# Joint Swing Energy for Skeleton-Based Gender Classification

**BEOM KWON**[ID]1 **AND SANGHOON LEE**[ID]2, **(Senior Member, IEEE)**
1Network Business, Samsung Electronics Company Ltd., Suwon 16677, South Korea
2Department of Electrical and Electronics Engineering, Yonsei University, Seoul 03722, South Korea

Corresponding author: Sanghoon Lee (slee@yonsei.ac.kr)

**ABSTRACT** Human gender classification in digital image content has received considerable attention from researchers for various applications, such as demographic research, video surveillance systems, and forensic science. In this study, we investigate three-dimensional (3D) human skeleton-based gender classification using a novel gait feature called joint swing energy (JSE). JSE is a kinematic gait feature that represents how distant a model skeleton's body joints are from anatomical planes while walking. However, anatomical planes are conventionally obtained from single static poses rather than dynamic motion. Therefore, in this study, we further investigate a novel method for obtaining transverse, frontal, and median planes from a 3D gait sequence. Using these planes, each joint's movement can be represented by coordinates centered on a human body rather than 3D Cartesian coordinates. Using the proposed methods, we extract JSEs of body joints from 3D gait sequences. We show that JSEs are different between walking men and women and propose the use of JSEs for machine classification of human gender. To demonstrate the effectiveness of the proposed JSE model on the gender classification task, we evaluate the performance of machine learning algorithms trained using JSE on four publicly available datasets, referred to as Datasets A, B, C, and D. Dataset A includes gait sequences from 164 persons between the ages of 17 and 45. Dataset B contains gait sequences from 104 persons between the ages of 17 and 36. In Dataset C, there are gait sequences for 30 persons between the ages of 23 and 55. In Dataset D, gait sequences for 30 persons between the ages of 21 and 55 are contained. The evaluation results demonstrate that the proposed technique achieves the highest classification accuracy and outperforms existing techniques for all datasets. These results suggest that human gender can be classified by JSEs extracted from the 3D gait sequence.

**INDEX TERMS** Feature extraction, gait, gender classification, human skeleton, machine learning.

## I. INTRODUCTION

Gender is one of the most important attributes used to describe people, and carries significant information related to male and female lifestyle characteristics. Gender classification by machine systems is a binary classification problem, with the object of distinguishing the gender of a human represented in digital information. Over the past decades, gender classification methods have been used in the fields of marketing, advertising, and sales strategy [1]–[3]. For example, product displays at stores are mainly based on statistics, including the number of male and female customers expected by date and time, their age, etc. [4]. Because such gendered

The associate editor coordinating the review of this manuscript and approving it for publication was Pengcheng Liu[ID].

product displays can draw more customers into retail stores and increase sales, classifying gender is very important and benefits retail management in presenting products for display effectively.

Furthermore, gender classification methods have been implemented in a wide variety of applications, including forensic science, human-computer interaction, biometrics, surveillance systems, demographic research, and games [5]–[8]. Due to these various applications, gender classification has received considerable attention in various fields of research, including pattern recognition and computer vision. Gender classification is thus considered a fundamental and important classification problem in these fields of research. Previous studies on gender classification have mainly focused on recognizing a person's gender from their

visual attributes, including body shape, eyebrows, footwear, face, clothes, and hairstyle [9]–[13].

For example, Cao *et al.* [9] proposed a method for recognizing human gender from full-body images taken from frontal or posterior views. To achieve this goal, the authors divided each full-body image into $3 \times 3$ blocks. They then extracted the histogram of oriented gradients (HOG) feature proposed in [14] from each block. A random forest algorithm was trained using these HOG features and tested on a publicly available database. According to their results, the random forest algorithm achieved an accuracy of approximately 75%. Dong and Woodard [10] investigated a method for predicting gender from eyebrow images. To achieve this goal, they segmented the eyebrow regions in facial images and extracted three shape features, called global shape, local area, and critical point features from the regions. They then trained a support vector machine (SVM) classifier using these features and demonstrated the feasibility of their classification approach. Yuan *et al.* [11] explored a method to classify gender from footwear images taken from top view. To achieve this goal, the authors extracted HOG features from footwear images and trained an SVM classifier using these features. They evaluated the SVM classifier on their own dataset consisting of 200 footwear images. According to their results, the SVM classifier achieved an accuracy of approximately 85%. Nazir *et al.* [12] proposed a method for classifying gender from facial images containing upper body clothes. To achieve this goal, the authors extracted facial features using a local binary pattern (LBP) descriptor. The authors also extracted clothing features using a color histogram and LBP. They then trained the SVM classifier using these facial and clothing features and evaluated their classification method on a publicly available face image dataset. According to their results, the SVM classifier achieved an accuracy of approximately 98%. Lee and Wei [13] also used facial images for gender classification. From the images, the authors extracted facial texture, hair geometry, and mustache features. Using these features, the authors trained the Adaptive Boosting (AdaBoost) classifier and evaluated it on a publicly available face image dataset. According to their results, the classifier achieved an accuracy of approximately 96%.

In most of the previous visual attribute-based gender classification methods, including [9], [10], [12], [13], it is necessary to use frontal images to classify human gender. Therefore, if a camera does not appropriately capture a front view of a person to be classified, these methods are not guaranteed to maintain the performance they achieved in research settings. Moreover, in a practical situation, part of the human face may be occluded by sunglasses, masks, and hats. In this case, even if a frontal image is provided, most of the features proposed for identification in the previous studies may not be effective.

Recently, along with rapid advancements in motion capture techniques, increasingly more convenient and inexpensive sensors have become commercially available. Enabled by these new sensors, many researchers have acquired three-dimensional (3D) human gait data and exploited them for various applications, including gait recognition, healthcare, personal identification, etc. [15]–[22]. In human gait data, humans are represented by a skeleton model consisting of some major body joints. Moreover, such gait data can be visualized as a temporal sequence of 3D positions of joints. Based on these 3D representations, it is possible to capture gait data even from side and back perspectives in addition to the frontal viewpoint. Because of this advantage, in many studies on gender classification, methods capable of discerning gender based on gait data have been studied. Gender classification in these studies was performed using features extracted from 3D gait sequences. Hence, the extracted features are called gait features. One gait feature widely used in gender classification is stride length, which is generally calculated as the distance between the right and left foot joints. The use of stride length in gender classification is based on the observation that women tend to have a shorter stride length than men [23]. To find other gait features in addition to stride length, many researchers have studied and analyzed human gaits. These studies are reviewed in Section II-C.

In this study, we propose a gender classification method using a new gait feature called joint swing energy (JSE). JSE is a kinematic gait feature that represents how distant the model skeleton's body joints are from anatomical planes while walking. The inspiration for JSE comes from observations that women and men walk differently [24]. According to the results of [24], pelvis and hip ranges of motion (ROM) in the frontal plane differ between women and men. Based on these findings, we decided to test whether ROM can be used for the gender classification task. However, we found that it was difficult to obtain anatomical planes, including transverse, frontal, and median planes, from representations of walking people, because anatomical planes are traditionally obtained from images of a person standing up straight. Therefore, we developed a novel method for obtaining anatomical planes from 3D gait sequence data. In general, ROM is measured by the angle between the joints at the flexed and extended positions. Following this definition of ROM, we measured right and left hips' ROM in frontal plane. Nevertheless, it was inaccurate to calculate the ROM from the ground truth. From the investigation, we found that the flexed (extended) position frequently jerked irregularly during the gait cycle, because the estimated body joints of the skeleton model in the 3D gait sequence contain noise and error, which leads to performance degradation. We also found that the movement of each joint in the transverse, frontal, and median planes while walking differs significanlty between women and men. Based on this observation, we introduced JSE as the average movement of each joint in each plane. JSE can be robust against noise and error because of the averaging process.

As shown in Fig. 1, during the study, we found that JSEs of body joints in walking sequences differ between men and women. Based on this observation, we assumed that JSE could be utilized as a feature to classify human gender.
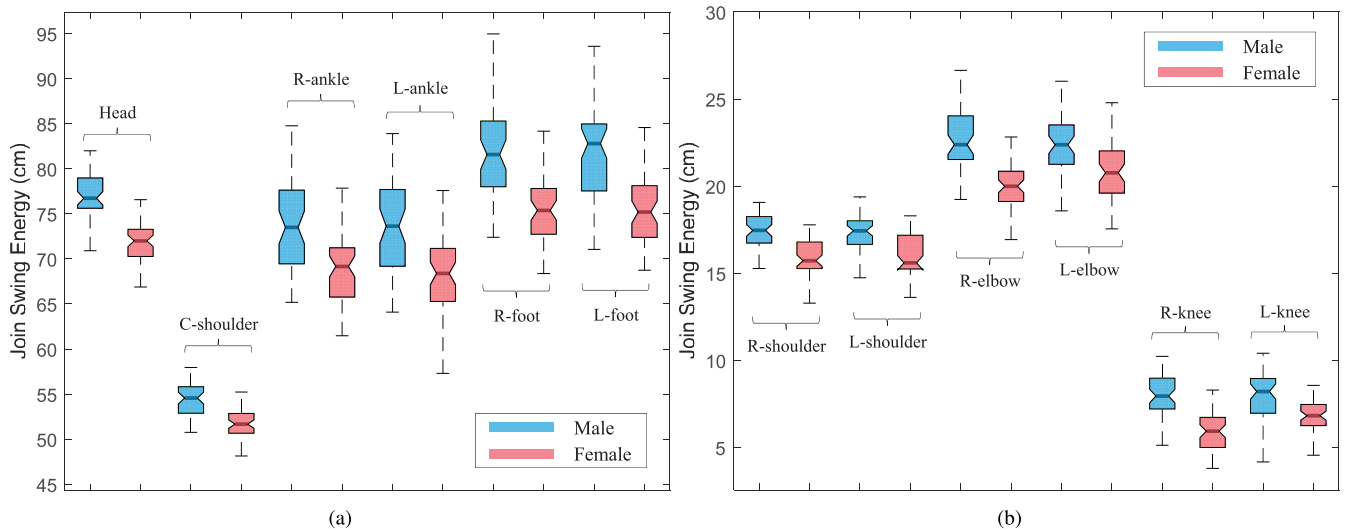
**FIGURE 1.** JSEs on a publicly available gait dataset, called Dataset B in this study. In the dataset, the number of female and male subjects are 54 and 50, respectively. (a) JSE calculated in transverse plane. (b) JSE calculated in median plane. In the joint names, "C," "L," and "R" indicate center, left, and right, respectively. The bottom and top of each box indicate the 25th and 75th percentiles of the sample, respectively. In addition, the black line in the middle of each box indicates the average of the set of samples.

To demonstrate the effectiveness of JSE, we conducted a performance comparison with other benchmarks over four publicly available gait datasets. The results show that the proposed gender classification method achieved the best performance on all datasets.

In summary, the objectives of this study were to (1) find features that differ between men and women from their gait sequences; (2) propose JSE as a new gait feature for classifying human gender; (3) assess the fesability of using JSE in the gender classification task; and (4) demonstrate that the machine learning algorithms using JSE achieve state-of-the-art performance.

The remainder of this paper is organized as follows. In Section II, we present a review of the relevant literature. In Section III, we describe the proposed gender classification method, including JSE feature extraction. In Section IV, we explain the four gait datasets used to evaluate JSE and the evaluation protocol. In Section V, we present the evaluation results of the proposed and existing methods. Finally, we summarize the study and present our conclusions in Section VI.

## II. RELATED WORK

### A. FACIAL IMAGE-BASED GENDER CLASSIFICATION

Research on discerning human gender from facial images goes back to 1990, when there were two studies [25], [26]. In these studies, grayscale images of faces were used, and two-stage networks were proposed to discriminate gender in the images. In the first stage, an autoencoder with a single hidden layer learned to find a compressed representation of the input facial image. In the second stage, hidden units in the hidden layer of the autoencoder were used as inputs to a neural network. The neural network output was compared

with human gender classification judgements, and the results demonstrated the feasibility of two-stage networks.

Motivated by the results of [25] and [26], Tamura *et al.* [27] used a multi-layer neural network to classify gender from grayscale face images at multiple resolutions. According to their results on 8-by-8 images, their proposed network achieved an average error rate of approximately 7%. Gutta *et al.* [28] explored a hybrid approach in which an ensemble of neural networks and decision trees was used to classify gender in grayscale face images. Their evaluation was conducted using the Face Recognition Technology (FERET) database. According to their results, the gender classification accuracy of their hybrid approach was 96%.

Instead of using neural networks, Moghaddam and Yang [29] used an SVM with a radial basis function (RBF) kernel for gender classification. The authors trained the SVM classifier using raw images, evaluating it on the FERET database and comparing the performance with those of traditional classifiers. According to their results, the SVM classifier outperformed the other classifiers and achieved an error rate of 3.38%. Instead of using raw images directly in the training phase, Abdi *et al.* [30] represented each raw image as a feature vector using principal component analysis (PCA). The authors then trained two classifiers, including RBF and perceptron networks, using PCA-based feature vectors. To demonstrate the effectiveness of the vectors in the gender classification task, the authors compared their classifiers with classifiers trained using raw images. According to their evaluation results on 160 facial images, classifiers trained using the PCA-based feature vectors achieved better performance.

Motivated by the results of [30], Sun *et al.* [31] also used PCA-based feature vectors as inputs to classifiers for

gender classification. To exclude irrelevant feature vectors and improve classification performance, the authors proposed a genetic algorithm (GA)-based feature selection method. The authors selected a feature subset from the set of PCA-based feature vectors by using the GA and used the selected subset as input to the classifier. To demonstrate the effectiveness of their method, the authors trained four classifiers, including Bayes, neural network, SVM, and linear discriminant analysis (LDA) classifiers, using the feature vectors selected by GA. According to their results, the error rates of the four classifiers were lower than those of classifiers trained using manually selected feature vectors.

Lian and Lu [32] proposed the use of facial texture information for gender classification. To this end, the authors divided the face area into several small regions and applied an LBP operator into each region to extract texture features. The extracted texture features, called LBP histograms, were concatenated into a single vector and used as input to an SVM classifier. According to their results on the Chinese Academy of Sciences-Pose, Expression, Accessories, and Lighting (CAS-PEAL) database, the SVM classifier achieved an accuracy of 96.75%. Sun *et al.* [33] also used an LBP operator to extract features in facial images. Instead of using the SVM classifier, the authors used the AdaBoost classifier and achieved an accuracy of 95.75% on the FERET database. Face images in the CAS-PEAL and FERET databases were taken under controlled conditions, such as being occlusion-free, having a clean background and consistent lighting, etc. Lian *et al.* [32] and Sun *et al.* [33] used these controlled images in their experiments. However, considering real-world applications, it is reasonable to design gender classification methods to perform well for images captured in uncontrolled conditions. To this end, Shan [34] used the Labeled Faces in the Wild home (LFW) database, which contains real-world face images. The author trained an SVM classifier using an LBP histogram extracted from real face images. According to his results on the LFW database, the SVM classifier achieved an accuracy of approximately 94.8%.

Recently, instead of using hand-crafted features from grayscale images, some research groups have focused on a machine learning methodology to automatically learn features extracted from RGB (that is, color) facial images. In this approach, convolutional neural networks (CNNs) have been mainly used. For example, in [35], a CNN model was proposed to directly learn features from facial images. Islam *et al.* [36] applied three existing CNN models, namely GoogleNet [37], SqueezeNet [38], and ResNet50 [39], to gender classification tasks. In addition, the authors demonstrated the feasibility of these models for classifying genders.

### B. SILHOUETTE IMAGE-BASED GENDER CLASSIFICATION
Some studies on gender classification have used human silhouette images. For example, Lee and Grimson [40] proposed the use of features extracted from silhouettes of the human gait sequence for gender classification. To this end, the authors divided each silhouette into seven regions and fitted

ellipses to each of the regions. For each of the seven ellipses, the centroid, aspect ratio of the major and minor axes, and the orientation of the major axis were calculated and used as features. Huang and Wang [41] extended the use of ellipse-based features to multi-view gender classification. To do so, they extracted the ellipse-based features from each silhouette image of multi-view gait sequences and combined them into a feature vector.

Li *et al.* [42] also used silhouette images of walking people for gender classification. The authors separated each silhouette image into seven components: head, arm, trunk, thigh, front leg, back leg, and feet. Then, they analyzed the contribution of the components to gender classification. They found that trunk and front leg forms appear to be important for gender classification. In contrast, head, back leg, and feet are not helpful for gender classification. Motivated by the results of [42], Yu *et al.* [43] explored an approach in which different weights are given to different components to improve gender classification accuracy. To achieve this goal, the authors carried out psychological experiments. In these experiments, subjects were asked to score the effectiveness of body components for gender classification. The authors also used the gait energy image (GEI) [44], which is defined as the average of silhouettes in a gait sequence. Based on the results of the psychological experiments, they segmented the GEI into five components: head and hair, chest, back, waist and buttocks, and legs. Each of the components was weighted by a corresponding score obtained in the psychological experiments. The authors then trained an SVM classifier using the weighted components. According to their results, the SVM classifier outperformed the methods of [40], [41], and [42] in terms of classification accuracy.

Motivated by the results of [43], Choudhary *et al.* [45], and El-Alfy and Binsaadoon [46] also used the GEI for gender classification. In [45], instead of using the original GEI, Choudhary *et al.* used a GEI with its dimension reduced by a PCA. In addition, the authors extracted spatiotemporal features, including cadence, speed, height, stride length, and stance period, from a gait sequence. In addition, the authors concatenated the spatiotemporal features with the GEI. They then trained SVM and artificial neural network (ANN) classifiers using the concatenated features. The SVM and ANN classifiers were tested on the Institute of Automation, Chinese Academy of Sciences (CASIA) database, and a maximum accuracy of 98.1% was achieved by the SVM classifier. In [46], El-Alfy *et al.* analyzed visual texture with the GEI. To this end, the authors proposed a fuzzy local binary pattern (FLBP) which calculates the relative intensities of each pixel compared to surrounding neighbor pixels. They then extracted a histogram by applying the FLBP operator to the GEI and used the histogram as features for gender classification.

Recently, with advancements in deep learning techniques, CNN-based methods that classify the gender of walking people from their calculated GEIs have emerged. For example, Kitchat *et al.* [47] proposed the creation of a GEI according to

camera angle. To this end, the authors used an existing CNN model, called VGG16 [48], to find the camera angle of an input silhouette image. By using the CNN model, the authors then categorized the silhouette images of a gait sequence into 11 groups according to camera angle. Silhouette images in the same group were used to create the GEIs. They trained the CNN model using GEIs and demonstrated the effectiveness of their method. Bei *et al.* [49] proposed a two-stream CNN that combined the GEI with temporal features for gender classification. The authors divided the whole gait sequence into several sub-sequences and created sub-GEIs for each of the sub-sequences. Based on adjacent sub-GEIs, the authors calculated optical flow maps to represent temporal variations. The authors trained the two-stream CNN using the GEI and optical flow maps and evaluated their method on public databases to demonstrate its effectiveness.

### C. SKELETON-BASED GENDER CLASSIFICATION
Several recent studies on gender classification have explored approaches utilizing information concerning the models of human skeletal joints. The idea that human skeleton information can be used for gender classification was first proposed by Yoo *et al.* in [50]. To this end, the authors used the Southampton (SOTON) database, which contains image sequences of walking people. The authors subtracted the background from each image in the sequence, and then extracted the contours of the body image. To estimate the positions of body joints in the contour, the author used the anatomical properties of human body parts studied in [51]. According to the anatomical properties, the vertical positions of the neck, shoulder, waist, pelvis, knee, and ankle were estimated based on body height. As a result, a two-dimensional (2D) skeleton model consisting of nine body joints was obtained from each image. The authors calculated the joint angles of the hip, knee, and ankle from the model and used them as features for gender classification. Motivated by the results of [50], Kastaniotis *et al.* [52] also explored a method for classifying gender from skeleton sequences. To achieve this goal, they used the Kinect v1 sensor, which provides 3D position information of 20 human body joints. Out of the 20 joints, the authors selected eight. They then calculated two Euler angles for each of the eight joints. To represent the distribution of values of each angle over time, the author constructed 16 histograms of 40 bins each. According to the values of angles calculated at each frame, the corresponding bins of the histograms were increased by one. These histogram-based features were thus 640-dimensional ($= 16 \times 40$). To reduce the dimensionality to 90, the authors used PCA. They trained an SVM classifier using these features and evaluated it on their own dataset. The evaluation results demonstrated the effectiveness of using these features in gender classification.

Andersson *et al.* [53] proposed the use of anthropometric features, including the average length of each body part and average body height, for gender classification. To demonstrate the usefulness of anthropometric features, the

authors tested three machine learning algorithms, including SVM, $k$-nearest neighbor (KNN), and multi-layer perceptron (MLP) classifiers, on their own skeleton dataset. According to their results, the MLP classifier showed the best performance, achieving a gender classification accuracy of approximately 93.1%. Miyamoto and Aoki [54] proposed a method using 3D position information of body joints obtained by the Kinect v2 sensor directly as a feature vector. The lengths of the feature vectors varied according to the length of the input skeleton sequences. The authors applied a linear interpolation to the sequences to make their lengths equal. They then trained an SVM classifier using the feature vectors and evaluated it on their own dataset consisting of twelve people (six males and six females). According to their results, the SVM classifier achieved a classification accuracy of 99.12%. Bachtiar *et al.* [55] captured skeleton sequences for 20 people (ten males and ten females) using the Kinect v1 sensor. Based on the analysis of the sequences, the authors proposed the use of features, including the average distance between right and left hand and foot joints over time.

Ahmed and Sabir [56] used 2D position information of 20 body joints obtained from the Kinect v1 sensor to classify human gender. To achieve this goal, at each frame, the authors calculated features, including step length and horizontal distance between right and left shoulders. In addition, the authors also calculated the vertical heights of several joints, including the head, wrists, shoulders, and ankles. For each, the authors calculated the average, standard deviation, and skewness over the entire frame. Then, the values were concatenated into a single feature vector and used as input to the classifiers.

Camalan *et al.* [57] focused on developing a method to discern human gender using height information. To this end, the authors used human skeleton data captured by the Kinect v1 sensor. At each frame of the sequences, the authors calculated the body height of an individual in three different ways. The first height, called height by skeleton (HbS), was calculated as the total length from foot to head. The second height, called height by estimation (HbE), was estimated by using the following formula. ''the distance between shoulders and knees $= 0.52 \times$ height.'' The third height, called height by wingspan (HbW), was calculated based on the idea that a person's wingspan (also referred to as arm span) is approximately equal to their height. Each of the three heights was averaged over the entire frame. Then, the average values were used as the input of classifiers.

Our work belongs to the category of skeleton-based gender classification. In the first category (i.e., facial image-based gender classification), most of the existing methods are limited in real-world applications because frontal face images are required to classify human gender. Furthermore, if some parts of the human face are occluded by masks, sunglasses, or hats, most of the methods may not be effective in gender classification. In the second category (i.e., silhouette image-based gender classification), most of the previous methods used GEIs for classifying human gender. In general, GEIs
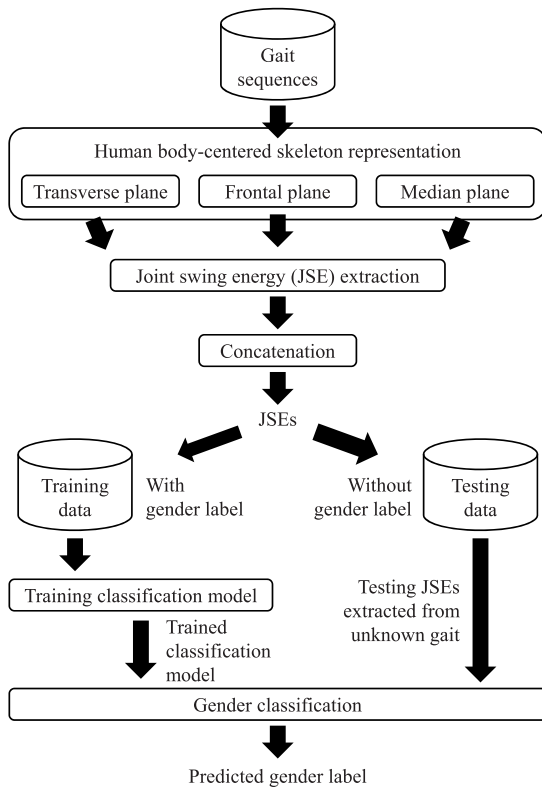
**FIGURE 2.** Overall process of the proposed gender classification method.



**FIGURE 3.** 3D human skeleton model and with joint information. "C," "L," and "R" indicate center, left and right, respectively.

are obtained from full-body images taken from side views. Therefore, if a camera does not capture a side view of person to be classified, most of the methods also are not guaranteed to maintain the performance they achieved in research settings. To avoid these drawbacks, in this study, we explore the use of 3D human skeleton on the gender classification task. During the study, we found that JSEs of body joints in walking sequences differ between men and women. Based on this observation, we propose a gender classification method using the JSE feature. In the next section, we explain the proposed method in details.

## III. PROPOSED METHOD

### A. OVERALL PROCESS

Fig. 2 shows the overall process of the proposed gender classification method. The process of gender classification is divided into three phases: (1) human body-centered skeleton representation, (2) JSE extraction, and (3) classification. In the first phase, from the gait sequence of a walking skeleton model, human body-centered coordinates, called transverse, frontal, and median planes, are calculated. In the second phase, the proposed JSE is extracted for each of the three planes. Then, the extracted JSEs are concatenated into a feature vector. In the final phase, the classification model is trained using the labeled training JSE data. After the training is completed, using the trained classification model, gender classification is conducted on the unlabeled testing JSE data.
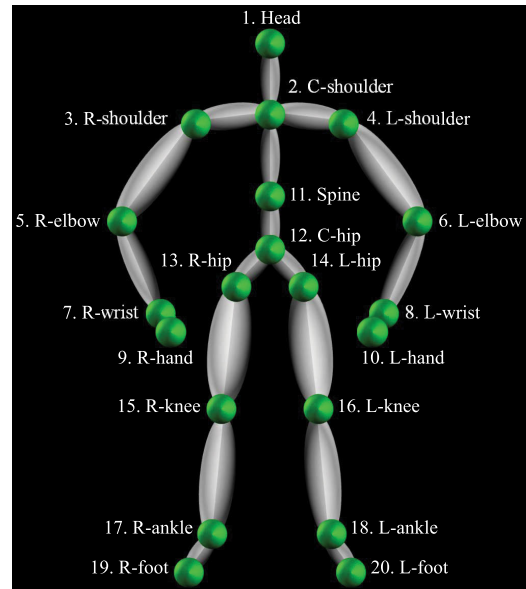
### B. HUMAN BODY-CENTERED SKELETON REPRESENTATION

In general, the 3D human skeleton is represented as 3D coordinates of $N_J$ points in a 3D Cartesian coordinate system. Here, each point indicates the position of the corresponding joint of the skeleton model. Let $x_j[n]$, $y_j[n]$, and $z_j[n]$ be the x-, y-, and z-coordinates of the $j^{th}$ joint at the $n^{th}$ frame, respectively. In addition, let $P_j[n] = (x_j[n], y_j[n], z_j[n])$ be the position of the $j^{th}$ joint at the $n^{th}$ frame in the 3D Cartesian coordinate system. The value of $N_J$ depends on the product specification of the motion capture sensors. For the Kinect v1 sensor, $N_T$ was 20. Fig. 3 shows the details of the joint information of the Kinect skeleton. In this study, based on the joint information in the figure, we introduce the proposed gender classification method.

The gait sequence is represented as a temporal stream of $P_j[n]$ for all joints. In practical situations, people would be reasonably expected to walk freely through the field of view of a motion capture sensor. For example, some people would walk from right to left, and vice versa. Therefore, even if the gait sequences are obtained from a single person, they may not be the same in terms of the walking direction. As shown in Fig. 4(b), to extract a viewpoint invariant gait feature from such sequences, we use anatomical planes that transect the human body. Using the anatomical planes, the skeleton's movement can be represented in human body-centered coordinates rather than in conventional 3D Cartesian coordinates. Based on this human body-centered skeleton representation, we analyze the positions of body joints and the direction of their movements. However, conventional anatomical planes are generally obtained from a human who stands up straight. To the best of our knowledge, a method to obtain anatomical planes from a walking person has not yet been developed.
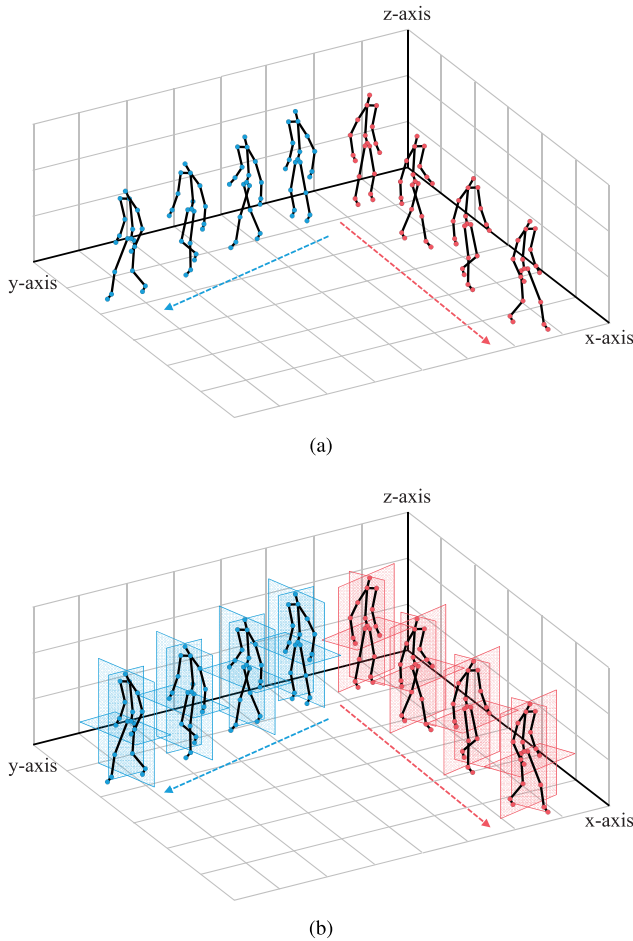
**FIGURE 4.** Illustration of two human skeleton models walking in different directions. (a) Original input gait sequences represented in 3D Cartesian coordinate system. (b) Human body-centered skeleton representation using anatomical planes.

Therefore, in this study, we propose a method for obtaining anatomical planes including transverse, frontal, and median planes, for the general skeleton model, as shown in Fig. 5.

Fig. 5(a) shows the transverse plane of the 3D human skeleton model. In this study, we define the transverse plane at the $n^{th}$ frame as a plane containing the point $P_{12}[n] = (x_{12}[n], y_{12}[n], z_{12}[n])$ with a normal vector $\vec{v_1} = (0, 0, 1)$. Then, the equation of the transverse plane at the $n^{th}$ frame can be written as

$$
\begin{aligned}
&\langle \vec{v_1}, (x - x_{12}[n], y - y_{12}[n], z - z_{12}[n]) \rangle \\
&= \langle (0, 0, 1), (x - x_{12}[n], y - y_{12}[n], z - z_{12}[n]) \rangle \\
&= 0 \cdot (x - x_{12}[n]) + 0 \cdot (y - y_{12}[n]) + 1 \cdot (z - z_{12}[n]) \\
&= z - z_{12}[n] = 0,
\end{aligned}
\tag{1}
$$

where $\langle , \rangle$ is the inner product operator.

As shown in Fig. 5(b), the frontal plane is defined as a plane that passes through the points $P_{13}[n] = (x_{13}[n], y_{13}[n], z_{13}[n])$ and $P_{14}[n] = (x_{14}[n], y_{14}[n], z_{14}[n])$ and is perpendicular to the xy-plane. Let $\vec{v_2}[n]$ be the normal vector of the frontal plane at the $n^{th}$ frame. Then, $\vec{v_2}[n]$ can be

found as

$$
\begin{aligned}
\vec{v_2}[n] &= \overrightarrow{P_{13}[n]P_{14}[n]} \times \vec{v_1} \\
&= (x_{14}[n] - x_{13}[n], y_{14}[n] - y_{13}[n], z_{14}[n] - z_{13}[n]) \\
&\quad \times (0, 0, 1) \\
&= (y_{14}[n] - y_{13}[n], x_{13}[n] - x_{14}[n], 0),
\end{aligned}
\tag{2}
$$

where $\times$ is the cross product operator. Then, the equation of the frontal plane at the $n^{th}$ frame can be written as follows.

$$
\begin{aligned}
&\langle \vec{v_2}[n], (x - x_{12}[n], y - y_{12}[n], z - z_{12}[n]) \rangle \\
&= (y_{14}[n] - y_{13}[n])(x - x_{12}[n]) \\
&\quad + (x_{13}[n] - x_{14}[n])(y - y_{12}[n]) + 0 \cdot (z - z_{12}[n]) \\
&= (y_{14}[n] - y_{13}[n])x + (x_{13}[n] - x_{14}[n])y \\
&\quad + x_{12}[n](y_{13}[n] - y_{14}[n]) + y_{12}[n](x_{14}[n] - x_{13}[n]) \\
&= 0.
\end{aligned}
\tag{3}
$$

As shown in Fig. 5(c), the median plane is defined as a plane containing the point $P_{12}[n]$ with a normal vector $\overrightarrow{P_{13}[n]P_{14}[n]}$. Then, the equation of the median plane at the $n^{th}$ frame can be written as follow.

$$
\begin{aligned}
&\langle \overrightarrow{P_{13}[n]P_{14}[n]}, (x - x_{12}[n], y - y_{12}[n], z - z_{12}[n]) \rangle \\
&= (x_{14}[n] - x_{13}[n])x + (y_{14}[n] - y_{13}[n])y \\
&\quad + (z_{14}[n] - z_{13}[n])z + x_{12}[n](x_{13}[n] - x_{14}[n]) \\
&\quad + y_{12}[n](y_{13}[n] - y_{14}[n]) + z_{12}[n](z_{13}[n] - z_{14}[n]) \\
&= 0.
\end{aligned}
\tag{4}
$$

### C. JOINT SWING ENERGY (JSE) EXTRACTION

We propose a new kinematic gait feature, called JSE. A JSE indicates how distant the skeleton's joints are from the three anatomical planes while walking. Fig. 6 shows the overall process of JSE extraction. For explanation, let $D_{j,T}[n]$ be the shortest distance from the $j^{th}$ joint to the transverse plane at the $n^{th}$ frame. Then, $D_{j,T}[n]$ can be calculated as

$$
\begin{aligned}
D_{j,T}[n] &= \frac{|0 \cdot x_j[n] + 0 \cdot y_j[n] + 1 \cdot z_j[n] - z_{12}[n]|}{\sqrt{0^2 + 0^2 + 1^2}} \\
&= |z_j[n] - z_{12}[n]|,
\end{aligned}
\tag{5}
$$

where $|\ |$ is the operator that returns the absolute value of its argument.

Let $R_{j,T}$ be the average distance from the $j^{th}$ joint to the transverse plane over the entire frame $N_F$. Using $D_{j,T}[n]$ in (5), $R_{j,T}$ is defined as

$$
R_{j,T} = \frac{1}{N_F} \sum_{n=1}^{N_F} D_{j,T}[n].
\tag{6}
$$

Let $D_{j,F}[n]$ be the shortest distance from the $j^{th}$ joint to the frontal plane at $n^{th}$ frame. Then, $D_{j,F}[n]$ can be calculated as

$$
D_{j,F}[n] = \frac{|(y_{14} - y_{13})(x_j - x_{12}) + (x_{13} - x_{14})(y_j - y_{12})|}{\sqrt{(y_{14} - y_{13})^2 + (x_{13} - x_{14})^2}}.
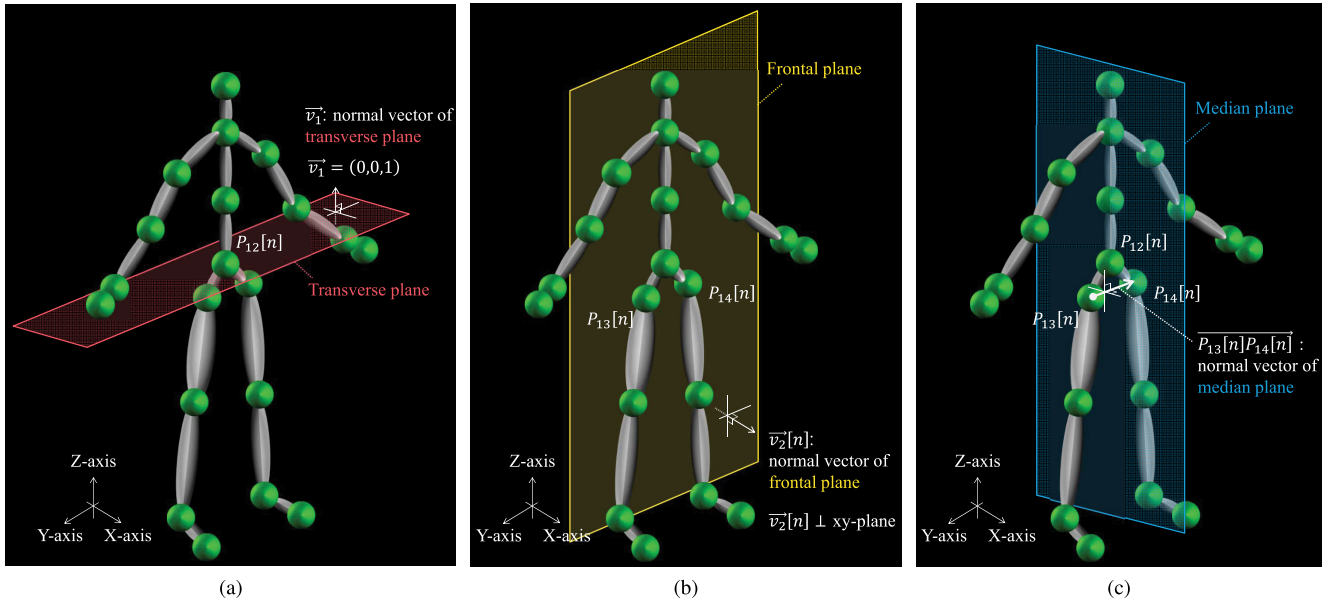\tag{7}
$$

**FIGURE 5.** Three anatomical planes of 3D human skeleton: (a) transverse plane (red), (b) frontal plane (yellow), and (c) median plane (blue).
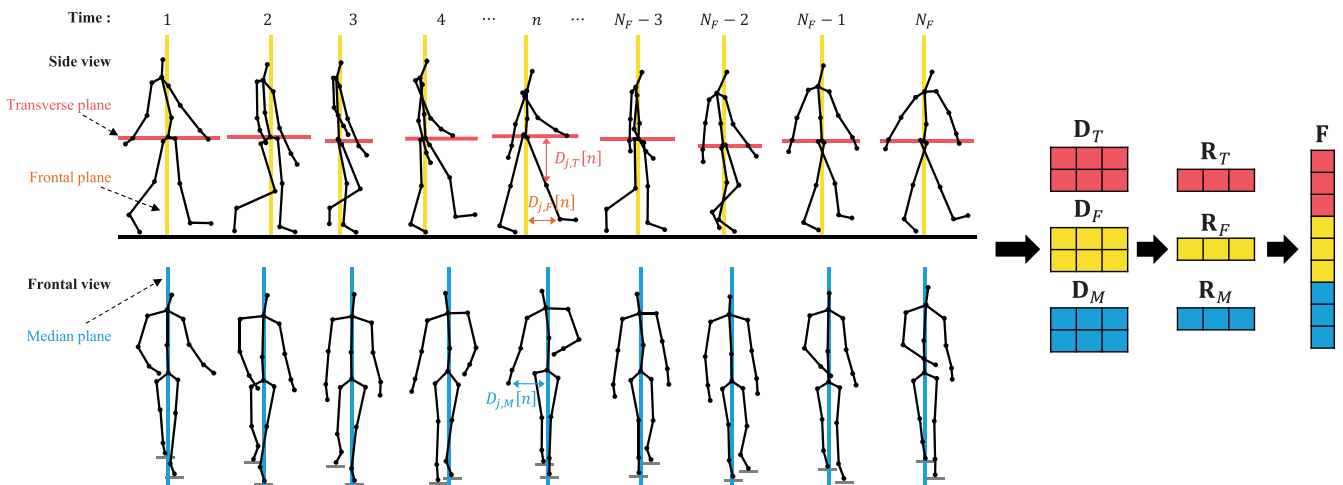


**FIGURE 6.** Overall process of JSE extraction. For clarity, we use two viewpoints: side and frontal. From the side view, the transverse and frontal planes each look like a line. From the frontal view, the median plane looks like a line. The red, yellow, and blue lines indicate transverse, frontal, and median planes, respectively.

In (7), for brevity, we use the simplified x-, y-, and z-coordinates $(x_{13}, y_{13}, z_{13}, x_{14}, y_{14}, z_{14}, x_j, y_j, z_j)$ after omitting the frame index $[n]$, from $(x_{13}[n], y_{13}[n], z_{13}[n], x_{14}[n], y_{14}[n], z_{14}[n], x_j[n], y_j[n], z_j[n])$.

Let $R_{j,F}$ be the average distance from the $j^{th}$ joint to the frontal plane over $N_F$ using $D_{j,F}[n]$ in (7). Then $R_{j,F}$ is defined as

$$R_{j,F} = \frac{1}{N_F} \sum_{n=1}^{N_F} D_{j,F}[n]. \qquad (8)$$

Let $D_{j,M}[n]$ be the shortest distance from the $j^{th}$ joint to the median plane at the $n^{th}$ frame. Then, $D_{j,M}[n]$ can be calculated as

$$D_{j,M}[n] = \frac{\left| (x_{14}-x_{13})x_j + (y_{14}-y_{13})y_j + (z_{14}-z_{13})z_j + d \right|}{\sqrt{(x_{14}-x_{13})^2 + (y_{14}-y_{13})^2 + (z_{14}-z_{13})^2}}, \qquad (9)$$

where $d = x_{12}[n](x_{13}[n] - x_{14}[n]) + y_{12}[n](y_{13}[n] - y_{14}[n]) + z_{12}[n](z_{13}[n] - z_{14}[n])$. In (9), for brevity, we use the simplified x-, y-, and z-coordinates $(x_{13}, y_{13}, z_{13}, x_{14}, y_{14}, z_{14}, x_j, y_j, z_j)$ after omitting the frame index $[n]$, from $(x_{13}[n], y_{13}[n], z_{13}[n], x_{14}[n], y_{14}[n], z_{14}[n], x_j[n], y_j[n], z_j[n])$.

Let $R_{j,M}$ be the average distance from the $j^{th}$ joint to the median plane over $N_F$. Using $D_{j,M}[n]$ in (9), $R_{j,M}$ is defined

**FIGURE 7.** Flowchart of the JSE extraction.

as

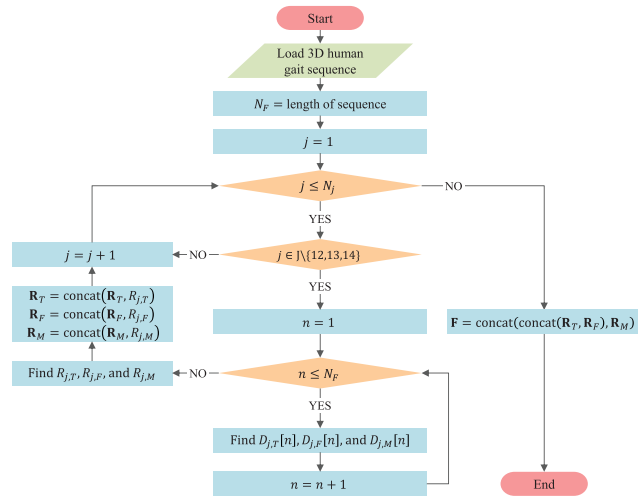$$R_{j,M} = \frac{1}{N_F} \sum_{n=1}^{N_F} D_{j,M}[n]. \qquad (10)$$

Considering the definition of the transverse, frontal, and median planes, regardless of gender, $D_{12,T}[n] = 0$, $D_{13,F}[n] = 0$, $D_{14,F}[n] = 0$, $D_{12,M}[n] = 0$ for all $n$. Therefore, for all joints except the twelfth (i.e., C-hip), thirteenth (i.e., R-hip), and fourteenth (i.e., L-hip) joints, we calculate $R_{j,T}$ in (6), $R_{j,F}$ in (8), and $R_{j,M}$ in (10) and then represent them as vectors:

$$R_T = \left[ R_{j,T} \text{ for } j \text{ in } \mathbb{J} \setminus \{12, 13, 14\} \right], \qquad (11)$$
$$R_F = \left[ R_{j,F} \text{ for } j \text{ in } \mathbb{J} \setminus \{12, 13, 14\} \right], \qquad (12)$$
$$R_M = \left[ R_{j,M} \text{ for } j \text{ in } \mathbb{J} \setminus \{12, 13, 14\} \right], \qquad (13)$$

where $\mathbb{J}$ is the set of joint indices (i.e., $\mathbb{J} = \{1, 2, \cdots, N_T - 1, N_T\}$) and $\setminus$ in $\mathbb{A} \setminus \mathbb{B}$ is an operator that returns the relative complement of $\mathbb{B}$ in $\mathbb{A}$. Then, each of $R_T$, $R_F$, and $R_M$ are 17-dimensional.

Using $R_T$ in (11), $R_F$ in (12), and $R_M$ in (13), we define the proposed JSE feature vector as

$$F = \text{concat}(\text{concat}(R_T, R_F), R_M), \qquad (14)$$

where $\text{concat}(I_1, I_2)$ is an operator that concatenates a vector $I_2$ to the end of a vector $I_1$. The dimension of $F$ in (14) is 51. The procedure for generating the JSE feature vector is described by the pseudocode in Algorithm 1. In addition, Fig. 7 shows the flowchart of the proposed JSE extraction.

## IV. DATASETS AND EVALUATION PROTOCOL

In this study, we evaluated the performance of our proposed gender classification method using the following four datasets:

- Dataset A: This dataset [58] provides both gait data and gender information for 164 subjects. Among the 164 subjects, 51 were female (31.1%) and 113 were

---

**Algorithm 1** Pseudocode for Generating the JSE Feature Vector

**Input:** 3D human gait sequence consisting of $N_F$ frames, where each frame is composed of a row vector of 3D coordinates of $N_j$ joints.

**Output:** JSE feature vector $F$

1: Create a $N_F$-by-1 column vector of zeros $D_T$
2: Create a $N_F$-by-1 column vector of zeros $D_F$
3: Create a $N_F$-by-1 column vector of zeros $D_M$
4: Create an empty vector $R_T$
5: Create an empty vector $R_F$
6: Create an empty vector $R_M$
7: **for** $j$ in $(\mathbb{J} \setminus \{12, 13, 14\})$ **do**
8:     **for** $n$ in range$(1, N_F + 1, 1)$ **do**
9:         Calculate $D_{j,T}[n]$ in (5)
10:         $D_T[n] = D_{j,T}[n]$
11:         Calculate $D_{j,F}[n]$ in (7)
12:         $D_F[n] = D_{j,F}[n]$
13:         Calculate $D_{j,M}[n]$ in (9)
14:         $D_M[n] = D_{j,M}[n]$
15:     **end for**
16:     Using $D_T$, calculate $R_{j,T}$ in (6) as mean$(D_T)$
17:     $R_T = \text{concat}(R_T, R_{j,T})$
18:     Using $D_F$, calculate $R_{j,F}$ in (8) as mean$(D_F)$
19:     $R_F = \text{concat}(R_F, R_{j,F})$
20:     Using $D_M$, calculate $R_{j,M}$ in (10) as mean$(D_M)$
21:     $R_M = \text{concat}(R_M, R_{j,M})$
22: **end for**
23: $F = \text{concat}(\text{concat}(R_T, R_F), R_M)$
24: **return** $F$

---

male (68.9%). The gait sequences for each subject were recorded five times using the Kinect v1 sensor. However, there are only three sequences for one subject denoted as "Person015." For three subjects denoted as "Person002," "Person158," and "Person164," there are only four sequences. In addition, there are six sequences for seven subjects denoted as "Person003," "Person034," "Person036," "Person052," "Person053," "Person074," and "Person096." As a result, the dataset contains a total of 822 sequences (i.e., 7 people × 6 sequences per person + 153 people × 5 sequences per person + 3 people × 4 sequences per person + 1 person × 3 sequences per person).

- Dataset B: This dataset [59] provides both gait data and gender information for 104 subjects. Among the 104 subjects, 54 were female (52%) and 50 were male (48%). Compared with Dataset A, Dataset B is well balanced in terms of the proportion of women and men. The gait data for each subject were recorded once using the Kinect v2 sensor. In addition, during the recording, each subject was asked to walk for approximately 30 seconds at a speed of approximately 1.2 m/s over a treadmill. For a fair comparison with the results on the other datasets where the Kinect v1 sensor was used,

we used only the 20 joints corresponding to the joints in the Kinect v1 skeleton.

- Dataset C: This dataset, called UPCVgaitK1 [52], [60], contains gait sequences from 30 subjects. Among the 30 subjects, 15 were female (50%) and 15 were male (50%). The gait sequences for each subject were recorded five times using the Kinect v1 sensor. As a result, the dataset contains a total of 150 sequences (i.e., 30 people, × 5 sequences per person).

- Dataset D: This dataset, called UPCVgaitK2 [61], [62], contains gait sequences from 30 subjects. Among the 30 subjects, 11 were female (36.7%) and 19 were male (63.3%). The gait sequences for each subject were recorded ten times using the Kinect v2 sensor. As a result, the dataset contains a total of 300 sequences (i.e., 30 people × 10 sequences per person). For a fair comparison with the results on the other dataset where the Kinect v1 sensor was used, we used only the 20 joints corresponding to the joints in the Kinect v1 skeleton.

For each dataset, we used leave-one-participant-out (LOPO) cross validation to assess the performance of the proposed method. The reason of using LOPO cross validation is to avoid that the same participants are included in both the training and testing data. If k-fold cross validation is carelessly used (i.e., training and testing data are carelessly splitted, the same participants can be included in both the training and testing data. In this case, machine learning algorithms may show artificialy high results. For example, suppose that, for Dataset C, a total number of 120 gait sequences (= 30 subjects × 4 sequences per subject) are used for the training data. In addition, the remaining 30 gait sequences (= 30 subjects × 1 sequence per subject) are used for the testing data. In this case, all subjects are both in the training and testing data. If a machine learining algorithm is trained using this training data and tested on the testing data, the algorithm may artificially achieve a high performance [52], [57], [63], [64]. To avoid this situation, in this study, we used LOPO cross validation. In each validation round, gait sequences for one person were used to test the model, while the sequences for the other people were used for training. As a result, many trained-classifiers were generated during the LOPO cross validation procedure. For example, for Dataset A, a total of 656 classifiers (= 4 classifiers × 164 rounds) were generated and tested. Since it was difficult to describe the hyperparameter values for all the models, we decided to provide all the trained models in our web-page. Our code and trained models are available at https://sites.google.com/view/beomkwon/gender-classification. In the web-page, we presented a step-by-step explanation of how to run our code.

## V. EXPERIMENTAL RESULTS
### A. PERFORMANCE MEASURES
Table 1 shows the gender classification confusion matrix used in this study. In this table, TP, FP, FN, and TN represent the

**TABLE 1.** Gender classification confusion matrix.

| Predicted \ True | Female | Male |
|---|---|---|
| Female | True positive (TP) | False positive (FP) |
| Male | False negative (FN) | True negative (TN) |

number of true positives, false positives, false negatives, and true negatives, respectively. By using these values, the gender classification accuracy (ACC) is calculated as

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \times 100\%. \quad (15)$$

Additionally, since the proportion of females and males in Datasets A and D were imbalanced, we also used the following statistical measures: true positive rate (TPR), positive predictive value (PPV), true negative rate (TNR), and negative predictive value (NPV). Here, the TPR is defined as

$$TPR = \frac{TP}{TP + FN} \times 100\%. \quad (16)$$

The PPV is defined as

$$PPV = \frac{TP}{TP + FP} \times 100\%. \quad (17)$$

The TNR is defined as

$$TNR = \frac{TN}{FP + TN} \times 100\%. \quad (18)$$

The NPV is defined as

$$NPV = \frac{TN}{FN + TN} \times 100\%. \quad (19)$$

### B. PERFORMANCE COMPARISON
To demonstrate the effectiveness of the proposed feature extraction method for gender classification, we compared the proposed method with four existing feature extraction methods. To this end, we implemented the four methods using MATLAB R2018a. For brevity, we named them F1, F2, F3, and F4.

- F1: This method is from Andersson *et al.* [53]. In this method, anthropometric features, including the average length of each body part and average height, are extracted from the subject's skeleton sequence.
- F2: This method is from Camalan *et al.* [57]. In this method, three features related to the subject's height, called HbS, HbE, and HbW, are extracted from the skeleton sequence.
- F3: This method is from Bachtiar *et al.* [55]. In this method, the average width between the right and left feet is extracted from the gait sequence. In addition, the average distance between the right and left hands was also extracted.
- F4: This method is from Ahmed and Sabir [56]. In this method, at each frame, the horizontal distance between the right and left shoulders is calculated. The step length, defined as the horizontal distance between the right and left ankles, is also calculated at each frame. In addition, the vertical heights are extracted for the joints of the

head, wrists, shoulders, and ankles. Then, for each, the average, standard deviation, and skewness are calculated over the whole frames and used as features.

The effectiveness of each feature extraction method is evaluated using four common machine learning algorithms: KNN, Naive Bayes (NB), SVM, and decision tree (DT) models. To this end, we implemented the four algorithms using the MATLAB functions "fitcknn," "fitcnb," "fitcsvm," and "fitctree," respectively. The hyperparameter configuration of each machine learning model was optimized using a hyperparameter optimization process supported by these functions. In the process, Bayesian optimization was used.

**TABLE 2.** Gender classification accuracy comparison between the proposed method and the existing methods.

| Techniques | Dataset A 164 Users | Dataset B 104 Users | Dataset C 30 Users | Dataset D 30 Users |
|---|---|---|---|---|
| **JSE-KNN** | **86.50** | 96.15 | **100.00** | 92.67 |
| **JSE-NB** | 80.41 | 89.42 | 96.00 | **94.33** |
| **JSE-SVM** | 84.18 | **98.08** | 96.67 | 92.67 |
| **JSE-DT** | 84.67 | 90.38 | 87.33 | 88.67 |
| F1-KNN | 68.13 | 93.27 | 96.00 | 76.00 |
| F1-NB | 74.94 | 92.31 | 88.67 | 81.33 |
| F1-SVM | 72.02 | 92.31 | 90.00 | 78.33 |
| F1-DT | 68.49 | 92.31 | 88.67 | 82.33 |
| F2-KNN | 73.11 | 89.42 | 78.67 | 77.00 |
| F2-NB | 76.76 | 81.73 | 85.33 | 74.33 |
| F2-SVM | 72.02 | 89.42 | 80.00 | 68.00 |
| F2-DT | 74.70 | 80.77 | 77.33 | 75.67 |
| F3-KNN | 71.90 | 61.54 | 85.33 | 71.00 |
| F3-NB | 72.02 | 67.31 | 82.00 | 79.00 |
| F3-SVM | 70.56 | 69.23 | 82.00 | 71.00 |
| F3-DT | 69.71 | 66.35 | 72.67 | 69.00 |
| F4-KNN | 68.49 | 84.62 | 88.00 | 74.67 |
| F4-NB | 71.17 | 84.62 | 86.67 | 64.67 |
| F4-SVM | 72.26 | 83.65 | 88.67 | 88.33 |
| F4-DT | 65.69 | 83.65 | 92.00 | 82.00 |

Table 2 shows the average classification accuracies of the proposed method and the existing methods (F1, F2, F3, and F4) in all datasets. In the table, the gender classification techniques are named according to the feature extraction method and classification model used. For example, in JSE-KNN, the proposed JSE feature extraction and KNN classifier were used. As another example, in F1-SVM, one of the existing feature extraction methods, F1, and an SVM classifier were used. The experimental results show that the proposed feature extraction method achieves the highest classification accuracy for all datasets. On Datasets A and C, the proposed technique, JSE-KNN, achieved the highest classification accuracy among all the techniques. On Dataset B, the proposed technique, JSE-SVM, outperformed the other techniques, while achieving a classification accuracy of 95.19%. In Dataset D, the proposed technique, JSE-NB, achieved the best performance.

Table 3 shows the TPR, PPV, TNR, and NPV of each technique in all datasets. In this table, TPR is a measure used to indicate the ability to correctly identify the gender of women as female. PPV indicates the ratio of correctly identified females to all females classified. TNR is a measure

used to indicate the ability to correctly identify the gender of men as male. NPV indicates the ratio of correctly identified males to all males classified.

- In Dataset A, the proposed technique, JSE-KNN, achieves the highest TPR among all the techniques. For PPV, the proposed technique, JSE-SVM, significantly outperforms the other techniques, while achieving a PPV of 82.5%. One can also see that the PPVs of the proposed techniques, JSE-KNN and JSE-DT, are greater than 80%. On the other hand, the PPVs of most of the existing techniques (except only F1-NB, F2-NB, and F2-DT) are lower than 60%. The proposed technique, JSE-SVM, achieves the highest TNR among all the techniques. Among the existing techniques, F3-SVM achieves a good TNR of 90.75%. For NPV, JSE-KNN outperformed the other techniques. The NPVs of our four techniques (JSE-KNN, JSE-NB, JSE-SVM, and JSE-DT) were above 90%. However, the NPVs of most of the existing techniques (except only F1-NB and F2-NB) are below 90%.

- In Dataset B, the proposed techniques, JSE-KNN and JSE-SVM, achieved the best TPR of 100%. Among the existing techniques, F1-SVM achieved the highest TPR of 94.44%. For PPV, achieving a PPV of 96.43%, our JSE-SVM outperforms the other techniques. In addition, F1-KNN comes in second with a PPV of 94.34%. For TNR, our JSE-SVM achieved the highest TNR of 96% among the techniques, while the lowest TNR of 56% was achieved by F3-KNN. For NPV, our JSE-KNN and JSE-SVM, achieved the best NPV of 100%. Among the existing techniques, F1-SVM achieved the highest TNR of 93.75%. For Dataset B, it is observed that our JSE-SVM achieved the best performance among all the techniques in terms of ACC, TPR, PPV, TNR, and NPV.

- In Dataset C, our JSE-KNN outperforms the other techniques, achieving the best TPR of 100%. Our JSE-NB places second with a TPR of 98.67%. For PPV, our JSE-KNN also achieved the best PPV of 100%. Among the other techniques, our JSE-SVM achieved the highest PPV of 97.3%. For TNR, our JSE-KNN ourperforms the other techniques, achieving the best TNR of 100%. Among the existing techniques, F1-KNN and F4-KNN achieved the highest TNR of 96%. For NPV, our JSE-KNN achieved the best NPV of 100% and JSE-NB ranked second with a NPV of 98.59%. Among the existing techniques, F1-KNN achieved the highest NPV of 96%. In summary, for Dataset C, our JSE-KNN achieves the best performance among all the techniques in terms of ACC, TPR, PPV, TNR, and NPV.

- In Dataset D, our JSE-NB achieves the highest TPR of 85.45% among all the techniques. In addition, it is observed that F4-SVM achieves the second highest TPR of 83.64%. Our JSE-KNN and JSE-SVM achieved TPRs of 81.82% and 80.91%, respectively. For PPV, our JSE-KNN, JSE-NB, and JSE-SVM outperform the

**TABLE 3.** TPR, PPV, TNR, and NPV of each technique in all datasets.

| Techniques | Dataset A | | | | Dataset B | | | |
|---|---|---|---|---|---|---|---|---|
| | TPR | PPV | TNR | NPV | TPR | PPV | TNR | NPV |
| **JSE-KNN** | **75.77** | 80.41 | 91.46 | **89.08** | **100.00** | 93.10 | 92.00 | **100.00** |
| **JSE-NB** | 68.85 | 69.11 | 85.77 | 85.61 | 90.74 | 89.09 | 88.00 | 89.80 |
| **JSE-SVM** | 63.46 | **82.50** | **93.77** | 84.73 | **100.00** | **96.43** | **96.00** | **100.00** |
| **JSE-DT** | 68.08 | 80.45 | 92.35 | 86.21 | 92.59 | 89.29 | 88.00 | 91.67 |
| F1-KNN | 43.46 | 49.56 | 79.54 | 75.25 | 92.59 | 94.34 | 94.00 | 92.16 |
| F1-NB | 61.15 | 60.23 | 81.32 | 81.90 | 92.59 | 92.59 | 92.00 | 92.00 |
| F1-SVM | 43.08 | 57.73 | 85.41 | 76.43 | 94.44 | 91.07 | 90.00 | 93.75 |
| F1-DT | 42.69 | 50.23 | 80.43 | 75.21 | 90.74 | 94.23 | 94.00 | 90.38 |
| F2-KNN | 48.85 | 59.07 | 84.34 | 78.09 | 87.04 | 92.16 | 92.00 | 86.79 |
| F2-NB | 61.15 | 63.86 | 83.99 | 82.37 | 85.19 | 80.70 | 78.00 | 82.98 |
| F2-SVM | 41.15 | 58.15 | 86.30 | 76.02 | 88.89 | 90.57 | 90.00 | 88.24 |
| F2-DT | 53.85 | 61.40 | 84.34 | 79.80 | 85.19 | 79.31 | 76.00 | 82.61 |
| F3-KNN | 36.15 | 59.12 | 88.43 | 74.96 | 66.67 | 62.07 | 56.00 | 60.87 |
| F3-NB | 38.46 | 58.82 | 87.54 | 75.46 | 75.93 | 66.13 | 58.00 | 69.05 |
| F3-SVM | 26.92 | 57.38 | 90.75 | 72.86 | 72.22 | 69.64 | 66.00 | 68.75 |
| F3-DT | 35.00 | 53.22 | 85.77 | 74.04 | 61.11 | 70.21 | 72.00 | 63.16 |
| F4-KNN | 43.46 | 50.22 | 80.07 | 75.38 | 87.04 | 83.93 | 82.00 | 85.42 |
| F4-NB | 43.46 | 55.67 | 83.99 | 76.25 | 87.04 | 83.93 | 82.00 | 85.42 |
| F4-SVM | 51.92 | 56.72 | 81.67 | 78.60 | 85.19 | 83.64 | 82.00 | 83.67 |
| F4-DT | 39.23 | 45.13 | 77.94 | 73.49 | 81.48 | 86.27 | 86.00 | 81.13 |

| Techniques | Dataset C | | | | Dataset D | | | |
|---|---|---|---|---|---|---|---|---|
| | TPR | PPV | TNR | NPV | TPR | PPV | TNR | NPV |
| **JSE-KNN** | **100.00** | **100.00** | **100.00** | **100.00** | 81.82 | 97.83 | 98.95 | 90.38 |
| **JSE-NB** | 98.67 | 93.67 | 93.33 | 98.59 | **85.45** | **98.95** | **99.47** | **92.20** |
| **JSE-SVM** | 96.00 | 97.30 | 97.33 | 96.05 | 80.91 | 98.89 | **99.47** | 90.00 |
| **JSE-DT** | 94.67 | 82.56 | 80.00 | 93.75 | 78.18 | 89.58 | 94.74 | 88.24 |
| F1-KNN | 96.00 | 96.00 | 96.00 | 96.00 | 61.82 | 69.39 | 84.21 | 79.21 |
| F1-NB | 82.67 | 93.94 | 94.67 | 84.52 | 71.82 | 75.96 | 86.84 | 84.18 |
| F1-SVM | 96.00 | 85.71 | 84.00 | 95.45 | 57.27 | 77.78 | 90.53 | 78.54 |
| F1-DT | 89.33 | 88.16 | 88.00 | 89.19 | 69.09 | 80.00 | 90.00 | 83.41 |
| F2-KNN | 76.00 | 80.28 | 81.33 | 77.22 | 58.18 | 73.56 | 87.89 | 78.40 |
| F2-NB | 77.33 | 92.06 | 93.33 | 80.46 | 56.36 | 68.13 | 84.74 | 77.03 |
| F2-SVM | 76.00 | 82.61 | 84.00 | 77.78 | 38.18 | 60.00 | 85.26 | 70.43 |
| F2-DT | 76.00 | 78.08 | 78.67 | 76.62 | 62.73 | 68.32 | 83.16 | 79.40 |
| F3-KNN | 86.67 | 84.42 | 84.00 | 86.30 | 43.64 | 65.75 | 86.84 | 72.69 |
| F3-NB | 85.33 | 80.00 | 78.67 | 84.29 | 56.36 | 80.52 | 92.11 | 78.48 |
| F3-SVM | 84.00 | 80.77 | 80.00 | 83.33 | 37.27 | 69.49 | 90.53 | 71.37 |
| F3-DT | 73.33 | 72.37 | 72.00 | 72.97 | 41.82 | 61.33 | 84.74 | 71.56 |
| F4-KNN | 80.00 | 95.24 | 96.00 | 82.76 | 69.09 | 64.41 | 77.89 | 81.32 |
| F4-NB | 88.00 | 85.71 | 85.33 | 87.67 | 60.00 | 51.56 | 67.37 | 74.42 |
| F4-SVM | 82.67 | 93.94 | 94.67 | 84.52 | 83.64 | 84.40 | 91.05 | 90.58 |
| F4-DT | 94.67 | 89.87 | 89.33 | 94.37 | 70.00 | 78.57 | 88.95 | 83.66 |

existing techniques, and their accuracies are above 97%. Among our techniques, JSE-NB achieved the highest PPV of 98.94%. For TNR, our JSE-NB and JSE-SVM achieved the best TNR of 99.47%. JSE-KNN also performs well, achieving a TNR of 98.95%. Among the existing techniques, F3-NB achieves the highest TNR of 92.11%. For NPV, our JSE-NB also achieved the best NPV of 92.2%. F3-NB achieves the second highest TNR of 90.58%. For Dataset D, it is observed that our JSE-NB achieves the best performance among all the techniques in terms of ACC, TPR, PPV, TNR, and NPV.

Table 4 shows the $F_\beta$ score of each technique in all datasets. In this table, $F_\beta$ score is calculated using (16) and (17) as follows:

$$F_\beta = (1 + \beta^2) \times \frac{\text{TPR} \times \text{PPV}}{\text{TPR} + \beta^2 \times \text{PPV}}, \qquad (20)$$

where $\beta$ is a positive real parameter that determines the weight of the TPR. In other words, in (20), TPR is $\beta$ times as

important as PPV. If $\beta$ is less than 1, more weight is applied to the PPV. On the other hand, if $\beta$ is greater than 1, more weight is applied to the TPR. If $\beta = 1$, $F_\beta$ becomes the harmonic mean of the TPR and PPV. In the experiments in Table 4, we set the values of $\beta$ 0.1, 0.5, 1, 1.5, and 2. In the cases of $\beta = 0.1$ and $\beta = 0.5$, the PPV performances of the techniques are considered more important than the TPR performances. In the cases of $\beta = 1.5$ and $\beta = 2$, the TPR performances of the techniques are considered more important than the PPV performances. In Dataset A, when $\beta = 0.1$, our JSE-SVM achieves the highest $F_\beta$ score among all the techniques. On the other hand, when $\beta \geq 0.5$, JSE-KNN outperforms the other techniques. In Dataset B, our JSE-SVM achieves the highest $F_\beta$ score among all the techniques, regardless of the value of $\beta$. Among the existing techniques, when $\beta \leq 1$, F1-KNN achieves the highest $F_\beta$ score. On the other hand, in the cases of $\beta = 1.5$ and $\beta = 2$, F1-SVM outperforms the other techniques. In Dataset C, our JSE-KNN outperforms the other techniques, achieving the best $F_\beta$ score

**TABLE 4.** $F_\beta$ score comparison between the proposed method and the existing methods.

| Techniques | Dataset A | | | | | Dataset B | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta = 0.1$ | $\beta = 0.5$ | $\beta = 1$ | $\beta = 1.5$ | $\beta = 2$ | $\beta = 0.1$ | $\beta = 0.5$ | $\beta = 1$ | $\beta = 1.5$ | $\beta = 2$ |
| **JSE-KNN** | 80.36 | **79.44** | **78.02** | **77.14** | **76.65** | 93.17 | 94.41 | 96.43 | 97.77 | 98.54 |
| **JSE-NB** | 69.11 | 69.06 | 68.98 | 68.93 | 68.90 | 89.11 | 89.42 | 89.91 | 90.23 | 90.41 |
| **JSE-SVM** | **82.26** | 77.83 | 71.74 | 68.31 | 66.53 | **96.46** | **97.12** | **98.18** | **98.87** | **99.26** |
| **JSE-DT** | 80.31 | 77.63 | 73.75 | 71.46 | 70.24 | 89.32 | 89.93 | 90.91 | 91.55 | 91.91 |
| F1-KNN | 49.49 | 48.21 | 46.31 | 45.17 | 44.56 | 94.32 | 93.98 | 93.46 | 93.12 | 92.94 |
| F1-NB | 60.24 | 60.41 | 60.69 | 60.87 | 60.97 | 92.59 | 92.59 | 92.59 | 92.59 | 92.59 |
| F1-SVM | 57.54 | 54.05 | 49.34 | 46.73 | 45.38 | 91.10 | 91.73 | 92.73 | 93.38 | 93.75 |
| F1-DT | 50.14 | 48.51 | 46.15 | 44.76 | 44.01 | 94.19 | 93.51 | 92.45 | 91.79 | 91.42 |
| F2-KNN | 58.95 | 56.70 | 53.47 | 51.59 | 50.60 | 92.10 | 91.09 | 89.52 | 88.55 | 88.01 |
| F2-NB | 63.83 | 63.30 | 62.48 | 61.96 | 61.68 | 80.74 | 81.56 | 82.88 | 83.75 | 84.25 |
| F2-SVM | 57.92 | 53.71 | 48.20 | 45.22 | 43.71 | 90.55 | 90.23 | 89.72 | 89.40 | 89.22 |
| F2-DT | 61.32 | 59.73 | 57.38 | 55.97 | 55.21 | 79.36 | 80.42 | 82.14 | 83.29 | 83.94 |
| F3-KNN | 58.75 | 52.46 | 44.87 | 41.06 | 39.20 | 62.11 | 62.94 | 64.29 | 65.18 | 65.69 |
| F3-NB | 58.52 | 53.19 | 46.51 | 43.05 | 41.32 | 66.21 | 67.88 | 70.69 | 72.62 | 73.74 |
| F3-SVM | 56.74 | 46.79 | 36.65 | 32.18 | 30.12 | 69.67 | 70.14 | 70.91 | 71.41 | 71.69 |
| F3-DT | 52.94 | 48.20 | 42.23 | 39.12 | 37.57 | 70.11 | 68.18 | 65.35 | 63.65 | 62.74 |
| F4-KNN | 50.14 | 48.71 | 46.60 | 45.34 | 44.66 | 83.96 | 84.53 | 85.45 | 86.06 | 86.40 |
| F4-NB | 55.51 | 52.71 | 48.81 | 46.61 | 45.45 | 83.96 | 84.53 | 85.45 | 86.06 | 86.40 |
| F4-SVM | 56.67 | 55.69 | 54.22 | 53.31 | 52.82 | 83.65 | 83.94 | 84.40 | 84.70 | 84.87 |
| F4-DT | 45.07 | 43.81 | 41.98 | 40.88 | 40.28 | 86.22 | 85.27 | 83.81 | 82.90 | 82.40 |

| Techniques | Dataset C | | | | | Dataset D | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta = 0.1$ | $\beta = 0.5$ | $\beta = 1$ | $\beta = 1.5$ | $\beta = 2$ | $\beta = 0.1$ | $\beta = 0.5$ | $\beta = 1$ | $\beta = 1.5$ | $\beta = 2$ |
| **JSE-KNN** | **100.00** | **100.00** | **100.00** | **100.00** | **100.00** | 97.64 | 94.14 | 89.11 | 86.16 | 84.59 |
| **JSE-NB** | 93.72 | 94.63 | 96.10 | 97.07 | 97.63 | **98.79** | **95.92** | **91.71** | **89.20** | **87.85** |
| **JSE-SVM** | 97.28 | 97.04 | 96.64 | 96.40 | 96.26 | 98.67 | 94.68 | 89.00 | 85.70 | 83.96 |
| **JSE-DT** | 82.66 | 84.73 | 88.20 | 90.58 | 91.97 | 89.45 | 87.04 | 83.50 | 81.37 | 80.22 |
| F1-KNN | 96.00 | 96.00 | 96.00 | 96.00 | 96.00 | 69.30 | 67.73 | 65.38 | 63.97 | 63.20 |
| F1-NB | 93.81 | 91.45 | 87.94 | 85.84 | 84.70 | 75.92 | 75.10 | 73.83 | 73.04 | 72.61 |
| F1-SVM | 85.81 | 87.59 | 90.57 | 92.58 | 93.75 | 77.50 | 72.58 | 65.97 | 62.33 | 60.46 |
| F1-DT | 88.17 | 88.39 | 88.74 | 88.97 | 89.10 | 79.88 | 77.55 | 74.15 | 72.12 | 71.03 |
| F2-KNN | 80.24 | 79.39 | 78.08 | 77.27 | 76.82 | 73.37 | 69.87 | 64.97 | 62.18 | 60.72 |
| F2-NB | 91.89 | 88.69 | 84.06 | 81.34 | 79.89 | 67.99 | 65.40 | 61.69 | 59.53 | 58.38 |
| F2-SVM | 82.54 | 81.20 | 79.17 | 77.92 | 77.24 | 59.66 | 53.85 | 46.67 | 42.99 | 41.18 |
| F2-DT | 78.06 | 77.66 | 77.03 | 76.63 | 76.41 | 68.26 | 67.12 | 65.40 | 64.35 | 63.77 |
| F3-KNN | 84.44 | 84.86 | 85.53 | 85.96 | 86.21 | 65.43 | 59.70 | 52.46 | 48.67 | 46.78 |
| F3-NB | 80.05 | 81.01 | 82.58 | 83.62 | 84.21 | 80.18 | 74.16 | 66.31 | 62.10 | 59.96 |
| F3-SVM | 80.80 | 81.40 | 82.35 | 82.98 | 83.33 | 68.90 | 59.25 | 48.52 | 43.47 | 41.08 |
| F3-DT | 72.38 | 72.56 | 72.85 | 73.03 | 73.14 | 61.05 | 56.10 | 49.73 | 46.36 | 44.66 |
| F4-KNN | 95.06 | 91.74 | 86.96 | 84.14 | 82.64 | 64.45 | 65.29 | 66.67 | 67.58 | 68.10 |
| F4-NB | 85.74 | 86.16 | 86.84 | 87.28 | 87.53 | 51.63 | 53.05 | 55.46 | 57.12 | 58.10 |
| F4-SVM | 93.81 | 91.45 | 87.94 | 85.84 | 84.70 | 84.40 | 84.25 | 84.02 | 83.87 | 83.79 |
| F4-DT | 89.92 | 90.79 | 92.21 | 93.14 | 93.67 | 78.48 | 76.69 | 74.04 | 72.43 | 71.56 |

of 100%. Our JSE-SVM also achieves a good $F_\beta$ score of 96% or above. Among the exsting techniques, F1-KNN achieves the highest $F_\beta$ score of 96%, while the lowest is achieved by F3-DT. In Dataset D, our JSE-NB outperforms the other techniques, regardless of the value of $\beta$. Among our techniques, JSE-DT shows the worst performance. However, JSE-DT outperforms the other existing techniques, except F4-SVM. Among the existing techniques, F4-SVM shows the best performance.

## VI. CONCLUSION

In this study, we developed a new gait feature, called JSE, for skeleton-based gender classification. The proposed gender classification method consists of three phases: (1) human body-centered skeleton representation, (2) JSE extraction, and (3) classification. In the first phase, we compute the transverse, frontal, and median planes from the input gait sequence for human body-centered skeleton representation. In the second phase, we extract the JSE feature for the three

planes. The extracted JSEs are then concatenated into a feature vector. In the final phase, the classification model is trained using the labeled JSE training data. After the training is completed, gender classification is conducted on the unlabeled JSE testing data using the trained classification model. To demonstrate the effectiveness of the proposed gender classification method, we trained four machine learning algorithms using JSE and evaluated them on four publicly available gait datasets. The results demonstrated that the proposed method achieved state-of-the-art performance in terms of accuracy, TPR, PPV, TNR, NPV, and $F_\beta$ in all the datasets. As a part of future work, we plan to investigate the effectiveness of our method for classifying the gender of people whose body parts are partially occluded. In the gait sequences used in the experiments, there was no invisible body joint. However, in practical situation, while walking, some body joints sometimes can be occluded by bag, clothes, and other objects, such as tree and wall. In the worst case scenario, only a few body joints are visible during the whole

time interval. In this case, for the occluded joints, it is difficult to extract the JSEs. Therefore, in future work, we plan to evaluate the performance of our method on gait sequences of people whose some body parts are occluded. In addition, analyzing the evaluation results, we plan to make our method robust to partial occlusion and data loss.

## REFERENCES

[1] H. Kraft and J. M. Weber, "A look at gender differences and marketing implications," *Int. J. Bus. Social Sci.*, vol. 3, no. 21, pp. 247–253, Nov. 2012.

[2] P. Keshari and S. Jain, "Effect of age and gender on consumer response to advertising appeals," *Paradigm*, vol. 20, no. 1, pp. 69–82, Jun. 2016.

[3] J. Andy Wood, J. Johnson, J. S. Boles, and H. Barksdale, "Investigating sales approaches and gender in customer relationships," *J. Bus. Ind. Marketing*, vol. 29, no. 1, pp. 11–23, Jan. 2014.

[4] A. Katrodia, M. J. Naude, and S. Soni, "Consumer buying behavior at shopping malls: Does gender matter?" *J. Econ. Behav. Stud.*, vol. 10, no. 1, pp. 125–134, Mar. 2018.

[5] E. Widjaja, G. H. Lim, and A. An, "A novel method for human gender classification using Raman spectroscopy of fingernail clippings," *Analyst*, vol. 133, no. 4, pp. 493–498, Feb. 2008.

[6] S. A. Khan, M. Ahmad, M. Nazir, and N. Riaz, "A comparative analysis of gender classification techniques," *Int. J. Bio-Sci. Bio-Technol.*, vol. 5, no. 4, pp. 223–243, Aug. 2013.

[7] M. Daneshmand, A. Aabloo, C. Ozcinar, and G. Anbarjafari, "Real-time, automatic shape-changing robot adjustment and gender classification," *Signal, Image Video Process.*, vol. 10, no. 4, pp. 753–760, Apr. 2016.

[8] F. Lin, Y. Wu, Y. Zhuang, X. Long, and W. Xu, "Human gender classification: A review," *Int. J. Biometrics*, vol. 8, nos. 3–4, pp. 275–300, Mar. 2017.

[9] L. Cao, M. Dikmen, Y. Fu, and T. S. Huang, "Gender recognition from body," in *Proc. 16th ACM Int. Conf. Multimedia MM*, 2008, pp. 725–728.

[10] Y. Dong and D. L. Woodard, "Eyebrow shape-based features for biometric recognition and gender classification: A feasibility study," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011, pp. 1–8.

[11] Y. Yuan, Y. Pang, and X. Li, "Footwear for gender recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 1, pp. 131–135, Jan. 2010.

[12] M. Nazir, A. Majid-Mirza, and S. Ali-Khan, "PSO-GA based optimized feature selection using facial and clothing information for gender classification," *J. Appl. Res. Technol.*, vol. 12, no. 1, pp. 145–152, Feb. 2014.

[13] C.-C. Lee and C.-S. Wei, "Gender recognition based on combining facial and hair features," in *Proc. Int. Conf. Adv. Mobile Comput. Multimedia - MoMM*, 2013, pp. 537–540.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.

[15] M. Deng and C. Wang, "Human gait recognition based on deterministic learning and data stream of microsoft kinect," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 12, pp. 3636–3645, Dec. 2019.

[16] P. Limcharoen, N. Khamsemanan, and C. Nattee, "View-independent gait recognition using joint replacement coordinates (JRCs) and convolutional neural network," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3430–3442, 2020.

[17] B. Kwon, D. Kim, J. Kim, I. Lee, J. Kim, H. Oh, H. Kim, and S. Lee, "Implementation of human action recognition system using multiple Kinect sensors," in *Proc. Pacific-Rim Conf. Multimedia (PCM)*, Gwangju, South Korea, Sep. 2015, pp. 334–343.

[18] B. Kwon, J. Kim, and S. Lee, "An enhanced multi-view human action recognition system for virtual training simulator," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA)*, Dec. 2016, pp. 1–4.

[19] B. Kwon, J. Kim, K. Lee, Y. K. Lee, S. Park, and S. Lee, "Implementation of a virtual training simulator based on 360° multi-view human action recognition," *IEEE Access*, vol. 5, pp. 12496–12511, 2017.

[20] B. Kwon and S. Lee, "Ensemble learning for skeleton-based body mass index classification," *Appl. Sci.*, vol. 10, no. 21, p. 7812, Nov. 2020.

[21] B. Kwon and S. Lee, "Human skeleton data augmentation for person identification over deep neural network," *Appl. Sci.*, vol. 10, no. 14, p. 4849, Jul. 2020.

[22] B. Kwon, J. Huh, K. Lee, and S. Lee, "Optimal camera point selection toward the most preferable view of 3-D human pose," *IEEE Trans. Syst., Man, Cybern. Syst.*, early access, Jul. 9, 2020, doi: 10.1109/TSMC.2020.3004338.

[23] S.-U. Ko, M. I. Tolea, J. M. Hausdorff, and L. Ferrucci, "Sex-specific differences in gait patterns of healthy older adults: Results from the baltimore longitudinal study of aging," *J. Biomech.*, vol. 44, no. 10, pp. 1974–1979, Jul. 2011.

[24] R. Frimenko, C. Whitehead, and D. Bruening, "Do men and women walk differently? A review and meta-analysis of sex difference in non-pathological gait kinematics," IST, A DCS Company, Dayton, OH, USA, Tech. Rep. AFRL-RH-WP-TR-2014-0016, Jan. 2014, pp. 1–34. Accessed: Aug. 16, 2020. [Online]. Available: https://apps.dtic.mil/sti/pdfs/ADA597428.pdf

[25] G. W. Cottrell and J. Metcalfe, "EMPATH: Face, emotion, and gender recognition using holons," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Denver, CO, USA, Nov. 1990, pp. 564–571.

[26] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski, "SEXNET: A neural network identifies sex from human faces," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Denver, CO, USA, Nov. 1990, pp. 572–577.

[27] S. Tamura, H. Kawai, and H. Mitsumoto, "Male/female identification from 8 × 6 very low resolution face images by neural network," *Pattern Recognit.*, vol. 29, no. 2, pp. 331–335, Feb. 1996.

[28] S. Gutta, H. Wechsler, and P. J. Phillips, "Gender and ethnic classification of face images," in *Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 1998, pp. 194–199.

[29] B. Moghaddam and M.-H. Yang, "Gender classification with support vector machines," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Mar. 2000, pp. 306–311.

[30] H. Abdi, D. Valentin, B. Edelman, and A. J. O'Toole, "More about the difference between men and women: Evidence from linear neural networks and the principal-component approach," *Perception*, vol. 24, no. 5, pp. 539–562, May 1995.

[31] Z. Sun, G. Bebis, X. Yuan, and S. J. Louis, "Genetic feature subset selection for gender classification: A comparison study," in *Proc. 6th IEEE Workshop Appl. Comput. Vis., . (WACV )*, Dec. 2002, pp. 165–170.

[32] H. C. Lian and B. L. Lu, "Multi-view gender classification using local binary patterns and support vector machines," in *Proc. Int. Symp. Neural Netw. (ISNN)*, Chengdu, China, May 2006, pp. 202–209.

[33] N. Sun, W. Zheng, C. Sun, C. Zou, and L. Zhao, "Gender classification based on boosting local binary pattern," in *Proc. Int. Symp. Neural Netw. (ISNN)*, Chengdu, China, May 2006, pp. 194–201.

[34] C. Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern Recognit. Lett.*, vol. 33, no. 4, pp. 431–437, Mar. 2012.

[35] G. Levi and T. Hassncer, "Age and gender classification using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 34–42.

[36] M. Islam, N. Tasnim, and J. H. Baek, "Human gender classification using transfer learning via Pareto frontier CNN networks," *Inventions*, vol. 5, no. 16, pp. 1–12, Apr. 2020.

[37] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[38] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5MB model size," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Apr. 2017, pp. 207–212.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[40] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2002, pp. 155–162.

[41] G. Huang and Y. Wang, "Gender classification based on fusion of multi-view gait sequences," in *Proc. Asian Conf. Comput. Vis.*, Nov. 2007, pp. 462–471.

[42] X. Li, S. J. Maybank, S. Yan, D. Tao, and D. Xu, "Gait components and their application to gender recognition," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 2, pp. 145–155, Mar. 2008.

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

[44] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.

[45] S. Choudhary, C. Prakash, and R. Kumar, "A hybrid approach for gait based gender classification using gei and spatio temporal parameters," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2017, pp. 1767–1771.

[46] E.-S.-M. El-Alfy and A. G. Binsaadoon, "Silhouette-based gender recognition in smart environments using fuzzy local binary patterns and support vector machines," *Procedia Comput. Sci.*, vol. 109, pp. 164–171, Jan. 2017.

[47] K. Kitchat, N. Khamsemanan, and C. Nattee, "Gender classification from gait silhouette using observation angle-based GEIs," in *Proc. IEEE Int. Conf. Cybern. Intell. Syst. (CIS) IEEE Conf. Robot., Autom. Mechatronics (RAM)*, Nov. 2019, pp. 485–490.

[48] Y. Qian and P. C. Woodland, "Very deep convolutional neural networks for robust speech recognition," in *Proc. IEEE Spoken Lang. Technol. Workshop (SLT)*, Dec. 2016, pp. 481–488.

[49] S. Bei, J. Deng, Z. Zhen, and S. Shaojing, "Gender recognition via fused silhouette features based on visual sensors," *IEEE Sensors J.*, vol. 19, no. 20, pp. 9496–9503, Oct. 2019.

[50] J. H. Yoo, D. Hwang, and M. S. Nixon, "Gender classification in human gait using support vector machine," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.*, Sep. 2005, pp. 138–145.

[51] W. T. Dempster and G. R. L. Gaughran, "Properties of body segments based on size and weight," *Amer. J. Anatomy*, vol. 120, no. 1, pp. 33–54, Jan. 1967.

[52] D. Kastaniotis, I. Theodorakopoulos, G. Economou, and S. Fotopoulos, "Gait-based gender recognition using pose information for real time applications," in *Proc. 18th Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2013, pp. 1–6.

[53] V. O. Andersson, L. S. Amaral, A. R. Tonini, and R. M. Araujo, "Gender and body mass index classification using a Microsoft Kinect sensor," in *Proc. Int. Florida Artif. Intell. Res. Soc. Conf.*, Apr. 2015, pp. 103–106.

[54] R. Miyamoto and R. Aoki, "Gender prediction by gait analysis based on time series variation on joint position," *J. Systemics, Cybern. Informat.*, vol. 13, no. 3, pp. 75–82, 2015.

[55] M. M. Bachtiar, F. F. Nuzula, and S. Wasista, "Gait with a combination of swing arm feature extraction for gender identification using Kinect skeleton," in *Proc. Int. Seminar Intell. Technol. Appl. (ISITIA)*, Jul. 2016, pp. 79–82.

[56] M. H. Ahmed and A. T. Sabir, "Human gender classification based on gait features using kinect sensor," in *Proc. 3rd IEEE Int. Conf. Cybern. (CYBCONF)*, Jun. 2017, pp. 1–5.

[57] S. Camalan, G. Sengul, S. Misra, R. Maskeliūnas, and R. Damaševičius, "Gender detection using 3D anthropometric measurements by Kinect," *Metrology Meas. Syst.*, vol. 25, no. 2, pp. 253–267, 2018.

[58] V. O. Andersson and R. M. Araujo, "Person identification using anthropometric and gait data from Kinect sensor," in *Proc. Assoc. Advancement Artif. Intell. (AAAI) Conf. Artif. Intell.*, Feb. 2015, pp. 425–431.

[59] D. A. Guffanti, A. Brunete, and M. Hernando, "Kinematic gait data using a microsoft Kinect V2 sensor during gait sequences over a treadmill," *IEEE Dataport*, to be published.

[60] D. Kastaniotis, I. Theodorakopoulos, C. Theoharatos, G. Economou, and S. Fotopoulos, "A framework for gait-based recognition using kinect," *Pattern Recognit. Lett.*, vol. 68, pp. 327–335, Dec. 2015.

[61] D. Kastaniotis, I. Theodorakopoulos, and S. Fotopoulos, "Pose-based gait recognition with local gradient descriptors and hierarchically aggregated residuals," *J. Electron. Imag.*, vol. 25, no. 6, pp. 1–9, Dec. 2016.

[62] D. Kastaniotis, I. Theodorakopoulos, G. Economou, and S. Fotopoulos, "Gait based recognition via fusing information from Euclidean and Riemannian manifolds," *Pattern Recognit. Lett.*, vol. 84, pp. 245–251, Dec. 2016.

[63] J. B. Flora, D. F. Lochtefeld, D. A. Bruening, and K. M. Iftekharuddin, "Improved gender classification using nonpathological gait kinematics in full-motion video," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 3, pp. 304–314, Jun. 2015.

[64] C. M. Williams, J. B. Flora, and K. M. Iftekharuddin, "Gender classification of running subjects using full-body kinematics," *Proc. SPIE*, vol. 9841, May 2016, Art. no. 984107.

**BEOM KWON** was born in Seoul, South Korea, in 1989. He received the B.S. degree in electrical and electronic engineering from Soongsil University, Seoul, in 2012, and the M.S. and Ph.D. degrees in electrical and electronic engineering from Yonsei University, Seoul, in 2018. From March 2018 to September 2019, he was a Senior Researcher with the Agency for Defense Development (ADD), Daejeon, South Korea. Since October 2019, he has been a Staff Engineer with Samsung Electronics Company Ltd., Suwon, South Korea. His research interests include wireless communication networks, computer vision, and machine learning.

**SANGHOON LEE** (Senior Member, IEEE) received the B.S. degree in E.E. from Yonsei University, Seoul, South Korea, in 1989, the M.S. degree in E.E. from KAIST, in 1991, and the Ph.D. degree in E.E. from The University of Texas at Austin, in 2000. From 1991 to 1996, he worked with Korea Telecom. From 1999 to 2002, he worked with Lucent Technologies. He joined the Faculty of the Department of Electrical and Electronics Engineering, Yonsei University, in March 2003, where he is currently a Full Professor. His research interests include deep learning, image and video quality of experience, computer vision, and computer graphics. He has been a member of the Technical Committees of the IEEE MMSP, since 2016, and IVMSP, since 2014. He received the 2015 Yonsei Academic Award, the 2010, 2015, 2016, 2017 Yonsei Outstanding Accomplishment Awards, the 2012 Special Service Award from the IEEE Broadcast Technology Society, the 2013 Special Service Award from the IEEE Signal Processing Society, the Humantech Thesis Award of Samsung Electronics, 2013, the IEEE Seoul Section Student Paper Contest Award, in 2012, the Qualcomm Innovation Award, Qualcomm, in 2012, and the Best Student Paper Award of International Conference on Quality of Multimedia Experience (QoMEX) 2018. Since 2011, he has been a Chair of the IEEE P3333.1 Quality Assessment Working Group. Since 2018, he has been serving as a Chair of the APSIPA IVM Technical Committee. He has participated in the international activities as a General Chair of the 2013 IEEE IVMSP Workshop, a Technical Program Co-chair of the IEEE International Conference on Multimedia and Expo (ICME) 2018, the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA) 2018, the International Conference on Information Networking (ICOIN) 2014, the Global 3D Forum 2012 and 2013, and the Exhibition Chair of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2018, and the Local Arrangements Chair of IEEE International Symposium On Broadband Multimedia Systems and Broadcasting (BMSB), 2012. From 2009 to 2015, he was an Editor of the *Journal of Communications and Networks* (JCN). From 2010 to 2014, he was an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING. He was a Guest Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, in 2013, and the *Journal of Electronic Imaging*, in 2015. Since 2014, he has been an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS.

• • •