Human Motion Recognition from Multiple Directions and Its Gait Cycles Analysis

Miki Ooba
Graduate School of Engineering, Kyushu Institute of Technology, 1-1 Sensui, Tobata-ku, Kitakyushu, 804-8550, Japan

Yui Tanjo
Faculty of Engineering, Kyushu Institute of Technology, 1-1 Sensui, Tobata-ku, Kitakyushu, 804-8550, Japan
Email: ooba.miki558@mail.kyutech.jp, tanjo@cntl.kyutech.ac.jp

Abstract

It is crucial for individuals to keep walking and stay healthy to prevent receiving nursing care. This paper proposes a method of recognizing walk motions and analyzing the gait cycle of a human focusing on his/her posture. We use 43 structural features defined from human joint coordinates obtained using OpenPose and 18 figural features from human domain images and their difference images. The feature vector containing these 61 features is used for the recognition of walk motion by Random Forest. In the experiment, we applied the method to recognizing six types of motions and analyzed the walk gait cycles of five persons, and obtained satisfactory results.

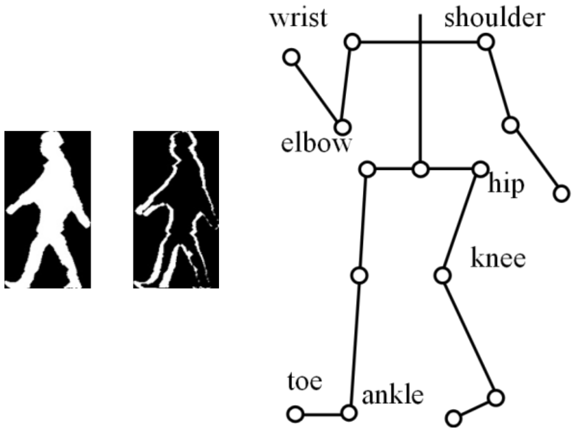
Keywords: OpenPose, images of human area, human walk motion, structural feature, figural feature, gait cycle

1. Introduction

According to the Ministry of Health, Labour and Welfare's 'Physical Activity Standards for Health Promotion' [1], increasing the amount of daily physical activity can reduce the risk of deterioration of life functions and other problems. Therefore, it is recommended to maintain a certain intensity of physical activity (walking or equivalent movements). The purpose of this research is to recognize human daily motions and analyze the information obtained from them.

Previous studies include posture and motion recognition methods that use a chest-mounted camera by the first person viewpoint called MY VISION [2][3], gait recognition by integrating the gait silhouette input into a CNN [4], and a method that evaluates movement based on the skeletal trajectory of the whole body using depth data [5]. There is a motion recognition method [6] using an extension of Motion History Image (MHI) called Triplet Motion Representation Images, with Histograms of Oriented Optical Flow. However, attaching a measurement device to a person may cause unnatural movements. It is necessary to consider an actual living environment.

In this paper, we propose a method of recognizing human posture and gait motion based on human joint positions and image features, as well as a method of analyzing gait motion based on the gait cycle, using a random forest classifier [7] based on a 61-dimensional feature vector.



(a) (b)

Fig. 1. Figural and structural features. (a) Human domain image, (b) joint coordinates

2. Foreground Extraction

To recognize a motion from the shape features of a human domain image, the human domain is extracted from the image as the foreground. In this paper, the sequential background estimation method based on the Gaussian mixture model [8][9] is used to extract a human domain corresponding to background change. Then, the noise on the extracted human domain is removed by expansion and contraction processing and it is trimmed to a rectangle of a specified scale. Figural features are extracted from the human domain using the binary and frame difference images shown in Fig. 1(a). After having the foreground, figural features are extracted by the step in 3.1

3. Feature Extraction

We use figural features and structural features to recognize human motion and its gait cycles.

3.1 Figural Features

Our aim is to extract effective features from human posture images. Eight kinds of figural features used in the proposed method can be categorized as follows.

1) Image aspect ratio: The ratio of the height to the width of the human domain image.

2) Percentage of a white pixel area in a rectangle: The percentage of a white pixel area in the human domain image.

3) Hand swing **(**upper body width): The difference between the rightmost and the leftmost points at 60% of the height of the upper body of the human domain image.

4) Shoulder height: The *y* coordinate at which the number of white pixels per line exceeds a threshold when the image is scanned from the upper left to the lower right. The threshold was experimentally set at 40 pixels to distinguish between a stretch out the back posture and others.

5) Length of the contour line at the bottom of the human region image: The length of the contour line at the bottom 40% of the height of the person area image.

6) Body asymmetry: The center of gravity of the lower 40% portion of the person area, and the percentage of the area to the left of the center of gravity in the white pixel area.

7) Position of the center of gravity: The distance of the center of gravity of the head from the centerline of the image.

8) Number of white pixels of the difference image: The number of white pixels in the 25% height area from the bottom of the difference image.

Since features 2), 3) and 5) take periodic values, their minimum and maximum values are also extracted and used as features. In this way, 18-dimensional figural features are obtained from the image of the human domain.

3.2 Structural Features

We extract the following six structural features (43 dimensions) using human joint coordinates provided from OpenPose [10].

1) Knee, ankle, and elbow angles: The knee angle θ_{knee}[deg] is calculated using the inner product of the hip 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. Using the same method, the ankle angles θ_{ankle} (the inner product of toe and knee)and elbow angle θ_{elbow} (the inner product of wrist and shoulder) are also obtained.

2) Ankle-toe, Hip-ankle angle, and Body tilt: The angle between the ankle and toe relative to the horizontal line is defined as the ankle-toe angle θ_{instep} [deg] and is given by;

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

where (x_{toe}, y_{toe}) is the coordinates of the toe as shown in Fig. 1(b). Similarly, we determine the hip-ankle angle $\theta_{hip-ankle}$ (The angle between the vertical line of the hip-ankle and the horizontal line of the left and the right hip) and the body inclination $\theta_{bodytilt}$ (The angle between the neck and the vertical line of the waist center).

3) Steps, Hip-ankle distance, and Wrist-shoulder distance

(a) Scale transformation: Using the length of a person's $thighs$ [cm], the parameter $scale$ [cm/pixel], which indicates how many centimeters one pixel corresponds to in the current frame, is calculated by the following;

$$scale = \frac{thighs}{\sqrt{(x_{hip} - x_{knee})^2 + (y_{hip} - y_{knee})^2}} \tag{5}$$

(b) Step width, Hip-ankle distance, Wrist-shoulder distance: The step width is calculated using the Euclidean distance between the left heel and the right heel as follows;

$$\begin{aligned} &step\_\mathrm{width} \\ &= scale \\ &\times \sqrt{(x_{Rheel} - x_{Lheel})^2 + (y_{Rheel} - y_{Lheel})^2} \end{aligned} \tag{6a}$$

where (x_{Rheel}, y_{Rheel}) and (x_{Lheel}, y_{Lheel}) are the coordinates of the right and the left heel, respectively. Concurrently, the Euclidean distance between hip-ankle and wrist-shoulder is calculated as follows;

$$*\_\# = scale \times \sqrt{(x_* - x_\#)^2 + (y_* - y_\#)^2} \tag{6b}$$

where $*$ indicates hip (or wrist), and # indicates ankle (or shoulder). (x_*, y_*) and $(x_\#, y_\#)$are the coordinates of hip and ankle or wrist and shoulder, respectively.

4) Walk speed: The feature $walk\_speed$ [cm/s] indicates how long 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;

$$walk\_speed = scale \times \frac{\sqrt{(x_{c0} - x_{cn})^2 + (y_{c0} - y_{cn})^2} \times frame\_rate}{n} \quad (7)$$

where (x_{c0}, y_{c0}), (x_{cn}, y_{cn}) are the coordinates of the center of the waist between the current frame and previous n frames, whereas the $frame\_rate$ is a unit that indicates the number of the images makeup in one second in a video.

5) 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 $scale$.

$$\begin{aligned} x &= (x_{c0} - x_{cn}) \times scale \\ y &= (y_{c0} - y_{cn}) \times scale \end{aligned} \quad (8)$$

6) Difference of the foot joint heights and wrist height: The height of the knee, and heel $height_{\_*}$ [cm] is calculated by the difference between the heights of the right and the left $height\_{*}$, as follows;

$$height\_{*} = |(y_{R*} - y_{L*})| \times scale \quad (9)$$

where * indicates knee, or heel, (x_{R*}, y_{R*}) and (x_{L*}, y_{L*}) are the coordinates of the right and the left knee, or heel, respectively. Moreover, the height from floor to toe and the height from floor to wrist are calculated in the same way.

Features such as knee and ankle angles are calculated with the left and the right foot, respectively. The feature values 1), 3), and 6) are then repeated for similar values in each gait cycle (e.g., in the walking motion, a person repeatedly bends and extends his/her knees). Therefore, in addition to the features 1) to 6) above in the current frame, the minimum and maximum values of these features in the past n frames are included as features. This results in 43 dimensional features.

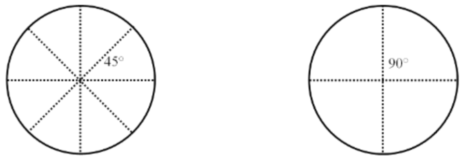
Finally, from the figural and structural features, we represent a human posture using a 18+43=61-dimensional feature vector.

4. Learning and Identification

The proposed method uses Random Forest [7] as a discriminator.

Table 1. The number of frames in the video.

Motion	The number of frames				
	A	B	C	D	E
Normal	2351	2101	1878	2124	2192
Forward leaning	3577	2830	2214	2290	2512
Fall	400	407	405	309	278
Help	394	327	511	375	565
Sit	1322	1039	1336	877	849
Sleep	1141	712	991	649	860
Total	9185	7416	7335	6624	7256



(a) walk (b) fall and sleep

Fig. 2. Direction of the motion

5. Experiment

5.1. Experimental Method

One fixed camera is set up at a height of 90 cm from an indoor floor to simulate a real-life situation. A person appears in the video, and the whole body is taken a video. The experiment was conducted on five subjects (22-23 years old). They are referred to as A, B, C, D, and E. Table 1 shows the number of frames contained in each video.

Six types of motions: Normal walking, forward leaning walk, falling, asking for help (beckoning to the camera), sitting, and sleeping are captured on video. Fig. 2 shows the direction.

Gait cycle [11]: Gait cycle (seven periods) recognition is performed for normal walking, i.e. when walking from the right to the left, the target of the gait cycle is left foot, whereas when walking from the left to the right, the target of the gait cycle is right foot. So we simply call it left-right foot. We manually select seven periods of the normal walk and put the selected periods into seven classes for training. The seven classes (periods) are Loading Response (LR), Mid Stance (MSt), Terminal Stance (TSt), Pre swing (PSw), Initial Swing (ISw), Mid Swing (MSw), Terminal Swing (TSw). In the gait cycle, 58 of the 61-dimensional features mentioned in section 3 are used for training. However, posture Initial Contact at the start of the cycle is omitted.

5.2. Method of Evaluation

Leave-one-out cross-validation was applied to the experimental data. The percentage of correct frames to all frames is used for evaluation. In the gait cycle analysis, left-right balance is evaluated by determining the left-right difference in the percentage of each period to the full video.

Table 2. Result of motion recognition

Motion	Precision [%]					
	A	B	C	D	E	**Ave**
Normal walk	97.0	99.4	93.8	99.6	98.8	97.8
Forward leaning	91.5	98.4	89.6	98.9	95.6	94.7
Fall	81.3	75.2	87.7	87.7	78.4	82.0
Help	79.7	82.0	95.1	99.2	99.3	92.1
Sit	100	100	92.5	100	100	98.2
Sleep	92.3	96.1	99.0	93.1	90.6	94.2

Table 3. Result of gait cycle recognition

Phases	Periods (class)	Precision [%]					
		A	B	C	D	E	**Ave**
Stance phase	LR	45.0	61.4	63.9	76.8	67.4	62.4
	MSt	88.7	90.7	90.8	94.5	85.3	89.9
	TSt	97.4	93.6	79.0	98.0	81.2	89.6
	PSw	55.8	77.6	77.3	68.5	67.8	67.8
Swing phase	ISw	87.9	90.4	86.8	87.4	75.2	85.4
	MSw	84.7	83.0	81.6	80.4	82.5	82.6
	TSw	91.8	81.0	66.1	90.9	78.5	81.9

6. Results and Discussion

6.1. Results

Tables 2 and 3 show the motion and gait cycle recognition results, respectively. Note that, help and sit motion are recognized without considering the direction. Fig. 3 shows some of the images of the recognition results. The upper left of the image shows the motion and gait cycle recognition results. Table 4 shows the different ratios of each period of the gait cycle of walking from the right to the left (the gait cycle of left foot) and the left to the right (the gait cycle of right foot) of each subject.

6.2. Discussion

Motion Recognition: The overall average accuracy was 95.2%. As shown in Table 2, Fall was less accurate than other motions, and the variation among subjects was greater than others. In fact, there are individual differences in the way the subject falls and shifts diagonally. Collecting more training data is needed.

Gait cycle: The overall average accuracy was 82.4%. In many cases, the gait cycle class was recognized as the class before or after the class. This is because walking is a continuous motion and the before and after postures are similar. From Table 4, it is expected that subject A's Mst is shorter than that of the left, indicating a slight decrease in the muscle strength of the right leg. Since analysis of a subject gait cycle is dependent on the recognition rate, the recognition accuracy in each class needs to be improved.

7. Conclusion

In this paper, we proposed a motion recognition and gait cycle analysis method using 61 features representing human posture and Random Forest as a discriminator. It was applied to the recognition of 6 motions and the analysis of 7 periods of gait cycles. Further study is needed to obtain more accurate values, such as step width, for human health, and to identify the cases where multiple movements and directions are mixed. If we can provide accurate and appropriate advices to individuals, we will be able to offer better health services.

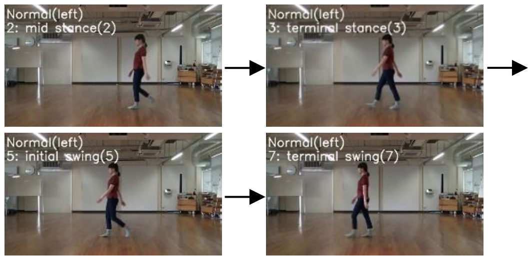


Fig. 3. Normal walk and the result of recognition

Table 4. The difference of left-right foot.

Phases	Periods (class)	left-right foot differences [%]				
		A	B	C	D	E
Stance phase	LR	0.3	-5.4	-2.4	2.0	2.5
	MSt	4.4	-0.3	0.3	0.8	-1.3
	TSt	-1.9	1.4	0.7	0.9	0.8
	PSw	-2.1	2.3	1.0	-1.1	1.1
Swing phase	ISw	0.6	-4.3	1.0	-3.7	0.4
	MSw	-0.4	1.4	0.1	2.7	0.7
	TSw	-0.8	4.8	-0.7	-1.7	-4.2

References

1. Cabinet Office, “Physical Activity Standards for Health Promotion 2013”, 2013.
2. J. K. Tan, 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.
3. Z. Liu, 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.
4. H. Chao, K. Wang, Y. He, 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.
5. B. B. Amor, A. Srivastava, P. Turaga, 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.
6. J. Cao, Y. Yamashita, 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.
7. L. Breman, “Random forests”, Machine Learning, Vol. 45, No. 1, pp. 5-32, 2001.
8. C. Stauffer, W. E. L. Grimson: “Adaptive background mixture models for real-time tracking”, Proceedings of Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 246-252, 1999.
9. A. Shimada, D. Arita, R. Taniguchi, “Increment and decrement of Gaussians in adaptive mixture-of-Gaussian background models”, MIRU2006, pp. 741-751, 2006.
10. Z. Cao, T. Simon, S. Wei, Y. Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields.”, Proc.

2017 IEEE Conference on Computer Vision and Pattern Recognition, No. 121, pp. 1302-1310, 2017.

11. S. Demura, "Fall prevention for community seniors from basic theory of falls to intervention practice", kyorin-shoin, pp.114-116, 2012. (in Japanese)

Authors Introduction

Ms. Miki Ooba

She received B.E. degree in Intelligent Control Engineering Course, Faculty of Engineering, Kyushu Institute of Technology. Her research interests include motion analysis of elderly people in order to improve their QOL.

Dr. Yui Tanjo

Dr. Tanjo 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, The Institute of Electronics, Information and Communication Engineers of Japan.