International Journal of Innovative Computing, Information and Control
Volume **20**, Number **4**, August **2024**
ICIC International ©2024 ISSN 1349-4198
pp. 1245–1256

A METHOD OF RECOGNIZING HUMAN WALK MOTION FROM MULTIPLE DIRECTIONS

Miki Ooba[1] and Yui Tanjo[2]

[1]Graduate School of Engineering
[2]Faculty of Engineering
Kyushu Institute of Technology
1-1, Sensui, Tobata, Kitakyushu, Fukuoka 804-8550, Japan
ooba.miki558@mail.kyutech.jp; tanjo@cntl.kyutech.ac.jp

Received December 2023; revised April 2024

Abstract. *It is an important issue for elderly people to maintain their walking ability in order to prevent falls and lead a healthy daily life. It may be of great help to develop a computer vision system which examines human daily walk posture and gives him/her some advices on the posture. This paper proposes a method of recognizing walk motions of a human focusing on his/her posture. In order to describe a human posture, we choose 39 structural features defined from human joint coordinates obtained using OpenPose and 19 figural features from human domain images and their different images. The feature vector containing these 58 features is used for recognizing a human walk motion by Random Forest. In the experiment, the method was applied to three walk motions of five persons each with eight walk directions and satisfactory results were obtained.*
Keywords: OpenPose, Image of human area, Human walk motion, Structural feature, Figural feature

1. **Introduction.** According to the statistics conducted by the Ministry of Health, Labour and Welfare in 2022 [1], the total population of Japan is 124.95 million, with the elderly accounting for 29.0% of the total population. According to the same statistics [2], the number of elderly people certified as requiring nursing care or support is also increasing. Fractures and falls accounted for 13.0% of the causes of their need for nursing care. The main factors of deterioration in walking ability with age are muscle strength, balance function, and decreased range of motion of joints. Characteristics of the walking posture of the elderly include decreased walking speed, shortened stride length, increased forward lean of the upper body, and decreased toe and heel elevation [3].

The ultimate goal of this research is to develop a robot that understands the feelings of care-receivers implicitly. As the first step, we aim at creating a viewpoint of a caregiver robot and understanding the information (images) obtained from the viewpoint.

Previous studies include gait data acquisition [4] and posture when exercising correction [5] using Kinect, and human behavior recognition [6] using inertial sensors. There is also measurement of walking using a multi-view camera [7] and recognition motion based on silhouette and optical flow [8]. There are methods to examine the gait by attaching a sensor to the objects that a person is wearing [9,10]. Another one is a posture and motion recognition method performed employing the perspective of a pedestrian himself, wearing a camera on his chest [11,12]. There is walk recognition using walk silhouettes input to the CNN and integrating them [13]. By using depth data, there are methods to represent motion from whole-body skeletal trajectories and to evaluate symmetry [14]. Using skeletal

DOI: 10.24507/ijicic.20.04.1245

information from OpenPose [15], there is an analysis of throwing motion [16], detection of a person trying to interrupt a conversation [17], and correction of weight training posture [18]. There are motion recognition methods [19] using triplet motion representation images, an extension of motion history image, and histograms of oriented optical flow. Other methods include action recognition by inputting low-resolution images captured with an encoded exposure camera into a CNN [20], and action recognition using a network that identifies relationships between ongoing actions and other information [21]. There is also a method that uses image reconstruction to perform posture estimation that is unaffected by the background and other factors by encoding only features of the foreground object in unsupervised learning [22]. Another method performs motion recognition by finding similarities to the current motion from a set of past motions, focusing on whole-body posture, body parts, and individual joints [23]. ResNet is used in [24] to perform skeletal detection by considering the position and rotation of joints. There is also a method for action recognition using dynamic vision sensors [25].

However, in order to recognize postures in daily life, it is undesirable to attach markers or sensors to a person or to the object he/she is wearing, as it leads to unnatural movements. In addition to that, it is better not to require large-scale devices such as multiple cameras. In order to analyze walking behavior, it is necessary to identify the type of movement and posture of a pedestrian, regardless of the direction in which he/she is moving. Therefore, the system should extract the features that can not only recognize walking behavior, but also analyze movements.

The objective of this research is to recognize everyday walk motions of a person. In this paper, we propose a recognition method of walk motions employing human joint locations and image characteristics. We choose 39 features defined by human joint coordinates obtained from OpenPose [26]. These features represent local functional features of a human posture. We also select 19 features from human domain images and their difference images. They express overall figural features of a human. Thus, a human posture is described by a 58-dimensional vector. The Random Forest classifier [27] is used for the human posture recognition based on the 58-D feature vectors.

In human motion recognition using a camera, it is necessary to recognize motions without being affected by a camera's direction or a person's walk direction. Considering three motions, i.e., normal walk, forward leaning walk and fall, we capture video images in eight directions (left, right, depth (2 ways), and oblique (2 diagonal ways and each 2 ways)) with the first two types of walk, and in four directions (left, right, depth) with the fall.

In the following, Section 2 gives problem statement and preliminaries. Section 3 introduces the proposed structural features. Section 4 explains how to extract a human area from a video, and describes the figural features derived from human domain images. In Section 5, experimental results are presented. Finally, Section 6 discusses the results and Section 7 concludes the paper.

2. **Problem Statement and Preliminaries.**

There are [4,5] that use a sensor called Kinect to capture motions of the walking of elderly people. The goal is to develop a system that can provide feedback to the subject by analyzing their joint coordinate data obtained from the motion and presenting the posture state. However, Kinect does not have a large recognition range for a subject, and there are problems of occlusion and the need to limit the walking range to ensure data accuracy.

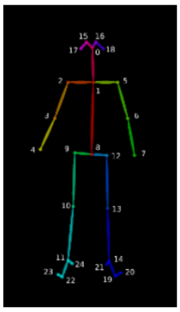
A study [6] used a data set containing four types of data: RGB video, depth video, skeletal position from the Kinect camera and inertial signals from a wearable inertial sensor. It shows how this database can be used to study a motion derived from both the depth camera data and the inertial sensor data. Although it states that this dataset would be beneficial for multimodality research activities being conducted for human behavior recognition, it admits that it is not realistic to use multiple cameras and wearable sensors to recognize everyday life activities. We believe that sensors attached to canes or shoe soles [9,10] are also difficult to use for the elderly for the same reason.

In the study on motion capture of outdoor pedestrians using multi-view cameras [7], the torso posture is obtained from manually set coordinates of limbs for some frames of video captured by four fixed cameras. The need for multiple cameras and the burden of manual set of the coordinates of the limbs would be troublesome, though.

In addition, as seen from [28], deep learning is widely used in current action and posture recognition techniques. However, the ultimate goal of this research is to give feedback on joint or partial body motions to the elderly in order to keep their healthy daily lives.

For this purpose, the emphasis is rather placed on the analysis of the recognized posture and we use a non-black-box method.

3. **Structural Features Using Joint Coordinates.** OpenPose [26] is a posture estimation library using convolutional neural networks. Joint coordinates extracted from monocular camera images are expressed by the (x, y) coordinates on the image. In the proposed method, 25 joint coordinates obtained from OpenPose (See Figure 1) are chosen to define joint-related features.



0 : Nose	9 : Rhip	18 : Lear
1 : Neck	10 : Rknee	19 : LBigToe
2 : Rshoulder	11 : Rankle	20 : LSmallToe
3 : Relbow	12 : Lhip	21 : Lheel
4 : Rwrist	13 : Lknee	22 : RBigToe
5 : Lshoulder	14 : Lankle	23 : RSmallToe
6 : Lelbow	15 : Reye	24 : Rheel
7 : Lwrist	16 : Leye	
8 : Midhip	17 : Rear	

FIGURE 1. Joint coordinates obtained by OpenPose [26]

In this section, we propose the following nine features (39 dimensions) to obtain human gait features.

1) Knee and ankle angle

The knee angle θ_{knee} [deg] is calculated using the inner product of the waist and knee vectors as follows:

$$\boldsymbol{a} = (x_{hip} - x_{knee}, y_{hip} - y_{knee}) \tag{1}$$

$$\boldsymbol{b} = (x_{ankle} - x_{knee}, y_{ankle} - y_{knee}) \tag{2}$$

$$\theta_{knee} = \frac{180}{\pi} \cos^{-1} \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\|\boldsymbol{a}\| \|\boldsymbol{b}\|} \tag{3}$$

where (x_{hip}, y_{hip}), (x_{knee}, y_{knee}), (x_{ankle}, y_{ankle}) are the coordinates of the hip, knee and ankle, respectively, as shown in Figure 2(a). Similarly, the ankle angles θ_{ankle} are obtained using the coordinates of the knee, ankle, and toe.

2) Ankle-toe angle

The angle between the ankle and toe with respect to a horizontal line is defined as the ankle-toe angle θ_{instep} [deg] and is given by the following equation:

$$\theta_{instep} = \frac{180}{\pi} \tan^{-1} \frac{y_{ankle} - y_{toe}}{x_{ankle} - x_{toe}} \tag{4}$$

Here, (x_{ankle}, y_{ankle}) and (x_{toe}, y_{toe}) are the coordinates of the ankle and toe, respectively, as shown in Figure 2(b).

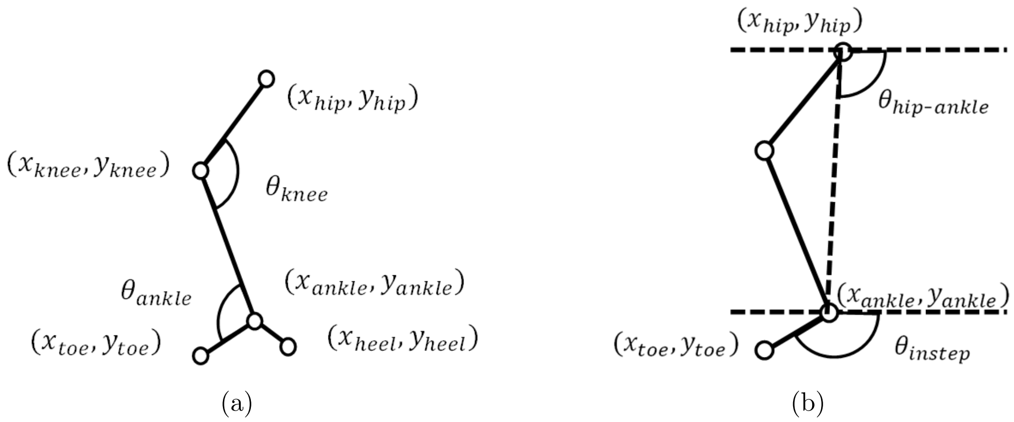


FIGURE 2. Foot joint coordinates: (a) Knee and ankle angle; (b) ankle-toe angle and hip-ankle angle

3) Hip-ankle angle

As in 2), the angle of the line connecting the hip and ankle is defined as the hip-ankle angle $\theta_{hip\text{-}ankle}$ [deg], which is given by the following equation:

$$\theta_{hip\text{-}ankle} = \frac{180}{\pi} \tan^{-1} \frac{y_{hip} - y_{ankle}}{x_{hip} - x_{ankle}} \tag{5}$$

where (x_{hip}, y_{hip}) and (x_{ankle}, y_{ankle}) are the coordinates of the hip and ankle, respectively, as shown in Figure 2(b).

4) Steps

(a) Scale transformation

The size of a person's vector in the image changes depending on the position of the person as seen from the shooting direction.

The parameter *scale* [cm/pixel], which indicates how many centimeters one pixel corresponds to in the current frame using the person's thigh length *thighs* [cm], is obtained by the following equation:

$$scale = \frac{thighs}{\sqrt{(x_{hip} - x_{knee})^2 + (y_{hip} - y_{knee})^2}} \tag{6}$$

(b) Measurement of stride length

The step size is calculated using the Euclidean distance between the left heel and the right heel as follows:

$$step\_size = scale \times \sqrt{(x_{Rheel} - x_{Lheel})^2 + (y_{Rheel} - y_{Lheel})^2} \tag{7}$$

where (x_{Rheel}, y_{Rheel}) and (x_{Lheel}, y_{Lheel}) are the coordinates of the right and left heel, respectively.

5) Walk speed

The feature walk speed [cm/s] indicates how much the coordinates of the waist center (x_{c0}, y_{c0}) in the current frame have moved from (x_{cn}, y_{cn}) in the previous n frames, and is given by the following equation:

$$walk\_speed = scale \times \frac{\sqrt{(x_{c0} - x_{cn})^2 + (y_{c0} - y_{cn})^2} \times frame\_rate}{n} \tag{8}$$

where (x_{c0}, y_{c0}), (x_{cn}, y_{cn}) are the coordinates of the center of the waist between the current frame and n frames ago.

6) Direction of motion

The direction of motion in the horizontal (x-axis) and in the vertical (y-axis) is obtained by vectorizing the displacement of the coordinates of the waist center during n frames and multiplying it by $scale$.

$$x = (x_{c0} - x_{cn}) \times scale, \quad y = (y_{c0} - y_{cn}) \times scale \tag{9}$$

7) Hip-ankle distance

As in 4), the hip-ankle distance is calculated using the Euclidean distance between the hip and the ankle, as shown in Figure 3(a).

$$\textit{hip-ankle} = scale \times \sqrt{(x_{hip} - x_{ankle})^2 + (y_{hip} - y_{ankle})^2} \tag{10}$$

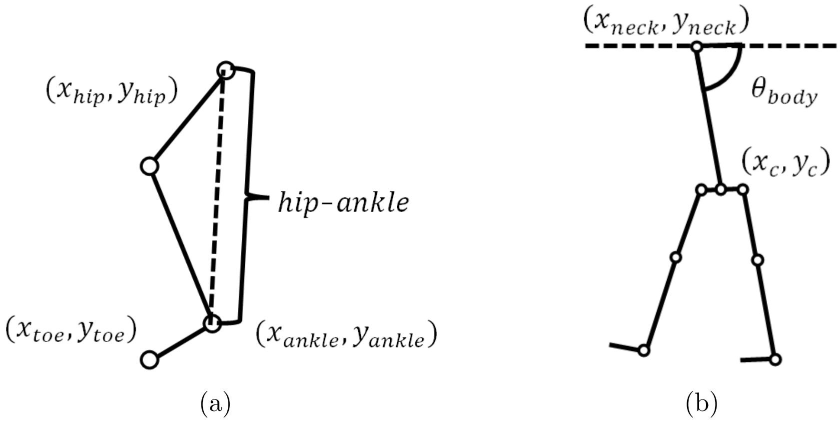


FIGURE 3. Joint coordinates: (a) Hip-ankle distance; (b) body tilt

8) Body tilt

As in 2), $\theta_{bodytilt}$ [deg] is defined as the angle of the line connecting the neck and waist as shown in Figure 3(b), and is given by the following equation:

$$\theta_{bodytilt} = \frac{180}{\pi} \tan^{-1} \frac{y_{neck} - y_c}{x_{neck} - x_c} \tag{11}$$

where (x_{neck}, y_{neck}) and (x_c, y_c) are the coordinates of the neck and center of the waist, respectively.

9) Difference of the foot joint heights

The height of the knee $height_{knee}$ [cm] is calculated by the difference between the heights of the right and the left knees as follows:

$$height_{knee} = |(y_{Rknee} - y_{Lknee})| \times scale \tag{12}$$

where (x_{Rknee}, y_{Rknee}) and (x_{Lknee}, y_{Lknee}) are the coordinates of the right and the left knees, respectively.

Moreover, the difference in height between the right and left heel as well as the height from floor to toe, is calculated in the same way.

Features such as knee and ankle angles are calculated for the left and right feet, respectively. And the feature values 1), 4) and 9) repeat similar values in each gait cycle (e.g., in the walking motion, a person repeatedly bends and extends the knee). Therefore, in addition to the above 1) to 9) features in the current frame, the minimum and the maximum values of these features in the past n frames are included as the features. This amounts to 39 dimensional features.

4. **Foreground Extraction.** To recognize a motion from the shape features of a human region image, the human region is extracted from the image as the foreground. In a real environment, there are changes in the background due to the movement of furniture or household goods.

The proposed method uses the sequential background estimation method based on a Gaussian mixture model [29] to extract human regions corresponding to background changes. In this method, a Gaussian mixture model is created for each pixel, and pixels that match one of the distributions are used as the background, while other pixels are used as the foreground. By updating each distribution using the information from the current frame [30], it is possible to respond to changes in the background.

After the foreground is extracted using the above mentioned method, noise is removed by expansion and contraction processing. The set of white pixels is then trimmed to a rectangle of a specified height as the person area. The binary image shown in Figure 4(a) is used to extract features from the human domain image. In addition, as shown in Figure 4(b), features are also extracted from the frame difference image between the previous frame image and the current frame image.

(a) (b)

Figure 4. Frame image: (a) The human region image; (b) the frame difference image

Figural features using human domain images. The proposed method uses the following eight features (19 dimensions) extracted from the human domain image.

1) Image aspect ratio

The aspect ratio of the image is the height to width of the human domain image.

2) Percentage of a white pixel area in a rectangle

The percentage of a white pixel area in a rectangle is calculated as a person area in the human domain image.

3) Hand swing (upper body width)

The size of the hand swing is calculated from the difference between the right and the left edges of the human domain at 60% of the *height* of the upper portion of the image.

4) Shoulder height

The shoulder height is the y coordinate at which the number of white pixels per line exceeds the threshold value when the image is scanned from the upper left. The threshold value was set at 40 pixels experimentally. By doing so, it distinguishes between a posture with a straight back and a bent back.

5) Length of the contour line of the lower part of the human region image

This is derived from calculating the length of the contour line of the bottom 40% of the *height* of the human domain image.

6) Difference between the left and the right sides of the body

This represents the asymmetry of the human body. First, the center of gravity (x_c, y_c) of the bottom 40% portion of the person area is calculated. Then, the ratio of the area to the left of the center of gravity in the area of the entire person area is calculated.

7) Position of the center of gravity of the head

The *head* indicates how far the head's center of gravity is from the centerline of the image.

8) Number of white pixels in the lower part of the difference image

The number of white pixels in the bottom 25% of the *height* of the difference image is used.

For features 2), 3), and 5), the values are periodic. Therefore, in addition to the value in the current frame, the minimum and maximum values in the past n frames are used as feature values. Thus, 19 features are obtained from the human domain image, including periodic features.

After all, from the structural features and figural features, the proposed method represents a human posture by a 58-dimensional feature vector.

After obtaining all the above features, the proposed method uses Random Forest [27] as a discriminator. Random Forest is characterized by its ability to output probabilistic predictions, generalization ability for unknown data, and efficiency due to its ability to run in parallel.

5. **Experiment.**

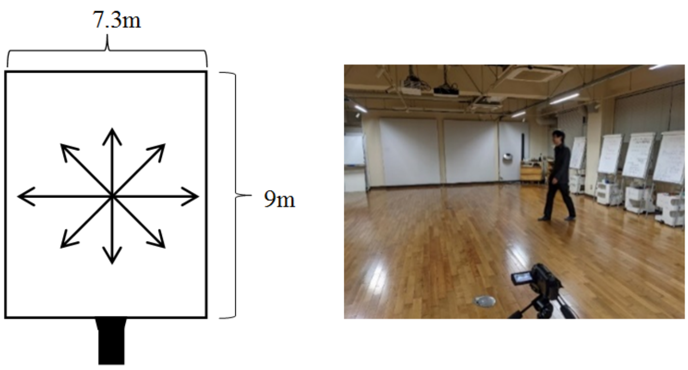
In this section, we describe the experiment on walk posture recognition based on the proposed method and evaluation of the experimental results. Five subjects perform three kinds of walk, each in eight directions. They are captured videos and recognized by the proposed method.

The experimental environment is as follows. The processed image size is 1280 × 720 [pixels]. The developmental environment is PyCharm 2021.2.3.

5.1. **Experimental method.**

Details of the experimental method are described below.

1) One fixed camera is installed at a height of 90 cm from the floor. Videos are taken in the environment shown in Figure 5. Since the actual scenario assumed indoor movements, the experiment was also conducted indoors.



(a) (b)

FIGURE 5. Environment of video taking: (a) Top view; (b) view of video taking

2) As a prerequisite, only one person appears in the video, and the entire body is captured.

3) Three movements, i.e., normal walk, forward leaning walk, and fall, are captured videos from several directions (20 moving images were taken for each person).

4) In the normal walk and forward leaning walk, eight directions' walk (right, left, front to back, back to front, front right to back left, back left to front right, front left to back right, and back right to front left) are captured. In order to discriminate between movements with different stride lengths, we vary the size of the stride lengths. First, subjects are asked to walk with their normal stride length, and then with a smaller stride length than that. The exact size of the stride was not specified.

5) With the fall movement, four directions' fall (right, left, front to back, and back to front) are captured. A subject is asked to walk with a normal stride length and then with a smaller stride length. It is noted that, with the fall movement, a subject acts the fall on a mat placed on the floor.

6) The experiment was conducted on five subjects (adults aged 21-22). They are identified by labels A, B, C, D and E.

Table 1 shows the number of frames for each video derived from the frame rate.

Table 1. The number of frames in the video

Motion	The number of frames					
	A	B	C	D	E	Total frame number of each motion
Normal walk	1317	1094	953	1069	1140	5573
Forward leaning walk	1833	1624	1195	1213	1464	7329
Fall	400	407	405	309	278	1799
Total frame number of respective subjects	3550	3125	2553	2591	2882	

5.2. **Method of evaluation.** We performed the experiment described in the previous section on five persons. Leave-one-out cross-validation was applied to the experimental data to obtain an average recognition rate.

The following equation is used to evaluate the results.

$$precision = \frac{number\ of\ correctly\ recognized\ frames\ of\ *}{total\ frame\ number\ of\ *} \times 100[\%] \tag{13}$$

5.3. **Results.** Table 2 compares the recognition accuracy when only joint coordinates are used, when only human shape is used, and when both of these features are used, where * in Equation (13) stands for each motion. Table 3, on the other hand, shows the recognition results with respect to each subject when both features are used, in which * in Equation (13) indicates respective subjects on each motion.

Figure 6 shows some of the images of the recognition results. The movement and direction of the recognition result are shown in the upper left corner of each image.

Table 2. Result of recognition

Method / Motion	Precision [%]		
	Structural features	Figural features	Both features
Normal walk	83.7	51.0	97.7
Forward leaning walk	80.6	62.2	94.1
Fall	84.1	78.8	87.5
Average	82.8	64.0	93.1

Table 3. Result of recognition with respective subjects

Motion	Precision [%]					
	A	B	C	D	E	**Ave**
Normal walk	95.0	99.1	97.0	98.3	99.1	97.7
Forward leaning walk	90.2	98.2	96.3	96.0	89.9	94.1
Fall	91.3	86.5	94.6	88.0	77.3	87.5

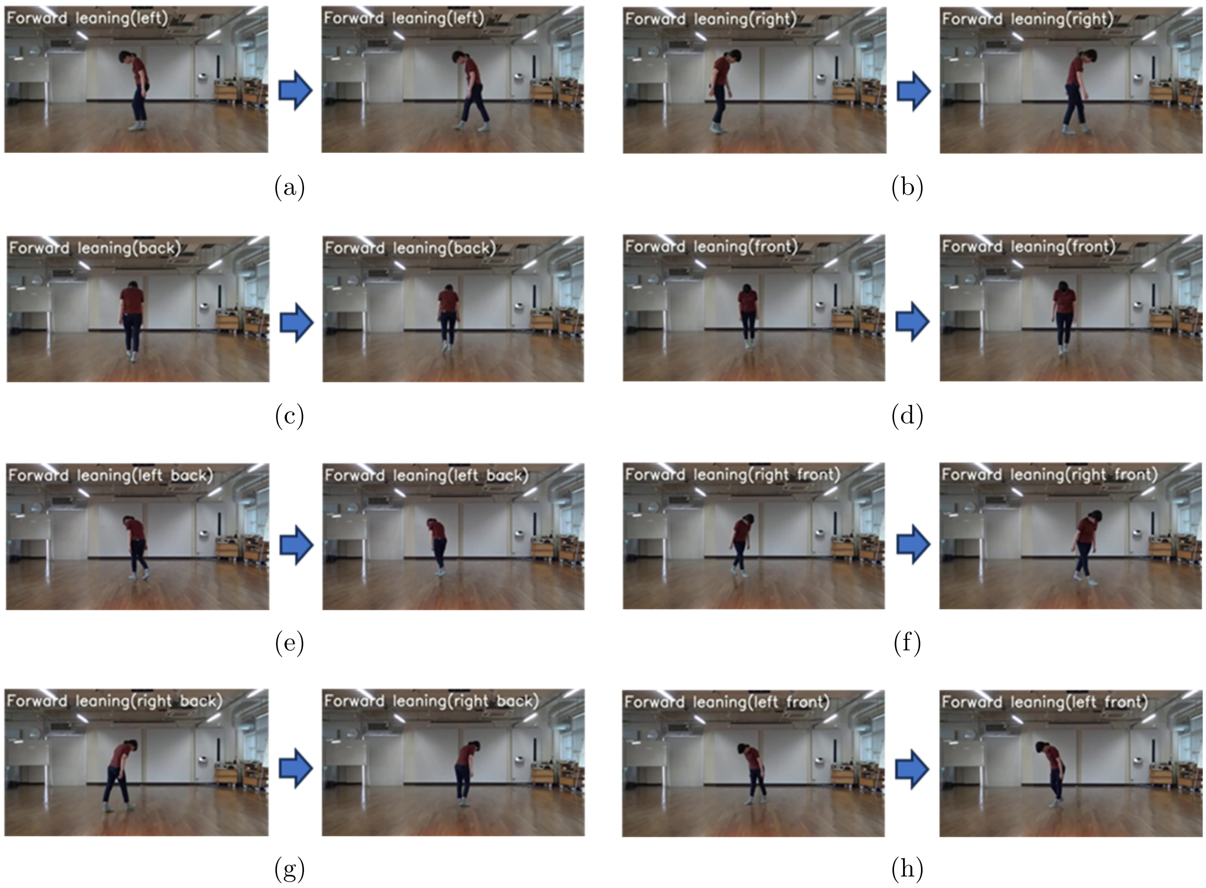


Figure 6. Forward leaning motion and the result of recognition: Walking to (a) the left, (b) the right, (c) the back, (d) the front, (e) the left back, (f) the right front, (g) the right back, and (h) the left front. The recognition results are shown in the upper left corner of each image.

6. **Discussion.** We proposed a method of recognizing a human walk posture using structural features defined by joint coordinates and figural features defined by human domain images. As shown in Table 3, the average precision of the recognition with the normal walk, forward leaning walk and fall was 97.7%, 94.1%, and 87.5%, respectively. The overall average accuracy was 93.1%. This is much better result than using the structural features or figural features independently as seen in Table 2. Although figural features may have a certain relation to structural features, the combined use of overall and local features contributed to satisfactory precision of the recognition.

In some cases, the motion of the initial frame of the videos was misrecognized. This may be due to the fact that appropriate values were not obtained for the features that use values from previous frames, such as the maximum and the minimum value, and speed. In other cases, we acquired images with missing human regions, and the background areas were also extracted as human regions. The incorrect foreground extraction was also a cause of the misidentification. These issues in video processing need to be improved.

As shown in Table 2, the motion, fall, has worse precision compared to the other two motions. Table 3 shows that the variance of the precision of fall is larger than other motions among the subjects. Actually, fall may be a little difficult to act for a subject. In practice, there were individual differences in the way of fall, such as the form of fall or fall in an oblique direction. Other causes of misrecognition are based on individual differences in the form and speed of walk and fall. Further consideration on data acquisition including collecting more training data needs to be done.

The proposed method uses a number of features for human motion posture recognition. By finding the features related to a biased posture caused by aging, a disease, or a habit, appropriate advice having a point will be given to the person concerned, which should be one of the abilities a future care robot possesses.

7. **Conclusion.** In this paper, we proposed a human motion recognition method using 58 features describing a human posture and Random Forest as a discriminator. The features include structural features defined by the joint coordinates and figural features defined by human domain images. The method was applied to the recognition of three types of motion, i.e., normal walk, forward leaning walk, and fall with eight directions of travel (left/right, depth and oblique). Experimental results showed that the average precision of the recognition was 93.1%. The main cause of misrecognition is individual differences in walk and fall style, such as walk speed, foot lift and body lean.

Future work includes treatment of initial frames in a video, more accurate foreground extraction and collecting more training data. Increasing the types of human motions and raising the precision of the recognition further also need to be investigated. The discrimination of the case in which part of the body is hidden by furniture and the case in which multiple movements and directions are mixed remain for further study.

REFERENCES

[1] Cabinet Office, *Current and Future Prospects of Aging Population*, Annual Report on the Aging Society, FY2022 1-1, https://www8.cao.go.jp/kourei/whitepaper/w-2023/zenbun/pdf/1s1s_01.pdf, Accessed on August 9, 2023.

[2] Cabinet Office, *Health and Welfare*, Annual Report on the Ageing Society, FY2022 1-2, https://www8.cao.go.jp/kourei/whitepaper/w-2023/zenbun/pdf/1s2s_02-1.pdf, Accessed on August 9, 2023.

[3] T. Kimura, H. Kobayashi, E. Nakayama and M. Hanaoka, Effects of aging on gait patterns in the healthy elderly, *Anthropological Science*, vol.115, pp.67-72, 2007.

[4] A. Kawai, S. Taniguchi, H. Li and S. Izumi, A physical condition measurement method for elderly people using motion sensor, *Multimedia, Distributed, Cooperative, and Mobile Symposium*, vol.2019, pp.587-594, 2019.

[5] F. Anderson, T. Grossman, J. Matejka and G. Fitzmaurice, YouMove: Enhancing movement training with an augmented reality mirror, *Proc. of the 26th Annual ACM Symposium on User Interface Software and Technology*, pp.311-320, 2013.

[6] C. Chen, R. Jafari and N. Kehtarnavaz, UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and wearable inertial sensor, *2015 IEEE International Conference on Image Processing*, pp.168-172, 2015.

[7] Y. Wada, S. Suzuki and M. Yamamoto, A 3D motion capture for outdoor pedestrian, *IPSJ SIG Technical Report (CG)*, vol.2010, no.4, pp.1-6, 2010.

[8] N. Zhang, Z. Hu, S. Lee and E. Lee, Human action recognition based on global silhouette and local optical flow, *Proc. of the International Symposium on Mechanical Engineering and Material Science (ISMEMS 2017)*, vo.134, pp.1-5, 2017.

[9] Y. Takahashi, Y. Otoda, M. Fujimoto and Y. Arakawa, Walking activity detection through a cane attached sensor, *IPSJ Transactions on CDS*, vol.8, no.2, pp.43-55, 2018.

[10] P. H. Truong, J. Lee, A.-R. Kwon and G.-M. Jeong, Stride counting in human walking and walking distance estimation using insole sensors, *Sensors (Basel)*, vol.16, no.6, pp.1-15, 2016.

[11] J. K. Tan and T. Kurosaki, Estimation of self-posture of a pedestrian using MY VISION based on depth and motion network, *Journal of Robotics, Networking and Artificial Life*, vol.7, no.3, pp.152-155, 2020.

[12] Z. Liu and J. K. Tan, Analysis of human walking posture using a wearable camera, *International Journal of Innovative Computing, Information and Control*, vol.19, no.3, pp.805-819, 2023.

[13] H. Chao, K. Wang, Y. He and J. Zhang, GaitSet: Cross-view gait recognition through utilizing gait as a deep set, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.44, no.7, pp.3467-3478, 2022.

[14] B. B. Amor, A. Srivastava, P. Turaga and G. Coleman, A framework for interpretable full-body kinematic description using geometric and functional analysis, *IEEE Transactions on Biomedical Engineering*, vol.67, no.6, pp.1761-1774, 2022.

[15] Z. Cao, T. Simon, S. Wei and Y. Sheikh, Realtime multi-person 2D pose estimation using part affinity fields, *Proc. of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, no.121, pp.1302-1310, 2017.

[16] J. Sakurai and M. Kobayashi, Proposal of analysis support system for throwing motion in elementary school physical education classes using cameras, *IPSJ Symposium Series*, vol.2021, pp.22-29, 2021.

[17] T. Oshima, R. Tanaka, R. Hamasaki and K. Nakayama, Preliminary examinations of support systems for dementia care, *IPSJ SIG Technical Report*, vol.2019-AAC-11, no.2, pp.1-7, 2019.

[18] Z. Oh, S. Takahashi and B. Siduki, Posture support system during body weight training using a smartphone, *IPSJ SIG Technical Report (HCI)*, vol.2019-HCI-183, no.5, pp.1-6, 2019.

[19] J. Cao, Y. Yamashita and J. K. Tan, Human motion recognition using TMRIs with extended HOOF, *Journal of Robotics, Networking and Artificial Life*, vol.7, no.4, pp.231-235, 2021.

[20] S. Kumawat, T. Okawara, M. Yoshida, H. Nagahara and Y. Yagi, Action recognition from a single coded image, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.45, no.4, pp.4109-4121, 2023.

[21] S. Lee, H. Eun, J. Moon, S. Choi, Y. Kim, C. Jung and C. Kim, Learning to discriminate information for online action detection: Analysis and application, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.45, no.5, pp.5918-5934, 2023.

[22] S. Honari1, V. Constantin, H. Rhodin, M. Salzmann and P. Fua, Temporal representation learning on monocular videos for 3D human pose estimation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.45, no.5, pp.6415-6427, 2023.

[23] L. Geng, W. Yang, Y. Jiao, S. Zeng and X. Chen, A multilayer human motion prediction perceptron by aggregating repetitive motion, *Machine Vision and Applications*, vol.34, no.98, pp.1-13, 2023.

[24] M. Fisch and R. Clark, Orientation keypoints for 6D human pose estimation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.44, no.12, pp.10145-10158, 2022.

[25] Y. Wang, X. Zhang, Y. Shen, B. Du, G. Zhao, L. Cui and H. Wen, Event-stream representation for human gaits identification using deep neural networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.44, no.7, pp.3436-3449, 2022.

[26] *OpenPose Doc-Output Pose Output Format (BODY_25)*, https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/02_output.md, Accessed on July 19, 2024.

[27] L. Breman, Random forests, *Machine Learning*, vol.45, no.1, pp.5-32, 2001.

[28] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz and M. Shah, Deep learning-based human pose estimation: A survey, *J. ACM*, vol.37, no.4, Article 111, pp.1-35, 2018.

[29] C. Stauffer and W. E. L. Grimson, Adaptive background mixture models for real-time tracking, *Proc. of Conference on Computer Vision and Pattern Recognition*, vol.2, pp.246-252, 1999.

[30] A. Shimada, D. Arita and R. Taniguchi, Increment and decrement of Gaussians in adaptive mixture-of-Gaussian background models, *MIRU2006*, pp.741-751, 2006.

Author Biography

Miki Ooba received the B.E. and M.E. degrees in Control Engineering Course from Kyushu Institute of Technology, in 2022 and 2024, respectively. She is currently a system engineer. Her research interests include motion analysis of elderly people in order to improve their QOL.

Yui Tanjo received the Ph.D. degree from Kyushu Institute of Technology. She is currently a professor with the Department of Mechanical and Control Engineering, Kyushu Institute of Technology. Her current research interests include ego-motion analysis by MY VISION, three-dimensional shape/motion recovery, human detection, and its motion analysis from video. She was awarded SICE Kyushu Branch Young Author's Award in 1999, the AROB Young Author's Award in 2004, the Young Author's Award from IPSJ of Kyushu Branch in 2004, and the BMFSA Best Paper Award in 2008, 2010, 2013 and 2015. She is a member of IEEE, The Information Processing Society, and The Institute of Electronics, Information and Communication Engineers of Japan.