Hybrid Activity Recognition for Ballroom Dance Exercise using Video and Wearable Sensor

Hitoshi Matsuyama*, Kei Hiroi*, Katsuhiko Kaji†, Takuro Yonezawa* and Nobuo Kawaguchi*

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

†Faculty of Information Science
Aichi Institute of Technology, Toyota, Japan
kaji@aitech.ac.jp

***Abstract*—In this paper, we propose a hybrid activity recognition method for ballroom dance exercise using video and wearable sensor. The purpose of our research is to design a mechanism to support ballroom dance exercise, and this paper reports the first part to design a mechanism is to support ballroom dance exercise, and this paper reports the first outcome to achieve the purpose - recognizing ballroom dance exercise. There are two conceivable ways to recognize dance exercise: videos and wearable sensor. However, each of them has its disadvantages. Using video is a good way to recognize the movement of the body. However, it cannot provide us accurate timing or strength of foot actions because the number of their flames per seconds is too small to recognize the fast movements of dancers. On the other hand, while a wearable sensor is good at recognizing foot timing and strength, it is not good at recognizing the movement of the whole body. Therefore we propose a hybrid recognition method utilizing the merits of both video and wearable sensor. This paper focuses to recognize four different types of steps in Latin American, a kind of ballroom dance. For each step, we record wearable sensing data and videos. As a result, it is found that the accuracy of step recognition is improved by adding wearable sensing data to video data shot from two different angles.**

***Index Terms*—Wearable sensors, Video signal processing, Machine learning, Activity Recognition**

I. Introduction

Ballroom dance is one of the popular sports regardless of ages or sex. Especially, the number of elderly people who play ballroom dance is greatly increasing 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 competition for the perfection of their performance. To enjoy ballroom dance, the improvement of the skill is one of the important purposes for the players.

Learning ballroom dance is difficult among various types of dancing. To learn dances, people usually check their performance by dancing in front of mirrors or watching videos. For example, in ballet, mirror is a central tool for dancer's education [2], and it is known that mirrors have various effects on ballet dancer's performance [3]

There are other related works on dance recognition using wearables. Some works focused on dancers' shoes. Paradiso et al. [4] combined their interest in wearables with performance interfaces for digital music, and they built highly instrumented shoes for interactive dance. Because of the growth of wearable computing, a more versatile human-computer interface for the foot has developed [5]. Additionally, Aylward et al. [6] described the design of a system for interactive dancing too. Their system is wireless, compact and can capture the expressive motion of dancers.

Motion capturing of dancing is also studied well. Wang et al. [7] worked for an approach to the human figure tracking in video and even made their dataset publicly-available. Their data contains three actions: dancing, walking, and jumping. The dancing action includes five types of dancing scenes. In the situation that there are many similar datasets, Sigal et al. developed a systematic quantitative evaluation of competing methods [8]. Their work supports the development of new articulated motion and pose estimation algorithms and provide a baseline for the evaluation and comparison of new methods.

To support the dancing practice, several works proposed dance recognition methods wearable sensors, especially focusing on dancers' shoes. Paradiso et al. [4] developed instrumented shoes which provide interactive dance performance with music. Paradiso et al. [5] also presented more versatile wearable foot sensor for the higher accuracy of dance recognitions. In addition, Aylward et al [6] proposed an interactive dancing too. These systems are designed using wireless and compact sensors so that they can capture the expressive motion of dancers. Video-based motion capturing of dancing is also well-studied. Wang et al. [7] presented a method for human figure tracking from video, and also provided their dataset as publicly-available. Their data contains three actions: dancing, walking and jumping. The dancing action includes five types of dancing scenes. In the situation that there are many similar datasets, Sigal et al. [8] developed a systematic quantitative evaluation of competing methods. Their work supports the development of new articulated motion and poses estimation algorithms, and also provides a baseline for the evaluation and comparison of new methods.

This paper presents the first part of the purpose – recognizing ballroom dance moves. As we described previously, there are two ways to recognize dancing performance - using

Fig. 1. Two types of ballroom dance

Fig. 2. Example of a step (Step A)

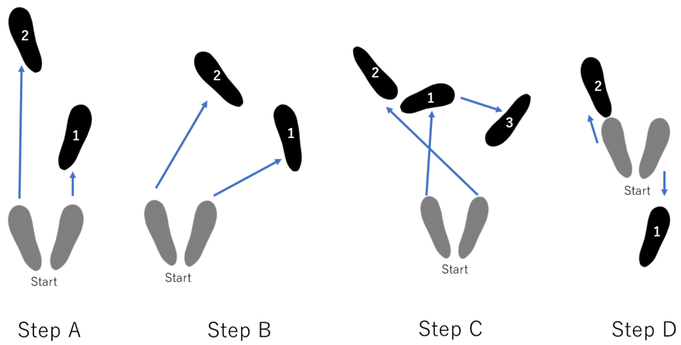

Fig. 3. Diagram of step

videos or wearable sensors. However, each of them has its disadvantages. Using video is a good way to recognize the movement of the whole body. However, it cannot provide accurate timing or strength of foot actions. On the other hand, while a wearable sensor is good at recognizing foot timing and strength, it cannot recognize the movement of the whole body. To solve the problem, we propose a hybrid recognition method utilizing the merits of both video and wearable sensor. We focus to recognize four different types of steps in Latin American, a kind of ballroom dance. For each step, we record wearable sensing data and videos. As a result, we found that the accuracy of step recognition is improved by adding wearable sensing data to video data shot from two different angles. As our future work, we also present several ideas on feedback ways and measurement methods on ballroom dance exercise for each recognized step.

II. Background Knowledge of Ballroom Dance

A. Types of ballroom dance

There are two types of ballroom dance, called Latin American and Standard (see Fig. 1[1 2]). Their dance music genre or movement differs from each other [9]. In this paper, we focus on the former, Latin American. Compared to Standard, Latin American dances are passionate and rhythmical. The roots of the movements and rhythms of Latin American originate in Latin America, Europe, and Africa.

B. Step

The performance of ballroom dance is composed of a combination of structural units, called steps [10]. Fig. 2 shows an example of foot actions of steps. There are numbers of steps in ballroom dance and they are categorized into "Basic figure", "Standard figure", "Named variation", "Variation" [11]. Except for "Variation", steps are charted and named individually. The word "step" in this paper is used to mention these named steps.

[1]https://commons.wikimedia.org/wiki/File:%D0%91%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B9%D1%82%D0%B0%D0%BD%D0%B5%D1%86.jpg

[2]https://www.publicdomainpictures.net/en/hledej.php ?hleda=ballroom

C. Timing

Ballroom dancers perform dance with music. The music count consists of "1, 2, 3, 4, 1, 2, 3, 4, 1, ..." and timings for moving the legs are specified for individual step according to these counts [10] [12](e.g., move right foot on count "4" then left foot on count "2").

D. Hip movement

Hip movement is the movement of the dancer's waist. Keeping hip movement is particularly important for Latin American [12]. The degree and method of hip movement are different for each step.

III. Classification of Steps

In this paper, we aim to recognize the segmented single-step data. Fig. 4 shows the process of our research. In order to verify the classification of the steps, an experienced dancer performed four types of steps. For simplicity, we named the steps as "Step A, Step B, Step C, Step D". The diagram of each step is shown in Fig. 3. Step A and Step B have the same number of foot actions, and other characteristics are also very similar except for the change amount of body orientations. Thus, these two steps have very similar movement. In contrast, the other two steps have different types of foot and body movements. The characteristics of each step are summarized in Table I.

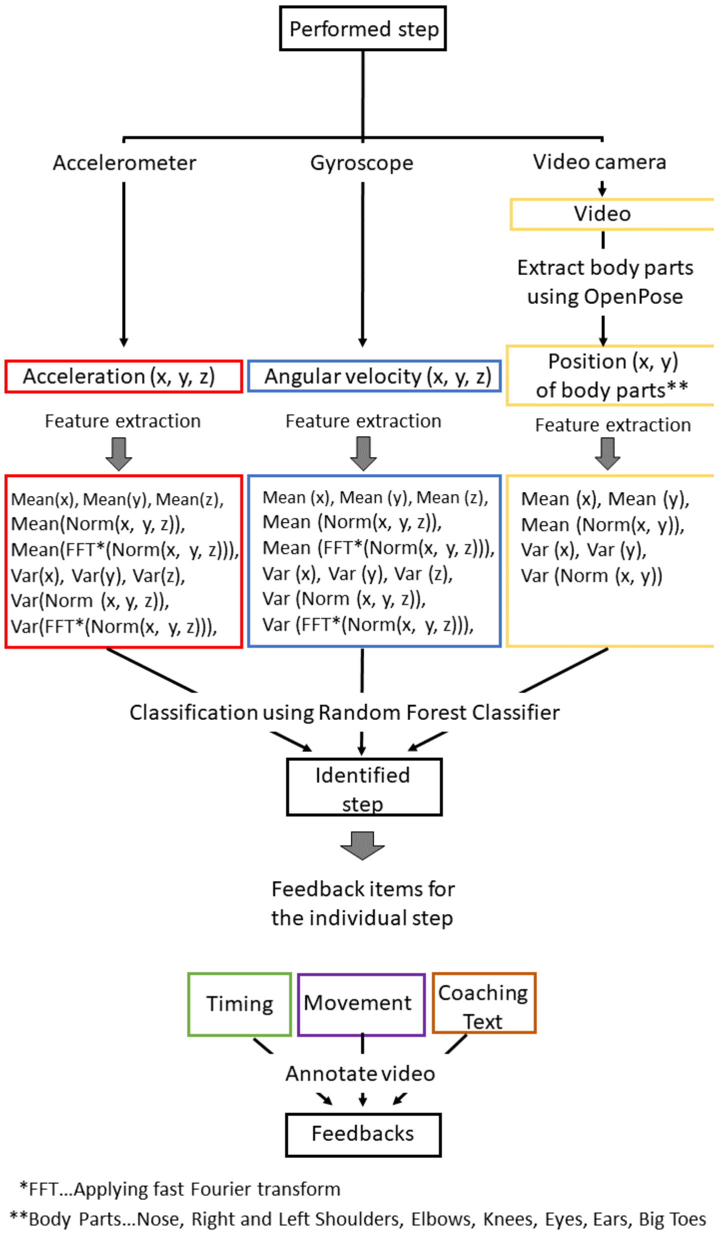

Fig. 4. Work process

A. Data Collection

For each step, both wearable sensor data and video data were acquired under the following conditions. The summary of data collection is shown in Table II.

TABLE I
CHARACTERISTICS OF EACH STEP

	Number of foot actions	Progressing direction	Change amount of body orientation	Amount of hip action
Step A	2 times	Forward	None	Big
Step B	2 times	Forward	Rotate 90 degree to the left	Big
Step C	3 times	Side	Rotate 315 degree to the right	Small
Step D	3 times	Backward	None	Big

TABLE II
SUMMARY OF DATA

	Wearable sensor and video position 1	Wearable sensor and video position 2
Step A	10 times	10 times
Step B	10 times	10 times
Step C	10 times	10 times
Step D	10 times	10 times

1) Wearable Sensor: We obtained the time variation data of acceleration and angular velocity from accelerometer and gyroscope of Nexus 5x worn inside the front right pocket of the pants. The sampling rate was 100 Hz.

2) Video: We obtained videos using Panasonic HDC-TM750 from two different positions. We first shot 10 videos per a step from position 1 and then shot another 10 videos per a step from position 2. The situation of shooting videos is shown in Fig. 5. The sampling rate was 30 fps.

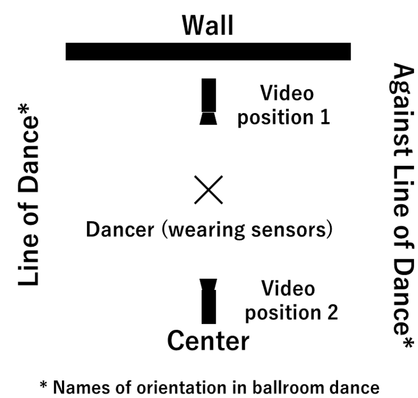

Fig. 5. Shooting videos

B. Classification method

In order to classify the steps, we use Random Forest Classifier [13] with following features. The summary of extracted features is shown in Table III.

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 norm of acceleration(x, y, z) applying the fast Fourier transform
- 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)
- mean and variance of the time change of norm of angular velocity(x, y, z) applying the fast Fourier transform

2) Feature Extraction from Videos: To recognize the characteristics of movements of each steps, we obtain the positions of joint points using OpenPose [14] [15] [16] [17]. Through analysis by OpenPose, two-dimentional positions of 25 joint points are obtained per a frame as shown in Fig. 6. From

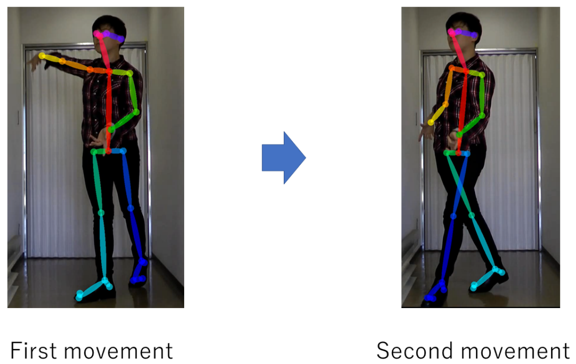

Fig. 6. Joint points from OpenPose

these positions of joint points, we select nose, RShoulder, RElbow, RKnee, REye, REar, RBigtoe, LShoulder, LElbow, LKnee, LEye, LEar, LBigtoe. For the time change data from these positions, we perform following feature extractions.

- mean and variance of the time change of selected joint positions(x, y)
- mean and variance of the time change of norm of selected joint positions(x, y)

TABLE III
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
			mean and variance of time variation
			applying the fast Fourier transform
	Angular velocity	x, y, z	mean and variance of time variation
		norm	mean and variance of time variation
			mean and variance of time variation
			applying the fast Fourier transform
Video	Selected points*	x, y	mean and variance of time variation
		norm	mean and variance of time variation

* nose, shoulders, elbows, knees, eyes, ears, bigtoes

C. Classification Result

To evaluate the accuracy of our classification method, we split data into 50 percent for training and 50 percent for testing. We first trained the classifier with the training data and then made a prediction for test data. Then we summarized the result of prediction as confusion matrixes. Also, we performed 4-fold cross-validation and showed the result. we used f1-score for the evaluation.

Table VI shows the result from videos of two different positions. Although we obtained f1-score 1.0 (shown in Table V) when we only shot videos from one direction, we obtained f1-score 0.87 when we shot from two different positions. Accordingly, we obtained features from wearable sensors and tried classification. We got f1-score 0.85 (shown in Table IV), which is almost the same as the previous result. Although these two results of f1-score are close, wearable sensors and videos captured different aspects of each type of steps. Accordingly, we performed classification using features from both videos and wearable sensors and finally, as shown in Table VII, we obtained f1-score 0.96.

TABLE IV
RESULT FROM VIDEOS OF TWO DIFFERENT POSITIONS

		Predicted			
		step A	step B	step C	step D
Actual	step A	10	0	0	0
	step B	1	5	4	0
	step C	0	0	10	0
	step D	0	0	0	10

TABLE V
RESULT FROM A VIDEO OF ONE POSITION

		Predicted			
		step A	step B	step C	step D
Actual	step A	5	0	0	0
	step B	0	5	0	0
	step C	0	0	5	0
	step D	0	0	0	5

TABLE VI
RESULT FROM WEARABLE SENSORS

		Predicted			
		step A	step B	step C	step D
Actual	step A	10	0	0	0
	step B	2	8	0	0
	step C	0	0	10	0
	step D	0	2	0	8

TABLE VII
RESULT FROM WEARABLE SENSORS AND VIDEO

		Predicted			
		step A	step B	step C	step D
Actual	step A	10	0	0	0
	step B	0	9	0	1
	step C	0	0	10	0
	step D	0	0	0	10

D. Discussion

Although most steps are classified correctly, there were some steps that are mistaken by the classifier. To improve it, we considered ways of feature extraction. In Random Forest, we can get the feature importance [18] of a classifier. Looking at the feature importance of our classifier shown in Table VIII, features from angular velocity (y) , the right and left ears mainly contributed to the classification. On the other hand, features from the big toes had almost no contribution to the classification.

TABLE VIII
FEATURE IMPORTANCES

	Best	Worst
1	Mean of angular velocity (y)	Variance of norm of left big toe
2	Variance of left ear (y)	Variance of right big toe (x)
3	Variance of right ear (y)	Variance of right big toe (y)

This result could be understood considering the characteristics of Latin American. Figure 7 shows the first foot actions in step A, B, C, and D. As shown in Figure 7

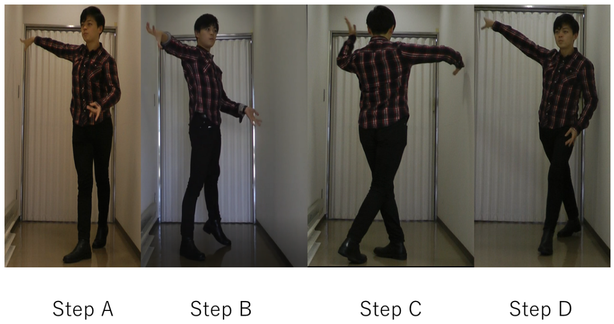

Fig. 7. First foot action in each step

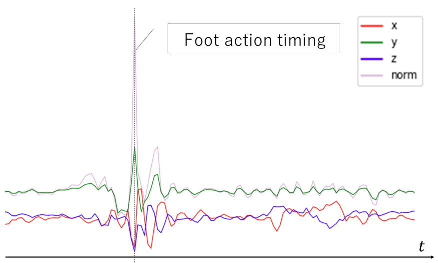

Fig. 8. Acceleration wave

and Table I, the progressing directions, change amounts of body orientation, and the amounts of body movements are characteristic in each step. Thus, the features which express the individual characteristics had many contributions. Observing misclassified steps, Step A and Step B sometimes classified incorrectly. This is because these two steps are consistent with features other than changes in body orientation. To improve the result, we should consider the way to extract the body orientation such as face, shoulders, and feet more accurately. Also, although we are currently attempting to classify in 4 classes, the types of actual dance steps are more diverse. By increasing the number of classes, it is expected that parts that can not be handled by the method of current feature extraction will appear. In order to actually utilize this research, we are expecting to increase the number of classes of steps and also expecting to consider additional methods of feature extraction.

IV. Feedback items to support ballroom dance exercise

Now we discuss the ideas of feedback items. Although there are many items to be exercised in the ballroom dance, we propose the following items which are expected to be provided by using video and wearable sensing data.

A. *Timing*

As shown in Fig. 8, the norm value of acceleration is remarkably large when the foot touched on the floor. Accordingly, calculation of norm value of acceleration will help to recognize timings. In order for this timing to be utilized in

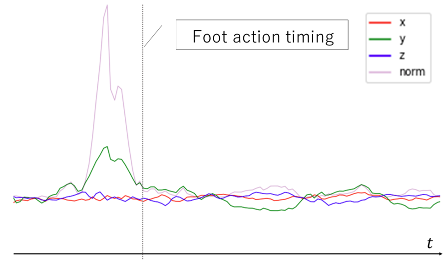

Fig. 9. Angular velocity wave

practice for ballroom dance exercises, we are considering the following feedback.

- The correct timing stated in the ballroom dance teaching materials
- Misalignment with respect to the correct timing

However, there is a need to consider how to handle the deviation between the detected timing and the sense of ballroom dancers. It is a main challenging problem to work in giving feedback on timing. In the future, we aim to recognize the timing of the foot movement with only video data. For that purpose, we should verify the relationships between the information about foot parts from video and acceleration sensor.

B. *Amount of hip movement*

As shown in Fig. 9, the norm value of angular velocity is remarkably large before the foot action occurs. As a hip action occurs before the foot action [12], calculation of norm value of angular velocity will help to recognize hip movements. In order for this hip movement to be utilized in practice for ballroom dance exercises, we are considering the following feedback.

- display the movement recommended by the ballroom dance textbook
- provide dancer's hip movement data by displaying time variation of acceleration and angular velocity

Although, since it is difficult to understand the relationship between time variations of sensing data and hip movement, it is necessary to investigate which motion affects which data. One possible idea is to investigate the relationship between the time variation of the angular velocity and the position of the joint points from OpenPose as shown in Fig. 6.

V. Conclusion

In this paper, we propose a hybrid activity recognition method for ballroom dance exercise using video and wearable sensor. This paper focuses to recognize four different types of steps in Latin American, a kind of ballroom dance. For each step, we record wearable sensing data and videos.

After collecting data, we extracted features from them and classified into individual steps using Random Forest Classifier. As a result, almost all steps are recognized correctly and we obtained f1-score 0.96. We also presented ideas on feedback items and measurement methods to users. Our next work is to construct the method to annotate video feedbacks on the timing and amount of hip movement while researching better ways of feature extraction to the better accuracy of classification of ballroom dance steps.

ACKNOWLEDGMENT

This work was partially supported by JSPS KAKENHI Grant Number JP17H01762.

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] 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.

[3] Sally A. Radell, Daniel D. Adame, and Stephen P. Cole. Effect of Teaching with Mirrors on Ballet Dance Performance. *In Proc. of Perceptual and Motor Skills*, 97(3):960–964, 2003. PMID: 14738364.

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

[5] 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.

[6] Ryan Aylward and Joseph Paradiso. *2006: Sensemble: A Wireless, Compact, Multi-user Sensor System for Interactive Dance*, pages 235–251. 2017.

[7] Ping Wang and J. M. Rehg. A modular approach to the analysis and evaluation of particle filters for figure tracking. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 790–797, 2006.

[8] 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.

[9] Simon Dixon, Elias Pampalk, and Gerhard Widmer. Classification of Dance Music by Periodicity Patterns. *International Conference on Music Information Retrieval*, 2003.

[10] Ballroom Guide. http://www.ballroomguide.com/index.html. [Online; accessed 12-January-2019].

[11] Dictionary of Ballroom Dance and the History of Aichi-ken Pro-dance Instructor Federal, 2016. [in Japanese].

[12] Latin American Rumba, 1999. [in Japanese].

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

[14] 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.

[15] 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.

[16] Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. Hand Keypoint Detection in Single Images using Multiview Bootstrapping. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[17] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional Pose Machines. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[18] L. Breiman. Using Models to Infer Machanisms. http://oz.berkeley.edu/users/breiman/wald2002-2.pdf, 2002. [Online; accessed 12-January-2019].