Gender Classification Using Video Sequences of Body Sway Recorded by Overhead Camera

Takuya Kamitani*, Yuta Yamaguchi*, Shintaro Nakatani*, Masashi Nishiyama* and Yoshio Iwai*
* Graduate School of Engineering, Tottori University, Tottori, 680–8552, Japan

***Abstract*—We investigate whether it is possible to classify the gender of a standing person based on a video sequence containing body sway recorded by an overhead camera. Existing methods that extract a feature from the movement of a walking person for gender classification cannot detect the slight movements of a standing person. In this paper, we propose a method for extracting a feature from the body sway of a standing person. We design a spatio-temporal feature for representing body sway using the frequency analysis of time-series signals derived from the local movements of the upper body. To evaluate the accuracy of our method, we acquired video sequences of body sway from 30 females and 30 males using an overhead camera. We found that our method obtained 90.3±1.3% accuracy for the gender classification of a standing person. We compared the accuracy of our method with that of parameters based on medical data. We found that the proposed spatio-temporal feature extracted from body sway significantly improved gender classification accuracy.**

I. Introduction

There is high demand for technology that can classify the gender of a person based on a video sequence [1], [2], [3]. Such gender classification has various applications, such as security surveillance and marketing planning. To accurately classify the gender of a person, the characteristics that distinguish between females and males must be obtained. The movements of a person in a video sequence have recently been considered for representing such characteristics.

In general, the movements of a person can be divided into walking movements (gait) and standing movements (body sway). Below, we review methods that classify the gender of a person based on walking or standing movements in a video sequence. We first consider gender classification based on gait. To distinguish between females and males based on gait, some methods [4], [5], [6] extract the gait energy image (GEI) as a feature for training a gender classifier. It has been reported that GEIs can be used to classify the gender of a walking person with high accuracy. However, methods based on GEIs are designed for classifying the gender of a walking person. To the best of our knowledge, there are no existing methods for gender classification based on body sway. Here, we propose a method for extracting a feature from body sway and investigate whether it can be used for gender classification.

We discuss whether body sway can be used to distinguish between females and males. Analytical research in the medical field has shown that there are differences between standing females and males in terms of body sway. Analytical studies [7], [8], [9] have used time-series signals of the center positions of the pressure of the feet acquired from a force plate placed on the floor. They demonstrated that there are significant differences between females and males in terms of the frequency characteristics and trajectories of the time-series signals. These studies focused on obtaining medical data on body sway and did not consider practical applications. To apply such medical data for gender classification, a contact-type sensor must be placed on the floor.

Here, we observe body sway using a camera, which is non-contact-type sensor. Previous studies [10], [11], [12], [13], [14] have measured body sway using a camera instead of a force plate for applications such as fall prevention assessment, avatar video generation, and person re-identification. However, the features of body sway were not used to distinguish between females and males.

Here, we investigate whether body sway can be used to classify the gender of a standing person by extracting a feature from a video sequence. We used an overhead camera attached to the ceiling in our experimental setting. We assumed that the head of a standing person makes larger movements than those of the legs and waist. An overhead camera can observe upper body sway, including that of the head. In our method, we estimate the upper-body region in a video sequence to obtain a silhouette sequence. We measure the time-series signals of body sway from the silhouette sequence and extract a feature for gender classification. We created a dataset of video sequences of the body sway of 60 participants to evaluate gender classification accuracy. We found that our method obtained $90.3 \pm 1.3\%$ accuracy for gender classification on our dataset. We also compared the accuracy of our method with that of features derived from medical data and found that our method has superior accuracy. To the best of our knowledge, the use of body sway in video sequences for gender classification has not been previously reported. Our main contribution is the development of a method for gender classification based on body sway. The remainder of this paper is organized as follows. Section II reviews related work. Section III describes our method and Section IV shows the experimental results of gender classification. Finally, Section V presents the conclusions.

II. Related Work

A. Video Sequences of Walking People for Gender Classification

To classify the gender of a walking person in a video sequence, some methods [4], [5], [6] use GEI features extracted from gait. A GEI feature [15] is represented by an average image calculated from a silhouette sequence containing the movements of arms and legs during one gait cycle. Shan et al. [4] applied GEI features directly to gender classification. Martín-Félez et al. [5] temporally divided one gait cycle into four intervals and extracted a GEI feature from each interval for gender classification. Yu et al. [6] assumed that the movements of arms and legs affect gender classification accuracy. Their method extracts a GEI feature that represents the movement of each body part and assigns an adaptive weight to each feature. Various methods [4], [5], [6] assume that arms and legs provide the most information. However, the body sway of a standing person rarely includes large movements of the arms and legs. In this paper, we extract a feature from body sway for gender classification.

B. Use of Single Images for Gender Classification

Some methods extract features from a single image for gender classification. Studies [16], [17], [18], [19] have proposed the use of low-level features derived from the colors and gradients in a single image. Other studies [20], [21], [22], [23] applied a convolutional neural network (CNN) to extract features from a single image in an end-to-end framework. These methods achieve high accuracy in gender classification when trained using a large number of images. Here, we increase gender classification accuracy by incorporating single images with temporal movements. Convolutional three-dimensional (C3D) [24] features are well-known spatio-temporal features. Xu et al. [25] and Liu et al. [26] reported that C3D features are useful for action recognition for classifying large movements, such as soccer shots, table tennis shots, and swimming strokes. However, C3D features are not designed for gender classification. We thus extract a spatio-temporal feature from body sway that are suitable for gender classification.

C. Analytical Research on Differences between Female and Male in Terms of Body Sway

Analytical studies [7], [8], [9] have been conducted to determine the differences between females and males in terms of body sway. These studies obtained time-series signals of body sway using a force plate placed on the floor and reported that there are significant differences in these signals between females and males in terms of frequency characteristics and trajectories [7], the elliptic approximated from trajectories [8], and the specific band of frequency characteristics [9]. However, they did not apply these parameters to gender classification. In preliminary experiments, we found that we could not achieve high gender classification accuracy using medical data. We thus extract a feature from body sway to accurately classify the gender of a standing person.

D. Applications of Body Sway in Video Sequences

Using a camera instead of a force plate to measure the body sway of a standing person has various applications [10], [11], [12], [13], [14]. Wang et al. [10] used body sway to evaluate the risk of falling. They observed a person from various directions using multiple cameras and obtained the time-series signals of three-dimensional centers. Nishiyama et al. [11] used body sway to generate a video sequence of an avatar of a person. They observed a person using a camera placed in front of the person and estimated the center position from time-series signals. Yeung et al. [12] and Lv et al. [13] applied body sway to evaluate a person's balance in the clinical field. They analyzed the time-series signals of body joints obtained from Microsoft Kinect. Kamitani et al. [14] applied body sway to identify people. They obtained the time-series signals of body sway recorded by an overhead camera and extracted the feature representing the identity of an individual. The above examples demonstrate that body sway can be used for various applications. However, body sway has not been applied to gender classification. In this paper, we investigate whether body sway can be used to classify the gender of a standing person.

III. Proposed Gender Classification Method

A. Overview

The proposed method can classify the gender of a standing person using a video sequence of body sway. We acquire a video sequence of a standing person using an overhead camera attached to the ceiling. The overhead camera is used to observe the upper body of a standing person, where the amount of movement is larger than that of the lower body. For the upper body, the head has the largest movement. We extract an informative feature from the upper body for gender classification by acquiring a video sequence of body sway.

Here, we discuss the camera setting used to view the upper body of a standing person. Ceiling height, which varies in real-world scenarios, affects the apparent size of a person and thus the amount of the movement observed from body sway. Fig. 1 shows examples of the apparent size of the upper body. Although Figs. 1 (a) and (b) show the upper body of the same female, the apparent sizes are completely different because the ceiling heights are different. The same tendency for males is shown in Figs. 1 (c) and (d). We thus develop a method for body sway measurement that does not depend on ceiling (camera) height.

The proposed method consists of the following three steps. We assume that a person is standing below the overhead camera and maintains the same posture. Fig. 2 shows an overview of the proposed method. In the first step, we acquire a video sequence of the standing person using the overhead camera and use it to estimate a silhouette sequence that represents the upper body. In the second step, we remove the variation of the apparent size of the upper body in the silhouette sequence due to the height of the overhead camera. In the third step, we measure the time-series signals of body sway from the silhouette sequence and extract a feature for gender classification. We determine the gender class using the extracted feature and a pre-trained classifier. We describe the removal of the variation in the apparent size of the body region in Section III-B and the extraction of a feature from body sway for gender classification in Section III-C.

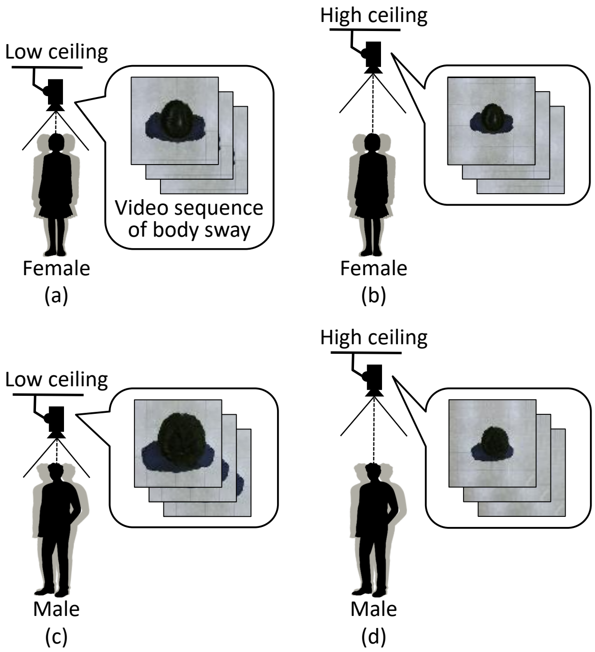


Fig. 1. Examples of the variation of the apparent size of the upper body in our experimental setting, where the camera height was randomly changed. Female recorded by (a) low and (b) high camera. Male recorded by (c) low and (d) high camera.

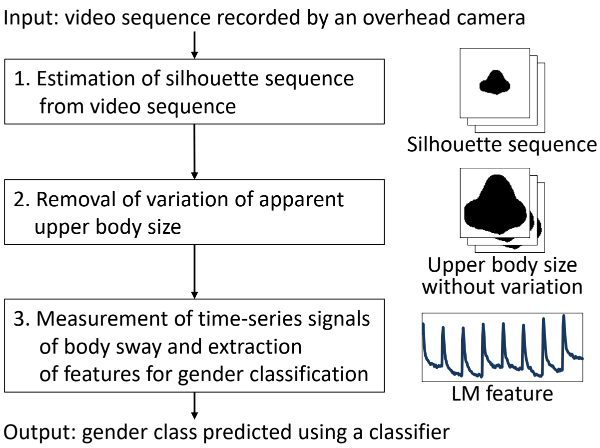


Fig. 2. Overview of proposed three-step method for gender classification. The input is a video sequence containing body sway recorded by an overhead camera and the output is a gender class predicted using a classifier and an extracted feature.

B. Removal of Variation in Apparent Size of Person in Silhouette Sequence

To accurately classify gender, the intra-class variation of appearance should be small. However, the apparent size of the upper body can increase this variation when the height of the overhead camera varies, as described in Section III-A. A silhouette sequence is also affected by the variation of apparent size. Fig. 3 shows examples of the apparent size of silhouette sequences of the upper body. The frames of the silhouette sequences in Figs. 3 (a) and (b) are estimated from

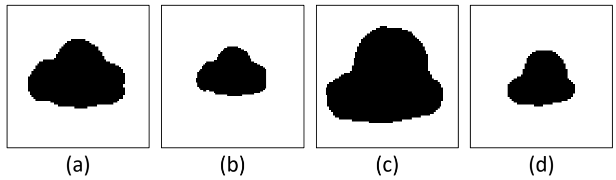


Fig. 3. Examples of silhouette sequence frames estimated from video sequences. The overhead camera was set at different heights. Female recorded by (a) low and (b) high camera. Male recorded by (c) low and (d) high camera. Black and white pixels respectively represent the upper body and background.

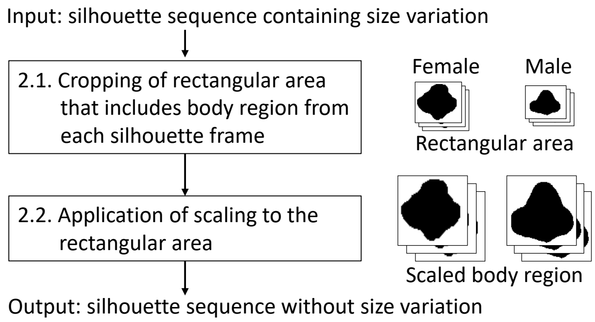


Fig. 4. Removal of the variation of the apparent size of the upper body in a silhouette sequence.

the same female but different camera heights. Although the silhouette sequences both belong to the female class, their apparent sizes of the upper body are different. The same tendency can be seen for the male class in Figs. 3 (c) and (d). If we do not consider the variation of the apparent size, gender classification accuracy will be low because the intra-class variation of appearance will be large.

We thus remove the variation of the apparent size in our method, as shown in Fig. 4. In this step, we crop a rectangular area from each input silhouette frame to remove the background region. The rectangular area includes the upper body and has a margin to prevent cutting off the upper body. We set this margin based on the maximum amount of movement. We determine the margin for each silhouette sequence. Finally, we apply a scaling technique so that the height and width of the rectangular area are equal to the reference values H and W, respectively.

C. Extraction of a Feature from Body Sway for Gender Classification

We now describe the extraction of a spatio-temporal feature from body sway for gender classification. Our method is inspired by the framework of an existing method [14] for person re-identification. Fig. 5 shows the feature extraction step. The upper body in each silhouette frame is radially divided into I local blocks to extract a spatial feature. Subtractions between a reference silhouette frame and silhouette frames are calculated to extract a temporal feature. We now describe the determination of the reference silhouette frame. The distances between all silhouette frames are calculated. The reference silhouette frame with the smallest distance is selected. In each

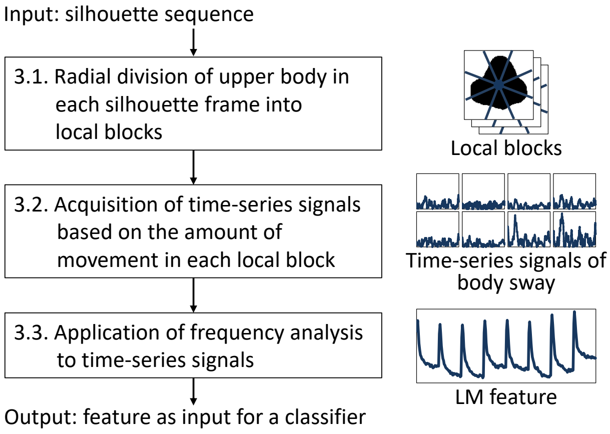


Fig. 5. Feature extraction step. An LM feature is extracted from a silhouette sequence of body sway.

local block, the amount of movement is calculated by summing the absolute values of all subtractions to obtain the time-series signals. Then, a window function of length L is convoluted into the time-series signals of the amount of movement in each local block. The power spectral density (PSD) [27] is estimated from the time-series signals. PSD consists of the component of the power value corresponding to each frequency. The number of components of PSD in each local block is $L/2$. Finally, the PSDs of all local blocks are combined into a feature vector for gender classification. The dimension of the feature is $IL/2$. The vector of PSDs is denoted as a local movement (LM) feature.

IV. Experiment

A. Dataset

We evaluated whether the gender of a standing person can be classified based on a video sequence of body sway recorded by an overhead camera. We acquired video sequences of the body sway of 60 participants (30 females and 30 males). Table I shows the details of the participants. The same instructions were given to all participants. We asked the participants to maintain an upright posture (Romberg's pose), shown in Fig. 6 (a), during the acquisition of their video sequence. We assumed a scenario where people wear the same work clothes in a factory. To reduce the changes in face orientation during the acquisition of a video sequence, we asked all participants to keep looking at a timer placed 3.0 m away. We set the height of the timer at 1.4 m. Fig. 6 (b) shows the experimental setting for the acquisition of video sequences of body sway. We randomly set the height of the overhead camera to between 2.0 and 4.0 m from the floor. The resolution and sampling rate of the overhead camera were 1920×1080 pixels and 30.0 Hz, respectively. We calibrated the overhead camera such that the optical axis coincided with the direction normal to the floor. The internal parameters of the overhead camera were fixed. We set the time length of each video sequence to 60 s. Fig. 7 shows examples of color images of females and males in video sequences acquired in

TABLE I
Details of the participants in our dataset of video sequences containing body sway.

	Female	Male
Number of participants	30	30
Average age (years)	22.4 ± 6.3	21.6 ± 1.3
Average height (cm)	158.7 ± 4.7	170.2 ± 6.4

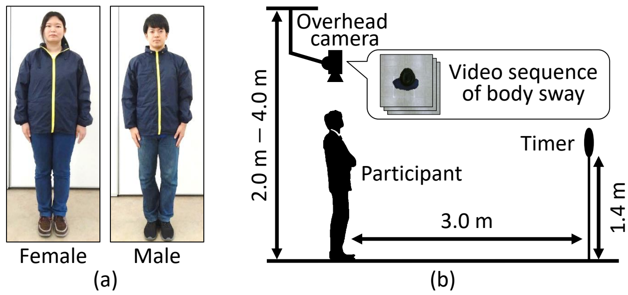


Fig. 6. Experimental setting for observing participants using an overhead camera. (a) Examples of a female and a male standing with an upright posture. (b) Camera setting for acquiring a video sequence of body sway.

our experimental setting. The apparent size of the upper body in the color images varies because of differences in camera height. The inter-class variation between females and males is small even though the intra-class variation of the apparent sizes is large.

B. Evaluation of Gender Classification Accuracy

We compared the accuracy of our method with those of three other methods in our experiment. The details of the methods are as follows.

Proposed method (LM): We used the LM features described in Section III to represent the body sway of a standing person. To extract an LM feature, we set the number of local blocks to $I = 8$ and the length of a window function to $L = 64$ (2.1 s). We estimated a silhouette sequence from a video sequence using a conventional background subtraction technique. We set $H = 100$ pixels and $W = 100$ pixels. We applied a linear support vector machine (SVM) as the classifier and set its regularization parameter to $C = 1.0$.

Alternative method 1 (GEI): We used the GEI [15] features reported in previous studies on the gender classification of a walking person. To extract a GEI feature, we calculated a temporal average image of all frames in a 60-s silhouette sequence. We used the same silhouette sequences as those in our method. We applied a linear SVM as the classifier and set its regularization parameter to $C = 1.0$.

Alternative method 2 (CNN): We used a CNN [22] with single images as a representative of conventional classification techniques. The structure of the CNN consisted of four two-dimensional convolutional layers and four two-dimensional pooling layers. We used 45000 images of females and 45000 images of males as training samples for the CNN. The size of the sample images was set to 100×100 pixels. Each pixel had RGB color values. Binary cross-entropy with the stochastic gradient descent was used.

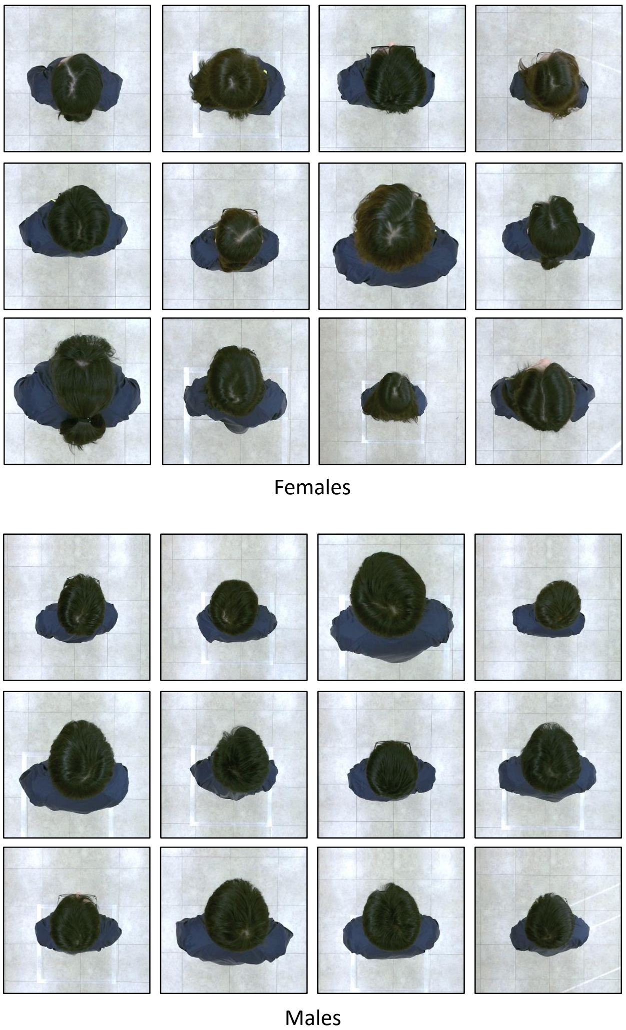


Fig. 7. Examples of color images of females and males in video sequences acquired by an overhead camera. These images show variation in the apparent size of the upper body due to camera height. The inter-class variation between females and males is small even though the intra-class variation of the apparent size is large.

Alternative method 3 (C3D): We used C3D [24] with short video sequences as a representative of spatio-temporal feature extraction. The structure of C3D consisted of four three-dimensional convolutional layers and four three-dimensional pooling layers. We used 2800 short video sequences of females and 2800 short video sequences of males as training samples. Each sample of a video sequence consisted of 16 frames. The size of each frame was set to 100×100 pixels. Each pixel of a frame had RGB color values. Binary cross-entropy with the stochastic gradient descent was used.

We randomly shuffled 60 participants and selected 50 participants as training samples and 10 participants as test samples. We completely separated the participants between the training samples and the test samples. We conducted the random shuffling 30 times and calculated the average and standard deviation of the gender classification accuracy for each method.

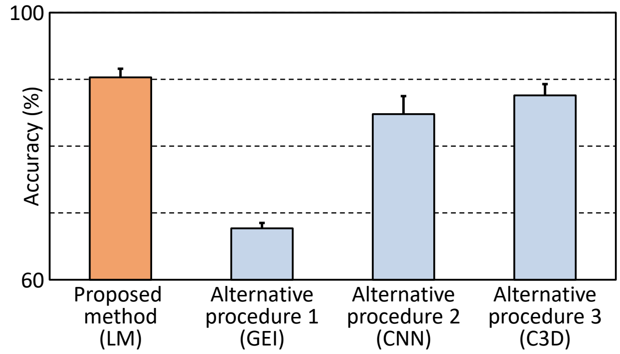


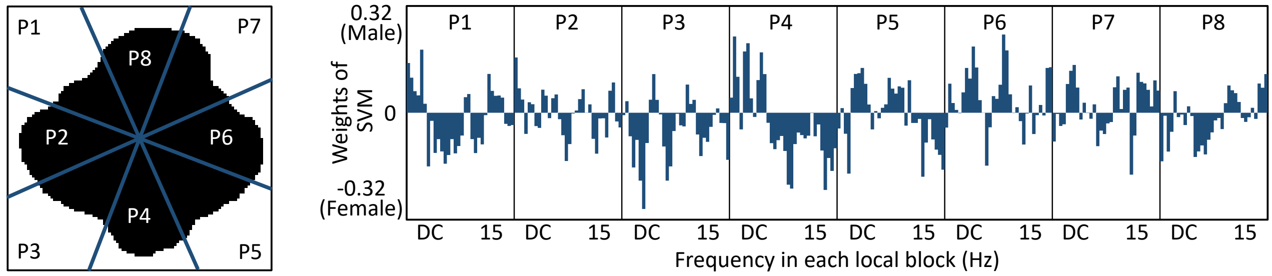
Fig. 8. Comparison of gender classification accuracy obtained using proposed LM, GEI, CNN, and C3D features.

Fig. 8 shows the gender classification accuracy for each method. The accuracy of our method is much higher than that of GEI features. GEI features were designed to represent the large movements of the arms and legs during walking and thus cannot accurately represent the slight movements of a standing person. Conversely, the proposed LM features were designed to represent body sway. They thus have higher gender classification accuracy compared with that of GEI features. Furthermore, the accuracy of our method was superior to that of the CNN. The CNN extracted only spatial features from single images. It was not designed to extract temporal features from movement. We used C3D features to extract spatio-temporal characteristics. The proposed LM features outperform these features. C3D features were not designed to represent body sway (their target is large movements during large movements) and thus have relatively low gender classification accuracy. The proposed LM features include better spatio-temporal characteristics for representing body sway.

C. Visualization of SVM Weights Calculated from LM Features

We visualized the SVM weights calculated when training a gender classifier to determine the most informative component of the proposed LM features for gender classification. We used $I = 8$ local blocks to extract an LM feature (see Section IV-B). The local blocks are labeled P1 to P8 as shown in Fig. 9 (a). Local blocks P2, P4, P6, and P8 correspond to the left hand, back, right hand, and face, respectively. Fig. 9 (b) shows the SVM weights of the LM features corresponding to the local blocks. The horizontal axis represents the frequency in each local block. The extreme left and right on the horizontal axis of each local block represent DC and 15 Hz, respectively. The vertical axis represents the weight of each component. A component with a negative (positive) weight contributes to the classification of females (males). The number of components of an LM feature was $I \times L/2 = 8 \times 64/2 = 256$.

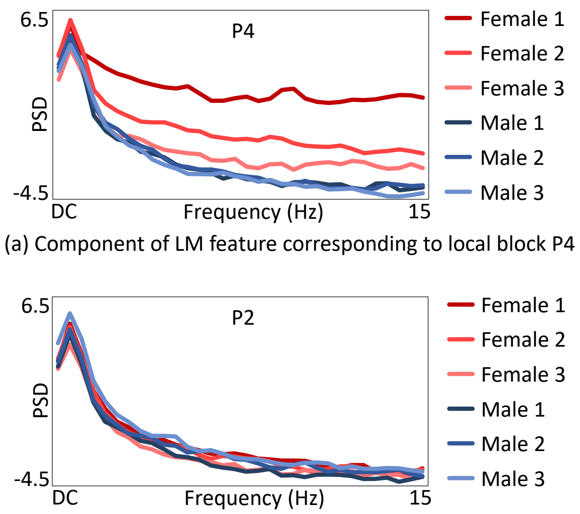
First, we identified the most informative local block for gender classification. We calculated the sum of the absolute SVM weights in each local block. The sum for local blocks P1 to P8 was 2.65, 1.72, 2.51, 3.44, 2.36, 2.36, 2.29, 2.11, and 1.81, respectively. A local block was more informative for gender classification when its sum was higher. Local block P4



(a) Definition of local blocks

(b) Visualization of SVM weights of LM features corresponding to each local block

Fig. 9. Visualization of SVM weights for determining the most informative component of proposed LM features for gender classification. (a) Definition of local blocks P1 to P8. (b) SVM weights of LM features corresponding to the local blocks.



(a) Component of LM feature corresponding to local block P4

(b) Component of LM feature corresponding to local block P2

Fig. 10. Examples of LM features corresponding to local block P2 or P4 for three females and three males.

(corresponding to a person's back) had the highest sum and was thus important for gender classification.

Next, we identified the most informative frequency band in local block P4 for gender classification. The high-frequency band is more informative than the low-frequency band for classifying the female class using the SVM weights corresponding to P4 in Fig. 9 (b). To determine the differences in LM features between females and males, some examples of the features corresponding to P4 are shown in Fig. 10 (a). Examples of those corresponding to P2 are shown in Fig. 10 (b). Local block P2 was not discriminative because it had the lowest sum of absolute SVM weights. We compared the features of P2 with those of P4. In P4, there are large differences in the high-frequency band (3.0 to 15.0 Hz), in which LM features of females are higher than those of males. Conversely, there are no remarkable differences between females and males in P2. We now discuss the reasons for the differences in P4. Local block P4 contained the back of the head, where females often have longer hair. We believe that the differences in the high-frequency band appear because the long hair moved during body sway.

D. Gender Classification Accuracy Obtained using Medical Data

We evaluated the gender classification accuracy of parameters derived from medical data. As described in Section II-C, analytical studies [7], [8], [9] have reported that the differences between females and males can be observed in the frequency characteristics and trajectories of time-series signals. Note that these studies used a force plate to acquire the parameters. In our experiment, we used an overhead camera instead of a force plate to acquire the time-series signals of the center positions of the upper body using silhouette sequences.

We used ten parameters, namely F1 to F6 for frequency characteristics and T1 to T4 for trajectories, reported in previous studies [7], [8], [9]. The parameters were set as follows:

F1 [7]: DC component of the power spectrum of center positions.

F2 [7]: Top accumulated frequency component of the power spectrum of center positions.

F3 [8]: DC component of the power spectrum of velocities.

F4 [9]: Sum of the vertical power spectrum at frequencies lower than 0.2 Hz.

F5 [9]: Sum of the horizontal power spectrum at frequencies lower than 0.2 Hz.

F6 [9]: Sum of the vertical power spectrum at frequencies higher than 2.0 Hz.

T1 [8]: Area of an ellipse approximated to the trajectory of center positions.

T2 [8]: Length of the major axis of the ellipse.

T3 [8]: Length of the minor axis of the ellipse.

T4 [8]: Length of the trajectory of center positions.

We tested each parameter as a one-dimensional feature vector for evaluating gender classification accuracy. We also combined the parameters into a ten-dimensional feature vector (**All**) to improve accuracy. The experimental conditions, except for the features, were the same as those for our method in Section IV-B.

Fig. 11 shows a comparison between the accuracy of each parameter derived from medical data and that of the proposed LM feature. The proposed LM feature had higher accuracy. The accuracy of the combined parameters was higher than that of individual parameters. The results show that the proposed

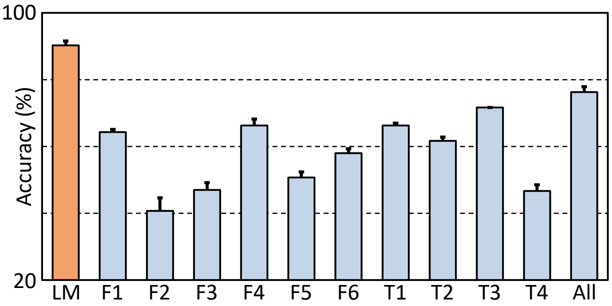


Fig. 11. Accuracy of parameters F1-F6 and T1-T4 derived from medical data and proposed LM feature. 'All' represents a feature that combines all parameters.

LM feature has higher gender classification accuracy than that of parameters derived from medical data.

V. Conclusions

We investigated whether the gender of a standing person can be classified by extracting a feature that represent body sway in a video sequence recorded by an overhead camera. Our method normalizes the apparent size of silhouette sequences of the upper body to remove variation. We divided the upper body into local blocks to represent spatial features and measured the time-series signals of body sway from each local block to represent temporal features. We acquired video sequences containing body sway for 60 participants to evaluate gender classification accuracy. The gender classification accuracy was $90.3\pm1.3\%$ for our spatio-temporal feature. We confirmed that body sway in a video sequence improves gender classification accuracy compared with that for parameters derived from medical data.

In future work, we intend to develop a method for extracting features that represent essential gender differences and are robust against posture changes. We will also investigate whether body sway can be used to classify attributes other than gender, such as age and clothing.

References

[1] X. Wang, S. Zheng, R. Yang, B. Luo, and J. Tang, "Pedestrian attribute recognition: A survey," *CoRR*, 2019.

[2] C. B. Ng, Y. H. Tay, and B. Goi, "Recognizing human gender in computer vision: a survey," *In Proceedings of the Pacific Rim International Conference on Artificial Intelligence*, pp. 335–346, 2012.

[3] S. Khan, M. Nazir, S. Akram, and N. Riaz, "Gender classification using image processing techniques: A survey," *In Proceedings of the IEEE 14th International Multitopic Conference*, pp. 25–30, 2011.

[4] C. Shan, S. Gong, and P. McOwan, "Fusing gait and face cues for human gender recognition," *Neurocomputing*, vol. 71, no. 10-12, pp. 1931–1938, 2008.

[5] R. Martín-Félez, R. Mollineda, and J. Sánchez, "Gender classification from pose-based geis," *In Proceedings of the International Conference on Computer Vision and Graphics*, pp. 501–508, 2012.

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

[7] J. Kim, G. Eom, C. Kim, D. Kim, J. Lee, B. Park, and J. Hong, "Sex differences in the postural sway characteristics of young and elderly subjects during quiet natural standing," *Geriatrics & Gerontology International*, vol. 10, no. 2, pp. 191–198, 2010.

[8] M. Plandowska, M. Lichota, and K. Górniak, "Postural stability of 5-year-old girls and boys with different body heights," *PLoS ONE*, vol. 14, no. 12, pp. 1–10, 2020.

[9] T. Kitabayashi, S. Demura, M. Noda, and T. Yamada, "Gender differences in body-sway factors of center of foot pressure in a static upright posture and under the influence of alcohol intake," *Journal of Physiological Anthropology and Applied Human Science*, vol. 23, no. 4, pp. 111–118, 2004.

[10] F. Wang, M. Skubic, C. Abbott, and J. Keller, "Body sway measurement for fall risk assessment using inexpensive webcams," *In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 2225–2229, 2010.

[11] 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*, pp. 155–162, 2016.

[12] L. Yeung, K. Cheng, C. Fong, W. Lee, and K. Tong, "Evaluation of the microsoft kinect as a clinical assessment tool of body sway," *Gait & Posture*, vol. 40, no. 4, pp. 532–538, 2014.

[13] Z. Lv, V. Penades, S. Blasco, J. Chirivella, and P. Gagliardo, "Evaluation of kinect2 based balance measurement," *Neurocomputing*, vol. 208, pp. 290–298, 2016.

[14] T. Kamitani, H. Yoshimura, M. Nishiyama, and Y. Iwai, "Temporal and spatial analysis of local body sway movements for the identification of people," *IEICE Transactions on Information and Systems*, vol. 102, no. 1, pp. 165–174, 2019.

[15] 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, 2006.

[16] L. Cao, M. Dikmen, Y. Fu, and T. Huang, "Gender recognition from body," *In Proceedings of the 16th ACM International Conference on Multimedia*, pp. 725–728, 2008.

[17] J. Tapia and C. Perez, "Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of lbp, intensity, and shape," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 3, pp. 488–499, 2013.

[18] M. Collins, J. Zhang, P. Miller, and H. Wang, "Full body image feature representations for gender profiling," *In Proceedings of the 12th International Conference on Computer Vision Workshops*, pp. 1235–1242, 2009.

[19] M. Yildirim, O. Ince, Y. Salman, J. Song, J. Park, and B. Yoon, "Gender recognition using hog with maximized inter-class difference," *In Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pp. 106–109, 2016.

[20] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 34–42, 2015.

[21] G. Ozbulak, Y. Aytar, and H. Ekenel, "How transferable are cnn-based features for age and gender classification?" *In Proceedings of the International Conference of the Biometrics Special Interest Group*, pp. 1–6, 2016.

[22] 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*, pp. 1263–1266, 2015.

[23] M. Nishiyama, R. Matsumoto, H. Yoshimura, and Y. Iwai, "Extracting discriminative features using task-oriented gaze maps measured from observers for personal attribute classification," *Pattern Recognition Letters*, vol. 112, pp. 241–248, 2018.

[24] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," *In Proceedings of the IEEE International Conference on Computer Vision*, pp. 4489–4497, 2015.

[25] H. Xu, A. Das, and K. Saenko, "R-c3d: Region convolutional 3d network for temporal activity detection," *In Proceedings of the IEEE International Conference on Computer Vision*, pp. 5783–5792, 2017.

[26] K. Liu, W. Liu, C. Gan, M. Tan, and H. Ma, "T-c3d: Temporal convolutional 3d network for real-time action recognition," *In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pp. 7138–7145, 2018.

[27] 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.