Poster: Towards Multi-Person Motion Forecasting: IMU Based Motion Capture Approach

Yasuo Katsuhara
Frontier Research Center, Toyota Motor Corp.
Susono, Shizuoka, Japan
yasuo_katsuhara@mail.toyota.co.jp

Hirotaka Kaji
Frontier Research Center, Toyota Motor Corp.
Susono, Shizuoka, Japan
hirotaka_kaji@mail.toyota.co.jp

ABSTRACT

Forecasting body motion has a lot of potential applications such as sports and entertainment. Previous studies have mainly employed cameras and optical motion captures to measure the joint positions of person, and predicted them about 0.5 seconds before by using deep neural networks. However, following two difficulties have to be solved to install the forecasting system into the real world: One is that camera and optical based methods have to take into account the environmental settings and occlusion problems, and the other is that previous studies have not considered plural persons. In this paper, we propose a multi-person motion forecasting system by using inertial measurement unit (IMU) motion captures to overcome these difficulties simultaneously, and demonstrate a preliminary result.

CCS CONCEPTS

• **Human-centered computing** → *HCI design and evaluation methods.*

KEYWORDS

Body motion forecasting; IMU based motion capture; multi-person; neural networks

ACM Reference Format:
Yasuo Katsuhara and Hirotaka Kaji. 2019. Poster: Towards Multi-Person Motion Forecasting: IMU Based Motion Capture Approach. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2019 International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct), September 9–13, 2019, London, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3341162.3343776

UbiComp/ISWC '19 Adjunct, September 9–13, 2019, London, United Kingdom

ACM ISBN 978-1-4503-6869-8/19/09.
https://doi.org/10.1145/3341162.3343776

1 INTRODUCTION

Forecasting body motion has a lot of potential applications such as sports and entertainment. In addition, anomaly motion prediction for factory workers and drivers is also a useful application area.

Previous studies have mainly employed cameras and optical motion captures to measure the joint positions of person, and predicted them about 0.5 seconds before by using deep neural networks. For instance, Horiuchi et al. have proposed a system to forecast jumping motion based on depth camera [2]. Optical motion capture and video datasets such as Human 3.6M [3] are widely used to train body motion prediction by using deep neural networks [1, 5]. As an application, Wu et al. have proposed a mixed reality martial arts training system with a RGB camera [7].

In this study, we focus on following two difficulties to install forecasting systems into the real world: One is that camera and optical based methods have to take into account the environmental settings and occlusion problems, and the other is that previous studies have not considered plural persons. In particular, forecasting multi-person motion with their interactions will be a challenging topic because a person's motion might affects others one and vice versa.

In order to overcome these difficulties simultaneously, we employ IMU based motion captures for body motion measurement. This approach allows us not to consider the camera setting and occlusion problems, and to easily measure plural persons, although there is a disadvantage that subjects have to wear them.

In this paper, at first, we develop a motion predictor based on IMU motion capture for two simple activities. Then, we construct a prototype of multi-person motion forecasting system and demonstrate a preliminary result.

2 MOTION PREDICTION

Data Collection

Eleven healthy males in their 20s to 40s participated in an experiment. We measured their three dimensional position (x, y, z) of 21 body joints by using the IMU motion capture system (Perception Neuron2.0 [6]) shown in Figure 1. The sampling interval was 50 milliseconds (20 frame per second).

Figure 1: The condition of experiment. The joint positions of two activities are measured by using IMU motion capture.

In this study, they also wore a wearable respiration sensor [4] at their abdomen, although it was not used for the following analysis.

In this experiment, the following two typical activities in daily living were measured.

Standing up: The subjects were asked to repeat standing up and sitting down 10 times for 3 minutes. The measurement was performed 2 trials for each subject.

Walking: The subjects were asked to walk around the rectangular table for 1 minute and executed it 4 trials (2 times clockwise, 2 times counterclockwise). The walking speed depended on each subject.

The design of the experiment was approved and conducted according to the ethical guidelines of Toyota Motor Corporation. We sufficiently explained the details and obtained informed consents from all the subjects.

Data Processing

To construct a motion predictor, we basically employed Horiuchi's manner [2]. The 3D positions (x, y, z) of 21 joints and center of gravity (CoG) of whole body are obtained every frame. At the present frame t, the feature vector $\boldsymbol{x}_t$ is composed by the 3D positions from $t-10$ to t frames. Hence, the dimension of feature vector is 22 (joints and CoG position) $\times 3$ (dimension) $\times 10$ (frames) $= 660$. In order to forecast the positions after 0.5 seconds, the predicted position vector $\boldsymbol{z}_t$ is defined by the 3D positions of $t+10$ frames (the dimention is $22 \times 3 = 66$).

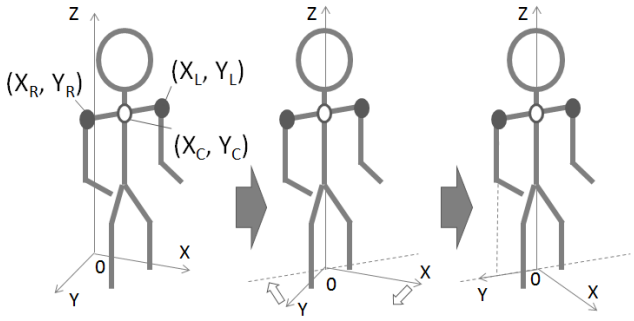

Figure 2: The conceptual diagram of axis rotation.

As the predictor, we simply employed a 3-layer neural networks (fully connected, input layer: 660, hidden layer: 20, output layer: 66). Sigmoid function was employed as the activation function for hidden units. The weight parameter $\boldsymbol{w}$ of the neural network was trained by the gradient descent method to minimize the mean squared error:

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{k=1}^{K}(\boldsymbol{y}_k - \boldsymbol{z}_k)^2, \tag{1}$$

where K is the number of training samples, $\boldsymbol{y}$ is the output of our model.

To predict human motion efficiently, we addressed regardless of the subject's direction. Figure 2 shows the conceptual diagram of axis rotation. Regarding the $x-y$ plane, the following procedure was executed at every frame. At first, the center coordinates $x_C = (x_R + x_L)/2$, $y_C = (y_R + y_L)/2$ were calculated, where (x_R, y_R) and (x_L, y_L) were the right and left shoulder positions in $x-y$ plane, respectively. Next, x_C and y_C was subtracted from all the positions in the dataset so that $(x_C, y_C) = (0, 0)$ Then, all the samples were rotated so that $x_r = x_L = 0$.

In order to evaluate the generalized performance of the predictor, the samples of nine subjects were used as the training data (standing up: 59,047 samples, walking: 36,033 samples), and that of the remaining two subjects were used as the test data (standing up: 13,310 samples, walking: 5,314 samples). We evaluated the prediction performance using the average prediction error of the predicted positions of whole body (21 joints + CoG) compared to the actual positions (after 0.5 seconds) of that by root mean square error (RMSE). RMSE is one of the commonly used indicators to evaluate forecasting performance [2, 7].

Result

Figures 3 and 4 show the prediction results of standing up and walking activities. We can confirm that the RMSE of standing up indicates the poor performance when the subject

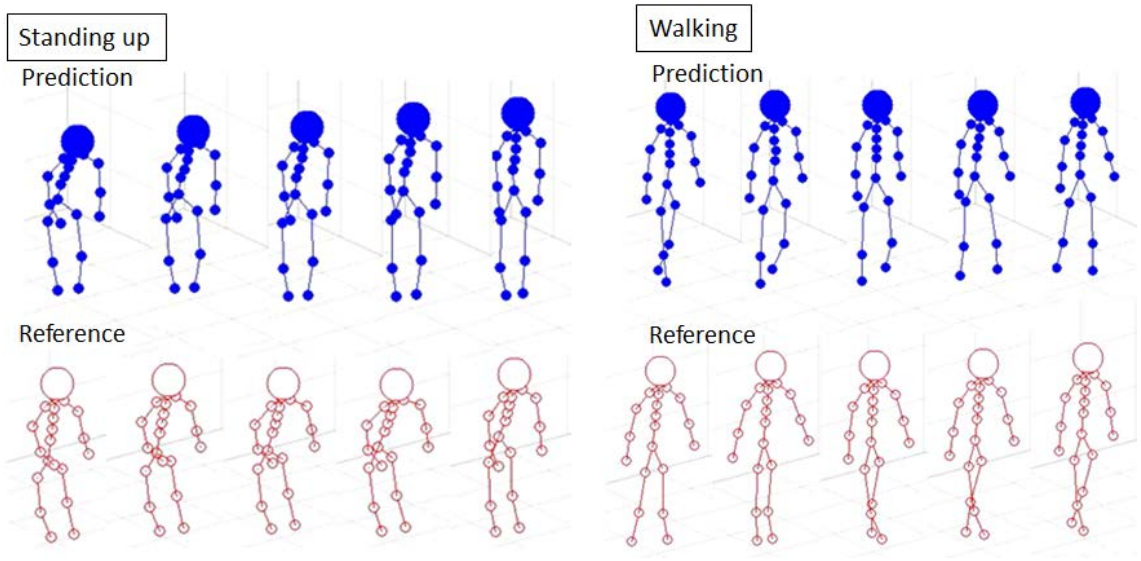

Figure 3: The prediction results of standing up and walking activities. Upper: prediction result, Lower: reference (current position).

was standing up and sitting down. In order to evaluate the performances of transient state more precisely, the RMSE in which the subject keep to standing up or sitting down for more than 2 seconds was excluded from the mean RMSE.

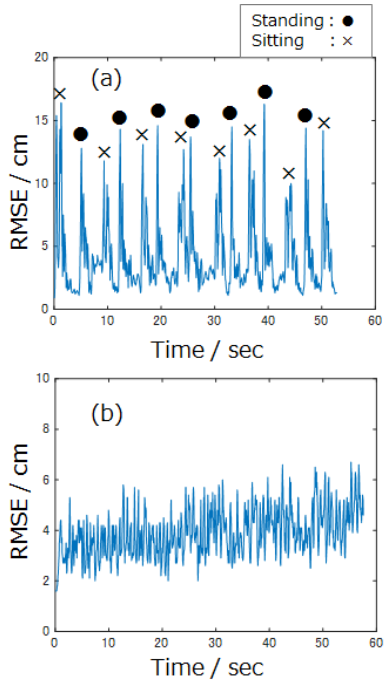

Figure 4: The transitions of mean RMSEs of (a) standing up and (b) walking activities. • and × indicate the timing of standing up and sitting down motion, respectively.

As a result, we confirmed that our model could forecast standing up and walking 0.5 seconds before with 5.96 cm and 4.43 cm mean RMSEs, respectively.

Through checking the prediction error for each joint, we found that the prediction errors for the hands and feet were the most influential. These joints which has high flexibility are considered to be due to the large influence of individual differences. Through this simple verification, we could confirm the generalized performance of our model.

3 SYSTEM IMPLEMENTATION

In order to forecast plural persons' motion simultaneously, we constructed a prototype of forecasting system with a couple of IMU motion captures. The block diagram of our system is shown in Figure 5.

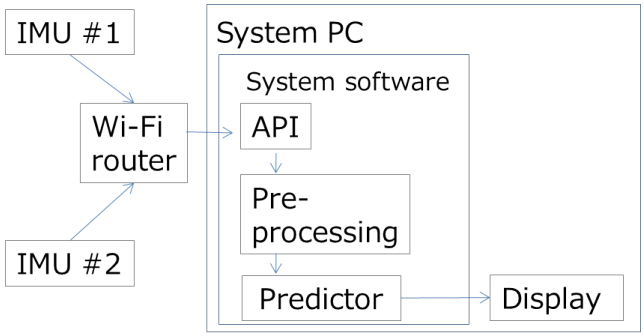

Figure 5: The block diagram of multi-person motion forecasting system.

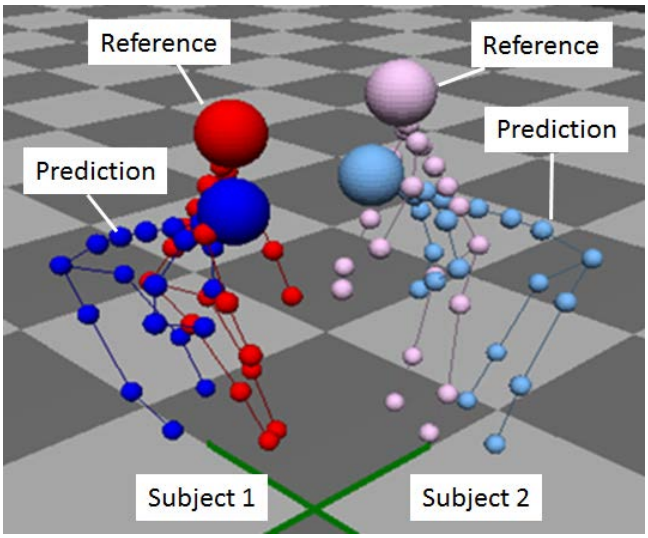

Figure 6: A snapshot of multi-person motion forecasting system. Blue and light blue human models: prediction result, red and pink human models: reference (current position).

We used a Microsoft Surface Pro 3 (RAM: 8GB, CPU: Intel Core i7) as the system PC. Different IDs were assigned to two IMU motion captures. Each positional data stream was transmitted to the system PC through a Wi-Fi network. The predictor trained in the previous section was installed into the system PC. The predictor and system software were implemented on c++, and the data transmission and prediction were performed every 50 milliseconds.

Figure 6 shows a snapshot of the multi-person motion forecasting system. The blue and light blue models indicate the prediction result of motion of after 0.5 seconds, and the red and pink human models represent the reference (current position). Here, the system simultaneously forecasts two subjects' motion of standing up in real time. We confirmed that the system PC can calculate them sufficiently fast within 50 milliseconds and it has a capability of handling several IMU motion captures.

4 CONCLUSION

In this paper, we proposed the multi-person motion forecasting system based on IMU motion captures. At first, we developed the IMU based motion prediction and indicated that the prediction error of our model was about 5 cm in standing up and walking motions with 11 subjects. Then, we demonstrated that our system could forecast two persons' motion in real time.

In the future, we plan to conduct an additional experiment to collect various activities including multi-person interaction with our system. In addition, exploring neural network structures to handle multi-person interaction will be a challenging task.

REFERENCES

[1] Emad Barsoum, John Kender, and Zicheng Liu. 2018. HP-GAN: Probabilistic 3D Human Motion Prediction via GAN. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[2] Yuuki Horiuchi, Yasutoshi Makino, and Hiroyuki Shinoda. 2017. Computational foresight: Forecasting human body motion in real-time for reducing delays in interactive system. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. ACM, 312–317.

[3] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. 2014. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 7 (jul 2014), 1325–1339.

[4] Hirotaka Kaji, Hayato Yamaguchi, Kazuhide Shigeto, and Hirokazu Kikuchi. 2018. Wearable Respiration Sensor Platform Using Ultrasound Transducer. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp '18)*. ACM, New York, NY, USA, 86–89. https://doi.org/10.1145/3267305.3267586

[5] Chen Li, Zhen Zhang, Wee Sun Lee, and Gim Hee Lee. 2018. Convolutional Sequence to Sequence Model for Human Dynamics. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[6] Noitom. 2019. perception neuron. Retrieved June 10, 2019 from https://neuronmocap.com/products/perception_neuron.

[7] Erwin Wu and Hideki Koike. 2019. FuturePose - Mixed Reality Martial Arts Training Using Real-Time 3D Human Pose Forecasting With a RGB Camera. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)* (2019), 1384–1392.