

Human Gait Gender Classification using 3D Discrete Wavelet Transform Feature Extraction

Kohei Arai¹

¹Graduate School of Science and Engineering
Saga University
Saga City, Japan

Rosa Andrie Asmara^{1,2}

²Informatics Management Department
State Polytechnics of Malang
Malang, Indonesia

Abstract—Feature extraction for gait recognition has been created widely. The ancestor for this task is divided into two parts, model based and free-model based. Model-based approaches obtain a set of static or dynamic skeleton parameters via modeling or tracking body components such as limbs, legs, arms and thighs. Model-free approaches focus on shapes of silhouettes or the entire movement of physical bodies. Model-free approaches are insensitive to the quality of silhouettes. Its advantage is a low computational costs comparing to model-based approaches. However, they are usually not robust to viewpoints and scale. Imaging technology also developed quickly this decades. Motion capture (mocap) device integrated with motion sensor has an expensive price and can only be owned by big animation studio. Fortunately now already existed Kinect camera equipped with depth sensor image in the market with very low price compare to any mocap device. Of course the accuracy not as good as the expensive one, but using some preprocessing method we can remove the jittery and noisy in the 3D skeleton points.

Our proposed method is to analyze the effectiveness of 3D skeleton feature extraction using 3D Discrete Wavelet Transforms (3D DWT). We use Kinect Camera to get the depth data. We use Iposoft mocap software to extract 3d skeleton model from Kinect video. From the experimental results shows 83.75% correctly classified instances using SVM.

Keywords—gender gait classification; 3D Skeleton Model; SVM; Biometrics; 3D DWT

I. INTRODUCTION

In recent years, there has been an increased attention on effectively identifying individuals for prevention of terrorist attacks. Many biometric technologies have emerged for identifying and verifying individuals by analyzing face, fingerprint, palm print, iris, gait or a combination of these traits [1]–[3].

Human Gait as the classification and recognition object is the famous biometrics system recently. Many researchers had focused this issue to consider for a new recognition system [4]–[11]. Human Gait classification and recognition giving some advantage compared to other recognition system. Gait classification system does not require observed subject's attention and assistance. It can also capture gait at a far distance without requiring physical information from subjects.

There is a significant difference between human gait and other biometrics classification. In human gait, we should use

video data instead of using image data as other biometrics system used widely. In video data, we can utilize spatial data as well as temporal data compare to image data.

There are 2 feature extraction method to be used in gait - classification: model based and free model approach [12]. Model-based approaches obtain a set of static or dynamic skeleton parameters via modeling or tracking body components such as limbs, legs, arms and thighs. Gait signatures derived from these model parameters employed for identification and recognition of an individual. It is obvious that model-based approaches are view-invariant and scale-independent. These advantages are significant for practical applications, because it is unlikely that reference sequences and test sequences taken from the same viewpoint. Model-free approaches focus on shapes of silhouettes or the entire movement of physical bodies. Model-free approaches are insensitive to the quality of silhouettes. Its advantage is a low computational costs comparing to model-based approaches. However, they are usually not robust to viewpoints and scale [13].

Gender classification along with human gait recognition has getting the researchers to find its best methods. Wide implementation make they seem so attractive research. The implementation will not only enhance existing biometrics systems but can also serve as a basis for passive surveillance and control in “smart area” (e.g., restricting access to certain areas based on gender) and collecting valuable demographics (e.g., the number of women entering a retail store, airports, post office, or public smoking area etc. on a given day)

Imaging technology developed quickly this decades. Motion capture (mocap) device integrated with motion sensor has an expensive price and can only be owned by big animation studio. Fortunately now already existed Kinect camera equipped with depth sensor image in the market with very low price compare to any mocap device. Of course the accuracy not as good as the expensive one, but using some preprocessing we can remove the jittery and noisy in the 3D skeleton points. Our proposed method is part of model based feature extraction and we call it 3D Skeleton model. 3D skeleton model for extracting gait itself is a new model style considering all the previous model is using 2D skeleton model. The advantages itself is getting accurate coordinate of 3D point for each skeleton model rather than only 2D point. We use Kinect to get the depth data. We use Iposoft mocap software to extract 3d skeleton model from Kinect video. Those 3D skeleton model exported to BVH animation standard format

file and imported to our programming tool which is Matlab. We use Matlab to extract the feature and use a classifier. We create our own gender gait dataset in 3D environment since there are not exist such a dataset before.

II. PROPOSED METHOD

The classification of gender gait quality in this paper consists of three part, preprocessing, feature extraction, and classification. Figure 1 shows the complete overview of proposed human disable gait quality classification.

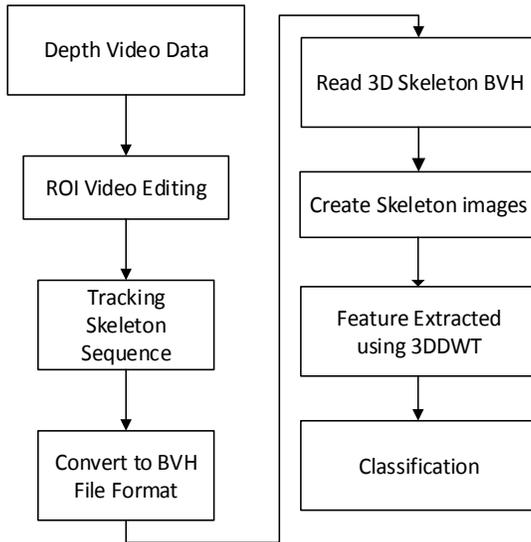


Figure 1: Proposed human gait gender classification

Using Kinect camera have one advantage compare to usual RGB camera. The skeleton created is in 3D space. One can get 2D images from different view angle using only single camera. Figure 2 below shows the 2D skeleton image created from different view angle at the same frame. This is useful to enhance the accuracy of the classification since some paper proposed using multi view image [10], [14]–[16]. However, these papers will only using one view for the analysis.

A. Preprocessing

First, take the Video data using Kinect and IpiRecorder to record the depth data along with RGB video data. To get the video data, there are some recommendation should be considered:

1. Using 9 by 5 feet room space to get best capture.
2. Object should be dressed in casual slim clothing, avoid shiny fabrics.
3. We should ensure that the whole body including arms and legs is visible during the recording states. Beginning from T-Pose and the recording can be started.

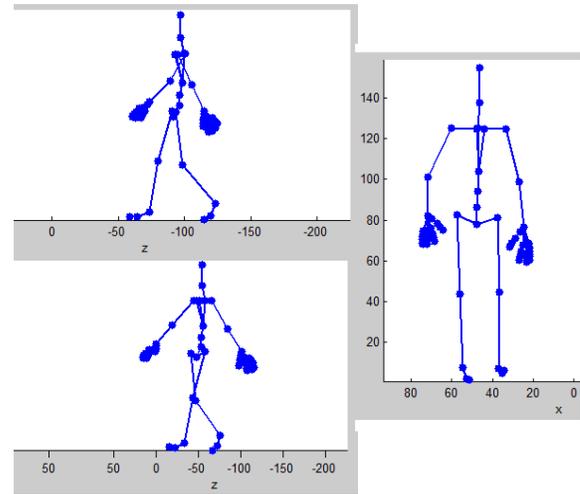


Figure 2. 2D skeleton image created from different view angle at the same frame

Second, processed the depth video data in IPISoft motion capture application. IPISoft will create the 3D skeleton model from video depth recorded using some tracking motion method. The first step is to take only the gait scene, and remove unimportant video scene or we call the Region of Interest (ROI) video. Figure below show the example of video recording.

Third, Create the skeleton 3d model using the tracking motion method, remove the jittery and noises, and export the skeleton model to BVH file format in IPISoft.

Fourth, Read the BVH file, extracted the feature, and classify the feature.

B. Dataset

Unfortunately, there are no Kinect Video Depth gait dataset exists until now. All exist gait dataset is using ordinary RGB camera like USF gait dataset, SOTON gait dataset, and CASIA gait dataset. Figure 3 shows the example of CASIA gait dataset.



Figure 3: Example of CASIA gait dataset

To conduct the experiment, we should prepare the dataset. We will use the Kinect Gait Dataset to analyze and classify gender using gait. The proposed research will search the capability of Kinect and 3D Skeleton model and use their 2D images for gait classification.

Figure 4 below shows the T-pose position before the video recording start. The top right image showing the RGB video sequence. T

he color gradient used to represents the depth in video data. Blue color means the object is close to the camera and red color means the object is far from camera.

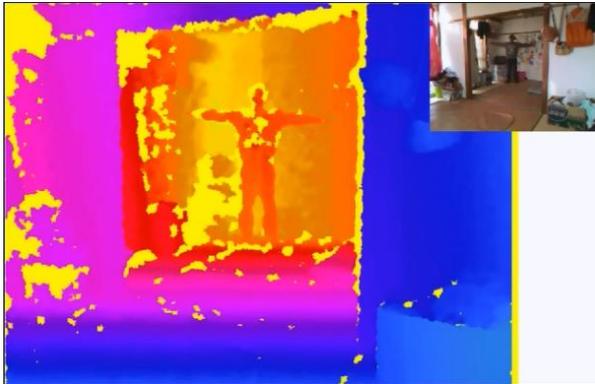


Figure 4: T-Pose Position before the recording begin

Figure 5 below show the 3D skeleton tracking motion sequence. First task is specifying subject's physical parameter like gender and height. IpiSoft will detect the ground plane automatically and provide the 3D skeleton in T-Pose position. Our next job is try to put the T-Pose skeleton in the same position with the subject T-Pose position in the first sequence of video.

This time also we should determine the Region of Interest video to be processed. Instead of all the video sequence that we use, we could only take the most important part of the video sequence. Once we put the skeleton to the same position with the subject, we can refitting pose using the application and start tracking. Jittery removal and Trajectory filtering can be done after the tracking finished.

The skeleton sequence result can be import to BVH file standard. Figure 6 and 7 below shows the BVH file result and preview in BVH file viewer and Matlab.

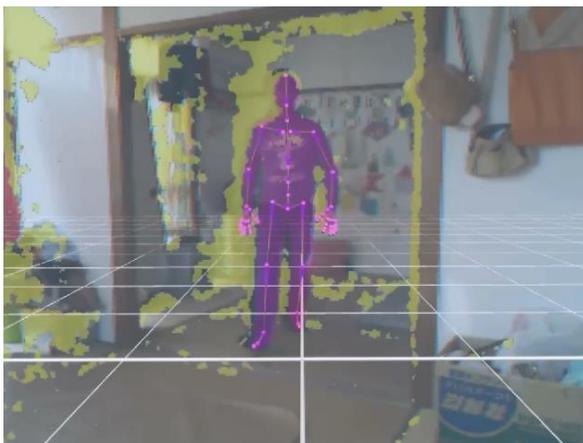


Figure 5: Skeleton motion tracking sequence

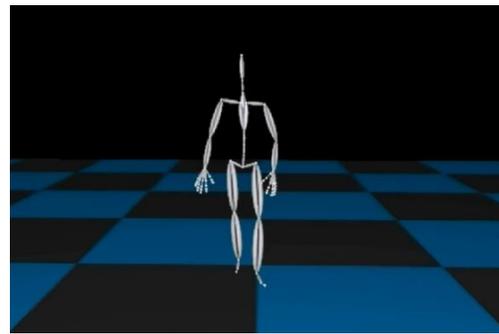


Figure 6: BVH skeleton results

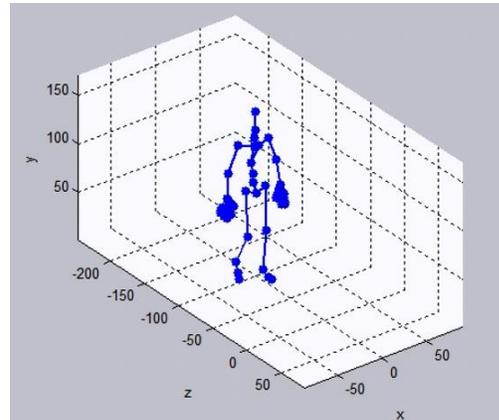


Figure 7: Skeleton model imported in MATLAB

C. 3D Discrete Wavelet Transforms (3D DWT)

3D version of Discrete Wavelet Transform is specially used in volume and video processing. In the 3D case, the 1D analysis filter bank is applied in turn to each of the three dimensions. If the data is of size N_1 by N_2 by N_3 , after applying the 1D analysis filter bank to the first dimension one have two sub band data sets, each of size $N_1/2$ by N_2 by N_3 . After applying the 1D analysis filter bank to the second dimension one have four sub band data sets, each of size $N_1/2$ by $N_2/2$ by N_3 . Applying the 1D analysis filter bank to the third dimension gives eight sub band data sets, each of size $N_1/2$ by $N_2/2$ by $N_3/2$. This is illustrated in the figure 8 below.

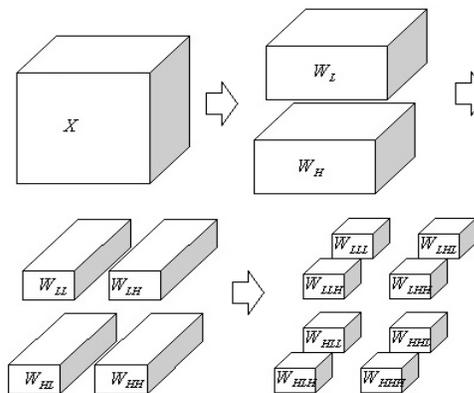


Figure 8. The resolution of a 3-D signal is reduced in each dimension

The block diagram of the 3D analysis filter bank is shown in Figure 9.

D. Feature Extraction

To extract the feature using 3D Discrete Wavelet Transform, we can prepare 2 kind of data. The first data is raw data and the second data is the resized data. The effectiveness and Classification accuracy of each data using statistical feature will be shown in this paper. The resized data have an advantage over the raw data. In Resize data, we can use whole sub bands or decomposition data and process them in the classifier directly.

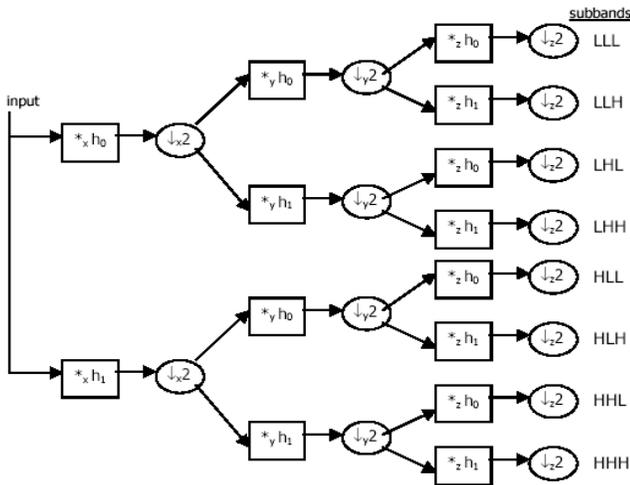


Figure 9. Block diagram of a single level decomposition for the 3D DWT

This can be done because all the dataset have the same dimension. This paper will cover the experimental result of resized data and extracted statistical feature, but this paper will not discuss analyzing of using whole the sub bands feature extraction. To extract the feature in the dataset created, one have to consider about the image size. If all the image is used, it will be costly. One have to extract the skeleton image only and not all the image, thus call it Region of Interest (ROI). One can do automatic ROI using simple image detection since the image is in binary space. After the ROI done, one can extract feature from the data directly or one can resize the data. Thus, there are two kind of data which is raw data and resized data.

Data have to same in image file and frame amount to be used as a resized data. The method used in this paper to create resized data is image resizing and frame cropping. In image resizing, biggest skeleton image ROI will be used as a reference because in this method we don't have to remove some amount of data and those removed data could be valuable information to the system. After finish the image resizing, one can start to crop the frame. This paper will crop the frame based on the smallest amount of video frame, thus all the data have same amount of frame. This paper using middle part of smallest frame amount as a cropped frames. If the smallest frame amount is x, then the video frame crop start at y is round $((total_frame - x)/2)$ and end in $z = y+x$.

This paper will used some famous classifier to compare and analyze their best correct classification rate. The decomposition result of Level 1 from 3D DWT of Haar Wavelet will get 8 sub bands which is LLL, LLH, LHL, LHH, HLL, HLH, HHL, and HHH. We will use 3 statistic feature that was used in previous research which is mean, standard deviation, and energy.

The formula for the energy used in Eq.(1) below.

$$E_3 = \sqrt{\sum_{k=1}^{tot_frame} \left(\sqrt{\sum_{j=1}^{tot_col} \left(\sqrt{\sum_{i=1}^{tot_row} (|Wave\ Coef|)^2} \right)^2} \right)^2} \quad (1)$$

From above explanation, we can extract 24 attribute from first and second data type. First 8 attributes is a mean for each sub bands. Second 8 attributes is a standard deviation for each sub bands, and third 8 attributes is an energy for each sub bands. All those 24 attributes namely as mean1_LLL, mean2_LLH, mean3_LHL, mean4_LHH, mean5_HLL, mean6_HLH, mean7_HHL, mean8_HHH, std1_LLL, std2_LLH, std3_LHL, std4_LHH, std5_HLL, std6_HLH, std7_HHL, std8_HHH, e3_LLL, e3_LLH, e3_LHL, e3_LHH, e3_HLL, e3_HLH, e3_HHL, and e3_HHH.

E. Classification

This paper will use two famous classifier which is Naïve Bayes and SVM to analyze and compared the results. SVM (Support Vector Machine) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier.

Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

III. EXPERIMENTAL RESULT

We start with using Resized data. From 80 dataset with 40 male and 40 female video, we will use SVM and Naïve Bayes as the classifier. First step is we are selecting the best feature in each classifier. There are two method to be use which is Wrapper Method and Ranked Method. We also conduct those two method in two kind of data preprocessing, which is without preprocessing and data after discretization filter preprocessing.

Table 1 is the result of the selected feature using both methods without preprocessing data (raw data type).

TABLE 1. FEATURE SELECTION WITHOUT PREPROCESSING DATA FOR RAW DATA TYPE

	Wrapper Method	Ranked Method
SVM	e3_LLL, e3_LLH	Std1_LLL, e3_HLH, e3_LLL, e3_HLL, e3_HHL, e3_HHH, e3_LHH
Naïve Bayes	mean2_LLH, mean3_LHL, mean6_HLH, mean7_HHL, mean8_HHH, std1_LLL, std3_LHL, std8_HHH, e3_LLL	

Table 2 is the result of the selected feature using both methods with discretization data (raw data type). Table 3 is the result of the selected feature using both methods without preprocessing data (resized data type). Table 4 is the result of the selected feature using both methods with discretization data (resized data type).

TABLE 2. FEATURE SELECTION WITH DISCRETIZATION FILTER DATA FOR RAW DATA TYPE

	Wrapper Method	Ranked Method
SVM	mean5_HLL, mean6_HLH, mean8_HHH, std1_LLL, e3_HLH	e3_HLH, std1_LLL, e3_HHH, e3_LLL, e3_HLL, e3_HHL, std2_LLH
Naïve Bayes	mean2_LLH, mean3_LHL, mean6_HLH, mean8_HHH, std1_LLL, e3_LLL, e3_HLH	

TABLE 3. FEATURE SELECTION WITHOUT PREPROCESSING DATA FOR RESIZED DATA TYPE

	Wrapper Method	Ranked Method
SVM	e3_LLL, e3_LLH, e3_HLL	std6_HLH, std8_HHH
Naïve Bayes	mean6_HLH, std8_HHH, e3_LHL	

TABLE 4. FEATURE SELECTION WITH DISCRETIZATION FILTER DATA FOR RESIZED DATA TYPE

	Wrapper Method	Ranked Method
SVM	mean5_HLL, mean8_HHH, std1_LLL, std7_HHL, std8_HHH, e3_HLL	std6_HLH, std8_HHH, std1_LLL, std4_LHH, e3_HLL, mean5_HLL, e3_LLL, std5_HLL
Naïve Bayes	mean3_LHL, mean4_LHH, mean5_HLL, mean8_HHH, std4_LHH, std8_HHH, e3_LHL, e3_LHH, e3_HHL	

Table 5 shows the result of Correct Classification Rate (CCR) for the selected features (raw data type). Table 6 shows the result of Correct Classification Rate (CCR) for the selected features (resized data type).

TABLE 5. CORRECT CLASSIFICATION RATE FOR EACH SELECTED FEATURE IN EACH CLASSIFIER FOR RAW DATA TYPE

	Wrapper Method		Ranked Method	
	No preprocess	Discretize	No preprocess	Discretize
SVM	68.75%	80%	48.75%	83.75%
Naïve Bayes	77.5%	68.75%	72.5%	76.25%

TABLE 6. CORRECT CLASSIFICATION RATE FOR EACH SELECTED FEATURE IN EACH CLASSIFIER FOR RAW AND RESIZED DATA TYPES

	Wrapper Method		Ranked Method	
	No preprocess	Discretize	No preprocess	Discretize
SVM	45%	76.25%	58.75%	75%
Naïve Bayes	56.25%	61.25%	67.5%	70%

As seen in the tables above, the best CCR is in raw data type using SVM and discretized data and ranked method selected feature. Table 7 is detail accuracy by class using SVM classifier.

TABLE 7. DETAIL ACCURACY BY CLASS USING SVM CLASSIFIER

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Male	0.775	0.1	0.886	0.775	0.827	0.838
Female	0.9	0.225	0.8	0.9	0.847	0.838
Weighted Average	0.838	0.163	0.843	0.838	0.837	0.838

IV. CONCLUSION

The proposed method uses Kinect depth sensor camera and Iposoft motion capture software to generate 3D skeleton model. Iposoft itself is special purpose application to create skeleton so user can use the motion to their computer generated character motion.

The 3D skeleton generated will then extract the 2D image in one view angle and create 2 model data type which is raw and resized video data type. Using Level 1 Haar 3D DWT, we got 8 sub bands and using 3 statistical feature for all 8 sub bands (Mean, Standard deviation, and Energy). By selecting the best feature and classify the results using SVM and Naïve Bayes, the result shows is Table 5 and Table 6. The best result achieved in raw data type using Ranked method feature selection and discretized data which is 83.75% CCR.

ACKNOWLEDGMENT

Portions of the research in this paper is funded by Saga University Japan. Authors would like to thank to Saga University for their providing of the Laboratory and Infrastructures.

REFERENCES

- [1] X. Qinhan, "Technology review - Biometrics-Technology, Application, Challenge, and Computational Intelligence Solutions," *IEEE Comput. Intell. Mag.*, vol. 2, pp. 5–25, 2007.
- [2] E. Yih, G. Sainarayanan, and A. Chekima, "Palmprint Based Biometric System: A Comparative Study on Discrete Cosine Transform Energy, Wavelet Transform Energy and SobelCode Methods," *Biomed. Soft Comput. Hum. Sci.*, vol. 14, no. 1, pp. 11–19, 2009.
- [3] Z. Yang, M. Li, and H. Ai, "An experimental study on automatic face gender classification," *Pattern Recognition, 2006. ICPR 2006. 18th Int. Conf.*, vol. 3, pp. 1099 – 1102, 2006.
- [4] N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis, "Gait recognition: A Challenging Signal Processing Technology For Biometric Identification," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 78–90, 2005.
- [5] M. S. Nixon and J. N. Carter, "Automatic Recognition by Gait," *Proc. IEEE*, vol. 94, pp. 2013–2024, 2006.

- [6] D. Cunado, M. S. Nixon, and J. N. Carter, "Automatic Extraction and Description of Human Gait Models for Recognition Purposes," *Comput. Vis. Image Underst.*, vol. 90, no. 1, pp. 1–41, 2003.
- [7] X. Li, S. Maybank, and S. Yan, "Gait components and their application to gender recognition," *Syst. Man, Cybern. Part C Appl. Rev. IEEE Trans.*, vol. 38, no. 2, pp. 145–155, 2008.
- [8] K. Arai and R. Asmara, "Human Gait Gender Classification using 2D Discrete Wavelet Transforms Energy," *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, pp. 62–68, 2011.
- [9] K. Arai and R. Asmara, "Human Gait Gender Classification in Spatial and Temporal Reasoning," *IJARAI Int. J. Adv. Res. Artif. Intell.*, vol. 1, no. 6, pp. 1–6, 2012.
- [10] G. Huang and Y. Wang, "Gender classification based on fusion of multi-view gait sequences," *Comput. Vision-ACCV 2007*, vol. 4843, pp. 462–471, 2007.
- [11] L. Lee and W. Grimson, "Gait Analysis for Recognition and Classification," in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 148–155.
- [12] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "A Review of Vision-Based Gait Recognition Methods for Human Identification," *2010 Int. Conf. Digit. Image Comput. Tech. Appl.*, pp. 320–327, Dec. 2010.
- [13] K. Arai and R. Asmara, "3D Skeleton model derived from Kinect Depth Sensor Camera and its application to walking style quality evaluations," *Int. J. Adv. Res. Artif. Intell.*, vol. 2, no. 7, pp. 24–28, 2013.
- [14] H.-C. Lian and B.-L. Lu, "Multi-view gender classification using multi-resolution local binary patterns and support vector machines.," *Int. J. Neural Syst.*, vol. 17, no. 6, pp. 479–87, Dec. 2007.
- [15] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Multi-view Gait Recognition Based on Motion Regression Using Multilayer Perceptron," *2010 20th Int. Conf. Pattern Recognit.*, pp. 2186–2189, Aug. 2010.
- [16] Y. Wang, S. Yu, and T. Tan, "Gait recognition based on fusion of multi-view gait sequences," *Adv. Biometrics*, vol. 3832, pp. 605–611, 2005.

AUTHORS PROFILE

Kohei Arai received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 and also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission A of ICSU/COSPAR since 2008. He wrote 29 books and published 500 journal papers.

Rosa A. Asmara received the B.E. degree in electronics engineering from Brawijaya University, and the M.S. degree in Multimedia engineering, from Institute of Technology Sepuluh Nopember, Surabaya, Indonesia, in 2004 and 2009, respectively. He is currently a researcher at Information Science in Saga University, Japan. His research interests include signal processing, image processing, parallel processing, pattern recognition, and computer vision.