# 3D Skeleton model derived from Kinect Depth Sensor Camera and its application to walking style quality evaluations

Kohei Arai[1]
1) Graduate School of Science and Engineering
Saga University
Saga City, Japan

Rosa Andrie Asmara[1,2]
2) Informatics 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 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 Ipisoft mocap software to extract 3d skeleton model from Kinect video. From the experimental results shows 86.36% correctly classified instances using SVM.**

*Keywords*—*Disable gait classification; 3D Skeleton Model; SVM; Biometrics*

## 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.

Gait therapist have a problem to calculate the quality improvement of the therapy that they did. They could calculate the gait quality using some device in the lab and in practical not too efficient. We propose a method that can measure gait disable quality and classify the result by only capturing the object walking in front of camera.

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 Ipisoft 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 gait disable dataset in 3D environment since there are not exist such a dataset before.

## II. PROPOSED METHOD

The classification of disable 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.
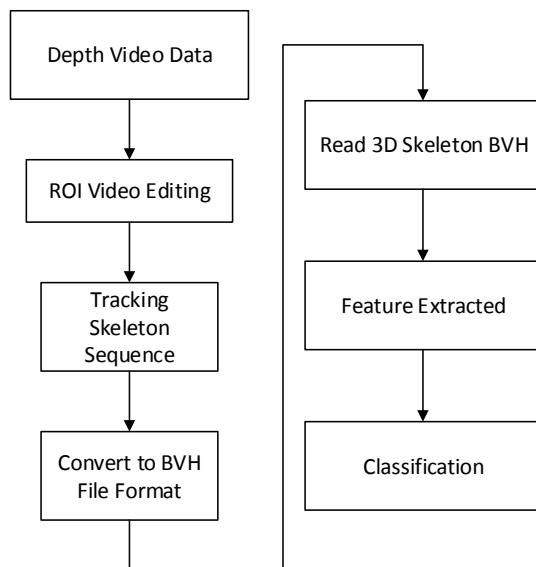
Fig.1.       Proposed human disable gait quality classification

### A. Preprocessing

First, the Video data using Kinect and IpiRecorder to record the depth data along with RGB video data is captured. 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.*

Second, the depth video data in IPISoft motion capture application is processed. 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 3 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 camera like USF gait dataset, SOTON gait dataset, and CASIA gait dataset. Figure 2 show the example of CASIA gait dataset.

Fig.2.       Example of CASIA gait dataset

To conduct the experiment, we should prepare the dataset. We will use the Kinect Gait Dataset to measure gait quality in disable person. The proposed research will analyze the capability of Kinect and 3D Skeleton model accuracy for gait classification. We also use fake gait patient to be the subject of disable gait person. The subject is analyzing neuropathic patient and classify the gait quality into 5 classes. The first class is for normal gait and calls it class A. The worst quality for neuropathy gait is called class E. The dataset provide 18 Videos each class, thus total data is 90. The dataset will provide knee left angle feature only because there will only left foot simulated.

Figure 3 shows the T-pose position before the video recording start. The top right image showing the RGB video sequence. The 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.
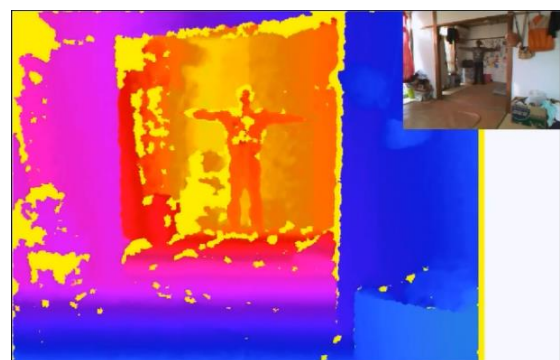
Fig.3.       T-Pose Position before the recording begin

Figure 4 shows 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 5 below show the BVH file result and preview in BVH file viewer.
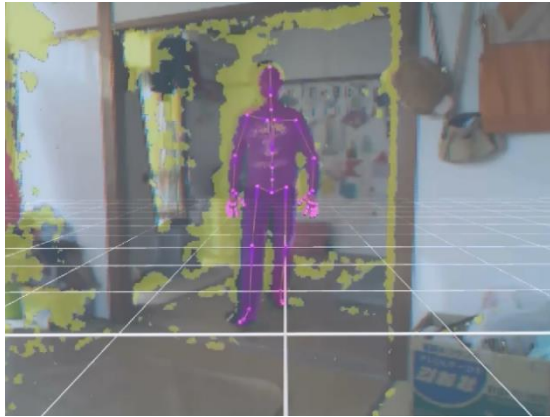


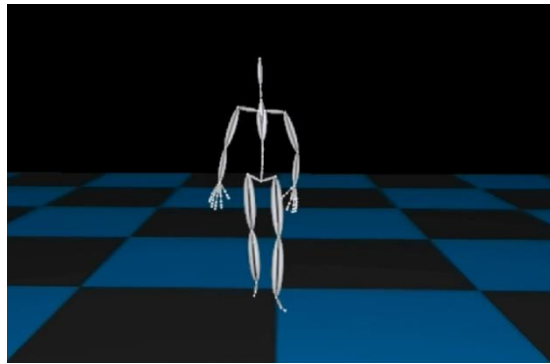Fig.4.        Skeleton motion tracking sequence



Fig.5.        BVH skeleton results

To get the 3D coordinates from skeleton model, we need to get the tree structure and the channel data first. Those data provided by BVH File. Figure 6 show the result of the skeleton model in 3D coordinates. Once we got the coordinates, we can calculate the angle of knee using 3 coordinate of Hip, Left Thigh, and Left Shin from the skeleton.

$$A = P_2 - P_1 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad [1]$$

$$B = P_3 - P_2 = \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2} \qquad [2]$$

$$C = P_3 - P_1 = \sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2} \qquad [3]$$

Thus the angle of θ is represented as follows:

$$\theta = \cos^{-1}\left(\frac{B^2 - A^2 - C^2}{2AC}\right) \qquad [4]$$

### C. Feature Extraction

Using the knee angle feature that we created, we can extract the features required for human gait classifications.
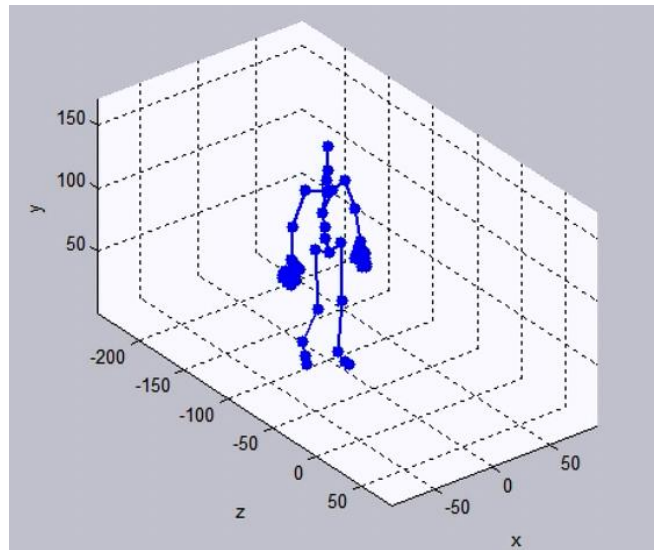


Fig.6.        Skeleton model imported in MATLAB

The angles of knee angle can be calculated using simple trigonometry formula as shown in Figure 7 and equation [1] – [4].
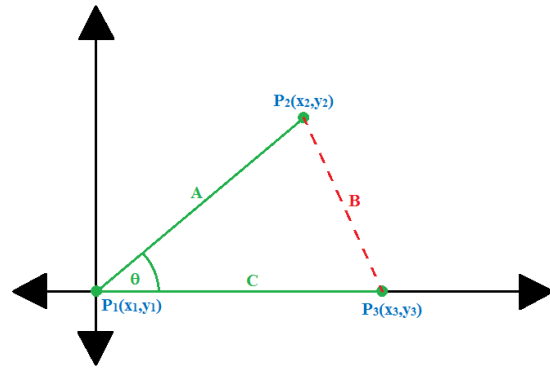


Fig.7.        Illustration to calculate angle between two lines using points coordinates

Figure 8 shows examples of the extracted features. In the Figure 8 there are seven peoples' knee angles of features. Their human gaits are evaluated with five grades from A to E. Gait cycles are different each other. Therefore, normalization is required with the longest human gait. Figure 9 shows the normalized human gait derived knee angles of features.

### D. Classification

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.
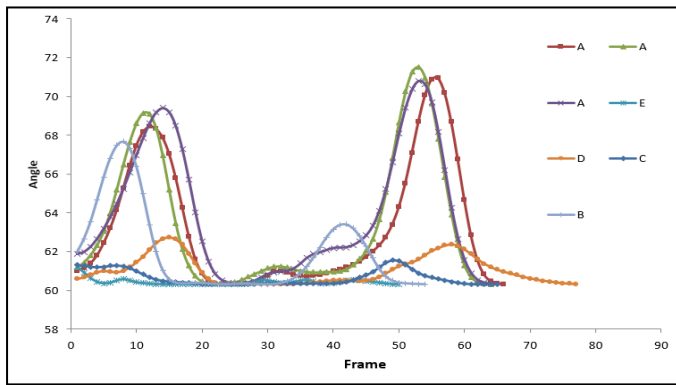
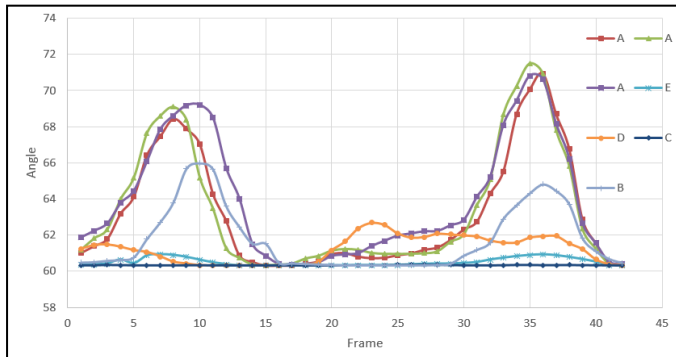Fig.8.        Sample from knee angle feature



Fig.9.        Figure 7 after normalization

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

Using C4.5 pruned tree, we got 77.27% correctly classified instances. Table 1 shows Confusion Matrix from C4.5 Classifier. Also Table 1 shows the detailed accuracy by class using C4.5 Classifier. The decision tree results from C4.5 is as follows,

```
8 <= 62.661977
|  43 <= 61
|  |  39 <= 60.577181
|  |  |  43 <= 51
|  |  |  |  8 <= 60.716972: B (5.0)
|  |  |  |  8 > 60.716972: D (2.0)
|  |  |  43 > 51: D (15.0/1.0)
|  |  39 > 60.577181: B (13.0)
|  43 > 61
|  |  35 <= 60.897472: D (2.0)
|  |  35 > 60.897472: C (17.0)
8 > 62.661977
|  36 <= 68.967621
|  |  32 <= 61.451288: A (3.0)
|  |  32 > 61.451288: E (19.0/1.0)
|  36 > 68.967621: A (12.0)
```

TABLE I.   Confusion Matrix from C4.5 Classifier

|   | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 11 | 0 | 0 | 0 | 5 |
| B | 0 | 14 | 1 | 3 | 0 |
| C | 0 | 0 | 16 | 1 | 1 |
| D | 0 | 4 | 3 | 11 | 0 |
| E | 2 | 0 | 0 | 0 | 16 |

TABLE II.        Detailed accuracy by class using C4.5 Classifier

| Class | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area |
|---|---|---|---|---|---|---|
| A | 0.688 | 0.028 | 0.846 | 0.688 | 0.759 | 0.858 |
| B | 0.778 | 0.057 | 0.778 | 0.778 | 0.778 | 0.937 |
| C | 0.889 | 0.057 | 0.8 | 0.889 | 0.842 | 0.924 |
| D | 0.611 | 0.057 | 0.733 | 0.611 | 0.667 | 0.833 |
| E | 0.889 | 0.086 | 0.727 | 0.889 | 0.8 | 0.879 |
| Weighted Avg | 0.773 | 0.058 | 0.775 | 0.773 | 0.769 | 0.887 |

Using SVM, we got 86.36% correctly classified instances.

TABLE III.        Confusion Matrix from SVM Classifier

|   | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 14 | 0 | 0 | 0 | 2 |
| B | 0 | 15 | 0 | 1 | 2 |
| C | 0 | 0 | 17 | 1 | 0 |
| D | 0 | 3 | 2 | 13 | 0 |
| E | 1 | 0 | 0 | 0 | 17 |

TABLE IV.        Detailed accuracy by class using SVM classifier

| Class | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area |
|---|---|---|---|---|---|---|
| A | 0.875 | 0.014 | 0.933 | 0.875 | 0.903 | 0.931 |
| B | 0.833 | 0.043 | 0.833 | 0.833 | 0.833 | 0.895 |
| C | 0.944 | 0.029 | 0.895 | 0.944 | 0.919 | 0.958 |
| D | 0.722 | 0.029 | 0.867 | 0.722 | 0.788 | 0.847 |
| E | 0.944 | 0.057 | 0.81 | 0.944 | 0.872 | 0.944 |
| Weighted Avg | 0.864 | 0.035 | 0.866 | 0.864 | 0.862 | 0.914 |

### IV.    CONCLUSION

The proposed method uses Kinect depth sensor camera and Ipisoft motion capture software to generate 3D skeleton model. Ipisoft itself is special purpose application to create skeleton so user can use the motion to their computer generated character motion. The skeleton generated will then extract the knee angle feature and use the feature to measure the gait disable quality. The purpose of this research is to analyze whether Kinect and Ipisoft can be used for extracting feature from 3D skeleton in gait related research. Experiments are done using knee angle feature and 18 video dataset in each class. From the experimental results shows 86.36% correctly classified instances using SVM.

R E F E R E N C E S

[1]  X. Qinhan, "Technology review - Biometrics-Technology, Application, Challenge, and Computational Intelligence Solutions," *IEEE Computational Intelligence Magazine*, 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," *Biomedical Soft Computing and Human Sciences*, 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 International Conference on*, 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 Processing Magazine*, vol. 22, no. 6, pp. 78–90, 2005.

[5]  M. S. Nixon and J. N. Carter, "Automatic Recognition by Gait," *Proceedings of the 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," *Computer Vision and Image Understanding*, vol. 90, no. 1, pp. 1–41, 2003.

[7]  X. Li, S. Maybank, and S. Yan, "Gait components and their application to gender recognition," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 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 International Journal of Computer Science and Network Security*, pp. 62–68, 2011.

[9]  K. Arai and R. Asmara, "Human Gait Gender Classification in Spatial and Temporal Reasoning," *IJARAI International Journal of Advanced Research in Artificial Intelligence*, vol. 1, no. 6, pp. 1–6, 2012.

[10]  G. Huang and Y. Wang, "Gender classification based on fusion of multi-view gait sequences," *Computer 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 International Conference on Digital Image Computing: Techniques and Applications*, pp. 320–327, Dec. 2010.

A U T H I O R S  P R O F I L E

**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 290 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 PhD Student at Information Science in Saga University, Japan. His research interests include signal processing, image processing, parallel processing, pattern recognition, and computer vision.