Ballroom Dance Step Type Recognition by Random Forest Using Video and Wearable Sensor

Hitoshi Matsuyama
Graduate School of Engineering, Nagoya University
Japan
hitoshi@ucl.nuee.nagoya-u.ac.jp

Kei Hiroi
Graduate School of Engineering, Nagoya University
Japan
k.hiroi@ucl.nuee.nagoya-u.ac.jp

Katsuhiko Kaji
Faculty of Information Science, Aichi Institute of Technology
Japan
kaji@aitech.ac.jp

Takuro Yonezawa
Graduate School of Engineering, Nagoya University
Japan
takuro@nagoya-u.jp

Nobuo Kawaguchi
Graduate School of Engineering, Nagoya University
Japan
kawaguti@nagoya-u.jp

ABSTRACT

The paper presents a hybrid ballroom dance step type recognition method using video and wearable sensors. Learning ballroom dance is very difficult for less experienced dancers as it has many complex types of steps. Therefore, our purpose is to recognize the various step types to support step learning. While the major approach to recognize dance performance is to utilize video, we cannot simply adopt it for ballroom dance because the dancers' images overlap each other. To solve the problem, we propose a hybrid step recognition method combining video and wearable sensors for enhancing its accuracy and robustness. We collect seven dancers' video and wearable sensors data including acceleration, angular velocity, and body parts location change. After that, we pre-process them and extract some feature values to recognize the step types. By adopting Random Forest for recognition, we confirmed that our approach achieved f1-score 0.760 for 13 step types recognition. Finally, we will open our dataset of ballroom dance to HASCA community for further research opportunities.

CCS CONCEPTS

• **Human-centered computing** → *Ubiquitous computing*.

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

ACM ISBN 978-1-4503-6869-8/19/09…$15.00
https://doi.org/10.1145/3341162.3344852

KEYWORDS

signal processing, machine learning, datasets

ACM Reference Format:
Hitoshi Matsuyama, Kei Hiroi, Katsuhiko Kaji, Takuro Yonezawa, and Nobuo Kawaguchi. 2019. Ballroom Dance Step Type Recognition by Random Forest Using Video and Wearable Sensor. 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, 7 pages. https://doi.org/10.1145/3341162.3344852

1 INTRODUCTION

Ballroom dance is one of the popular sports regardless of ages or sex. In particular, the number of elderly people who play ballroom dance is greatly increased with the progress of the aging population in developed countries. This is because ballroom dance is also good for health – it contributes to preventing physical and cognitive decline [1]. Ballroom dance is popular as a communication means, and people sometimes enjoy the competition for the perfection of their performance. To enjoy ballroom dance, the improvement of the dancing skill is one of the important purposes for the players. However, learning ballroom dance is sometimes difficult for less experienced dancers as it has many complex types of steps.

The main forms of dance learning are (1) lessons taught by instructors, and (2) individual exercise by looking at the players themselves in mirrors or videos. In each practice form, several methods have been proposed to improve dance skills by assisting the dance exercise, such as transmitting the movement of the instructor to the recipient [2] [3] [4], or the system itself that plays the role of instructor [5] [6] [7]. However, those systems are not handling the information of step type.

In dance, especially ballroom dance, understanding step types is an important method to clarify how to improve the performance because each step type has its correct way of dancing. Ballroom dance has many complex step types in which the direction, timing of the foot and the orientation of the body are defined, and dancers are recommended to compliant them while they are not so experienced. For less experienced dancers, however, it is very difficult to remember and understand the step types. Therefore, we aim to recognize the complex step types automatically to support step learning and understanding. Automatic step recognition must lead several useful applications such as dance support system including automatic step teaching and improving.

Ballroom dance has the following characteristics: it has a wide range of movements, and there are over 100 types of steps, which can be danced in vacant places as well as crowded environments, such as dance studios and dance halls where many people practice. We have proposed a hybrid ballroom dance performance recognition method[8]. In the method, we utilized video to capture whole body movement and smart-phone sensor to acquire body movement information independent of shooting environments. However, the study has limitation that there is only four types of steps from 1 dancer. Thus, in this study, we show the usefulness of the proposed method by collecting more kinds of step data of multiple dancers. We use a high sampling rate video camera and wearable sensors, and acquire 7 dancers' sequential dance stepping data. The detail of data collection will be stated in section 3, DATA COLLECTION.

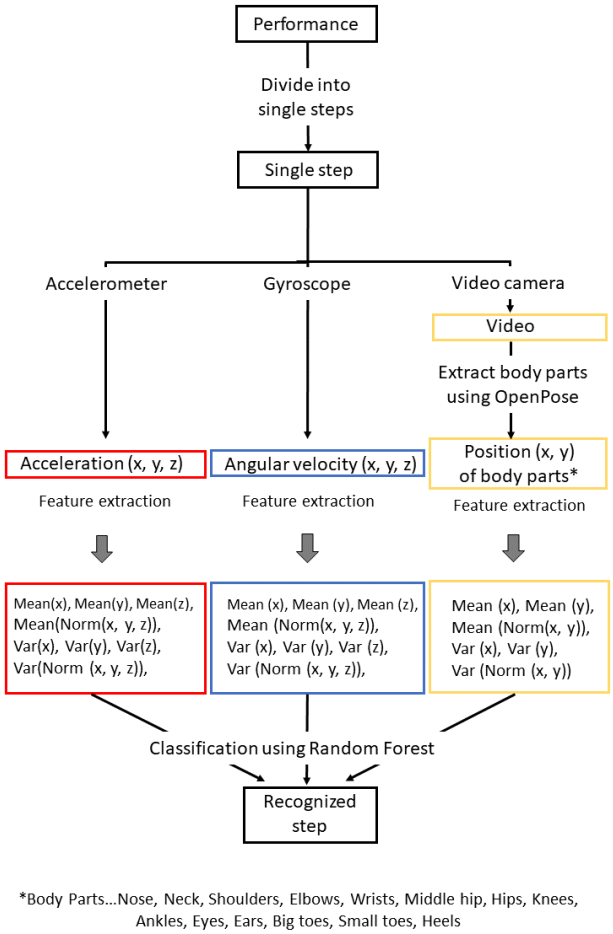


Figure 1: Method overview

2 RELATED WORKS

In this section, we will show some related works of dance performance recognition and supporting method. A major approach is to transmit the posture or movement information of an instructor to a participant. For example, Fujimoto et al. proposed a visual-based system to support dance exercise[2] using Kinect. In the system, participants can know how to move their bodies by looking at the skeleton location of the instructor, which is overlapped onto the participants' images. In addition to using Kinect, Yamauchi et al. utilized a wireless-mouse, and developed a more accomodating dance supporting system[3]. There are also some footwear-based supporting approaches. Narazani et al developed a dance-skill transfer system by foot-base interaction[4]. Other footwear devices developments are known too[9] [10].

On the other hand, some researchers aimed to construct the system itself that plays the role of instructor. For example, Anderson et al. developed an augmented mirror to support ballet exercise[5]. Milka et al. took their focuses on an augmented mirror too, and developed a visual and verbal feedback system for augmented mirror[6]. Not only designing an augmented mirror, there is also a work that constructed a virtual instructor. Huang et al analyzed the ballroom dance lesson system, and divided the lesson time into some parts. Finally, they developed a virtual ballroom dance instructor system[7].

Based on these researches and considering the characteristics of ballroom dance, we have proposed a hybrid recognition method using video and wearable sensor. In that paper, we have also proposed some feedback ideas to support dance exercise[8]. And in this study, we show the usefulness of the proposed method by collecting more kinds of step data of multiple dancers.

3 DATA COLLECTION

Fig. 1 shows the process of our proposed method for recognizing the segmented single-step data. Fig. 2 shows some

Table 1: Attribute of dancers

	Sex	Height (cm)	Experience (year)
Dancer 1	Male	173	17
Dancer 2	Male	176	5
Dancer 3	Male	182	1
Dancer 4	Male	160	1
Dancer 5	Male	171	4
Dancer 6	Male	175	3
Dancer 7	Male	177	5

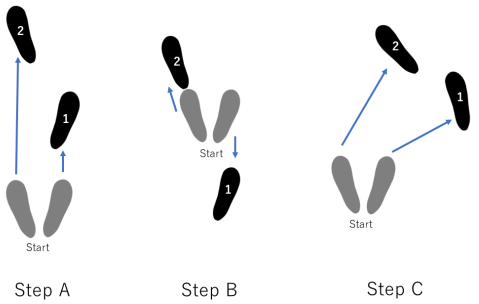


Figure 2: Step diagrams

examples of step diagrams. In order to verify the classification of the steps, seven dancers performed four types of steps. For simplicity, we name the steps as "Step A, Step B, ..., Step M". The characteristics of each step are stated in Table 3 and the states of step types are shown in Fig. 3.

We prepare a sequence of dance performance by connecting the steps, and participants perform the sequential steps. The order of step types is "Step A, B, C, D, E, F, E, G, A, H, I, B, C, J, K, L, K, M". There are seven dancers who have participated in the data collection work, and the attributes of them are shown in the Table 1. The experiences of all of them are not less than 1 year, which means that all of them can perform the dance steps almost correctly. The height and experience of dancers vary from 160cm to 182cm, and 1 year to seventeen years. You can see that we collected only male performances. This is because the male's and female's steps are usually totally different in the same step name.

While performing, video and wearable sensor data are acquired. The data collection environments are shown in the Fig. 6 and Fig. 7. Six wearable sensors (ATR-Promotions, TSND151, sampling rate = 125Hz) are worn on arms, hips, and ankles. Each axis of the sensors and wearing condition are shown in Fig. 4. While this figure is showing the example of the sensor on feet, the other sensors are worn in the same way. About all wearable sensors, axis z towards the center of a body. Simultaneously, video (SONY FDR-AX60, sampling rate = 120fps) of the dance performance is shot. We prepared two shooting positions. For each dancer, the first half of the performances are shot from position 1, and the other half of them are shot from position 2. In total, 20 performance data are obtained for one dancer. The detail of acquired data is shown in the Table 2. While similar datasets including dance data are known[11], this dataset is peculiar as it contains sensor information of six body parts with 125 Hz (not 120 because the sensor control software only allowed specification in integer milliseconds like 7ms, 8ms or 9ms) sampling rate, and 120 fps video. Also, it is rare to find the dataset collecting only ballroom dance step data. The main reason why we select such high-function devices is we consider it is important to develop a high-quality system in the beginning regardless of the device expensiveness.

All collected data will be public. The data format of video and wearable sensor data are as follows:

- Video data:
  JSON file put out by OpenPose[1], that contains 25 body parts location data of each frame, and also some raw video data of dancers.
- Wearable sensor data:
  CSV file that contains the time variation data of accelerometer and gyroscope. There are six CSV files for each one dance performance: Left ankle, right ankle, left hip, right hip, left arm, and right arm.

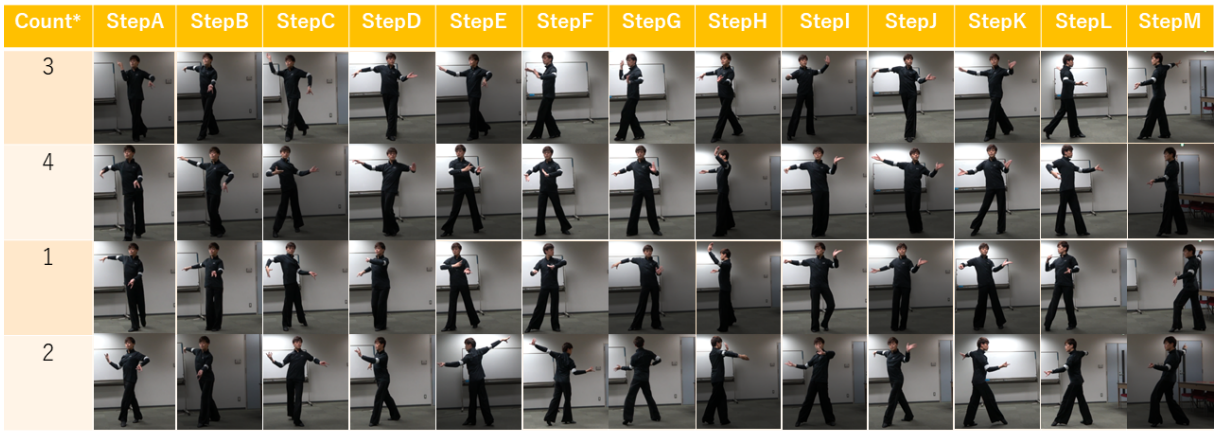
4 RECOGNITION METHOD

In our previous work, the basic recognition method is stated[8]. However, in this work, there are some differences in the ways of obtaining data, such as the difference in sampling rates. Therefore, we perform some preprocessing on the data, and then extract feature values.

Preprocessing

- Resampling:
  As the sampling rates of video and wearable sensor are different, we apply scypy.signal.resample to wearable sensor data, and resampled them from 125 to 120Hz.
- Elimination of the movement distance in the frame:
  The dancer changes position within the video frame during the performance. However, the movement distance and direction differ depending on the person, and also depending on the start position. Therefore, we eliminate such effects by correcting the coordinates of other body parts to relative positions based on the neck coordinate (See Fig. 5).
- Extraction of each step position in the sequential dance performance:

[1]https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/output.md



* Count: Beat of music. Each foot movement in a step has its proper timing to the count.

Figure 3: Step types

Table 2: Data detail

	Number of data	Location	Sampling rate	Video/Sensors
Video	20 times	Arms, hips, and ankles	120 fps	HD(1920 x 1080)
Wearable sensor	20 times	Two positions	125 Hz	Accelerometer, gyroscope

Table 3: Characteristics of each step

	Number of foot actions	Progressing direction	Change amount of body orientation	Amount of free arm
Step A	3 times	F	None	Free
Step B	3 times	F and B	None	Free
Step C	3 times	S	90 degree ACW	Free
Step D	3 times	F and B	90 degree CW	Free
Step E	3 times	F	90 degree CW	Free
Step F	3 times	F	90 degree ACW	Free
Step G	3 times	F	360 degree ACW	Free
Step H	3 times	S	315 degree CW	Holding
Step I	3 times	S	None	Holding
Step J	3 times	F and B	90 degree CW	Free
Step K	3 times	B	90 degree ACW	Free
Step L	3 times	B	90 degree CW	Free
Step M	2 times	B	None	Free

F: Forward, B: Backward, S: Side, CW: Clockwise, ACW: Anticlockwise

Because each performance data are sequential and consists of different step types, we extract each step position by looking at each video and recording them. After recording, we divide video and sensor data into each step. In the recognition part, we aim to recognize these segmented data performing only single steps.

- Interpolating missing value:

Sometimes there are missing values in the data. To handle them, we utilize the interpolate function of the pandas in python.

Feature extraction

In order to classify the steps, we use Random Forest Classifier[12] with following features. The summary of extracted features is also shown in Table 4.

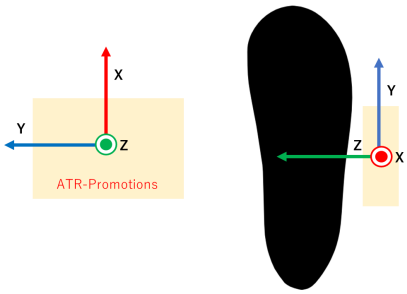


Figure 4: Wearable sensor

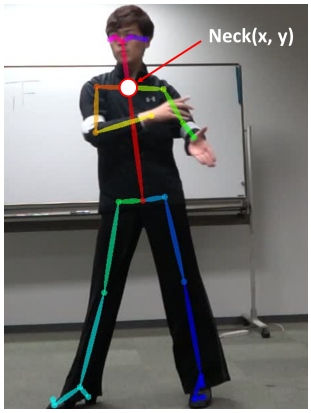


Figure 5: Position of neck and other parts

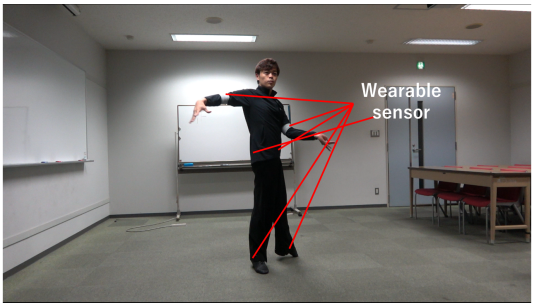


Figure 6: Data Collection (Direction 1)

Feature Extraction from Wearable Sensors: For the time change data from the wearable sensors in the smartphone, we perform following feature extractions.

- mean and variance of the time change of acceleration for each axis (x, y, z)
- mean and variance of the time change of norm of acceleration(x, y, z)
- mean and variance of the time change of angular velocity for each axis (x, y, z)
- mean and variance of the time change of norm of angular velocity(x, y, z)

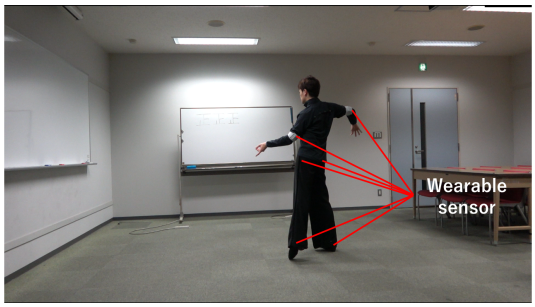


Figure 7: Data Collection (Direction 2)

Feature Extraction from Videos: To recognize the characteristics of movements of each steps, we obtain the positions of body parts using OpenPose[13][14]. Processed by OpenPose, two-dimentional positions of 25 body parts are obtained. From these position data, we calculate the feature values below:

- mean and variance of the time change of body part positions(x, y)
- mean and variance of the time change of norm of body part positions(x, y)

5 EVALUATION

In order to evaluate the result of classification, we split data into 80 percent for training and 20 percent for testing. We first train the classifier (Random Forest, n_estimators = 2500, criterion ="gini") with the training data and then make a prediction for test data. We also perform 5-fold cross-validation and 1-subject-out cross validation and show the result.

Result

Performing classification using features from both videos and wearable sensors, as shown in Fig. 8, we obtained f1-score 0.760. We summarized the result of prediction as confusion matrix(Fig. 8). In addition, we performed 5-fold cross-validation and calculated mean of the five results, and got f1-score 0.760. Furthermore, we performed 1-subject-out cross validation(Fig. 6). To discuss the feature extraction method, the feature importances of the classification are summarized in Table 5. Compared to our previous work, the number of step types is increased from four to thirteen. Thus, it is not unnatural that the f1-score decreased from 0.96 to 0.760. However, there are some specific mistakes.

From the matrix, we see that many mistakes occur among continuous steps. For example, Step A, B, and C were performed continuously, and some of them were mistakenly recognized with each other. There are also some step types mistaken each other despite they are not connected each other like step D and J. Then about the feature extraction, the

Table 4: Summary of extracted features

Sensor type		Features	
Wearable sensor	Acceleration	x, y, z	mean and variance of time variation
		norm	mean and variance of time variation
	Angular velocity	x, y, z	mean and variance of time variation
		norm	mean and variance of time variation
Video	Body parts	x, y	mean and variance of time variation
		norm	mean and variance of time variation

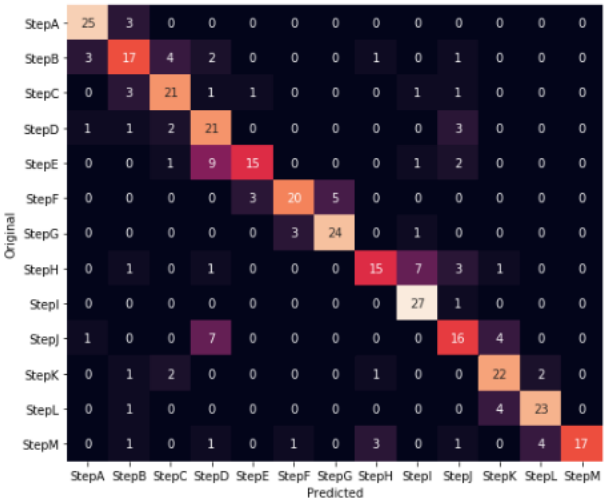


Figure 8: Confusion matrix for the step classification

Table 5: Five important features

1	Mean(Gyroscope x of left knee)
2	Mean(Gyroscope x of left ankle)
3	Mean(Gyroscope x of left hip)
4	Mean(Gyroscope x of right hip)
5	Mean(Gyroscope x of right ankle)

table shows that the features from gyroscope x are more important compared to others. As x-axis captures vertical angular velocity change, it was found that knowing how much the body tilts with respect to the vertical axis is useful for the ballroom dance step recognition. On the other hand, feature values from video were not so important, which means that we must develop the preprocessing and feature extraction method to get more accurate recognition.

Discussion

One of the big problems is the mistakes in continuous step types. This problem happens because the frame around the joint part of the steps may be interpreted as both steps. Thus, we should take much care to handle the step joints, not only in the feature extraction, but also in labeling and segmenting data into steps. One way out we have is to let each step overlap each other to some extent so that we can consider the frames around the step joint parts as both step types. However, in order to recognize steps in time series in the future, we need some contrivance to use the method.

The second thing is that there are some step types that are so similar to each other. For example, step D and step J have almost the same characteristics exclude the body orientation change (See Table 3) though they have different step type names. Thus, sometimes it was hard for the classifier to recognize them correctly. For the higher accuracy, we' ll need to consider some features that grasp the characteristics of body orientation and other specific points of each step type. On the other hand, it was surprising that some symmetrical step types, such as Step K and Step L, did not perplex the classifier so much.

In addition, it is found that there are some subject dependencies problems by running leave-one-subject-out method. This is because there are some differences in performing ways among dancers. Some of the dancers have basic ways, some have special ways. We consider such tendencies caused the subject dependencies.

6 CONCLUSION

The paper presented a hybrid ballroom dance step type recognition method using video and wearable sensors. Our purpose is to recognize the various step types to support step learning, and proposed a hybrid step recognition method combining video and wearable sensors for enhancing the accuracy and robustness of step type recognition. We collected seven dancers' video and wearable sensors data including acceleration, angular velocity, and body parts location change. After that, we pre-processed them and extracted some feature values to recognize the step types. By adopting Random Forest for recognition, we confirmed that our approach achieved f1-score 0.760 for 13 step types recognition. Finally, we will open our dataset of ballroom dance to HASCA community for further research opportunities.

Table 6: Leave one subject out

Left subject	Dancer 1	Dancer 2	Dancer 3	Dancer 4	Dancer 5	Dancer 6	Dancer 7
F1 score	0.82	0.68	0.88	0.92	0.82	0.67	0.53

ACKNOWLEDGMENT

This work was partially supported by JSPS KAKENHI Grant Number JP17H01762. The ballroom dance performances were given by the members of Nagoya University Ballroom Dance Club and its alumni.

REFERENCES

[1] Dafna Merom, Robert Cumming, Erin Mathieu, Kaarin J. Anstey, Chris Rissel, Judy M. Simpson, Rachael L. Morton, Ester Cerin, Catherine Sherrington, and Stephen R. Lord. Can Social Dancing Prevent Falls in Older Adults? a Protocol of the Dance, Aging, Cognition, Economics (DAnCE) Fall Prevention Randomised Controlled Trial. *BMC Public Health*, 13(1):477, 2013.

[2] Minoru Fujimoto, Masahiko Tsukamoto, and Tsutomu Terada. A Dance Training System that Maps Self-Images onto an Instruction Video.

[3] Masashi Yamauchi, Ryo Shinomoto, Eriko Nishiwaki, Risa Onozawa, and Tetsuro Kitahara. Development of Dance Training Support System Using Kinect and Wireless Mouse. *The Sysmposium of Entertainment Computing*, 2013:332–338, 2013.

[4] Marla Narazani, Katie Seaborn, Atsushi Hiyama, and Masahiko Inami. StepSync: Wearable skill transfer system for real-time foot-based interaction, 2018.

[5] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. YouMove: Enhancing Movement Training with an Augmented Reality Mirror. *In Proc. of UIST 2013 Conference:ACM Symposium on User Interface Software and Technology*, pages 311–320, 2013.

[6] Milka Trajkova and Francesco Cafaro. Takes Tutu to Ballet: Designing Visual and Verbal Feedback for Augmented Mirrors. *In Proc. of ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(1):1–30, 2018.

[7] Hung-Hsuan Huang, Masaki Uejo, Yuki Seki, Joo-Ho Lee, and Kyoji Kawagoe. Construction of a Virtual Ballroom Dance Instructor. *The Japanese Society for Artificial Intelligence*, 28(2):187–196, 2013.

[8] Hitoshi Matsuyama, Kei Hiroi, Katsuhiko Kaji, Takuro Yonezawa, and Nobuo Kawaguchi. Hybrid Activity Recognition for Ballroom Dance Exercise using Video and Wearable Sensor. In *International Conference on Activity and Behavior Computing*, 2019.

[9] Paradiso Joseph, Hu Eric, and Hsiao Kai yuh. The CyberShoe: A Wireless Multisensor Interface for a Dancers Feet. 03 1999.

[10] J. A. Paradiso, K. Hsiao, A. Y. Benbasat, and Z. Teegarden. Design and Implementation of Expressive Footwear. *IBM Systems Journal*, 39(3.4):511–529, 2000.

[11] L. Sigal, A. Balan, and M. J. Black. HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion. *International Journal of Computer Vision*, 87(1):4–27, 2010.

[12] L. Breiman. Random Forests. *Machine Learning*, 2(1):5–32, 2001.

[13] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: Realtime Multi-person 2D Pose Estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*, 2018.

[14] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.