# Multisensor-Based 3D Gesture Recognition for a Decision-Making Training System

Tse-Yu Pan, *Member, IEEE,* Chen-Yuan Chang, Wan-Lun Tsai, *Student Member, IEEE,* and Min-Chun Hu, *Member, IEEE,*

***Abstract*—This paper demonstrates a gesture recognition method using multiple Inertial Measurement Unit (IMU) sensors, which can record acceleration and rotation information for hand joints. The proposed gesture recognition method comprises frequency ConvNet and TemporalNet to extract the representative features within a sliding window of IMU signals for recognizing various types of hand gestures. To validate the proposed gesture recognition method, basketball official referee signals (ORSs), which comprise sixty-five types of gestures including both large motion hand movement and subtle motion hand movement, are utilized as the main recognition task to evaluate the proposed method. The evaluation results reveal the proposed recognition model can achieve convinced performance, which outperforms other existing works. In addition, the satisfied performance of the proposed recognition model encourages us to develop a decision-making training (DMT) system for cultivating basketball referees. The results of subjective evaluations by the recruited 20 participants indicate the training system based on the proposed gestures recognition method can efficiently strengthen the decision-making skills of users.**

***Index Terms*—Gesture Recognition, Wearable Sensor, Hybrid Neural Network, Sports Training, Decision-Making, Training System.**

## I. Introduction

DECISION making is defined as a cognitive operation involved in the selection of a response from a range of available responses in circumstances where an action is required, and it has been investigated in many fields. For example, in fast-paced sports, an athlete or a team might have to make snap decisions to determine which tactics should be employed to win the game, or a referee might have to make snap decisions when there is a foul. Immediate feedback on the correctness of decisions without further explanations is sufficient for raising decision accuracy. Therefore, researchers in sport psychology investigate different efficient training methods by which to improve the perceptual skills of athletes and referees, such as sports vision training (SVT) and pressure simulation [1]–[3] to help them make better decision under realistic simulated situations of dynamic, fast-paced sports. However, accompanied with an improvement in perception skills, the correctness of cognitive behavior should be given more attention to ensure the decision can achieve its intended goal. Over the past decades, advancements of hand gesture recognition technology has led the growth of the researches in human-computer interaction (HCI), which can facilitate the development of a suitable decision-making training (DMT) system for cultivating the skills of users. To develop a beneficial DMT system, basketball Official Referee's Signals (ORSs) [4], which comprise sixty-five types of gestures including both large motion hand movement and subtle motion hand movement, are considered as an example of hand gesture recognition task. Then, the proposed hand gesture recognition method will be used for developing of basketball referee DMT system. In a basketball game, when a violation event, a foul event, or a scoring event happens, referee plays an indispensable role to correctly transmit the judgement to the scorer or players, or the game will be delayed, and the players may dispute the call. To become an authoritative referee, the new learner had better practice frequently or experience being a referee in profession-level games. However, there is only extremely few opportunities to be the referee of formal games. Therefore, in this work, we focus on proposing a hand gesture recognition method, which can be further tackle the recognition of ORSs and applied to a basketball referee DMT system.

The hand gesture recognition approach can be utilized through the use of vision camera or wearable sensors. However, the use of only an RGB camera may easily suffer from the complexity of the background environment such as colors or textures [5]. The combinations of an RGB camera with additional techniques of 2D human pose estimation [6], [7], or a single depth camera [8], [9] also have problems of the object occlusion issue and the limited distance between the user and the camera, resulting the restriction of applications [10]. By contrast, wearable sensors such as temperature sensors, heart rate sensors, electrocardiogram sensors, flexible sensors, textile sensors, accelerometers, gyroscopes, and IMU sensors are widely utilized to monitor different human activities [11]–[13]. Particularly, IMU sensors are cost-effective, compact, low-power and non-invasive, and therefore can be directly attached on human body to obtain relatively accurate body motion [14], [15]. In this work, we focus on estimating human hand gestures consisting of arm or finger movements based on IMUs, which have outstanding performance for accurate and portable motion tracking. Therefore, with the aid of wearable

Manuscript received Month Date, Year; revised Month Date, Year.

This research was supported by the Ministry of Science and Technology (contracts MOST-108-2221-E-007 -106-MY3 and MOST-105-2221-E-006-066-MY3), Taiwan. This research received funding from the Headquarters of University Advancement at the National Cheng Kung University to the Intelligent Manufacturing Research Center (iMRC), National Cheng Kung University, which is sponsored by the Ministry of Education, Taiwan, ROC.

T.-Y. Pan and M.-C Hu are with the Department of Computer Science, National Tsing Hua University, Hsinchu 300, Taiwan (e-mail: {typan@mx.nthu.edu.tw; anitahu@cs.nthu.edu.tw)

W.-L. Tsai and C.-Y. Chang are with the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 701, Taiwan (e-mail: {lookoutking, harrychang}@mislab.csie.ncku.edu.tw)

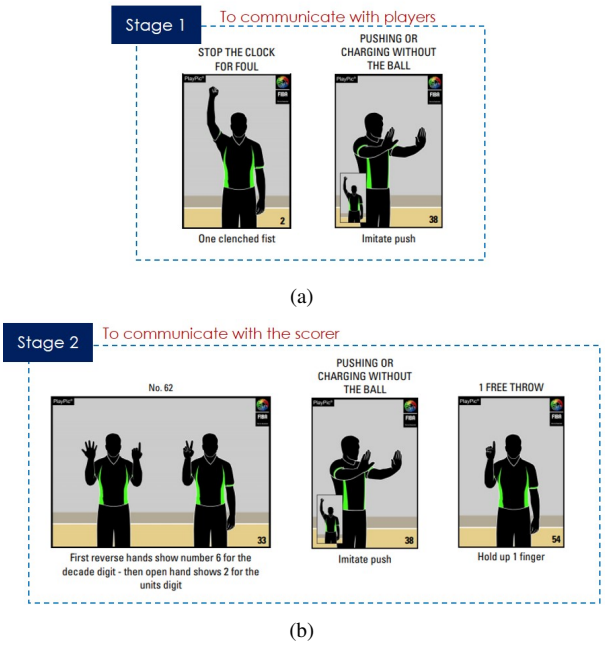


Fig. 1. An example sequence of basketball ORSs to make judgment while a foul happens. (a) In the first stage, the referee must to pause the game by raising the clenched fist, then a "PUSHING" gesture is performed to present what the foul was to players. (b) In the second stage, a sequence of gestures are performing for revealing the player number of the player who fouled, which foul, and the penalty to the scorer and players.

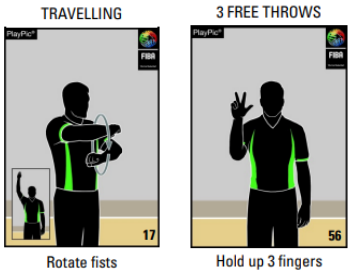


Fig. 2. (a) An example of large motion basketball ORS, where the referee has to rotate his/her arms and fist for a while. (b) An example of subtle motion basketball ORS, where the referee only holds up his/her fingers statically.

sensor technology, a wearable sensor suit equipped with multiple IMU sensors called the Perception Neural [1] device is utilized to provide a more realistic interactive experience.

There are some challenges arose in the task of ORSs recognition for DMT system as follows: (1) When the referee make a judgment, typically, a sequence of ORSs should be correctly and proficient performed to transmit the decision to the scorer and players. Fig. 1 illustrates an example of the two stages when a foul happens. In the first stage, a pause signal of foul events is made toward players followed by a signal demonstrating what kind of foul event is. In the second stage, a sequence of ORS indicating the player number of who made the foul, the kind of the foul is, and the penalty are made toward the scorer in that order. According to the requirement, the proposed gesture recognition method should possess the ability of tackling continuous IMU signals. (2) ORSs comprise sixty-five types of gestures including both large motion hand movement and subtle motion hand movement as shown in Fig. 2. Hence, it should be hard to find proper features that distinguish among all types of gestures. To develop a real-time basketball referee DMT system, the recognition method cannot be too complicated. (3) The placements of multiple IMU sensors might have slight offset for each user because of the individual physical characteristics. The proposed basketball referee DMT system should be flexible for general use without the complex calibration procedure.

In our preliminary studies for of ORSs recognition [16], [17], the improvements of gesture recognition performance is necessary. In addition, the consideration of developing basketball referee DMT system should be addressed. Therefore, the contributions of this work are highlighted as follows. (1) A multisensor-based gesture recognition method is presented, in which a suit of wearable IMU sensors is utilized to not only record acceleration and rotation information for hand joints but also to determine which basketball ORS is executed by the user. (2) The proposed gesture recognition method which involves frequency ConvNets, fusion ConvNet, and TemporalNet can extract more representative features in continuous IMU signals, and distinguish among sixty-five types of ORSs. To improve the feasibility of the proposed recognition method, a semi-supervised scheme called Ladder Network [18] is attached to the proposed recognition method to benefit the performance by using unlabeled data. (3) An interactive video-based basketball referee DMT system based on the proposed gesture recognition method is further demonstrated, and twenty basketball players are recruited to figure out simple violations or fouls in order to conduct the effectiveness of the proposed DMT system.

## II. Related Work

### A. Wearable Sensors for Gesture Recognition

Over the past decade, wearable sensor technology has led the development of different sensors for gesture recognition such as flex sensors [19], biopotential sensors (e.g., EMG sensors and EEG sensors) [20], [21], and inertial sensors [22]. Particularly, advancements of inertial sensors has been applied in many fields according the placements of sensors such as body health monitoring [23], activity recognition [15], sign language recognition [24], and gait detection [25]. However, inertial sensors might be affected by metal sources and the nearby electronics, resulting in poor signals for recognition. In the past, the researchers attempted to consider the hand-crafted features to facilitate the recognition/classification tasks. Nowadays, deep learning algorithm has been proved that it has a strong ability to find more representative correlations among data in many research areas, and a gesture recognition task is no exception. For example, Ha *et al.* [26] utilized the idea of Convolutional Neural Network (CNN) to deal with IMU signals for recognition task. To glean the specific characteristic and normal characteristic in intra and inter modality sensor data, both of CNN-pf and CNN-pff were proposed to combine partial and full weight sharing strategy to pursue

[1]https://neuronmocap.com/

the accurate recognition results. Kim *et al.* [27] presented an improved restricted column energy (RCE) neural network with the aid of dynamic time warping technique, which can tackle time-dependent data captured from an IMU sensors. Moreover, in accordance with literature review of human activity recognition based on deep learning model [28], they indicated that the more complicated deep learning model (e.g., the conjunction of convolutional model and recurrent model) may achieve better performance. For instance, Ordonez *et al.* [29] proposed a framework composed of convolution layers and Long Short-Term Memory Network (LSTM) recurrent layers, which has the more satisfied results for learning the feature representations and temporal dependencies necessary than only consider convolutional model or recurrent model. Inspired by the advantage of the conjunction of convolutional model and recurrent model, the proposed recognition model followed the similar idea to extract the representative local features and temporal dependencies pursuing the superior recognition ability. In addition, to develop a real-time DMT system, the proposed recognition model should be not too complicated.

### B. Decision-Making Training (DMT) System

Video technologies such as 2D videos, 3D panorama videos, and computer-simulated virtual environments have been widely used for perceptual skill training in sports. For example, Larkin *et al.* [30] proposed a video-based training program to cultivate Australian football umpires, and evaluated the proposed program using three test sessions (i.e., a pre-test and one-week and three-week retention assessments). This work indicates that a video-based DMT program can provide more impressive effects when make decisions in the formal games, especially for those umpires who has less experience to play the formal game. Schweizer *et al.* [31] proposed that establishing standards, which can be established by video training programs, may help referees not only deal with ambiguous situations but also may improve the quality of their decisions. Hohmann *et al.* [32] conducted the the effectiveness of learning tactic between using 3D video-based decision training and using tactical board in a national youth handball team. The results showed that 3D training approach can lead to a greater degree of correctness and less decision time than other alternatives. However, the above-mentioned works only focused on perceptual skill training without precise pose/gesture measurements. In the proposed scenario, the advantage of wearable sensor technology was considered to measure the correctness of the human gestures under consideration, providing a more convenient DMT system for the new learner to understand how to perform a sequence of ORSs correctly.

## III. Gesture Recognition Methodology

Fig. 3 illustrates the proposed gesture recognition model, which is composed into pre-processing, high frequency ConvNet, low frequency ConvNet, fusion ConvNet, TemporalNet, and output loss. In this work, a multi-IMUs suit called the Perception Neuron, which is powered by Noitom Ltd., was utilized to record the raw signals of IMUs. Perception Neuron provides 32 IMUs for upper and lower body. Since all the referee signals are performed by hand gestures, the IMUs for lower body can be ignored. Moreover, referee signals including player number information and arm movements, which means the movements of all fingers and both arms are important. Therefore, 18 sensors on the upper body joints (as shown in Fig. 3) were used to capture the motion information in our experiments. Eighteen IMUs which comprise a three-axis accelerometer as well as a three-axis gyroscope were utilized to record the information of acceleration and rotation from two upper arms, two fore arms, two palm backs, and ten fingers. There were totally 108 channels (= 18 sensors $*$ 2 modalities $*$ 3 axis) of raw signals for the input data. Note that size of each IMU is 12.5mm x 13.1mm x 4.3mm*, dynamic range is 360 deg, accelerometer range is 16g, gyroscope range is 2000 dps, and resolution is 0.02 deg.

The sliding window strategy is first applied to segment the input data into several overlapping clips. Since the time length of each gesture is around 1 3 seconds, we set the sliding window as 1 second. The sampling rate of Perception Neuron is 60 fps, and therefore the sample size of each sliding window is 60. Moreover, the sliding step is empirically set as 5 samples. Then, each sliding window was divided into four fragments to facilitate the recognition procedure. In this section, the detail for each components of the proposed recognition model will be described.

### A. Pre-processing

In order to facilitate the following feature extraction procedure, Fourier transform (FFT) was utilized in each fragments to transfer the time series signals into frequency domain signals. In this work, the first ten FFT Coefficients were taken as inputs of the high/low frequency ConvNet module. Each FFT Coefficient has the real part and the imaginary part, resulting in twenty dimensions for each fragment in each channel. Note that data with the same modality were placed together, which means channels of accelerometers were arranged together, followed by channels of gyroscopes.

### B. Frequency ConvNets

In accordance with conventional IMU data analysis method [33], complementary filter is a trick to deal with noisy or missing signals, which utilized a low-pass filter to better the attitude estimated by the accelerometer and a high-pass filter to enhance attitude estimated by the gyroscope. Inspired by the above-mentioned work, inputs of FFT Coefficients were separated into high-order frequencies and low-order frequencies as two streams to extract features of local dependencies, indicating the high frequency ConvNet and the low frequency ConvNet in Fig. 3. To be more precise, first-order to fifth-order FFT Coefficients were considered as low-order frequencies, while sixth-order to tenth-order FFT Coefficients were considered as low-order frequencies. Since features of high frequencies and low frequencies may have distinct dependencies, the frequency-based partial weight sharing strategy has been

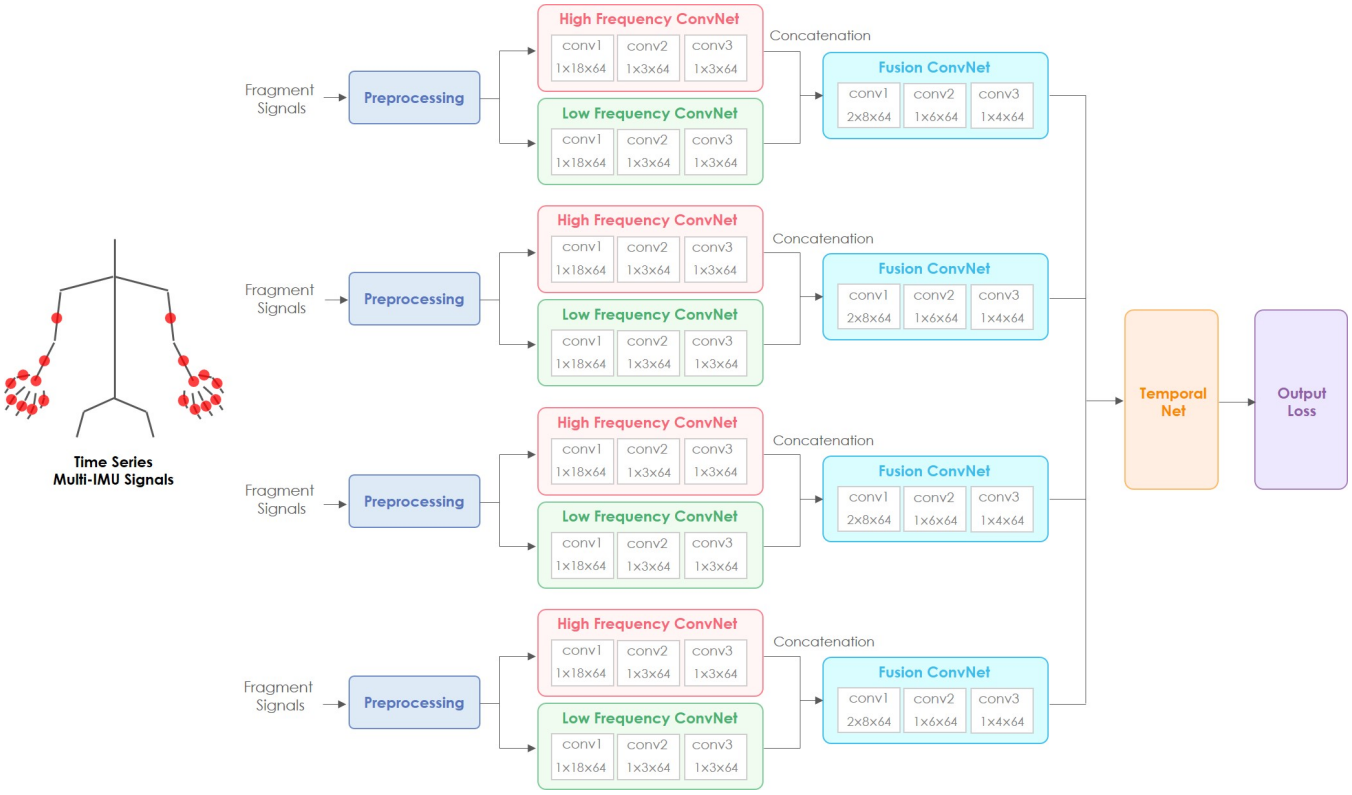


Fig. 3. The proposed gesture recognition model, which is composed into pre-processing module, high frequency ConvNet module, low frequency ConvNet module, fusion ConvNet module, TemporalNet module, and output loss module.

applied to high-order frequencies and low-order frequencies individually to reduce the computational complexity.

In two-stream frequency ConvNets, the first convolution layer with $(1, 18)$ kernel filters (i.e., 3 sensors $*$ 3 axis $*$ 2 for the real part and the imaginary part) were applied to extract the local patterns in both frequencies. Then, two $(1*3)$ kernel filters were used to gain a much more symbolic dependency in the coming two convolutional layers.

### C. Fusion ConvNet

The outputs of two-stream frequency ConvNets were concatenated to form 2D feature maps as the inputs for the fusion ConvNet. Note that a 2D feature map was formed by the outputs of low frequency ConvNet in the first row and outputs of high frequency ConvNet in the second row. Inspired by [34], convolution fusion strategy can achieve best performance rather than other strategies (e.g., sum fusion, max fusion, or bilinear fusion) when fusing outputs of two streams. Therefore, in the fusion ConvNet, the first convolution layer with 2D filter shaped $(2 * 8)$ was utilized to fuse the relationship from the dependency of both high frequencies and the low frequencies. Later, two convolution layers with $(1 * 6)$ kernel filters and $(1 * 4)$ kernel filters were used to obtain high-level features layer by layer.

### D. TemporalNet

In order to model the temporal relationship among signals of four fragments, Gated Recurrent Units (GRUs) were utilized to handle the long-term dependency with outputs of fusion ConvNet of four segments. Compared to Long Short-Term Memory (LSTM) [35], GRU is faster and more flexible to extract the time sequence features. Therefore, one-layer GRU was employed to connect the four outputs of fusion ConvNet within a sliding window as TemporalNet. It should be noted that flatten strategy was applied to outputs of the fusion ConvNet before connecting to TemporalNet.

### E. Implementation and Training

The softmax loss was applied as the output loss, and trained the proposed recognition model by minimizing the cross-entropy between the output gesture category and its ground truth category. Since the number of gesture category is significant, the center loss was further combined to ensure the same gestures can be closer in the embedding space if they belong to the same label. In this way, the distance in the embedding space between the data and the center of others with the same label can be minimized. Therefore, the softmax loss function and the center loss functions was combined as the total output loss, which can be formulated as follows:

$$\begin{aligned} C_{supervised} &= C_{softmax} + \lambda * C_{center} \\ &= -\frac{1}{N}\sum_{i=1}^{N} logP(y_i = t_i|x_i) + \frac{\lambda}{2}\sum_{i=1}^{N}||x_i - c_{t_i}||_2^2, \end{aligned} \tag{1}$$

where $x_i$ is feature vector of the $i$-th input data; $t_i$ is the target label; $y_i$ is the recognition result, and $c_{t_i}$ is the feature vector

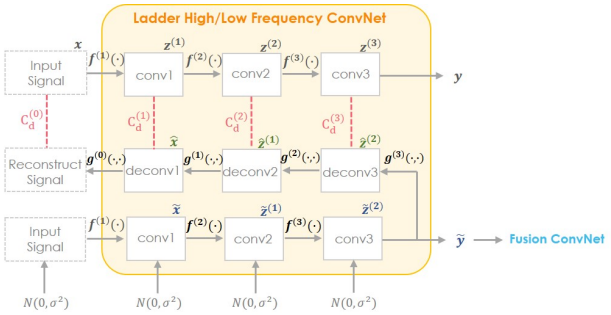


Fig. 4. The frequency ConvNet combined with the Ladder Network.

center of the category $t_i$. $\lambda$ is defined as a parameter to control the strength of the center loss function. In this work, $\lambda$ is set at 0.003. It should be noted that the dropout strategy as well as batch normalization strategy were carried out to facilitate the training procedure.

### F. Semi-Supervised Scheme

To build a powerful recognition model, enormous labeled data can definitely benefit the training procedure for supervised learning. Since there is no public ORSs dataset recorded by multiple IMU sensors, a semi-supervised strategy was employed to improve the recognition performance of the proposed recognition model by only using a few labeled data. Therefore, the idea of the Ladder Network proposed by [18] was further engaged in the frequency ConvNet of the proposed recognition model. The proposed frequency ConvNet with semi-supervised scheme was illustrated in Fig. 4. Unlabeled data were added Gaussian noise and encoded in each convolution layer, and then reconstruct each layers from the latest layers to the first layer. The reconstruction layer $\hat{z}^{(l)}$ of the decoder was derived by the latest layer $\hat{z}^{(l+1)}$ and the its corresponding layer $\tilde{z}^{(l)}$ in the noisy encoder. The loss was calculated layer by layer by measuring the distance between the feature vector in each reconstructed layer $\hat{z}^{(l)}$ of the decoder and the layer $z^{(l)}$ of the clean encoder for unlabeled data, which can be formulated as follows:

$$\begin{aligned} C_{semi} = & \quad C_{supervised} + \sum_{l=0}^{3} \lambda_l C_d^{(l)} \\ = & \quad C_{supervised} + \frac{1}{M} \sum_{i=1}^{M} \sum_{l=0}^{3} \lambda_l ||\hat{z}_i^{(l)} - z_i^{(l)}||_2^2, \end{aligned} \tag{2}$$

where $C_{supervised}$ is the cost described in Eq.1, and $M$ is the number of unlabeled data. In this work, $\lambda_0$, $\lambda_1$, $\lambda_2$, and $\lambda_3$ were set as 5, 0.05, 0.005, and 0.0005, respectively. In this way, it can generate more useful features to distinguish gestures.

## IV. Design of Decision-Making Application

In this section, how to employ the proposed gesture recognition model to develop a basketball referee decision-making training system is revealed. The sliding step was set as 5 samples, so that the system could predict results 12 times in a second. Since the proposed recognition model is a light-weight model, it can efficiently execute in both GPU and CPU mode. To be more precise, the computation time of the proposed method was investigated when executed in the CPU and GPU, it spent 0.012 and 0.003 seconds, on average. An algorithm was proposed and applied to update the current gesture state from neighboring prediction results, where the current state of a gesture would change to the next gesture only if neighboring predictions demonstrated adequate confidence. More details about the updating algorithm are described in Algorithm 1. The probability threshold $T$, which controlled the stability of proposed system, and was set at 0.6.

**Algorithm 1** The algorithm about real-time gesture updating.

**Input:** The previous gesture $G_{t-1}$; The last three recognition results $R_t$, $R_{t-1}$, $R_{t-2}$ and there probabilities $P_t$, $P_{t-1}$, $P_{t-2}$; The probability threshold $T$.

**Output:** Current gesture $G_t$.

```
if all of recognition results R are same then
    if minimum value of probabilities P > T then
        G_t ← R_t;
    else if mean value of probabilities P > T then
        G_t ← G_{t-1};
    else
        G_t ← "idle gesture";
    end if
else
    G_t ← "idle gesture";
end if
    return G_t;
```

## V. Performance Evaluation

In this section, the performance of the proposed gesture recognition model is evaluated. Several experiments, such as model component evaluation, different dataset evaluation, and different methodologies evaluation, were conducted to prove the proposed gesture recognition model can achieve satisfied performance.

### A. Dataset and Evaluation Metric

In our scenario, the Perception Neural device is utilized to not only record the acceleration and rotation information but also facilitate our gesture recognition procedure. However, to the best of our knowledge, there is no public dataset recorded by the Perception Neural device. Therefore, two ORS datasets using the Perception Neural device with multiple IMU sensors were collected by our ourselves to examine the proposed method. In this work, the instruction of Perception Neuron was directly followed to place the sensors on the middle of each skeleton because of the following reason. When performing a gesture, each body joint controls the movement of its connecting body skeleton. If IMUs are placed on body joints, motion information might not be adequately captured. IMUs can also be put near the finger tips, but the IMUs might fall off easily. In addition, a public dataset called MHealth Dataset [36] was further considered to evaluate the proposed

method. More details about these three datasets are described as follows:

**Labeled ORS Dataset:** Five participants were asked to make the sixty-five types of ORSs, and each type of ORS was performed five times. Note that all the participants are under supervision of the professional referee when performing ORSs. Then, the recorded multiple IMU signals was manually labeled with their corresponding ORS for the supervised learning tasks. The sliding window strategy was applied to set the size as 60 samples as well as the sliding step as 5 samples. In consequence, about 6,000 labeled windows can be obtained for each participant to evaluate performance of the proposed recognition model. It should be noticed that those signals which are not belong to any ORSs was further treated as "Idle" gesture, and the same sliding window strategy was applied to those signals.

**Unlabeled ORS Dataset:** In order to evaluate the semi-supervised scheme, another fourteen participants were recruited to perform the sixty-five types of ORSs. These unlabeled ORS signals were adopted the sliding window strategy to segment the signals into the size of 60 samples with the sliding step of 30 samples. As a result, about 7,400 unlabeled windows can be acquired for each participant.

**MHealth Dataset:** The MHealth Dataset [36] is a public dataset for human activities. Three IMU sensors were placed on the subject's chest, right wrist, and left ankle and used to record the acceleration and gyroscope information. This dataset comprised 10 subjects and 12 types of physical activity.

In the following experiments, Leave-One-Participant-Out Cross Validation (LOPOCV) evaluation metric was chosen for calculating the accuracy, which means all the sliding windows were split into $N$ folds by different subjects to evaluation in turn. In other words, when a certain subject is considered as the testing set, the remaining subjects are considered the training set to train the parameters of the proposed recognition model.

### *B. Performance of Model Ablation*

**Pre-processing:** The frequency domain signals and the time domain signals were taken as inputs to observe which one could achieve better performance. Note that there is no weight sharing strategy was applied in this experiment to evaluate their efficiency. As can be seen in Fig. 5, the blue curve indicates that the accuracy of frequency-based input data is better in different training iterations.

**Fragment Numbers:** Each sliding window was separated into different numbers of fragments (i.e., three fragments, four fragments and five fragments) to conduct the evaluation. Fourier Transform was employed to gain representative patterns of each fragment. Fig. 6 shows the results using different numbers of fragments, where it can be observes that there are no significant differences among the three different numbers of fragments. Therefore, in this work, a sliding window was split into four fragments in order to facilitate the generation of the sliding windows.

**Partial Weight Sharing Strategies:** Four kinds of partial weight sharing strategies were designed for the proposed ConvNet in the recognition model, which are traditional partial

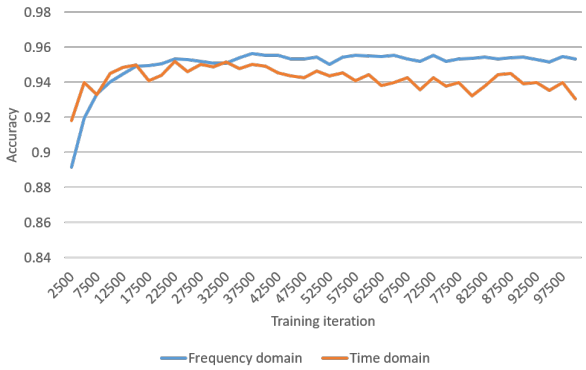


Fig. 5. The accuracy curves for time domain inputs and the frequency domain inputs.

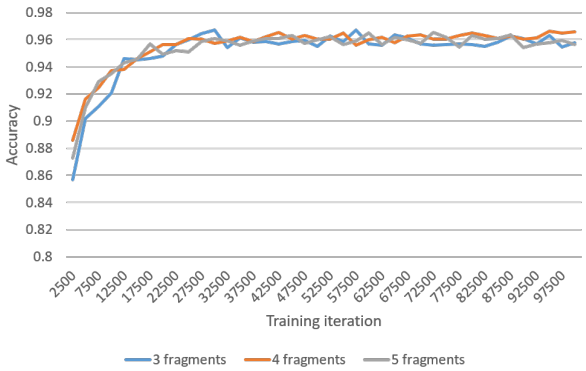


Fig. 6. The accuracy curves for different number of fragments.

weight sharing, frequency domain partial weight sharing, modality domain partial weight sharing, and fusion partial weight sharing. The performance for these four kinds of partial weight sharing strategies are shown in Table I, where it can be seen that the frequency domain partial sharing strategy can facilitate the proposed recognition model to gain more representative features among distinguish hand gestures. In addition, it can also prove our idea of separate the input signals into the high frequency ConvNet and the low frequency ConvNet is convincing.

**TemporalNet:** Several different configurations were conducted for the proposed TemporalNet. The experiment shows a single GRU can lead to the best results, as shown in Table II.

**Loss Functions:** The efficiency of the center loss function was investigated to prove that when combing both the softmax loss function and the center loss function can achieve superior performance. Fig. 7 shows the experiment results of two accuracy curves. We can observe that the accuracy curves with both the softmax loss function and the center loss function is better in different training iterations rather than only using softmax loss function.

### *C. Performance Comparison with Existing Works*

In this section, the performance of the proposed method was compared with existing methods using the Labeled ORS

TABLE I
THE PERFORMANCE FOR DIFFERENT PARTIAL WEIGHT SHARING STRATEGIES FOR CONVNETS.

| Weight Sharing Strategy | Accuracy(%) | Variance |
|---|---|---|
| Traditional | 95.7 | 4.0 |
| Frequency domain | **96.6** | 2.0 |
| Modality domain | 95.0 | 7.1 |
| Fusion domain | 95.8 | 2.3 |

TABLE II
THE PERFORMANCE FOR DIFFERENT TEMPORALNET CONFIGURATIONS.

| Recurrent layer | Accuracy(%) | Variance |
|---|---|---|
| Single LSTM | 95.8 | 2.4 |
| Single GRU | **96.6** | 2.0 |
| Two Stacked GRU | 95.3 | 3.4 |

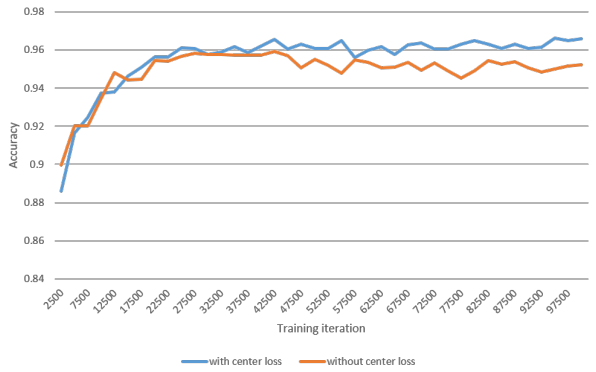


Fig. 7. The accuracy curves for two different loss functions w/o the center loss function.

Dataset and the public MHealth Dataset, respectively. The experiment results evaluated in the Labeled ORS Dataset are shown in Table III, and it can be seen that the proposed recognition model outperformed the existing methods. The accuracy curves of the proposed model and existing methods were further illustrated in different numbers of training iterations. It can be shown that the proposed model is robust enough to achieve better accuracy, where it is difficult to overfit. On the other hand, to prove the proposed method can be utilized in other sensor datasets, the proposed method was further evaluated in the MHealth Dataset, which is a public dataset for human activities. Table IV shows that the proposed model outperformed other existing works, and the proposed model is flexible for other sensor dataset.

### *D. Performance of Semi-Supervised Scheme*

In order to evaluate the proposed semi-supervised scheme, the Unlabeled ORS Dataset was utilized to conduct the following experiment. In this experiment, Subject 1 was set as the labeled data and the other subjects (Subject 2 to Subject 14) as the unlabeled data. The accuracy was obtained on the basis of a 13-fold LOPOCV evaluation metric. For example, when unlabeled Subject 2 was tested to measure accuracy, labeled Subject 1 and unlabeled Subject 3 to 14 were training data into the proposed semi-supervised scheme. Fig. 9 provides

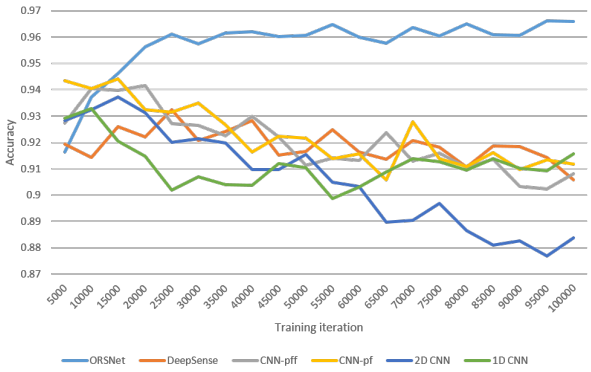


Fig. 8. The accuracy curves comparing existing works when there are increases in the training iterations.

TABLE III
THE ACCURACIES (%) OF THE PROPOSED MODEL AND EXISTING WORKS EVALUATED USING LOPOCV IN THE LABELED ORS DATASET.

| Method | Sub. 1 | Sub. 2 | Sub. 3 | Sub. 4 | Sub. 5 | Avg. | Var. |
|---|---|---|---|---|---|---|---|
| 1D CNN [37] | 90.6 | 93.6 | 92.3 | 93.7 | **96.3** | 93.3 | 3.5 |
| 2D CNN [38] | 95.5 | 93.9 | 92.4 | 95.5 | 91.3 | 93.7 | 2.8 |
| CNN-pf [26] | 96.4 | 93.0 | 91.9 | 95.0 | 95.7 | 94.4 | 2.9 |
| CNN-pff [26] | 95.2 | 93.5 | 92.0 | 96.3 | 93.9 | 94.2 | 2.2 |
| DeepSense [39] | 87.6 | 94.6 | 93.5 | 97.0 | 93.5 | 93.2 | 9.6 |
| ORSNet | **97.8** | **97.0** | **94.8** | **98.4** | 95.2 | **96.6** | 2.0 |

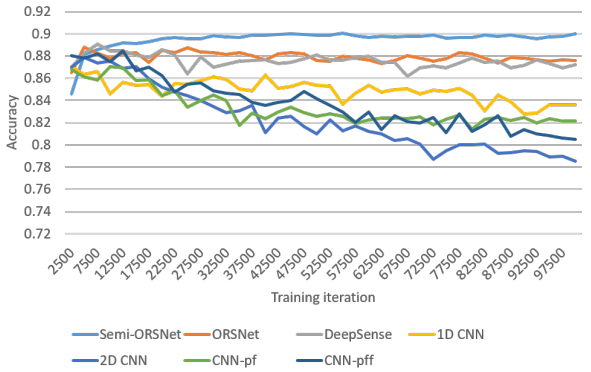


Fig. 9. The accuracy curves comparing existing works and the proposed semi-supervised scheme using limited labeled data.

a comparison of semi-supervised scheme, where it can be seen that the performance improves when unlabeled signals were added as training data into the proposed semi-supervised scheme, which can also reduce the cost incurred when labeling huge amounts of data.

## VI. PERFORMANCE EVALUATION ON THE DECISION-MAKING SYSTEM

### *A. Evaluation Pipeline*

In order to evaluate the usefulness of the proposed recognition model applied on a basketball referee decision-making training application, an interactive video-based decision-making training pipeline was proposed, where a multisensor-based IMU sensors as used to record hand trajectory information. Twenty participants (fifteen males and five females,

TABLE IV
THE ACCURACIES (%) OF THE PROPOSED MODEL AND EXISTING WORKS IN THE MHEALTH DATASET.

| Method | Sub. 1 | Sub. 2 | Sub. 3 | Sub. 4 | Sub. 5 | Sub. 6 | Sub. 7 | Sub. 8 | Sub. 9 | Sub. 10 | Avg. | Var. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HMM | 61.30 | 63.58 | 68.24 | 71.92 | 59.28 | 70.40 | 69.07 | 70.02 | 64.15 | 67.27 | 66.62 | 16.04 |
| SVM | 58.79 | 54.50 | 77.37 | 76.06 | 76.05 | 63.92 | 54.31 | 50.64 | 76.22 | 66.18 | 65.40 | 99.36 |
| HCRF | 62.86 | 74.04 | 81.07 | 67.10 | 73.36 | 70.83 | 54.22 | 49.91 | 81.99 | 70.61 | 68.60 | 98.74 |
| 1D CNN | 81.99 | 86.29 | 86.83 | 85.01 | 85.98 | 88.92 | **99.29** | 95.29 | 98.85 | 95.75 | 90.43 | 35.41 |
| 2D CNN | 84.68 | 80.81 | 83.99 | 84.07 | 87.98 | 87.97 | 89.87 | 96.34 | **100** | 95.66 | 89.14 | 35.70 |
| CNN-pf | 87.19 | 87.66 | 84.24 | 84.08 | 96.68 | 90.44 | 89.69 | 97.81 | **100** | 95.48 | 91.33 | 29.99 |
| CNN-pff | 85.45 | 87.06 | 83.48 | 82.95 | 97.04 | 89.77 | 98.76 | **99.27** | **100** | 95.66 | 91.94 | 42.81 |
| DeepSense | 93.16 | **92.03** | **90.79** | **91.90** | 98.48 | 85.13 | 98.75 | 89.66 | 99.29 | 97.56 | 93.68 | 19.90 |
| ORSNet | **96.02** | 88.00 | 87.69 | 91.56 | **98.65** | **94.41** | 90.54 | 96.52 | 99.91 | **99.28** | **94.26** | 18.81 |

aged from 19 to 24) were recruited from a local university to validate this application. Since our goal was to cultivate beginners to understand how to perform ORSs and its corresponding performing opportunity, none of these participants had referee experience. However, to ensure participants have a sense of fouls or violations in basketball games, they were asked to have basketball team experience for at least six months in their department of a local university. Individual differences in the frequency of playing basketball every week is Mean = 2.15 and SD = 1.39, while the years of playing basketball is Mean = 3.25 and SD = 2.40. 10 basketball game scenarios were designed and recorded that basketball referees have to perform a sequence of ORSs immediate to conduct the experiment.

Before the subjective experiment, was told the goal of the experiment is to evaluate the effectiveness of the training system. Then, a pre-test of basketball ORSs was carried out, where the participant was asked to watch 10 scenario videos on an LCD and to wear an IMU suit while performing ORSs in the pre-test. Note that these 10 scenarios are the most common scenarios in formal basketball games suggested by a professional basketball referee. The correctness of the performing results was recorded to compare with the results of the post-test. Fig. 10 shows an example of one participant using the proposed training system. After the pre-test, three training tasks were executed to cultivate their decision-making skills. The first training task was to follow the ORS performance videos as well as their corresponding example figures in the FIBA textbook and to correctly perform the procedure 5 times. In the second training task, the participant was asked to observe the name of the ORS and to correctly perform its corresponding hand movement. In the third training task, the 10 scenario videos were played (which were not the same as the video in the pre-test). The videos were paused to allow the participant to determine the required movement while looking at a sequence of the ORS example figures. The video was played again until the participants were able to perform the correct hand movement. To ensure all the participants could effectively enhance the decision-making skills, they were permitted to repeat the third training task until they could smoothly make a judgement. The next procedure is that the participant was asked to complete the post-test, in which the scenario videos were the same as those in the pre-test. At the end of the experiment, each participant was asked to fill out evaluation questionnaires. The number of correct ORSs in both the pre-test and the post-test were recorded

Fig. 10. A participant wearing a suit with multiple IMU sensors experiences the proposed basketball referee decision-making training system.

to evaluate the knowledge improvement of each participant. The filled questionnaires, the System Usability Scale (SUS) questionnaire and the Technology Acceptance Model (TAM) questionnaire (cf. Appendix), were utilized to analyze the subjective evaluation performance.

### B. Performance of Knowledge Improvement

In the 10 scenario videos, there were 38 ORSs that the participants have to perform. If the participant performed the correct gestures, the participant can obtain score of 1. In the subjective experiment, the score of the pre-test was Mean = 5 and SD = 4.75, while the score of the post-test was Mean = 30.8 and SD = 5.49. Therefore, it can be obviously observed that all the participants made significant progress after using the decision-making training system based on the proposed gesture recognition model.

### C. Results of the System Usability Scale (SUS) Questionnaire

The SUS questionnaire [40] was applied to evaluate whether the basketball referee decision-making training system based on the proposed gesture recognition method led to each participant gaining fluency in the required operations. The SUS questionnaire is one of the most widely used standards to understand the feasibility and usability for a system. The SUS questionnaire is scored using a 5-point Likert Scale [41]. To facilitate the analysis, Bangor *et al.* [42] proposed a scale with 5 levels to explore the SUS score (i.e., Level A to Level F), where a higher score indicates better usability. In our experiment, a result of Mean = 83.25 and SD = 7.9

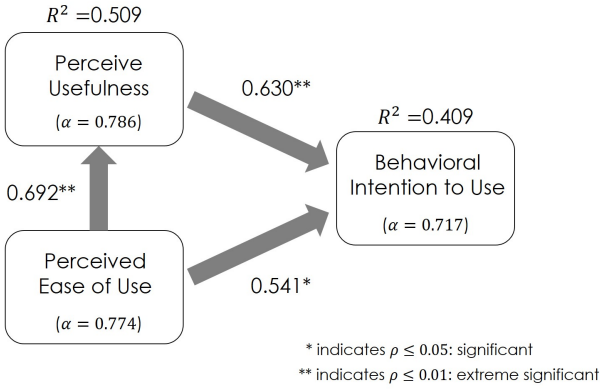


Fig. 11. The TAM correlation model and the regression analysis of the relationships among the measures.

(Level B) can be obtained, which means the proposed gesture recognition model can be successfully utilized to develop a decision-making training system.

### D. Results of the Technology Acceptance Model (TAM) Questionnaire

The TAM questionnaire is a system theory which provide six factors to simulate how a subject accepts and uses a technology. This model is a fairly free questionnaire. The spirit is that when users come into contact with new technologies, there will be many factors affecting how and when they are used. To simplify qualitative study, a preliminary study was conducted to conceive three factors, which is the perceived usefulness, ease of use, and behavioral intentions toward use of the ORS decision-making training system using the 7-point Likert Scale [41]. The perceived usefulness indicated that the proposed gesture recognition method could correctly recognize the ORSs performed by the participants and that the proposed decision-making training system improves their skill. Perceived ease of use was defined as a measure of how intuitive the proposed system was for the participants. Behavioral intentions was utilized to understand the intention of participates to use the proposed system to improve their decision-making skills. The Cronbach's Alpha for each of these measures was 0.786, 0.774, 0.717, respectively. All the Cronbach's Alpha coefficients were considered acceptable. The means and standard deviations of each measure were 6.54$\pm$0.3, 6.51$\pm$ 0.4, and 6.66$\pm$0.28, respectively. Spearman correlation was used to measure the correlation between each of these factors. Moreover, a regression analysis was employed to understand the relationships among them. The results of the regression analysis were shown in Fig. 11, which indicated that the perceive usefulness was a slightly significant predictor of behavioral intentions, and that perceived usefulness can be significantly influenced by the perceived ease of use. Therefore, the result of TAM questionnaire also proved that the basketball referee decision-making training system based on the proposed gesture recognition method encouraged users to practice their skills.

### E. Limitation

The limitation of this work is that wearing and calibrating the IMU suit takes around 5 to 10 minutes, which might be inconvenient for the user. To find the proper placements of IMUs, the suggestion from the instruction of the Perception Neuron device should be followed. In addition, the Perception Neuron device may be affected by nearby electronics, resulting in poor signals for recognition. Therefore, the additional environmental magnetic field should be avoided when using the IMU suit. Since the rule of Institutional Review Board (IRB), the limited numbers of participants can be recruited and demonstrated in the subjective evaluation. However, the result still can prove that the proposed decision-making system can benefit the user's skills.

## VII. Conclusion and Future Work

In this paper, a 3D gesture recognition model was proposed that can accurately recognize the ORSs performed by a user wearing a suit equipped with multiple IMU sensors used to record hand movements. The recorded acceleration and rotation information of hand joints from the IMU sensors could be recognized by the proposed ORSNet model, which involved fequency ConvNet and TemporalNet to extract the representative features and correlations in the time serial signals, respectively. To solve the problem of limited labeled data, a semi-supervised scheme was utilized to facilitate the model training procedure. The experimental results indicated that the proposed recognition model outperforms the existing methods, which encouraged the development of a basketball referee training system. Therefore, a basketball referee decision-making training system based on the proposed gesture recognition system was designed, and the results of a knowledge evaluation as well as subjective evaluations (i.e., SUS and TAM questionnaires) showed that the training system can cultivate the beginners of basketball referees. In the future, inspired by Mokhlespour's work [43], the feasibility of integrating our recognition algorithm with the smart textile system will be investigated to provide more comfortable experience for the users of the proposed referee training system. In addition, researches on visualization of representative features extracted by deep learning models are more and more popular, and we will try to explain the representative features by methods like LIME and GRAD CAM [44] in the future.

## References

[1] C. Stinson and D. A. Bowman, "Feasibility of training athletes for high-pressure situations using virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 4, pp. 606–615, 2014.
[2] L. G. Appelbaum and G. Erickson, "Sports vision training: A review of the state-of-the-art in digital training techniques," *International Review of Sport and Exercise Psychology*, vol. 11, no. 1, pp. 160–189, 2018.
[3] L. Ross2018-Stewart, J. Price, D. Jackson, and C. Hawkins, "A preliminary investigation into the use of an imagery assisted virtual reality intervention in sport," *Journal of Sports Science*, vol. 6, 2018.
[4] "Official basketball rules 2014," https://www.fiba.com/documents.
[5] T.-Y. Pan, L.-Y. Lo, C.-W. Yeh, J.-W. Li, H.-T. Liu, and M.-C. Hu, "Real-time sign language recognition in complex background scene based on a hierarchical clustering classification method," in *Proceedings of the IEEE International Conference on Multimedia Big Data*, 2016.

[6] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *CVPR*, vol. 1, no. 2, 2017, p. 7.

[7] G. Liang, X. Lan, J. Wang, J. Wang, and N. Zheng, "A limb-based graphical model for human pose estimation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 7, pp. 1080–1092, 2018.

[8] L. M. Inc., "Leap motion," 2014. [Online]. Available: https://www.leapmotion.com/

[9] Microsoft, "Kinect," 2010. [Online]. Available: https://developer.microsoft.com/zh-tw/windows/kinect

[10] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: A review," *IEEE transactions on cybernetics*, vol. 43, no. 5, pp. 1318–1334, 2013.

[11] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE sensors journal*, vol. 15, no. 3, pp. 1321–1330, 2014.

[12] M. I. M. Esfahani and M. A. Nussbaum, "A "smart" undershirt for tracking upper body motions: Task classification and angle estimation," *IEEE sensors Journal*, vol. 18, no. 18, pp. 7650–7658, 2018.

[13] J. A. Schrack, R. Cooper, A. Koster, E. J. Shiroma, J. M. Murabito, W. J. Rejeski, L. Ferrucci, and T. B. Harris, "Assessing daily physical activity in older adults: unraveling the complexity of monitors, measures, and methods," *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, vol. 71, no. 8, pp. 1039–1048, 2016.

[14] A. Filippeschi, N. Schmitz, M. Miezal, G. Bleser, E. Ruffaldi, and D. Stricker, "Survey of motion tracking methods based on inertial sensors: A focus on upper limb human motion," *Sensors*, vol. 17, no. 6, p. 1257, 2017.

[15] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga, "A survey of online activity recognition using mobile phones," *Sensors*, vol. 15, no. 1, pp. 2059–2085, 2015.

[16] C.-W. Yeh, T.-Y. Pan, and M.-C. Hu, "A sensor-based official basketball referee signals recognition system using deep belief networks," in *International Conference on Multimedia Modeling*. Springer, 2017, pp. 565–575.

[17] T.-Y. Pan, C.-Y. Chang, W.-L. Tsai, and M.-C. Hu, "Orsnet: A hybrid neural network for official sports referee signal recognition," in *Proceedings of the 1st International Workshop on Multimedia Content Analysis in Sports*. ACM, 2018, pp. 51–58.

[18] A. Rasmus, M. Berglund, M. Honkala, H. Valpola, and T. Raiko, "Semi-supervised learning with ladder networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 3546–3554.

[19] W.-C. Chuang, W.-J. Hwang, T.-M. Tai, D.-R. Huang, and Y.-J. Jhang, "Continuous finger gesture recognition based on flex sensors," *Sensors*, vol. 19, no. 18, p. 3986, 2019.

[20] Z. Zhang, K. Yang, J. Qian, and L. Zhang, "Real-time surface emg pattern recognition for hand gestures based on an artificial neural network," *Sensors*, vol. 19, no. 14, p. 3170, 2019.

[21] S. Saha, A. Konar, A. Saha, A. K. Sadhu, B. Banerjee, and A. K. Nagar, "Eeg based gesture mimicking by an artificial limb using cascade-correlation learning architecture," in *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016, pp. 4680–4687.

[22] P. Muhammad and S. A. Devi, "Hand gesture user interface for smart devices based on mems sensors," *Procedia Computer Science*, vol. 93, pp. 940–946, 2016.

[23] S. Majumder, T. Mondal, and M. J. Deen, "Wearable sensors for remote health monitoring," *Sensors*, vol. 17, no. 1, p. 130, 2017.

[24] J. Wu, L. Sun, and R. Jafari, "A wearable system for recognizing american sign language in real-time using imu and surface emg sensors," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 5, pp. 1281–1290, 2016.

[25] S. Glowinski, A. Blazejewski, and T. Krzyzynski, "Inertial sensors and wavelets analysis as a tool for pathological gait identification," in *Innovations in Biomedical Engineering*. Springer, 2017, pp. 106–114.

[26] S. Ha and S. Choi, "Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors," in *IEEE International Joint Conference on Neural Networks*, 2016, pp. 381–388.

[27] M. Kim, J. Cho, S. Lee, and Y. Jung, "Imu sensor-based hand gesture recognition for human-machine interfaces," *Sensors*, vol. 19, no. 18, p. 3827, 2019.

[28] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019.

[29] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[30] P. Larkin, C. Mesagno, J. Berry, M. Spittle, and J. Harvey, "Video-based training to improve perceptual-cognitive decision-making performance of australian football umpires," *Journal of sports sciences*, vol. 36, no. 3, pp. 239–246, 2018.

[31] G. Schweizer, H. Plessner, and R. Brand, "Establishing standards for basketball elite referees' decisions," *Journal of Applied Sport Psychology*, vol. 25, no. 3, pp. 370–375, 2013.

[32] T. Hohmann, H. Obelöer, N. Schlapkohl, and M. Raab, "Does training with 3d videos improve decision-making in team invasion sports?" *Journal of sports sciences*, vol. 34, no. 8, pp. 746–755, 2016.

[33] T.-Y. Pan, C.-H. Kuo, H.-T. Liu, and M.-C. Hu, "Handwriting trajectory reconstruction using low-cost imu," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 3, pp. 261–270, 2018.

[34] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1933–1941.

[35] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017.

[36] O. Banos, R. Garcia, J. A. Holgado-Terriza, M. Damas, H. Pomares, I. Rojas, A. Saez, and C. Villalonga, "mhealthdroid: a novel framework for agile development of mobile health applications," in *International workshop on ambient assisted living*. Springer, 2014, pp. 91–98.

[37] S. Duffner, S. Berlemont, G. Lefebvre, and C. Garcia, "3d gesture classification with convolutional neural networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014.

[38] S. Ha, J.-M. Yun, and S. Choi, "Multi-modal convolutional neural networks for activity recognition," in *IEEE International Conference on Systems, Man, and Cybernetics*, 2015, pp. 3017–3022.

[39] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, "Deepsense: A unified deep learning framework for time-series mobile sensing data processing," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 351–360.

[40] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[41] F. D. Davis, "User acceptance of information technology: system characteristics, user perceptions and behavioral impacts," *International Journal of Man-machine Studies*, vol. 38, no. 3, pp. 475–487, 1993.

[42] A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: Adding an adjective rating scale," *Journal of usability studies*, vol. 4, no. 3, pp. 114–123, 2009.

[43] M. I. Mokhlespour Esfahani and M. A. Nussbaum, "Preferred placement and usability of a smart textile system vs. inertial measurement units for activity monitoring," *Sensors*, vol. 18, no. 8, p. 2501, 2018.

[44] S. M. Mathews, "Explainable artificial intelligence applications in nlp, biomedical, and malware classification: A literature review," in *Intelligent Computing-Proceedings of the Computing Conference*. Springer, 2019, pp. 1269–1292.


**Tse-Yu Pan** received the B.S. degree in computer science and information engineering from National Taiwan University of Science and Technology, Taipei, Taiwan, in 2013, and the Ph.D. degree from computer science and information engineering at National Cheng Kung University, Tainan, Taiwan, in 2018. He is a postdoctoral fellow at National Tsing Hua University now. His research interests include human-computer interaction, digital signal processing, pattern recognition, computer vision, and multimedia information system. He was awarded the Doctoral Fellowship from Pan Wen Yuan Foundation and Student Travel Grand of Multimedia Conference from ACM SIG MM in 2016 and 2017, respectively. He was awarded the Excellent Reviewer Award from IEEE VCIP 2018. He is a member of Phi Tau Phi.

**Chen-Yuan Chang** received the B.S. degree in computer science and information engineering from National Taipei University of Science and Technology, Taiwan, in 2016 the M.S. degree in computer science and information engineering, Cheng Kung University, Tainan, Taiwan, in 2018. His research interests include digital signal processing and machine learning.

**Wan-Lun Tsai** received the B.S. degree from the Department of Computer Science and Information Engineering, National Cheng Kung University (NCKU), Tainan, Taiwan, in 2016. Since 2016, she has been pursuing the Ph.D. degree in the Multimedia Information System Lab, Department of Computer Science and Information Engineering, National Cheng Kung University. Her research interests include computer graphics, virtual reality, and digital signal processing.

**Min-Chun Hu** is also known as Min-Chun Tien and Ming-Chun Tien. She received the B.S. and M.S. degrees in computer science and information engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2004 and 2006, respectively, and the Ph.D. degree from the Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan, in 2011. She is an Associate Professor in the Department of Computer Science, National Tsing Hua University, Taiwan. From 2012 to 2017, she was an assistant professor in the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan, and she was promoted as an associate professor in 2018. She has published more than 75 papers in international journals and conferences. She was awarded the Exploration Research Award from Pan Wen Yuan Foundation, the Outstanding Youth Award from the Computer Society of the Republic Of China (CSROC), and the Best Young Professional Member Award of IEEE Tainan Section in 2015, 2017, and 2018, respectively. Her research interests include digital signal processing, multimedia content analysis, machine learning, computer vision, computer graphics, virtual reality and augmented reality.