https://doi.org/10.2352/ISSN.2470-1173.2020.17.3DMP-002

Deadlift Recognition and Application based on Multiple Modalities using Recurrent Neural Network

Shih-Wei Sun[b] , Ting-Chen Mou[a], and Pao-Chi Chang[a]
[a]Dept. of Communication Engineering, National Central University, Taoyuan City, Taiwan, ROC
[b]Dept. of New Media Art, Taipei National University of the Arts, Taipei, Taiwan ROC

Abstract

To improve the workout efficiency and to provide the body movement suggestions to users in a ``smart gym'' environment, we propose to use a depth camera for capturing a user's body parts and mount multiple inertial sensors on the body parts of a user to generate deadlift behavior models generated by a recurrent neural network structure. The contribution of this paper is trifold: 1) The multimodal sensing signals obtained from multiple devices are fused for generating the deadlift behavior classifiers, 2) the recurrent neural network structure can analyze the information from the synchronized skeletal and inertial sensing data, and 3) a Vaplab dataset is generated for evaluating the deadlift behaviors recognizing capability in the proposed method.

I. Introduction

Recognizing workout behaviors is important for users in a gym environment. Recognizing valid movement behavior can help users effectively have a workout, weight training, and get in good shape. On the other hand, with the recognized workout behaviors, the following behavior suggestion from a system is more and more popular. To name a few, many fitness apps are designed for autonomous training, recording diet, and suggesting a fixed training menu. The apps are often connected with smartwatch, smart belt, and other wearable smart devices. For example, FITVISOR [1] automatic assist system is developed by RONFIG and installed in a Sydney ``smart gym'' for providing suggestions to the gym users.

Deadlift is one of the most representative workout behaviors for weight training. Therefore, in this paper, we focus on recognizing deadlift behaviors. It is a challenging task for recognizing deadlift behaviors from a camera-based approach, due to the possible severe self-occlusion situations. On the other hand, the sensing signal amplitude changing measured from wearable inertial sensors can identify possible moving time instant and period, but global moving directional information cannot be revealed. Therefore, in this paper, we propose to adopt a Kinect depth camera [2] for obtaining the skeletal information of a user, and mount multiple X-OSC [3] inertial sensors on the body parts of a user for obtaining the inertial sensing data for deadlift recognition. In addition, a recurrent neural network structure is used for training the behavior classifiers from multi-modal sensors. Specifically, a deep learning process is adopted for generating the deadlift behavior models.

Once the deadlift behaviors can be recognized, the obtained sensing data are evaluated by the proposed system. The scores and the suggestions are displayed to the users to improve the workout behavior in the following possible movements. To validate the proposed prototyping system, we generate a Vaplab dataset by recording the sensing data from multiple devices for a group of gym users. The experimental results demonstrated that the proposed system can effectively recognize deadlift behaviors.

II. Proposed method

The system block diagram of the proposed deadlift recognition system is shown in Fig. 1. The depth data and the inertial sensor signals are captured and recorded on the left part of Fig. 1. The skeleton features and the IMU features are extracted from the depth camera modality and the inertial sensor modality, correspondingly. After the pre-processing form the raw data of multiple modalities, the feature vectors of different modalities are synchronized in time and combined as a vector in each time frame. Based on the obtained feature vectors, at the bottom of the right side of Fig. 1, a long short-term memory (LSTM) recurrent neural network (RNN) architecture is applied to generate the deadlift behavior classifiers. Moreover, the recognized behavior results with the feature vectors are further analyzed for movement evaluation. The proposed will feedback a user for the score assessment result and provide the movement suggestion texts to a user. The details of the block diagram of Fig. 1 are described in the following subsections.

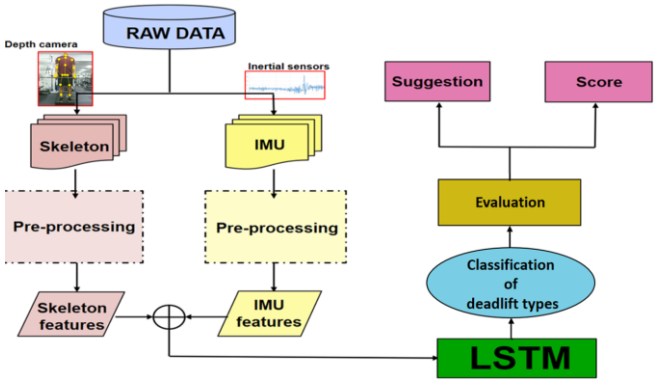


Fig. 1. The block diagram of the proposed deadlift recognition system.

A. Feature extraction from a depth camera and multiple inertial sensors

Because the sampling rate from multiple sensing devices might be different, to temporally synchronize the signals obtained from multimodal devices, a resample process for all of the data obtained from all the devices is necessary. A difference operation from the current time instant to the previous ones of the centroid position, gyro sensing signal,

and accelerometer signal are calculated as the equations (1), (2) and (3):

$$C_{d,t} = \sqrt{(c_{x,t} - c_{x,t-1})^2 + (c_{y,t} - c_{y,t-1})^2 + (c_{z,t} - c_{z,t-1})^2} \quad (1)$$

$$G_{d,t,k} = \sqrt{(g_{x,t,k} - g_{x,t-1,k})^2 + (g_{y,t,k} - g_{y,t-1,k})^2 + (g_{z,t,k} - g_{z,t-1,k})^2} \quad (2)$$

$$A_{d,t,k} = \sqrt{(a_{x,t,k} - a_{x,t-1,k})^2 + (a_{y,t,k} - a_{y,t-1,k})^2 + (a_{z,t,k} - a_{z,t-1,k})^2} \quad (3)$$

According to the criteria defined in [4], the maximum relative difference of 5% in equations (1), (2) and (3) are used as the threshold for signal temporal segmentation.

For each temporal segment, all the sensing signals are empirically resampled to 120 sampling points in this paper. For the depth camera modality, the obtained frames are resampled to 120 frames for each temporal segment. In addition, we adopt a time-variant skeleton vector projection method [5] (our previous work) to extract the feature in a frame. According to the obtained 25 skeleton joints (from Kinect v2 official sdk), the shoulder vector S and the foot vector F can be obtained. According to a cross-product from the vector S and vector F, a normal vector N (the yellow arrow) can be obtained, with a mutually orthogonal property. Once a joint is obtained, e.g. j_t^{hr} in Fig. 3, its projective vectors to the bases N, F, and S, can be obtained, generating the feature vectors. Therefore, in each temporal segment, for the depth camera modality, the amount of skeleton feature data is 120 (frame number) * 25 (number of nodes) * 3 (base vector), a total of 9,000 values of data.

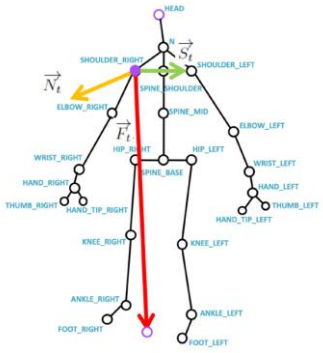


Fig. 2. Skeleton base vector diagram [5].

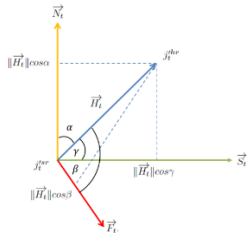


Fig. 3. Projection volume [5].

For the inertial sensor modality, the obtained data is resampled to 120 sampling points. As shown in Fig. 4, each resampled temporal segment is divided into six intervals (each section with the length 20). For one of the six intervals, the average value μ, the standard deviation σ, and the variation number σ^2 in x-, y-, and z- directions are obtained as the features [6]. Therefore, in each temporal segment, for the inertial sensor modality, the amount of the inertial sensor feature data is 4 (number of devices) * 2 (accelerometer sensor and gyro sensor) * 9 (features numbers) * 6 (interval numbers), a total of 432 values of data.

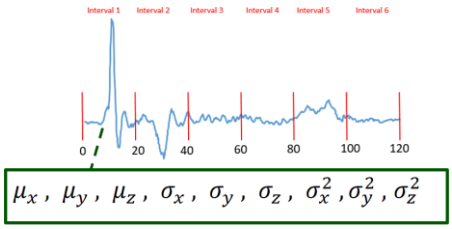


Fig. 4. Features in a temporal segment of the inertial sensor modality.

B. *Deadlift classifier training from a recurrent neural network structure*

Once the features from the depth camera modality and the inertial sensor modality are obtained, as shown in the right bottom part of Fig. 1, an LSTM recurrent neural network architecture is adopted for training the deadlift classifiers. Before sending into the LSTM, the obtained features are concatenated and flattened as a one-dimensional vector as shown in Fig. 5. To keep the temporal synchronized manner, the features in the depth camera is also divided into six equal partitions. Finally, the obtained concatenated feature vector is as shown in Fig. 5.

Improved from the conventional RNN, in this paper, we applied LSTM [7] to observe the received information with long-term memory for the sequential data from the depth camera modality and the inertial sensor modality, with the LSTM parameter setting in [8]. The hidden layer size is related to the complexity of the observed data, according to the empirical tests for the layer size of one, two, and three, the hidden layer size is determined to two (LSTM_1 and LSTM_2), as shown in Fig. 6, which is the training process for obtaining the LSTM-based classifiers.

$$\begin{aligned} features\ concatenated : & Sensor\_1_1 , Sensor\_2_1 , Sensor\_3_1 , Sensor\_4_1 , \\ & Kinect_1 , Kinect_2 , \ldots\ldots , Kinect_{1499} , Kinect_{1500} , \\ & Sensor\_1_2 , Sensor\_2_2 , Sensor\_3_2 , Sensor\_4_2 , \\ & Kinect_{1501} , Kinect_{1502} , \ldots\ldots , Kinect_{2999} , Kinect_{3000} , \\ & Sensor\_1_3 , Sensor\_2_3 , Sensor\_3_3 , Sensor\_4_3 , \\ & Kinect_{3001} , Kinect_{3002} , \ldots\ldots , Kinect_{4499} , Kinect_{4500} , \\ & Sensor\_1_4 , Sensor\_2_4 , Sensor\_3_4 , Sensor\_4_4 , \\ & Kinect_{4501} , Kinect_{4502} , \ldots\ldots , Kinect_{5999} , Kinect_{6000} , \\ & Sensor\_1_5 , Sensor\_2_5 , Sensor\_3_5 , Sensor\_4_5 , \\ & Kinect_{6001} , Kinect_{6002} , \ldots\ldots , Kinect_{7499} , Kinect_{7500} , \\ & Sensor\_1_6 , Sensor\_2_6 , Sensor\_3_6 , Sensor\_4_6 , \\ & Kinect_{7501} , Kinect_{7502} , \ldots\ldots , Kinect_{8999} , Kinect_{9000} \end{aligned}$$

Fig. 5. The complete features from multiple modalities.

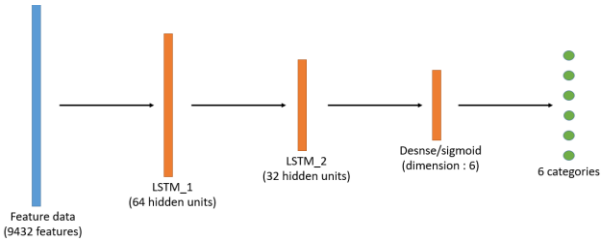


Fig. 6. The training process of the proposed LSTM-based scheme.

III. Experimental Results

In this paper, a deadlift behavior dataset is generated and used to evaluate the performance of the proposed method. In addition, the textual feedback from the proposed system (the upper-right part of Fig. 1) will be discussed.

A. Deadlift behavior dataset

In this paper, we generate a dataset for obtaining deadlift behaviors, called `` VAPLAB Multi-Modality Fitness Behavior Dataset''. As shown in Fig. 7, a Kinect depth camera is installed in front of the user about 2.1 meters. 4 x-OSC inertial sensors are worn on the wrists and ankles on both hands of the user. In the dataset, 6 behaviors are operated by the professional workout users in a gym, including: Conventional Dead Lift (CDL), Sumo Deadlift (SDL), Romanian Deadlift (RDL), Stiff-legged Deadlift (SLDL), Block pull Deadlift (BPDL), Deficit Deadlift (DDL). The snapshots of the representative deadlift behaviors in the color frames are shown in Fig. 8. Furthermore, in `` VAPLAB Multi-Modality Fitness Behavior Dataset'', the inertial data, depth frames, color frames, and skeleton joints are recorded, as shown in Fig. 9.

Fig. 7. The snapshot of recording the deadlift behavior of a user.

Fig. 8. Representative deadlift behaviors in the dataset.

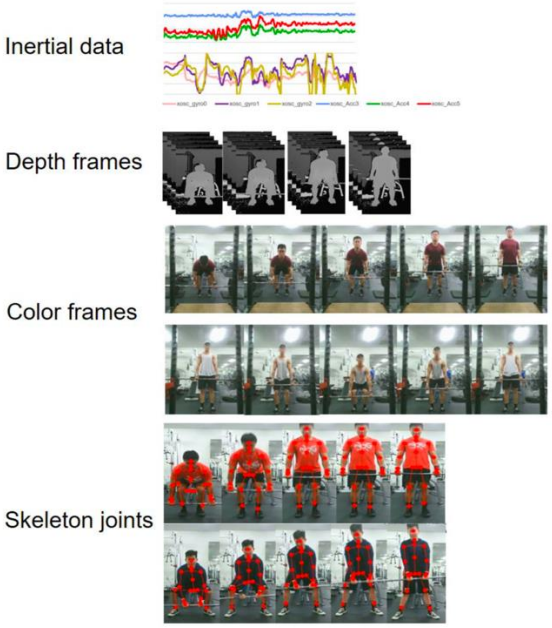


Fig. 9. Representative deadlift behaviors in the dataset.

B. Deadlift recognition results

In the experimental results, an Intel Core i7 CPU, 8GB ram computer is used. We used python 3.6 with tensor flow [9] and Keras [10] to implement the RNN algorithm. There were 6 professional workout users invited to operate the CDL, SDL, RDL, SLDL, BPDL, and DDL behaviors. In the experimental results, we adopt a leave one out cross-validation to train the RNN models. For each behavior, a user-operated the same behaviors for 10 times. In the RNN training process, we implemented 500 epochs, and the recognition accuracy of the confusion matrix is shown in Table 1, with the average accuracy 79.99%, average training time 531.9 seconds for training one behavior model.

In the confusion matrix shown Table I, the accuracy of CDL is only 60%. In this case, 20% DDL samples are wrongly predicted as CDL. In addition, 14.44% SLDL are also wrongly predicted as CDL. The representative sample color frames and depth frames of CDL, SLDL, and DDL are shown in the first row and the second row of Fig. 10. The depth frames captured from the depth camera are similar for the CDL, SLDL, and DDL. From the depth camera modality, it has chances to wrongly recognize the behaviors to the others, as shown by the last row (yellow rectangular area) of Fig. 10.

On the other hand, for the inertial sensor modality, the IMU features (the third row of Fig. 10) have different signal properties, as shown by the green rectangular areas and red rectangular areas in Fig. 10. Therefore, by combining the inertial sensor modality into the feature vectors, we still have the chances to correctly recognize the CDL behavior with the average accuracy 60%.

TABLE I. ACCURACY OF EACH ACTION

ACTUAL \ PREDICT	BPDL	CDL	DDL	RDL	SDL	SLDL
BPDL	90%	3.33%	1.11%	2.22%	0	3.33%
CDL	1.11%	60%	13.33%	2.22%	0	23.33%
DDL	0	20%	64.44%	1.11%	0	14.44%
RDL	0	0	0	100%	0	0
SDL	0	1.11%	2.22%	0	96.66%	0
SLDL	4.44%	14.44%	12.22%	0	0	68.88%

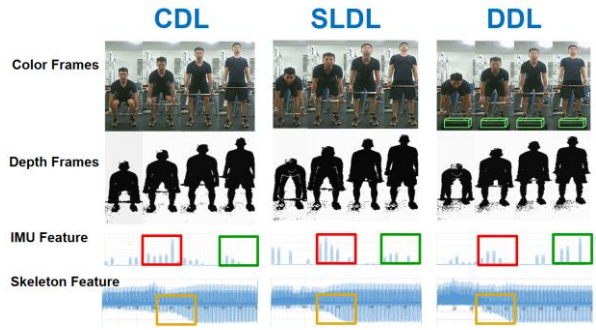


Fig. 10. CDL, SLDL, DDL comparison diagram

C. *Behavior evaluation from the system*

On a deadlift behavior can be roughly divided into five phases, as shown in Table II. For example, in the beginning, the CDL behavior is with the ``Starting tension'' phase (Phase 1 in Table II), and an example is shown by Fig. 11. The red oval area marked in Fig. 11 depicts two peak values (measured from the sensor modality) occur when the user was trying to use his muscle to move the object. Next, in CDL behavior, a ``Humpback'' phase (phase 2 in Table II) should be operated by the user, and an example is shown by Fig. 12. The red line segments depicted in Fig. 12 demonstrated a good ``Humpback'' pose in the most left part of Fig. 12. The joints of spine_shoulder, spine_mid, and spine_base should be in the common line segment, and it can be revealed from a camera modality.

TABLE II. JUDGMENT CRITERIA FOR EACH ACTION

	1	2	3	4	5
CDL	Starting tension	Humpback	Barbell drop	Stand erect	Shoulders overcompensation
SDL	Starting tension	Humpback	Barbell drop	Stand erect	Knee Valgus
RDL	Hip-driven	Joint lock	ROM	Over bending	Hip thrust
SLDL	Humpback	Hip position	Proper force producer	Stand erect	Lose balance
BPDL	Starting tension	Humpback	Hip thrust	Over arching	Arms overcompensation
DDL	Starting tension	Humpback	Barbell drop	Stand erect	Shoulders overcompensation

Fig. 11. Starting tension

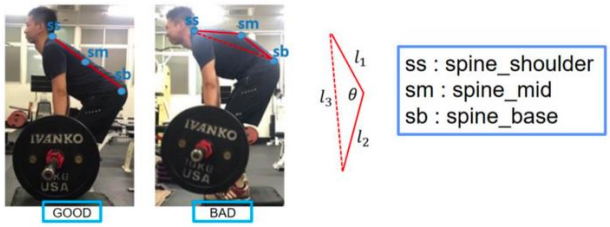


Fig. 12. Humpback

Based on the sensor modality and the camera modality, in the proposed system, a deadlift behavior operated by a user is evaluated and scored by the criteria of the competition rules defined by International Powerlifting Federation (IPF) [11]. The reasons to be judged as a fail case [11] are: "1) Any downward movement of the bar before it reaches the final position. 2) Failure to stand erect with the shoulders back. 3) Failure to lock the knees straight at the completion of the lift." In our system, the deadlift behaviors are scored by the definitions in [11]. For example, in Table III, the users of case 1 and case 2 can obtain the corresponding scores and textual suggestions in the prototype of the proposed smart gym system.

TABLE III. EACH JUDGING STANDARD CALCULATES THE SCORE PROPORTIONALLY

	CASE 1	CASE 2
RDL_1	Hip-driven (20%)	Hip-driven (20%)
RDL_2	θ = 152.98 (9.08%)	θ = 147.85 (11.72%)
RDL_3	Standard ROM (20%)	Standard ROM (20%)
RDL_4	247 times (15.06%)	430 times (11.40%)
RDL_5	Value = 0.1754 (6.17%)	Value = 0.3021 (20%)
Score	70.32	83.12
Suggestion	1. Bend knee and push hip back to protect your knee joint. 2. Strengthen latissimus doris and keep the tension during the movement. 3. Do a barbell hip thrust to improve pushing power of hip.	1. Bend knee and push hip back to protect your knee joint. 2. Strengthen latissimus doris and keep the tension during the movement.

IV. Conclusion

In this paper, we proposed a prototyping smart gym system. A depth camera and four inertial sensors are used to capture the body movements and measure the joint changing movements, correspondingly. According to the information obtained from the camera modality and the sensor modality, the deadlift behavior models are generated by a recurrent neural network structure. There are three contributions in this paper: 1) we fused multimodal sensing signals to obtain the deadlift behavior classifiers, 2) the temporal synchronized skeletal and inertial sensing data are used to train deadlift models, and 3) we proposed a Vaplab deadlift dataset to be used for evaluation.

However, the proposed deadlift behavior recognition system is still in the begging phase for developing a complete system. In many workout applications, the feelings from the muscles are more important than the appearance of a pose. In the future, muscle sensors, e.g., MYO, can be applied for further study and research.

References

[1] [Online]. Available: https://www.ronfic.com/

[2] [Online]. Available: https://www.xbox.com/xbox-one/kinect

[3] [Online]. Available: http://x-io.co.uk/x-osc/

[4] G. Zhu, L. Zhang, P. Shen, and J. Song, “An online continuous human action recognition algorithm based on the Kinect sensor”, Sensors, vol. 16, no. 2, pp. 161, 2016.

[5] C. H. Kuo, P. C. Chang, and S. W. Sun, "Behavior Recognition Using Multiple Depth Cameras Based on a Time-Variant Skeleton Vector Projection," in IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 1, no. 4, pp. 294-304, Aug. 2017.

[6]] S.W. Sun, T.C. Mou, C.C. Fang, P.C. Chang, K.L. Hua, and H.C. Shih, “Baseball Player Behavior Classification System Using Long Short-Term Memory with Multimodal Features,” Sensors SCI journal, pp.1425, 2019.

[7] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural computation, 9(8):1735–1780, 1997

[8] Chih-Chieh Fang, Ting-Chen Mou, Shih-Wei Sun, Pao-Chi Chang, “Maching-Learning Based Fitness Behavior Recognition from Camera and Sensor Modalities,” in IEEE International Conference on Artificial Intelligence and Virtual Reality(AIVR),17 January 2019.

[9] [Online]. Available: https://www.tensorflow.org/

[10] [Online]. Available: https://keras.io/

[11] [Online].Available: https://www.powerlifting.sport/rulescodesinfo/technical-rules.html

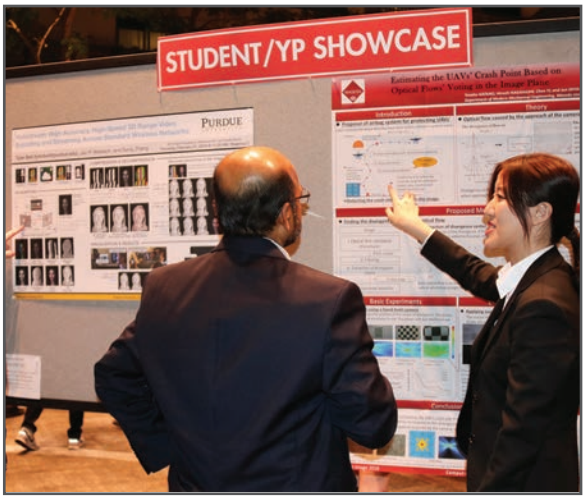
JOIN US AT THE NEXT EI!

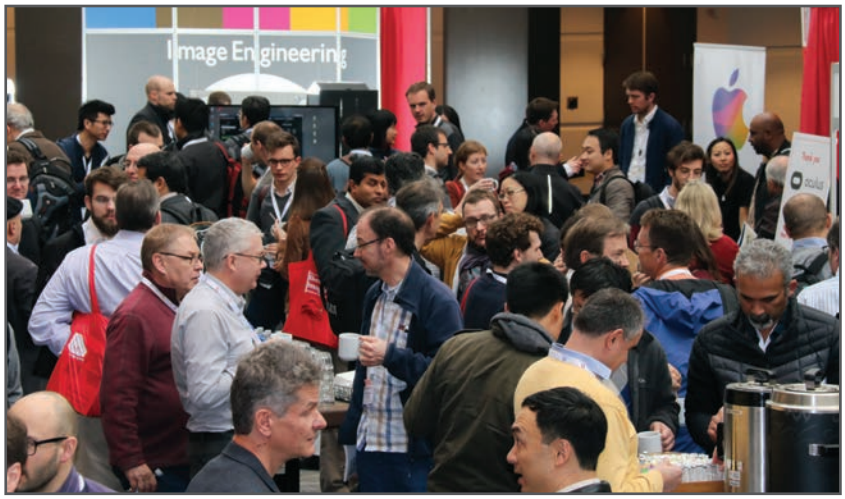
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org