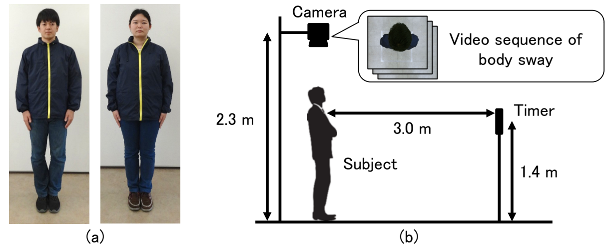

Figure 3. Setup for acquiring video sequences of body sway.

Algorithm 3 Extracting the feature $\boldsymbol{f}$

```
for i = 1 to I do
  compute the PSD from {d_{i,t} | t ∈ 1, ..., T}
  take the logarithm of the PSD for each frequency
  select the PSD f_i of M low frequencies
end for
concatenate {f_i | i ∈ 1, ..., I} to f
```

3. Experiments

3.1. Dataset of video sequences of body sway

To evaluate the identification performance of our method, we collected video sequences of body sway acquired for 118 participants (average age of 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, such as 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 floor normal. Each participant stood under the camera as shown in Figure 3 (b). A marker was set to position a heel of the participant at the standing position. We asked each participant to look at a timer set at a position 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 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 move-

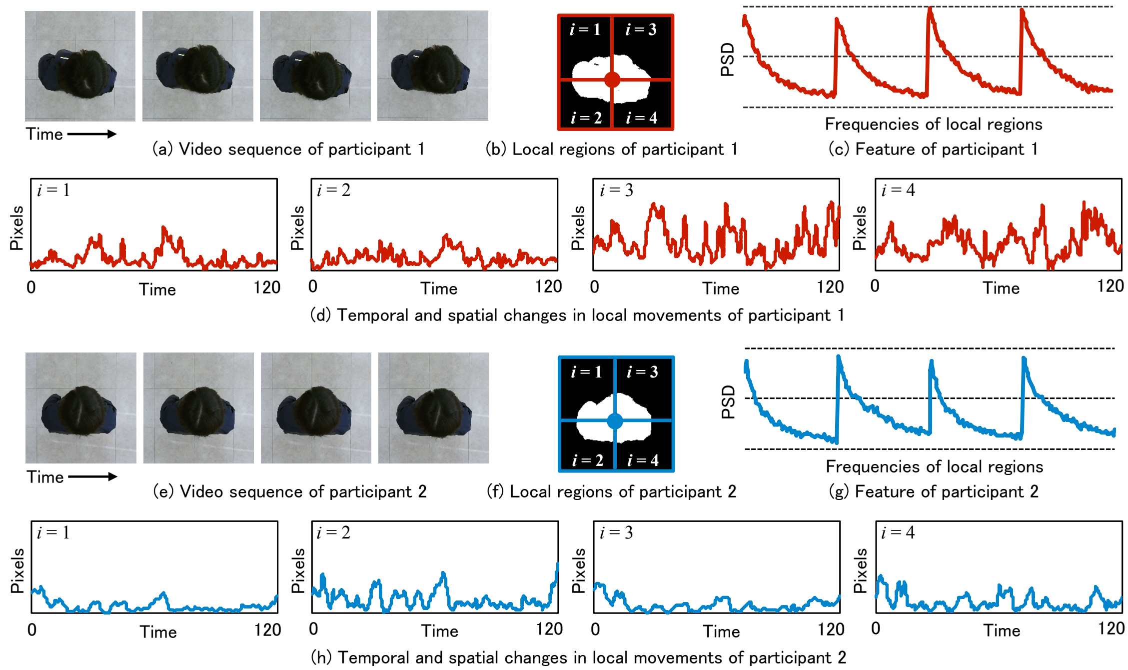

Figure 4. Examples of temporal and spatial changes in local movements measured from video sequences of two subjects.

ments. Each participant was observed three times. The participant sat and rested each time shooting ended. To generate 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). Features in (c) and (g) were extracted from the temporal and spatial changes of the local movements in (d) and (h). We find that features differ between the participants even though the video sequences appear to be almost the same.

We evaluated the identification performance while changing the number of local regions I and the range of the low frequencies M, separately. We used one query video sequence and two target video sequences for each participant. We used the correct match rate (%) for the identification performance. Figure 5 (a) and (b) shows the identification performance. We used a nearest-neighbor algorithm for identification. To compute the PSD, we used a Hann window and 512 signal samples for the fast Fourier transformation. We set $M = 128$ in (a) and $I = 30$ in (b). In (a), $I = 1$ corresponds to no division into local regions. We see that the identification performance was improved by increasing the number of local regions. When the number of local regions exceeded 20 in (a), the identification performance was almost constant. We also see that

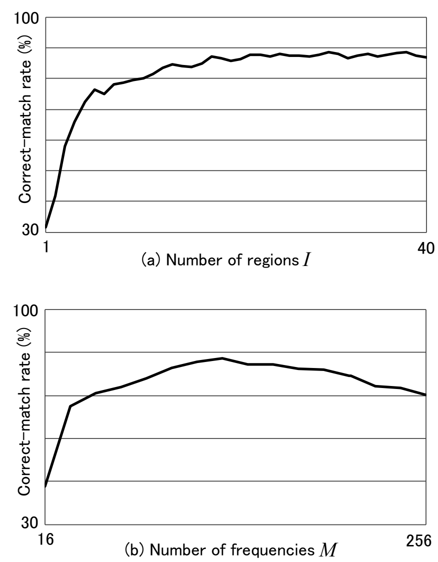

Figure 5. Identification performance when changing the parameters of our method: (a) number of local regions, (b) range of low frequencies.

high-frequency components reduce the identification performance in (b). $M = 128$ provided the best performance. We believe that high-frequency components do not represent features sufficient for identification.

We combined our local movement feature extraction with the nearest-neighbor (**NN**) algorithm and the following classifiers.

Table 1. Identification performance (%) when combining our feature with existing classifiers.

NN	SVM	LDA	LR
87.3 ± 1.8	87.6 ± 1.4	92.4 ± 4.2	$\mathbf{94.6 \pm 1.4}$

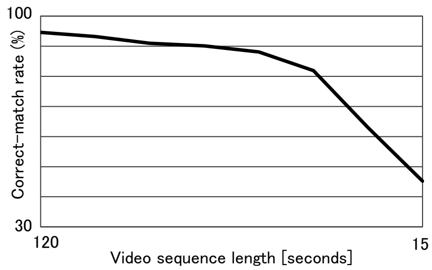

Figure 6. Identification performance when changing the time length of video sequences.

- **SVM**: support vector machine [5] (penalty parameter $C = 1$, use of RBF kernels),
- **LDA**: linear discriminant analysis [9] (dimension reduced to 117 using principal component analysis),
- **LR**: logistic regression [6] (regularization parameter of 1).

Table 1 presents the identification performance when combining our feature with existing classifiers. We used the same query and target sequences as in the above experiment. We see that LR had higher identification performance than the other classifiers.

Figure 6 shows the identification performance when reducing the time length of video sequences from 120 s ($T = 3600$). When the time length was reduced to 1/2 or 1/4, the performance reduced by 6.5 or 31.6 points, respectively. We believe that the degradation of the performance is related to the periodicity of body sway. We should conduct further examinations 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 feature obtained using our method and features obtained using existing methods.

- **LM** (Local Movements): We computed a feature using our method. We set the parameters $I = 30, M = 128$.
- **GEI** (Gait Energy Image) [10]: We assumed a walking cycle T. We computed a feature by averaging the mask images as $\Sigma_{t=1}^{T} \boldsymbol{m}_t / T$.
- **MHI** (Motion History Image) [3]: We assigned a weight $\tau = t/T$ for each time at a position where movement was generated. We temporally added the weights at each position.

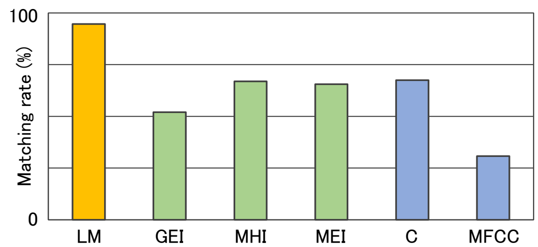

Figure 7. Comparison of the identification performance achieved with local movement feature obtained using our method and the performances of features obtained using existing methods.

- **MEI** (Motion Energy Image) [3]: We set positions where movement was generated as $\cup_{t=2}^{T} |\boldsymbol{m}_t - \boldsymbol{m}_{t-1}|$.
- **C** (Cepstrum) [11]: We applied cepstrum analysis to the temporal change in local movement. We used quefrencies from the DC component to the 1100-th component for a feature.
- **MFCC** (Mel-frequency Cepstrum Coefficients) [7]: We computed a feature using 40 coefficients.

We used 120-s video sequences of 118 participants and a logistic regression classifier. The query and target sequences had the same experimental conditions described in Section 3.2. Note that C and MFCC were computed from 30 local regions as in our method.

Figure 7 shows the identification performance achieved using features extracted using our method and existing methods. We see that LM outperformed GEI, MHI, and MEI. We believe that the performances of GEI, MHI, and MEI were lower because these methods cannot represent small movements of the body; the purpose of GEI was gait recognition when people move their limbs largely while the purpose of MHI and MEI was action recognition when people dynamically move their bodies. The performances of MHI and MEI were almost equivalent, while the performance of GEI was lower. Table 2 compares LM and GEI in terms of the numbers of correctly and wrongly identified queries. LM had more correctly identified queries than did GEI. Returning to Figure 7, we see that LM outperformed C and MFCC. We believe that the performances of C and MFCC were lower because the purpose of these methods was gaze authentication when people abruptly move their eyes within a short time and these methods are thus unable to stably represent body sway that is characterized by low-frequency components over a long time. We confirmed that our method outperforms existing methods in identifying people using body sway.

3.4. 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 data for 10 par-

Table 2. Comparison of the numbers of correctly and wrongly identified queries. The total number of queries was $118 \times 3 = 354$.

	GEI (Correct)	GEI (Wrong)
LM (Correct)	176	**159**
LM (Wrong)	**8**	11

ticipants (average age of 22.6 ± 1.3 years; nine males and one female). We acquired three target video sequences for each participant. After 128 days, we acquired three query video sequences for each participant. We used the same experimental conditions described in Section 3.3. The identification performance of LM was $55.0 \pm 8.3\%$, while the performances of GEI, HMI, and HEI were $31.7 \pm 1.7\%$, $38.3 \pm 8.3\%$, and $43.3 \pm 3.3\%$, respectively. Although the identification performance of our method was higher than the performances of existing methods, variation remained over the long term. We need to improve the performance in future work.

4. Conclusions

We proposed a method of identifying people using video sequences of body sway. We designed a feature for identification by measuring temporal and spatial changes in local movements. We evaluated our method on a dataset of body sway collected for 118 participants. Our method of identification was highly accurate compared with existing representative methods. To the best knowledge of the authors, this is the first work in the area of person recognition that focuses on the use of body sway to extract informative features. As part of our future work, we intend to evaluate the identification performance for various postures of people standing.

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

References

[1] G. Antipov, S. Berrani, N. Ruchaud, and J. Dugelay. Learned vs. hand-crafted features for pedestrian gender recognition. In *Proceedings of the 23rd ACM International Conference on Multimedia*, pages 1263–1266, 2015. 1

[2] 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*, 58(3):335–340, 1995. 1

[3] A. Bobick and J. Davis. The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3):257–267, 2001. 2, 5

[4] 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 OnLine*, 8(35), 2009. 1

[5] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. 5

[6] D. Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B*, 20(2):215–242, 1958. 5

[7] N. Cuong, V. Dinh, and L. Ho. Mel-frequency cepstral coefficients for eye movement identification. In *Proceedings of 24th International Conference on Tools with Artificial Intelligence*, volume 1, pages 253–260, 2012. 2, 5

[8] A. Dantcheva, C. Velardo, A. D'Angelo, and J.-L. Dugelay. Bag of soft biometrics for person identification. *Multimedia Tools and Applications*, 51(2):739–777, 2011. 1

[9] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 2nd edition, 1990. 5

[10] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):316–322, Feb. 2006. 1, 2, 5

[11] P. Kasprowski and J. Ober. Eye movements in biometrics. In *Proceedings of International Workshop on Biometric Authentication*, pages 248–258, 2004. 2, 5

[12] 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*, 32(5):253–259, 1992. 1

[13] C. H. Kuo, S. Khamis, and V. Shet. Person re-identification using semantic color names and rankboost. In *Proceedings of IEEE Workshop on Applications of Computer Vision*, pages 281–287, 2013. 1

[14] 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*, pages 134–146, 2015. 1

[15] 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 9th European Conference on Computer Vision*, pages 151–163, 2006. 1, 2

[16] 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*, 16(3):325–330, 1991. 1

[17] M. Nishiyama, T. Miyauchi, H. Yoshimura, and Y. Iwai. Synthesizing realistic image-based avatars by body sway analysis. In *Proceedings of the Fourth International Conference on Human Agent Interaction*, pages 155–162, 2016. 2

[18] H. Tang, H. Liu, and W. Xiao. Gender classification using pyramid segmentation for unconstrained back-facing video sequences. In *Proceedings of the 23rd ACM International Conference on Multimedia*, pages 1183–1186, 2015. 1

[19] 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*, 15(2):70–73, 1967. 3

[20] 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*, 37(1):103–107, 2013. 1